

**LECTURES ON UNDECIDABILITY AND
INCOMPLETENESS
(Draft of 5/23/09)**

Timothy J. Carlson

Spring 2009

In these notes, we will establish Gödel's incompleteness theorems:

First Incompleteness Theorem. If A is a decidable set of sentences true in the standard model of arithmetic, $(\omega; 0, S, <, +, \cdot, E)$, then $\text{Cn } A$, the theory of logical consequences of A , is not complete.

Second Incompleteness Theorem. The consistency of mathematics cannot be demonstrated by elementary means.

Clearly, both theorems have to be stated more precisely. In particular, for the First Incompleteness Theorem, we need a mathematical analogue of *decidable* and for the Second Incompleteness Theorem, we need to explain what it would mean to prove the consistency of a theory T within T . Dealing with the first issue will involve us in a discussion of *Church's Thesis* which proposes a precise definition of *decidable*. The second issue has a natural solution: for any reasonable mathematical theory T we can find a sentence Con_T which will naturally express the consistency of T though Con_T is a statement about natural numbers. One instance of the Second Incompleteness Theorem is

$$ZFC \not\vdash \text{Con}_{ZFC}$$

where ZFC is Zermelo-Fraenkel Set Theory with the Axiom of Choice (the foundation of present day mathematics).

1 Preliminaries

Our basic reference is

A Mathematical Introduction to Logic by Herbert Enderton (2nd Ed.)

except that for first-order logic:

- We will use connectives \neg , \wedge and \vee (rather than $\bar{\neg}$ and $\bar{\rightarrow}$).
- We will have quantifiers \forall and \exists (rather than only \forall).

As Enderton does, we will require that $=$ always be interpreted as equality (as opposed to our treatment in *Notes on Completeness* where $=$ was treated as any other binary predicate symbol). Large parts of these notes are taken almost directly from Enderton's text.

We will use a more general notion of substitution than Enderton does. Suppose φ is a formula, x_1, \dots, x_n are variables and t_1, \dots, t_n are terms.

$$\varphi(x_1, \dots, x_n | t_1, \dots, t_n)$$

is the formula obtained by simultaneously replacing each free occurrence of x_i by t_i for $i = 1, \dots, n$. The reader should supply a formal definition by recursion. Notice that this operation does not generally agree with substituting t_i for x_i iteratively since that substitution may introduce new free occurrences of x_j for some $j > i$. We will write $\varphi(t_1, \dots, t_n)$ for $\varphi(v_1, \dots, v_n | t_1, \dots, t_n)$ (where v_1, v_2, \dots is the standard enumeration of the variables).

One also defines when t_1, \dots, t_n are substitutable for x_1, \dots, x_n in φ to mean that no occurrence of a variable in some t_i becomes bound after the substitution. Once again, the reader should supply a recursive definition. The version of the substitution lemma below is proved by induction.

As usual, for any formula φ , variables x_1, \dots, x_n and terms t_1, \dots, t_n there is an alphabetic variant φ^* of φ such that t_1, \dots, t_n are substitutable for x_1, \dots, x_n in φ^* .

Substitution Lemma. Assume \mathcal{A} is a structure, t_1, \dots, t_n are terms and x_1, \dots, x_n are variables.

1. Suppose s is an assignment in \mathcal{A} . If t is a term then

$$\bar{s}^{\mathcal{A}}(t(x_1, \dots, x_n | t_1, \dots, t_n)) = \overline{s(x_1, \dots, x_n | \bar{s}^{\mathcal{A}}(t_1), \dots, \bar{s}^{\mathcal{A}}(t_n))}^{\mathcal{A}}(t)$$

2. If φ is a formula such that t_1, \dots, t_n are substitutable for x_1, \dots, x_n in φ then

$$\mathcal{A} \models \varphi(x_1, \dots, x_n | t_1, \dots, t_n)[s] \text{ iff } \mathcal{A} \models \varphi[s(x_1, \dots, x_n | \bar{s}^{\mathcal{A}}(t_1), \dots, \bar{s}^{\mathcal{A}}(t_n))]$$

We will write $\mathcal{N}_{\mathcal{E}}$ for $(\omega, 0, S, <, +, \cdot, E)$. For a sentence σ in $\mathcal{L}_{\mathcal{E}}$, we will often simply say that σ is *true* when σ is true in $\mathcal{N}_{\mathcal{E}}$.

2 Cantor's Theorem

Recall

Cantor's Theorem. For any set A , $\text{card}(A) < \text{card}(\mathcal{P}(A))$.

Cantor's theorem is equivalent to

For any set A , there is no surjection of A onto $\mathcal{P}(A)$.

Which in turn is equivalent to

Cantor's Theorem (2nd version). *For any binary relation R on A , there is a subset of A which is not of the form R_a for any $a \in |A|$.*

Proof. The proof is by *diagonalization*.

Let D be the subset of A defined by

$$a \in D \text{ iff } (a, a) \notin R$$

i.e.

$$a \in D \text{ iff } a \notin R_a$$

Clearly, $D \neq R_a$ for all $a \in A$.

QED

Cantor's diagonalization argument can be adapted to many other situations, as we shall see.

3 Tarski's Theorem

One use of Cantor's diagonalization argument is

Diagonalization Lemma for Definable Sets. Assume \mathcal{A} is a structure for a language \mathcal{L} . For any binary relation R on $|A|$ which is definable in \mathcal{A} , there is a subset of $|A|$ which is not of the form R_n for $n \in |A|$.

Proof. The proof is identical with the proof of Cantor's Theorem (the 2nd version) given above noting that D is definable in \mathcal{A} if R is. **QED**

When studying formal theories mathematically, it is natural to code expressions by natural numbers. Such a coding is referred to as a *Gödel numbering*. We require that a Gödel numbering $\varepsilon \mapsto \#(\varepsilon)$ is 1-1. We would also like it to have the following properties:

1. $\varepsilon \mapsto \#(\varepsilon)$ is computable.
2. The range of $\varepsilon \mapsto \#(\varepsilon)$ is decidable.

Notice that if $\varepsilon \mapsto \#(\varepsilon)$ has properties 1 and 2 then for any set A of expressions, A is effectively enumerable iff $\#A$ is. Hence, A is decidable iff $\#A$ is.

Given a Gödel numbering $\varepsilon \mapsto \#\varepsilon$, define the binary operation sub_1 on ω by

$$sub_1(\#\alpha, n) = \#\alpha(\bar{n})$$

for any formula α and

$$sub_1(a, n) = 0$$

whenever a is not of the form $\#\alpha$ for a formula α (the subscript 1 indicates that the variable v_1 is being substituted for).

Exercise.* Assume $\varepsilon \mapsto \#(\varepsilon)$ satisfies conditions 1 and 2 above. Show that sub_1 is computable.

We will argue later that the graph of any computable function is definable in \mathcal{N}_E , which will imply, by the exercise, that the graph of sub_1 is definable in \mathcal{N}_E . Therefore, the assumption of the contingent version of Tarski's Theorem, below, is natural.

A Contingent Version of Tarski's Theorem. Assume $\varepsilon \mapsto \#(\varepsilon)$ is a Gödel numbering. If sub_1 is definable in \mathcal{N}_E then $\#Th(\mathcal{N}_E)$ is not definable in \mathcal{N}_E .

Proof. Argue by contradiction and assume sub_1 and $\#Th(\mathcal{N}_E)$ are both definable in \mathcal{N}_E . Let θ define $\#Th(\mathcal{N}_E)$ in \mathcal{N}_E . Define the binary relation R on ω by

$$R(a, n) \text{ iff } \mathcal{N}_E \models \theta[sub_1(a, n)]$$

Notice that for any formula α with at most v_1 free

$$\begin{aligned} R(\#\alpha, n) & \text{ iff } \mathcal{N}_E \models \theta[sub_1(\#\alpha, n)] \\ & \text{ iff } \mathcal{N}_E \models \theta[\#\alpha(\bar{n})] \\ & \text{ iff } \mathcal{N}_E \models \alpha(\bar{n}) \\ & \text{ iff } \mathcal{N}_E \models \alpha[n] \end{aligned}$$

Therefore, every subset of ω which is definable in \mathcal{N}_E is of the form R_a . By the previous theorem, R cannot be definable in \mathcal{N}_E . However, letting " $sub_1(v_1, v_2) = v_3$ " be a formula which defines the graph of sub_1 in \mathcal{N}_E we have

$$R(a, n) \text{ iff } \mathcal{N}_E \models \forall v_3 (\text{"sub}_1(v_1, v_2) = v_3" \rightarrow \theta(v_3))[a, n]$$

for all $a, n \in \omega$. This implies that R is definable in \mathcal{N}_E – contradiction. **QED**

In a later section, we will prove Tarski's Theorem which says that for a wide family of Gödel numberings $\#$ we have that $\#Th\mathcal{N}_E$ is not definable in \mathcal{N}_E . To prove this, we will use the theorem above and establish that sub_1 is definable in \mathcal{N}_E .

4 Σ and Δ

Suppose \mathcal{L} is a first-order language which contains the binary relation symbol $<$. The collection of Δ_0 formulas of \mathcal{L} is the set of formulas which is generated from the atomic formulas by \neg, \wedge, \vee and both *universal bounded quantification* and *existential bounded quantification* i.e. if α is in Δ_0 , x is a variable and t is a term not containing x then both $(\forall x < t)\alpha$ and $(\exists x < t)\alpha$ are in Δ_0 . Here we are using $(\forall x < t)\alpha$ to abbreviate the formula $\forall x(x < t \rightarrow \alpha)$ and we are using $(\exists x < t)\alpha$ to abbreviate the formula $\exists x(x < t \wedge \alpha)$. The collection of Σ formulas of \mathcal{L} is the collection of formulas generated from Δ_0 by \wedge, \vee , bounded universal quantification and existential quantification.

Given a structure \mathcal{A} for a language \mathcal{L} and a set of formulas Φ , an n -ary relation R on $|\mathcal{A}|$ is said to be Φ -definable in \mathcal{A} if there is a formula $\varphi \in \Phi$ and a list x_1, \dots, x_n of distinct variables containing the free variables of φ such that φ defines R in \mathcal{A} with respect to x_1, \dots, x_n . We say that R is Δ -definable in \mathcal{A} if it and its complement are both Σ -definable in \mathcal{A} .

A relation R is Σ -definable, respectively Δ -definable, if it is Σ -definable, respectively Δ -definable, in \mathcal{N}_E . A function $f : \omega^k \rightarrow \omega$ is Σ -definable if its graph is.

We will argue later that the Σ -definable functions are a precise mathematical analogue of the computable functions. This is a form of *Church's Thesis*.

Lemma 4.1 *If R is a k -ary relation on ω which is Σ definable and x_1, \dots, x_k are distinct variables then there is a Σ formula which defines R in \mathcal{N}_E with respect to x_1, \dots, x_k .*

Proof. One can use induction on the complexity of the defining formula of R . Another approach begins with choosing a Σ formula φ which defines R with respect to some y_1, \dots, y_k . If no x_i is among y_1, \dots, y_k then $\exists y_1 \cdots \exists y_k (x_1 = y_1 \wedge \cdots \wedge x_k = y_k \wedge \varphi)$ is a Σ formula which defines R with respect to x_1, \dots, x_k . By doing this twice, one obtains a proof for arbitrary x_1, \dots, x_k . **QED**

Lemma 4.2 *The Σ -definable relations are closed under rearranging variables (including adding dummy variables), intersection, union, bounded universal quantification and existential quantification.*

Proof. Straightforward. For example, if $R \subseteq \omega^n$ is defined by the Σ formula φ then the relation $Q \subseteq \omega^n$ given by

$$Q(a_1, \dots, a_n) \quad \text{iff} \quad \text{for all } b < a_i, P(a_1, \dots, a_{i-1}, b, a_{i+1}, \dots, a_n)$$

is defined by $(\forall x < v_i) \varphi(v_1, \dots, v_{i-1}, x, v_{i+1}, \dots, v_n)$. **QED**

Lemma 4.3 *The Σ -definable relations are closed under substitution by Σ -definable functions.*

Proof. By the previous two lemmas. **QED**

We will write $Th_{\Sigma} \mathcal{N}_E$ for $Th \mathcal{N}_E \cap \Sigma$.

Theorem 4.4 *Assume \mathcal{A} is a structure for \mathcal{L}_E . The following are equivalent.*

1. \mathcal{A} is a model of $Th_{\Sigma} \mathcal{N}_E$.
2. \mathcal{A} is a model of all literals in $Th \mathcal{N}_E$ along with the sentences

$$\forall x \ x \neq 0$$

and

$$(\forall x < \overline{n+1})(x = 0 \vee x = \bar{1} \vee \cdots \vee x = \bar{n})$$

for all $n \in \omega$.

3. The standard part of \mathcal{A} is isomorphic to \mathcal{N}_E and is closed downward with respect to $<^{\mathcal{A}}$ i.e. for any $n \in \omega$ and any $a \in |\mathcal{A}|$, $a <^{\mathcal{A}} \bar{n}^{\mathcal{A}}$ iff there exists $m < n$ such that $a = \bar{m}^{\mathcal{A}}$.

Proof. (1 \Rightarrow 2) Trivial.

(2 \Rightarrow 3) Straightforward.

(3 \Rightarrow 1) Let h be the isomorphism of the \mathcal{N}_E with the standard part of \mathcal{A} . Clearly, $h(n) = \bar{n}^{\mathcal{A}}$. Let U be the set of formulas which are upward absolute with respect to h i.e. U is the set of formulas α such that for any assignment s in \mathcal{N}_E

$$\mathcal{N}_E \models \alpha[s] \implies \mathcal{A} \models \alpha[h \circ s]$$

Claim: U contains all atomic formulas, is closed under conjunction, disjunction, bounded universal quantification and existential quantification.

The Claim follows by the induction principal using the homomorphism theorem for the case of atomic formulas. For simplicity, one might first prove the claim for the case when \mathcal{N}_E is a substructure of \mathcal{A} i.e. when h is the identity (notice that one can always reduce to this case by replacing \mathcal{A} with an isomorphic copy).

Since a formula is absolute iff it and its negation are upward absolute, the claim implies that the absolute formulas contain the atomic formulas and is closed under negation, conjunction, disjunction and bounded quantification. Therefore, every Δ_0 formula is absolute. Applying the claim again, every Σ formula is upward absolute. This implies that \mathcal{A} is a model of $Th_{\Sigma}\mathcal{N}_E$. **QED**

5 Decidability and Church's Thesis

Many attempts have been made to define the terms *computable*, *effectively enumerable* and *decidable* precisely with the result that the classes of functions and relations have been shown to be identical. This lead Church to hypothesize that these families correspond to our intuitive notions of computable, effectively enumerable and decidable. We will discuss several of these approaches now, focusing on the notion of *computable function* on the natural numbers.

We first recall a theorem from Enderton's text (Theorem 33H of section 3.3).

Theorem 5.1* *The following three conditions on a function $f : \omega^k \rightarrow \omega$ are equivalent:*

1. f is computable.
2. The graph of f is effectively enumerable.
3. The graph of f is decidable.

5.1 Σ -Definable Functions

Exercises.

1. Show that the collection of Δ_0 formulas and the collection of Σ formulas are each closed under substitution.
2. Show that

$$\mathcal{N}_E \models (\forall x < t)\alpha \text{ iff } \mathcal{N}_E \models \alpha(x|\bar{n}) \text{ for all } n < t^{\mathcal{N}_E}$$

and

$$\mathcal{N}_E \models (\exists x < t)\alpha \text{ iff } \mathcal{N}_E \models \alpha(x|\bar{n}) \text{ for some } n < t^{\mathcal{N}_E}$$

for any formula α , variable x and ground term t such that α has at most x free. Remark: A *ground term* is a term without an occurrence of any variable.

The exercise allows us to compute when Δ_0 sentences are true in \mathcal{N}_E . A finite sequence $\langle \varphi_0, \dots, \varphi_{n-1} \rangle$ of Δ_0 sentences is a *truth verification* if for all $i < n$ one of the following is true

- φ_i is a literal which is true in \mathcal{N}_E .
- *to be continued ...*

Because the quantifiers in a sentence of \mathcal{L}_E which is in Δ_0 have finite bounds when interpreted in \mathcal{N}_E we have

Theorem 5.2* *The set of true (i.e. satisfied in \mathcal{N}_E) Δ_0 sentences is decidable.*

A useful subcollection of the Σ formulas is the collection of Σ_1^0 formulas which have the form

$$\exists x \theta$$

where θ is Δ_0 .

Lemma 5.3 *Any Σ formula φ of \mathcal{L}_E is equivalent in \mathcal{N}_E to a formula φ^* in Σ_1^0 .*

Proof. By the induction principle. **QED**

The natural proof of the lemma shows that φ^* can be found effectively from φ in the lemma. This and the previous theorem imply

Theorem 5.4* *The set of true Σ sentences is effectively enumerable.*

Corollary 5.5* *Every Σ -definable operation on ω is computable.*

Proof. Immediate from the theorem. **QED**

5.2 Turing Computable Functions

A *Turing Machine* is a tuple $M = (Q, \Gamma, \delta, q_0, B)$ where

- Q is a finite set whose elements are the *states* of M .
- Γ is a finite set whose elements are the *tape symbols* of M including the *blank symbol* B and 1 (where $B \neq 1$).
- δ is a partial function from $Q \times \Gamma$ to $Q \times \Gamma \times \{R, L\}$ called the *next move function* of M .
- q_0 is an element of Q which is called the *start state* of M .

We imagine a Turing Machine M computing on an infinite 1-way tape divided into an infinite number of cells each of which contains a tape symbol and all but finitely many of which contain the blank symbol. At every stage of the computation, M is in one of the states from Q and is reading the contents of one of the cells. At the beginning of the computation, M is in the start state q_0 , M is reading the first cell of the tape and the tape contents must be of the form $1^{i_1} B 1^{i_2} B 1^{i_3} \dots B 1^{i_k} B^\infty$ where i_1, \dots, i_k are positive integers. In this case we say M has *input* (i_1, \dots, i_k) . The move function explains how the computation proceeds: if at a stage in the computation M is in state q , reading a cell with contents a and $\delta(q, a)$ is defined and equals (q', a', X) then M replaces the contents of the cell it is reading with a' , changes its state to q' and moves one cell in the direction X (where R stands for right and L stands for left). The computation comes to a halt when either M the next move function is undefined or the first cell is being read and the move function says to move to the left (off the tape). If M comes to a halt and the tape has contents $1^m B^\infty$ then the *output* of the computation is m .

Given a Turing machine M , for each $k \in \omega$, M *computes* a partial function f from ω^k to ω given by $f(i_1, \dots, i_k) = m$ iff M halts on input (i_1, \dots, i_k) and has output m . A partial operation on ω is *Turing computable* if there is a Turing machine which computes it.

5.3 Recursive Functions

We will use c_b^m to denote the m -ary constant operation on ω with value b :

$$c_b^m(a_1, \dots, a_m) = b$$

The collection of c_b^m are called *constant functions*. I_k^m is the m -ary operation on ω given by

$$I_k^m(a_1, \dots, a_m) = a_k$$

The collection of I_k^m are called *projection functions*. The binary function K is the characteristic function of the ordering relation on ω i.e. K takes only the values 0 and 1 and $K(m, n) = 1$ iff $m < n$.

Assume R is an $n + 1$ -ary relation on ω and $a_1, \dots, a_n \in \omega$ such that there exists $b \in \omega$ with $R(a_1, \dots, a_n, b)$. We write $\mu x.R(a_1, \dots, a_n, x)$ for the least such b

Suppose \mathcal{F} is a family of operations on ω .

\mathcal{F} *closed under compositions* if whenever $f \in \mathcal{F}$ is k -ary and $g_1, \dots, g_k \in \mathcal{F}$ are n -ary then the n -ary operation h on ω given by

$$h(a_1, \dots, a_n) = f(g_1(a_1, \dots, a_n), \dots, g_k(a_1, \dots, a_n))$$

is in \mathcal{F} .

\mathcal{F} is *closed under the μ -operator* if whenever $g \in \mathcal{F}$ is $(n + 1)$ -ary and for all $a_1, \dots, a_n \in \omega$ there is $b \in \omega$ such that $g(a_1, \dots, a_n, b) = 0$ then $f \in \mathcal{F}$ where f is the n -ary operation on ω given by

$$f(a_1, \dots, a_n) = \mu x.g(a_1, \dots, a_n, x) = 0$$

Definition 5.6 The family of *recursive functions* is the smallest family of operations on ω which contains the constant functions, the projection functions, $+$, \cdot and K and is closed under composition and the μ -operator.

5.4 Representable Functions

Definition 5.7 Assume T is a satisfiable theory in a language \mathcal{L} including the constant 0 and the unary function symbol S and R is a k -ary relation on ω . Also, suppose φ is a formula all of whose free variables are among x_1, \dots, x_k . We say that φ *represents R in T with respect to x_1, \dots, x_k* if for all $n_1, \dots, n_k \in \omega$

$$\langle n_1, \dots, n_k \rangle \in R \implies T \models \varphi(x_1, \dots, x_k | \overline{n_1}, \dots, \overline{n_k})$$

and

$$\langle n_1, \dots, n_k \rangle \notin R \implies T \models \neg\varphi(x_1, \dots, x_k | \overline{n_1}, \dots, \overline{n_k})$$

Exercise 3. With T , R , φ and x_1, \dots, x_n as in the definition above, we say that φ *weakly represents R in T with respect to x_1, \dots, x_n* if for all $n_1, \dots, n_k \in \omega$

$$R(n_1, \dots, n_k) \iff T \models \varphi(x_1, \dots, x_k | \overline{n_1}, \dots, \overline{n_k})$$

Show that if T is satisfiable then φ represents R in T with respect to x_1, \dots, x_k iff φ weakly represents R in T^+ with respect to x_1, \dots, x_k for every satisfiable theory T^+ extending T .

Definition 5.8 A relation R on ω is *representable in T* if φ represents R in T with respect to x_1, \dots, x_n for some φ and x_1, \dots, x_n . R is *representable* if it is representable in T for some finitely axiomatizable theory T . An operation f on ω is *representable in T* if its graph is representable in T . f is *representable* if the graph of f is representable.

Theorem ?? shows that $Th_{\Sigma}\mathcal{N}_{\mathbb{E}}$ says little about nonstandard elements. We would like a Δ_0 formula which defines the standard elements as well as certain nonstandard elements which share some properties of the standard elements. For our purposes the formula

$$(0 = v_1 \vee 0 < v_1) \wedge (\forall v_2 < v_1)(Sv_2 < v_1 \vee Sv_2 = v_1)$$

which we denote by ρ . Notice that if $\mathcal{A} \models Th_{\Sigma}\mathcal{N}_{\mathbb{E}}$ then any standard element will be in the set defined by ρ and if a is nonstandard and in the set defined by ρ then $\bar{n}^{\mathcal{A}} <^{\mathcal{A}} a$ for all $n \in \omega$.

Theorem 5.9 *Suppose T is a theory in $\mathcal{L}_{\mathbb{E}}$ which contains all true Σ sentence. If R is Δ -definable relation on ω then R is representable in T (in fact, there is a Σ formula which represents R in T).*

Proof. For notational simplicity, assume $R \subseteq \omega$ is Δ -definable. Let $\exists v_2\varphi$ and $\exists v_2\psi$ define R and the complement of R respectively where φ and ψ are Δ_0 . The formula

$$\exists v_2(\rho(v_2) \wedge \varphi(v_1, v_2) \wedge (\forall v_3 < v_2)(\neg\psi(v_1, v_3)))$$

represents R in T .

QED

5.5 λ -Definable Functions

5.6 Church's Thesis

The 5 families of functions described above can be shown to be identical. Church proposed that they capture the notion of computable function:

Church's Thesis has three equivalent forms:

(version 1) An operation $f : \omega^k \rightarrow \omega$ is computable iff it's graph is Σ -definable.

(version 2) A relation on ω is effectively enumerable iff it is Σ -definable.

(version 3) A relation on ω is decidable iff it is Δ -definable.

Exercise 4.* Give an informal proof showing that the three versions of Church's Thesis are equivalent. Hint: Theorem 4.4* implies the right to left direction in each version.

Diagonalization Lemma for Δ -definable Relations. If R is a binary relation on ω which is Δ -definable then there is a Δ -definable subset of ω which is not of the form R_n for any $n \in \omega$.

Proof. Define the diagonalizing set D as in the proof of the Contingent Version of Tarski's Theorem. A fairly easy argument shows that D is Δ -definable. **QED**

Theorem 5.10* (A Version of the Halting Problem) *Assume Church's Thesis. The set of true Σ sentences is not decidable.*

Proof. Argue by contradiction and assume the set of true Σ sentences is decidable. Let $\alpha_0, \alpha_1, \dots$ be an effective enumeration of all Σ formulas of \mathcal{L}_E with at most v_1 free. Define the binary relation R on ω by

$$R(m, n) \text{ iff } \mathcal{N}_E \models \alpha_m(\bar{n})$$

By our assumption, R is decidable. Assuming Church's thesis, R is Δ -definable contradicting the previous lemma. **QED**

6 Arithmetization of Syntax

As we mentioned earlier, we would like to identify formulas with natural numbers by a Gödel numbering. The following exercise shows that our choice is not particularly important given that our choice satisfies the conditions 1 and 2 of section 3.

Exercise 1.* (CT) Assume $\#_1$ and $\#_2$ are Gödel numberings which satisfy conditions 1 and 2. Show that for any set of expressions A

- $\#_1 A$ is Σ -definable iff $\#_2 A$ is Σ -definable.
- $\#_1 A$ is Δ -definable iff $\#_2 A$ is Δ -definable.

There are many ways to set up a Gödel numbering, but the most natural is to begin with a *numbering* of the symbols. Fix a language \mathcal{L} . A *numbering* of \mathcal{L} is a 1-1 function which maps the symbols of \mathcal{L} into ω so that h maps the symbols

$$(\cdot), \neg, \wedge, \vee, \forall, \exists, v_1, v_2, v_3, \dots$$

onto the odd numbers greater than 1 in the order given. Actually, all we really need is that h restricted to these symbols is computable and maps this family of symbols 1-1 onto a decidable set of natural numbers with infinite complement.

A numbering h is *Δ -definable* if both of the relations

$$\{(m, n) \mid m = h(P) \text{ for some } n\text{-place relation symbol } P\}$$

and

$$\{(m, n) \mid m = h(f) \text{ for some } n\text{-place function symbol } f\}$$

are Δ -definable. Notice that if \mathcal{L} has only finitely many predicate and function symbols (we're including constants as 0-place function symbols) then both sets are finite and, hence, Δ -definable. Also, in this case h is clearly computable.

Notice that for a Δ -definable numbering h , it may be the case that neither the set of $h(P)$ for predicate symbols P or the set of $h(f)$ for function symbols f is Δ -definable. Nevertheless, such h will be sufficient for our purposes.

Assuming h is a Δ -definable numbering, we define the *Gödel numbering generated by h* so that for an expression $\varepsilon = s_0 s_1 \cdots s_n$

$$\#(\varepsilon) = \langle h(s_0), \dots, h(s_n) \rangle$$

where we define

$$\langle n_0, \dots, n_k \rangle = p_0^{n_0+1} p_1^{n_1+1} \cdots p_k^{n_k+1}$$

The important properties of this *tupling function* mapping the collection of finite sequences of natural numbers into the natural numbers are

- It is 1-1.
- It is computable.
- It has decidable range.

The last condition implies that the range is Δ -definable by Church's Thesis. We'll give a formal proof of this fact later.

We extend $\#$ to sequences of expressions so that

$$\#(\varepsilon_0 \cdots \varepsilon_n) = \langle \#(\varepsilon_0), \dots, \#(\varepsilon_n) \rangle$$

For $n \in \omega$, we say that n is a *sequence number* if there are $k \in \omega$ and m_0, \dots, m_{k-1} such that $n = \langle m_0, \dots, m_{k-1} \rangle$.

Exercise 2. Assume h is a numbering such that the parameters of \mathcal{L} are mapped to numbers which are not sequence numbers (notice that no odd number greater than 1 is a sequence number).

Define the set of *hyperexpressions* to be the smallest collection containing the symbols of \mathcal{L} that is closed under taking finite sequences. Assume that no symbol of \mathcal{L} is a finite sequence of hyperexpressions.

- a. Show that the set of hyperexpressions is freely generated from the set of symbols by the sequence operations.
- b. Extend h to a function $\#$ defined on all hyperexpressions so that

$$\#(\varepsilon_0 \cdots \varepsilon_n) = \langle \#(\varepsilon_0), \dots, \#(\varepsilon_n) \rangle$$

Show that $\#$ is 1-1.

7 The Main Technical Lemma

In this section we state the main technical lemma that we will need for the sequel. This lemma says that various functions and relations connected with syntax and semantics are either Σ -definable or Δ -definable. Proofs will be given in a later section.

For this section, fix a Δ -definable numbering h for a language \mathcal{L} . Since \mathcal{L} has only finitely many parameters, h is Δ -definable.

Main Technical Lemma

1. sub_1 is Σ -definable.
2. There is a Σ -definable function $impl$ such that

$$impl(\#\alpha, \#\beta) = \#(\alpha \rightarrow \beta)$$

for all formulas α and β .

3. There is a Σ definable function neg such that

$$neg(\#\alpha) = \#(\neg\alpha)$$

for any formula α .

4. $\#A$ is Δ -definable where A is the set of sentences.
5. If A is a set of sentences such that $\#A$ is Σ -definable then $\#Cn A$ is Σ -definable.

Exercise 1.* Give an informal proof of the Main Technical Lemma assuming Church's Thesis. Hint: Show that $\#$ satisfies conditions 1 and 2 of section 3.

8 Tarski's Theorem and Gödel's First Incompleteness Theorem

In this section, we will prove Tarski's Theorem and Gödel's First Incompleteness assuming the Main Technical Lemma.

For this section, fix a Δ -definable numbering h for \mathcal{L}_E .

By the contingent version of Tarski's Theorem in Section 3 and part 1 of the Main Technical Lemma, we have

Tarski's Theorem $\#Th \mathcal{N}_E$ is not definable in \mathcal{N}_E .

The theorem immediately the following corollaries.

Corollary 8.1 $Th \mathcal{N}_E$ is not Δ -definable.

Corollary 8.2* Assume Church's Thesis. $Th \mathcal{N}_E$ is not decidable.

Corollary 8.3* Assume Church's Thesis. $Th \mathcal{N}_E$ is not axiomatizable.

The corollaries are more properly viewed as a versions of Gödel's First Incompleteness Theorem. The fact that it follows simply from Tarski's Theorem is due to our choice of Δ -definability as the formal analogue of decidable. If we had chosen one of the other approaches, we would need a more elaborate argument e.g. we could show that the other approach is equivalent to using Δ -definability.

Gödel's First Incompleteness Theorem If A is a set of sentences which are true in \mathcal{N}_E and $\#A$ is Σ -definable then $Cn A$ is not a complete theory.

Proof. Argue by contradiction and assume A is a set of true sentences such that $\#A$ is Σ -definable but with the property that $Cn A$ is complete. By part 5 of the Main Technical Lemma, $\#Cn A$ is Σ -definable. Since $Cn A$ is complete and contained in $Th \mathcal{N}_E$, $Cn A = Th \mathcal{N}_E$. Therefore, $\#Th \mathcal{N}_E$ is definable in \mathcal{N}_E contradicting Tarski's Theorem. **QED**

The First Incompleteness Theorem leaves open some interesting questions.

If we focus on $Th \mathcal{N}_E$, we know that it has no decidable set of axioms but we can ask how simple a set of axioms might be. We will see later that we can't do much better than the complexity of the theory itself since $\#A$ is not definable in \mathcal{N}_E for any set A of axioms for $Th \mathcal{N}_E$.

If we focus on the notion of decidability rather than the theory $Th \mathcal{N}_E$ itself, we might ask if there are interesting subtheories T of $Th \mathcal{N}_E$ such that $\#T$ is Δ -definable. We will see this is not the case: there is no subtheory T of $Th \mathcal{N}_E$ such that $\#T$ is Δ -definable.

We will close this section with a related result which implies much of the result just described but doesn't require as many technical details (one only needs the lemma which says that sub_1 is Σ -definable).

We first give a formal version of the undecidability of a the version of the Halting Problem discussed previously.

Theorem 8.4 *If A is the set of true Σ sentences then $\#A$ is not Δ -definable.*

Proof. Argue by contradiction and assume $\#A$ is Δ -definable where A is the set of true Σ sentences. Define the binary relation R on ω by

$$R(m, n) \text{ iff } sub_1(m, n) \in \#A$$

Our assumption implies that R is Δ -definable. Moreover, every Δ -definable subset of ω is of the form R_m . This contradicts the Diagonalization Lemma for Δ -Definable Relations in Section 4. **QED**

Theorem 8.5 *If T is a satisfiable theory in \mathcal{L}_E which contains all true sentences in Σ then $\#T$ is not Δ -definable.*

Proof. Argue by contradiction and assume the theorem is false. By Theorem 5.9, every Δ -definable relation is representable in T . Arguing as in the previous theorem, we can define a binary relation R on ω which is Δ -definable such that every Δ -definable subset of ω has the form R_n for $n \in \omega$ – a contradiction. **QED**

The following is a common generalization of the previous two theorems.

Theorem 8.6 *Assume A is a set of Σ sentences containing all true Σ sentences but disjoint from the collection of Σ sentences σ such that $Th_{\Sigma} \mathcal{N}_E \models \neg \sigma$. $\#A$ is not Δ -definable.*

Proof. Similar to the proof of the previous theorem replacing T by $CnTh_{\Sigma}\mathcal{N}_{\mathbb{E}}$.
QED

Lemma 8.7 *Assume T is a complete theory in a language \mathcal{L} with a Δ -definable numbering. If $\#T$ is Σ -definable then $\#T$ is Δ -definable.*

Proof. Since T is complete, for a sentence α we have $\alpha \notin T$ iff $\neg\alpha \in T$. So for $n \in \omega$, $n \notin \#T$ iff either n is not the Gödel number of a sentence or $neg(n) \in \#T$. By Main Technical Lemma 3, $\#T$ has a Σ -definable complement. **QED**

Theorem 8.8 *Assume T is a theory in the language $\mathcal{L}_{\mathbb{E}}$ containing all true Σ sentences. If $\#T$ is Σ -definable then $\#T$ is not complete.*

Proof. By the previous lemma and Theorem ?? **QED**

9 A Subtheory of Number Theory

In this section we study a finite set of axioms named $A_{\mathbb{E}}$. The crucial properties of $A_{\mathbb{E}}$ are

1. Every sentence in $A_{\mathbb{E}}$ is true in $\mathcal{N}_{\mathbb{E}}$.
2. $A_{\mathbb{E}}$ is finite.
3. For every sentence σ in Σ ,

$$\mathcal{N}_{\mathbb{E}} \models \sigma \implies A_{\mathbb{E}} \models \sigma$$

Let $A_{\mathbb{E}}$ be the set of sentences consisting of

- (L1) $\forall x \ x \not< 0$
- (L2) $\forall x \forall y \ (x < \mathbf{S}y \leftrightarrow x < y \vee x = y)$
- (A1) $\forall x \ x + 0 = x$
- (A2) $\forall x \forall y \ x + \mathbf{S}y = \mathbf{S}(x + y)$
- (M1) $\forall x \ x \cdot 0 = 0$
- (M2) $\forall x \forall y \ x \cdot \mathbf{S}y = x \cdot y + x$
- (E1) $\forall x \ x \mathbf{E}0 = \mathbf{S}0$
- (E2) $\forall x \forall y \ x \mathbf{E} \mathbf{S}y = (x \mathbf{E} y) \cdot x$

Clearly, the elements of $A_{\mathbb{E}}$ are true in $\mathcal{N}_{\mathbb{E}}$.

Exercise 1. Show that $\forall x \ \mathbf{S}x \neq 0$ follows from $A_{\mathbb{E}}$.

Exercise 2.* Find a nonstandard model of $A_{\mathbb{E}}$ which is decidable.

Theorem 9.1 *Assume \mathcal{A} is a model of A_E .*

1. *For $n \in \omega$, for any $a \in |\mathcal{A}|$, $a <^{\mathcal{A}} \bar{n}^{\mathcal{A}}$ iff $a = \bar{m}^{\mathcal{A}}$ for some $m \in \omega$ with $m < n$.*
2. *If $h : \omega \rightarrow |\mathcal{A}|$ is defined by*

$$h(n) = \bar{n}^{\mathcal{A}}$$

then h is an embedding of \mathcal{N}_E into \mathcal{A} .

Proof. Part 1 can be proved by induction on n using the fact that (L1) and (L2) are true in \mathcal{A} .

For part 2, first notice that h preserves S . Part 1 implies that h is 1-1 and preserves $<$. We next prove h preserves $+$, \cdot and E in that order.

To see that h preserves $+$, first fix $n \in \omega$. Now show that $h(n + m) = h(n) +^{\mathcal{A}} h(m)$ by induction on $m \in \omega$ using the fact that (A1) and (A2) are true in \mathcal{A} .

Similar arguments hold for \cdot and E .

QED

Theorem 9.2 $A_E \models Th_{\Sigma} \mathcal{N}_E$

Proof. By the previous theorem and Theorem 4.9.

QED

10 The Strong Undecidability of A_E and Church's Theorem

A satisfiable set of sentences A is *strongly undecidable* if there is no theory T with $T \cup A$ satisfiable such that $\#T$ is Δ -definable.

Lemma 10.1 *Assume T is a theory and A is a finite set of sentences. If $\#T$ is Δ -definable then $\#Cn(T \cup A)$ is Δ -definable.*

Proof. For any $n \in \omega$, $n \in \#Cn(T \cup A)$ iff either $n \notin \#(\text{sentences})$ or $\text{imp}(\#\alpha, n) \in \#T$ where α is the conjunction of the elements of A .

QED

Theorem 10.2 A_E is strongly undecidable.

Proof. Argue by contradiction and assume T is a theory such that $T \cup A_E$ is satisfiable and $\#T$ is Δ -definable. By Lemma 9.1, we may assume that T contains A_E and, hence $Th_{\Sigma} \mathcal{N}_E$. This contradicts Theorem ??.

QED

Corollary 10.3 *If T is a theory in \mathcal{L}_E such that $T \cup A_E$ is satisfiable and $\#T$ is Σ -definable then T is incomplete.*

Proof. Immediate from the previous theorem and Lemma ??.

QED

Corollary 10.4 Church's Theorem *The set of Gödel numbers of valid sentences in \mathcal{L}_E is not Δ -definable.*

Proof. Immediate from the theorem.

QED

One can ask for which finite languages Church's Theorem holds. One condition which suffices is that the language have an n -ary relation symbol for some $n \geq 2$.

11 More Properties of Σ -Definable and Δ -definable Relations and Functions

Lemma 11.1 *The collection of Δ -definable relations contains the collection of Δ_0 -definable relations, including $=$ and $<$, and is closed under complement, intersection, union, bounded universal quantification and bounded existential quantification.*

Proof. By lemma ??.

QED

We can now use Lemmas 10.2?? and 10.3?? to establish that various relations are Δ -definable rather than explicitly talk about formulas.

Lemma 11.2 *If a function is Σ -definable then it is Δ -definable.*

Proof. Assume $f : \omega^k \rightarrow \omega$ is Σ -definable. We need to see that the complement of the graph of f is Σ -definable. This follows from the fact that $f(a_1, \dots, a_k) \neq b$ iff there exists $c \in \omega$ such that $f(a_1, \dots, a_k) = c$ and $b \neq c$.

QED

Lemma 11.3 *The Σ -definable functions include all the constant functions, the projection functions, the successor function, the predecessor function, addition, multiplication and exponentiation and are closed under compositions and the μ -operator.*

Proof. That the specific functions mentioned are Σ -definable is straightforward e.g. I_i^n is defined by $v_{n+1} = v_i$.

To verify closure under compositions, assume $h : \omega^k \rightarrow \omega$ and $g_i : \omega^n \rightarrow \omega$ ($i = 1, \dots, k$) are Σ -definable. Let $f : \omega^n \rightarrow \omega$ be given by $f(a_1, \dots, a_n) = h(g_1(a_1, \dots, a_n), \dots, g_k(a_1, \dots, a_n))$. We have $f(a_1, \dots, a_n) = b$ iff there are $c_1, \dots, c_k \in \omega$ such that $g_i(a_1, \dots, a_n) = c_i$ for $i = 1, \dots, k$ and $h(c_1, \dots, c_k) = b$.

To verify closure under the μ -operator, assume $g : \omega^{n+1} \rightarrow \omega$ and for all $a_1, \dots, a_n \in \omega$ there is $b \in \omega$ such that $g(a_1, \dots, a_n, b) = 0$. Define $f : \omega^n \rightarrow \omega$ by $f(a_1, \dots, a_n) = \mu b. g(a_1, \dots, a_n, b) = 0$. We see that $f(a_1, \dots, a_n) = b$ iff $g(a_1, \dots, a_n, b) = 0$ and for all $c < b$, $g(a_1, \dots, a_n, c) \neq 0$. Since g is Δ -definable by the previous lemma, we see that f is Σ -definable.

QED

Lemma 11.4 *A relation is Δ -definable iff its characteristic function is Σ -definable.*

Proof. By Lemma 10.4??.

QED

Lemma 11.5 *The Σ -definable functions are closed under definition by Δ -definable cases i.e. if $f_i : \omega^k \rightarrow \omega$ is Σ -definable for $i = 1, \dots, n$ and $R_i \subseteq \omega^k$ is Δ -definable for $i = 1, \dots, n$ then $f : \omega^k \rightarrow \omega$ is Σ definable where $f(a_1, \dots, a_k) = f_i(a_1, \dots, a_k)$ whenever $R_i(a_1, \dots, a_k)$.*

Proof. $f(a_1, \dots, a_k)$ is given by

$$f_1(a_1, \dots, a_k) \cdot \text{Char}_{R_1}(a_1, \dots, a_k) + \dots + f_n(a_1, \dots, a_k) \cdot \text{Char}_{R_n}(a_1, \dots, a_k)$$

QED

12 Talking about Sequences in Arithmetic

In this section we develop the results about sequences needed to handle syntax.

Recall the definition

$$\langle n_0, \dots, n_{k-1} \rangle = p_0^{n_0+1} p_1^{n_1+1} \dots p_{k-1}^{n_{k-1}}$$

where p_k is the k^{th} prime number. We interpret the case $k = 0$ to mean $\langle \rangle = 1$. A number is a *sequence number* if it is of the form $\langle n_0, \dots, n_{k-1} \rangle$. We define several auxiliary functions. $lh(\langle n_0, \dots, n_{k-1} \rangle) = k$ and $lh(a) = 0$ if a is not a sequence number. $(\langle n_0, \dots, n_{k-1} \rangle)_i = n_i$ if $i < k$, $(\langle n_0, \dots, n_{k-1} \rangle)_i = 0$ if $k \leq i$ and $(a)_i = 0$ if a is not a sequence number. $\langle n_0, \dots, n_{k-1} \rangle | i$ is $\langle n_0, \dots, n_{i-1} \rangle$ if $i < k$, $\langle n_0, \dots, n_{k-1} \rangle | i = \langle n_0, \dots, n_{k-1} \rangle$ if $k \leq i$ and $a | i = 0$ if a is not a sequence number. $\langle n_0, \dots, n_{k-1} \rangle * \langle m_0, \dots, m_{j-1} \rangle = \langle n_0, \dots, n_{k-1}, m_0, \dots, m_{j-1} \rangle$ and $a * b = 0$ if either a or b is not a sequence number.

Lemma 12.1

1. *The relation consisting of all pairs (a, b) of natural numbers such that a divides b is Δ -definable.*
2. *The set of primes is Δ -definable.*
3. *The set of sequence numbers is Δ -definable.*
4. *The function $n \mapsto p_n$ is Σ -definable.*
5. *The function lh is Σ -definable.*
6. *The function $a, i \mapsto (a)_i$ is Σ -definable.*
7. *The function $a, i \mapsto a | i$ is Σ -definable.*
8. *The function $a, b \mapsto a * b$ is Σ -definable.*

Proof. Except for part 4, straightforward using the lemmas of the previous section.

For part 4, let X be the set of numbers of the form $p_0^1 \cdot p_1^2 \cdots p_k^{k+1}$. Using parts 1 and 2, we see that X is Δ -definable. This implies that the graph of $n \mapsto p_n$ is Σ -definable. **QED**

Definition 12.2 Assume $f : \omega^{k+1} \rightarrow \omega$. Define $\bar{f} : \omega^{k+1} \rightarrow \omega$ by

$$\bar{f}(a, b_1, \dots, b_k) = \langle f(0, b_1, \dots, b_k), \dots, f(a-1, b_1, \dots, b_k) \rangle$$

Lemma 12.3 If $f : \omega^{k+1} \rightarrow \omega$ is Σ -definable then so is \bar{f} .

Proof. Straightforward. **QED**

Lemma 12.4 The family of Σ -definable functions is closed under primitive recursion i.e. if $h : \omega^{k+2} \rightarrow \omega$ is Σ -definable then the function $f : \omega^{k+1} \rightarrow \omega$ is Σ -definable where

$$f(a, b_1, \dots, b_k) = h(\bar{f}(a, b_1, \dots, b_k), a, b_1, \dots, b_k)$$

Proof. First show that \bar{f} is Σ -definable.

To see that f is Σ -definable, notice that $f(a, b_1, \dots, b_k) = (\bar{f}(a+1, b_1, \dots, b_k))_a$.

QED

Exercise. Assume $g : \omega^k \rightarrow \omega$ and $h : \omega^{k+2} \rightarrow \omega$ are Σ -definable. Use the previous lemma to show that $f : \omega^{k+1} \rightarrow \omega$ is Σ -definable where

$$f(0, b_1, \dots, b_k) = g(b_1, \dots, b_k)$$

and

$$f(a+1, b_1, \dots, b_k) = h(f(a, b_1, \dots, b_k), a, b_1, \dots, b_k)$$

for $a \in \omega$.

Define

$$*(\langle n_0, \dots, n_k \rangle) = n_0 * \cdots * n_k$$

$$*(\langle \rangle) = 1$$

and

$$*(a) = 0$$

if a is not a sequence number.

Lemma 12.5 The function $a \mapsto *(a)$ is Σ -definable.

Proof. From the previous exercise and part 8 of Lemma 11.1. **QED**

13 Proofs of the Main Technical Lemmas

We begin by commenting that the relations and functions described in items 1-9 of section 3.4 of Enderton's text are Δ -definable. This includes parts 1 and 4 of the Main Technical Lemma. The proofs are straightforward but somewhat tedious and follow the same lines as in the text. Many of the concepts we have introduced by induction can be dealt with by primitive recursion. For parts 2 and 3 of the Main Technical Lemma, we can take

$$\text{imp}(n, m) = \langle h(\langle \rangle) * n * \langle h(\rightarrow) \rangle * m * \langle h(\langle \rangle) \rangle$$

and

$$\text{neg}(n) = \langle h(\langle \rangle, h(\neg)) * n * \langle h(\langle \rangle) \rangle$$

Part 5 of the Main Technical Lemma is more complicated – it is a formal version of the Completeness Theorem.

Lemma 13.1 *Assume \mathcal{L} is a language with a Δ -definable numbering. $\#V$ is Σ -definable where V is the set of all sequences of formulas $\alpha_0, \dots, \alpha_{k-1}, \beta$ such that $\{\alpha_0, \dots, \alpha_{k-1}\} \models \beta$.*

Proof. We will not give all the details, but we begin by assuming that the reader is aware of an effective sequent calculus for which the completeness theorem can be proved i.e. a sequent has a proof in the calculus iff it is valid. If the completeness theorem is proved properly, the rules of the sequent calculus will be decidable. This means that the set of Gödel numbers of rules should be Δ -definable. To make this precise, we need to explain what the Gödel number of a rule is. Our approach will be such that a rule can have several Gödel numbers. First, $\langle \#\alpha_0, \dots, \#\alpha_{k-1}, \#\beta \rangle$ is a *Gödel number* of the sequent $\{\alpha_0, \dots, \alpha_{k-1}\} \vdash \beta$. Notice that a sequent which has at least one premise has infinitely many Gödel numbers. If S_i ($i = 1, \dots, n-1$) and S are sequents with Gödel numbers a_1, \dots, a_{n-1} and b respectively then the rule with premises S_1, \dots, S_{n-1} and conclusion S will have Gödel number $\langle a_1, \dots, a_{n-1}, b \rangle$. We can now state precisely

Claim: The set of Gödel numbers of rules in our calculus is Δ -definable.

The proof will generally be very tedious but straightforward.

The claim easily implies the lemma. **QED**

Part 5 of the Main Technical Lemma is now straightforward since $\sigma \in Cn(A)$ iff there is a finite list $\alpha_0, \dots, \alpha_{k-1}$ of elements of A such that $\{\alpha_0, \dots, \alpha_{k-1}\} \models \sigma$.

14 A Fixed Point Theorem for A_E

In this section we will prove

Fixed Point Theorem for A_E . For any formula θ with at most v_1 free, there is a sentence σ such that

$$A_E \models \sigma \leftrightarrow \theta(\overline{\#\sigma})$$

Notice that the proof of Tarski's Theorem actually shows that for any formula θ with at most v_1 free there is a sentence σ such that

$$\mathcal{N}_E \models \sigma \leftrightarrow \neg\theta(\overline{\#\sigma})$$

which clearly implies the following weaker version of the result we seek.

Fixed Point Lemma for \mathcal{N}_E . For any formula θ with at most v_1 free, there is a sentence σ such that

$$\mathcal{N}_E \models \sigma \leftrightarrow \theta(\overline{\#\sigma})$$

Can σ be found effectively given θ ? Yes, by looking carefully at the proof of the theorem. We define a binary relation R on ω by

$$R(a, k) \text{ iff } \mathcal{N}_E \models \theta[\text{sub}_1(a, k)]$$

and then define the set D by

$$k \in D \text{ iff } R(k, k)$$

In other words

$$k \in D \text{ iff } \mathcal{N} \models \theta[\text{sub}_1(k, k)]$$

We see that if we let φ be any formula which defines D in \mathcal{N}_E and let $a = \#\varphi$ then

$$\begin{aligned} \mathcal{N} \models \varphi(\bar{a}) & \text{ iff } a \in D \\ & \text{ iff } \mathcal{N}_E \models \theta[\text{sub}_1(a, a)] \\ & \text{ iff } \mathcal{N}_E \models \theta[\overline{\#\varphi(\bar{a})}] \\ & \text{ iff } \mathcal{N}_E \models \theta(\overline{\#\varphi(\bar{a})}) \end{aligned}$$

So σ may be taken to be $\varphi(\bar{a})$ i.e. $\varphi(\overline{\#\varphi})$. Our task is to choose φ so that we have a fixed point for θ in A_E rather than just \mathcal{N}_E . We may assume without loss of generality that v_2 is substitutable for v_1 in θ . Letting $\gamma(v_1, v_2)$ be a formula which defines the function $k \mapsto \text{sub}_1(k, k)$ in \mathcal{N}_E , we could let φ be

$$\forall v_2(\gamma(v_1, v_2) \rightarrow \theta(v_2))$$

or

$$\exists v_2(\gamma(v_1, v_2) \wedge \theta(v_2))$$

This gives us two fixed points in \mathcal{N}_E :

$$\forall v_2(\gamma(\bar{a}_1, v_2) \rightarrow \theta(v_2))$$

where $a_1 = \#\forall v_2(\gamma \rightarrow \theta(v_2))$ and

$$\exists v_2(\gamma(\overline{a_2}, v_2) \wedge \theta(v_2))$$

where $a_2 = \#\exists v_2(\gamma(v_1, v_2) \wedge \theta(v_2))$. Will either of these formulas work for A_E ? Notice that the first would give the left to right implication, but not clearly the right to left implication, while the second would give the right to left but not clearly the left to right. The problem is that the formula γ may not determine a function in all models of A_E . To remedy this we need

Lemma 14.1 *Suppose $f : \omega \rightarrow \omega$ is a Σ -definable operation. There is a Σ formula $\gamma(v_1, v_2)$ which defines f in \mathcal{N}_E such that*

$$A_E \models \forall v_2(\gamma(\overline{n}, v_2) \leftrightarrow v_2 = \overline{f(n)})$$

for any $n \in \omega$.

Proof. Let φ be a formula which represents (the graph of) f in A_E . Let γ be the formula

$$\varphi(v_1, v_2) \wedge (\forall v_3 < v_2)(\neg\varphi(v_1, v_3))$$

Suppose \mathcal{A} is a model of A_E . Also let $n \in \omega$ and let a be an arbitrary element of $|\mathcal{A}|$. We leave it to the reader to show that $\mathcal{A} \models \gamma(\overline{n}, v_2)[v_2|a]$ iff $a = \overline{f(n)}^{\mathcal{A}}$.

QED

We remark that the lemma can be proved with $Th_{\Sigma}\mathcal{N}_E$ in place of A_E .

Proof of the Fixed Point Lemma for A_E . Choosing the formula γ by the lemma above, one can show that either of the choices for φ above leads to a fixed point for θ in A_E . We'll consider the second example of φ .

As above, we let φ be $\exists v_2(\gamma(v_1, v_2) \wedge \theta(v_2))$, let $a = \#\varphi$ and let $\sigma = \varphi(\overline{a})$.

To show that σ is a fixed point for θ in A_E , assume \mathcal{A} is a model of A_E . First notice that

$$\mathcal{A} \models \theta(\overline{\#\sigma}) \text{ iff } \mathcal{A} \models \theta[\overline{\#\sigma}^{\mathcal{A}}]$$

We also have

$$\begin{aligned} \mathcal{A} \models \sigma & \text{ iff } \mathcal{A} \models \exists v_2(\gamma(\overline{a}, v_2) \wedge \theta(v_2)) \\ & \text{ iff } \mathcal{A} \models \gamma(\overline{a}, v_2)[v_2, b] \text{ and } \mathcal{A} \models \theta(v_2)[v_2, b] \text{ for some } b \in |\mathcal{A}| \\ & \text{ iff } b = \overline{sub_1(a, a)}^{\mathcal{A}} \text{ and } \mathcal{A} \models \theta[b] \text{ for some } b \in |\mathcal{A}| \\ & \text{ iff } \mathcal{A} \models \theta[\overline{\#\sigma}^{\mathcal{A}}] \end{aligned}$$

The last equivalence follows from $sub_1(a, a) = sub_1(\#\varphi, \#\varphi) = \#\varphi(\overline{\#\varphi}) = \#\sigma$.

QED

15 The Derivability Conditions

We have seen that for any set of sentences A in a language with a Δ -definable numbering such that $\#A$ is Σ definable, $\#Cn A$ is also Σ -definable. Let Pr_A be a Σ formula which defines $\#Cn A$ in \mathcal{N}_E (Pr_A stands for *provable from A*). There are many ways to choose Pr_A , but the natural approach would be have it clearly say “there is a valid sequent whose conclusion is v_1 and all of whose premises are in A ” and use Main Technical Lemma 3 which says there is a Σ definition of the codes of the valid sequents.

We will write

$$Pr_A\alpha$$

for $Pr_A(\overline{\#\alpha})$ when α is a formula.

We see that for sentences α and β

1. If $A \models \alpha$ then $\mathcal{N}_E \models Pr_A\alpha$.
2. $\mathcal{N}_E \models Pr_A(\alpha \rightarrow \beta) \wedge Pr_A\alpha \rightarrow Pr_A\beta$
3. $\mathcal{N}_E \models Pr_A\alpha \rightarrow Pr_A Pr_A\alpha$ provided A implies any true Σ sentence e.g. A extends A_E .

This leads us to the following *Derivability Conditions* for a set of sentences A in a language \mathcal{L} extending \mathcal{L}_E with a Δ -definable numbering and a choice of Pr_A .

- D1. For any sentence α , if $A \models \alpha$ then $A \models Pr_A\alpha$.
- D2. For any sentences α and β , $A \models Pr_A(\alpha \rightarrow \beta) \wedge Pr_A\alpha \rightarrow Pr_A\beta$.
- D3. For any sentence α , $A \models Pr_A\alpha \rightarrow Pr_A Pr_A\alpha$.

Notice that any A which extends A_E will satisfy D1. While D2 and D3 are fairly mild conditions, they apparently do not hold for A_E . D1-D3 follow if we add restricted families of induction axioms i.e. axioms of the form

$$\forall v_2 \dots \forall v_k (\varphi(0) \wedge \forall v_1 (\varphi(v_1) \rightarrow \varphi(Sv_1)) \rightarrow \forall v_1 \varphi(v_1))$$

and we make a reasonable choice for Pr_A , but we will not prove that fact in these notes.

16 The Second Incompleteness Theorem

Let Con_A to be the formula $\neg Pr_A 0 \neq 0$. If we have chosen Pr_A appropriately, Con_A is a number-theoretic statement, i.e. a sentence in \mathcal{L}_E , which expresses that A is consistent.

Gödel’s Second Incompleteness Theorem Assume A is a set of sentences in the language \mathcal{L}_E extending A_E and Pr_A is chosen so that the derivability conditions hold. If A is satisfiable then

$$A \not\models Con_A$$

Proof. Assume A is satisfiable.

Since A extends A_E , there is a sentence σ such that

$$A \models (\sigma \leftrightarrow \neg Pr_A \sigma)$$

Notice that $A \not\models \sigma$ since otherwise we would have $A \models Pr_A \sigma$ (by D1) and $A \models \neg Pr_A \sigma$ (by choice of σ) implying A is not satisfiable.

We claim that $A + Con_A \models \sigma$, from which the theorem follows. This is equivalent to showing $A + \neg \sigma \models \neg Con_A$, i.e. $A + \neg \sigma \models Pr_A 0 \neq 0$. We have

$$A + \neg \sigma \models Pr_A \sigma$$

by choice of σ . By D3 we have

$$A + \neg \sigma \models Pr_A Pr_A \sigma$$

Applying D1 we have

$$A + \neg \sigma \models Pr_A (\sigma \rightarrow \neg Pr_A \sigma)$$

again using the choice of σ . Since we know $A + \neg \sigma \models Pr_A \sigma$,

$$A + \neg \sigma \models Pr_A (\neg Pr_A \sigma)$$

by D2. Since D1 implies that $A \models Pr_A (\varphi \rightarrow (\neg \varphi \rightarrow 0 \neq 0))$ for any formula φ we have

$$A + \neg \sigma \models Pr_A 0 \neq 0$$

by applying D2 twice. **QED**

The proof of the Second Incompleteness Theorem may seem mysterious. What led to the choice of the fixed point? We'll sketch a proof of the Second Incompleteness Theorem under slightly strengthened assumptions which can lead to the discovery of σ in the proof if analyzed carefully.

We assume A is a set of sentences in the language \mathcal{L}_E which contains A_E and Pr_A is a Σ formula which satisfies the derivability conditions and the following strong version of D3:

For any Σ sentence σ , $A \models \sigma \rightarrow Pr_A \sigma$.

Notice that $\mathcal{N}_E \models \sigma \rightarrow Pr_A \sigma$ for any Σ sentence σ if A contains A_E and Pr_A defines $\#Cn(A)$.

Let X be a maximal collection of Σ sentences such that $A \cup X$ is satisfiable. We will show that $Pr_A 0 \neq 0$ is in X which implies that $A \cup \neg Con_A$ is satisfiable. Argue by contradiction and assume $Pr_A 0 \neq 0 \notin X$. By the derivability conditions, for any sentence σ , in particular any Σ sentence σ , it is not possible that both $Pr_A \sigma$ and $Pr_A (\neg \sigma)$ are in X . An easy argument shows that if σ is a Σ sentence and $\sigma \notin X$ then $A \cup X \models \neg \sigma$ and, by the strengthened derivability conditions (and the Compactness Theorem), $A \cup X \models Pr_A (\neg \sigma)$. Hence

$$A \cup X \models \sigma \leftrightarrow Pr_A \sigma$$

and

$$A \cup X \models \neg\sigma \leftrightarrow Pr_A(\neg\sigma)$$

This implies that

- The complement of any relation on ω which is representable in $A \cup X$ by a Σ formula is representable in $A \cup X$ by a Σ formula.

Define $R \subseteq \omega \times \omega$ by

$$R(m, n) \iff A \cup X \models Pr_A(\overline{sub_1(m, n)})$$

One can show that

- R is representable in $A \cup X$ by a Σ formula.
- Every subset of ω which is representable in $A \cup X$ by a Σ formula is of the form R_a .

We now get a contradiction by noticing that the diagonalizing set for R should be of the form R_a .

The argument above can be modified to prove the version of the Second Incompleteness Theorem given earlier. Rather than let X be a maximal collection of Σ sentence such that $A \cup X$ is satisfiable, let X be a maximal collection of sentences of the form $Pr_A\varphi$ such that $A \cup X$ is satisfiable. One can also give the argument a more model theoretic flavor if one works with a model of $A \cup X$ and considers the collection of relations on ω which are “representable” in the structure.