A Magician Looks at Gödel’s Proof

Raymond Smullyan

USA
Part I

An Incompleteness Theorem
An Incompleteness Theorem

We consider a mathematical system in which certain expressions are called **predicates**, and by a **sentence** is meant any expression of the form $HX$, where $H$ is a predicate and $X$ is any expression in the language of the system.
Intuition:

We might think of a predicate $H$ as being the name of a property of expressions, and $HX$ as a sentence asserting that $X$ has the property named by $H$. 
Some definitions

- The system has a negation symbol $N$, and for any sentence $X$, the expression $NX$ is also a sentence called the **negation** of $X$.

- The system is called **consistent** iff no sentence and its negation are both provable.

- A sentence is called **undecidable** iff neither it nor its negation is provable.
Two conditions

The system obeys the following two conditions:

\( C_1 \) There is a predicate \( P \) (called a provability predicate) such that for any sentence \( X \), the sentence \( X \) is provable iff \( PX \) is provable.

\( C_2 \) There is a predicate \( R \) such that for any expression \( X \), the sentence \( RX \) is provable iff \( PXXX \) is provable.
An amazing consequence

From just these two conditions, something quite amazing follows:

**Theorem** $G_0$ *If the system is consistent, then there is an undecidable sentence.*

Moreover, assuming consistency, one can exhibit an undecidable sentence using any of the symbols $P, N, R.$
Problem 1: Exhibit such a sentence.

Solution: Assuming consistency, such a sentence is \( RNR \).

Proof: By condition \( C_2 \), for any expression \( X \), the sentence \( RX \) is provable iff \( PXX \) is provable, and by condition \( C_1 \), \( PX X \) is provable iff \( XX \) is provable.
A problem, and the solution

Therefore $RX$ is provable iff $XX$ is provable. Since this holds for every expression $X$, it holds taking $X$ to be the expression $NR$, and so $RNR$ is provable iff $N RNR$ is provable. Thus either $RNR$ and its negation $N RNR$ are both provable, or neither is provable. By the assumption of consistency, they are not both provable, hence neither one is provable, and thus $RNR$ is undecidable.
Part II

Omega consistency
Proving at stages

We next consider a system in which the provable sentences are proved at various stages (we can imagine a computer proving these sentences in a certain order). We are given predicates $P$, $R$, and infinitely many predicates $P_1, P_2, \ldots P_k, \ldots$ [As before, $PX$ means that $X$ is provable, and now $P_kX$ means that $X$ is provable at stage $k$].
Conditions

We are now given the following conditions:

\( G_1 \) If \( X \) is provable at stage \( k \), then \( P_k X \) is provable, and if \( X \) is not provable at stage \( k \), then \( \neg P_k X \) is provable.

\( G_2 \) If for some \( k \), the sentence \( P_k X \) is provable, then \( P X \) is provable.

\( G_3 \) [Same as \( C_2 \) of Part I] \( R X \) is provable iff \( P X X \) is provable.
Omega consistency

The system is called \textit{omega consistent} iff it is consistent and also it is the case that there is no sentence $X$ such that $PX$ is provable, and at the same time all the sentences $NP_1X, NP_2X, \cdots NP_kX, \cdots$ are provable.
Gödel’s theorem

Theorem G (After Gödel) If the system is omega consistent, then there is an undecidable sentence.

Problem 2: Assuming omega consistency, exhibit such a sentence (again using any of the symbols \( P, R \) and \( N \)).
A solution (But more interesting)

The sentence is the same as before — $RNR$ — but the proof now is, I believe, more interesting. We already know that condition $C_2$ of Part I holds, and we will now see that omega consistency implies that condition $C_1$ also holds.
Well, even without the assumption of omega consistency, if $X$ is provable, so is $P_X$, because if $X$ is provable, it must be proved at same stage $k$, and hence $P_k X$ is provable (by $G_1$), and then $P X$ is provable (by $G_2$). Thus if $X$ is provable, so is $P X$. 
A solution (But more interesting)

Now for the converse: Suppose $PX$ is provable. If $X$ were not provable, then it would not be provable at any stage, hence by $G_1$, the sentences $NP_1X, NP_2X, \ldots, NP_kX \ldots$ would all be provable, and since $PX$ is, the system would not be omega consistent.
Thus if the system is omega consistent, then the provability of $PX$ implies the provability of $X$, and so then $X$ is provable iff $PX$ is provable, which is condition $C_1$ of Part I. The rest then follows as seen in Part I.
Remarks

In the systems studied by Gödel, predicates are not considered as names of properties of expressions, as I have done, but rather as properties of (natural) numbers. To each predicate $H$ and each number $n$, is associated a sentence $H(n)$ to be thought of as asserting that $n$ has the property denoted by $H$. 

A Magician Looks at Gödel's Proof – p. 19
Remarks

Indeed I have found it convenient to associate to every expression $X$ and number $n$, an expression $X(n)$ such that if $X$ is a predicate, $n$ is a sentence.

Now, Gödel assigned to each expression a number, subsequently called the Gödel number of the expression. Now, to get away from always referring to Gödel numbers, the following schema is quite useful.
For any expressions $X$ and $Y$, I define $XY$ to be $X(y)$, where $y$ is the Gödel number of $Y$. Also I define $XYZ$ to be $X(YZ)$. Now, using this meta notion, in the systems to which Gödel’s argument goes through, there really are predicates $P, R, P_1, P_2, \cdots P_k, \cdots$ satisfying conditions $G_1, G_2,$ and $G_3$!