Incompleteness and Undecidability

Dan Hathaway

March 30, 2022

1 Introduction

The purpose of this document is to describe how to prove various results related to Gödel's Incompleteness Theorems and Tarski's Undefinability of Truth.

For simplicity, instead of working with the standard model of arithmetic $\Omega = \langle \omega, 0, 1, <, +, \cdot \rangle$, we work with the hereditarily finite sets together with the set membership relation $\langle H(\omega), \in \rangle$. Here is how these two structures are related:

Definition 1.1. *$\mathcal{L}_\in$ is the language that has the binary relation symbol $\dot{\in}$ as the only non-logical sysmbol. $H(\omega)$ is the set of heriditarily finite sets. We abuse notation and write $H(\omega)$ for the $\mathcal{L}_\in$-structure $\langle H(\omega), \in \rangle$.*

Let $\mathcal{L}_\Omega$ be the language for the structure Ω. In either $\mathcal{L}_\in$ or $\mathcal{L}_\Omega$, for each $n \in \omega$, let $\dot{n}$ be the canonical term that evaluates to n. That is, $\dot{n}^{H(\omega)} = n$ and $\dot{n}^\Omega = n$. These can also be interpreted in non-standard models. We will sometimes have an injection gn from the sentences or formulas of $\mathcal{L}_\in$ to ω ("gn" stands for "Gödel number"). Once the gn function is fixed, we let $\ulcorner \phi \urcorner$ denote $\dot{n}$ where $n = \mathrm{gn}(\phi)$.

For $\mathcal{L}_\in$, we also write $<$ instead of $\in$ when the elements are in the "ω part" of the $\mathcal{L}_\in$ model. This should not cause confusion.

Fact 1.2. *There is a function $F : Form_{\mathcal{L}_\in} \to Form_{\mathcal{L}_\Omega}$ taking n-ary $\mathcal{L}_\in$-formulas to n-ary $\mathcal{L}_\Omega$-formulas (for each n) as well as a function $f : H(\omega) \to \omega$ such that the following hold for each $\varphi \in Form_{\mathcal{L}_\in}$:*

1) *$H(\omega) \models \varphi(a_1, ..., a_n)$ iff $\Omega \models F(\varphi)(f(a_1), ..., f(a_n))$;*
2) *ZF - Infinity $\vdash \varphi(\dot{a}_1, ..., \dot{a}_n)$ iff PA $\vdash F(\varphi)(\dot{f}(a_1), ..., \dot{f}(a_n))$.*

There are also analogous functions $G : Form_{\mathcal{L}_\Omega} \to Form_{\mathcal{L}_\in}$ and $g : \omega \to H(\omega)$.

We also assume that formulas and terms are canonically coded by elements of ω so that given any set S of $\mathcal{L}_\in$-formulas, we may reasonably ask whether or not S is computable. We will sometimes absuse terminology and speak of a formula or a term as if it is a number. Assume that formulas and terms are formed in the standard way so that functions such as $\langle \phi, \psi \rangle \mapsto \phi \wedge \psi$ are computable. We have that the map $n \mapsto \dot{n}$ is computable.

2 A First Look

Suppose a theory T is *computably axiomatizable*, meaning there is a computable set of statements A such that $T = \mathrm{Th}(A)$. Then T is computably enumerable. Also, the set $T^- := \{\varphi : T \vdash \neg\varphi\}$ is computably enumerable. Thus, if T is complete, then since this implies that T and T^- are complements, we have that T is computable. Hence, if we know that T is not computable, then it must be that T is not complete. How can we show that T is not computable? One way is to show there is a complicated set $R \subseteq \omega$ and a formula ϕ such that for all n,

$$n \in R \Leftrightarrow \phi(\dot{n}) \in T.$$

Then, the set R is many-one reducible to T, which implies that T is at least as complicated as R. In particular, if R is not computable, then neither is T, so T is incomplete.

3 Basic Definitions

Definition 3.1. *A theory T is ω-**consistent** if it is consistent and whenever $\phi(\dot{n}) \in T$ for all $n \in \omega$, $\exists x \neg\phi(x) \notin T$. A theory T is ω-**complete** if it is consistent, complete, and whenever $\phi(\dot{n}) \in T$ for all $n \in \omega$, $\forall x \phi(x) \in T$.*

All ω-complete theories are ω-consistent. Any theory that has Ω as a model is ω-consistent.

4 Basic Representability Definitions

Definition 4.1. *Fix an $\mathcal{L}_\in$-theory T. A relation $R \subseteq {}^k\omega$ is T-**representable** iff there exists an $\mathcal{L}_\in$-formula ϕ such that*

$$R(n_1, ..., n_k) \Rightarrow \phi(\dot{n}_1, ..., \dot{n}_k) \in T$$

and

$$\neg R(n_1, ..., n_k) \Rightarrow \neg\phi(\dot{n}_1, ..., \dot{n}_k) \in T.$$

Definition 4.2. *A relation R is **weakly** T-**representable** iff there exists a formula ϕ such that*

$$R(n_1, ..., n_k) \Leftrightarrow \phi(\dot{n}_1, ..., \dot{n}_k) \in T.$$

Definition 4.3. *A function $f : {}^k\omega \to {}^l\omega$ is T-**representable** iff its graph is T-representable by a formula ϕ and such that for all $n_1, ..., n_k \in \omega$,*

$$T \vdash (\exists! y_1, ..., y_l)\, \phi(\dot{n}_1, ..., \dot{n}_k, y_1, ..., y_l).$$

Note that a function being T-representable is stronger than its graph being representable. The extra ingredient is that T proves it is one-valued, not many-valued.

5 When is a Relation T-Representable?

Fact 5.1. *Let T be any theory and R be any relation on ω. If R is T-representable, then $\neg R$ is T-representable.*

Fact 5.2. *If $T_1 \subseteq T_2$ are two consistent theories and R is T_1-representable, then R is T_2-representable.*

Fact 5.3. *If T is computably axiomatized and R is T-representable, then R is computable.*

So the question is now, for a computably axiomalized theory T, when are all computable relations T-representable?

Fact 5.4. *Let Q be Robinson arithemetic. Then every comutable relation on ω is T-representable.*

Corollary 5.5. *Let $T \supseteq Q$ be consistent and computably axiomatized. Then the relations on ω that are T-representable are precisely the computable ones.*

6 When is a Relation weakly T-representable?

Fact 6.1. *If T is a computably axiomatized theory and R is weakly T-representable, then R must be computably enumerable.*

So the question is, for a computably axiomatized theory T, when are all computably enumerable relations T-representable? Here is a partial result:

Proposition 6.2. *If every computable set is T-representable and T is ω-consistent, then every c.e. set is weakly T-representable.*

Proof. Let $E \subseteq \omega$ be an arbitrary c.e. set. Let $R \subseteq \omega \times \omega$ be some computable set such that $E = \{n \in \omega : (\exists m \in \omega)\langle n, m\rangle \in R\}$. Let ϕ be a formula that T-represents R. We claim that for each $n \in \omega$,

$$n \in E \quad \Leftrightarrow \quad (\exists x)\phi(\dot{n}, x) \in T.$$

The $(\Rightarrow)$ direction is immediate. The $(\Leftarrow)$ direction follows by the ω-completeness of T. $\square$

Here is a antother partial result:

Fact 6.3. *Let $T \supseteq Q$ be consistent. Then all computable relations on ω are weakly T-representable.*

Proof. See Fact 5.4 and Fact 8.1. $\square$

But back to the main question: when are all computable enumerable relations T-representable? We have the following:

Theorem 6.4. *Let T be a theory such that*

1) T is consistent,

2) $T \supseteq Q$,

3) T is computably axiomatizable.

Then every computably enumerable relation is weakly T-representable.

Proof. To simplify the proof, fix a computably enumerable *set* $X \subseteq \omega$. Let $S \subseteq \omega \times \omega$ be computable such that

$$X = \{x \in \omega : (\exists y \in \omega)\,(x, y) \in S\}.$$

Let ϕ T-represent S. Since T is a computable enumerable set, fix a computable set $S' \subseteq \omega \times \omega$ such that

$$T = \{p : (\exists y \in \omega)\,(gn(p), y)\}.$$

more!!! □

Here we can put together some previous results:

Corollary 6.5. *Let $T \supseteq Q$ be a consistent theory. Assume either*

(A) T is ω-consistent, or

(B) T is c.e..

Then every c.e. set is weakly T-representable.

Note that (A) happens if $T \subseteq \mathrm{Th}(\Omega)$, because then T has Ω as a model.
Question: if T is a consistent theory and every comutably enumerable relation is weakly T-representable, then is every computable relation T-representable?

7 When is a Function T-representable?

Fact 7.1. *Every computable function $f : \omega^k \to \omega^l$ is Q-representable.*

Corollary 7.2. *Let $T \supseteq Q$ be consistent and computably axiomatized. Then the functions $f : \omega^k \to \omega^l$ that are T-representable are precisely the computable ones.*

8 Normal vs Weak Representability

Fact 8.1. *Assume T is a consistent theory. Let R be a relation. Then if R is T-representable, then R and $\neg R$ are both weakly T-represntable.*

Question 8.2. *Let T be a consistent theory. Let R be a relation. If R and $\neg R$ are both weakly T-representable, then is R T-representable? What is a counterexample?*

9 A Theory Cannot Represent Itself

There can exist an $R \subseteq \omega$ that is weakly T-representable even though $\neg R$ is not.

If T is a complete theory, then any weakly T-representable relation is also T-representable.

We have the following diagonal argument:

Theorem 9.1. *Let $\langle \chi_n : n \in \omega \rangle$ be any enumeration of all formulas with one free variable in the language $\mathcal{L}_\in$. For any consistent $\mathcal{L}_\in$-theory T, the relation $R = \{\langle n, m \rangle : \chi_n(\dot{m}) \in T\}$ is not T-representable. In fact, $\neg R$ is not weakly T-representable.*

Proof. Suppose towards a contradiction that R is T-representable. This implies that $\neg R$ is T-representable. Since T is consistent, $\neg R$ is weakly T-representable. Let ϕ be a formula that weakly T-represents $\neg R$. That is,

$$\chi_n(\dot{m}) \notin T \Leftrightarrow \phi(\dot{n}, \dot{m}) \in T.$$

Consider the formula $\phi(x, x)$. This is a formula with one free variable x, so $\phi(x, x) = \chi_p(x)$ for some $p \in \omega$. We have

$$\chi_p(\dot{p}) \notin T \Leftrightarrow \phi(\dot{p}, \dot{p}) \in T \Leftrightarrow \chi_p(\dot{p}) \in T.$$

This is a contradiction because T is consistent. □

On the other hand, it is quite possible for the R in the above theorem to be weakly T-representable, as T could be computably axiomalizable and T could weakly represent every computably enumberable relation.

We may remember the above theorem by thinking "a theory cannot represent itself", although this is not exactly what the theorem says. The question of whether a theory can represent the set of all Gödel numbers of its true *sentences* is addressed in the next section. For our purposes now, the above theorem is cleaner and it has various important consequences:

Corollary 9.2 (Weak Form of Tarski's Undefinability of Truth)**.** *Let $\mathfrak{A}$ be any $\mathcal{L}_\in$-structure. Let $\langle \chi_n : n \in \omega \rangle$ be any enumeration of all formulas with one free variable in the language $\mathcal{L}_\in$. The relation $\{\langle n, m \rangle : \chi_n(\dot{m}) \in Th(\mathfrak{A})\}$ is not definable over $\mathfrak{A}$.*

Proof. A relation is definable over $\mathfrak{A}$ iff it is $\mathrm{Th}(\mathfrak{A})$-representable. □

Note that the usual statement of Tarski's Undefinability of Truth is that the set of Gödel number of *statements* true in $\mathfrak{A}$ is not definable in $\mathfrak{A}$. We will explain how to derive this result in the next section. The following is the connection with undecidability:

Corollary 9.3. *If T is any consistent $\mathcal{L}_\in$-theory such that every computable relation on ω is T-representable, then T is not computable.*

Proof. Suppose towards a contradiction that T is computable (computable). Let $\langle \chi_n : n \in \omega \rangle$ be an enumeration of the $\mathcal{L}_\in$-formulas with one free variable such that the map $\langle n, m \rangle \mapsto \chi_n(\dot{m})$ is computable. Since T is computable, so is the relation $R := \{\langle n, m \rangle : \chi_n(\dot{m}) \in T\}$. Since every computable relation is T-representable, R is T-representable, but this contradicts the above theorem. $\square$

Corollary 9.4. *The theory Th$(H(\omega))$ is not computable.*

Proof. Every computable relation on ω is $\mathrm{Th}(H(\omega))$-representable. $\square$

From this last corollary, it follows by completeness that $\mathrm{Th}(H(\omega))$ is not computably axiomatizable. Hence, the theory generated by any computable subset of $\mathrm{Th}(H(\omega))$ is incomplete.

Notice that if a relation is T_1-representable and $T_2 \supseteq T_1$, then it is also T_2 representable. We can sharpen the corollary above by finding a subtheory T of $\mathrm{Th}(H(\omega))$ such that every computable relation on ω is T-representable. The $\mathcal{L}_\in$-theory that does the trick is *Robinson Arithmetic*, which we shall denote by Q. There are only two properties of Q which are important:

1) Q is finitely axiomatizable;

2) Any model $\mathfrak{A}$ of Q has an initial segment that is isomorphic to $H(\omega)$. Moreover, given any $n \in \omega$, $\dot{n}^{\mathfrak{A}}$ is in this initial segment and

$$\{a \in \mathfrak{A} : a < \dot{n}^{\mathfrak{A}}\} = \{\dot{m}^{\mathfrak{A}} : m < n\}.$$

Proposition 9.5. *Every computable relation on ω is Q-representable. In fact, every computable function $f : {}^k\omega \to {}^l\omega$ is Q-representable.*

Proof. It suffices to show that every function is Q-representable, because a representation for the characteristic function of a relation immediately yields a representation for the relation itself. We will not go into details, but a key property of Q that makes the result possible is 2) above, which implies that if $\phi(x)$ is a formula and $n \in \omega$ is such that $\phi(\dot{n}) \in \mathrm{Q}$ but $\neg\phi(\dot{m}) \in \mathrm{Q}$ for all $m < n$, then

$$\forall x[[\phi(x) \wedge (\forall y < x)\neg\phi(y)] \leftrightarrow x = \dot{n}] \in \mathrm{Q}.$$

$\square$

Corollary 9.6. *Any consistent $\mathcal{L}_\in$-theory $T \supseteq Q$ is not computable.*

Proof. Since $T \supseteq \mathrm{Q}$ and every computable relation is Q-representable, every computable relation is T-representable. Since also T is consistent, we may apply Corollary 9.3 to get that T is not computable. $\square$

Since Q is finitely axiomatizable, we have the following striking generalization of the last corollary:

Corollary 9.7. *If T is any $\mathcal{L}_\in$-theory such that $T \cup Q$ is consistent, then T is not computable. In particular, the set of all logical validities of the language $\mathcal{L}_\in$ is not computable.*

Proof. Let θ_{Q} be the conjunction of a finite set of axioms for Q. For any sentence ϕ, we have

$$\phi \in \mathrm{Th}(T \cup \mathrm{Q}) \Leftrightarrow (\theta_{\mathrm{Q}} \rightarrow \phi) \in T.$$

If T was decidable, then since the map $\phi \mapsto (\theta_{\mathrm{Q}} \rightarrow \phi)$ is computable, $T \cup \mathrm{Q}$ would be be decidable as well. This contradicts the previous corollary. □

This tells us something about incompleteness:

Corollary 9.8. *If T is any computably enumerable $\mathcal{L}_{\in}$-theory such that $T \cup Q$ is consistent, then T is incomplete.*

Proof. By the previous corollary, T is undecidable. Since T is computably enumerable, so is $\{\phi : \neg\phi \in T\}$. Since T is not computable, these sets cannot be compliments of each other, so T is incomplete. □

The last two corollaries are generalizations of Gödel's First Incompleteness Theorem. We can further generalize these by considering relative interpretability of theories. For details, see [1].

10 Tarski's Undefinability of Truth

In the previous section we proved a weak version of Tarski's Undefinability of Truth. One consequence of the definition of a function being representable is the following:

Proposition 10.1. *Let $f : {}^k\omega \rightarrow {}^l\omega$ be T-representable. If $A \subseteq {}^l\omega$ is T-representable, then $f^{-1}(A)$ is T-representable. If $B \subseteq {}^l\omega$ is weakly T-representable, then $f^{-1}(B)$ is weakly T-representable.*

Proof. If T is inconsistent, then every relation is T-representable and no relation is weakly T-representable, so there is nothing to prove. Thus, assume T is consistent. Let f be T-represented by the formula ϕ. If A is T-represented by the formula ψ, then

$$\eta(x_1, ..., x_k) := (\exists y_1, ..., y_l)\, \phi(x_1, ..., x_k, y_1, ..., y_l) \wedge \psi(y_1, ..., y_l)$$

T-represents $f^{-1}(A)$. If B is weakly T-represented by the formula ψ, then the same formula η weakly T-represents $f^{-1}(B)$. □

We can now apply this to Theorem 9.1.

Theorem 10.2. *Let T be a consistent $\mathcal{L}_{\in}$-theory. Let gn be any injection from the set of all $\mathcal{L}_{\in}$-sentences to ω. If there exists an enumeration $\langle \chi_n : n \in \omega \rangle$ of the formulas with one free variable in the language L such that the function $\langle n, m \rangle \mapsto gn(\chi_n(\dot{m}))$ is T-representable, then the set $S = \{gn(\phi) : \phi \in T\}$ is not T-representable.*

Proof. If S was T-representable, then by the above proposition the relation $R := \{\langle n, m\rangle : \chi_n(\dot{m}) \in T\}$ would be T-representable, but this would contradict Theorem 9.1. $\square$

This result implies that if a theory represents enough functions, it will be unable to represent its set of true sentences.

Corollary 10.3. *Let T be a consistent $\mathcal{L}_\in$-theory. Let gn be a computable injection from the set of all $\mathcal{L}_\in$-sentences to ω. If every computable function is T-representable, then the set $S = \{gn(\phi) : \phi \in T\}$ is not T-representable.*

Proof. There exists a computable enumeration $\langle \chi_n : n \in \omega\rangle$ of the formulas with one free variable in the language $\mathcal{L}_\in$ and the function $\langle n, m\rangle \mapsto \chi_n(\dot{m})$ is computable. Hence, the function $\langle n, m\rangle \mapsto \mathrm{gn}(\chi_n(\dot{m}))$ is computable, and therefore T-representable. So by the above theorem, S is not T-representable. $\square$

Recall that $A \subseteq \omega$ is definable over a model $\mathfrak{A}$ iff A is $\mathrm{Th}(\mathfrak{A})$-representable.

Corollary 10.4. *[Tarski's Undefinability of Truth] Let gn be a computable injection from the set of all $\mathcal{L}_\in$-sentences to ω. If $\mathfrak{A}$ is an $\mathcal{L}_\in$-structure such that every computable function is Th($\mathfrak{A}$)-representable, then the set $S = \{gn(\phi) : \phi \in Th(\mathfrak{A})\}$ is not definable over $\mathfrak{A}$.*

Thus, identifying $\mathrm{Th}(H(\omega))$ with $\{\mathrm{gn}(\phi) : \phi \in \mathrm{Th}(H(\omega))\}$, we have that $\mathrm{Th}(H(\omega))$ is not definable over $H(\omega)$.

Using Parameters To Define Truth

For this subsection, we will use some notation defined at the beginning of section 3. A natural question is whether a model $\mathfrak{A}$ can ever be such that there exists a formula θ and a parameter $a \in |\mathfrak{A}|$ such that

$$\mathfrak{A} \models \phi \leftrightarrow \theta(\ulcorner \phi \urcorner, a)$$

for every sentence ϕ. This is certainly the case when $\mathfrak{A} = H(\omega_1)$. It is also possible to construct models with this property using a compactness argument.

That is, let $\mathcal{L}_\in^+$ be the language $\mathcal{L}_\in$ with the additional constant symbol c. The $\mathcal{L}_\in$-theory given by

$$\begin{array}{rl} & \mathrm{Th}(H(\omega)) \\ \cup & \{\ulcorner \phi \urcorner \in c : \phi \in \mathrm{Th}(H(\omega))\} \\ \cup & \{\ulcorner \phi \urcorner \notin c : \phi \notin \mathrm{Th}(H(\omega))\} \end{array}$$

is finitely consistent, because every finite subset of this set of sentences has $(H(\omega), s)$ as a model for some $s \in H(\omega)$. By compactness, the theory is consistent. If we let $\mathfrak{A}$ be any model of this theory, $a = c^{\mathfrak{A}}$, and $\theta(x, y)$ be the formula $\theta(x, y) := x \in y$, then we have for all ϕ that $\mathfrak{A} \models \phi \leftrightarrow \theta(\ulcorner \phi \urcorner, a)$.

The next natural question is whether parameters can be used to define the truth of formulas *with parameters*. This is *not* the case. That is, given a model $\mathfrak{A}$ such that every computable function is $\mathrm{Th}(\mathfrak{A})$-representable, it is not the case that there exists a formula θ and a parameter $b \in |\mathfrak{A}|$ such that

$$\mathfrak{A} \models \phi(a) \leftrightarrow \theta(\ulcorner \phi \urcorner, a, b)$$

for all ϕ and $a \in |\mathfrak{A}|$. To see why, suppose towards a contradiction that this is possible. Given such an $\mathcal{L}_\in$-structure $\mathfrak{A}$, a formula θ, and a parameter b, by letting $a = b$, we have $\mathfrak{A} \models \phi(b) \leftrightarrow \theta(\ulcorner \phi \urcorner, b, b)$ for all ϕ. Expand $\mathcal{L}_\in$ to the language $\mathcal{L}'_\in := \mathcal{L}_\in \cup \{c\}$ where c is a new constant symbol and let $(\mathfrak{A}, b)$ be the expansion of $\mathfrak{A}$ where $c^{(\mathfrak{A},b)} = b$. We still have that every computable function is representable in the $\mathcal{L}'_\in$-structure $(\mathfrak{A}, b)$. Now Corollary 10.4 applies not only to $\mathcal{L}_\in$-structures but to $\mathcal{L}'_\in$-structures. We get a contradiction to Corollary 10.4 because the $\mathcal{L}'$-formula $\theta'(x) := \theta(x, c, c)$ is defining the truth of all $\mathcal{L}'_\in$-sentences.

Note that the argument we gave might seem to only work because we had a single parameter b. If instead we have many parameters $b_1, ..., b_n$, we could modify the previous proof as long as $\mathfrak{A}$ can sufficiently code tuples of elements into single elements, which is a mild assumption.

11 The Self Reference Lemma

One can think of the Self Reference Lemma as the analogue of the recursion theorem from computability theory. Their proofs are similar, but I see no way of proving both from a common result.

Lemma 11.1 (Self Reference Lemma)**.** *Let T be a theory in which every computable function is T-representable. Given any formula θ with one free variable, there is a sentence η such that*

$$T \vdash \eta \leftrightarrow \theta(\ulcorner \eta \urcorner).$$

Proof. Let $f : \omega \to \omega$ be the function defined by $f(n) := \mathrm{gn}(\mathrm{gn}^{-1}(n)(\dot{n}))$. This function is computable, so by hypothesis it is T-representable. Let ϕ be a formula that T-represents R. That is, $f(n) = m \Rightarrow T \vdash \phi(\dot{n}, \dot{m})$, $f(n) \neq m \Rightarrow T \vdash \neg\phi(\dot{n}, \dot{m})$, and for all $n \in \omega$, $T \vdash \exists! y \phi(\dot{n}, y)$. Let $n \in \omega$ be such that $\mathrm{gn}^{-1}(n)$ is the formula $\forall y[\phi(x, y) \to \theta(y)]$. This means that $\mathrm{gn}^{-1}(n)(\dot{n})$ is the sentence $\forall y[\phi(\dot{n}, y) \to \theta(y)]$, so of course

$$T \vdash \mathrm{gn}^{-1}(n)(\dot{n}) \leftrightarrow \forall y[\phi(\dot{n}, y) \to \theta(y)].$$

Let $m := f(n)$. Since $T \vdash \phi(\dot{n}, \dot{m})$ and $T \vdash \exists! y \phi(\dot{n}, y)$, any sentence ψ that has $\phi(\dot{n}, y)$ as a subformula is provably equivalent over T to the sentence that is ψ but with the subformula $\phi(\dot{n}, y)$ replaced with $y = \dot{m}$. Hence,

$$T \vdash \mathrm{gn}^{-1}(n)(\dot{n}) \leftrightarrow \forall y[y = \dot{m} \to \theta(y)].$$

Since $\forall y[y = \dot{m} \to \theta(y)]$ is logically equivalent to $\theta(\dot{m})$, we have

$$T \vdash \mathrm{gn}^{-1}(n)(\dot{n}) \leftrightarrow \theta(\dot{m}).$$

Let us define $\eta := \mathrm{gn}^{-1}(n)(\dot{n})$. We have $\ulcorner \eta \urcorner = \dot{m}$, so

$$T \vdash \eta \leftrightarrow \theta(\ulcorner \eta \urcorner).$$

□

12 Sentences Witnessing Incompleteness

In section 1, we got a lot of mileage out of the diagonal argument that a theory cannot represent itself. In this section, we will show how to construct particular sentences that are independent. One reason for doing this is so that in the future, more natural statements can be shown to be independent by relating them to the ones given here.

We will now, once and for all, fix a Gödel numbering gn of all L-formulas.

Definition 12.1. *Given any formula ϕ, let $gn(\phi)$ be the Gödel number of ϕ. Assume gn is computable, 1-1, and maps onto ω. Assume also that for each k, the map*

$$\langle n, m_1, ..., m_k \rangle \mapsto gn[(gn^{-1}(n))(\dot{m}_1, ..., \dot{m}_k)]$$

is computable. In the above, if ϕ is a formula with $l < k$ free variables, then let $\phi(\dot{m}_1, ..., \dot{m}_k) := \phi(\dot{m}_1, ..., \dot{m}_l)$. For each formula ϕ, let $\ulcorner \phi \urcorner := \dot{n}$ where $n = gn(\phi)$.

The following is a way to get a sentence witnessing incompleteness:

Theorem 12.2. *Let T be a consistent c.e. theory such that every c.e. set is weakly T-representable. If ψ is any formula that weakly T-represents the set $X := \{n : (gn^{-1}(n))(\dot{n}) \in T\}$, then the sentence $\psi(\ulcorner \neg\psi \urcorner)$ is independent of T.*

Proof. The set X is c.e. because of our requirements on the function gn. Since every c.e. set is weakly T-representable, there is some formula ψ that weakly T-represents X. We have the following:

$$\psi(\ulcorner \neg\psi \urcorner) \in T \Leftrightarrow \mathrm{gn}(\neg\psi) \in X \Leftrightarrow \neg\psi(\ulcorner \neg\psi \urcorner) \in T.$$

The first ($\Leftrightarrow$) is because ψ weakly T-represents X. The second ($\Leftrightarrow$) is the definition of X. Since T is consistent, neither $\psi(\ulcorner \neg\psi \urcorner)$ nor $\neg\psi(\ulcorner \neg\psi \urcorner)$ is in T. □

In order to apply this theorem to a theory T, we need to show that every c.e. set is T-representable. We showed several ways to do this in Section 6 (see for example Corollary 6.5). Here is what we now have:

Corollary 12.3. *Let $T \supseteq Q$ be a consistent c.e. theory. Then every c.e. set is weakly T-representable, and hence the formula $\psi(\ulcorner \neg\psi \urcorner)$ defined in the previous theorem is independent of T.*

more!!!

13 The Second Incompleteness Theorem

more!!!

References

[1] Hinman, Peter G. *Fundamentals of Mathematical Logic*. A K Peters, Ltd, Wellesley, MA, 2005.