Gödel's 2nd Incompleteness Theorem

Robert Rynasiewicz
Mathematical Logic II

Spring 2022

The Consistency Sentence

For a numerical theory T, let

$$Con_T =_{df} \exists x(Wff(x) \wedge \neg Prov_T(x)).$$

Then Con_T expresses (indirectly) the consistency of T. For if T were inconsistent, then every wff would be provable from T. However, since Gödel it has become common, especially if $T = PA$, to define Con_T slightly differently, e.g.

$$Con_T =_{df} \neg Prov_T(\ulcorner 0 = 1 \urcorner),$$

or even

$$Con_T =_{df} \neg Prov_T(\ulcorner \perp \urcorner).$$

Which of these we choose doesn't really matter.

A Prerequisite for the 2nd Theorem

What does matter is that not only can we express the first part of the first incompleteness theorem (i.e., that if PA is consistent, then G is not provable in PA) in the object language, but we can also prove it in PA. Now, we have Con_{PA} to express that PA is consistent and $\neg Prov(\ulcorner G \urcorner)$ to express that $PA \nvdash G$. So what we want is:

$$(*) \quad PA \vdash Con_{PA} \rightarrow \neg Prov_{PA}(\ulcorner G \urcorner).$$

Gödel didn't argue this, but took it as obvious.[1] If we do so as well, we have a very simple proof of the 2nd incompleteness theorem, viz.,

[1] Gödel was working with 2nd-order PA in the context of Russell and Whitehead's *Principia Mathematica*, which, to avoid Russell's paradox, adopts an infinite hierarchy of types.

Proof of the 2nd Theorem Assuming (*)

Gödel's 2nd Incompleteness Theorem. $T \nvdash Con_T$ for any consistent, axiomatizable theory T at least as strong as PA.

Pf. Suppose that $T \vdash Con_T$. Then, by (*), which holds for any axiomatizable T at least as strong as PA if it holds for PA, we have that $T \vdash \neg Prov_T(\ulcorner G_T \urcorner)$. But we were also able to show that for any such theory

$$T \vdash G_T \leftrightarrow \neg Prov_T(\ulcorner G_T \urcorner).$$

Hence, $T \vdash G_T$. But we had established that if T is consistent, then $T \nvdash G_T$. ■

But can we really take (*) for granted?

Can PA Prove Its Own **In**consistency?

Let's set the status of (*) aside for a moment. Without it we can easily show:

Theorem. If T is any axiomatic extension of Q and T is ω-consistent, then $T \nvdash \neg Con_T$.

Pf. Suppose T is ω-consistent. For each $n \in \mathbb{N}$,

$$T \vdash \neg Proof_T(\overline{n}, \ulcorner \bot \urcorner).$$

Hence, by ω-consistency,

$$T \nvdash \exists x \; Proof_T(x, \ulcorner \bot \urcorner),$$

i.e.,

$$T \nvdash \neg Con_T. \qquad \blacksquare$$

With (*), Consistency Entails ω-Incompleteness

From here out, suppose that we restrict our discussion to axiomatic extensions of PA. And suppose indeed that $T \nvdash Con_T$. Then, as we saw on the last slide,

$$T \vdash \neg Proof_T(\overline{n}, \ulcorner \bot \urcorner),$$

for each $n \in \mathbb{N}$. But since $T \nvdash Con_T$,

$$T \nvdash \forall x \neg Proof_T(x, \ulcorner \bot \urcorner).$$

In other words, T is ω-incomplete.

Corollary of the 2nd Incompleteness Theorem. If T is a consistent axiomatic extension of PA, then T is ω-incomplete.

But What About (*)?

Recall

$$(*) \quad PA \vdash Con_{PA} \to \neg Prov_{PA}(\ulcorner G \urcorner).$$

There are three "derivability" conditions, called the *Hilbert-Bernays-Löb* conditions, that are jointly sufficient to prove (*). Rather than focusing just on PA, consider arbitrary numerical theories T and write $\Box\varphi$ for $Prov_T(\ulcorner \varphi \urcorner)$.

C1 If $T \vdash \varphi$, then $T \vdash \Box\varphi$.

C2 $T \vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

C3 $T \vdash \Box\varphi \to \Box\Box\varphi$

The first of these we have already established for PA in "Further Results on Incompleteness," i.e.,

Theorem. If $T \vdash \varphi$, then $T \vdash Prov_T(\ulcorner \varphi \urcorner)$ (assuming T is an axiomatic extension of Q).

Proof of C2

Theorem. If T is an axiomatic extension of Q, then

$$T \vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi).$$

Pf. Note that if k is the super g.n. of a proof of $(\varphi \to \psi)$ and ℓ the super g.n. of a proof of φ, then $m = k \star \ell \star 2^{\#\psi+1}$ is the super g.n. of a proof of ψ. Arguing now inside of T, assume $\exists x\, Proof(x, \ulcorner(\varphi \to \psi)\urcorner)$ and $\exists x\, Proof(x, \ulcorner\varphi\urcorner)$. By existential instantiation with a and b respectively, we have $Proof(a, \ulcorner(\varphi \to \psi)\urcorner)$ and $Proof(b, \ulcorner\varphi\urcorner)$. As an exercise, one can dig through the definitions of these to arrive at

$$Proof(a * b * 2^{\ulcorner\psi\urcorner+1}, \ulcorner\psi\urcorner),$$

where $*$ represents the concatenation function in T. Then existentially generalize on $a * b * 2^{\ulcorner\psi\urcorner+1}$ to get $\exists x\, Proof(x, \ulcorner\psi\urcorner)$. ■

Proof of C3

Theorem. If T is an axiomatic extension of PA, then

$$T \vdash \Box\varphi \to \Box\Box\varphi.$$

Proof Sketch. $\Box\varphi$ is a Σ_1 wff, viz., $\exists x\, Proof(x, \ulcorner\varphi\urcorner)$, where $Proof(x, y)$ is Δ_0. So, it suffices to show that if ψ is Σ_1 then $T \vdash \psi \to \Box\psi$. Now, Q, and hence PA, is Σ_1 complete. The plan of attack is to formalize the proof of that inside of PA. We say PA and not Q, since since that proof required a course of value induction on the complexity of wffs, so the induction schema is necessary to pull off the formalization. ∎

Proof of (*) Using the Derivability Conditions

Theorem. If T is an axiomatic extension of PA, then

$$T \vdash Con_T \to \neg\Box G_T.$$

Pf. For the sake of ease, we'll take Con_T to be $\neg\Box\bot$. The following will hold for any wff φ in virtue of the definitions.

$$\vdash \neg\varphi \to (\varphi \to \bot).$$

By deductive monotonicity,

$$T \vdash \neg\varphi \to (\varphi \to \bot).$$

From C1 we get

$$T \vdash \Box(\neg\varphi \to (\varphi \to \bot)),$$

and from C2 then

$$(**) \quad T \vdash \Box\neg\varphi \to \Box(\varphi \to \bot).$$

Proof of (*) Using the Derivability Conditions (cont.)

Dropping the subscript T on Con_T and G_T:

1.	$T \vdash G \to \neg\Box G$	from 1st Incompl. Th.
2.	$T \vdash \Box(G \to \neg\Box G)$	from 1 and C1
3.	$T \vdash \Box G \to \Box\neg\Box G$	from 2 and C2
4.	$T \vdash \Box\neg\Box G \to \Box(\Box G \to \bot)$	Instance of (**)
5.	$T \vdash \Box G \to \Box(\Box G \to \bot)$	3,4
6.	$T \vdash \Box G \to (\Box\Box G \to \Box\bot)$	5, C2
7.	$T \vdash \Box G \to \Box\Box G$	C3
8.	$T \vdash \Box G \to \Box\bot$	6,7, Hilbert Axiom Type II
9.	$T \vdash \neg\Box\bot \to \neg\Box G$	8 contraposition
10.	$T \vdash Con \to \neg\Box G$	defn of *Con*

Equivalence with Gödel's Definition of Con

Recall that Gödel's original definition of Con was

$$\exists x(Wff(x) \wedge \neg Prov(x)).$$

It's easy to show now:

Theorem. If T is consistent, then there is no wff φ s.t. $T \vdash \neg\Box\varphi$.

Pf. We will show that $T \vdash \Box\bot \to \Box\varphi$, i.e., $T \vdash \neg\Box\varphi \to \neg\Box\bot$, so that if $T \vdash \neg\Box\varphi$, then $T \vdash \neg\Box\bot$, which we've shown is impossible.

- $\bot \to \varphi$ is a tautology,
- and so trivially $T \vdash \bot \to \varphi$.
- By C1, $T \vdash \Box(\bot \to \varphi)$,
- and by C2, $T \vdash \Box\bot \to \Box\varphi$. ■

Formalizing the 2nd Theorem

Theorem. If T is an axiomatic extension of PA, then

$$T \vdash Con_T \to \neg\Box Con_T.$$

Proof. Dropping the subscript T from Con_T and G_T:

1.	$T \vdash Con \to \neg\Box G$	(*)
2.	$T \vdash \neg\Box G \to G$	from 1st incompleteness theorem
3.	$T \vdash Con \to G$	1,2
4.	$T \vdash \Box(Con \to G)$	C1
5.	$T \vdash \Box Con \to \Box G$	C2
6.	$T \vdash \neg\Box G \to \neg\Box Con$	5, contraposition
7.	$T \vdash Con \to \neg\Box Con$	1,6 ■

Another Fixed Point

Earlier we had shown that for any wff φ that $T \vdash \neg\Box\varphi \rightarrow \neg\Box\bot$. This holds in particular for Con_T. Thus, since we've been taking $\neg\Box\bot$ to be Con_T,

$$T \vdash \neg\Box Con_T \rightarrow Con_T.$$

Putting this together with the Formalized 2nd Theorem yields

$$T \vdash Con_T \leftrightarrow \neg\Box Con_T,$$

i.e.

$$T \vdash Con_T \leftrightarrow \neg Prov_T(\ulcorner Con_T \urcorner).$$

In other words:

Cor. Con_T is a fixed point of $\neg Prov_T(v_0)$.

All Fixed Points of $\neg Prov_T(v_0)$ Are Equivalent

Theorem. If T is an axiomatic extension of PA, then any fixed point of $\neg Prov_T(v_0)$ is equivalent in T to Con_T.

Proof. Next slide.

1.	$T \vdash \gamma \leftrightarrow \neg\Box\gamma$	supposition
2.	$T \vdash \gamma \rightarrow (\Box\gamma \rightarrow \bot)$	1
3.	$T \vdash \Box(\gamma \rightarrow (\Box\gamma \rightarrow \bot))$	C1
4.	$T \vdash \Box\gamma \rightarrow \Box(\Box\gamma \rightarrow \bot)$	C2
5.	$T \vdash \Box\gamma \rightarrow (\Box\Box\gamma \rightarrow \Box\bot)$	C2
6.	$T \vdash \Box\gamma \rightarrow \Box\Box\gamma$	C3
7.	$T \vdash \Box\gamma \rightarrow \Box\bot$	5,6, Hilbert Axiom Type II
8.	$T \vdash \bot \rightarrow \gamma$	Meaning of $\bot$
9.	$T \vdash \Box(\bot \rightarrow \gamma)$	C1
10.	$T \vdash \Box\bot \rightarrow \Box\gamma$	C2
11.	$T \vdash \Box\bot \leftrightarrow \Box\gamma$	7,10
12.	$T \vdash \gamma \leftrightarrow \neg\Box\bot$	1,11 ■

Fixed Points of $Prov_T(v_0)$?

By the diagonalization lemma, there is a sentence H s.t.

$$T \vdash H \leftrightarrow Prov_T(\ulcorner H \urcorner),$$

and so H appears to say 'I am provable'. But is it?

Löb's Theorem. If $T \vdash \Box\varphi \rightarrow \varphi$, then $T \vdash \varphi$.

So the answer is 'yes', since by hypothesis $T \vdash H \leftrightarrow \Box H$.

Proof of Löb's Theorem. Pick φ and suppose that $T \vdash \Box\varphi \rightarrow \varphi$. Take the predicate for the diagonalization lemma to be $Prov_T(v_0) \rightarrow \varphi$. Then there exists a sentence γ s.t.

$$T \vdash \gamma \leftrightarrow (Prov_T(\ulcorner \gamma \urcorner) \rightarrow \varphi).$$

Löb's Theorem (cont.)

1.	$T \vdash \gamma \leftrightarrow (\Box\gamma \rightarrow \varphi)$	Diagonalization Lemma
2.	$T \vdash \gamma \rightarrow (\Box\gamma \rightarrow \varphi)$	1
3.	$T \vdash \Box(\gamma \rightarrow (\Box\gamma \rightarrow \varphi))$	C1
4.	$T \vdash \Box\gamma \rightarrow \Box(\Box\gamma \rightarrow \varphi)$	C2
5.	$T \vdash \Box\gamma \rightarrow (\Box\Box\gamma \rightarrow \Box\varphi)$	C2
6.	$T \vdash \Box\gamma \rightarrow \Box\Box\gamma$	C3
7.	$T \vdash \Box\gamma \rightarrow \Box\varphi$	5,6, Hilbert Type II
8.	$T \vdash \Box\varphi \rightarrow \varphi$	Premise
9.	$T \vdash \Box\gamma \rightarrow \varphi$	7,8
10.	$T \vdash \gamma$	1,9
11.	$T \vdash \Box\gamma$	10, C1
12.	$T \vdash \varphi$	9,11 ■

Curry's Paradox

- Let φ be 'the moon is made of green cheese', and
- let $\Box\varphi$ be interpreted as 'it is true that φ.
- The resulting interpretations of C1-C3 seem sound,
- as well as the premise $\Box\varphi \rightarrow \varphi$.
- If there is a sentence γ that is equivalent to: if γ is true, then the moon is made of green cheese,
- then we can prove the moon is made of green cheese.

Is PA Consistent?

- If T is weaker than PA (and consistent), then $T \nvdash Con_{PA}$.
- $Q \nvdash Con_Q$ (Bezboruah and Shepherdson, 1976)
- If T is stronger than PA (and consistent), then, although it might be that $T \vdash Con_{PA}$, $T \nvdash Con_T$.
- Gentzen (1936) proves PA consistent using a system with an induction schema only Δ_0 but with the assumption of ϵ_0 as a well-founded transfinite ordinal (neither weaker nor stronger than PA). No indication that Gentzen is consistent.
- Existence of self-verifying systems (Willard 2001), but, e.g., cannot show multiplication a total function.
- Is there a fact of the matter about Con_{PA}?