An un-rigorous introduction to the incompleteness theorems[*]

PHIL 43904
Jeff Speaks

October 8, 2007

1 Soundness and completeness

When discussing Russell's logical system and its relation to Peano's axioms of arithmetic, we distinguished between the *axioms* of Russell's system, and the *theorems* of that system. Roughly, the axioms of a theory are that theory's basic assumptions, and the theorems are the formulae provable from the axioms using the rules provided by the theory.

Suppose we are talking about the theory A of arithmetic. Then, if we express the idea that a certain sentence p is a theorem of A — i.e., provable from A's axioms — as follows:

$$\vdash_A p$$

Now we can introduce the notion of the *valid* sentences of a theory. For our purposes, think of a valid sentence as a sentence in the language of the theory that can't be false. For example, consider a simple logical language which contains some predicates (written as upper-case letters), 'not', '&', and some names, each of which is assigned some object or other in each interpretation. Now consider a sentence like

$$Fn$$

In most cases, a sentence like this is going to be true on some interpretations, and false in others — it depends on which object is assigned to 'n', and whether it is in the set of things assigned to 'F'. However, if we consider a sentence like

---

[*]The source for much of what follows is Michael Detlefsen's (much less un-rigorous) "Gödel's theorems", available at http://www.rep.routledge.com/article/Y005. There you can also find a bibliography of other work on this topic.

not (Fn & not Fn)

we can see that it is true, no matter what object is assigned to 'n.' So this is an example of a valid sentence in the language of the theory: it's a sentence which can't be false.

Now we can ask, for any theory, how the theorems of the theory relate to its valid sentences. When we talk about soundness and completeness, we are talking about this relationship.

If every theorem of a theory is valid, we say that the theory is *sound*. This means, basically, that if you derive a sentence from the axioms of a sound theory, you never get a falsehood. We usually only work with sound theories. To get an example of an unsound theory, suppose that we added a new connective, 'tonk', to our usual basic logical language, and that 'tonk' was governed by the following two rules:

$$\frac{p}{p \text{ tonk } q}$$

$$\frac{p \text{ tonk } q}{q}$$

Can you see why a logic involving 'tonk' would be unsound?

If every theorem is valid, the theory is sound; if every valid sentence is a theorem, we say that the theory is *complete*. This means that any sentence which is true in the language of a complete theory is provable from the theory's axioms.

Gödel's incompleteness theorems are, as the name would indicate, proofs that certain mathematical and logical theories — one of which is Peano arithmetic — are not complete. That is, Gödel showed that there were certain valid sentences of those theories which were not provable from their axioms.

2   Gödel's first incompleteness theorem

2.1   Gödel numbering and 'provable'

First we note that we can use the natural numbers to come up with 'names' for every formula of arithmetic. We do this by assigning every formula what is now called a Gödel number. There are many ways to do this. One is by first assigning a natural number to every basic symbol of the language of arithmetic, such as, for example, 0, 1, +, *, =, (, ... Then imagine that we have some formula of arithmetic, e.g.

0+1=1

We assign this formula a Gödel number by multiplying the Gödel number for the first digit of the formula (0) by 2, the Gödel number for the second digit of the formula (+) by 3, the Gödel number of the third digit (1) by 5, and so on, multiplying the Gödel number for the nth digit by the nth prime number. The sum of these n products is the Gödel number for the formula as a whole.

The point of using only prime numbers is that, this way, no two formulae will ever have the same Gödel number. The key points here are that every formula has a Gödel number, and that no two formulae every have the same Gödel number.

Rather than actually writing out the Gödel numbers for formulae, if we're talking about some sentence p, we'll use the following symbol for the Gödel number of p: $\ulcorner p \urcorner$.

The point of this is to give us a way, using the language of arithmetic, of talking about the sentences of that language.

Now assume that we can also define a predicate 'provable' such that, for the theory A of arithmetic,

$$\vdash_A p \iff \vdash_A \text{provable}(\ulcorner p \urcorner)$$

i.e., 'provable($\ulcorner p \urcorner$)' is a theorem of A if and only if 'p' is. That is, 'provable(x)' is a a theorem of the theory just in case the formula whose Gödel number 'x' is is a theorem. Intuitively, formulae involving 'provable' talk about sentences of the language of arithmetic.

2.2   The fixed point theorem

Gödel showed that we could, for the theory A of arithmetic, find a sentence q such that

$$\vdash_A (q \iff \neg \, \text{provable}(q))$$

In other words, he showed that there was a formula such that it was provable in A that the formula was true iff it was not provable.

Suppose first that the formula is true. Then it follows from the above that it is not provable. Suppose instead that the formula is false. We are assuming that theory in question is sound, and hence that falsehoods are never provable in the theory. So, whether the theory is true or false, it is unprovable. Hence,

by the fixed point theorem, it is true, and its negation is false. So there is some formula which is unprovable, and whose negation is also unprovable.

This is the moral of the first incompleteness theorem: there is some formula in the language of arithmetic which can neither be proved or refuted (i.e., have its negation proven) by the axioms.

3   Gödel's second incompleteness theorem

The second incompleteness theorem establishes that we cannot prove, in the language of arithmetic, the consistency of the axioms of arithmetic. The route this takes is a proof of the conditional claim

A is consistent $\rightarrow q$

Since this conditional is provable, if the consistency of A were provable, 'q' would be provable as well. But we just saw above (in the proof of the first incompleteness theorem) that it isn't. So, the consistency of A is not provable in A.

4   (Possible) consequences of the incompleteness theorems

Many consequences have been claimed for the incompleteness theorems, and most of these are still a matter of dispute. Some of the alleged consequences are as follows:

- *Logicism.* The incompleteness theorems show that there is no set of axioms from which all the truths of arithmetic can be proven. So, if we think of logicism as the view that all mathematical truths are disguised versions of truths provable in some system of logic, it seems that Gödel has shown that logicism is false.

- *Mathematical truth and provability.* Gödel's results have also been taken to show that mathematical truth is something independent of human mathematical activity, rather than being in some sense a 'creation of the mind', if you think of the latter as essentially involving the construction of proofs.

- *The computational model of the mind.* It has also been argued that Gödel's theorems show that the mind is importantly unlike machines such as computers. The idea here is that if our minds were like computers, we could know only what was provable. But we can know, for example, that our beliefs are consistent, which (given certain assumptions) is not provable. So, we know things that no computer could figure out.