Gödel's incompleteness theorem

Bengt Ringnér

Centre for Mathematical Sciences, Lund University, Lund, Sweden.
Homepage: http://www.maths.lth.se/∼/bengtr

December 8, 2008

1 A remarkable equation

At the previous turn of century it was generally believed, at least among mathematicians, that for every equation like

$$x^n + y^n = z^n, \quad x, y, z > 0,\ n > 2$$

or

$$x^4 + y^4 + z^4 = u^4, \quad x, y, z, u > 0$$

concerning integers, one would sooner or later find one of

1. a solution[1], or
2. a proof that no solution exists[2].

Since all formulas and proofs can be represented by strings of symbols, they can be numbered according to some rule, and a function B can be defined by

$$B(x) = \begin{cases} y & \text{if string no } x \text{ is a proof ending with string no } y \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, also the second problem can be regarded as solving an equation, so one has to do one of

[1] To the second equation, posed by Euler, a solution $x = 2\,682\,440, y = 15\,365\,639, z = 187\,960, u = 20\,615\,673$ was found in 1988.

[2] Fermat's last theorem states the that the first equation has no solution. A proof of this was obtained recently.

1. find a solution to $f(x) = y_0$, or
2. find a solution to $B(x) = y_1$ where y_1 is the number of the string "$f(x) \neq y_0$ for all x".

Of course, one might have $f = B$. But what happens if one succeeds in finding an y_0 such that $y_1 = y_0$?

- Then the problems of finding a solution to $B(x) = y_0$ and of finding a proof that it has no solution are the same, so, unless contradictions can occur in mathematics, neither will ever be found[3].

  It might seem paradoxical that we have nevertheless concluded that the equation is unsolvable. The reason is that we have used a new assumption which is not taken into account in the B-function. If it were so, and if B_κ denotes that a set κ of new rules have been incorporated, we would find ourselves in the same situation over and over again.

- As a consequence, is impossible to prove that contradictions can never occur, since, if it were possible, then a proof of the unprovable $B(x) \neq y_0$ would emerge.

Gödel found such an y_0. His finding has stimulated a great number of more or less philosophical dissertations on self-referential statements like a Kretan saying "all Kretans are liars". However, the present one is not more esoteric than the following example.

Example Let z be the number of signs in the string

"This sentence contains ?? signs".

Then $z = 32$, so the sentence

"This sentence contains 32 signs",

makes a statement about itself. Of course, it is not because it says so that it contains 32 signs, it is because anyone can count the number.

□

Gödel defined a substitution function by $S_y(x, z)$ equal to the number of the string that results whan one takes string no x and replaces, whenever

[3] It is still conceivable that, meanwhile, someone might come up with a proof that a solution exists witout actually exhibiting one. This has to do with ω-consistency to be treated below.

possible, the symbol "y" with the number z written out with digits. If the first string in the example has number p, then the second one has number $S_{??}(p, z)$.

Now, let p be the number of

$$\text{“}B(x) \neq S_y(y, y) \text{ for all } x\text{”}.$$

Suppose that $p = 123$, say.

- Then $S_y(123, 123)$ is the number of

$$\text{“}B(x) \neq S_y(123, 123) \text{ for all } x\text{”},$$

so $y_0 = S_y(123, 123)$ serves as the desired y_0.

2 The first incompleteness theorem

In order to convince oneself that the above is not merely a play with words, one has to distinguish between mathematics and metamathematics. Solving the equation $B(x) = S_y(123, 123)$ belongs to mathematics, while the question whether there is a proof ending with "$B(x) \neq S_y(123, 123)$ for all x" belongs to the latter. Gödel went one step further and and obtained a purely metamathematical result.

2.1 The result

Suppose a proof were found that the equation $B(x) = y_0$ has no solution. Then the number of the string constituting the proof is a solution to the equation, which counts as a proof that the equation has a solution. Therefore,

- if mathematics is consistent, that is free of contradictions, then it is impossible to prove that $B(x) = y_0$ has no solution.

Next, if one could prove that the equation has some solution while at the same time there are proofs of $B(1) \neq y_0, B(2) \neq y_0, \ldots$ for all integers, a weaker form of inconsistency, called ω-inconsistence by Gödel, would occur. On the other hand, we know that, if mathematics is consistent, then each computation of $B(n)$ would give $B(n) \neq y_0$, so

- if mathematics is ω-consistent then it is impossible to prove that $B(x) = y_0$ has a solution.

The result is Gödel's first incompleteness theorem; provided that mathematics is ω-consistent, he exhibits a proposition that can be neither proved nor disproved.

2.2 Technical details

In formal mathematics one starts with *axioms* concerning natural numbers, denoted by $0, \$0, \$\$0, \$\$\$0, \ldots$ (the "\$" sign denotes the next number), and deduces cryptic formulas like

$$\forall Add\left(\left(\forall a \forall b \forall c\left(\left(Add(a,b,c) \Leftrightarrow Add(a,\$b,\$c)\right) \wedge Add(a,0,a)\right)\right) \Rightarrow Add(\$\$0,\$\$0,\$\$\$\$0)\right) \tag{1}$$

and

$$\neg\forall Add\left(\left(\forall a \forall b \forall c\left(\left(Add(a,b,c) \Leftrightarrow Add(a,\$b,\$c)\right) \wedge Add(a,0,a)\right)\right) \Rightarrow Add(\$\$0,\$\$0,\$\$\$\$\$0)\right). \tag{2}$$

A deduction starts with some axioms, which are given formulas, and rewrites them according to given *deduction rules*. As far as I have understood, a mathematical proof is only valid if it can be transformed into a formal deduction by writing out every minute detail, and it is tacitly understood that this can be done with our ordinary proofs.

In *metamathematics* one considers the above formulas as pure abracadabra and studies what can be achieved with the manipulations, for instance,

1. whether the axioms are *consistent* (with respect to the deduction rules), that is, one can never deduce both a formula and the same formula prepended with a "¬" sign, and

2. whether the axioms are *complete*, that is for any syntactically correct formula, one can deduce either the formula or its negation.

When doing this, ordinary mathematics can be used. Gödel showed, with ordinary induction, that there exists a formula, something like

$$\forall Add\left(\left(\forall a \forall b \forall c\left(\left(Add(a,b,c) \Leftrightarrow Add(a,\$b,\$c)\right) \wedge Add(a,0,a)\right)\right) \Rightarrow Add(x,y,z)\right), \tag{3}$$

such that if the relation $x + y = z$ holds, then one can replace the x, y, and z with a "0" prepended by the corresponding number of "\$" signs, and deduce the resulting formula while, if the relation does not hold, the formula prepended with a "¬" sign is deducible. Formulas (1) and (2) are examples of this for, respectively, $2 + 2 = 4$ and $2 + 2 \neq 5$.

- More generally, he showed, Thm V page 186 in [1], that for all relations which, like $x + y = z$, can be defined recursively, there exist corresponding formulas. This gives a link from metamathematics to mathematics.

Since all formulas and deductions can be represented by strings of symbols, they can be numbered, and functions can be defined according to

$$B(x) = \begin{cases} y & \text{if string no } x \text{ is a deduction ending with string no } y \\ 0 & \text{otherwise.} \end{cases}$$

and substitution functions[4]

$$Sb_y(x, z)$$

which takes string no x, replaces, whenever possible, the symbol in the subscript with a "0" prepended with z "$" signs, and returns the number of the resulting string. For instance, if formula (3) has number n, then the formula with number

$$Sb_x(Sb_y(Sb_z(n, z), y), x)$$

is deducible if $x + y = z$ while the formula prepended with a "$\neg$" is deducible if $x + y \neq z$.

Next he defined a relation Q between two nonnegative integers, x and y by

$$Q(x, y) \Leftrightarrow B(x) \neq Sb_y(y, y).$$

It turns out to be recursive, so it has a corresponding formula

$$\forall Q\Bigg(\Big(\forall a \forall b\Big(Q(a, b) \Leftrightarrow \ldots\Big)\Big) \Rightarrow Q(x, y)\Bigg),$$

where the "..." symbolises a large amount of recursions reflecting what is meant by Q. Let q be the number of this formula, p the number[5] of

$$\forall x \, \text{formula}_q,$$

[4] Gödel uses the notation

$$Sb\left(x \begin{array}{c} 19 \\ Z(z) \end{array}\right) = Sb_y(x, z),$$

since variables $x, y, \ldots$ are numbered by the primes, $17, 19, \ldots$ and $Z(n)$ denotes the number of the string consisting of a "0" prepended by n successor signs.

[5] If $Gen_x(y)$ denotes the number of the formula $\forall x \text{formula}_y$, we can write $p = Gen_x(q)$. Gödel writes this $p = 17 \, \text{Gen} \, q$.

and try to solve the following equation[6]:

$$B(x) = Sb_y(p,p). \tag{4}$$

Since $Sb_y(p,p)$ is the number of[7]

$$\forall x \forall Q\biggl(\Bigl(\forall a \forall b\bigl(Q(a,b) \Leftrightarrow \dots\bigr)\Bigr) \Rightarrow Q(x, \underbrace{\$\dots\dots\$}_{\substack{p \\ \text{times}}} 0)\biggr),$$

the following emerges:

1. Suppose $Sb_y(p,p)$ were deducible. Then (4) would have a solution, $x = n$, so $Q(n,p)$ would be false, and, hence,

   $$\neg\forall Q\biggl(\Bigl(\forall a \forall b\bigl(Q(a,b) \Leftrightarrow \dots\bigr)\Bigr) \Rightarrow Q(\underbrace{\$\dots\$}_{n} 0, \underbrace{\$\dots\$}_{p} 0)\biggr)$$

   deducible, while, on the other hand, the deduction of $Sb_y(p,p)$ could be continued into one of

   $$\forall Q\biggl(\Bigl(\forall a \forall b\bigl(Q(a,b) \Leftrightarrow \dots\bigr)\Bigr) \Rightarrow Q(\underbrace{\$\dots\$}_{n} 0, \underbrace{\$\dots\$}_{p} 0)\biggr).$$

   Therefore:

   - Unless it is possible to deduce both a formula and its negation, $Sb_y(p,p)$ cannot be deduced and (4) has no solution.

---

[6] If formula no $Sb_y(p,p)$ is interpreted as

$$B(x) \neq Sb_y(p,p) \quad \text{for all } x,$$

then the equation amounts to finding a a proof of its own unsolvability. However, in metamathematics we have to go through the link established by Gödel's Thm V.

[7] For the reader who wants to check with Gödel's original work on page 188, we note that $Sb_y(p,p) = Gen_x(r)$ (17 Gen r) where $r = Sb_y(q,p)$ is the number of

$$\forall Q\biggl(\Bigl(\forall a \forall b\bigl(Q(a,b) \Leftrightarrow \dots\bigr)\Bigr) \Rightarrow Q(x, \underbrace{\$\dots\$}_{p} 0)\biggr),$$

since it does not matter in which order we substitute a value for y and prepend $\forall x$, cf. Gödel's formula 13 and footnote 44 on page 187.

2. Assuming consistency, we have just concluded that $B(x) \neq Sb_y(p,p)$ for all x. Consequently, all formulas

$$\forall Q\Bigg(\Big(\forall a \forall b \Big(Q(a,b) \Leftrightarrow \dots\Big)\Big) \Rightarrow Q(\underbrace{\$\dots\$}_{n}0, \underbrace{\$\dots\$}_{p}0)\Bigg), \qquad n = 0, 1, \dots$$

are deducible. If nevertheless

$$\neg\forall x \forall Q\Bigg(\Big(\forall a \forall b \Big(Q(a,b) \Leftrightarrow \dots\Big)\Big) \Rightarrow Q(x, \underbrace{\$\dots\$}_{p}0)\Bigg),$$

that is the negation of formula no $Sb_y(p,p)$, were deducible, then what Gödel calls an ω-contradiction would occur, so

- if the axioms are ω-consistent, then the negation of formula no $Sb_y(p,p)$ cannot be deducible.

3 The second incompleteness theorem

3.1 The result

As noted earlier, we cannot prove that the axioms of mathematics are consistent.

3.2 Technical details

As proved above, if the axioms are consistent, it is impossible to deduce formula no $Sb_y(p,p)$, namely

$$\forall x \forall Q\Bigg(\Big(\forall a \forall b \Big(Q(a,b) \Leftrightarrow \dots\Big)\Big) \Rightarrow Q(x, \underbrace{\$\dots\$}_{p}0)\Bigg).$$

For this part ω-consistency was not needed. On the other hand, the formula "says" that it is not deducible. Therefore, if we add an axiom "saying" that the original axioms are consistent, it might be deducible.

The new axiom is based on the following observation: If the axioms are inconsistent, then all formulas can be proved, so consistency is the same as: There exists a formula that cannot be proved, namely either f or $\neg f$. Consider the relation

$W(x,y) \Leftrightarrow B(x) \neq y$ and y is the number of a syntactically correct formula.

It is recursive, so it corresponds to a formula

$$\forall W\left(\left(\forall a \forall b\Big(W(a,b) \Leftrightarrow \circ\circ\circ\Big)\right) \Rightarrow W(x,y)\right).$$

Gödel undertook to deduce, in a fortcoming paper

$$\exists y \forall x \forall W\left(\left(\forall a \forall b\Big(W(a,b) \Leftrightarrow \circ\circ\circ\Big)\right) \Rightarrow W(x,y)\right)$$

$$\Rightarrow$$

$$\forall x \forall Q\left(\left(\forall a \forall b\Big(Q(a,b) \Leftrightarrow \dots\Big)\right) \Rightarrow Q(x,\underbrace{\$\dots\$}_{p}0)\right),$$

which would prove that it is impossible to deduce

$$\exists y \forall x \forall W\left(\left(\forall a \forall b\Big(W(a,b) \Leftrightarrow \circ\circ\circ\Big)\right) \Rightarrow W(x,y)\right),$$

which can is interpreted as the axioms being consistent. This was never done, presumably because von Neumann, when hearing Gödel talk on his first theorem, draw this conclusion. As I collect, a formal proof was made much later by Turing in connection with Turing machines.

4 On ω-consistency

For obvious reasons, we can never observe an ω-contradiction. But, if the axioms are consistent and we add the negation of formula $Sb_y(p,p)$ to the axioms, then the result will still be consistent, but not ω-consistent.

References

[1] K Gödel. Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I. *Monatsh. für Math. und Phys.*, 38:173–198, 1931. English translations in 'Collected works of Kurt Gödel' and in 'From Frege to Gödel' Ed. van Heijenoort, J.