

□

This Lecture Will Surprise You: When Logic is Illogical Transcript

Date: Monday, 19 January 2015 - 6:00PM

Location: Barnard's Inn Hall

□

19 January 2015

This Will Surprise You: When Logical is Illogical

Dr Tony Mann

Before I say anything tonight, I should like to welcome you to my first Gresham College lecture of 2015.

If you found that introduction mildly amusing, that augurs well for this evening's lecture, which is going to explore the role of paradox in mathematics. Paradoxes and self-referential constructions such as the self-defeating statement I just made are, for many of us, amusing. What I hope to show is that there is more to it than that: that the kind of logical paradox which has entertained us for millennia turns out to have serious, important and valuable implications for mathematics and for computing.

This is the first of three lectures I am giving this term which present different angles on paradoxes. This evening I will be looking at paradoxes in logic. Next month I will be talking about a paradox which has fascinated me since I was first introduced to it almost 40 years ago, and which in recent years, with the help of computer simulation, has unexpectedly turned out to provide valuable insights into human behaviour and important issues of trust and co-operation. In March I will be telling you about a more recently discovered paradox, an astonishing example which shows that, if one has two games both of which one expects to lose, they can sometimes be combined into one that you expect to win. This too-good-to-be-true scenario doesn't seem to allow you to break the bank at Monte Carlo, but it does have applications in chemistry and genetics. One of the aspects which I find interesting is that while mathematicians were astonished by this paradox, the physicist who invented it, Juan Parrondo, didn't consider it at all surprising!

Anyway, back to tonight and to logical paradoxes. I want to remind you of my title - "This lecture will surprise you", and the claim I made in the Gresham College programme, "I guarantee that you will be surprised". I'd like you to keep this claim of mine in mind during the talk.

The Chinese philosopher Zhuang Zhou once dreamt that he was a butterfly, and thereafter was never sure whether he was not a butterfly dreaming that he was Zhuang Zhou. How does one ever know that one is not dreaming? One of my greatest rhetorical triumphs came when I persuaded a roomful of philosophers that I could be absolutely certain that, as I was talking to them, I was not dreaming. Unfortunately, when I woke up I found that I had forgotten my argument. (This joke came from the logician Raymond Smullyan, whose mind-twisting books such as *What is the name of this book?* have entertained and inspired me for many years and which are the source of several of my examples tonight.)

A paradox is, according to Wikipedia, "a statement that apparently contradicts itself and yet might be true". Since logic proceeds by avoiding contradictions, a paradox is a serious challenge to the logic on which mathematics is based. Many mathematical arguments - those involving "proof by contradiction" are based on the premise that a contradiction cannot arise so if an assumption leads to a contradiction, that assumption cannot be valid.

For example, suppose I want to prove that if n is an integer such that n^2 is an odd number, then n itself must be odd. I prove this by assuming I have a contrary example and aim to derive a contradiction from that assumption. So I assume that n is an even integer with this property, so that n is $2k$ for some integer k . Then n^2 is an odd number by assumption, but $n^2 = (2k)^2 = 4k^2$ is divisible by two, so n^2 is clearly an even number. So n^2 is both odd and even, which is a contradiction, so our initial assumption that n was an even number cannot be valid, we have shown that n cannot be even and therefore n must be odd.

A paradox is a challenge to this kind of proof, because our argument depends on the fact that a contradiction must be impossible. You might think that in an ideal world there would be no paradoxes. But in that case medical care would be rather limited, because in such an ideal world there can only be one

doctor! Why? Because if there are two doctors then we have a pair o' docs - and we said there can be no paradox in our ideal world!

We need not dwell on that example (thankfully). But there are other cases which are less easy to dispose of. Raymond Smullyan tells the story that when he was a student he needed extra income to support his studies and he applied for a job as a vacuum-cleaner salesman (a rather dubious employment opportunity then available to students). During the interview he was asked if he would ever be prepared to tell a lie. Now, he was a very honest young man and there were no circumstances in which he would be prepared to lie. But he really needed the job, and he knew that if he said "No", then he wouldn't get it, so he replied "Yes". Was he lying? If he was lying, then clearly there was a situation in which he would lie - this interview situation - and so he was prepared to lie and his statement was true. On the other hand, if he was telling the truth, then he was never prepared to lie, so his answer was a lie. We can go round this loopy argument for ever! An "Alex" cartoon by Charles Peattie and Russell Taylor from 1990 has a variation on this theme: the banker Clive is asked in an interview if he would be prepared to lie in the interest of seeing through a business deal. He answers truthfully "yes", he would lie, revealing that he is too honest for the job - the correct answer would have been a dishonest "absolutely not"!

Such paradoxes go back a long time. You are probably familiar with the paradoxical statement "This sentence is false". If it is true, then since it asserts truthfully that it is false, then it must be false. If on the other hand it is false, then since it claims falsely that it is false, it must be true. In either case we have a contradiction.

This is a version of the Cretan paradox - someone who is themselves a Cretan says "All Cretans are liars". Indeed, it is found in this form in the Bible, in the New Testament Book of Titus, Chapter I verse 12:

One of themselves, even a prophet of their own, said, The Cretians are always liars, evil beasts, slow bellies.

A more recent example of this came when the confidence trickster and later FBI informant Mel Weinberg (the inspiration for the film *American Hustle*) was testifying under oath. He admitted that he had once described himself as "the world's biggest liar", but said that he had been lying at the time.

So does this paradox have any value, any significance, or is it just something we can write off as a harmless joke, noting perhaps that self-referential statements can be dangerous in this way? Of course, not all self-referential statements are problematic. If I say "This sentence is true", my claim may be true, or it may be false, but in neither case is there a paradox.

In a similar way the first sentence of my talk tonight - "Before I say anything ..." is self-defeating. Was it possible for me to say that? We'll find out when the Gresham College staff edit the recording and put this lecture online.

Not all such statements are quite so self-defeating. Recently I had to find my passport. Very annoyingly, it was in the last place I looked. But that was only to be expected, because as soon as I found it, I stopped looking!

This effect can be misleading. A possible survey technique is to ask people about their most recent experience of whatever one is interested in. If I want to know what proportion of trains run to time, it might make sense to select a random sample of people and ask whether the last train they caught was late.

I might try to conduct a similar survey about, say, golf. In fact, I have checked the data on the golfers who competed in last year's British Open Championship. If I consider 136 golfers who competed at Hoylake and look at the last shot that they played in the tournament, it seems that 136 of them holed that shot. So my

survey demonstrates the excellence of today's golfers: every shot they play ends up in the hole!

I also surveyed tennis players at Wimbledon. Out of 256 contestants in the Gentlemen's and Ladies' Singles, no fewer than 254 lost their last match in the tournament! So any tennis player playing a match at Wimbledon has a less than 1 in 100 chance of winning it. For professional sportsmen this seems a staggeringly bad success rate! So it would appear that top golfers are better at their job than top tennis players. Of course, there is an obvious flaw in this analysis.

Let's return from that digression. I'm going to see if someone can be inspired to make a successful guess. I'd like a volunteer from the audience, please.

My next slide is going to reveal my prediction about an event which will take place shortly. My prediction may be true or it may be false. My volunteer has no idea what my prediction is, and therefore has no way of knowing whether it will be true or false. I am going to ask my volunteer to write on this card "Yes" if they think (or guess) the event I have predicted WILL occur, and "No" if they think it WON'T. Clearly there is a fifty-fifty chance that my volunteer will get it right, so if, for example, I offered odds of 3:1 for a correct answer, that would obviously be a rather good bet for you to accept.

Here is my prediction:

The volunteer will write "No" on the card.

So what happened? If my volunteer wrote "No" on the card, then they thought my prediction would fail, but in fact I was right. If on the other hand they wrote "Yes", then they thought my prediction would be accurate, but it turned out that it was wrong. So whatever they wrote, I could be sure that my volunteer would guess wrong!

This brings me to one of the heroes of this evening's talk: the early fourteenth-century philosopher John Buridan, who explored paradoxes of self-reference. Buridan's name is associated with an unlucky donkey. In this parable, sometimes used by political cartoonists, Buridan's ass was presented with two equally enticing piles of hay, exactly equally far away. With no rational reason to prefer one pile to the other, the unfortunate animal, unable to choose which food supply to eat first, starved to death. (This story does not appear in Buridan's surviving works, and it is not clear whether he originated the example, or whether perhaps it was invented by one of his opponents as an argument against his philosophy.)

Buridan deserves to be remembered for much more than cruelty to animals. The stories of his life are colourful, although modern scholars doubt their veracity. As a young man he was competing with one Pierre Roger for the affections of the wife of a shoe-maker. During this affair he supposedly hit Roger over the head with a shoe, as a result of which Roger developed the phenomenal memory for which he was famous when he later became Pope Clement VI. According to the stories, Buridan's rather too close friendship with the Queen of France resulted in the King's having him tied up in a sack and thrown into the river Seine: Buridan survived thanks to the help of a former student. Some version of this incident led to Buridan's being name-checked by Villon in one of the most famous of French poems:

Où est la très sage Heloïs,
Pour qui fut castré et puis moyne
Pierre Esbaillart à Saint-Denys?
Pour son amour eut cest essoyne.
Semblablement, où est la royne

Qui commanda que Buridan
Fust jetté en ung sac en Seine?
Mais où sont les neiges d'antan!

“Where are the snows of yesteryear?”

Buridan's romantic exploits may be mythical but his scientific achievements were real and equally extraordinary. In mechanics he developed a theory of impetus which in some ways anticipated the ideas of Newton 250 years later, and to him has been attributed the creation of the modern theory of money. I can't resist quoting an example from his economic analysis which shows something of his provocative approach. He borrows the names of Socrates and Plato for his characters, and posits that Socrates, willingly and with her consent, allows Plato to commit adultery with his wife in exchange for ten books. Buridan argues that, while all the parties may have suffered moral injury, in terms of the trade, both have gained something they value more than what they have paid for it: in terms of modern game theory, this is not a zero-sum game, but one from which both participants benefit.

John Buridan's study of self-reference seems to me to be remarkably modern, and I am going to show you a couple of his examples. First of all, consider the following statement:

I say that I am the greatest mathematician in the world.

Now, let it be clearly understood that I am not really the greatest mathematician in the world. But is the statement above true or false? Well, Buridan argues that it is definitely true: it says that I say I am the greatest mathematician in the world, and I did just say that.

You might think that this is a rather silly example, but the stakes are high, as Buridan goes on to show. For Psalm 14 begins

The fool hath said in his heart, There is no God.

and for Buridan what the fool says is certainly false. So if my sentence is false, then the first sentence of the Psalm must equally be false. But that would mean that David, in writing these words, was lying, which is unthinkable. So it is a matter of some importance that my sentence be true, even if a more natural interpretation is otherwise.

Buridan also analysed another rather nice paradox. Consider the proposition

Someone at this moment is thinking about a proposition and is unsure whether it is true or false

Please think about this proposition. Is it true or false? You can't be sure it's false, because someone in Australia might be thinking, for example, about the Riemann Hypothesis, establishing the truth of which is an unsolved mathematical problem. But if you are unsure, then the proposition in that case is true, so you have established it and you aren't unsure. Therefore the proposition must be true: and yet as far as any of us know you might be the only person in the world who is thinking about a proposition at this moment, and you are not unsure of the truth of that proposition, because you have just established that it is true.

The reason I have introduced Buridan was to show that another of his self-referential sophisms anticipated my prediction which my volunteer got wrong. In this example, Socrates wishes to cross a bridge which is guarded by Plato. Plato tells Socrates that, if Socrates makes a true statement, Plato will let him cross, but if he makes a false statement, then Plato will throw him in the river. Now if all Socrates wanted to do was to cross the bridge he could simply say "I exist", or "2 plus 2 equals 4", but in fact Socrates chooses to say "You will throw me in the river". Plato is now nonplussed: if he throws Socrates in the river, then Socrates was telling the truth and should have been allowed to cross; but if he lets Socrates cross, then Socrates's statement was a lie and he should have been thrown in the river. What this shows is that Plato has made a promise he cannot keep.

Buridan's bridge story is well known, partly because it appears in chapter 51 of *Don Quixote*. Similar stories go back to classical times. The second-century writer Lucian has a crocodile seize a baby and offer to restore it to its mother if she correctly predicts whether the crocodile will return the baby or eat it. Whichever she answers, the crocodile can argue that the baby should be eaten.

These ideas have given rise to lots of logical puzzles, some of remarkable ingenuity, involving statements which can be true or false. Typically these puzzles are set on an island where every individual either always tells the truth or always lies (but you do not know which). Raymond Smullyan is the master of these: here is a simple example.

You meet two people from this island, A and B. A says "At least one of us is a liar." What are A and B?

To solve this, first suppose A is a liar. Then at least one of the two **is** a liar, so A's statement was true, contradicting our assumption that A was lying. So A must have been telling the truth, in which case one of the two must be a liar, and that must be B. So A is a truth-teller and B is a liar.

Here's another one, which shows the intricacies to which the scenario can give rise.

I found two of the islanders, C and D, sitting together. I asked "Is either of you a truth-teller?" When C answered, I could deduce what each of them was. How?

This one is more complex. If C said "yes" then the answer could have been given truthfully by a truth-teller or falsely by a liar, and that wouldn't allowed me to make my deduction. But if C's answer was "no", it can't have been given by a truth-teller, who would have had to answer "yes". So C must be a liar, in which case the answer "no" was a lie, which means one of them must be a truth-teller (and, since that isn't C, that must be D). So from the puzzle you can deduce that the answer was "no", C is a liar and D is a truth-teller.

If you want a puzzle to think about on your journey home, here's one whose answer I won't tell you.

E and F are two islanders. E said "We are both of the same type" (meaning that both are liars or both are truth-tellers) and F said "We are of opposite types." What are E and F?

Is any of this of any practical value, or does it just provide a source of puzzles for nerds? After all, there probably aren't any islands where some of the inhabitants always tell the truth and others always lie. But in fact exactly this situation did interest French lawyers in the sixteenth century, in the context of examining suspected witches. The theology of the time indicated that witches were possessed by the Devil, and either always lied or at least often lied. The lawyers worked out that if a witch always lies, then clever questioning can elicit the information that is sought, but that if they only lie some of the time, and are trying to mislead you, then they can answer your questions in such a way that you can't rely on their evidence at all.

A much earlier court case had presented a logical paradox in court. The story is that the Greek philosopher Protagoras (who lived in the 5th century BCE) gave legal training to the student Euathlus with the terms that half the fee was payable in advance and the balance when Euathlus won his first court case. (There are varying versions of the precise condition.) After completing the training, Euathlus delayed taking on any cases and eventually Protagoras, impatient for the balance of his fee, sued for it. Protagoras argued that if he won the case, the ruling would require him to be paid, while if he lost the case, Euathlus, having won the case, would have to pay under the terms of the original agreement. On the other hand, Euathlus argued that if his plea was successful, then by the judgment he wouldn't have to pay, while if he lost the case, under the terms of the agreement there would be no occasion to pay. We seem to have a paradox: whatever judgment is made, there are convincing reasons why Protagoras should get the money, but equally strong arguments that he shouldn't. (Perhaps the correct judgment would be in favour of the student Euathlus, allowing Protagoras to sue again on the grounds that having won his case Euathlus now definitely owes the fee. That would seem to follow the important legal principle that the more court cases there are and the more work there is for lawyers, the better.) Incidentally, although one might have thought that an argument of this type would appeal particularly to those with mathematical inclinations, Protagoras disapproved of the study of mathematics, saying, "The subject matter is unknowable and the terminology distasteful".

Again, this paradox seems more like an entertaining mind-twister than a real legal problem. But in fact it was cited as a precedent in a court case in Ohio in 1946. A physician called Jones was charged with carrying out an illegal abortion. The only evidence implicating him came from the woman, Harris, on whom he had allegedly carried out the operation. Now the law specified firstly that a woman who voluntarily arranged an abortion was an accomplice to the illegal act, so if Jones were found guilty then Harris must also be found guilty, and secondly that someone cannot be convicted solely on the word of a criminal accomplice. These conditions set up a paradox. If Jones is guilty then Harris is an accomplice and her evidence cannot be used against Jones, who must therefore be acquitted, in which case Harris's evidence is valid and Jones should be convicted. The judge in fact ordered that if the jury found Jones guilty, then they must by law find Harris to have been his accomplice, which put the jury in the position of being unable to reach a non-paradoxical verdict. Nevertheless the jury convicted, and the court of appeal concurred, noting that Jones's argument (that Harris's evidence should be disregarded because she was an accomplice) actually presupposed that he was guilty.

So we have seen a number of examples which show that these paradoxes are not only of academic interest. But what about mathematics and computing? Paradoxes have prompted mathematical breakthroughs. The apparent paradox that the squares $\{1, 4, 9, 16, \dots\}$ are a proper subset of the positive integers $\{1, 2, 3, 4, \dots\}$, which contain all the squares and many more numbers, and therefore has more members, and yet the two sets are also in one-to-one correspondence, so they have the same number of elements. This had puzzled mathematicians from Galileo's time until Cantor resolved the paradox with his theory of infinite sets. But it was at the beginning of the twentieth century that a plethora of paradoxes emerged which led to sensational developments in the logic underlying pure mathematics.

At the end of the nineteenth century mathematicians like Gottlob Frege were trying to put mathematics on a solid foundation through set theory. The idea was that the concept of a set - a collection of objects - was so simple that we could be sure no contradiction could arise. But the English logician Bertrand Russell showed that even the Arcadia of set theory was not free of paradox.

Russell's Paradox starts with the observation that some sets are members of themselves. For example, the set of all sets is itself a set, so it is a member of itself, whereas the set of all teapots is not itself a teapot, so it is not a member of itself. So "being a member of itself" is a property of some sets but not others. This means that we can consider the set of all sets that are not members of themselves - let's call it S . Is this set S a member of itself? If it is, then since it contains itself, S is not a member of the set of all sets that do not contain themselves, which means S is not a member of itself. But if S is not a member of itself, then it must be a member of the set of all sets that do not contain themselves, so it is a member of itself. We have a loop very similar to that of Raymond Smullyan's interview lie.

Russell's paradox was a disaster for Frege, who had to insert a note into his two-volume book currently in press noting that Russell's example had refuted the whole work (although some people now think this was an over-reaction). It entered popular as well as mathematical culture, often recast as the Barber Paradox: in a certain village the barber shaves all and only those who do not shave themselves. So who shaves the barber? If he shaves himself, then by the proposition he is not shaved by the barber, so he doesn't shave himself: if he doesn't shave himself, then the condition states that he is shaved by the barber so he does

shave himself.

For me the Barber paradox is rather different from the set-theory one. The Barber story simply proves that there cannot be a village in which that situation arises. If I said that “in a certain village $2+2 = 5$ ”, I would be lying: there is no such village. The same is true of the village in which the Barber Paradox is supposedly set. On the other hand, Russell’s set theory paradox seems to me to be a genuine paradox in the context of naïve set theory.

Other paradoxes were examined at around the same time. A linguistic analogue of Russell’s paradox is due to Grelling and Nelson. Some adjectives, like “short” or “polysyllabic” describe themselves – “short” is a short word and “polysyllabic” has many syllables. Others, like “long” and “monosyllabic”, do not describe themselves – “long” is not a long word and “monosyllabic” has more than one syllable. Let’s call the former, those adjectives that describe themselves, “autologous”, and the others “heterologous”. Now, is the adjective “heterologous” itself heterologous? If so, then it doesn’t describe itself, and if it doesn’t describe itself, then by definition it is heterologous and so it does describe itself. On the other hand, if it is not heterologous, then that means it doesn’t describe itself, so it is heterologous.

My favourite is Berry’s paradox. This was originally formulated in terms of the numbers of syllables available in English but I’m going to put it in terms of Twitter – a technology which wasn’t available to Berry and Russell in 1906.

In Twitter a tweet can use up to 140 characters, and there are only finitely many available symbols. So clearly the number of different possible tweets is finite (in fact it is N^{140} where N is the number of available symbols). Now there are infinitely many positive integers. These can be identified in different ways – I could tweet “1” or “one” or “zero factorial” or “ $4 - 3$ ” to identify the number one, for example, but clearly I can identify only at most N^{140} different integers in tweets, which means that not all integers can be identified unambiguously in a tweet. Now since some integer less than N^{140} cannot be identified in a tweet, there is a smallest such integer – we can find it by testing every integer in turn, which is a finite process. So we can find “The smallest integer that cannot be identified in a tweet of fewer than 160 characters”. But that tweetable phrase now identifies it!

These paradoxes all involve self-reference. But Willard Quine created a paradoxical statement that doesn’t refer to itself:

“Yields falsehood when preceded by its quotation” yields falsehood when preceded by its quotation.

And Raymond Smullyan has a nice reformulation of Russell’s Paradox. Somebody who pretends to be something they are not – a bogus doctor, or lawyer, or mathematician – is called a charlatan. So suppose someone pretends to be a charlatan. Is this bogus charlatan a charlatan, or not?

If my Buridanesque claim that “I say that I am the world’s greatest mathematician” was truthful didn’t convince you, perhaps I can prove the same thing another way. Consider these two statements:

- A: Both these statements are false
- B: I am the world’s greatest mathematician

Now, A cannot be true, because it asserts that it is false. So A must be false: if B were also false then A would be true, which is a contradiction, but if A is false and B is true there is no contradiction. So we have shown that B must be true. Note how this differs from “This sentence is false”, which has no consistent reading – here we have a perfect mathematical proof which considers all possible combinations of truth

values for the two statements and shows that all but one lead to contradictions, thus proving the correctness of the only non-contradictory scenario.

I can prove my credentials as a mathematician in other ways too. Consider the statement

If there were a Nobel Prize for mathematics then, as the greatest mathematician in the world, I would deserve to win it.

Now, there is no Nobel Prize in mathematics (the persistent story is that Nobel didn't create one because his wife had an affair with a mathematician: the persistence of this myth is baffling because it has been convincingly refuted, and in fact Nobel never married). So the first part of this "If A, then B" proposition is false. Now, in logic a proposition of the form "If A, then B" or "A implies B" is true unless A is true and B is false. In my example, A is false, which means that the whole proposition is true. Of course, mathematically the statement doesn't have the connotations that it has in natural language: since there is no Nobel Prize for mathematics, logically it says absolutely nothing about my mathematical ability.

So perhaps we need another proof. For this one I will use Haskell Curry's paradox from 1941. Consider the statement

If this statement is true, then I am the greatest mathematician in the world

This statement has the form "If A then B", and we establish the truth of such a proposition by testing it in the case when A is true. Here the truth of A is simply the truth of the statement as a whole, so if we assume A then we are also assuming "If A then B". And if we are told that both A and "If A then B" are true, then we can deduce that B is true. So our proposition is certainly true, and if it is true then I am the greatest mathematician in the world.

For the avoidance of doubt, I should again make it clear that even though I have proved it in four different ways, I am not the world's greatest mathematician! And that is not just false modesty. And neither was that last sentence. Nor that last one. And so on...

Incidentally, Lewis Carroll, in his 1895 story "What the Tortoise said to Achilles", presents another paradox regarding "If A then B", which we can rephrase as "A implies B". In this case Achilles knows that both A and "A implies B" are true, and he wants to deduce that B must be true. He believes that he doesn't need anything more to do so, but the Tortoise points out that, to make that deduction, he does need to know something else, that

(C) If A and "A implies B" are both true, then B must be true.

The Tortoise will grant that, but Achilles needs something more, that

(D) If A, "A implies B" and "Together A and 'A implies B' imply B" are all true, then B is true.

Again, even if D is true, Achilles needs a further step, and so on for ever. Carroll's dialogue ends with the suggestion that Achilles should be renamed "Taught-Us" while the Tortoise should be "A Kill-Ease", which shows that our emeritus Gresham Professor Robin Wilson isn't the only mathematician who makes dreadful puns.

But back to the serious maths. Are there contradictions in mathematics? Many of us have the romantic idea that the whole point of mathematical proof is to establish facts beyond all possible doubt. For us romantics, a single contradiction in mathematics would put the whole edifice in doubt. We want certainty. The great German mathematician David Hilbert said in his famous address to the International Congress of Mathematicians in Paris in 1900, "In mathematics, there is no *ignorabimus*" – there is no "We shall not know" – and thirty years later he still maintained, "We must know — we shall know!" We might not yet have found a proof for the Riemann Hypothesis or for the Goldbach Conjecture that every even number is the sum of at most two primes, but if these things are true, a proof exists, and hopefully some ingenious future mathematician will find that proof.

But two results published in 1931 by Kurt Gödel presented a blow to this view of mathematics. Suppose that we have a system of logical rules which tells us how we can make deductions, and that this system allows us to count (without which our mathematics would be very limited!) Can we use this system to show that there cannot be a contradiction amongst the results we can prove using the system? Well, Gödel showed that the answer can be "yes", which might seem like good news, except that the answer is "yes" if and only if the system does contain a contradiction. In other words, if our system is consistent – that is, it doesn't contain contradictions – then we cannot prove that it is consistent. But if our system is not consistent, then we can prove that it is consistent!

So suppose you are in the market for a good basis for doing mathematics. You want a system which is free of contradictions. What you are offered is a system which can prove it has no contradictions if and only if it actually does have contradictions. Does that sound like a good deal?

That was Gödel's Second Incompleteness Theorem. His First Incompleteness Theorem tells us that in a consistent mathematical system like this, there are results that are true which we cannot prove to be true. So, for example, the Goldbach Conjecture might be true but there might be no way to prove it. (It cannot be false but unprovable, because if it is false there is some even integer which is not the sum of two primes, and in that case it would be possible to find that integer and prove the theorem to be false. So if the conjecture is unprovable, it must be true: we certainly won't be able to prove that it is unprovable!)

How did Gödel prove these devastating theorems? Essentially he created a version of the Liar Paradox "This statement is false". Using a clever coding system to represent mathematics by integers, he effectively creates a proposition that says "This statement is not provable within our system". If that sentence could be proved, we would have a contradiction, so the statement cannot be proved, therefore it is true.

What was the outcome of Gödel's discovery? One might have thought that mathematicians many of whom over the centuries have claimed that their subject has a unique relationship with truth, might be rather worried that they can prove their systems have no contradictions only if the reverse is in fact true. Or that the search for proof of longstanding conjectures might seem less attractive if there is a possibility that no such proof exists. (Admittedly, if a statement is unprovable within a logical system, one can always create a new system in which it is provable, but is that good mathematics?) But in fact, although supposedly the great mathematician John von Neumann came out of a seminar at which Gödel had presented his results saying "It's all over", Gödel's theorems don't seem to have had a major effect on the practice of pure mathematics. Indeed, arguably they are very helpful: when someone challenges a pure mathematician "How can you spend time proving things when you aren't sure that your logical system doesn't contain contradictions?" they can now answer "It's impossible to prove that, so we don't need to waste time worrying about it."

But lots of people have misused Gödel's Theorems to make claims about life in general by applying it to systems which don't meet the tight conditions of the theorems. For example, they have been abused to attack religion: one finds on the internet claims such as

Gödel's Incompleteness Theorem demonstrates that it is impossible for the Bible to be both true and complete.

This is nonsense, because there is no reason to suppose that the Bible forms a logical system of the type to which the theorem applies. Gödel himself created a logical proof of the existence of God, though he possibly did that more as an exercise in logic than for theological purposes.

A perhaps more plausible (though in my opinion equally mistaken) argument is that our human ability to go outside a formal system – to see that a result that is unprovable within a system is in fact true within that system – is evidence that our consciousness is different from that of a computer, which supposedly cannot escape its logical specification.

Arguments reminiscent of the liar paradox also occur in the theory of computing. Alan Turing showed that the Halting Problem – telling whether or not a particular computer programme with a specified input will compute for ever or will eventually stop – is undecidable, and his proof bears some resemblance to Gödel's theorems. Similar ideas are now used frequently in the theory of computational complexity. So the liar paradox is fundamental to one of the most important and exciting areas of intellectual enquiry in the 21st century.

I'm coming to the conclusion of this lecture. You will remember that I guaranteed that you would be surprised. Well, perhaps you were surprised by something I have said tonight, in which case my guarantee has been met. But perhaps nothing in my lecture has surprised you. In that case, you were expecting a surprise guaranteed by a visiting Professor of Gresham College, and you found that guarantee wasn't fulfilled. That is surely a surprising outcome, and so in that case too, you have had the promised surprise!

Thank you for listening.

Further reading:

Douglas Hofstadter, *Gödel, Escher, Bach: an Eternal Golden Braid* (Penguin, 20th anniversary edition, 2000)

Raymond Smullyan, *What is the Name of this Book?* (Prentice-Hall, 1978: Dover, 2011) and *The Gödelian Puzzle Book: Puzzles, Paradoxes and Proofs* (Dover, 2013)

Francesco Berto, *There's Something About Gödel!: The Complete Guide to the Incompleteness Theorem* (Wiley-Blackwell, 2009)

Scott Aaronson, *Quantum computing Since Democritus* (Cambridge University Press, 2013)