Gödel's Incompleteness Theorem

by Yurii Khomskii

We give three different proofs of Gödel's First Incompleteness Theorem. All three proofs are essentially variations of one another, but some people may find one of them more appealing than the others. We leave out all details about representability.

First of all, we fix some axiomatization N of arithmetic. This can be the way it was done in Piet's notes, it can be Robinson's arithmetic Q, or simply Peano Arithemtic (PA). Which one we chose has no relevance for the rest of the proof.

1. Definition. Let $R \subseteq \mathbb{N}^k$ be a k-ary relation. We say that R is **representable** (in N) if there is a formula $\phi(x_1, \ldots, x_k)$ in the language of arithmetic, such that for all $n_1, \ldots, n_k$:

$$(n_1, \ldots, n_k) \in R \implies N \vdash \phi(\underline{n_1}, \ldots, \underline{n_k})$$

$$(n_1, \ldots, n_k) \notin R \implies N \vdash \neg\phi(\underline{n_1}, \ldots, \underline{n_k})$$

2. Main Theorem. *If $R \subseteq \mathbb{N}^k$ is a computable relation, then R is representable.*

Proof. This requires some technical work, using e.g. the β-function as in the notes. We won't go into the details. □

There is a weaker notion called semi-representability:

3. Definition. Let $R \subseteq \mathbb{N}^k$ be a k-ary relation. We say that R is **semi-representable** (in N), or N-definable, if there is a formula $\phi(x_1, \ldots, x_k)$ in the language of arithmetic, such that for all $n_1, \ldots, n_k$:

$$(n_1, \ldots, n_k) \in R \iff N \vdash \phi(\underline{n_1}, \ldots, \underline{n_k})$$

Now we are going to encode first order logic, and all the operations it involves, into natural numbers.

4. Definition.

- First, we assign Gödel numbers to formulae in an effective way. If ϕ is a formula, $\ulcorner\phi\urcorner$ denotes its Gödel number. If $e = \ulcorner\phi\urcorner$ then we write $\phi = \phi_e$. Thus we have an effective listing of all first order formulae: $\phi_0, \phi_1, \phi_2, \ldots$.
- Next, we assign Gödel numbers to finite sequences of formulas, in an effective way.

- Now we can define a couple of relations and functions:
  - $Ax_T := \{n \mid \phi_n \in T\}$ (the axioms of T)
  - $Th_T := \{n \mid T \vdash \phi_n\}$ (the theorems of T)
  - $Ref_T := \{n \mid T \vdash \neg\phi_n\}$ (the sententes refutable from T)
  - $Proof_T := \{(s, n) \mid s$ codes a proof of ϕ_n from $T\}$.
  - $Sub : \mathbb{N}^2 \to \mathbb{N}$, a function defined by $Sub(n, m) := \ulcorner\phi_n(\underline{m})\urcorner$.
- We say that a theory T is **computably axiomatized**, or **computably formalized**, or **effectively formalized**, if Ax_T is a computable set.

5. Theorem. *If T is computably axiomatized then $Proof_T$ is a computable relation.*

Proof. Given (s, n), recover the finite sequence $\phi_{n_1}, \ldots, \phi_{n_k}$ coded by s. Check whether $n_1 \in Ax_T$ which can be done by assumption. Furthermore, check whether each step $\phi_{n_i} \mapsto \phi_{n_{i+1}}$ is a correct derivation according to the rules of logic. This operation is also computable. Finally, check whether $n_k = n$. If all of this is satisfied, then s indeed codes a proof of ϕ_n from T, i.e. $(s, n) \in Proof_T$, and otherwise $(s, n) \notin Proof_T$. □

Using this theorem, it immediately follows that many interesting sets regarding consequences of T are c.e. In particular:

6. Theorem. *If T is computably axiomatized then the following sets are all c.e.:*

1. Th_T
2. Ref_T
3. For every $\phi(x)$, the set $\{n \mid T \vdash \phi(\underline{n})\}$.
4. Every set A semi-representable in T (called "T-definable" in the notes).

Proof.

1. We have $Th_T = \{n \mid \exists s\ Proof_T(s, n)\}$, and by the previous theorem, $Proof_T(s, n)$ is a computable relation, so Th_T is Σ_1, hence c.e.
2. Here we note that the function $Negate : \mathbb{N} \to \mathbb{N}$ defined by $Negate(n) := \ulcorner\neg\phi_n\urcorner$ is obviously computable. Then $Ref_T = \{n \mid \exists s\ Proof_T(s, Negate(n))\}$, which is c.e.
3. Here, let e be the Gödel number of ϕ. Then the set in question is equal to $\{n \mid \exists s\ Proof_T(s, Sub(e, n))\}$, which is c.e.
4. This follows immediately from the previous point. □

One consequence of point 4 above is the converse of Main Theorem 2: if R is representable, then R is computable. This follows because if R is representable by ϕ, then R is semi-representable by ϕ and $\overline{R}$ is semi-representable by $\neg\phi$ (assuming that N is consistent), so, since N is obviously computably axiomatized, by the above both R and $\overline{R}$ are c.e., hence R is computable.

But more important is this: now we know that the theory of T (i.e. the set of sentences provable from T) is c.e. Is it also computable?

7. Definition. We say that T is **decidable** if Th_T is computable.

8. Theorem. *If T is computably axiomatized and complete, then T is decidable.*

Proof. If T is inconsistent, then it is trivially decidable since every formula follows form T. If T is consistent and complete, then for all ϕ we have that $T \nvdash \phi$ iff $T \vdash \neg\phi$. Therefore $\overline{Th_T} = \{n \mid T \nvdash \phi_n\} = \{n \mid T \vdash \neg\phi_n\} = Ref_T$. But both Th_T and Ref_T are c.e. by Theorem 6., i.e., both Th_T and its complement are c.e., so Th_T is computable. □

Now we can give the first (in a sense the most direct) proof of the incompleteness theorem.

9. Gödel's First Incompleteness Theorem. *If T is a computably axiomatized, consistent extension of N, then T is undecidable and hence incomplete.*

First Proof. Let $D := \{n \mid T \vdash \phi_n(\underline{n})\}$. Assume, towards contradiction, that T is decidable, i.e., Th_T is computable. Then D is also computable, since it can be written as $\{n \mid Sub(n,n) \in Th_T\}$ and Sub is a computable function. Hence $\overline{D}$ is also computable. Then by Main Theorem 2, $\overline{D}$ is representable (in N), say, by ϕ_e. But then:

$$e \in D \;\Rightarrow\; e \notin \overline{D} \;\Rightarrow\; N \vdash \neg\phi_e(\underline{e}) \;\Rightarrow\; T \vdash \neg\phi_e(\underline{e}) \;\Rightarrow\; T \nvdash \phi_e(\underline{e}) \;\Rightarrow\; e \notin D$$

$$e \notin D \;\Rightarrow\; e \in \overline{D} \;\Rightarrow\; N \vdash \phi_e(\underline{e}) \;\Rightarrow\; T \vdash \phi_e(\underline{e}) \;\Rightarrow\; e \in D$$

which is a contradiction. □

*

You may notice that the above proof is quite similar to our proof that K is not computable. We can give another proof of Gödel's incompleteness theorem which builds more directly on what we already know about basic recursion theory. This requires the additional assumption of ω-consistency (although there may be a way to avoid that.)

10. Definition. A theory T in the language of arithmetic is **ω-consistent** if for all formulas $\phi(x)$ the following holds:

$$\text{If for all } n \in \mathbb{N},\ T \vdash \phi(\underline{n}),\ \text{then } T \not\vdash \exists x \neg\phi(x)$$

For ω-consistent extensions of N we can prove the converse of (6.4.)

11. Lemma. *Suppose T is an ω-consistent extension of N. Then every c.e. set $A \subseteq \mathbb{N}$ is semi-representable in T.*

Proof. Let $A = \{n \mid \exists m\ R(n,m)\}$, with R a computable relation. Suppose R is represented by $\psi(x,y)$ (in N). Let

$$\phi(x) \equiv \exists y\ \psi(x,y)$$

We will show that ϕ semi-represents A. Indeed, if $n \in A$ then for some m, $R(n,m)$ holds, so $N \vdash \psi(\underline{n},\underline{m})$ and hence $N \vdash \exists y\ \psi(\underline{n},y)$, i.e., $N \vdash \phi(\underline{n})$. Hence $T \vdash \phi(\underline{n})$. On the other hand, if $n \notin A$ then for all m, $R(n,m)$ does not hold and so for all m we have $N \vdash \neg\psi(\underline{n},\underline{m})$, and hence $T \vdash \neg\psi(\underline{n},\underline{m})$. Then by ω-consistency, $T \not\vdash \exists y\ \psi(\underline{n},y)$, i.e., $T \not\vdash \phi(\underline{n})$. This proves that ϕ semi-represents A in T. □

As simple as the Lemma looks, it has the following consequence:

12. Theorem. *Let T be an ω-consistent extension of N. Then for all c.e. sets A we have*

$$A \leq_m Th_T$$

(i.e., Th_T is Σ_1-complete.)

Proof. Suppose ϕ_e semi-represents A in T. Let f be the function defined by $f(n) := Sub(e,n) = \ulcorner\phi_e(\underline{n})\urcorner$. Obviously f is computable. Then

$$n \in A \ \Rightarrow\ T \vdash \phi_e(\underline{n})\ \Rightarrow f(n) \in Th_T$$

$$n \notin A \ \Rightarrow\ T \not\vdash \phi_e(\underline{n})\ \Rightarrow\ f(n) \notin Th_T$$

□

Now, of course, it follows immediately that Th_T is not computable: choose any non-computable c.e. set, such as K, and the result follows from $K \leq_m Th_T$. Therefore:

Second proof of Gödel's Theorem: If T is an ω-consistent extension of N then it follows from the above discussion that T is undecidable. Moreover, if T is computably axiomatized, then it follows by Theorem 8 that T is incomplete.

*

Both proofs above proceed via **decidability**, but that is not actually the way Gödel originally proved his theorem. What he did instead was, roughly speaking, the following:

Third proof of Gödel's Theorem. Let T be the ω-consistent, computably axiomatized extension of N. Suppose $\phi_{Sub}(x, y, z)$ represents the computable function Sub and $\phi_{Proof_T}(x, y)$ represents the computable relation $Proof_T$. Let $\psi(x) \equiv \neg\exists s\ \exists y\ (\phi_{Proof_T}(s, y) \wedge \phi_{Sub}(x, x, y))$. So $\psi(x)$ codes the statement "$T \nvdash \phi_x(\underline{x})$". Let e be the Gödel number of ψ.

Then if $T \vdash \phi_e(\underline{e})$, there is a proof of $\phi_e(\underline{e})$ from T, so there is an s such that $Proof_T(s, Sub(e, e))$. Then $N \vdash \exists y\ (\phi_{Proof_T}(\underline{s}, y) \wedge \phi_{Sub}(\underline{e}, \underline{e}, y))$ and hence $N \vdash \exists s\ \exists y\ (\phi_{Proof_T}(s, y) \wedge \phi_{Sub}(\underline{e}, \underline{e}, y))$, i.e. $N \vdash \neg\psi(\underline{e})$. Then also $T \vdash \neg\psi(\underline{e})$. But $\psi = \phi_e$, so $T \vdash \neg\phi_e(\underline{e})$. Since T is consistent, $T \nvdash \phi_e(\underline{e})$—contradiction.

On the other hand, suppose $T \vdash \neg\phi_e(\underline{e})$, then, since T is consistent, $T \nvdash \phi_e(\underline{e})$. Then for all s, $Proof_T(s, Sub(e, e))$ is false. Therefore, for all s, $N \vdash \neg[\exists y\ (\phi_{Proof_T}(\underline{s}, y) \wedge \phi_{Sub}(\underline{e}, \underline{e}, y))]$, and so T proves the same. Now by ω-consistency of T, $T \nvdash \exists s\ \exists y\ (\phi_{Proof_T}(s, y) \wedge \phi_{Sub}(\underline{e}, \underline{e}, y))$, i.e., $T \nvdash \neg\psi(\underline{e})$, i.e., $T \nvdash \neg\phi_e(\underline{e})$—contradiction.

Therefore $T \nvdash \phi_e(\underline{e})$ and $T \nvdash \neg\phi_e(\underline{e})$, so T is incomplete. □

The above proof can be stated somewhat colloquially which, though techanically imprecise, may give a better impression of the essence of the proof:

- Let e be the Gödel number of the formula $\phi(x) \equiv$"$T \nvdash \phi_x(x)$".
- Then, in particular, $\phi_e(e) \equiv$"$T \nvdash \phi_e(e)$".
- Now, if $T \vdash \phi_e(e)$ holds "in reality", then by **representability** its coded version holds in T, so $T \vdash$"$T \vdash \phi_e(e)$", so $T \vdash \neg\phi_e(e)$, so by **consistency**, $T \nvdash \phi_e(e)$.
- And if $T \vdash \neg\phi_e(e)$ holds, then by **consistency** $T \nvdash \phi_e(e)$, so by **representability** $T \vdash$"s is not the code of the proof '$T \vdash \phi_e(e)$' ", for all $s \in \mathbb{N}$. Then by ω-**consistency**, $T \nvdash$"$T \vdash \phi_e(e)$", so $T \nvdash \neg\phi_e(e)$.

.