# Incompleteness and Undecidability

Dan Hathaway

May 11, 2012

## 0. Purpose

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. Care has been taken to isolate theorems that use diagonalization arguments. In sections 1 and 2, only Theorem 1.2 uses diagonalization, and this result is heavily exploited.

## 1. A Theory Cannot Represent Itself

Every theory we will consider will be in a language that has a name $\dot{n}$ for each $n \in \omega$. For example, the name for $2 \in \omega$ in the language $L_\Omega$ of arithmetic is the term "$1+1$". The results stated will only apply to countable languages. This requirement is implicit in the hypothesis of the existence of an enumeration of all formulas with one free variable, for example. We also assume that formulas and terms are literally finite sequences of numbers, so that given any set $S$ of $L$-formulas, we may reasonably ask whether $S$ is recursive or not. 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 recursive. Assume also that the map $n \mapsto \dot{n}$ is recursive.

**Definition 1.1.** *A relation $R \subseteq {}^k\omega$ is $T$-**representable** iff there exists an $L$-formula $\phi$ such that*

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

*and*

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

*A relation $R$ is **weakly** $T$-**representable** iff there exists a formula $\phi$ such that*

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

If $R$ is $T$-representable, then so is $\neg R$. If $T$ is a consistent theory, then any $T$-representable relation is also weakly $T$-representable. The following is a related concept that we will have use for later:

**Definition 1.2.** *A function $f : {}^k\omega \to {}^l\omega$ is $T$-**representable** iff its graph is $T$-representable by a formula $\phi$ such that for all $n_1, ..., n_k \in \omega$,*

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

We have the following diagonal argument:

**Theorem 1.3.** *Let $\langle \chi_n : n \in \omega \rangle$ be any enumeration of all formulas with one free variable in the language $L$. For any consistent $L$-theory $T$, the relation $R = \{\langle n, m \rangle : \chi_n(\dot{m}) \in T\}$ is not $T$-representable.*

*Proof.* Suppose, towards a contradiction, that $R$ is $T$-representable. This means that $\neg R$ is $T$-representable. Since $T$ is consistent, $\neg R$ is weakly $T$-representable. Let $\phi$ 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. □

What is really shown in the above theorem is that for any consistent theory $T$, the relation $\neg R$ is not weakly $T$-representable. On the other hand, it is possible for $R$ to be weakly $T$-representable (take for instance PA which can weakly represent every r.e. set).

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, the above theorem is cleaner and it has various deep consequences which we will go into now.

**Corollary 1.4** (Weak Form of Tarski's Undefinability of Truth)**.** *Let $\mathfrak{A}$ be any $L$-structure. Let $\langle \chi_n : n \in \omega \rangle$ be any enumeration of all formulas with one free variable in the language $L$. 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 in $T$ is not $T$-representable. We will explain how to derive this result in the next section. The following is the connection with undecidability:

**Corollary 1.5.** *If $T$ is any consistent $L$-theory such that every recursive relation is $T$-representable, then $T$ is undecidable.*

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

**Corollary 1.6.** *The theory Th($\Omega$) is undecidable, where $\Omega$ is the standard model of arithmetic.*

*Proof.* Every recursive relation is Th($\Omega$)-representable. $\square$

From this last corollary, it follows by completeness that Th($\Omega$) is not recursively axiomatizable. Hence, the theory generated by any recursive subset of Th($\Omega$) is incomplete.

Notice that if a relation is $T_1$-representable and $T_2 \supseteq T_1$, then it is also $T_2$ representable. Thus, we can sharpen the above corollary by finding a subtheory $T$ of Th($\Omega$) such that every recursive relation is $T$-representable. The theory that does the trick is Robinson Arithmetic, which we shall denote by Q.

**Proposition 1.7.** *Every recursive relation is Q-representable. In fact, every recursive function 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 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

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

$\square$

A good way of understanding why the above proposition is true is to remember that every model of Q has an initial segment isomorphic to the standard model of arithmetic.

**Corollary 1.8.** *Any consistent $L_\Omega$-theory $T \supseteq Q$, is undecidable.*

*Proof.* Since $T \supseteq \mathrm{Q}$ and every recursive relation is Q-representable, every recursive relation is $T$-representable. Since also $T$ is consistent, we may apply Corollary 1.4 to get that $T$ is undecidable. $\square$

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

**Corollary 1.9.** *If $T$ is any $L_\Omega$-theory such that $T \cup Q$ is consistent, then $T$ is undecidable. In particular, the set of all logical validities of the language $L_\Omega$ is undecidable.*

*Proof.* Let $\theta_{\mathrm{Q}}$ be the conjunction of the finitely many axions of Q. For any sentence $\phi$, we have

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

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

Of course, we have the following application of that last corollary to incompleteness:

**Corollary 1.10.** *If $T$ is any r.e. $L_\Omega$-theory such that $T \cup Q$ is consistent, then $T$ is incomplete.*

*Proof.* By the previous corollary, $T$ is undecidable. Since $T$ is r.e., so is $\{\phi : \neg\phi \in T\}$. Since $T$ is not recursive, these sets cannot be compliments of each other, so $T$ is incomplete. $\square$

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].

# 2. Tarski's Undefinability of Truth

In the previous section we proved a weak version of Tarski's Undefinability of Truth. The notion of a function being representable will allow us to use Theorem 1.2. One consequence of the definition of a function being representable is the following:

**Proposition 2.1.** *Let $f : {}^k\omega \to {}^l\omega$ 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 $\phi$. If $A$ is $T$-represented by the formula $\psi$, 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 $\psi$, then the same formula $\eta$ weakly $T$-represents $f^{-1}(B)$. $\square$

We can now apply this to Theorem 1.2.

**Theorem 2.2.** *Let $T$ be a consistent $L$-theory. Let $gn$ be any injection from the set of all $L$-sentences to $\omega$. 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 1.2. $\square$

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

**Corollary 2.3.** *Let $T$ be a consistent $L$-theory. Let $gn$ be a recursive injection from the set of all $L$-sentences to $\omega$. If every recursive function is $T$-representable, then the set $S = \{gn(\phi) : \phi \in T\}$ is not $T$-representable.*

*Proof.* We may assume $L$ is countable, otherwise there could be no such injection gn. Because of this, we have that there exists a recursive enumeration $\langle \chi_n : n \in \omega \rangle$ of the formulas with one free variable in the language $L$ such that the function $\langle n, m \rangle \mapsto \chi_n(\dot{m})$ is recursive. Hence, the function $\langle n, m \rangle \mapsto \mathrm{gn}(\chi_n(\dot{m}))$ is recursive. Since every recursive function is $T$-representable, so is this one. By the above theorem, $S$ is not $T$-representable. $\square$

Applying this to definability, we have the following:

**Corollary 2.4** (Tarski's Undefinability of Truth). *Let gn be a recursive injection from the set of all $L$-sentences to $\omega$. If $\mathfrak{A}$ is an $L$-structure such that every recursive function is Th$(\mathfrak{A})$-representable, then the set $S = \{gn(\phi) : \phi \in Th(\mathfrak{A})\}$ is not definable over $\mathfrak{A}$.*

*Proof.* This follows immediately from the last corollary. $\square$

In particular, given any such function $gn$, the set $\{\mathrm{gn}(\phi) : \phi \in \mathrm{Th}(\Omega)\}$ is not definable over $\Omega$.

## 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 $\theta$ and a parameter $a \in |\mathfrak{A}|$ such that

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

for every sentence $\phi$. This is certainly the case when $\mathfrak{A}$ is the standard model of *second* order arithmetic. It is also possible to have this with a model of first order arithmetic using a compactness argument. That is, let $L$ be the language of arithmetic together with the constant symbol $c$. The $L$-theory given by

$$\begin{array}{ll} & \mathrm{Th}(\Omega) \\ \cup & \{\text{“}\dot{n} \text{ divides } c\text{”} : n = \mathrm{gn}(\phi) \text{ and } \phi \in \mathrm{Th}(\Omega)\} \\ \cup & \{\text{“}\dot{n} \text{ does not divide } c\text{”} : n = \mathrm{gn}(\phi) \text{ and } \phi \notin \mathrm{Th}(\Omega)\} \end{array}$$

is finitely consistent, because every finite subset of this set of sentences has $(\Omega, m)$ as a model for some $m \in \omega$. By compactness, the theory is consistant. The reduct $\mathfrak{A}$ of this structure to the language or arithmetic is the desired model with $a = c^{\mathfrak{A}}$ and $\theta(x, y) :=$“$x$ divides $y$”.

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 recursive function is Th$(\mathfrak{A})$-representable, it is not the case that there exists a formula $\theta$ and a parameter $b \in |\mathfrak{A}|$ such that

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

for all $\phi$ and $a \in |\mathfrak{A}|$. To see why, suppose towards a contradiction that this is possible. Given such an $L$-structure $\mathfrak{A}$, a formula $\theta$, and a parameter $b$, by letting $a = b$, we have $\mathfrak{A} \models \phi(b) \leftrightarrow \theta(\ulcorner\phi\urcorner, b, b)$ for all $\phi$. Expand $L$ to the language $L' := L \cup \{c\}$ 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 recursive function is representable in the $L'$-structure $(\mathfrak{A}, b)$. Hence, we get a contradiction by Corollary 2.4. because the $L'$-formula $\theta'(x) := \theta(x, c, c)$ is defining the truth of all $L'$-sentences.

Note that the argument we gave seems 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.

# 3. 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 3.1.** *Given any formula $\phi$, let $gn(\phi)$ be the Gödel number of $\phi$. Assume gn is recursive, 1-1, and maps onto $\omega$. 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 recursive. In the above, if $\phi$ 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 $\phi$, let $\ulcorner\phi\urcorner := \dot{n}$ where $n = gn(\phi)$.*

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

**Theorem 3.2.** *Let $T$ be a consistent r.e. theory such that every r.e. set is weakly $T$-representable. If $\psi$ 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 r.e. because of our requirements on the function gn. Since every r.e. set is weakly $T$-representable, there is some formula $\psi$ 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 $\psi$ 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 $T$. $\square$

In order to apply this theorem to a theory $T$, we need to show that every r.e. set is $T$-representable. Examples of theories for which this is true are $\omega$-consistent theories in which all recursive sets are representable and recursively axiomatizable theories extending Q. We will consider $\omega$-consistent theories first.

**Definition 3.3.** *A theory $T$ is $\omega$-**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 $\omega$-**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 $\omega$-complete theories are $\omega$-consistent. Any theory that has $\Omega$ as a model is $\omega$-consistent.

**Proposition 3.4.** *If every recursive set is $T$-representable and $T$ is $\omega$-consistent, then every r.e. set is weakly $T$-representable.*

*Proof.* Let $E \subseteq \omega$ be an arbitrary r.e. set. Let $R \subseteq \omega \times \omega$ be some recursive set such that $E = \{n \in \omega : (\exists m \in \omega)\langle n, m \rangle \in R\}$. Let $\phi$ 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 $\omega$-completeness of $T$. $\square$

At this point, we can apply the above theorem to any theory $T$ satisfying $\mathrm{Q} \subseteq T \subseteq \mathrm{Th}(\Omega)$ (every recursive set must be $T$-representable and $T$ must be $\omega$-consistent). To get an analogue of the above proposition for recursively axiomatizable theories extending Q, we need the following technical lemma which will exploit the fact that a particular recursive *function* is Q-representable.

**Lemma 3.5** (Self Reference Lemma)**.** *Let $T$ be a theory in which every recursive function is $T$-representable. Given any formula $\theta$ with one free variable, there is a sentence $\eta$ 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 recursive, so by hypothesis it is $T$-representable. Let $\phi$ 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 $\psi$ that has $\phi(\dot{n}, y)$ as a subformula is provably equivalent over $T$ to the sentence that is $\psi$ 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

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

$\square$

more!!!

## 4. The Second Incompleteness Theorem

more!!!

## References

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