Incompleteness Theorems

Theodoros Papamakarios
papamakarios@uchicago.edu
March 12, 2019

Abstract

The aim of this text is to present gradually stronger versions of Gödel’s incompleteness theorem. Starting with a version based on the fact that arithmetical truth cannot be expressed by a $\Sigma_1$ formula, Gödel’s original result based on $\omega$-consistency, and Rosser’s theorem, which replaces the $\omega$-consistency condition with plain consistency, are presented. Extra care was given to showcase the heart of the proofs, suspending technical details. Our exposition generally follows [2].

1 What is a proof?

Well, whatever it is, it should be easily verifiable, where for our purposes, “easily” means “in an algorithmic way”. More precisely, if for a proof system $\mathcal{S}$, $P_{\mathcal{S}}$ is a predicate such that

$$P_{\mathcal{S}}(x, y) \iff y \text{ is a proof of } x \text{ in } \mathcal{S},$$

then it must be the case that $P_{\mathcal{S}}$ is recursive. $P_{\mathcal{S}}$ being recursive implies that

$$Q_{\mathcal{S}} := \{ x : \exists y P_{\mathcal{S}}(x, y) \},$$

the set of all provable in $\mathcal{S}$ formulas is recursively enumerable. Now consider $T$, the set of all formulas in the language of arithmetic, which are true in the standard model. We will see that $T$ is not recursively enumerable. $T$ is not r.e., but $Q_{\mathcal{S}}$ is r.e. for any $\mathcal{S}$, so there cannot be a proof system for which $Q_{\mathcal{S}}$ is the same as $T$. In particular, for a system that proves only true (in the standard model) formulas, there must be a true but not provable formula, and this is a weak form of Gödel’s result.

2 Tarski’s theorem

A way to show that $T$ is not r.e. is to reduce a known non r.e. set (like the complement of the halting set) to it. Such a reduction is described in Lemma 39.1 of [1]. We will show something stronger. That not only $T$ is not r.e., but does not belong to any level of the arithmetical hierarchy.

First to set the stage: By number, we mean natural number. By formula, we mean formula in the language of arithmetic, i.e. the language containing the constant symbol 0, the function symbols $+$, $\cdot$ and $s$, and the predicate symbols $=$ and $\leq$. Saying that a formula is true, we mean true in the standard model of arithmetic, i.e., the model with domain the set of natural numbers, where 0 is interpreted as zero, $+$ as addition, $\cdot$ as multiplication, $s$ as the successor function etc. We say that a set $A$ of numbers is arithmetic if there is a formula expressing it, i.e. a formula $F(v)$ with only $v$ as its free variable such that for any number $n$,

$$F(\overline{n}) \text{ is true } \iff n \in A,$$
where $F(\overline{n})$ is $F$ where all free occurrences of $v$ are replaced by the expression

$$\overline{n} \overset{\text{def}}{=} \underbrace{s \ldots s}_{n \text{ times}} 0.$$  

Similarly, we say that a relation $R$ between numbers is arithmetic if there is a formula $F(v_1, \ldots, v_k)$ with $v_1, \ldots, v_k$ as its only free variables such that for any numbers $n_1, \ldots, n_k$,

$$F(\overline{n_1}, \ldots, \overline{n_k}) \text{ is true } \iff R(n_1, \ldots, n_k).$$

Tarski’s theorem says that the set of all numbers encoding true formulas is not arithmetic.

We write $F_x$ for the formula encoded by the number $x$. Let $d(x, y)$ be the number encoding the formula

$$F_x[\overline{y}] \overset{\text{def}}{=} \forall v \ (v = \overline{y} \rightarrow F_x).$$

Notice that $F_x[\overline{y}]$ is equivalent to $F_x(\overline{y})$. The function $d$ plays a crucial role in incompleteness theorems. An important fact is:

**Proposition 2.1.** The relation $z = d(x, y)$ is arithmetic.

**Proof sketch.** If our language contained exponentiation the proof would be straightforward. Exponentiation allows us to express number concatenation, denoted with $\ast$. And having such power we could first express $\overline{y}$, setting for convenience the digit for $b - 1$ in our base $b$ number system to encode $s$, as $g(y) \overset{\text{def}}{=} (b^y - 1) \ast c[0]$, and then express $z = d(x, y)$ as

$$\exists w \ (w = g(y) \land z = k \ast w \ast c[\rightarrow] \ast x),$$

where $c[\sigma]$ is the number encoding the string $\sigma$, and $k$ is $c[\forall v \ (y = \cdot)]$. But our language does not contain exponentiation. One way to circumvent that difficulty would be to augment our language and model to include exponentiation. Or we could try to express exponentiation using $+$ and $\cdot$ alone. This is possible, but not simple; it amounts to expressing $d(x, y)$ directly with $+$ and $\cdot$; see Chapter 4 of [2] on how it can be done.

For any set of numbers $A$, define $A^*$ as

$$n \in A^* \overset{\text{def}}{=} d(n, n) \in A.$$  

**Proposition 2.1** allows us to prove:

**Proposition 2.2.** If $A$ is arithmetic, then $A^*$ is arithmetic.

**Proof.** We can express $A^*$ as

$$\exists y \ (y = d(x, x) \land y \in A).$$

From **Proposition 2.1**, there is a formula expressing $y = d(x, x)$, so if there is a formula expressing $A$, there is a formula expressing $A^*$.

Now Tarski’s theorem can be proved as follows. Let $T$ be the set of all numbers encoding the true formulas. We claim that there cannot be a formula expressing $\overline{T}$: If $F(v)$ is such a formula, then for any number $n$,

$$F(\overline{n}) \text{ is true } \iff n \in \overline{T}$$

$$\iff d(n, n) \in T$$

$$\iff d(n, n) \notin T$$

$$\iff F_n(\overline{n}) \text{ is not true}$$

$$\iff F_n(\overline{n}) \text{ is not true}.$$  

But for the number $f$ encoding $F(v)$, we get

$$F(\overline{f}) \text{ is true } \iff F(\overline{f}) \text{ is not true},$$

which is absurd. Therefore $\overline{T}$ is not arithmetic. So from **Proposition 2.2**, $\overline{T}$ is not arithmetic, and since the negation of a formula expressing $A$ expresses $\overline{A}$, $T$ cannot be arithmetic.
3 Gödel’s proof

Systems the set of provable formulas of which is r.e. are often called recursively axiomatizable. We saw that Tarski’s theorem implies that for a system which

1. proves only true formulas,
2. is recursively axiomatizable,

there must be a true but not provable theorem. Gödel’s proof is sharper in at least two ways. First, it replaces assumption 1 with a weaker assumption, that of ω-consistency: A proof system is ω-inconsistent if for some formula \( F(v) \), it proves \( \exists v F(v) \) and also proves all formulas \( \neg F(0), \neg F(1), \ldots \); it is ω-consistent if it is not ω-inconsistent. Secondly, it explicitly constructs a true but not provable formula. Gödel’s proof is often presented for the particular system of Peano arithmetic, but the argument can be applied to any system that

1. is recursively axiomatizable,
2. can prove all valid formulas (i.e. formulas true in all models),
3. has modus ponens,
4. can prove all true formulas with only bounded quantifiers.

A formula with only bounded quantifiers is a formula all quantifiers of which are of the form \( (\forall v \leq c) \) or \( (\exists v \leq c) \), where \( (\forall v \leq c) F \) is an abbreviation for \( \forall v (v \leq c \rightarrow F) \), \( (\exists v \leq c) F \) is an abbreviation for \( \exists v (v \leq c \wedge F) \) and \( c \) is either a variable or a term. Fix for the rest of this section a proof system \( S \) that satisfies conditions 1–4 above. Saying that a formula \( F \) is provable, which we write as \( \vdash F \), we mean provable in \( S \).

Let us start with some definitions. We call the formulas with only bounded quantifiers \( \Sigma_0 \) formulas. A set is a \( \Sigma_0 \) set if it can be expressed by a \( \Sigma_0 \) formula. Formulas of the form \( \exists v F \), where \( F \) is a \( \Sigma_0 \) formula are called \( \Sigma_1 \) formulas, and the sets expressed by \( \Sigma_1 \) formulas are called \( \Sigma_1 \) sets. The \( \Sigma_1 \) sets are precisely the r.e. sets. We say that a formula \( F(v_1, v_2) \) enumerates a number set \( A \) in \( S \) if for any number \( n \), the following two conditions hold:

1. If \( n \in A \), then there is a number \( m \) such that \( \vdash F(n, m) \).
2. If \( n \notin A \), then for every number \( m \), \( \vdash \neg F(n, m) \).

We say that \( A \) is enumerable in \( S \) if there is a formula that enumerates it in \( S \).

**Lemma 3.1.** All \( \Sigma_1 \) sets are enumerable in \( S \).

**Proof.** Suppose \( A \) is a \( \Sigma_1 \) set. Being \( \Sigma_1 \) means that there is a formula of the form \( \exists v_2 F(v_1, v_2) \), where \( F \) is \( \Sigma_0 \), that expresses it. We claim that \( F(v_1, v_2) \) enumerates \( A \). To see this, notice that:

1. If \( n \in A \), then \( \exists v_2 F(\bar{\pi}, v_2) \) is true, so \( F(\bar{\pi}, \bar{m}) \) is true for some number \( m \). But \( S \) can prove all true \( \Sigma_0 \) formulas; in particular, it can prove \( F(\bar{\pi}, \bar{m}) \).
2. If \( n \notin A \), then \( \exists v_2 F(\bar{\pi}, v_2) \) is false. Hence for any number \( m \), \( F(\bar{\pi}, \bar{m}) \) is false, so \( \neg F(\bar{\pi}, \bar{m}) \) is true, and being a \( \Sigma_0 \) formula, is provable. \( \square \)

We say that a formula \( F(v) \) represents a set \( A \) in \( S \) if for any number \( n \),

\[ n \in A \iff \vdash F(\bar{\pi}). \]

\( A \) is representable in \( S \) if there is a formula that represents it in \( S \). Lemma 3.1 gives:
Lemma 3.2. If $S$ is $\omega$-consistent, then all $\Sigma_1$ sets are representable in $S$.

Proof. Suppose that $A$ is a $\Sigma_1$ set. From Lemma 3.1, there is a formula, say $F(v_1, v_2)$, that enumerates $A$. We claim that $\exists v_2 F(v_1, v_2)$ represents $A$ in $S$. To see this, notice that:

1. If $n \in A$, then, because $F$ enumerates $A$, there is a number $m$ such that $\vdash F(\overline{n}, \overline{m})$. Now since $F(\overline{n}, \overline{m}) \rightarrow \exists v_2 F(\overline{n}, v_2)$ is a valid formula, it is the case that $\vdash \exists v_2 F(\overline{n}, v_2)$.

2. If $n \notin A$, then, because $F$ enumerates $A$, we have that $\vdash \neg F(\overline{n}, \overline{m})$ for any $m$. And since $S$ is $\omega$-consistent, it cannot be the case that $\vdash \exists v_2 F(\overline{n}, v_2)$. \hfill $\square$

Gödel’s theorem is:

Theorem 3.1. If $S$ is $\omega$-consistent, then there is a formula $G$ such that neither $G$ nor $\neg G$ is provable in $S$.

Proof. The set $P$ of all provable in $S$ formulas is a $\Sigma_1$ set. Now it holds that if a set $A$ is $\Sigma_1$, then also $A^*$ is $\Sigma_1$: This should be clear from Propositions 2.1 and 2.2 in the case where our language contains exponentiation; if it doesn’t, then it can be seen (but it is not simple) that we can express $z = d(x, y)$ by $+$ and $\cdot$ with a $\Sigma_1$ formula. So $P^*$ is also $\Sigma_1$, and from Lemmas 3.1 and 3.2, there is a formula of the form $\exists v_2 F(v_1, v_2)$, where $F$ is $\Sigma_0$ that represents $P^*$ in $S$. It is the case that $\vdash \exists v_2 F(v_1, v_2)$ if and only if $\vdash \neg \forall v_2 \neg F(v_1, v_2)$, so $\forall v_2 \neg F(v_1, v_2)$ is a formula whose negation represents $P^*$ in $S$. Call that formula $G(v_1)$.

For any number $n$,

$$n \in P^* \iff \vdash \neg G(\overline{n}).$$

In particular, for the number $g$ that encodes $G(v_1)$,

$$g \in P^* \iff \vdash \neg G(\overline{g}).$$

But it is also the case that

$$\vdash G(\overline{g}) \iff \vdash G[\overline{g}] \iff g \in P^*.$$

The first line follows because $G(\overline{g}) \equiv G[\overline{g}]$ is a valid formula, and the second line is from the definition of $P^*$. Therefore we have that

$$\vdash G(\overline{g}) \iff \vdash \neg G(\overline{g}),$$

meaning that either both $G(\overline{g})$ and $\neg G(\overline{g})$ are provable, or they are both non-provable. The former case violates the $\omega$-consistency condition (for since $S$ is closed under modus ponens and proves all valid formulas, proving a contradictory formula means that all formulas are provable), therefore it must be the case that neither $G(\overline{g})$ nor $\neg G(\overline{g})$ is provable. \hfill $\square$

Let us end this section with some remarks on Theorem 3.1. A proof system is called consistent if it doesn’t prove a formula and its negation. At the end of Theorem 3.1, we noticed that $\omega$-consistency implies consistency.

Proposition 3.1. If $S$ is consistent, then all provable in $S$ $\Sigma_0$ formulas are true.

Proof. This is a consequence of the fact that $S$ proves all true $\Sigma_0$ formulas. For if $S$ proved a false $\Sigma_0$ formula $F$, then since $\neg F$ is true and $\Sigma_0$, $S$ would prove $\neg F$, violating the consistency assumption. \hfill $\square$

Proposition 3.2. Let $G \equiv \forall v_2 \neg F(\overline{g}, v_2)$ be the formula of Theorem 3.1. If $S$ is consistent, then $G$ is not provable, but true.
Proof. Theorem 3.1 says that if $S$ is $\omega$-consistent, then $G$ is not provable. But in fact, only simple consistency is needed for showing that. Suppose that $\forall v_2 \neg F(g, v_2)$ is provable. Then $g \in P^*$. And since $F(v_1, v_2)$ enumerates $P^*$ in $S$, there must be a number $m$ such that $\vdash F(\overline{g}, \overline{m})$. Hence $\vdash \exists v_2 F(\overline{g}, v_2)$, so $\vdash \neg \forall \neg v_2 F(\overline{g}, v_2)$, in other words $\vdash \neg G$. Therefore if $G$ is provable then so its negation, violating the consistency assumption.

To see that $G$ is true, we know that $G$ is not provable, hence $g \notin P^*$. Since $F(v_1, v_2)$ enumerates $P^*$ and $g \notin P^*$, for any number $n$, $\vdash \neg F(g, n)$. But $\neg F(\overline{g}, \overline{m})$ is $\Sigma_0$, so by Proposition 3.1, for any $n$, $\neg F(\overline{g}, \overline{m})$ is true. Therefore $G$ is true. Proposition 3.2 reveals an astonishing fact. All formulas $\neg F(g, 0)$, $\neg F(g, 1)$, $\ldots$ are true and $\Sigma_0$, hence $S$ can prove them. Yet, if $S$ is consistent, it cannot prove the universal formula $\forall v \neg F(g, v)$!}

4 Rosser’s proof

We saw that simple consistency was enough to show that there is a true formula, not provable in $S$. However, to prove Theorem 3.1, we needed the stronger assumption of $\omega$-consistency. In this section we shall see how to replace $\omega$-consistency with consistency in Theorem 3.1. To do that we will construct a different undecidable formula. The construction is due to Rosser.

Weakening the assumption of $\omega$-consistency, comes with a price, actually making Gödel’s and Rosser’s theorems incomparable in strength. Namely, the conditions 1–4 of the previous section that our system has to satisfy must be strengthened by changing 2 and 3 to

2’. the system must be able to prove all valid formulas of first-order logic with equality (i.e. formulas true in all models where the equality symbol is interpreted as equality),

3’. the system contains besides modus ponens, the generalization rule, viz. from $F$ infer $\forall v F$,

and adding the new condition:

5. the system must prove the formulas

i. $\forall v \left( v \leq \overline{m} \equiv (v = \overline{0} \lor \cdots \lor v = \overline{m}) \right)$,

ii. $\forall v \left( v \leq \overline{m} \lor \overline{m} \leq v \right)$.

So fix for the rest of this section a system $S$ that satisfies conditions 1–4 of the previous section, changing 2 and 3 to 2’ and 3’, and adding condition 5. Again, the provability predicate $\vdash F$ is with respect to $S$.

Let $R$ be the set of all formulas $F$ such that $\vdash \neg F$. The set $R$ (and more specifically $R^*$) will play a crucial role in what follows. To give a rough outline, Gödel’s proof amounts to constructing a formula whose negation represents $P^*$ in $S$ (or alternatively a formula which represents $R^*$, see [2]). Rosser’s proof represents a superset of $R^*$ disjoint from $P^*$; for doing this, only the assumption of consistency is needed.

We say that a formula $F(v)$ separates a set $A$ from a set $B$ in $S$ if for all $n$,

$n \in A \implies \vdash F(\overline{n})$,

$n \in B \implies \vdash \neg F(\overline{n})$.

The following lemma plays the same role that Lemma 3.2 played in the previous section.

**Lemma 4.1.** For any two disjoint $\Sigma_1$ sets $A$ and $B$, there is a formula that separates $B$ from $A$ in $S$.
Proof. From Lemma 3.1, A and B are enumerable in S. Let A(x, y) and B(x, y) be two formulas that enumerate them. We show that the formula
\[ \forall y \ (A(x, y) \rightarrow (\exists z \leq y) B(x, y)) \]  
separates B from A in S.

1. Suppose \( n \in B \). Then \( \vdash B(\pi, \bar{k}) \) for some \( k \). Also \( n \notin A \), because A and B are disjoint, so for every \( m \leq k \) (in fact for every \( m \)), \( \vdash \neg A(\pi, \bar{m}) \). From the latter and using condition 5i, we get
\[ \vdash (\forall y \leq \bar{k}) \neg A(\pi, y). \]
This in turn gives \( \vdash y \leq \bar{k} \rightarrow \neg A(\pi, y) \) and so \( \vdash A(\pi, y) \rightarrow \neg (y \leq \bar{k}) \). Using 5ii, it follows that \( \vdash A(\pi, y) \rightarrow \bar{k} \leq y \), and since \( \vdash B(\pi, \bar{k}) \), we get that
\[ \vdash A(\pi, y) \rightarrow \bar{k} \leq y \land B(\pi, \bar{k}). \]
This gives
\[ \vdash A(\pi, y) \rightarrow (\exists z \leq y) B(\pi, z), \]
from which the provability of (1) follows by the generalization rule.

2. Suppose \( n \in A \). Then \( \vdash A(\pi, \bar{k}) \) for some \( k \) and for all \( m \leq k \), \( \vdash \neg B(\pi, \bar{m}) \). The latter gives \( \vdash (\forall z \leq \bar{k}) \neg B(\pi, z) \), so
\[ \vdash A(\pi, \bar{k}) \land (\forall z \leq \bar{k}) \neg B(\pi, z), \]
\[ \vdash \neg (A(\pi, \bar{k}) \rightarrow (\exists z \leq \bar{k}) \neg B(\pi, z)), \]
\[ \vdash \neg (A(\pi, \bar{k}) \rightarrow (\exists z \leq \bar{k}) \neg B(\pi, z)), \]
\[ \vdash \neg \forall y \ (A(\pi, \bar{k}) \rightarrow (\exists z \leq \bar{k}) B(\pi, z)). \]
Rosser’s theorem is:

**Theorem 4.1.** If \( S \) is consistent, then there is a formula \( H \) such that neither \( H \) nor \( \neg H \) is provable in \( S \).

Proof. \( S \) is recursively axiomatizable, so both \( P \) and \( R \) are \( \Sigma_1 \) sets. As in Theorem 3.1, this implies that \( P^* \) and \( R^* \) are also \( \Sigma_1 \). Now the consistency of \( S \) says that \( P \) and \( R \) are disjoint, therefore \( P^* \) and \( R^* \) are also disjoint, and Lemma 3.2 gives a formula \( H(\nu) \) that separates \( R^* \) from \( P^* \). Let \( R' \), the set \( H(\nu) \), represents in \( S \). \( R' \) is the set for which for all \( n \)
\[ n \in R' \iff \vdash H(\nu) \]
Notice that \( R' \) is disjoint from \( R^* \) and \( R^* \subseteq R' \). This is so, because if \( n \in R^* \), then \( \vdash H(\nu), \) so \( n \in R' \). Also if \( n \in R' \cap P^* \), then \( \vdash F(\nu) \) and \( \vdash \neg F(\nu) \) which contradicts the consistency of \( S \). Now let \( h \) be the number which encodes \( H(\nu) \). We have that
\[ h \in R' \iff \vdash H(\nu), \]
\[ \iff h \in P^*. \]
Since \( R' \) is disjoint from \( P^* \), \( h \notin R' \) for \( h \notin P^* \). Since \( h \notin P^* \), \( H(\nu) \) is not provable. Since \( h \notin R' \) and \( R^* \subseteq R' \), \( h \notin R^* \) and so \( \neg H(\nu) \) is not provable. \( \Box \)
5 Epilogue and final remarks

The first incompleteness result presented followed showing that the set of all true formulas can not be expressed by a $\Sigma_1$ formula (in fact by any formula), and assuming that our system can prove only true formulas. Gödel’s proof boils down to showing that the set $P^*$ can be represented in our system, and for this the $\omega$-consistency assumption was needed. Rosser’s proof replaces the $\omega$-consistency assumption with consistency by representing a superset of $R^*$ disjoint from $P^*$. Actually it is possible to directly represent $P^*$ assuming simple consistency in suitable systems. This is due to Ehrenfeucht and Feferman, who showed that all $\Sigma_1$ sets can be represented in every recursively axiomatizable and consistent system that extends Robinson’s arithmetic (see Chapter 7 of [2]).

References
