Definability and its limits

Professor Cian Dorr

15th November 2022

New York University

Gödel's incompleteness theorem, again

Gödel's first incompleteness theorem (zeroth pass)

No theory is

1. consistent
2. negation-complete
3. "sufficiently simple"
4. "sufficiently strong"

Gödel's incompleteness theorem, again

Gödel's first incompleteness theorem (first pass)

No theory in **a signature extending the standard string signature Str** is

1. consistent
2. negation-complete
3. finitely axiomatizable
4. includes Min

Gödel's incompleteness theorem, again

Gödel's first incompleteness theorem (second pass)

No theory in any signature is

1. consistent
2. negation-complete
3. finitely axiomatizable
4. **interprets** Min

T_1 interprets T_2 iff T_1 has a **finite definitional extension** that extends T_2 or a notational variant of T_2.

Gödel's incompleteness theorem, again

To get from the first-pass version to the second version, note first that the first-pass version implies that no theory is consistent, negation-complete, finitely axiomatizable, and includes a *notational variant* of Min: if we had such a theory, we could just substitute different non-logical constants to turn it into the kind of theory whose existence is ruled out by the first-pass version.

Now suppose for contradiction that T in signature Σ is consistent, negation-complete, finitely axiomatizable, and has a finite definitional extension T^+ in a signature Σ^+ that extends a notational variant Min' of Min in signature Str'. T^+ is also finitely axiomatizable, since it is axiomatized by any axiomatization of T together with the finitely many definitions that we add to get to T^+. It is also consistent (by the fact that definitional extensions are conservative extensions), and it's negation-complete (since for every Σ^+-sentence P there is a Σ-sentence P' such that $T^+ \vDash P \leftrightarrow P'$). This is impossible by the first-pass version of the theorem.

Definitions of predicates

Definition

Where F is an n-ary predicate that is in Σ^+ but not in Σ, a **definition of F in Σ** is a sentence of the form $\forall v_1 \ldots \forall v_n (F(v_1, \ldots, v_n) \leftrightarrow P)$, where P is a formula of Σ and $v_1, \ldots, v_n$ are distinct variables.

Definition

Where f is an n-ary function symbol that occurs in Σ^+ but not in Σ, a **definition of f in Σ** is a sentence of the form $\forall v_1 \ldots \forall v_{n+1} (v_{n+1} = f(v_1, \ldots, v_n) \leftrightarrow P)$, where P is a formula of Σ and $v_1, \ldots, v_{n+1}$ are distinct variables.

Definition

Where S is a structure for Σ, a definition of a predicate or function symbol not in Σ is **legitimate in** S iff either it's a definition of a predicate, or it's a definition $\forall v_1 \ldots \forall v_{n+1}(v_{n+1} = f(v_1, \ldots, v_n) \leftrightarrow P))$ of a function symbol such that $S \Vdash \forall v_1 \ldots v_n \exists! v_{n+1} P$.

Definability in a structure

Definable sets and relations

Definition

Let S be a structure for Σ. An n-ary relation R of S's domain is **definable in** S iff there is a definitional expansion of S that includes a n-ary predicate F with R as its interpretation.

This is equivalent to the claim that there is a Σ-formula P with n free variables $v_1, \ldots, v_n$ such that for all $x_1, \ldots, x_n$ in the domain of S,
$S, [v_1 \mapsto x_1, \ldots, v_n \mapsto x_n] \Vdash P$ iff $Rx_1 \ldots x_n$.

Definable sets and relations

Definition

Let S be a structure for Σ. An n-ary relation R of S's domain is **definable in** S iff there is a definitional expansion of S that includes a n-ary predicate F with R as its interpretation.

This is equivalent to the claim that there is a Σ-formula P with n free variables $v_1, \ldots, v_n$ such that for all $x_1, \ldots, x_n$ in the domain of S,
$S, [v_1 \mapsto x_1, \ldots, v_n \mapsto x_n] \Vdash P$ iff $Rx_1 \ldots x_n$.

- I'll sometimes abbreviate "formula with n free variables" as "n-formula".
- Say that n-formula P is **true of** $x_1, \ldots, x_n$ in S iff $S, [v_1 \mapsto x_1, \ldots, v_n \mapsto x_n] \Vdash P$, where $v_1, \ldots, v_n$ are the free variables of P *in alphabetical order*.

Definable functions

For an an n-ary function f on S's domain, we could also say that f is definable in S iff there is a definitional expansion of S that includes an n-ary function symbol g with f as its interpretation.

But since n-ary functions are $n+1$-ary relations, this actually comes to the same thing: for an n-ary function on S's domain, it's the interpretation of an $n+1$-ary predicate in a definitional expansion of S iff it's the interpretation of an n-ary function symbol in a definitional expansion of S.

Definable partial functions

Note that if an n-ary partial function f is definable in a structure S, there is a definitional expansion of S that includes an n-arey function symbol g with some total extension of f as its interpretation.

The existence of undefinable sets

Since there are only countably many formulae in the language, the set of definable subsets of a structure's domain is always countable.

If the domain is infinite, then by Cantor's theorem it follows that not all subsets of the domain are definable.

A set of strings undefinable in $\mathbb{S}$

For the standard string structure (or any other structure whose domain includes the formulae of its signature), we can give an example.

Recall that in proving Cantor's theorem, we proved that an arbitrary function $f : A \to \mathcal{P}A$ is not surjective by showing that $\{x \mid x \notin fx\}$ isn't in its range. For the assumptiont that $fy = \{x \mid x \notin fx\}$ implies the contradictory

$$y \in fy \text{ iff } y \notin fy$$

A set of strings undefinable in $\mathbb{S}$

For the standard string structure (or any other structure whose domain includes the formulae of its signature), we can give an example.

Recall that in proving Cantor's theorem, we proved that an arbitrary function $f : A \to \mathcal{P}A$ is not surjective by showing that $\{x \mid x \notin fx\}$ isn't in its range. For the assumptiont that $fy = \{x \mid x \notin fx\}$ implies the contradictory

$$y \in fy \text{ iff } y \notin fy$$

The same reasoning establishes

Undefinability Fact 1

The set of all 1-formulae that are not true of themselves in $\mathbb{S}$ is not definable in $\mathbb{S}$.

For the assumption that P is true of a 1-formula Q iff Q is not true of Q implies the contradictory

$$P(v) \text{ is true of } P(v) \text{ iff } P(v) \text{ is not true of } P(v)$$

Another undefinable set

Now that we have one undefinable set, we can give more examples.

Undefinability Fact 2

The set of all 1-formulae that *are* true of themselves in $\mathbb{S}$ is not definable in $\mathbb{S}$.

Proof: suppose there were a 1-formula P true in $\mathbb{S}$ of all and only such 1-formulae. Then $\neg P$ would be a 1-formula true in $\mathbb{S}$ of exactly the 1-formulae that are *not* true of themselves in $\mathbb{S}$, which can't happen by Fact 1.

- Here we are using the fact that $\neg P$ is true of d, i.e. true on $v \to d$ iff P isn't true of d, i.e. isn't true on $v \to d$.

An undefinable relation

Undefinability Fact 3

The set of all pairs of 1-formulae P, Q such that P is true of Q is not definable in $\mathbb{S}$.

Proof: suppose there were a 2-formula $A(x, y)$ true in $\mathbb{S}$ of $\langle P, Q \rangle$ iff P is true of in $\mathbb{S}$. Then $A(x, x)$—i.e., $A(x, y)[y \mapsto x]$ would be a 1-formula true of P iff P is true of P, which can't happen by Fact 2.

Tarski's Undefinability Theorem

Here are two facts we will take on trust for now.

Promissory Note 1

The *standard label* function $\langle \cdot \rangle$ is definable in $\mathbb{S}$.

Promissory Note 2

For any variable v, *substitution* function that takes a formula P and a term t and returns $P[t/v]$ is definable in $\mathbb{S}$.

Tarski's Undefinability Theorem

Here are two facts we will take on trust for now.

Promissory Note 1

The *standard label* function $\langle \cdot \rangle$ is definable in $\mathbb{S}$.

Promissory Note 2

For any variable v, *substitution* function that takes a formula P and a term t and returns $P[t/v]$ is definable in $\mathbb{S}$.

Given these two facts, we can establish

Tarski's Undefinability Theorem

$\mathrm{Th}\,\mathbb{S}$ is not definable in $\mathbb{S}$.

Tarski's Undefinability Theorem

Proof: Suppose for contradiction that there's a 1-formula True(x) true in $\mathbb{S}$ of exactly the sentences true in $\mathbb{S}$. Consider a definitional expansion of $\mathbb{S}$ with a singulary function symbol label interpreted by the labelling function, and a binary function symbol subst interpreted by the substitution function. Now consider the 2-formula

$$\mathsf{True}(\mathsf{subst}(x, \mathsf{label}(y)))$$

It would be true of $\langle P, Q\rangle$ iff True is true of $P[\langle Q\rangle/x]$, iff $P[\langle Q\rangle/x]$ is true in $\mathbb{S}$. Since $[\![\langle Q\rangle]\!]_{\mathbb{S}} = Q$, this is the case iff P is true of Q in $\mathbb{S}$. This is ruled out by Undefinability Fact 3.

Tarski's Undefinability Theorem

Proof: Suppose for contradiction that there's a 1-formula True(x) true in $\mathbb{S}$ of exactly the sentences true in $\mathbb{S}$. Consider a definitional expansion of $\mathbb{S}$ with a singulary function symbol label interpreted by the labelling function, and a binary function symbol subst interpreted by the substitution function. Now consider the 2-formula

$$\mathsf{True}(\mathsf{subst}(x, \mathsf{label}(y)))$$

It would be true of $\langle P, Q\rangle$ iff True is true of $P[\langle Q\rangle/x]$, iff $P[\langle Q\rangle/x]$ is true in $\mathbb{S}$. Since $[\![\langle Q\rangle]\!]_{\mathbb{S}} = Q$, this is the case iff P is true of Q in $\mathbb{S}$. This is ruled out by Undefinability Fact 3.

Or for a more direct proof, we can move straight to the 1-formula

$$\neg\, \mathsf{True}(\mathsf{subst}(x, \mathsf{label}(x)))$$

and observe that it would be true of P iff True is not true of $P[\langle P\rangle/x]$, iff P is not true of itself—ruled out by Undefinability Fact 1.

Definability of the standard labelling function

Cashing out Promissory Note 1

Recall: $\langle$`dog`$\rangle$ = `⊕("d",⊕("o",⊕("g","")))`.

(i) Any finite set or relation is definable in the standard string structure (since it's explicit). So, in particular, the partial function that takes each one-character string to the corresponding constant (a three-character string) that denotes it in $\mathbb{S}$ is definable in $\mathbb{S}$. Let's definitionally extend $\mathbb{S}$ with a function symbol constantOf(x) whose extension is some total extension of this function.

(ii) The *equally long* as relation is definable in $\mathbb{S}$, with definition

$$\forall x \forall y (\text{EquallyLong}(x, y) \leftrightarrow x \leq y \wedge y \leq x)$$

(iii) The *twice as as long as* relation is definable, with definition

$$\forall x \forall y (\text{TwiceAsLong}(x, y) \leftrightarrow \exists z_1 \exists z_2 (x = z_1 \oplus z_2 \wedge \text{EquallyLong}(y, z_1) \wedge \text{EquallyLong}(y, z_2))$$

Similarly we could write down a definition for SixTimeAsLong.

(iv) The set of all strings that consist entirely of right parentheses is definable, with definition

$$\forall x(\text{AllRightParens}(x) \leftrightarrow \forall y_1 \forall y_2 \forall y_3((x = y_1 \oplus y_2 \oplus y_3 \wedge \mathit{EquallyLong}(y_2, "a")) \rightarrow y_2 = \text{rpa}))$$

(v) So, the labelling function can be defined as follows:

$$
\begin{aligned}
\forall x \forall y (y = \mathsf{label}(x) \leftrightarrow \exists y_1 \exists y_2 (y = y_1 \oplus \mathit{quo} \oplus \mathit{quo} \oplus y_2 \\
&\wedge \mathit{EquallyLong}(y_2, x) \\
&\wedge \mathit{AllRightParens}(y_2) \\
&\wedge \forall x_1 \forall x_2 \forall x_3 (x = x_1 \oplus x_2 \oplus x_3 \wedge \mathit{LengthOne}(x_2) \rightarrow \\
&\quad \exists z_1 \exists z_2 \exists z_3 (y_1 = z_1 \oplus z_2 \oplus z_3 \\
&\quad \wedge \mathit{SixTimesAsLong}(z_1, x_1) \\
&\quad \wedge \mathit{SixTimesAsLong}(z_3, x_3) \\
&\quad \wedge z_2 = \text{"} \oplus \text{"} \oplus \mathit{lpa} \oplus \mathit{constantOf}(x_2) \oplus \mathit{com})))
\end{aligned}
$$