

THE IMPORTANCE OF MATHEMATICS

W. T. Gowers

It is with some disbelief that I stand here and prepare to address this gathering on the subject of the importance of mathematics. For a start, it is an extraordinary honour to be invited to give the keynote address at a millennium meeting in Paris. Secondly, giving a lecture on the significance of mathematics demands wisdom, judgment and maturity, and there are many mathematicians far better endowed than I am with these qualities, including several in this audience. I hope therefore that you will understand that my thoughts are not fully formed: if I had been asked to speak on this subject five years ago, I would have given a completely different lecture, and I am confident that in five years' time it would again have changed.

My title (which I did not actually choose myself, though I willingly agreed to it) also places on me a great burden of responsibility. After all, I am speaking to an audience which contains not just mathematicians but journalists and other influential non-mathematicians. If I fail to convince you that mathematics is important and worthwhile, I will be letting down the mathematical community, and also letting down Mr Clay, whose generosity has made this event possible and is benefiting mathematics in many other ways as well.

Unfortunately, if one surveys in a superficial way the vast activity of mathematicians around the world, it is easy to come away with the impression that mathematics is *not* actually all that important. The percentage of the world's population, or even of the world's university-educated population, who could accurately state a single mathematical theorem proved in the last fifty years, is small, and smaller still if Fermat's last theorem is excluded. If you ask a mathematician to explain what he or she works on, you will usually be met with a sheepish grin and told that it is not possible to do so in a short time. If you ask whether this mysteriously complicated work has practical applications (and we all get asked this from time to time), then there are various typical responses, none of them immediately impressive.

One is the line taken by the famous Cambridge mathematician G. H. Hardy, who was perfectly content, indeed almost proud, that his chosen field, Number Theory, had *no* applications, either then or in the foreseeable future. For him, the main criterion of mathematical worth was beauty. At the other end of the spectrum there are mathemati-

cians who to work in areas such as theoretical computer science, financial mathematics or statistics, areas of acknowledged practical importance. Mathematicians in these areas can point to ideas that have had a big impact, such as the Black-Scholes equation for derivative pricing, which has transformed the operation of the financial markets, and the public-key cryptosystem invented by Rivest, Shamir and Adelman, which is now the basis for security on the internet, and which, as has been pointed out many times, is an application of number theory that Hardy certainly did not expect.

Also at the applied end of the spectrum there are many mathematicians whose work has intimate connections with theoretical physics. Actually, it is not obvious that unifying General Relativity and Quantum Mechanics would have direct *practical* applications, since today's physics already provides us with predictions that are accurate to within the limits we can measure. But one never knows, and in any case such a breakthrough would be of absolutely fundamental interest to science, or indeed anybody with the slightest intellectual curiosity. If mathematicians can make a contribution to this area, then they will at least be able to point to a huge *external* application of mathematics.

Most mathematicians, including me, lie somewhere in the middle of the spectrum, when it comes to our attitude to applications. We would be delighted if we proved a theorem that was found to be useful outside mathematics - but we do not actively seek to do so. Given the choice between an interesting but purely mathematical problem and an uninteresting problem of potential benefit to engineers, computer scientists or physicists, we will opt for the former, though we would certainly feel awkward if *nobody* worked on practical problems.

Actually, this attitude is held even by many of those who work in more practical-seeming areas. If you press such a person, asking for a specific example of an application in business, industry or science of their *own work* as opposed to an application of a result in their general area, you will often, though not invariably, witness an uncomfortable reaction. It turns out that a great deal of the research in even the so-called practical areas is in fact not practical at all. I am not trying to draw attention to some sort of scandal by saying this - as I hope to demonstrate, this phenomenon is a natural and desirable consequence of what it means to view the world mathematically.

The reason for it is that mathematics is a two-stage process. Rather than studying the world directly, mathematicians create so-called *models* of the world, and study them.

This applies even to the simplest mathematics. After the age of four or five we do not study addition by actually combining groups of objects and counting them. Instead we use an abstract mathematical construction, or model, known as the positive integers (that is, the numbers 1,2,3,4,5 and so on). Similarly, we do not do basic geometry by cutting shapes out of paper, partly because it is not necessary and partly because in any case the resulting shapes would *not be* exact squares, triangles or whatever they were supposed to be. Once again, we study a model, a sort of idealized world that contains things that we do not come across in everyday life, such as infinitely thin lines that stretch away to infinity, or absolutely perfect circles, and does not contain untidy, worldly things like hamburgers, chairs or human beings.

If one works in a practical area of mathematics, then there will be two conflicting criteria for what makes a good model. On the one hand, the model should be accurate enough to be useful, and on the other, it should be simple and elegant enough to generate realistic and interesting mathematical problems. It is tempting, as a mathematician, to attach far more importance to the second criterion - mathematical interest and elegance - than to the first - accuracy - even if this means not immediately contributing to the gross national product of one's country.

A good example of this attitude comes from computer science. Consider the network shown in Figure 1, and imagine that we have been asked to colour the nodes of the network with two colours, red and blue, in such a way that no two nodes of the same colour are ever linked. Such a colouring is called a *proper* colouring of the network.

If we start with a single node, such as the one marked A, then it doesn't matter whether we colour it red or blue, since the roles of the two colours are interchangeable. However, once we have decided that A should be red, say, then the two nodes marked B have to be blue, since they are linked to A and therefore must not have the same colour as A. Having established this, we then see that the nodes marked C must all be coloured red (since they are linked to blue nodes), and then that all nodes marked D must be coloured blue. But now we have hit a problem, which is that two of the nodes marked D are linked to one another. Since all our choices of colour were forced from the moment we coloured the node A, it follows that it is impossible to give a proper colouring of the nodes with only two colours.

Now this argument does more than merely establish that one particular network can-

not be properly coloured. We have used a general procedure, or *algorithm*, as it is called in mathematics and computer science, for testing whether *any* given network can be properly coloured with two colours. Briefly, this procedure can be described as follows: colour one node arbitrarily and then continue by colouring nodes whenever the choice of colour is forced. If you are eventually forced to give two linked nodes the same colour then the network cannot be properly coloured, and if you are not then it can.

This procedure is sufficiently well determined that a computer can easily be programmed to carry it out. Obviously, the larger the network, the longer the procedure takes, and hence the longer the computer will take to run the program. A careful analysis shows that if the network has n nodes, then the number of steps needed by the computer is proportional to n^2 . (It may seem to be only linear, but this is because for a small network represented visually one can see at a glance where the neighbours of any given node are. For a large network encoded as a string of bits this is no longer the case.) To give some idea of what this means, if the network has 100 nodes, then the number of computer steps will be around 10,000, and if it has 1,000 nodes then the number of steps rises to around 1,000,000.

Now let us modify our problem slightly. Figure 2 shows a new network. It can be shown quite easily that this network cannot be properly coloured with two colours (for example, consider the triangles towards the left of the network), but what happens if we allow ourselves three colours? Is there a proper colouring?

This question turns out to be much harder, in general. The reason is that if one tries to colour the network by colouring nodes in turn, then many of the choices one makes are not forced in the way that they were with two colours. For example, if one wishes to colour a node which is linked to two other nodes both of which have been coloured red, one can do so with either green or blue. Then, if one runs into difficulty later, it may be that this difficulty was an indirect consequence of this bad decision made much earlier. To make things worse, there may have been several other decisions, of which some were bad and some good, with no easy way to tell which was which. As a result, it is very hard to establish conclusively that a proper colouring with three colours does not exist, and if it does exist it can be very hard to find.

One method we could try is simply to examine all possible choices of colours and see if one of them works. Of course, this would be very tedious, but isn't tedious repetitive

calculation what computers are good at? The answer is yes, but only if the number of repetitions is not too large. For this problem, the amount of time needed by the computer would be prohibitively large. If a network has n nodes, then the number of ways of assigning each node one of three colours is 3^n , and for each assignment of the colours it takes the computer about n^2 steps (at worst) to check whether there are two linked nodes of the same colour. Hence, the number of steps needed by the computer is something like $3^n \cdot n^2$. Again, one can understand what this means by considering a specific value of n such as 100. For a network with 100 nodes, the number of steps needed is $3^{100} \cdot 100^2$, or

$$5153775207320113310364611297656212727021075220010000.$$

It would take all the world's computers combined far longer than the universe has existed to perform this number of steps.

It follows that a second procedure I have just outlined, namely check all possible colourings and see if one of them is a proper colouring, is impractical in the extreme. As it happens, there is *no* known practical way of determining whether a general network can be properly coloured with three colours. However, you may be interested to know that the network in Figure 2 can be: see Figure 3.

Obviously, if one is programming a computer, it is worth knowing whether one's program is likely to run in a reasonably short time. Hence, it is important in theoretical computer science to establish what one means by a practical algorithm, or procedure. The most widely used convention is that an algorithm is regarded as efficient if the number of steps is no worse than a polynomial function of the size of the input. For example, if somebody were to find a way of determining whether a network with n nodes could be properly coloured with three colours using no more than $100n^8 + 73n^6 + 12n^3 + n + 1$ computational steps, then this would be regarded as the first practical solution of the problem and a breakthrough of the first importance.

However, a computer programmer could make only limited use of such a breakthrough. For example, when $n = 100$,

$$100n^8 + 73n^6 + 12n^3 + n + 1 = 1000073000012000101$$

which is still far too many steps for a computer. So the practical solution would only be 'practical in theory' as opposed to 'practical in practice'.

The point I wish to make is that theoretical computer scientists have a notion of practicality of algorithms which is different from genuine practicality, though related to it. The reason they use this notion is that for mathematical reasons it is natural, elegant and convenient. From a mathematical point of view, this definition of practicality leads to questions which are often more interesting than those that arise from the genuine needs of computer programmers. Thus, even in a very practical subject, practical problems are not always of the highest priority. (I should add that theoretical computer scientists are well aware of what I am saying, and are not indifferent to questions of genuine practicality.)

To summarize what I have said so far, most mathematicians, including those who work in useful-sounding branches of mathematics, do not work on problems with direct practical applications. It would be dishonest of me to argue for the importance of mathematics by trying to pretend that this was not so. Instead, my task will be to explain why, despite this fact, mathematics is a worthwhile endeavour, and why it should be supported. I will give two arguments, the first based on the practical utility of mathematics (despite what I have just said) and the second on its cultural value.

It may look as though I have been trying to convince you that mathematics is a useless subject, but in fact all I have claimed is that a typical *mathematician* does not actively try to be useful. These are two very different statements. They are different because there is an important distinction between the collective result of an activity and the individual motives of the participants. Let me give an example of this from outside mathematics. Some capitalist economies are based on the premise that individual greed and selfishness, to use somewhat emotive terms, can act for the collective good of society. The greed causes people to strive to become wealthy, and this benefits the entire economy in many ways, such as increasing the tax revenue for the government, which can then be spent on hospitals, schools, public transport and so on, or causing companies to be set up, which provide livelihoods for many people. The *individuals* need have absolutely no interest in whether other members of society have satisfactory lives, provided that sufficient social order is maintained, but in an indirect way their activity does benefit others.

Of course, not everybody agrees that economies such as I have briefly described are a good thing, and the last thing I wish to do is let politics intrude on a mathematical lecture. However, it is surely not controversial to state that individual selfishness *can* lead to public good, and I wish to argue that something similar happens in mathematics. Although

individual mathematicians are motivated primarily by a subtle mixture of ambition and intellectual curiosity, and *not* by a wish to benefit society, nevertheless, mathematics as a whole *does* benefit society. I would now like to examine in more detail why this is.

One straightforward answer is this: mathematics is cheap, and occasionally produces breakthroughs of enormous economic benefit, either directly, as in the case of public-key cryptography, or indirectly, as a result of providing the necessary theoretical underpinning for science. If you were to work out what mathematical research has cost the world in the last 100 years, and then work out what the world has gained, in crude economic terms, then you would discover that the world has received an extraordinary return on a very small investment. And I haven't even mentioned the fact that those who engage in mathematical research also teach very bright students, many of whom do not themselves become mathematicians, but rather use their mathematical training in ways that directly contribute to the world economy.

Taken as a whole, then, mathematics is undeniably important. However, a cost-cutting finance minister will notice a gap in the above argument: might it not be possible to achieve the same benefits more cheaply? If the benefits of mathematics come from teaching and a few breakthroughs, while most mathematicians get on with their interesting but useless research, then why not cut the research funding to the useless areas and just support the teaching and the more practically oriented mathematics? One of my main objectives today is to expose the fallacy, or rather fallacies, that would lie behind such a proposal.

The first one is the idea that it is possible to identify the areas of mathematics that will turn out to be useful. In fact, it is notoriously hard to predict this, and the history of mathematics is littered with examples of areas of research that were initially pursued for their own sake and later turned out to have a completely unexpected importance. I could mention the RSA algorithm yet again. A more fundamental example is the non-Euclidean geometry of Gauss, Bolyai and Lobachevsky, which is internally consistent despite such apparently paradoxical phenomena as the existence of triangles with angles not adding to 180° . This paved the way for Riemannian geometry, which seemed to be an example of pure mathematics *par excellence* until it turned out to be exactly what Einstein needed for his general theory of relativity.

A recent and celebrated example is provided by the theory of knots. Figure 3 shows seven examples of what mathematicians call knots. These differ from ordinary knots in

that the ends of the knotted string are fused together - perhaps ‘knotted loops’ would be a more accurate term. A particularly simple knot is shown at the bottom. This is known as the unknot, because, as one can easily see, it can be untwisted into a simple loop. The other knots are more genuinely knotted, but this is surprisingly hard to prove, and to this day there is no known practical method (that is, algorithm again) of deciding whether or not a more complicated diagram of the kind appearing in Figure 3 represents a knot that can be untied into a simple loop. Another central problem in knot theory is to decide when two diagrams in fact represent the same knot, in the sense that one can be twisted into the other. For example, though it is not obvious from the diagram, a good mental gymnast can eventually see that the fourth and fifth diagrams (reading from the top, and from left to right) represent the same knot.

Once again, these looked like amusing puzzles until, after work of Vaughan Jones and Edward Witten, it was realized that knot theory had fundamental connections with theoretical physics.

So - mathematicians can tell their governments - if you cut funding to pure mathematical research, you run the risk of losing out on *unexpected* benefits, which historically have been by far the most important.

However, the miserly finance minister need not be convinced quite yet. It may be very hard to identify *positively* the areas of mathematics likely to lead to practical benefits, but that does not rule out the possibility of identifying *negatively* the areas that will quite clearly be useless, or at least useless for the next two hundred years. In fact, the finance minister does not even need to be certain that they will be useless. If a large area of mathematics has only a one in ten thousand chance of producing economic benefit in the next fifty years, then perhaps that at least could be cut.

You will not be surprised to hear me say that this policy would still be completely misguided. A major reason, one that has been commented on many times and is implied by the subtitle of this conference, “A Celebration of the Universality of Mathematical Thought”, is that mathematics is very interconnected, far more so than it appears on the surface. The picture in the back of the finance minister’s mind might be something like Figure 4. According to this picture, mathematics is divided into several subdisciplines, of varying degrees of practicality, and it is a simple matter to cut funding to the less practical ones.

A more realistic picture, though still outrageously simplified, is given in Figure 5. (Just for the purposes of comparison, Figure 6 shows Figures 4 and 5 superimposed.) The nodes of Figure 5 represent small areas of mathematical activity and the lines joining them represent interrelationships between those areas. The small areas of activity form clusters where there are more of these interrelationships, and these clusters can perhaps be thought of as subdisciplines. However, the boundaries of these clusters are not precise, and many of the interrelationships are between clusters rather than within them.

In particular, if mathematicians work on difficult practical problems, they do not do so in isolation from the rest of mathematics. Rather, they bring to the problems several tools - mathematical tricks, rules of thumb, theorems known to be useful (in the mathematical sense), and so on. They do not know in advance which of these tools they will use, but they hope that after they have thought hard about a problem they will realize what is needed to solve it. If they are lucky, they can simply apply their existing expertise straightforwardly. More often, they will have to adapt it to some extent. Perhaps it will be helpful if I show another two pictures, this time illustrating what it is like to solve a mathematical problem.

Figure 7 is a naive view: you start at the boundary of what is known, with a definite goal in mind. You then have a succession of brilliantly clever ideas after which the solution pops out. A more realistic view, which I have tried to represent pictorially in Figure 8, takes into account numerous important features of mathematical research such as false starts, promising ideas that lead nowhere or insights that unexpectedly solve a different problem.

Thus, a good way to think about mathematics as a whole is that it is a huge body of knowledge, a bit like an encyclopaedia but with an enormous number of cross-references. This knowledge is stored in books, papers, computers and the brains of thousands of mathematicians round the world. It is not as convenient to look up a piece of mathematics as it is to look up a word in an encyclopaedia, especially as it is not always easy to specify exactly what it is that one wants to look up. Nevertheless, this “encyclopaedia” of mathematics is an incredible resource. And just as, if one were to try to get rid of all the entries in an encyclopaedia, or, to give a different comparison, all the books in a library, that nobody ever looked up, the result would be a greatly impoverished encyclopaedia or library, so, any attempt to purge mathematics of its less useful parts would almost certainly be very damaging to the *more* useful parts as well.

So far I have simply *stated* that mathematics is full of surprising connections. Any mathematician will happily confirm that statement and be able to give examples from his or her experience. Indeed, discovering surprising connections is one of the great joys of the subject. I would like to illustrate the interconnectedness of mathematics with an example in which I played a small part. To do this I must briefly describe a few unsolved mathematical problems.

The first one is simple enough that I can explain it precisely. Consider the following three sequences of numbers:

5	11	17	23	29
	7	19	31	43
11	41	71	101	131

They have two important features in common. First, they go up in regular steps: thus, the first one goes up by 6 each time, the second by 12 and the third by 30. Such a sequence is called an *arithmetic progression*. Secondly, and more interestingly, they all consist only of prime numbers.

If you try to extend one of these sequences in the natural way, then it will no longer consist solely of primes. For example, $29 + 6 = 35$, which is 5×7 . Similarly, $55 = 5 \times 11$ and $161 = 7 \times 23$. In fact, it is relatively straightforward to show that any arithmetic progression, if continued far enough, contains numbers that are not prime. However, this observation still leaves open the following two questions:

Problem 1. *Are there infinitely many arithmetic progressions of length four consisting of prime numbers?*

Problem 2. *Can arithmetic progressions of primes have any (finite) length or is there some upper limit to how long they can be?*

I should say in passing that if you replace ‘four’ by ‘three’ in question (1), then the answer is known to be yes, but the proof is by no means easy. Secondly, the longest known arithmetic progression consisting of prime numbers has length in the early twenties and was found by a computer. Of course, no such example leaves us any the wiser about question (2).

The next problem I wish to describe is more geometrical in character. Figure 9a shows a triangle divided up into eight tall thin triangles. This is done in the most obvious way:

the base of the triangle is divided into eight equal lengths and these form the bases of the small triangles, which all have as their top corner the top corner of the original triangle.

Now imagine that we are allowed to slide the small triangles about horizontally, and even to do so in such a way that they overlap. Suppose that we wish to do this in such a way that the area of the shape formed by the overlapping triangles is as small as possible. One approach to this challenge is the following. First, group the triangles into overlapping pairs, as shown in Figure 9b. Having done that, group the pairs themselves into pairs, obtaining two groups of four overlapping triangles (Figure 9c). Finally, slide these groups so that they overlap and form the tree-like picture in Figure 9d.

It turns out that by a method of this kind, one can make the area of the final figure as small as one likes, provided that one divides the original triangle into enough pieces. However, it is also known that the number of pieces must be large: roughly speaking if you want to end up with an area of one n^{th} of the area of the original triangle, you will have to divide its base into 2^n pieces. (This means that to reduce the area to five percent of what it was to start with, the number of thin triangles will have to be over a million.)

Now look at Figure 10. This illustrates the corresponding three-dimensional problem, so that instead of a triangle, one begins with a pyramid, and instead of dividing the base of the triangle into small lines, one divides the base of the pyramid into small squares. These form tall thin pyramids that point in various directions. Also in Figure 10 I have drawn what it might look like if four of these smaller pieces were moved horizontally until they overlapped.

By a method very similar to the method I described for two dimensions it can be shown that, once again, if you divide the original pyramid into enough pieces, then you can slide these pieces horizontally until they overlap enough for their combined volume to be as small as you like. However, now the question of how many pieces are needed to obtain a specified decrease in the volume is wide open. Let me state this question, which is known as the *Keakeya* problem, more formally.

Problem 3. *Suppose that you wish to divide a pyramid into N tall thin pyramids as described above, then slide the thin pyramids horizontally so that they overlap and form a new shape with volume $1/n$ times that of the original pyramid. Then how large do you need N to be?*

The method that worked in two dimensions can be used to show that around $N = 4^n$ pieces will be enough. However, in three dimensions there is much more flexibility to move the pieces, and it is not at all obvious that there is not some much more efficient method of doing it. For the benefit of those with more mathematical experience, let me say that the question of greatest interest is whether N has a power-like dependence on n , and one would like to know this for every dimension. If N could be bounded above by n^{100} , say, then many important conjectures would be disproved. On the other hand, showing that N is greater than any power of n would be a big step towards proving those conjectures.

Next, I would like to describe a theorem rather than an unsolved problem. Figure 11 shows a sequence of numbers, and I have drawn attention to certain groups of four of them. For example, I have picked out the numbers 8, 14, 29 and 35 in one group, and I have done so for the simple reason that $14 - 8 = 35 - 29$. Similarly, $5 - 3 = 29 - 27$, and so on. Let us call such a group of four numbers a *special quadruple*.

Now forget about special quadruples for a moment and consider the following set of numbers: $\{1, 2, 4, 6, 23, 29\}$. (It is traditional to enclose mathematical collections, or sets as they are known, in curly brackets.) To help us think about this set, let us give it a name, A . Then A has size 6, in the sense that it consists of six numbers. Let us define a new set, denoted $A + A$, be the set of all numbers that you can make by adding two numbers (which are allowed to be the same) from A . A brief calculation shows that

$$A + A = \{2, 3, 4, 5, 6, 7, 8, 10, 13, 24, 25, 27, 29, 30, 31, 33, 35, 46, 52, 58\} .$$

For example, 27 belongs to $A + A$ because it can be written $4 + 23$ and both 4 and 23 belong to A . Similarly, $7 = 1 + 6$ and $58 = 29 + 29$. The size of $A + A$ is 20, which is over three times larger than the size of A . However, if we drop the numbers 23 and 29 from A , we obtain a new set $B = \{1, 2, 4, 6\}$, and then

$$B + B = \{2, 3, 4, 5, 6, 7, 8, 10, 12\} ,$$

which has size 9 and is therefore only just over twice the size of B .

Thus, although $A + A$ was considerably larger than A , we were able to form a new set B , consisting of a reasonable proportion of the numbers in A , in such a way that $B + B$ was not all that much larger than B .

Suppose we now choose a new set of numbers as our set A and that $A + A$ is much larger than A . Will we always be able to pick out a reasonably large set B in such a way that $B + B$ is not too much larger than B ? In general, the answer is no, as can be shown quite easily. However, a theorem of Balog and Szemerédi tells us that the answer is yes if A has a particular property, and this property brings us back to the special quadruples defined earlier. Let me give an imprecise statement of their result.

Theorem. *If A is a set of numbers and if A contains many special quadruples, then A contains a reasonably large set B for which $B + B$ is not too much bigger than B .*

A precise statement of the above theorem would require me to say how my notions of ‘reasonably large’ and ‘not too much bigger than’ depend on my notion of ‘many’. However, the details need not concern us here.

Notice that there is a certain initial plausibility about the theorem. If $B + B$ is to be small, then there must be several numbers that can be written in many different ways as a sum of two numbers in B . However, every time a number m can be written both as $x + y$ and as $z + w$ we then have $x + y = z + w$, and therefore $x - z = w - y$. In other words, we have a special quadruple. The interest in the theorem is that in a certain sense this argument can be reversed: starting with the special quadruples one finds a set B for which $B + B$ is small.

The final problem I wish to discuss is not so much a single problem as an entire area of research, which I shall describe in only the vaguest terms. It has been realized since Newton that huge numbers of physical phenomena can be described by means of what are known as *partial differential equations*. Some examples of these phenomena are the behaviour of water waves (or indeed light waves, sound waves, and many other kinds of waves), the flow of heat, the motion of fluids, the growth of populations, the spread of disease and the wave functions of quantum mechanics.

Partial differential equations can be divided into two kinds: the so called *linear* ones, which are (relatively) easy to analyse and can often be solved completely, and the *non-linear* ones, which are much harder to analyse and can almost never be solved completely. Unfortunately (or fortunately - it depends on your attitude) many extremely interesting and important physical phenomena are best modelled by the non-linear kind. One example is turbulence, which is ubiquitous, but very hard to understand mathematically.

Because of the difficulties in analysing non-linear partial differential equations, mathematicians are forced not to set their sights too high. Instead of searching for a formula that perfectly describes the physical situation, they try to answer more qualitative questions. For example, one might wish to know whether the physical quantity under investigation will become more and more violent and unstable as time goes on, or whether it will gradually die out. At an even more basic level, instead of trying to find a solution, one might be content merely to determine whether a solution exists. Thus, one could describe research in partial differential equations as an attempt to answer the following general problem.

Problem 4. *Given a partial differential equation, does it have a solution, and if so, how does that solution behave?*

I have now described a somewhat miscellaneous collection of four problems and one theorem. Why did I do so? Because they are not so miscellaneous after all. In fact, despite apparently coming from completely different areas of mathematics, they are intimately related to each other. Figure 12 shows the various links in diagrammatic form. As in my earlier, more fanciful diagram I have linked two areas of investigation with a line if there is a close connection between them. Let me describe briefly these links, which are typical of the surprising connections that occur all over mathematics.

In order to investigate arithmetic progressions in the primes, it is very natural and fruitful to think about arithmetic progressions in more general sets of numbers. A famous conjecture of Erdős and Turán from 1936, finally proved by Endre Szemerédi 40 years later, asserts that every reasonably large set of numbers (for the mathematician, this means every set of integers with positive lower density) contains arithmetic progressions of every length. A few years after Szemerédi proved this theorem, Hillel Furstenberg discovered a completely different proof, relating the theorem to ergodic theory, a branch of mathematics with close connections to physics. Much more recently, I found a third proof of Szemerédi's theorem, and for this proof I needed to use the Balog-Szemerédi theorem that I mentioned earlier. However, I had to improve the theorem a little, in the sense that my 'reasonably large' was larger than theirs and my 'not too much bigger' was smaller.

Jean Bourgain then realized that he could adapt my proof of the Balog-Szemerédi theorem and use the adaptation to improve greatly what is known about the Kakeya problem in large dimensions. Subsequently, Nets Katz and Terence Tao found a different

argument that improved Bourgain's result and did not use my work. So in a sense the connection is weakened, but it is not entirely invalidated - that, after all, is one of the ways that mathematics progresses.

The Kakeya problem is closely related to problems in harmonic analysis that have a direct bearing on the behaviour of partial differential equations, as Terence Tao will explain in his lecture tomorrow. And as I have already said, the study of partial differential equations includes a high proportion of physics. Thus we have found a chain from a very purely mathematical question about prime numbers all the way to the heart of physics.

It is amusing to note that there is a route back from the Kakeya problem to the primes. A fascinating discovery of Bourgain is that the Kakeya problem is closely related, via a conjecture of Montgomery, to the distribution of the zeros of the famous Riemann zeta function. Why is the Riemann zeta function so famous? Because it is intimately related to the distribution of the prime numbers. Indeed, it is so closely related that many questions about the prime numbers are actually equivalent to questions about the zeros of the Riemann zeta function. Finally, if one wants to know about the existence of arithmetic progressions in the primes, then information about how the primes are distributed is of course helpful.

I should make it clear that I am *not* saying that research into arithmetic progressions in the primes has had a direct impact, via physics, on our everyday lives, since after two or three links in the chain, the connection becomes somewhat tenuous. The point I wish to make is simply that the way mathematics as a whole progresses is complicated, and often depends on unexpected links between apparently very different areas.

As a final cautionary tale for the hypothetical finance minister, let me also demonstrate that problems that look like nothing more than fun and games can turn out to be of direct practical importance after all. To do this, I shall look at the undoubtedly real-world problem of timetabling. As an example of this problem, let us imagine that seven students are about to take some examinations, and they have each made a different choice of papers. I have numbered the candidates from 1 to 7, labelled the papers A to H, and laid out the choices of the candidates in the table in Figure 13.

How should we schedule the examinations? Well, it is important not to do so in such a way that some candidate is expected to take two papers at once, and it is also important (to avoid any possibility of cheating) that all candidates taking a given paper do so at the

same time. One way of meeting these requirements is simply to have the examinations for all the papers at different times, but it may be expensive to hire an examination hall, and if it is, we can save money by having two papers taken simultaneously, provided that no candidate has signed up for both.

The question that now arises is this. For how few sessions do we need to book the examination hall? Going back to Figure 13, we could note that no candidate is taking both paper A and paper D, so there is nothing to stop the examinations for these two papers being simultaneous. Before we go any further though, let us try to represent the problem in a more visual way. In Figure 14 I have drawn a network, with nodes labelled from A to H representing the papers. I have joined two of these nodes with a line when the corresponding two papers are not allowed to be at the same time. Thus, node A is joined to node C because candidates 1 and 5 are taking both papers A and C. Similarly, node B is joined to node D because of candidate 2, and so on.

Now we obviously cannot get away with hiring the examination hall only twice, since candidate 1 is taking three papers. Can we manage with three sessions? If we can, then we will be able to divide the papers into three sets, and put the first set of papers into session 1, the second into session 2 and the third into session 3. What we need to do then is put a number between 1 and 3 next to each node (telling us in which session the paper will be taken) in such a way that nodes that are joined do not have the same number.

To make this still more visual, we could represent the three sessions not by numbers but by the colours red, blue and green. (For example, if paper C is to be taken in session 2 then we could colour node C blue.) Now what we are trying to do is colour the nodes A to H with three colours in such a way that nodes that are joined never have the same colour. This is, as you will undoubtedly have noticed, exactly the problem that I discussed near the beginning of this lecture. Figure 15 shows that the colouring is possible, and hence that it is only necessary to hire the examination hall three times, one for papers A, D and F, one for papers B, E and H and once for papers C and G. Thus, whereas I earlier showed that seemingly very different problems can be closely connected, this example shows that seemingly very different problems are sometimes precisely the same problem. There are many other examples of this phenomenon throughout mathematics.

I think I have said enough on the economic case for supporting pure mathematical research in more or less its current form, and I would now like to turn to the cultural

case. I, as I have said, am one of those individual mathematicians who do not directly strive to swell the national coffers, so let me spend some time explaining what it is about mathematical research that so appeals to me. Since mathematics is a very broad subject, and tastes and mathematical styles differ widely from one branch of this subject to another, this part of the lecture will inevitably be less universal and more personal than the first. However, I believe my experience is fairly typical.

I hope that in France, of all places, a cultural argument will be taken seriously. My home country, England, was famously described by Napoleon as ‘une nation de boutiquiers’ - a nation of shopkeepers. In England, the word ‘intellectual’ is sometimes regarded as virtually a synonym for ‘pretentious’, whereas in France intellectuals are widely admired, and abstract thought greatly appreciated, or at least so one reads.

The cultural case, in brief, is that knowledge is worth pursuing for its own sake. Just as one of the rewards for *individual* mathematical or other cultural success is a form of immortality, so *entire societies*, ancient Greece being the most obvious example, are remembered for their contributions to knowledge long after their political and economic influence has faded. A society that deliberately turns its back on the pursuit of knowledge is, to put it bluntly, a boring society. [After I gave the lecture, Ilan Vardi suggested Rome as a counterexample to this assertion.]

On the other hand, it would be foolish to suggest that all knowledge, or even all mathematical knowledge, is of equal value. What I shall do now is talk a little about a theorem that I definitely do find worth knowing for its own sake. Perhaps one day it will be useful as well, but for now I present it as an example of *beauty* in mathematics. The idea that a piece of mathematics could have an aesthetic appeal puzzles those who have never experienced it, or perhaps I should say never knowingly experienced it. The theorem I have chosen as my example is very well known in the right mathematical circles, but it is not one of the giant theorems of the twentieth century, and it is completely unknown outside mathematics. This is a deliberate decision on my part, because I want to make the point that mathematical beauty is not confined to one or two spectacular and famous theorems, but lurks everywhere, waiting to surprise and delight us.

Figure 16 is, as you can see, a table of numbers. If you examine it for a few seconds you will see that the numbers on the left of each column are simply the numbers 2,3,4,5 and so on, all the way to 100, but they are written as products of prime numbers. Thus,

for example, the number 63 appears as $3 \times 3 \times 7$ and 64 as $2 \times 2 \times 2 \times 2 \times 2$. It has been known since Euclid that every number can be written in this way, and that for each number there is only one way of doing it (except that you can change the order in which you write the primes, but this is not a genuine difference). Hence, for each number n we can ask, ‘How many prime factors does n have?’ Here, if a prime number is repeated I count it multiply, so for example 64 has six prime factors.

In isolation, this is not a particularly interesting question, since all you can do to answer it is work out the prime factorization of n and count how many primes you used. On the right hand side of the columns of the table I have shown the result of this calculation for every n between 2 and 100, and as can be seen the numbers do not follow any recognizable pattern.

A mathematically unsophisticated response to this situation would be to hope that one day somebody will discover a beautiful formula that gives, as if by magic, the numbers in the right of the columns. However, it is almost certain that no such formula exists, so it is more sensible to set our sights lower. If there is not a clean way to generate these numbers, then we can instead ask less precise questions about how they behave. Notice that this situation is similar to the situation for most non-linear partial differential equations - even if you can’t solve them with a crisp formula, you may still be able to show that a solution exists and that it has certain properties.

What properties should we look for in our strange sequence of numbers? To help answer this question, I have compiled a new table in Figure 17, which shows, for each number between 1 and 6, how many numbers between 2 and 99 have that number of prime factors. In other words, for each number I have shown how many times it appears on the right in the previous table. For example, 5 appears four times, because the only numbers between 2 and 99 with exactly five prime factors are 32,48,72 and 80. Below the new table is a diagram which gives exactly the same information but in the form of a bar chart. You will notice that the height of the bars rises to a peak and then tails away again.

You may have been wondering what sort of less precise question should be asked about the fairly random seeming numbers in Figure 16. I can now give an example. Having seen the bar chart, it is natural to ask (and mathematicians, who are trained as much to ask questions as to provide answers, would do so instinctively) whether the shape of this bar chart is just a fluke, or whether it is a sign of some underlying mathematical fact. To be

more precise still, suppose we were to repeat the experiment but with a far larger number. For instance, we could get a computer to calculate the number of prime factors of every number up to a billion, and provide us with a bar chart similar to the one in Figure 17, telling us how many times each number had occurred. Would the new bar chart also rise to a peak and then tail away? If so, roughly what shape would it have?

There is one shape that bar charts often seem to have, sometimes called the bell curve. Mathematicians call such phenomena *normally distributed*. It seems a little outrageous to suggest it, but might the number of factors of numbers up to n be normally distributed when n gets large. We have something a bit like a bell (though unfortunately much of the left of it has been cut off) in Figure 17, but this is hardly conclusive, or even persuasive, evidence.

Remarkably, however, this guess is correct, as was proved by Erdős and Kac in 1939. I find this result beautiful for at least five reasons. First, the shape of the normal distribution is itself aesthetically satisfying, though this is true of many mathematical curves. Secondly, the theorem is unexpectedly simple. If you are not a mathematician then this may not be immediately apparent, since the bell curve is defined by the formula $y = ae^{-b(x-c)^2}$. However, the normal distribution has all sorts of properties that make it particularly easy to deal with, and it occurs throughout mathematics and science. In a certain precise sense, it is the simplest of all distributions.

A third reason for the theorem being beautiful is that it is unexpected. Behind the disorder and irregular behaviour of the primes there lies the simplicity and regularity of the normal distribution. This is particularly surprising because the primes are defined deterministically (there is no choice about whether a given number is a prime or not) while the normal distribution usually describes very random phenomena.

A fourth reason is that the phenomenon uncovered by Erdős and Kac is not one we can directly experience. If you wanted to produce a bar chart that gave a good approximation to the normal distribution, you would have to calculate the prime factorizations of more numbers than a computer could handle. Thus, the result of Erdős and Kac is, in a sense, purely theoretical. Though we can appreciate the pattern, we have to do this by mathematical thought, rather than by experiment. (We can of course *imagine* an experiment, and, thanks to Erdős and Kac, we know exactly what the result would be.)

The fifth reason is that the *proof* of the theorem is very satisfying. Let me give a very

brief outline of it. For the benefit of those who are afraid of logarithms, when I say $\log n$, then you will not lose much if you interpret this as $2 \cdot 3$ times the number of digits of n . In particular, when n is a large number, $\log n$ is much smaller than n . I shall divide the proof into four steps.

Step 1. When n is large, most numbers near n have roughly $\log \log n$ prime factors. (That is, with a few exceptions, if m is near n then you can approximate the number of prime factors of m by taking its logarithm twice.) This result was proved by Hardy and Ramanujan in 1920, and again in 1934, with a much simpler argument, by Paul Turán.

Step 2. Therefore, most prime factors of most numbers near n are small. This follows because a significant number of large prime numbers would multiply to a number bigger than n .

Step 3. If m is chosen to be a random number near n , then the events ‘ m is divisible by p ’, where p is a small prime, are roughly independent. For example, if you know that m is divisible by 3 and 5, but not by 11, it gives you almost no information about whether m is divisible by 7. By a technique known as the Brun sieve, this means that if we think of the events as being exactly independent, then the conclusions we draw from this will be approximately correct.

Step 4. If these events *were* exactly independent, then a normal distribution would result, because (subject to certain technical conditions that hold here) it always arises when one counts how many of a large number of independent events have occurred.

An important comment about the above proof is that it *works*. The steps may be reasonably convincing (at least to a mathematician) but there are many vague words such as ‘large’, ‘small’, ‘approximately’ and so on. As with the Balog-Szemerédi theorem, one must make these precise by saying exactly how large counts as large, what is regarded as a good approximation and so on. The way the Erdős-Kac theorem was proved is one of the great romantic stories of mathematics: Kac gave a seminar at the Institute for Advanced Study in Princeton in which he suggested that the theorem (as yet unproved) might be true; Erdős, apparently not concentrating, was in fact thinking hard; intimately familiar with the Brun sieve, he rapidly conceived of the above proof outline; at the end of the seminar Erdős went up to Kac and said that he had a proof. There are many other stories

of a similar kind: mathematicians who are at the top of a mountain when the solution to a problem suddenly hits them, and so on. However, if one does have a bright idea, it is important to sit down and check the details thoroughly. If my own experience is anything to go by, the great majority of bright ideas are either wrong, or have been had by hundreds of people before.

I hope that, even if the Erdős-Kac theorem has not just given you an intense aesthetic experience, you can at least see how it might do so for those who spend their lives devoted to mathematics. Clearly, mathematical beauty is not the same as the beauty of a painting, but then neither is the beauty of a painting the same as that of a piece of music, or a tree, or a poem, or a human face. Though I would hesitate to define beauty, there are certain features commonly associated with beauty that definitely occur within mathematics. Some of these are symmetry, balance (for this to be good, excessive symmetry may well be undesirable), and tension between simplicity and complexity, between regularity and irregularity, and between predictability and unpredictability. In general, beauty has to arouse our interest, for which reason it often involves patterns that we can appreciate but not fully understand.

These concepts make sense even in very simple situations. For example, Figure 18 shows four grids of 13×13 squares, each with some squares filled in. Most people, if asked to rank the four patterns in order of beauty (or if that is too strong a word, of aesthetic preference), will have opinions on the matter. For example, the first pattern, though similar to the second, is somehow more satisfying. There is even a mathematical reason for this, which is that the first one represents the eight times table in base 13, and 8 and 13 are consecutive Fibonacci numbers. This means that $13/8$ approximates the famous golden ratio, which (for reasons I do not have time to explain) is directly connected with the fact that the first pattern does not have the ‘ugly’ almost vertical stripes of the second. What is it about those stripes that is ugly? Perhaps it is that we somehow apprehend them too easily, and they force us to look at the pattern in only one way, making it less mysterious than the first.

Of course the first isn’t particularly mysterious itself, and I don’t want to sound too aesthetically excited about any of them. Nevertheless, it is interesting that some sort of aesthetic discussion is possible. If you want to know how the other patterns were generated, the second fills in every 9th square rather than 8th. The third is the graph of $y = x^2$ modulo 13 (and would perhaps look better if I removed zero), while for the fourth I chose each

square entirely randomly with probability $1/10$. (In my opinion, although the resulting pattern does not yield up a deeper meaning, its composition is not too bad.)

I would like to end by returning to the word “importance”, because, interestingly enough, the beauty of some mathematics *contributes* to its importance. This is not just because we want as much beauty in the world as possible - after all, the beauty of higher mathematics is appreciated by only a tiny minority of the world’s population. A more serious reason is that there is a remarkable correlation between mathematics that is beautiful, and mathematics that is important.

This is partly for reasons internal to mathematics. There are two ways in which the subject develops. One is the solution of problems, which involves clever, unexpected ideas and hard, technical work (in varying proportions). Now if mathematics were *nothing but* problem-solving, then as it grew and grew it would become more and more chaotic, specialized and difficult. Indeed, to some extent this does in fact happen. Fortunately, there is a tendency in the opposite direction as well. It often happens that there are similarities between the solutions to problems, or between the structures that are thrown up as part of the solutions. Sometimes, these similarities point to more general phenomena that simultaneously explain several different pieces of mathematics. These more general phenomena can be very difficult to discover, but when they are discovered, they have a very important simplifying and organizing role, and can lead to the solutions of further problems, or raise new and fascinating questions. I have just argued that an important component of aesthetic appreciation is the feeling that a complicated pattern has been generated in a simple way, but not so that one can immediately apprehend it. It should not come as a total surprise, therefore, that when a number of loosely related specific results turn out to be consequences of the same general one, mathematicians should have an aesthetic response.

In fact, an appreciation of beauty is essential, or at least very useful indeed, for solving problems as well. Often, one can reject a line of attack on the grounds that it is simply not elegant enough to work. One may be wrong, but it is still efficient to look for beautiful solutions first, and settle for ugly ones only as a last resort. Similarly, finding the solution of a problem involves a great deal of rather vague reasoning, following hunches, making guesses and so on. How does one know whether one of these guesses will survive a later, more rigorous scrutiny? Well, one doesn’t, as I stressed earlier, but it is a good rule of

thumb that the more beautiful the guess, the more likely it is to survive.

My final illustration comes from a children's book, and I give it as an example of a picture that I find distinctly lacking in beauty. An surprised elephant is squirting water out of its trunk and over a train. I hope you will agree with me that the shape of the jet of water looks wrong. Indeed, it *is* wrong - it should be a parabola, which it clearly isn't, but at a more basic level, if the water comes out of the elephant's trunk almost vertically, then it should continue more or less vertically, rather than miraculously going a short horizontal distance and then dropping vertically again. However, the point I wish to make is not so much that the artist was ignorant of physics as that one's first reaction against the picture is an aesthetic one. It just looks wrong, and unsatisfying, an initial impression that can easily be backed up with elementary physics later.

There are also external reasons for the importance of beauty. The mathematician Eugene Wigner once gave a famous lecture entitled "The Unreasonable Effectiveness of Mathematics" (in fact, so famous that it is something of a cliché in mathematical circles to mention it). He was referring to the fact that the physical world obeys very simple mathematical laws - such as Newton's inverse square law of gravitation - and it is very hard to explain why this should be. Could one not conceive of a world in which science was impossible, because there simply wasn't enough regularity? Mathematics would still be possible in such a world, though whether it would be pursued is another matter because simple and elegant mathematical models would have no physical counterparts.

I don't think this philosophical question has been satisfactorily answered, but we can be grateful that in *our* world it *is* possible to use simple mathematical models. These can describe, or even explain, the great complexities of physics and to a lesser extent the other sciences. Once again, complexity arises from simplicity, and, once again, beauty reveals itself to be important. Thanks to this piece of good fortune, we can be confident that mathematicians, if they are given the freedom to pursue the subject that gives them so much pleasure, will continue to produce a body of work that is important in every sense of the word.