**Chapter 5. Gödel’s Second Incompleteness Theorem**

Gödel informally explained his First Incompleteness Theorem noting that the analogy to the antinomy of the Liar “leaps to the eye (GCW-I, 149).” Whereas the Liar sentence asserts of itself that it is *untrue*, the Gödel sentence says of itself that it is *unprovable* in a precisely specified formal system such as *Principia Mathematica*:

G: G is not provable in system PM

Gödel further remarks that *any* epistemological antinomy can be used to motivate his proof but chooses the Richard Paradox, which we discussed in Chapter 1. Gödel’s choice was prescient: the Berry Paradox is a simplification of the Richard Paradox, and can be used to elegantly prove the First Incompleteness Theorem showing its connection with Chaitin’s interpretation Gödel Incompletehes in terms of algorihic randomness.

Gödel’s First Incompleteness Theorem: if a formal system is consistent and its axiom system has enough arithmetic so that its theorems can be listed by some mechanical procedure, then there exists an *undecidable* sentence in that formal system, which is therefore *incomplete*.[[1]](#footnote-1)

**Provability Modal Logics**

□(P → Q) → (□P → □Q)

*~~Show~~* Q

*~~Show~~* P → Q *~~Show~~* P

An intriguing conceptual issue: ∃ vs. ∀?

□ = ⊢ provability = the ∃xistence of a proof?

□ = ⊨ semantic validity = the ∀niversality or truth in ∀ll possible worlds

Why is it legitimate to use the □ for both notions?

Elegant proofs of Gödel’s Second Incompleteness Theorem were discovered in *modal provability logics,* which emerged from the 1950s-1970s. These logics were anticipated by Gödel’s [1933*f*] *“An interpretation of intuitionistic propositional calculus.”* Gödel’s insight was that intuitionistic *truth* was characterized in terms of *proof*, which is a kind of *necessity*, and so modal axioms could be used to formalize the *properties of provability*:

(**T**) □P → P *What is provable is \_\_\_\_\_\_\_\_\_.*

(**K**) □(P → Q) → (□P → □Q) *Whatever follows from what is provable is \_\_\_\_\_\_\_\_\_.*

(**4**) □P → □□P *What is provable is provably \_\_\_\_\_\_\_\_\_\_\_.*

Leon Henkin [1952] posed the intriguing question whether the *positive* Gödelian sentence “I am *provable*” is provable. Martin Löb [1955] answered Henkin’s question in the affirmative by showing that Peano Arithmetic proves a counterpart to Löb’s Axiom:

(**L**) □(□P → P) → □P *Löb’s axiom restricts* (**T**) *to what is provable*

A *Gödel-Löb modal probability logic* (**GL**) results from adding (**L**) to (**K**) and

1. the rule of *necessitation* or *universal derivation* [UD] (*i.e.,* if **GL** ⊢ P then **GL** ⊢ □P),
2. *modus ponens* [MP]
3. a rule for proving all *tautologies* or the KM2 system of natural deduction.

Note: adding axiom (**4**) turns out to be redundant. In 1975 Howard de Jongh proved that Axiom (**4**) is derivable from Löb’s axiom and (**K**) using the substitution of ‘(□P ∧ P)’ for ‘P’).

A *provability system* consists minimal modal logic **K** with the additional axiom:

(**W**) □(□ϕ → ϕ) → □ϕ .

*Theorem* (Howard de Jongh, 1970s)[[2]](#footnote-2): Axiom (**4**) holds in the Gödel-Lob Provability Logic.

*Proof*:

Replacing ϕ by (□ϕ ∧ ϕ) in Axiom (**W**) we obtain:

□(□(□ϕ ∧ ϕ) → (□ϕ ∧ ϕ)) → □(□ϕ ∧ ϕ)

It happens that the following lemma is derivable in minimal modal logic **K** (see exercises below):

□ϕ → □(□(□ϕ ∧ ϕ)) → □(□ϕ ∧ ϕ))

It therefore follows that

□ϕ → □(□ϕ ∧ ϕ)

and hence by (**T**)

□ϕ → □□ϕ .

Lemma:

1. *~~Show~~* □ϕ → □(ϕ → (□(□ϕ ∧ ϕ) → (□ϕ ∧ ϕ))

2. □ϕ Assume (\_\_\_)

3. □(□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ) → □(□ϕ ∧ ϕ) (L)

4. *~~Show~~* □(ϕ → (□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ)) 5, \_\_\_\_

5. *~~Show~~* ϕ → (□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ) 7, \_\_\_\_

6. ϕ Assume (\_\_)

7. *~~Show~~* □(□ϕ ∧ ϕ) → □ϕ ∧ ϕ 11, \_\_\_\_

8. □(□ϕ ∧ ϕ) Assume (\_\_\_)

9. □□ϕ ∧ □ϕ Distribution □/∧

10. □ϕ 9, S

11. □ϕ ∧ ϕ 10, 6 ADJ

Next comes the Löb-Gödel Theorem, which we shall call the “Magical Modal Mystery Tour”:[[3]](#footnote-3)

if ⊢ □ϕ → ϕ, then ⊢ ϕ

1. ⊢ □ϕ → ϕ Assumption (**T**)
2. ⊢ σ ↔ (□σ → ϕ) Gödel Fixed-Point Theorem
3. *~~Show~~* ϕ 20, \_\_\_
4. *~~Show~~* □σ 16, \_\_\_
5. *~~Show~~* □σ → ϕ 14, \_\_\_
6. □σ Assume (\_\_\_)
7. *~~Show~~* □[σ ↔ (□σ → ϕ)] 8, \_\_\_
8. σ ↔ (□σ → ϕ) 2, Gödel Fixed-Point Theorem
9. □σ → □(□σ → ϕ)] 7, T81, IE, MD, S, Axiom (**K**), MP
10. □(□σ → ϕ) 9, 6 \_\_\_
11. □□σ → □ϕ 10, Axiom (\_\_\_), MP
12. □σ → □□σ Axiom (\_\_\_)
13. □ϕ 12, 6 \_\_\_, 11 \_\_\_
14. ϕ 13, (\_\_\_\_)
15. (□σ → ϕ) → σ 2, \_\_\_
16. σ 15, 5 \_\_\_
17. σ → (□σ → ϕ) 2, \_\_\_
18. σ 4, (\_\_\_)
19. □σ → ϕ 17, 18 \_\_\_
20. ϕ 19, 4 \_\_\_

Now we may obtain a modal version of Gödel’s First Incompleteness Theorem. Gödel’s famous arithmetical version of the Liar G intuitively says, “I am not provable”:

G ↔ ~□G

Note the consistency of Peano Arithmetic may be formulated as: ~□ (1 = 2) (von Neumann) or ~□ 0 ≠ 0.

We can sketch an elegant proof in modal provability logic of Gödel’s Second Incompleteness Theorem. First, we have a modal counterpoint to the *fixed-point theorem* that yields the *Gödel sentence*:

⊢ G ↔ ~□G .

In his letter von Neumann noted that the consistency of Peano Arithmetic (PA) can be expressed by the formula that (1 = 2) is not provable:

Cons(PA): = ~□ (1 = 2) .

Now the gist of the First Incompleteness Theorem is the demonstration that:

if ⊢ G ↔ ~□G, then ⊢ ~□(1 = 2) → ~□G .

By the fixed-point theorem, ~□G is *logically equivalent* to G, so we have ⊢ ~□ 1 = 2 → G:

~□ 1 = 2 → ~□G

*fixed-point theorem*

G

By the rule of necessitation, we may prefix a **□** and then distribute, using (**K**), the **□** over the conditional:

⊢ **□**~□ 1 = 2 → **□**G

According to the First Incompleteness Theorem, the provability of the Gödel sentence implies the inconsistency of the system, so:

**□**~□ 1 = 2 → **□**G

*First Incompleteness Theorem*

**□** 1 = 2

In short, we have:

⊢ **□**~□ 1 = 2 → **□** 1 = 2 ,

which, by contraposition, yields:

⊢ ~□ 1 = 2 → ~□~□ 1 = 2 .

Since ~□ 1 = 2, by definition, is Cons(PA), we have:

Gödel’s Second Incompleteness Theorem. ⊢ (Cons(PA) → ~□ Cons(PA)), *i.e.*, if Peano Arithmetic is *consistent*, then it cannot prove its own *consistency*.

In a lecture to a joint meeting of the Mathematical Association of America and the American Mathematical Society, Gödel summarized the significance of his result for Hilbert’s program: the hope of finding *“…a proof for freedom from contradiction undertaken by Hilbert and his disciples” had “vanished entirely in view of some recently discovered facts. It can be shown quite generally that there can exist no proof of the freedom of contradiction of a formal system S which could be expressed in terms of the formal system S itself ….”* ([\*1933*o*], GCW-III, 52).

Solutions to Problems:

1. *~~Show~~* □ϕ → □(ϕ → (□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ)) 3, CD

2. □ϕ Assume (CD)

3. □(□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ) → □(□ϕ ∧ ϕ) Axiom (L)

4. *~~Show~~* □(ϕ → (□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ)) 5, ND

5. *~~Show~~* ϕ → (□(□ϕ ∧ ϕ) → □ϕ ∧ ϕ) 7 CD

6. ϕ Assume (CD)

7. *~~Show~~* □(□ϕ ∧ ϕ) → □ϕ ∧ ϕ 11, CD

8. □(□ϕ ∧ ϕ) Assume (CD)

9. □□ϕ ∧ □ϕ 8, (**K**), MP

10. □ϕ 9, S

11. □ϕ ∧ ϕ 10, 6 ADJ (note we do not cross a □-derivation)

1. ⊢ □ϕ → ϕ Assumption (**T**)
2. ⊢ σ ↔ (□σ → ϕ) Gödel Fixed-Point Theorem
3. *~~Show~~* ϕ 20, DD
4. *~~Show~~* □σ 16, ND
5. *~~Show~~* □σ → ϕ 14, CD
6. □σ Assume (CD)
7. *~~Show~~* □[σ ↔ (□σ → ϕ)] 8, ND
8. σ ↔ (□σ → ϕ) 2, Gödel Fixed-Point Theorem
9. □σ → □(□σ → ϕ)] 7, T81, IE, MD, S, Axiom (K), MP
10. □(□σ → ϕ) 9, 6 MP
11. □□σ → □ϕ 10, Axiom (**K**), MP
12. □σ → □□σ Axiom (**4**)
13. □ϕ 12, 6 MP, 11 MP
14. ϕ 13, (**T**)
15. (□σ → ϕ) → σ 2, BC
16. σ 15, 5 MP
17. σ → (□σ → ϕ) 2, BC
18. σ 4, (**T**)
19. □σ → ϕ 17, 18 MP
20. ϕ 19, 4 MP

1. The formal system is also *essentially* *incomplete*, *i.e.,* one can add the undecidable Gödel sentence as a new axiom and the resulting system will have a new undecidable sentence, which is also undecidable in the original system. [↑](#footnote-ref-1)
2. At the Tenth International Tbilisi Symposium on Language, Logic and Computation in 2013, one of the speakers referred to this theorem part of the folklore of the field not realizing that de Jongh was in the audience. [↑](#footnote-ref-2)
3. van Benthan [2010], p. 245 describes this theorem “as a piece of ‘magical’ modal reasoning that has delighted generations.” [↑](#footnote-ref-3)