Gödel's First Incompleteness Theorem

Unit 6: Gödel's First Incompleteness Theorem

Lavinia Picollo

1 Self-reference in arithmetic

1.1 Diagonalization

Definition (diagonalization)**.** *If $\varphi(\mathsf{v})$ is a wff of L_A its diagonalization is $\forall \mathsf{v}\,(\mathsf{v} = \overline{\ulcorner\varphi(\mathsf{v})\urcorner} \to \varphi(\mathsf{v}))$.*

The diagonalization of $\varphi(\mathsf{v})$ is logically equivalent to $\varphi(\overline{\ulcorner\varphi(\mathsf{v})\urcorner})$ – i.e. the diagonalization of $\varphi(\mathsf{v})$ says of $\varphi(\mathsf{v})$ that it holds of itself (i.e. that it's a φ).

Exercise 1. *Show that the diagonalization function, diag, that maps the code of a wff $\varphi(\mathsf{v})$ of L_A to the code of its diagonalization, is p.r.*

Theorem 47 (Diagonalization Lemma)**.** *Let $\mathsf{Q} \subseteq \mathsf{T}$. For each wff $\varphi(\mathsf{v})$ of the language of T there is a sentence ψ also in the language s.t.*

$$T \vdash \varphi(\overline{\ulcorner\psi\urcorner}) \leftrightarrow \psi$$

For each wff $\varphi(\mathsf{v})$ there is a sentence ψ that 'says of itself' – that is equivalent in T to a sentence that says of ψ – that it's a φ.

Proof. Let $\mathsf{Diag}(\mathsf{x},\mathsf{y})$ capture in Q the diagonalization function and consider the wff $\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))$. It's diagonalization is

$$\forall \mathsf{v}\,(\mathsf{v} = \overline{\ulcorner\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))\urcorner} \to \forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))) \tag{1}$$

This says that if something is identical to $\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))$ and x is its diagonalization, then x is a φ: i.e. that the diagonalization of $\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))$ – i.e. itself! – is a φ. To see it, note that (1) is logically equivalent to

$$\forall \mathsf{x}\,(\mathsf{Diag}(\overline{\ulcorner\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))\urcorner},\mathsf{x}) \to \varphi(\mathsf{x})) \tag{2}$$

Since $\mathsf{Diag}(\mathsf{x},\mathsf{y})$ captures the function *Diag* in Q, we can also prove that

$$\mathsf{Diag}(\overline{\ulcorner\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))\urcorner}, \overline{\ulcorner(1)\urcorner}) \wedge \forall \mathsf{y}\,(\mathsf{Diag}(\overline{\ulcorner\forall \mathsf{x}\,(\mathsf{Diag}(\mathsf{v},\mathsf{x}) \to \varphi(\mathsf{x}))\urcorner},\mathsf{y}) \to \mathsf{y} = \overline{\ulcorner(1)\urcorner}) \tag{3}$$

(3) entails that (2) is equivalent in Q – and, a fortiori, in T – to

$$\varphi(\overline{\ulcorner(1)\urcorner}) \tag{4}$$

Therefore, we have that (1) is equivalent to (4), i.e. (1) is the sentence ψ we were looking for. □

1.2 Gödel sentences

Definition (primitive recursive axiomatization). *A theory T formulated in L_A p.r. axiomatized iff the property of being the g.n. of a non-logical axiom of T is p.r.*

Let T be a p.r. axiomatized theory formulated in L_A. Let $\mathsf{Prf}_T(\mathsf{x},\mathsf{y})$ be a Σ_1 wff expressing in L_A and capturing in Q the p.r. relation $Prf_T(m,n)$ (cf. Unit 5, Theorem 26 and Exercise 5), that holds when m is the s.g.n. of a proof in T of the sentence with g.n. n, and let

$$\mathsf{Prov}_T(\mathsf{x}) =_{def} \exists \mathsf{y}\, \mathsf{Prf}_T(\mathsf{y},\mathsf{x})$$

$\mathsf{Prov}_T(\mathsf{x})$ is a Σ_1 formula that expresses in L_A the numerical property that holds of a number iff it's the g.n. of a theorem of T.

Exercise 2. *Show that if $T \vdash \varphi$, then $\mathsf{Q} \vdash \mathsf{Prov}_T(\overline{\ulcorner \varphi \urcorner})$.*

The Diagonalization Lemma applied to $\neg\mathsf{Prov}_T(\mathsf{x})$ entails that there is a sentence, G_T, s.t.

$$\mathsf{Q} \vdash \neg\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner}) \leftrightarrow \mathsf{G}_T$$

We call G_T the – a – *Gödel sentence* of T. G_T 'says of itself' that it is not provable in T.

Exercise 3. *Show that G_T is Π_1.*

Note that, e.g. Goldbach's Conjecture, that every even number greater than 2 can be expressed as the sum of two primes, can be expressed by a Π_1 wff too:

$$\forall \mathsf{x}\, (\mathsf{Even}(\mathsf{x}) \wedge \overline{3} \leq \mathsf{x} \to \exists \mathsf{y} \leq \mathsf{x} \exists \mathsf{z} \leq \mathsf{x}\, (\mathsf{Prime}(\mathsf{y}) \wedge \mathsf{Prime}(\mathsf{z}) \wedge \mathsf{x} = \mathsf{y} + \mathsf{z}))$$

Mathematicians have been searching for a proof of this conjecture or of its negation since the 18th century. Maybe it's undecidable.

2 First Incompleteness Theorem: the semantic version

Theorem 36. *If PA is sound, then G_{PA} is true but undecidable in PA, so PA is negation incomplete.*

Proof. Assume that PA is sound and that G_{PA} is not true. Then, what G_{PA} says is not the case. Since G_{PA} says of itself that it is not provable in PA, then G_{PA} must be provable in PA. But then, PA would prove a false statement, which is impossible, as it is sound. Therefore, we must conclude that G_{PA} is true.

But then, G_{PA} cannot be provable in PA, for in that case it would be false. Its negation cannot be provable either, for then PA would prove $\mathsf{Prov}_{\mathsf{PA}}(\overline{\ulcorner \mathsf{G}_{\mathsf{PA}} \urcorner})$. Since PA is sound, that would mean that G_{PA} is provable in PA. But we already showed this is not possible. As a consequence, neither $\mathsf{PA} \vdash \mathsf{G}_{\mathsf{PA}}$ nor $\mathsf{PA} \vdash \neg\mathsf{G}_{\mathsf{PA}}$: G_{PA} is undecidable in PA. □

This is not a problem specific to PA but to any sound and p.r. – and also just recursively – axiomatized theory formulated in L_A.

Theorem 38. *Let T be a p.r. axiomatized theory formulated in L_A. If T is sound, then G_T is true but undecidable in T, so T is negation incomplete.*

Proof. Assume that T is sound and that G_T is not true. Then, what G_T says is not the case. Since G_T says of itself that it is not provable in T, then G_T must be provable in T. But then, T would prove a false statement, which is impossible, as it is sound. Therefore, we must conclude that G_T is true.

But then, G_T cannot be provable in T, for in that case it would be false. Its negation cannot be provable either, for then T would prove $\mathsf{Prov}_T(\overline{\ulcorner \mathsf{G}_T \urcorner})$. Since T is sound, that would mean that G_T is provable in T. But we already showed this is not possible. As a consequence, neither $T \vdash \mathsf{G}_T$ nor $T \vdash \neg\mathsf{G}_T$: G_T is undecidable in T. $\square$

This means, for instance, that adding G_{PA} to PA will not result in a complete theory, because $\mathsf{PA} + \mathsf{G}_{\mathsf{PA}}$ will have its own undecidable Gödel sentence.

Actually, the negation incompleteness result extends to any theory p.r. – and also just recursively – axiomatized theory formulated in an effective extension of L_A.

Theorem 37. *Let L be a first-order language s.t. $L_A \subseteq L$ and there is an effective basic coding of the non-logical symbols of L by numbers. Let T be a p.r. axiomatized theory formulated in L. If T is sound, then there is a Gödel sentence G_T of L that is true but undecidable in T, so T is negation incomplete.*

Exercise 4. *Prove Theorem 37.*

Exercise 5. *Let T be as in Theorem 37. Is $T + \neg G_T$ consistent? What about $T + G_T$? Justify your answers.*

Exercise 6. *Is there a p.r. axiomatized theory formulated in L_A that is negation complete? Justify your answer.*

Exercise 7. *Give a true sentence of L_A that says (or is equivalent to the claim) that* PA *is consistent.*

Exercise 8 (*)**.** *Let $T(n)$ be the numerical property that holds of n just in case n is the g.n. of a true sentence of L_A. Show that there is no wff in L_A that expresses this property. Is T a p.r. property?*

3 First Incompleteness Theorem: the syntatic version

3.1 ω-completeness and ω-consistency

Definition (ω-incompleteness)**.** *Let T be a theory formulated in L_A. T is ω-incomplete iff there is a wff $\varphi(\mathsf{x})$ s.t. T proves $\varphi(\overline{\mathsf{n}})$ for each natural number n, but $T \nvdash \forall \mathsf{x}\, \varphi(\mathsf{x})$.*

For instance, Q is ω-incomplete, for it proves $\mathsf{0} + \overline{\mathsf{n}} = \overline{\mathsf{n}}$ for each natural number n but it cannot prove $\forall \mathsf{x}\,(\mathsf{0} + \mathsf{x} = \mathsf{x})$ (cf. Unit 2, Theorem 12). Note that this means that, despite being truth preserving (in L_A's intended interpretation(s)), an inference rule that allows us to go from all the instances $\varphi(\overline{\mathsf{n}})$ of a formula $\varphi(x)$ to $\forall \mathsf{x}\, \varphi(\mathsf{x})$ is *not* available. This is known as the *ω-rule.* Adding such a rule to our theories would make some proofs infinitely long. This would mean we cannot code them using s.g.n., and no proof predicate will be available in the language for us to obtain a Gödel sentence. In fact, adding the ω-rule to, e.g. Q – and, a fortiori, also PA – results in a negation complete – but not effectively axiomatized or axiomatizable – theory.

Definition (ω-inconsistency)**.** *Let T be a theory formulated in L_A. T is ω-inconsistent iff there is a wff $\varphi(\mathsf{x})$ s.t. T proves $\varphi(\overline{\mathsf{n}})$ for each natural number n and, at the same time, $T \vdash \neg\forall \mathsf{x}\, \varphi(\mathsf{x})$.*[1]

For instance, $\mathsf{Q} + \neg\forall \mathsf{x}\,(0 + \mathsf{x} = \mathsf{x})$ is ω-inconsistent. Nonetheless, it is consistent, for otherwise we would have that Q proves $\forall \mathsf{x}\,(0 + \mathsf{x} = \mathsf{x})$.

Exercise 9. *Show that if T is ω-consistent, then it is consistent.*

Theorem 39. *Let T be formulated in L_A. If T is ω-inconsistent then it is unsound.*

Exercise 10. *Prove Theorem 39.*

3.2 PA's incompleteness

Theorem 40. *If PA is consistent, then $\mathsf{PA} \nvdash \mathsf{G_{PA}}$.*

Proof. Assume $\mathsf{G_{PA}}$ is provable in PA, so $\mathsf{PA} \vdash \neg\exists \mathsf{x}\, \mathsf{Prf_{PA}}(\mathsf{x}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$. Then, there must be a number n that is the s.g.n. of a proof of $\mathsf{G_{PA}}$. Thus, $\mathsf{PA} \vdash \mathsf{Prf_{PA}}(\overline{\mathsf{n}}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$, which implies that $\mathsf{PA} \vdash \exists \mathsf{x}\, \mathsf{Prf_{PA}}(\mathsf{x}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$. But that contradicts our assumption that PA is consistent. $\square$

Theorem 41. *If PA is consistent, then it is ω-incomplete.*

Proof. If PA is consistent, by the previous theorem $\mathsf{PA} \nvdash \mathsf{G_{PA}}$, so $\mathsf{PA} \nvdash \neg\exists \mathsf{x}\, \mathsf{Prf_{PA}}(\mathsf{x}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$. The latter entails that PA doesn't prove $\forall \mathsf{x}\, \neg\mathsf{Prf_{PA}}(\mathsf{x}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$. But since $\mathsf{G_{PA}}$ isn't provable in PA either, no number is the s.g.n. of a proof of $\mathsf{G_{PA}}$ in PA, so PA proves all instances of $\neg\mathsf{Prf_{PA}}(\overline{\mathsf{n}}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$ for each natural number n. $\square$

Theorem 42. *If PA is ω-consistent, then $\mathsf{PA} \nvdash \neg\mathsf{G_{PA}}$.*

Proof. Assume $\mathsf{PA} \vdash \neg\mathsf{G_{PA}}$, so $\mathsf{PA} \vdash \exists \mathsf{x}\, \mathsf{Prf_{PA}}(\mathsf{x}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$. The latter entails that PA proves $\neg\forall \mathsf{x}\, \neg\mathsf{Prf_{PA}}(\mathsf{x}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$. But since ω-consistency entails consistency, by Theorem 40 we have that $\mathsf{G_{PA}}$ isn't provable in PA either. Thus, PA proves $\neg\mathsf{Prf_{PA}}(\overline{\mathsf{n}}, \overline{\ulcorner \mathsf{G_{PA}} \urcorner})$ for each natural number n. But then PA would be ω-inconsistent, contradicting our assumptions. $\square$

Theorem 43. *If PA is ω-consistent, it is negation incomplete.*

Exercise 11. *Prove Theorem 43.*

3.3 Generalizing incompleteness

Theorem. *Let T be a p.r. axiomatized theory formulated in L_A that contains Q. If T is consistent, then there is a Π_1 sentence G_T s.t. $T \nvdash \mathsf{G}_T$, and if T is ω-consistent, $T \nvdash \neg\mathsf{G}_T$, so T is negation incomplete.*

Exercise 12. *Prove the theorem above.*

Theorem 44. *Let L be a first-order language s.t. $L_A \subseteq L$ and there is an effective basic coding of the non-logical symbols of L by numbers. Let T be a p.r. axiomatized theory formulated in L that contains Q. If T is consistent, then there is a Π_1 sentence G_T s.t. $T \nvdash G_T$, and if T is ω-consistent, $T \nvdash \neg G_T$, so T is negation incomplete.*

Exercise 13. *Prove Theorem 44.*

[1] We can also define ω-incompleteness and ω-inconsistency for theories formulated in an extension of L_A, but one might have to relativize the quantifiers in the definitions to a predicate $\mathsf{N}(\mathsf{x})$ expressing the property of being a natural number.