§9. Gödel's Incompleteness Theorem I: Represening Primitive Recursive Functions

The following three problems can be considered the basic problems in the foundations of mathematics.

1) Every mathematical truth about the natural numbers should have a meaningful finitistic proof. While it is arguable what a "finitistic proof" is, one precise way of saying this is that there is a natural set of axioms from which we can derive all truths about the natural numbers. For example Peano Arithmetic would be a good candidate.

2) **Hilbert's Conservation Program**: If a mathematical truth can be proved by strong methods (say using set theoretic methods), then it can be proved by finitistic methods.

3) **Hilbert's Consistency Program**: We should be able to give a finitistic proof of the consistency of our methods.

Gödel showed that all of this is impossible.

Theorem 9.1. i) (First Incompleteness Theorem) There is a sentence ϕ such that $\mathbf{N} \models \phi$ and $PA \not\vdash \phi$. Indeed if T is a recursive theory such that $T \supseteq PA$ and $\mathbf{N} \models T$, then there is a sentence ϕ such that $T \not\vdash \phi$ and $T \not\vdash \neg\phi$.

ii) (Second Incompleteness Theorem) Let T be as above. Then T does not prove the consistency of T.

The first completeness theorem shows that i) will fail. The second incompleteness theorem shows that iii) fails. Since using set theory we can show that PA is consistent, iii) shows that the conservation program fails as well.

We begin by outlining the main steps of Gödel's proof. Let $\mathcal{L}$ be a recursive language conaining $+, \cdot, <, 0, 1$ the language of arithmetic and let $\mathbf{T}$ be an $\mathcal{L}$-theory. $\mathbf{T}$ will be our candidate for an axiomatization of a large part of mathematics. We assume $\mathbf{T}$ is sufficiently strong by requiring that $\mathbf{T}$ contains $\mathbf{T}_0$, where $\mathbf{T}_0$ is a theory which axiomatizes enough number theory for our purposes and $\mathbf{N} \models \mathbf{T}_0$. We will take $\mathbf{T}_0 = PA$, though a much weaker theory would suffice. For $\mathbf{T}$ we can take PA, the Zermelo-Frankel axioms for set theory, or some other strong system.

The following lemma is the key technical idea. For $n \in \mathbf{N}$, let $\widehat{n}$ denote the $\mathcal{L}$-term $\underbrace{1 + \ldots + 1}_{n-\text{times}}$. If $m_1, \ldots, m_n \in \mathbf{N}$, we let $\widehat{\bar{m}}$ denote $(\widehat{m}_1, \ldots, \widehat{m}_n)$.

Lemma 9.2: (Representation Lemma) We will show that if $f(\bar{x})$ is a primitive recusive function, then there is a formula $\phi_f(\bar{x}, y)$ such that

i) $f(\bar{n}) = m \Leftrightarrow \mathbf{T}_0 \vdash \phi_f(\widehat{\bar{n}}, \widehat{m}) \Leftrightarrow \mathbf{N} \models \phi_f(\widehat{\bar{n}}, \widehat{m})$.

ii) $\mathbf{T}_0 \vdash \ \forall \bar{x} \exists y \ \phi_f(\bar{x}, y)$,and

iii) $\mathbf{T}_0 \vdash \ \forall x \forall y \forall z \ ((\phi_f(\bar{x}, y) \wedge \phi_f(\bar{x}, z)) \rightarrow y = z)$.

Lemma 9.2 and the results of §7 already allow us to deduce the impossibility of i). Suppose $R \subseteq \mathbf{N}^m$ is recursive. Suppose $R = W_e$. Then

$$\bar{x} \in R \Leftrightarrow \exists s\ T(e, \bar{x}, s)$$

where T Kleene's T-predicate.

Suppose $A \subseteq \mathbf{N}$ is Σ_n. Say

$$x \in A \Leftrightarrow \exists y_1 \forall y_2 \ldots Q y_n\ R(x, \bar{y})$$

where R is recursive and Q is $\exists$ if n is odd and $\forall$ if n is even.

Suppose without loss of generality that n is odd. Then

$$x \in A \Leftrightarrow \exists y_1 \forall y_2 \ldots Q y_n \exists s\ T(e, x, \bar{y}, s)$$

Let $\psi_A(v)$ be the formula

$$\exists y_1 \forall y_2 \ldots Q y_n \exists s\ \phi_T(\widehat{e}, v, \bar{y}, s)$$

where ϕ_T is the formula which represents T.

It is clear that

$$n \in A \Leftrightarrow \mathbf{N} \models \psi_A(\widehat{n})$$

Let $Th(\mathbf{N}) = \{\psi : \mathbf{N} \models \psi\}$. The function

$$n \mapsto \psi_A(\widehat{n})$$

gives a many-one reduction of A to $Th(\mathbf{N})$. Thus we have proved:

Lemma 9.3. $Th(\mathbf{N})$ is not arithmetic.

Proof.

Suppose $Th(\mathbf{N})$ were Σ_n. By 7.15, there is a Σ_{n+1} set A which is not Σ_n. By the above arguments $A \leq_m Th(\mathbf{N})$. But by 7.11 vi), if $Th(\mathbf{N}) \in \Sigma_n$, then $A \in \Sigma_n$, a contradiction.

In particular we can now deduce the first incompleteness theorem.

Corollary 9.4: If $\mathbf{N} \models \mathbf{T}$ and $\mathbf{T}$ is recursive, then there is a sentence ϕ such that $\mathbf{N} \models \phi$ and $\mathbf{T} \not\vdash \phi$.

proof:

If $\mathbf{T}$ is recursive, then $\{\phi : \mathbf{T} \vdash \phi\}$ is a recursively enumerable subset of $Th(\mathbf{N})$. Since $Th(\mathbf{N})$ is not recursively enumerable, there is a setence ϕ such that $\mathbf{N} \models \phi$ but $\mathbf{T} \not\vdash \phi$.

Indeed, if $\mathbf{T} \subseteq Th(\mathbf{N})$ is Σ_n, then $\{\phi : \mathbf{T} \vdash \phi\}$ is Σ_{n+1} and the above argument shows that there is a true sentence ϕ such that $\mathbf{T} \not\vdash \phi$.

Our first task then is to prove the representation lemma. This will require yet another method of coding finite sequences. The goal is to have a method of coding finite sequences so that one can express properties of the codes easily in the language of arithmetic. We will use Gödel's beta-function.

Let Seq be the set of finite sequences of elements of $\mathbf{N}$. We define $\beta : \mathbf{N}^3 \to Seq$, such that $\beta(u,v,w)$ is the sequence $(a_0,\ldots,a_{w-1})$ where $a_i = u(\mathrm{mod}\ (i+1)v+1)$ for $i = 0,\ldots,w-1$ (and $\beta(u,v,w)$ is the empty sequence for $w=0$).

Let $\Psi(u,v,w,i,x)$ be the formula

$$i < w \wedge 0 \le x < (i+1)v+1 \wedge \exists y \le u\ y((i+1)v+1)+x = u.$$

Then $\Psi(u,v,w,i,x)$ expresses that x is the i^{th} element in the sequence $\beta(u,v,w)$. We will write

$$\beta(u,v,w)_i = x$$

for $\Psi(u,v,w,i,x)$. While it is easy to express that $\beta(u,v,w)_i = x$, it is not so easy to see that every sequence is coded in this way.

The proof uses the following simple lemma from number theory.

Lemma 9.5. (Chineese Remainder Theorem) Suppose $m_1,\ldots,m_n$ are relatively prime. Then for any $a_1,\ldots,a_n$ there is an x such that $x \equiv a_i(\mathrm{mod}\ m_i)$ for $i=1,\ldots,n$.

Proof.

Let

$$M_i = \prod_{j \ne i} m_j.$$

Since M_i and m_i are relatively prime, we can find b_i such that $b_iM_i \equiv 1(\mathrm{mod}\ m_i)$. Let

$$x = \sum_{i=1}^{n} a_ib_iM_i.$$

Since $m_i|M_j$ for $j \ne i$, $x \equiv a_ib_iM_i(\mathrm{mod}\ m_i)$. Thus $x \equiv a_i(\mathrm{mod}\ m_i)$ for $i=1,\ldots,n$.

Lemma 9.6. For any sequence $\sigma = (a_0,\ldots,a_{w-1})$ there are u and v such that $\beta(u,v,w) = \sigma$.

Proof.

Let $n = \max(w,a_0,\ldots,a_{w-1})$ and let $v = n!$. We claim that $v+1, 2v+1,\ldots,wv+1$ are relatively prime. Suppose p is prime and $p|iv+1$ and $p|jv+1$ where $j > i > 0$ Then $p|(j-i)v$ Thus $p|(j-i)$ or $p|v$ and, since $(j-i)|v$, $p|v$. But then $p \not| iv+1$. Thus $v+1,\ldots,wv+1$ are relatively prime.

By the Chinese remainder theorem there is a number u such that $u \equiv a_i(\mathrm{mod}\ (i+1)v+1)$ for $i=0,\ldots,w-1$.

The proof above is a very simple number theoretic argument that can easily be formalized in Peano Arithmetic (in particular the usal proof of the Chineese remainder theorem can be formalized in PA. Thus for any w

$$PA \vdash \forall a_0 \ldots \forall a_{w-1} \exists u \exists v \forall i < w \bigwedge_{i=0}^{w-1} \beta(u,v,w)_i = a_i.$$

We will use PA as the weak base theory T_0. If we wanted to use a weaker base theory we would need to use a more subtle coding.

We can now prove the Representation Lemma. If $f : \mathbf{N}^m \to \mathbf{N}$ is a primitive recursive function we say that $\phi(x_1, \ldots, x_m, y)$ *represents* f in PA if and only if

i) $f(\bar{m}) = n$ if and only if $PA \vdash \phi(\widehat{\bar{m}}, n)$ if and only if $\mathbf{N} \models \phi(\widehat{\bar{m}}, n)$ for $\bar{m}, n \in \mathbf{N}$.

ii) $PA \vdash \forall \bar{x} \exists y \; \phi(\bar{x}, y)$, and

iii) $PA \vdash \forall \bar{x} \forall y \forall z \; ((\phi(\bar{x}, y) \wedge \phi(\bar{x}, z)) \to y = z)$.

Conditions ii) and iii) assert that in any model of PA the formula ϕ defines the graph of a function. Condition i) asserts that on the standard part of the model that function agrees with f.

We will prove that any primitive recursive function is represented in PA. This will be done by induction on the complexity of the definition of f. We begin with the basic functions.

- If z is the zero function, let $\phi_z(x, y)$ be the formula
- If s is the succesor function, let $\phi_s(x, y)$ be the formula $y = x + 1$.
- If π_i^n is the function $(x_1, \ldots, x_n) \mapsto x_i$, then $\phi_{\pi_i^n}(x_1, \ldots, x_n, y)$ is the formula $y = x_i$.

It should be clear if f is a basic function then ϕ_f represents ϕ.

• Suppose $g_1, \ldots, g_n : \mathbf{N}^m \to \mathbf{N}$ and $h : \mathbf{N}^n \to \mathbf{N}$ are primitive recursive and $f : \mathbf{N}^m \to \mathbf{N}$ by $f(\bar{x}) = h(g_1(\bar{x}), \ldots, g_n(\bar{x}))$. Suppose by induction that there are formulas $\phi_{g_1}, \ldots, \phi_{g_n}, \phi_h$ representing $g_1, \ldots, g_n$, and h respectively. Let $\phi_f(\bar{x}, y)$ be the formula

$$\exists z_1 \ldots \exists z_n \bigwedge_{i=1}^{n} \phi_{g_i}(\bar{x}, z_i) \; \wedge \phi_h(z_1, \ldots, z_n, y).$$

Suppose $a_1, \ldots, a_m \in \mathbf{N}$ and $h(\bar{a}) = b$. There are $c_1, \ldots, c_n \in \mathbf{N}$ such that $g_i(\bar{a}) = c_i$ for $i = 1, \ldots, n$ and $h(\bar{c}) = b$. Then

$$PA \vdash \bigwedge_{i=1}^{n} \phi_{g_i}(\widehat{\bar{a}}, \widehat{c}_i) \wedge \phi_h(\widehat{\bar{c}}, \widehat{b})$$

so $PA \vdash \phi_f(\widehat{\bar{a}}, \widehat{b})$. On the other hand suppose $a_1, \ldots, a_m, b \in \mathbf{N}$ and $PA \vdash \phi_f(\widehat{\bar{a}}, \widehat{b})$. Then $\mathbf{N} \models \phi_f(\widehat{\bar{a}}, \widehat{b})$. Thus there are $c_1, \ldots, c_n \in \mathbf{N}$ such that

$$\mathbf{N} \models \phi_{g_i}(\widehat{\bar{a}}, \widehat{c}_i) \wedge \phi_h(\widehat{\bar{c}}, \widehat{b}).$$

Since PA proves that ϕ_{g_i} and ϕ_h define graphs of functions which agree on $\mathbf{N}^m$ and $\mathbf{N}^n$ with g_i and h respectively, $g_i(\bar{a}) = c_i$ and $h(\bar{c}) = b$ thus $f(\bar{a}) = b$.

Since PA proves that ϕ_{g_i} for $i = 1, \ldots, n$ and ϕ_h define the graphs of a functions, PA proves that ϕ_f defines the graph of a function.

• Suppose $g : \mathbf{N}^{m-1} \to \mathbf{N}$ and $h : \mathbf{N}^{m+1} \to \mathbf{N}$ are primitive recursive $f(0, \bar{x}) = g(\bar{x})$ and $f(n+1, \bar{x}) = h(n, \bar{x}, f(n, \bar{x}))$. Let ϕ_g represent g and ϕ_h represent h.

Let $\phi_f(z, \bar{x}, y)$ be the formula asserting "there is σ coding a sequence of length $z+1$, say $\sigma = (u_0, \ldots, u_z)$ where $\phi_g(\bar{x}, u_0)$ and $\phi_h(i, \bar{x}, u_i, u_{i+1})$ for $i < z+1$ and $y = u_z$".

To make this precise we use the beta-function. Let $\phi_f(z, \bar{x}, y)$ be the formula

$$\exists u, v\ [(\forall w \le u \beta(u, v, z+1)_0 = w \to \phi_g(\bar{x}, w)) \wedge (\forall i \le z \forall w_1, w_2 \le u\ (\beta(u, v, z+1)_i = w_1 \wedge \beta(u, v, z+1)_{i+1} = w_2) \to \phi_h(i, \bar{x}, w_1, w_2))].$$

Using the facts above it is easy to see that ϕ_f has the desired interpretation.

Exercise: Show that ϕ_f represents f.

This concludes the proof of the Representation Lemma.

We can get a much sharper version of the representation lemma.

Exercise: We say that a formula in the language of arithmetic is a Δ_0 formula if it is in the smallest collection of formulas containing the atomic formulas and closed under $\wedge, \vee, \neg$ and bounded quantification $\forall x < t$ and $\exists x < t$ where t is a term. For example $\forall x < y^2 \exists z < 2xy\ z = y - x$ is a Δ_0 formula.

We say that $\psi(\bar{x})$ is a Σ_1-formula if it is of the form $\exists \bar{y} \phi(\bar{x}, \bar{y})$ where ϕ is Δ_0. We say that a formula θ is Σ_1^{PA} if $PA \vdash \theta \to \psi$ for some Σ_1-formula Ψ.

i) If ϕ_0 and ϕ_1 are Σ_1^{PA}-formulas show that there $\phi \wedge \psi$ and $\phi \vee \psi$ are Σ_1^{PA}.

ii) Suppose $\phi(\bar{x}, y)$ is Σ_1^{PA} then so are $\exists \bar{y}\ \phi(\bar{x}, \bar{y})$ and $\forall y < t(\bar{x})\ \phi(\bar{x}, y)$ for any term t.

iii) Show that any primitive recursive function is represented by a Σ_1^{PA}-formula.

We can say a bit more. Suppose $P(\bar{x})$ is a primitive recursive predicate. Then there is a formula $\psi_P(\bar{x})$ such that $P(\bar{n})$ if and only if $PA \vdash \psi_P(\bar{n})$ for all $\bar{n} \in \mathbf{N}$. [Let $\Psi_P(\bar{x})$ be $\phi_f(\bar{x}, 1)$ where f is the primitive recursive characteristic function for P.] The same is true for any recursively enumberable set.

Corollary 9.7. Let $A \subseteq \mathbf{N}$ be recursively enumerable. There is a formula $\psi_A(x)$ such that $n \in A$ if and only if $PA \vdash \psi_A(n)$ for all $n \in \mathbf{N}$.

Proof.

Let $A = W_e$. Then $x \in A$ if and only if $\exists s\ T(e, x, s)$ where T is Kleene's T-predicate. Let $\psi_T(u, v, w)$ be a formula representing the primitive recursive predicate T. Let $\psi_A(x)$ be the formula

$$\exists y\ \psi_T(\widehat{e}, x, y).$$

If $n \in A$, then there is an $s \in \mathbf{N}$ such that $T(e, n, s)$. Since ψ_T represents T, $PA \vdash \psi_T(\widehat{e}, \widehat{n}, \widehat{s})$. But then $PA \vdash \exists y\ \psi_T(\widehat{e}, \widehat{n}, y)$.

On the other hand if $PA \vdash \exists y\ \psi_T(\widehat{e}, \widehat{n}, y)$, then $\mathbf{N} \models \exists y\ \psi_T(\widehat{e}, \widehat{n}, y)$. Choose $s \in \mathbf{N}$ such that $\mathbf{N} \models \psi_T(\widehat{e}, \widehat{n}, \widehat{s})$. But then $T(e, n, s)$ holds and $n \in A$.

We conclude this section with a stronger version of the first incompleteness theorem. We will replace the assumption that $\mathbf{N} \models \mathbf{T}$ by the weaker assumption that $\mathbf{T} \supseteq PA$ is consistent.

Let $P(\mathbf{T}) = \{\phi : \mathbf{T} \vdash \phi\}$ and let $R(\mathbf{T}) = \{\phi : \mathbf{T} \vdash \neg\phi\}$. Intuitively $P(\mathbf{T})$ are the sentences provable from $\mathbf{T}$ and $R(\mathbf{T})$ are the sentences refutable from $\mathbf{T}$.

Theorem 9.8. The sets $P(PA)$ and $R(PA)$ are recursively inseparable.

Proof.

Since PA is consistent $P(PA) \cap R(PA) = \emptyset$ and since PA is recursively axiomatized they are recursively enumerable.

For $i = 0, 1$ let $A_i(e, x, s)$ be the primitive recursive predicate asserting "the computation of $\phi_e(x)$ halts in at most s steps with output i". Let $\psi_i(e, x, s)$ represent A_i.

Let $\theta_0(x)$ be the formula

$$\exists y\ \psi_0(x, x, s) \wedge \forall z < y\ \neg\psi_1(x, x, z)$$

and let $\theta_1(x)$ be the formula

$$\exists y\ \psi_1(x, x, s) \wedge \forall z \leq y\ \neg\psi_0(x, x, z)$$

claim i) $PA \vdash \forall x\ \neg(\phi(x) \wedge \psi(x))$

ii) $PA \vdash \forall x\ [(\exists y(\psi_0(x, x, y) \vee \psi_1(x, x, y))) \rightarrow (\theta_0(x) \vee \theta_1(x))]$.

For any x, if there is a y such that one of $\psi_0(x, x, y)$ or $\psi_1(x, x, y)$ holds, then, by induction, there is y_0 such that one of the $\psi_i(x, x, y_0)$ holds and $\neg\psi_0(x, x, z)$ and $\neg\psi(x, x, z)$for $z < y_0$. If $\psi_0(x, x, y_0)$ holds, then $\theta_0(x)$; otherwise $\theta_1(x)$.

Let $A = \{e : \phi_e(e) = 0\}$ and $B = \{e : \phi_e(e) = 1\}$ be the recursively inseparable sets of 8.4.

Suppose C is a recursive set of sentences such that $P(PA) \subset C$ and $R(PA) \subset C$. Let $D = \{e :$ the sentence $\theta_0(\widehat{e}) \in C\}$. Clearly D is recursive. We claim that D separates A and B.

Let $e \in A$. Then there is an $s \in \mathbf{N}$ such that $A_0(e, e, s)$. Thus $PA \vdash \psi_0(\widehat{e}, hate, \widehat{s})$ and for all $t < s$ $PA \vdash \neg\psi_1(\widehat{e}, \widehat{e}, \widehat{s})$. Since

$$PA \vdash t < \widehat{s} \rightarrow \bigwedge_{i=0}^{s-1} t = \widehat{i},$$

$PA \vdash \psi_0(\widehat{e}, \widehat{e}, \widehat{s}) \wedge \forall z < \widehat{s}\ \neg\phi_1(\widehat{e}, \widehat{e}, \widehat{s})$. Thus $PA \vdash \theta_0(\widehat{e})$. So $\theta(\widehat{e}) \in R(PA) \subseteq C$ and $e \in D$.

Similarly if $e \in B$, then $PA \vdash \theta_1(\widehat{e})$. Thus $PA \vdash \neg\theta_0(\widehat{e})$. Thus $\theta_0(\widehat{e}) \in R(PA) \subseteq \neg C$ and $e \in \neg D$. Since D separates A and B, D is not recursive, a contradiction.

Corollary 9.9. (Rosser's Incompleteness Theorem) Suppose $\mathbf{T} \supseteq PA$ is a consistent recursively axiomatized theory. Then $P(\mathbf{T})$ and $R(\mathbf{T})$ are recursively inseparable. In particular $\mathbf{T}$ is incomplete.

Proof.

Since $\mathbf{T}$ is consistent and recursively axiomatized, $P(\mathbf{T})$ and $R(\mathbf{T})$ are disjoint recursively enumberable sets. Since $\mathbf{T} \supseteq PA$, $P(\mathbf{T}) \supseteq P(PA)$ and $R(\mathbf{T}) \supseteq R(PA)$. If we could recusively separate $P(\mathbf{T})$ and $R(\mathbf{T})$ we would also separate $P(PA)$ and $R(PA)$ contradicting 9.8.

If $\mathbf{T}$ were complete, then $P(\mathbf{T})$ would itself be recursive (by 4.8) contradicting 9.8.

§10. Gödel's Incompleteness Theorem II: Arithmetization of Syntax:

In §9 we gave a proof of the first incompleteness theorem based on basic recursion theoretic ideas. In this section we give a second proof which follows Gödel more closely. We will also sketch the ideas behind the proof of the second incompleteness theorem. The new idea of this section is the idea of Gödel codes for formulas.

We will assign a number $\lceil \phi \rceil$ to each formula ϕ. We call $\lceil \phi \rceil$ the *Gödel code* for ϕ. Gödel coding allows us to talk about properties of formulas in the language of arithmetic. Gödel showed that there are amazing possibilities for self reference. In particular he proved the following striking lemma.

Lemma 10.1. (Diagonalization Lemma) Let $\phi(v)$ be a formula in the language of arithmetic with one free variable v. There is a sentence ψ such that

$$PA \vdash \ \psi \leftrightarrow \phi(\lceil \psi \rceil).$$

Intuitively the sentence ψ says "My code has property ϕ". Strictly speaking we should write $PA \vdash \psi \leftrightarrow \phi(\widehat{\lceil \phi \rceil})$ but we will drop the $\widehat{}$ when no confusion arises.

We will begin shortly the work needed to prove the diagonalization lemma and to deduce the incompleteness theorem from it, but first let us deduce one simple and important corollary.

A formula $\Gamma(v)$ is called a *truth definition* if and only if

$$\mathbf{N} \models \psi \ \text{ if and only if } \quad \mathbf{N} \models \Gamma(\lceil \psi \rceil).$$

for all sentences ψ

Corollary 10.2. (Tarski's Undefinability of Truth) There are no truth definitions.

Proof.

Suppose $\Gamma(v)$ is a truth definition. Apply the diagonalization lemma to $\neg\Gamma$ to obtain a sentence ψ such that $PA \vdash \ \psi \leftrightarrow \neg\Gamma(\lceil \psi \rceil)$. Clearly ψ shows that Γ is not a truth definition.

We now begin the mechanics of coding. We fix a primitive recursive method of coding finite sequences. We let $\langle a_1, \ldots, a_m \rangle$ be the code for the sequence $(a_1, \ldots, a_m)$. We choose the coding so that:

i) every natural number codes a sequence,

ii) $n \mapsto l(n)$ is primitive recursive, where $l(n)$ is the length of the sequence coded by n, and

iii) $(n, i) \mapsto (n)_i$ is primitive recursive, where $(n)_i$ is the i^{th}-element of the sequence coded by n if $i \leq l(n)$ and $(n)_i = 0$ if $i > l(n)$.

For example we could use the coding τ described in §6.

Let us assume that our language is $\mathcal{L} = \{+, \cdot, <, 0, 1\}$ and that we use only the connectives $\wedge$ and $\neg$, the quantifier $\exists$ and variables $v_0, v_1, \ldots$. We assign each symbol a code as follows.

$\ulcorner 0 \urcorner = \langle 0, 0 \rangle$	$\ulcorner 1 \urcorner = \langle 0, 1 \rangle$	$\ulcorner v_i \urcorner = \langle 1, i \rangle$
$\ulcorner + \urcorner = \langle 2, 0 \rangle$	$\ulcorner \cdot \urcorner = \langle 2, 1 \rangle$	$\ulcorner < \urcorner = \langle 3, 0 \rangle$
$\ulcorner = \urcorner = \langle 3, 1 \rangle$	$\ulcorner \wedge \urcorner = \langle 4, 0 \rangle$	$\ulcorner \neg \urcorner = \langle 4, 1 \rangle$
$\ulcorner \exists \urcorner = \langle 5, 0 \rangle$		

We inductively define coding of terms as follows. If t_1 and t_2 are terms then

$\ulcorner t_1 + t_2 \urcorner = \langle \ulcorner + \urcorner, \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner \rangle$ and
$\ulcorner t_1 \cdot t_2 \urcorner = \langle \ulcorner \cdot \urcorner, \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner \rangle$.

If t_1 and t_2 are terms, we code atomic formulas as follows.

$\ulcorner t_1 = t_2 \urcorner = \langle \ulcorner = \urcorner, \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner \rangle$ and
$\ulcorner t_1 < t_2 \urcorner = \langle \ulcorner < \urcorner, \ulcorner t_1 \urcorner, \ulcorner t_2 \urcorner \rangle$.

Finally if ϕ and ψ are formulas then

$\ulcorner \neg\phi \urcorner = \langle \ulcorner \neg \urcorner, \ulcorner \phi \urcorner \rangle$,
$\ulcorner \phi \wedge \psi \urcorner = \langle \ulcorner \wedge \urcorner, \ulcorner \phi \urcorner, \ulcorner \psi \urcorner \rangle$ and
$\ulcorner \exists v_i \phi \urcorner = \langle \ulcorner \wedge \urcorner, \ulcorner v_i \urcorner, \ulcorner \phi \urcorner \rangle$.

We will see that all basic syntactic properties of formulas are primitive recursive. It is easy to see for example that the maps $\ulcorner \phi \urcorner \mapsto \ulcorner \neg\phi \urcorner$ and $(\ulcorner \phi \urcorner, \ulcorner \psi \urcorner) \mapsto \ulcorner \phi \wedge \psi \urcorner$ are primitive recursive.

Lemma 10.3. The predicates "n codes a term" and "n codes a formula" are primitive recursive.

Proof.
Let

$$T(x) = \begin{cases} 1 & x = \ulcorner 0 \urcorner > \text{ or } x = \ulcorner 1 \urcorner \\ 1 & l(x) = 3, (x)_1 = \langle 2, 0 \rangle \text{ or } \langle 2, 1 \rangle,\ T((x)_2) = 1 \text{ and } T((x)_3) = 1 \\ 0 & \text{otherwise} \end{cases} \cdot$$

Clearly T is primitive recursive and $T(n) = 1$ if and only if n codes a term. Let

$$F(x) = \begin{cases} 1 & l(x) = 3, (x)_1 = \ulcorner = \urcorner \text{ or } \ulcorner < \urcorner,\ T((x)_1) = 1 \text{ and } T((x)_2) = 1 \\ 1 & l(x) = 2, (x)_1 = \ulcorner \neg \urcorner \text{ and } F((x)_2) = 1 \\ 1 & l(x) = 3, (x)_1 = \ulcorner \wedge \urcorner \text{ and } F((x)_2) = F((x)_3) = 1 \\ 1 & l(x) = 3, (x)_1 = \ulcorner \exists \urcorner,\ \exists i < x\ (x)_2 = \langle 1, i \rangle \text{ and } F((x)_3) = 1 \\ 0 & \text{otherwise} \end{cases} .$$

Then F is primitive recurisive and $F(n) = 1$ if and only if n is the code for a formula.

The next lemmas will be the key to proving the diagonalization lemma.

Lemma 10.4. There is a primitive recursive function s such that if t is a term and $i, y \in \mathbf{N}$, then $s(\ulcorner t \urcorner, i, y)$ is the code for the term obtained by replacing all occurences of v_i in t by the term $\widehat{y}$ (where $\widehat{y}$ is the term $\underbrace{1 + \ldots + 1}_{y-\text{times}}$).

Proof. We define s by:

$$s(x, i, y) = \begin{cases} x & x = \ulcorner 0 \urcorner, x = \ulcorner 1 \urcorner \text{ or } x = \ulcorner v_j \urcorner \text{ where } i \neq j \\ \ulcorner \widehat{y} \urcorner & x = \ulcorner v_i \urcorner \\ \langle +, s(t_1, i, y), s(t_2, i, y) \rangle & x = \langle +, t_1, t_2 \rangle \\ \langle \cdot, s(t_1, i, y), s(t_2, i, y) \rangle & x = \langle \cdot, t_1, t_2 \rangle \\ 0 & \text{otherwise} \end{cases}$$

Clearly s is primitive recursive and s is the desired function.

Lemma 10.5. There is a primitive recursive function sub such that $sub(\ulcorner \phi \urcorner, i, y) = \ulcorner \psi \urcorner$ where ψ is the formula obtained by substituting $\widehat{y}$ for each free occurence of v_i in ϕ. .

Proof.

We may define sub by

$$sub(x, i, y) = \begin{cases} \langle \ulcorner = \urcorner, s(t_1, i, y), s(t_2, i, y) \rangle & x = \langle \ulcorner = \urcorner, t_1, t_2 \rangle \\ \langle \ulcorner < \urcorner, s(t_1, i, y), s(t_2, i, y) \rangle & x = \langle \ulcorner < \urcorner, t_1, t_2 \rangle \\ \langle \ulcorner \neg \urcorner, sub(\ulcorner \phi \urcorner, i, y) \rangle & x = \langle \ulcorner \neg \urcorner, \ulcorner \phi \urcorner \rangle \\ \langle \ulcorner \wedge \urcorner, sub(\ulcorner \phi \urcorner, i, y), sub(\ulcorner \psi \urcorner, i, y) \rangle & x = \langle \ulcorner \wedge \urcorner, \ulcorner \phi \urcorner, \ulcorner \psi \urcorner \rangle \\ \langle \ulcorner \exists \urcorner, \ulcorner v_j \urcorner, sub(\ulcorner \phi \urcorner, i, y) \rangle & x = \langle \ulcorner \exists \urcorner, \ulcorner v_j \urcorner, \ulcorner \phi \urcorner \rangle \text{ and } i \neq j \\ \langle \ulcorner \exists \urcorner, \ulcorner v_i \urcorner, \ulcorner \phi \urcorner \rangle & x = \langle \ulcorner \exists \urcorner, \ulcorner v_i \urcorner, \ulcorner \phi \urcorner \rangle \\ 0 & \text{otherwise.} \end{cases}$$

We are now ready to prove the diagonalization lemma 10.1. Let $\phi(v_0)$ be an $\mathcal{L}$-formula with one free variable v_0. Let $S(x, y, z, w)$ be an $\mathcal{L}$-formula representing the primitive function sub.

Let $\theta(v_0) = \exists y \ (S(v_0, 0, v_0, y) \wedge \phi(y))$ That is, $\theta(v_0)$ asserts $\phi(sub(v_0, 0, v_0))$. Let $m = \ulcorner \theta(v_0) \urcorner$ and let $\psi = \theta(m)$.

Then

$$\begin{aligned} PA \vdash \psi &\leftrightarrow \theta(m) \\ &\leftrightarrow \exists y \ S(m, 0, m, y) \wedge \phi(y) \\ &\leftrightarrow \exists y \ S(\ulcorner \theta(v_0) \urcorner, 0, m, y) \wedge \phi(y) \\ &\leftrightarrow \exists y \ y = \ulcorner \theta(m) \urcorner \wedge \phi(y) \\ &\leftrightarrow \phi(\ulcorner \theta(m) \urcorner) \\ &\leftrightarrow \phi(\ulcorner \psi \urcorner) \end{aligned}$$

as desired.

Our coding will be slightly easier if we assume that the theory $\mathbf{T}$ has a primitive recursive axiomatization. The next lemma shows that this is no loss of generality.

Lemma 10.6. (Craig's trick) Suppose T is a recusively axiomatized $\mathcal{L}$-theory. Then there is a primitive recursively axiomatized $\mathcal{L}$-theory T^* such that T and T^* have the same consequences (ie. $T \vdash \phi \Leftrightarrow T^* \vdash \phi$ for any $\mathcal{L}$-sentence ϕ).

Proof.
Suppose $W_e = \{\ulcorner\phi\urcorner : \phi \in T\}$. Let

$$T^* = \{\underbrace{\phi \wedge \ldots \wedge \phi}_{s-\text{times}} : \ulcorner\phi\urcorner \in W_e^s\}.$$

It is easy to see that T and T^* have the same logical consequences. On the other hand, since "$x \in W_e^s$" is a primitive recursive predicate, $\{\ulcorner\psi\urcorner : \psi \in T^*\}$ is primitive recursive.

Note that in fact Craig's trick works just as well for recursively enumerable theories T.

Lemma 10.7. Let $\mathbf{T}$ be a primitive recursive $\mathcal{L}$-theory. Let $Prov_{\mathbf{T}}(x, y)$ be the predicate "x is a proof from $\mathbf{T}$ if the formula with Gödel code y". The predicate $Prov_{\mathbf{T}}$ is primitive recursive.

Proof.
Basicly, $Prov_{\mathbf{T}}(x, y)$ if and only if $\forall i \leq l(x) \quad (x)_i \in \mathbf{T}$ or $(x)_i$ follows from previous $(x)_j$ be an inference rule.
This can be coded in a primitive recursive way. We leave the details to the reader.

Let $\psi_{Prov_T}(x, y)$ be an $\mathcal{L}$-formula representing $Prov_T$ in PA and let

$$Pr_{\mathbf{T}}(y) \Leftrightarrow \exists x \ \psi_{Prov_{\mathbf{T}}}(x, y)$$

Using $Pr_{\mathbf{T}}$ we can give Gödel's proof of the first incompleteness theorem. Let $\mathbf{T}$ be a consistent primitive recursive theory extending PA. By the diagonalization lemma there is a sentence ϕ such that $PA \vdash \phi \leftrightarrow \neg Pr_{\mathbf{T}}(\phi)$. We call ϕ the *Gödel sentence* for $\mathbf{T}$.

Theorem 10.8 (First Incompleteness theorem) Let ϕ be the Gödel sentence for $\mathbf{T}$. Then $\mathbf{T} \not\vdash \phi$. Moreover if $\mathbf{N} \models \mathbf{T}$, then $\mathbf{T} \not\vdash \neg\phi$.

Proof.
If $\mathbf{T} \vdash \phi$, then there is an $n \in \mathbf{N}$ such that $Prov_{\mathbf{T}}(n, \ulcorner\phi\urcorner)$. But then $PA \vdash \psi_{Prov_{\mathbf{T}}}(n, \ulcorner\phi\urcorner)$ and $PA \vdash Pr_{\mathbf{T}}(\ulcorner\phi\urcorner)$ and $PA \vdash \neg\phi$. Thus $\mathbf{T} \vdash \neg\phi$, contradicting the consistency of $\mathbf{T}$. Hence $\mathbf{T} \not\vdash \phi$.

If $\mathbf{N} \not\models \phi$, then $\mathbf{N} \models Pr_{\mathbf{T}}(\ulcorner\phi\urcorner)$ and hence there is $m \in \mathbf{N}$ such that $\mathbf{N} \models \psi_{Prov_{\mathbf{T}}}(m, \ulcorner\phi\urcorner)$ and m really is the code for a proof of ϕ from $\mathbf{T}$. But then $\mathbf{T} \vdash \phi$ and $\mathbf{N} \models \phi$ a contradiction, thus $\mathbf{N} \models \phi$. Hence if $\mathbf{N} \models \mathbf{T}$, then $T \not\vdash \neg\phi$.

A slightly different diagonalization gives a different proof of the incompleteness result from corollary 9.9.

Theorem 10.9. Let $\mathbf{T}$ be a recursively axiomatized consistent extension of PA, then $\mathbf{T}$ is incomplete.

Proof.

Let $\theta(x, y)$ be an $\mathcal{L}$-formula representing the primitive recursive relation "x and y are Gödel codes for formulas and x codes the negation of the formula coded by y".

Let $Pr^*_{\mathbf{T}}(v)$ be the formula

$$\exists y \ (Prov_{\mathbf{T}}(y, v) \wedge \forall z \ (\theta(z, v) \rightarrow \forall x < y \ \neg Prov_{\mathbf{T}}(x, z))).$$

Thus $Pr^*_{\mathbf{T}}(\ulcorner \phi \urcorner)$ asserts that there is x coding a proof of ϕ and no $y < x$ codes a proof of $\neg\phi$.

By the diagonalizaiton lemma there is a sentence ϕ such that

$$PA \vdash \phi \leftrightarrow \neg Pr^*_{\mathbf{T}}(\ulcorner \phi \urcorner).$$

We call ϕ a *Rosser sentence*.

Suppose $\mathbf{T} \vdash \phi$. Then there is a natural number n coding a proof of ϕ and since $\mathbf{T}$ is consistent if $m < n$, then m does not code a proof of $\neg\phi$. But then if $M \models T$, then $M \models Pr^*_{\mathbf{T}}(\ulcorner \phi \urcorner))$ and $M \models \neg\phi$ a contradiction. Thus $\mathbf{T} \not\vdash \phi$.

Suppose $T \vdash \neg\phi$. Then there is a natural number n coding a proof of $\neg\phi$ and if $m < n$, then m does not code a proof of ϕ. Thus if $M \models \mathbf{T}$, then $M \models \neg Pr^*_{\mathbf{T}}(\ulcorner \phi \urcorner)$, so $M \models \phi$, a contradiction. Thus $T \not\vdash \neg\phi$.

The next lemma summarizes the facts about provability that one must verify in PA to prove the second incompleteness theorem.

Lemma 10.10. Let $\mathbf{T} \supseteq PA$ be a primitive recursive theory and let ϕ and ψ be $\mathcal{L}$-sentences. Then the following *derivability conditions hold*

D1. If $\mathbf{T} \vdash \phi$, then $PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner)$.

D2. If $PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner)$, then $PA \vdash Pr_{\mathbf{T}}(\ulcorner Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \urcorner)$.

D3. $PA \vdash (Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \wedge Pr_{\mathbf{T}}(\ulcorner \phi \rightarrow \psi \urcorner)) \rightarrow Pr_{\mathbf{T}}(\ulcorner \psi \urcorner)$

D4. $PA \vdash (Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \wedge Pr_{\mathbf{T}}(\ulcorner \psi \urcorner)) \rightarrow Pr_{\mathbf{T}}(\ulcorner \phi \urcorner)$

D5. $PA \vdash Pr_{\mathbf{T}+\psi}(\ulcorner \phi \urcorner) \leftrightarrow Pr_{\mathbf{T}}(\ulcorner \psi \rightarrow \phi \urcorner)$.

Proof.

D1 is easy. If n codes a proof of ϕ,then $PA \vdash Prov_{\mathbf{T}}(n, \ulcorner \phi \urcorner)$, so $PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner)$. D3, D4 and D5 are proved by constructing a primitive recursive function which takes the two given proofs and produces the desired proof. D2 requires more work. See Boolos and Jeffries, *Computability and Logic* for details.

Theorem 10.11 (Second Incompleteness Theorem) Let $\mathbf{T}$ be a consistent recursively axiomatized theory such that $\mathbf{T} \supseteq PA$. Let $Con(\mathbf{T})$ be the sentence $\neg Pr_{\mathbf{T}}(\ulcorner 0 = 1 \urcorner)$. Then $\mathbf{T} \not\vdash Con(\mathbf{T})$.

Proof.

Let ϕ be the Gödel sentence such that $PA \vdash \phi \leftrightarrow \neg Pr_{\mathbf{T}}(\ulcorner \phi \urcorner)$.

We will show that $PA \vdash \phi \leftrightarrow Con(\mathbf{T})$. Then, by 10.8, $PA \not\vdash \phi$, so $PA \not\vdash Con(\mathbf{T})$.

Since we can derive anything from a contradiction, $\mathbf{T} \vdash 0 = 1 \rightarrow \phi$. Thus by D1,

$$PA \vdash Pr_{\mathbf{T}}(\ulcorner 0 = 1 \rightarrow \phi \urcorner).$$

Thus by D3, $PA \vdash \neg Con(\mathbf{T}) \rightarrow Pr_{\mathbf{T}}(\ulcorner \phi \urcorner)$. By choice of ϕ, we have

$$PA \vdash \neg Con(\mathbf{T}) \rightarrow \neg \phi$$

and taking the contrapositive

$$PA \vdash \phi \rightarrow Con(\mathbf{T}).$$

On the otherhand, by D2

$$PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \rightarrow Pr_{\mathbf{T}}(\ulcorner Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \urcorner).$$

By choice of ϕ, $PA \vdash Pr_{\mathbf{T}}(\phi) \rightarrow \neg \phi$. Thus by D1 and D3

$$PA \vdash Pr_{\mathbf{T}}(\ulcorner Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \urcorner) \rightarrow Pr_{\mathbf{T}}(\ulcorner \neg \phi \urcorner).$$

Thus

$$PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \rightarrow Pr_{\mathbf{T}}(\ulcorner \neg \phi \urcorner).$$

Using D4 we see that

$$PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \rightarrow Pr_{\mathbf{T}}(\ulcorner \phi \wedge \neg \phi \urcorner).$$

But then by D1 and D2

$$PA \vdash Pr_{\mathbf{T}}(\ulcorner \phi \urcorner) \rightarrow Pr_{\mathbf{T}}(\ulcorner 0 = 1 \urcorner).$$

So $PA \vdash Con(\mathbf{T}) \rightarrow \phi$, as desired.

By the diagonalization lemma there are sentences ϕ such that $PA \vdash \phi \leftrightarrow Pr_{\mathbf{T}}(\phi)$. Henkin asked if such a sentence is provable? The following result shows that it is.

Corollay 10.12. (Löb's Theorem) Let $\mathbf{T}$ be a consisitent recursively axiomatized theory extending PA and let ϕ be any sentence. Then

$$\mathbf{T} \vdash Pr_{\mathbf{T}}(\ulcorner\phi\urcorner) \rightarrow \phi \quad \Leftrightarrow \mathbf{T} \vdash \phi.$$

Proof.

($\Leftarrow$) This is clear since if $\mathbf{T} \vdash \phi$, then $T \vdash \psi \rightarrow \phi$ for any sentence ψ.

($\Rightarrow$). Suppose $\mathbf{T} \nvdash \phi$. Then $T + \neg\phi$ is consistent and by the second incompleteness theorem

$$\mathbf{T} + \neg\phi \nvdash Con(\mathbf{T} + \neg\phi).$$

By D5,

$$\mathbf{T} + \neg\phi \nvdash \neg Pr_{\mathbf{T}}(\ulcorner\neg\phi \rightarrow 0 = 1\urcorner).$$

Since $\mathbf{T} \vdash (\neg\phi \rightarrow 0 = 1) \rightarrow \phi$, by D1,

$$\mathbf{T} \vdash Pr_{\mathbf{T}}(\ulcorner(\neg\phi \rightarrow 0 = 1) \rightarrow \phi\urcorner).$$

Thus, by D3,

$$\mathbf{T} \vdash Pr_{\mathbf{T}}(\ulcorner\neg\phi \rightarrow 0 = 1\urcorner) \rightarrow Pr_{\mathbf{T}}(\phi).$$

Thus, $\mathbf{T} + \neg\phi \nvdash \neg Pr_{\mathbf{T}}(\phi)$. So $\mathbf{T} \nvdash Pr_{\mathbf{T}}(\phi) \rightarrow \phi$, as desired.