Philosophical Communications, Web Series, No. 35, pp. 155-162
Dept. of Philosophy, Göteborg University, Sweden
ISSN 1652-0459

Kvantifikator för en Dag

Essays dedicated to Dag Westerståhl on his sixtieth birthday

A proof of Gödel's theorem

Per Lindström

Let **N** = (N, +, ·, 1) be the standard model of arithmetic and let Th(**N**), the theory of **N**, be the set of sentences of first-order arithmetic true in **N**. Let c.e. be short for "computably enumerable". (We apply the notions of computability theory to sets of, relations between, etc. syntactic expressions. The natural number n is identified with the sequence 1^{n+1} of of n+1 1's.)

The following result is a form of Gödel's (first incompleteness) theorem.

Theorem 1. Th(**N**) is not c.e.

In this note we give a quite simple proof of this result, a straightforward combination of (more than half a century) old ideas.

We begin by sketching a proof of the Church-Turing theorem that first-order logic is undecidable.

Let X be a finite set of symbols, e.g. X = {0,1,h,q} (as in Lemma 1, below). An X-*word* is a finite string of (occurrences of) members of X. uv is the concatenation of the words u and v. It is convenient to assume that there is an *empty* word ø: uø = øu = u for all u. Let X* be the set of X-words (including ø).

A *production* (over X) is an expression of the form

$$\alpha u \beta \rightarrow \alpha v \beta.$$

Here α, β are variables and $u,v \in X^*$. The expression $w_0 \rightarrow w_1$ is an *instance* of this production if there are $w,w' \in X^*$ such that $w_0 = wuw'$ and $w_1 = wvw'$. A *combinatorial system* C (over X) is determined by a finite set of productions (over X) and an *initial word* (axiom) $w^i \in X^*$. A word w is *derivable* in C, C |– w, if there is a sequence $w_0,...,w_m$ of X-

words such that (a) $w_0 = w^i$, (b) for $k < m$, $w_k \rightarrow w_{k+1}$ is an instance of a production of C, and (c) $w_m = w$.

We borrow the following result from computability theory (cf. Davis (1958), Kleene (1952)); it is essentially just another way of saying that there is a Turing Machine whose halting problem is (computably) unsolvable.

Lemma 1. There is a finite set of productions

(P) $\alpha u_i \beta \rightarrow \alpha v_i \beta, \quad i \leq p,$

over {0,1,h,q} such that the following holds. For every n, let C_n be the combinatorial system with these productions and whose initial word is $hq1q1^{n+1}h$. Then

(i) the set $\{n: C_n \vdash 0\}$ is not computable,

(ii) $C_n \vdash 0$ iff $\{w: C_n \vdash w\}$ is finite.

Since $\{n: C_n \vdash 0\}$ is c.e., it follows that $\{n: \text{not } C_n \vdash 0\}$ is not c.e.

Lemma 1 can be applied to prove the undecidability of first-order logic as follows; the proof is essentially Turing's original proof.

Let F be a one-place predicate, o a two-place function symbol, and let ø, 0, 1, h, q be individual constants. (The interpretation of o we have in mind is concatenation). For $w = s_0 s_1 ... s_n$, where the s_i are single symbols, let $[w] := (s_0 o (s_1 o (s_2 o ... o s_n)...)$; $[s_i] := s_i$.

Let the productions (P) be as in Lemma 1. For $i \leq p$, let

$$\pi_i(x,y) := \exists zz'(x = (z \circ [u_i]) \circ z' \wedge y = (z \circ [v_i]) \circ z').$$

Next, let θ be the conjunction of the following sentences:

$$\forall xyz((x \circ y) \circ z = x \circ (y \circ z)),$$

$$\forall x(x \circ ø = ø \circ x = x),$$

$$\forall xy(Fx \wedge \bigvee\{\pi_i(x,y): i \leq p\} \rightarrow Fy).$$

Let $t(x) := [hq1q] \circ (x \circ h)$. Let $\chi(x) :=$

$$\theta \wedge Ft(x) \rightarrow F0.$$

Finally, let $\underline{n} := [1^{n+1}]$.

It is now not difficult to see that for every n, $\chi(\underline{n})$ is valid iff for every set Y of X-words, if $t(\underline{n}) \in Y$ and Y is closed under the productions (P), then $0 \in Y$. And so

$\chi(\underline{n})$ is valid iff $C_n \vdash 0$.

From this and Lemma 1(i) it follows that the set of valid sentences of {F, o, ø, 0, 1, h, q} is not computable, as desired.

Modifying this proof we can now prove the following result of Trakhtenbrot (cf. Ebbinghaus *et al*. (1984)). A sentence φ is *finitely valid* if it is true in all finite models.

Theorem 2. There is a finite language L such that the set of finitely valid sentences of L is not c.e.

Proof (sketch)**.** Let L = {F, o , ø, 0, 1, h, q, e}, where F, o , ø, 0, 1, h, q are as above and e is an individual constant. Let $\sigma(x)$:=

$x = 0 \vee x = 1 \vee x = h \vee x = q$.

Let θ^+ be the conjunction of θ, above, and the following sentences:

$\forall x(x \circ y = ø \rightarrow x = ø)$,

$\forall x(x \circ e = e)$,

$\forall xyz(x \circ y = x \circ z \neq e \rightarrow y = z)$,

$\forall xyzu(\sigma(x) \wedge \sigma(y) \wedge x \circ z = y \circ u \neq e \rightarrow x = y)$,

¬Fe and "ø, 0, 1, h, q, e are all different".

As before let t(x) := [hq1q] o (x o h). Let $\chi^+(x)$:=

$\theta^+ \wedge Ft(x)$.

Then

(+) $\chi^+(\underline{n})$ has a finite model iff $C_n \vdash 0$.

First, suppose not $C_n \vdash 0$. Let **A** be a model of $\chi^+(\underline{n})$. By Lemma 1(ii), {w: $C_n \vdash w$} is infinite. Moreover, $[u] \neq [v]$ is true in **A** whenever $C_n \vdash u$, $C_n \vdash v$, and $u \neq v$. It follows that $F^{\mathbf{A}}$ is infinite, and so **A** is infinite, as desired.

Next, suppose $C_n \vdash 0$. By Lemma 1(ii), $\{w: C_n \vdash w\}$ is finite. For every word w, let $|w|$ be the length of w; $|\text{ø}| = 0$. Let $r = \max\{|w|: C_n \vdash w\}$. Let $A' = \{w \in \{0,1,h,q\}^*: |w| \leq r\}$. Let $A = A' \cup \{e\}$. Let $F^{\mathbf{A}} = \{w: C_n \vdash w\}$. For $u, v \in A$, let

$$u \circ^{\mathbf{A}} v = uv \text{ if } uv \in A',$$

$$= e \text{ otherwise.}$$

Finally, let $\mathbf{A} = (A, F^{\mathbf{A}}, \circ^{\mathbf{A}}, \text{ø}, 0, 1, h, q, e)$. Then $\mathbf{A}$ is a finite model of $\chi^+(\underline{n})$.

This proves (+). By (+), for every n, $\neg\chi^+(\underline{n})$ is finitely valid iff not $C_n \vdash 0$. It follows that the set of finitely valid sentences of L is not c.e., as desired. ■

Function symbols and individual constants can be replaced by predicates and (universally quantified) variables, respectively, preserving (finite) validity. Thus, Theorem 2 has the following:

Corollary 1. There is a finite set L of predicates such that the set of finitely valid sentences of L is not c.e.

Actually, this holds with $L = \{P\}$, where P is a two-place predicate.

We can now derive Theorem 1 from Corollary 1 as follows.

The following number-theoretic lemma is essentially Gödel's lemma on the so-called β-fuction; it is proved in the same way (cf. Davis (1958), Kleene (1952)).

Lemma 2. For all m, n, and k_i, $i \leq n$, there are r, s such that for all $i \leq n$ and all $k \leq m$, $k = k_i$ iff $\exists q \leq r(r = q(1+(i+1)s) + k)$.

Lemma 3. Let R be any two-place relation on $\{k: k \leq m\}$. There are then numbers n, r, s, r', s' such that

$$R = \{(k,k'): k, k' \leq m \;\&\; \exists i \leq n(\exists q \leq r(r = q(1+(i+1)s + k) \;\&\; \exists q' \leq r'(r' = q'(1+(i+1)s' + k'))\}.$$

Proof. Let n, k_i, k_i' be such that $R = \{(k_i,k_i'): i \leq n\}$. Let r, s be as in Lemma 2 and let r', s' be as in that lemma with k_i replaced by k_i'. ■

Clearly, there is a result similar to Lemma 3 for n-place relations for every n.

For simplicity we now add the two-place predicate ≤ to the language of arithmetic. Let **N'** = (**N**, ≤). Clearly, Th(**N**) is c.e. iff Th(**N'**) is.

In what follows let $\exists x{\leq}y\varphi := \exists x(x \leq y \wedge \varphi)$ and $\forall x{\leq}y\varphi := \forall x(x \leq y \rightarrow \varphi)$.

Lemma 4. For any sentence φ containing no function symbols or constants, we can construct an arithmetical sentence φ* such that φ* is true in **N'** iff φ is finitely valid.

Proof. We explicitly deal only with the case where φ contains only one two-place predicate P; the extension to the general case is straightforward. We assume that the variables y, y', z, z', u, u', v, v', w do not occur in φ. Let φ'(u) be the result of relativizing all quantifiers in φ to "≤ u", i.e., replacing ∃x by ∃x≤u and ∀x by ∀x≤u. Let ρ(y,z,y',z',u',x,x') :=

$$\exists w{\leq}u'(\exists v{\leq}y(y = v{\cdot}(1+(w+1){\cdot}z + x) \wedge \exists v'{\leq}y'(y' = v'{\cdot}(1+(w+1){\cdot}z' + x'))$$

(compare Lemma 3).

Next, replace Pxx', for any variables x, x', everywhere in φ'(u) by ρ(y,z,y',z',u',x,x'). Let φ''(y,z,y',z',u,u') be the result.

Now let m be any number and let R be any two-place relation on {k: k ≤ m}. Let n, r, s, r', s' be as in Lemma 3. Then

$$R = \{(k,k'): k, k' \leq m \;\&\; \rho(r,s,r',s',m,k,k') \text{ is true in } \mathbf{N}'\}.$$

It follows that

φ is true in ({k: k ≤ m}, R) iff φ''(r,s,r',s',m,n) is true in **N'**.

Finally, let φ* :=

$$\forall yy'zz'uu'\varphi''(y,z,y',z',u,u').$$

By Lemma 3, if φ* is true in **N'**, then φ is finitely valid. The converse implication follows, since all quantifiers of φ'(u) are relativized to "≤ u". ■

Proof of Theorem 1. The function mapping φ on φ* as in Lemma 4 is computable. Suppose Th(**N**) is c.e. Then so is Th(**N'**). But then, by Lemma 4, the set of finitely valid sentences would be c.e., contradicting Corollary 1. It follows that Th(**N**) is not c.e. ■

An arithmetical formula φ is *bounded* if every quantifier expression in φ is bounded, i.e., of the form $\forall x \leq y$ or $\exists x \leq y$, where x, y are any variables. φ is *essentially universal* (*e.u.*) (or Π_1) if it is of the form $\forall x_1 \ldots x_n \psi$, where ψ is bounded. (Here we may take n to be 1, since, if x does not occur in ψ, then $\forall x_1 \ldots x_n \psi$ is equivalent to $\forall x \forall x_1 \leq x \ldots \forall x_n \leq x \psi$.)

The formula φ" defined in the proof of Lemma 4 is bounded and so φ* is e.u. Thus, (the proof of) Theorem 1 has the following:

Corollary 2. The set of e.u. sentences true in **N**' is not c.e.

Let PA be Peano Arithmetic. False e.u. sentences are disprovable in PA (in fact, in a very weak finite fragment of PA). Thus, if T is a consistent c.e. extension of PA, every e.u. sentence provable in T is true. Also, the set of e.u. sentences provable in T is c.e. Thus, from Corollary 2 we get:

Corollary 3. If T is a consistent c.e. extension of PA, there is a true e.u. sentence not provable in T.

An example of a true e.u. sentence not provable in T can be obtained as follows. Let W_0, W_1, W_2, ... be the standard enumeration of the c.e. sets. Let $K = \{n: n \in W_n\}$. Then K is c.e. There are productions (P), as in Lemma 1, such that $K = \{n: C_n \vdash 0\}$. (This is true of every c.e. set in place of K.) In the proof of Theorem 1 we effectively construct for every n an e.u. sentence ψ_n such that $C_n \vdash 0$ iff ψ_n is false. It follows that

(1) $n \in W_n$ iff ψ_n is false.

$\{n: T \vdash \psi_n\}$ is c.e. In proving this you actually compute a number e such that

(2) $n \in W_e$ iff $T \vdash \psi_n$.

From (1), (2) it follows that

(3) $T \vdash \psi_e$ iff ψ_e is false.

Now suppose T |– ψ_e. Then ψ_e is false. But then, ψ_e being e.u., T |– $\neg\psi_e$. And so T is inconsistent, contrary to assumption. Thus, not T |– ψ_e. It follows that ψ_e is true. And so ψ_e is as desired.

To obtain a proof of Gödel's second incompleteness theorem from this we have to translate the syntax of first-order arithmetic and computability theory into the language {+, ·, 1} of first-order arithmetic and then show that the translation of (3) is provable in PA.

Another result, essentially equivalent to Theorem 1, which is particularly easy to prove on the present approach, is the following theorem.

Let HF be the set of hereditarily finite sets, i.e., finite sets whose members are finite, whose members of members are finite, etc. Let **HF** = (HF,∈).

Theorem 3. Th(**HF**) is not c.e.

Proof. Let φ be any sentence as in the proof of Theorem 1. Suppose the variables u, v do not occur in φ. Let φ'(u) be the result of relativizing all quantifiers in φ to "∈ u". As usual let (x,y) = {{x},{x,y}}. Let φ''(u,v) be obtained from φ'(u) by replacing Pxy by (x,y)∈ v. Let φ* := ∀uvφ''(u,v). It is then clear that φ* is true in **HF** iff φ is finitely valid. And so Th(**HF**) is not c.e. ■

Per Lindström
Göteborg University
Gatan 3
SE-448 50 Tollered
Sweden
per.lindstrom@mbox334.swipnet.se

References

Davis, M. (1958). Computability and Unsolvability, McGraw-Hill.
Ebbinghaus, H.-D., Flum J., Thomas, W. (1984). Mathematical Logic, Springer-Verlag.
Kleene, S. C. (1952). Introduction to Metamathematics, North-Holland.