Logic, Computability and Incompleteness

Gödel's Second Theorem, Löb's Theorem and the Logic of Provability

Hilbert’s Program (again)

- As we saw before, Hilbert’s Formalist Program for the foundation of mathematics advocated an approach in which all of mathematics is deducible in an axiomatizable formal theory where the axioms themselves are provably consistent.
- In particular, Hilbert sought an ‘internal’ and finitary consistency proof for the axioms of elementary number theory.
- Gödel’s First Theorem can be seen as refuting Hilbert’s goal of reducing arithmetic to an ‘inventory of provable formulas’, while Gödel’s Second Theorem can be seen as undermining Hilbert’s quest for an internal consistency proof for the axioms of arithmetic.

The Unprovability of Consistency

Gödel's First Incompleteness Theorem (roughly): if formal arithmetic is consistent, then neither $\boldsymbol{S}$ nor $\neg\, \boldsymbol{S}$ is provable,

where $\boldsymbol{S}$ is constructed such that $\vdash\ \boldsymbol{S} \leftrightarrow \neg\, \boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner)$.

So if arithmetic is consistent then $\boldsymbol{S}$ is unprovable, hence **true**

Is formal arithmetic consistent?

Gödel's Second Incompleteness Theorem (roughly):

if formal arithmetic <u>is</u> consistent,

then it <u>cannot</u> prove its own consistency.

A **basic fact** of Classical Logic: if formal arithmetic is a consistent theory, then at least one sentence is <u>unprovable</u>.

The Unprovability of Consistency

So let the consistency of arithmetic be expressed

in arithmetic (**!**) by the sentence:

$$\neg\ \boldsymbol{Prov}\ (\ulcorner \mathbf{o} = \mathbf{o}' \urcorner) \qquad (\underline{\mathbf{con}})$$

Gödel's Second Incompleteness Theorem:

if $\vdash$ <u>**con**</u> *then* arithmetic is inconsistent.

As above, Gödel's First Theorem states that

(#) *if* arithmetic is consistent, *then* it is <u>*not*</u> provable that ***S***

Since <u>**con**</u> is a sentence of the object language that expresses the consistency of arithmetic, then a formalization of the <u>First</u> <u>Theorem</u>, in terms of (#), would yield:

$$\vdash\ \underline{\mathbf{con}} \rightarrow \neg\ \boldsymbol{Prov}\ (\ulcorner \boldsymbol{S} \urcorner) \qquad (\mathbf{!})$$

The Unprovability of Consistency

i.e $\vdash$ **con** $\rightarrow$ ***S***

So *if* this formalization of the First Theorem is indeed provable in formal arithmetic, *then* it follows (on the assumption of consistency) that not $\vdash$ **con**.

Why? Because $\vdash$ **con** $\rightarrow$ ***S*** entails that

if $\vdash$ **con** *then* $\vdash$ ***S***

And if arithmetic is consistent, then (by 1st Theorem)

not $\vdash$ ***S***, and contraposition on the above yields not $\vdash$ **con**.

So *if* the conditional **con** $\rightarrow$ ***S*** is a theorem of arithmetic,

then if arithmetic is consistent,

then the consistency sentence **con** cannot be provable in arithmetic.

Gödel's Second Incompleteness Theorem

Thus to prove Gödel's Second Incompleteness Theorem need to show that $\vdash$ **con** $\to$ $\boldsymbol{S}$, and the underivability of the consequent $\boldsymbol{S}$ will yield the underivability of the antecedent **con**.

proof: will require the diagonal lemma and characteristics (i) – (iii) of a proof predicate.

As before, the diagonal lemma gives

$\vdash\ \boldsymbol{S} \leftrightarrow \neg\, \boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner)$, which yields $\vdash\ \boldsymbol{S} \to \neg\, \boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner)$

and then $\vdash\ \neg\neg\, \boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner) \to \neg\, \boldsymbol{S}$ and finally

(0) $\vdash\ \boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner) \to \neg\, \boldsymbol{S}$

Recall (i) if $\vdash \boldsymbol{A}$, then $\vdash \boldsymbol{Prov}\,(\ulcorner \boldsymbol{A} \urcorner)$, which applied to (0)

gives $\vdash \boldsymbol{Prov}\,(\ulcorner \underline{\boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner) \to \neg\, \boldsymbol{S}} \urcorner)$

Gödel's Second Incompleteness Theorem

Recall (ii) $\vdash \mathit{Prov}(\ulcorner A \rightarrow B \urcorner) \rightarrow (\mathit{Prov}(\ulcorner A \urcorner) \rightarrow \mathit{Prov}(\ulcorner B \urcorner))$

Which gives $\vdash \mathit{Prov}(\ulcorner \underline{\mathit{Prov}(\ulcorner S \urcorner) \rightarrow \neg S} \urcorner) \rightarrow$

$(\mathit{Prov}(\ulcorner \mathit{Prov}(\ulcorner S \urcorner) \urcorner) \rightarrow \mathit{Prov}(\ulcorner \neg S \urcorner))$

Now modus ponens gives

(1) $\vdash \mathit{Prov}(\ulcorner \mathit{Prov}(\ulcorner S \urcorner) \urcorner) \rightarrow \mathit{Prov}(\ulcorner \neg S \urcorner)$

Similarly, take $\vdash (S \wedge \neg S) \rightarrow \mathbf{o} = \mathbf{o}'$ and apply (i) to get

$\vdash \mathit{Prov}(\ulcorner \underline{(S \wedge \neg S) \rightarrow \mathbf{o} = \mathbf{o}'} \urcorner)$ then apply (ii) to get

$\vdash \mathit{Prov}(\ulcorner \underline{(S \wedge \neg S) \rightarrow \mathbf{o} = \mathbf{o}'} \urcorner) \rightarrow$

$(\mathit{Prov}(\ulcorner (S \wedge \neg S) \urcorner) \rightarrow \mathit{Prov}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner))$ and MP to get

($) $\vdash \mathit{Prov}(\ulcorner (S \wedge \neg S) \urcorner) \rightarrow \mathit{Prov}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner)$

Gödel's Second Incompleteness Theorem

Note that $\vdash \mathit{Prov}(\ulcorner (S \wedge \neg S) \urcorner) \leftrightarrow (\mathit{Prov}(\ulcorner S \urcorner) \wedge \mathit{Prov}(\ulcorner \neg S \urcorner))$

Substitution of provable equivalents in ($) yields

(2) $\vdash (\mathit{Prov}(\ulcorner S \urcorner) \wedge \mathit{Prov}(\ulcorner \neg S \urcorner)) \longrightarrow \mathit{Prov}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner)$

Recall (iii) $\vdash \mathit{Prov}(\ulcorner A \urcorner) \longrightarrow \mathit{Prov}(\ulcorner \mathit{Prov}(\ulcorner A \urcorner) \urcorner)$

which gives: (3) $\vdash \mathit{Prov}(\ulcorner S \urcorner) \longrightarrow \underline{\mathit{Prov}(\ulcorner \mathit{Prov}(\ulcorner S \urcorner) \urcorner)}$

Recall (1) $\vdash \underline{\mathit{Prov}(\ulcorner \mathit{Prov}(\ulcorner S \urcorner) \urcorner)} \longrightarrow \mathit{Prov}(\ulcorner \neg S \urcorner)$

which in combination with (3) yields

(4) $\vdash \mathit{Prov}(\ulcorner S \urcorner) \longrightarrow (\mathit{Prov}(\ulcorner S \urcorner) \wedge \mathit{Prov}(\ulcorner \neg S \urcorner))$

With (4) and (2) we get $\vdash \mathit{Prov}(\ulcorner S \urcorner) \longrightarrow \mathit{Prov}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner)$

And by contraposition

(5) $\vdash \neg \mathit{Prov}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner) \longrightarrow \neg \mathit{Prov}(\ulcorner S \urcorner)$.

Gödel's Second Incompleteness Theorem

Since **con** has been defined as the sentence $\neg$ ***Prov*** $(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner)$

and $\vdash$ $\boldsymbol{S} \leftrightarrow \neg$ ***Prov*** $(\ulcorner \boldsymbol{S} \urcorner)$,

Rewriting (5) $\vdash$ $\neg$ ***Prov*** $(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner) \rightarrow \neg$ ***Prov*** $(\ulcorner \boldsymbol{S} \urcorner)$

as $\vdash$ **con** $\rightarrow \boldsymbol{S}$ yields the desired result

and if arithmetic is consistent then not $\vdash$ **con** $\square$

This is another incompleteness result, because if arithmetic is consistent then **con** is ***true***,

so if arithmetic is consistent then **con** is yet another **unprovable truth**, and the wedge between

truth and provability that started with the First Theorem is driven even deeper.

Löb's Theorem

We've just seen a 'direct' proof of Gödel's Second Incompleteness Theorem. However, it is also possible to prove this theorem as a corollary of the closely related but more general Löb's Theorem, which is motivated as follows.

Another way to think of provable consistency is in terms of the characteristic

for all sentences $\boldsymbol{A}$,

(v) $\quad \vdash \boldsymbol{Prov}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{A}$

Which 'asserts that' if a sentence is provable, then it is true,

so that it's provable in the formal theory that only truths are provable.

Löb's Theorem

Löb's Theorem: if $\boldsymbol{B}(y)$ is a proof predicate for some theory $\mathbf{T}$ that extends $\boldsymbol{Q}$, then for any sentence $\boldsymbol{A}$ in the language of $\mathbf{T}$

if $\vdash_{\mathbf{T}} \boldsymbol{B}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{A}$ *then* $\vdash_{\mathbf{T}} \boldsymbol{A}$

proof: assume (1) $\vdash_{\mathbf{T}} \boldsymbol{B}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{A}$.

Let $\boldsymbol{D}(y)$ be the formula $\boldsymbol{B}(y) \rightarrow \boldsymbol{A}$.

The diagonal lemma guarantees a sentence $\boldsymbol{C}$ such that

$\vdash_{\mathbf{T}} \boldsymbol{C} \leftrightarrow \boldsymbol{D}(\ulcorner \boldsymbol{C} \urcorner)$, i.e.

(2) $\vdash_{\mathbf{T}} \boldsymbol{C} \leftrightarrow (\boldsymbol{B}(\ulcorner \boldsymbol{C} \urcorner) \rightarrow \boldsymbol{A})$

(1) and (2) in combination with (i) – (iii) yield $\vdash_{\mathbf{T}} \boldsymbol{A}$

(see B&J p. 187 for the detailed steps).

Löb's Theorem

Reformulation of Gödel's Second Incompleteness Theorem:

if $\boldsymbol{B}(y)$ is a proof predicate for some theory **T** that extends ***Q***,
then **not** $\vdash_{\mathbf{T}} \neg \boldsymbol{B}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner)$

proof: suppose $\vdash_{\mathbf{T}} \neg \boldsymbol{B}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner)$.

Then $\vdash_{\mathbf{T}} \boldsymbol{B}(\ulcorner \mathbf{o} = \mathbf{o}' \urcorner) \to \mathbf{o} = \mathbf{o}'$ (by prop. logic)

and by Löb's Theorem $\vdash_{\mathbf{T}} \mathbf{o} = \mathbf{o}'$

But $\vdash_{\boldsymbol{Q}} \neg \mathbf{o} = \mathbf{o}'$, and **T** is inconsistent □

So *if* **T** is consistent, *then* it can't prove its own consistency.

The Henkin Sentence

Historically, Löb's Theorem was used to answer a question posed by Henkin with regard to the the 'Henkin Sentence' ***H***. Unlike the Gödel sentence ***S***, ***H*** 'asserts its own **provability**'.

The diagonal lemma guarantees an ***H*** such that

$$\vdash \ \boldsymbol{H} \leftrightarrow \boldsymbol{Prov}\,(\ulcorner \boldsymbol{H} \urcorner)$$

Is ***H*** <u>provable</u> (and hence **true**)?

It follows directly from Löb's Theorem that $\vdash \ \boldsymbol{H}$.

Contrast with the 'truth teller', which is contingent.

Modal Logic of Provability

The defining characteristics of a provability predicate have very clear analogues in modal logic:

(i) *if* $\vdash \boldsymbol{A}$, *then* $\vdash \boldsymbol{Prov}\,(\ulcorner \boldsymbol{A} \urcorner)$

corresponds to the modal inference rule of *Necessitation*

if $\vdash \boldsymbol{A}$, *then* $\vdash \Box\, \boldsymbol{A}$

(ii) $\vdash \boldsymbol{Prov}\,(\ulcorner \boldsymbol{A} \rightarrow \boldsymbol{B} \urcorner) \rightarrow (\boldsymbol{Prov}\,(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{Prov}\,(\ulcorner \boldsymbol{B} \urcorner))$

corresponds to the **K** axiom schema

$\Box\, (\mathbf{A} \rightarrow \mathrm{C}) \rightarrow (\Box\, \mathbf{A} \rightarrow \Box\, \mathbf{C})$

Modal Logic of Provability

(iii) $\vdash \boldsymbol{Prov}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{Prov}(\ulcorner \boldsymbol{Prov}(\ulcorner \boldsymbol{A} \urcorner) \urcorner)$

corresponds to the **S4** axiom schema

$$\Box \boldsymbol{A} \rightarrow \Box \Box \boldsymbol{A}$$

(v) $\vdash \boldsymbol{Prov}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{A}$

corresponds to the **T** axiom schema

$$\Box \boldsymbol{A} \rightarrow \boldsymbol{A}$$

The modal theory **S4** is the closure of all the **K**, **S4** and **T** axioms under logical consequence and the rule of *Necessitation*.

Modal Logic of Provability

Hence **S4** is too strong to represent the logic of arithmetical proof, as shown by the Gödel-Löb results.

Instead, need to replace **T** axiom schema with the formalized and then modalized version of Löb's Theorem.

Recall Löb's Theorem:

$$\textit{if} \quad \vdash \boldsymbol{\mathit{Prov}}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{A} \quad \textit{then} \vdash \boldsymbol{A}$$

Löb's Theorem formalized in arithmetic:

$$\vdash \quad \boldsymbol{\mathit{Prov}}(\ulcorner \boldsymbol{\mathit{Prov}}(\ulcorner \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{A} \urcorner) \rightarrow \boldsymbol{\mathit{Prov}}(\ulcorner \boldsymbol{A} \urcorner)$$

The modal version of the formalization yields the **G** axiom schema:

$$\Box(\Box \boldsymbol{A} \rightarrow \boldsymbol{A}) \rightarrow \Box \boldsymbol{A}$$

Modal Logic of Provability

- The modal theory **G** is the closure of all the **K**, **S4** and **G** axioms under logical consequence and the rule of *Necessitation*.
- Hence the modal theory **G** represents the logic of provability in formal arithmetic.

Montague and Predicate Modal Logic

Suppose the modal concept of **Necessity** is formalized as a 1-place metalinguistic predicate ***N***(x)

attaching to **names** of formulas,

rather than as an **operator** on formulas, i.e. □.

So the assertion that it is necessarily the case that Φ

is formalized as ***N***$(\ulcorner\Phi\urcorner)$ rather than as □Φ

Then if the modal theory in question incorporates formal arithmetic and the comparatively weak modal structure of the rule of *Necessitation* (i) and the **T** axiom schema (v),

then the theory is inconsistent….

Montague and Predicate Modal Logic

proof: the diagonal lemma guarantees a sentence $\boldsymbol{M}$ such that

$\vdash\ \boldsymbol{M} \leftrightarrow \neg \boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner)$ So …

(1)	$\vdash$	$\boldsymbol{M} \leftrightarrow \neg \boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner)$	by diagonal lemma
(2)	$\vdash$	$\boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner) \rightarrow \neg \boldsymbol{M}$	from (1)
(3)	$\vdash$	$\boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner) \rightarrow \boldsymbol{M}$	by (v)
(4)	$\vdash$	$\neg \boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner)$	prop log on (2), (3)
(5)	$\vdash$	$\boldsymbol{M}$	by (1), (4)
(6)	$\vdash$	$\boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner)$	(i) applied to (5)
(7)	$\vdash$	$\boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner) \wedge \neg \boldsymbol{N}(\ulcorner \boldsymbol{M} \urcorner)$	from (4), (6)

And the modal theory is inconsistent □

Leibniz’s Law

Leibniz’s Law is the principle that the *truth-value of a statement should be preserved under the substitution of co-referential terms*.

It can be seen as a direct corollary of Frege’s principle of compositionality:

the semantic value of the whole is a function of the semantic values of the relevant parts and their mode of combination.

Leibniz’s Law holds in all purely extensional contexts.

However, the law can **fail** in propositional attitude contexts such as *knowledge* and *belief*.

Failure of Leibniz's Law

For example, consider the following:

(i) Aristotle knew **that** 9 > 7 (true)

(ii) 9 = the number of planets (true)

(iii) Aristotle knew **that** the number of planets > 7 (false)

Or consider an example using belief

(i) Frank believes **that** gold is valuable (true)

(ii) Gold is the element with atomic number 79 (true)

(iii) Frank believes **that** the element with atomic number 79 is valuable (*de dicto* false, if Frank is unaware of (ii))

Failure of Leibniz's Law in Metamathematics

Let $\boldsymbol{S}$ be the Gödel sentence, so that (again)

$$\vdash \boldsymbol{S} \leftrightarrow \neg \textbf{\textit{Prov}} (\ulcorner \boldsymbol{S} \urcorner).$$

If arithmetic is consistent then

$\boldsymbol{S}$ is true and <u>unprovable</u>, while

$\neg \boldsymbol{S}$ is false and <u>unprovable</u>.

Suppose $\ulcorner \boldsymbol{S} \urcorner = \mathbf{a}$.

Then $\boldsymbol{S}$ is equivalent to the sentence $\neg \textbf{\textit{Prov}} (\mathbf{a})$.

Let $\boldsymbol{d}$ be the definite description (using Russell's variable-binding, term-forming iota operator $(\boldsymbol{\iota} x)$,

to be read as 'the x such that…')

$$\boldsymbol{d} = (\boldsymbol{\iota} x) ((\neg \textbf{\textit{Prov}} (\mathbf{a}) \rightarrow x = \mathbf{a}) \wedge (\textbf{\textit{Prov}} (\mathbf{a}) \rightarrow x = \ulcorner \mathbf{0} = \mathbf{0}' \urcorner))$$

Failure of Leibniz's Law in Metamathematics

If arithmetic is consistent then ***d* = a**.

But this truth is not provable in arithmetic,

because it would require proving ¬ ***Prov*** (**a**).

Hence, we have the situation:

(1) It is provable in arithmetic **that a = a** (true)

(2) **a = *d*** (true)

(3) It is provable in arithmetic **that a = *d*** (false)

Where (3) is derived from (1) by the substitution of co-referential terms given in (2).

So the context 'It is provable in arithmetic **that** … '

violates Leibniz's Law, **if** arithmetic is consistent.