Logic, Computability and Incompleteness

Undecidability, Indefinability and Gödel’s First Theorem

Definability and Decidability

- Some important technical concepts and terminology:

Definability: a set of natural numbers Θ is

definable in a theory **T** iff

there is a *formula* $\boldsymbol{B}(x)$ in the language of **T** such that for any number k,

if $k \in \Theta$ then $\vdash_{\mathbf{T}} \boldsymbol{B}(\mathbf{k})$, and

if $k \notin \Theta$ then $\vdash_{\mathbf{T}} \neg\, \boldsymbol{B}(\mathbf{k})$

in which case the formula $\boldsymbol{B}(x)$ defines the set Θ in **T**.

Decidability: a set of expressions is decidable if the set of

Gödel numbers of its members is a recursive set, i.e.

if the characteristic function of the set is recursive.

Definability and Decidability

Thus, if Θ^{Gn} is the set of Gödel numbers of expressions in Θ

and if $f_{\Theta Gn}$ is the characteristic function of Θ^{Gn}

then $f_{\Theta Gn}(n) = 1$ iff $n \in \Theta^{Gn}$ (and = 0 otherwise)

and if $f_{\Theta Gn}$ is recursive then the set of expressions Θ is decidable

So a theory **T** is **decidable** iff

the set of Gödel numbers of its theorems is a **recursive** set.

- Connection between the two notions:

  if a set of expressions Θ is **decidable** then the respective characteristic function is **recursive**

  and hence is representable in ***Q***

Definability and Decidability

Which in turn means that the set of Gödel numbers of expressions in Θ is **definable** in $\boldsymbol{Q}$.

This is because if the characteristic function

$f_{\Theta^{Gn}}$ of Θ^{Gn} is recursive

and the formula $\boldsymbol{A}_{f_{\Theta Gn}}(x,y)$ represents $f_{\Theta Gn}$ in $\boldsymbol{Q}$,

then $\boldsymbol{A}_{f_{\Theta Gn}}(x,\mathbf{1})$ defines Θ^{Gn} in $\boldsymbol{Q}$ (!)

So if a theory **T** is **decidable** then

the set of Gödel numbers of its theorems is **definable** in $\boldsymbol{Q}$.

Definability and Decidability

- **Lemma**: if **T** is a consistent extension of $\boldsymbol{Q}$, then the set of Gödel numbers of theorems of **T** is **not definable** in **T**.
- **proof**: by *reductio*, using basic template furnished by the diagonal lemma.

Let $\boldsymbol{C}(y)$ define the set of Gödel numbers of theorems of **T**.

The function **diag** is representable in **T** and $\neg\, \boldsymbol{C}(y)$ is a formula with only the variable y free.

So by the diagonal lemma there is a sentence $\boldsymbol{G}$ such that

$$(*) \quad \vdash_{\mathbf{T}} \boldsymbol{G} \leftrightarrow \neg\, \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner).$$

Suppose $\boldsymbol{gn}[\boldsymbol{G}] = k$, so $\ulcorner \boldsymbol{G} \urcorner = \mathbf{k}$. Then

$$\text{(i)} \quad \vdash_{\mathbf{T}} \boldsymbol{G} \leftrightarrow \neg\, \boldsymbol{C}(\mathbf{k}).$$

Definability and Decidability

It follows by (sub) reductio that $\vdash_{\mathbf{T}}$ $\boldsymbol{G}$, for if **not** $\vdash_{\mathbf{T}} \boldsymbol{G}$,
then, since $\boldsymbol{C}(y)$ defines the set of theorems of **T**,
we get $\vdash_{\mathbf{T}} \neg\, \boldsymbol{C}(\mathbf{k})$ and hence $\vdash_{\mathbf{T}}$ $\boldsymbol{G}$ by (i) [going R to L].
So $\vdash_{\mathbf{T}}$ $\boldsymbol{G}$. Thus $k \in \Theta$ and $\vdash_{\mathbf{T}} \boldsymbol{C}(\mathbf{k})$.
By (i) we get $\vdash_{\mathbf{T}}$ $\boldsymbol{G} \rightarrow \neg\, \boldsymbol{C}(\mathbf{k})$
contraposition yields $\vdash_{\mathbf{T}} \neg\neg\, \boldsymbol{C}(\mathbf{k}) \rightarrow \neg\, \boldsymbol{G}$, which yields
$\vdash_{\mathbf{T}}$ $\boldsymbol{C}(\mathbf{k}) \rightarrow \neg\, \boldsymbol{G}$, and finally by modus ponens $\vdash_{\mathbf{T}} \neg\, \boldsymbol{G}$.
So both $\vdash_{\mathbf{T}}$ $\boldsymbol{G}$ **and** $\vdash_{\mathbf{T}} \neg\, \boldsymbol{G}$, rendering **T** inconsistent,
contrary to initial hypothesis.

Conclusion: *there can be no such* $\boldsymbol{C}(y)$ ■

Undecidability of FOL (from a different angle)

- Bigger conclusion: **no consistent** extension of ***Q*** is **decidable**.
  Why?
- Because if the theory **T** were **decidable**, then its set of theorems would be **definable** in ***Q*** and hence in **T**
- **Church's Theorem**: FOL is undecidable.
- **proof**: we have just established that ***Q*** is undecidable,
  since it is a consistent extension of itself.
  Let Φ be the single sentence formed by conjoining
  all of the 7 axioms of ***Q***.
  Then a sentence ***S*** is a theorem of ***Q*** iff the conditional
  $\Phi \rightarrow$ ***S*** is a theorem of FOL.

Undecidability of FOL (from a different angle)

In other words

$$\vdash_{\boldsymbol{Q}} \boldsymbol{S} \quad \text{iff} \quad \vdash_{\text{FOL}} (\Phi \rightarrow \boldsymbol{S})$$

Hence (intuitively) *if* FOL were decidable *then* so would ***Q*** be.

To carry out this *reductio* **proof** more formally,

let $gn[\Phi] = q$ and let the function f be defined such that

$$f(n) = 1 * (q * (39999 * (n * 2)))$$

f is recursive (by construction)

and if n is the Gödel number of the sentence $\boldsymbol{S}$,

then $f(n)$ is the Gödel number of the sentence $(\Phi \rightarrow \boldsymbol{S})$

Undecidability of FOL (from a different angle)

- Let Θ be the set of Gödel numbers of theorems of FOL.

  If Θ is <u>recursive</u> then so is $\{n: f(n) \in \Theta\}$.

  But $\{n: f(n) \in \Theta\}$ is the set of Gödel numbers of theorems of ***Q***, which has just been shown not to be decidable.
- Thus Θ is not <u>recursive</u> and FOL is not decidable ■

Indefinability of Arithmetical Truth [optional start]

- **Tarski's Theorem**: the set of Gödel numbers of true sentences of arithmetic is not definable in arithmetic.
- **proof**: suppose some formula $\boldsymbol{C}(y)$ defined the set of truths. Then for all sentences $\boldsymbol{S}$ in the language of arithmetic:

  (i) if $\boldsymbol{S}$ then $\vdash_{\boldsymbol{Q}} \boldsymbol{C}(\ulcorner \boldsymbol{S} \urcorner)$ and

  (ii) if $\neg \boldsymbol{S}$ then $\vdash_{\boldsymbol{Q}} \neg \boldsymbol{C}(\ulcorner \boldsymbol{S} \urcorner)$

  By the diagonal lemma there is a sentence $\boldsymbol{G}$ such that

  (*′) $\vdash_{\boldsymbol{Q}} \boldsymbol{G} \leftrightarrow \neg \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$.

  $\boldsymbol{G}$ is either true or false, and since $\boldsymbol{C}(y)$ defines the set of Gödel numbers of true sentences, exactly one of

  $\vdash_{\boldsymbol{Q}} \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$ or $\vdash_{\boldsymbol{Q}} \neg \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$ must obtain.

Indefinability of Arithmetical Truth [optional]

Suppose $\vdash_{Q} \neg \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$. Then $\vdash_{Q} \boldsymbol{G}$ by (*′)

and $\vdash_{Q} \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$ by (i), and $\boldsymbol{Q}$ is inconsistent.

Suppose $\vdash_{Q} \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$. Then

$\vdash_{Q} \boldsymbol{G} \rightarrow \neg \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$ by (*′) , contraposition yields

$\vdash_{Q} \neg\neg \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner) \rightarrow \neg \boldsymbol{G}$, then $\vdash_{Q} \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner) \rightarrow \neg \boldsymbol{G}$

MP yields $\vdash_{Q} \neg \boldsymbol{G}$, and finally $\vdash_{Q} \neg \boldsymbol{C}(\ulcorner \boldsymbol{G} \urcorner)$ by (ii),

and $\boldsymbol{Q}$ is inconsistent.

So *if* $\boldsymbol{Q}$ is consistent *then* <u>there is no such</u> $\boldsymbol{C}(y)$

and the set of Gödel numbers of true sentences

of arithmetic is not definable in arithmetic ■

Indefinability of ‘True-in-*L*’ in *L* [optional]

- More general version of **Tarski’s Theorem**: suppose $\boldsymbol{Tr}(x)$ is a formula in a language $\boldsymbol{L}$ attaching to names of formulas of $\boldsymbol{L}$, and $\boldsymbol{Tr}(x)$ is intended to be a truth predicate for $\boldsymbol{L}$, in which case it must satisfy the Tarski biconditional schema:

for all sentences $\boldsymbol{S}$ of $\boldsymbol{L}$,

$$\vdash_{\boldsymbol{L}} \boldsymbol{Tr}(\ulcorner \boldsymbol{S} \urcorner) \leftrightarrow \boldsymbol{S}$$

The Tarski biconditional schema is famously illustrated by the example:

The sentence ‘Snow is white’ is true iff snow is white.

Suppose further that the diagonal function is representable in in $\boldsymbol{L}$.

Indefinability of ‘True-in-L’ in L [optional finish]

Since $\boldsymbol{Tr}(x)$ is a formula of $\boldsymbol{L}$, so is $\neg\, \boldsymbol{Tr}(x)$

and by the diagonal lemma there is a sentence $\boldsymbol{G}$ of $\boldsymbol{L}$ such that

$$(*'') \quad \vdash_{L} \boldsymbol{G} \leftrightarrow \neg\, \boldsymbol{Tr}(\ulcorner \boldsymbol{G} \urcorner)$$

$\boldsymbol{G}$ is the notorious ‘liar’ sentence that ‘asserts its own falsity’

Since $\boldsymbol{G}$ is a sentnce of $\boldsymbol{L}$, the Tarski biconditional schema must apply to $\boldsymbol{G}$, yielding

$\vdash_{L} \boldsymbol{Tr}(\ulcorner \boldsymbol{G} \urcorner) \leftrightarrow \boldsymbol{G}$ which, in combination with

$(*'')$ $\vdash_{L} \boldsymbol{G} \leftrightarrow \neg\, \boldsymbol{Tr}(\ulcorner \boldsymbol{G} \urcorner)$ yields the contradiction

$$\vdash_{L} \boldsymbol{Tr}(\ulcorner \boldsymbol{G} \urcorner) \leftrightarrow \neg\, \boldsymbol{Tr}(\ulcorner \boldsymbol{G} \urcorner)$$

Conclusion: if $\boldsymbol{L}$ is consistent then it cannot contain its own truth predicate ■

Gödel's First Incompleteness Theorem

- A formal theory **T** is (negation) **complete** iff
  for all sentences ***S*** in the language of **T**, either $\vdash_{\mathbf{T}} \boldsymbol{S}$ or $\vdash_{\mathbf{T}} \neg \boldsymbol{S}$.
- So a formal theory **T** is **incomplete** iff it is **not** the case that
  for all sentences ***S*** in the language of **T**, either $\vdash_{\mathbf{T}} \boldsymbol{S}$ or $\vdash_{\mathbf{T}} \neg \boldsymbol{S}$.

- **Gödel's First Incompleteness Theorem** (1931):
  If formal arithmetic is **consistent**, **then** it is **incomplete**.

Gödel's First Incompleteness Theorem

- **proof**: will construct a Gödel sentence ***S*** that

  'asserts its own unprovability',

  and demonstrate that neither ***S*** nor ¬ ***S*** is provable

  if the formal theory of arithmetic is consistent.

  To do this, will first need to scrutinize (and then 'arithmetize') the structure of formal proofs.

  For present purposes we'll think of axiomatic ('Hilbert style') formal proofs.

  Basic ingredients required for an axiomatic system **AX**:

  a set of axioms and a set of inference rules.

Formal Axiomatic Proofs

Then a proof of some conclusion $\boldsymbol{C}$

from <u>premises</u> $\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n$

is a finite sequence of formulas,

$$\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_k$$

where $\boldsymbol{F}_k$ is the conclusion $\boldsymbol{C}$,

and where each $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots,$ in the sequence is either one of the premises $\boldsymbol{B}_i$, or is one of the axioms, or is obtained from some earlier $\mathbf{F}_i$'s in the sequence by using a rule of inference.

If there is such a proof sequence, then we write

$$\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n \vdash_{\mathbf{AX}} \boldsymbol{C}$$

Formal Axiomatic Proofs

For convenience, a proof sequence can also be written vertically, as follows:

1. $\boldsymbol{F_1}$

2. $\boldsymbol{F_2}$

$\vdots$

***k*.** $\boldsymbol{F_k}$ **(i.e., *C*)**

- We will be concerned with axiomatic proofs of theorems of ***Q*** (and axiomatic extensions),

where **FOL=** can be formalized in terms of finite collection of **axiom schemas** and the *single* **inference rule**

of modus ponens (MP).

Formal Axiomatic Proofs

Here is such an axiomatic proof system for Propositional Logic using just the connectives $\neg$ and $\rightarrow$

Logical Axioms Schemas

(I) $\mathbf{A \rightarrow (B \rightarrow A)}$

(II) $\mathbf{((A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C)}$

(III) $\mathbf{(\neg B \rightarrow \neg A) \rightarrow ((\neg B \rightarrow A) \rightarrow B)}$

Any *instance* of an Axiom Schema is a **logical axiom**

Rule of Inference: Modus Ponens

If you have formulas **A** and $\mathbf{A \rightarrow B}$ at some point in the proof sequence (in either order), then you can add **B** at a later point in the proof sequence.

[This axiomatic system for propositional logic is **complete**]

Formal Axiomatic Proofs

Here is an axiomatic proof for

$\vdash P \rightarrow P$ (so no premises involved)

1. $(P \rightarrow ((P \rightarrow P) \rightarrow P)) \rightarrow ((P \rightarrow (P \rightarrow P)) \rightarrow (P \rightarrow P))$
instance of (II)

2. $P \rightarrow ((P \rightarrow P) \rightarrow P)$ instance of (I)

3. $(P \rightarrow (P \rightarrow P)) \rightarrow (P \rightarrow P)$ MP 1, 2

4. $P \rightarrow (P \rightarrow P)$ instance of (I)

5. $P \rightarrow P$ MP 3, 4

Formal Axiomatic Proofs

- To extend this Propositional fragment to a formalization of full **FOL=** we need the following **Logical Axiom Schemas**:

(IV) $\forall \mathbf{x}(\Phi \rightarrow \Psi) \rightarrow (\forall \mathbf{x}\Phi \rightarrow \forall \mathbf{x}\Psi)$

(V) $\Phi \rightarrow \forall \mathbf{x}\Phi$ if **x** not free in Φ

(VI) $\forall \mathbf{x}\Phi \rightarrow \Phi(x/t)$ where **t** replaces all free occurrences of **x** in Φ

(VII) $\mathbf{t} = \mathbf{t}$ for any term **t**

(VIII) $\mathbf{s}_1 = \mathbf{t}_1 \rightarrow (\mathbf{s}_2 = \mathbf{t}_2 \rightarrow (\ldots \rightarrow \mathbf{s}_n = \mathbf{t}_n \rightarrow (\mathbf{f}(\mathbf{s}_1, \ldots \mathbf{s}_n) = \mathbf{f}(\mathbf{t}_1, \ldots \mathbf{t}_n)))$

for any n ≥ 1, and 2*n* terms $\mathbf{s}_1, \ldots \mathbf{s}_n, \mathbf{t}_1, \ldots \mathbf{t}_n$ and *n*-place function symbol **f**

(IX) $\mathbf{s}_1 = \mathbf{t}_1 \rightarrow (\mathbf{s}_2 = \mathbf{t}_2 \rightarrow (\ldots \rightarrow \mathbf{s}_n = \mathbf{t}_n \rightarrow (\mathbf{P}(\mathbf{s}_1, \ldots \mathbf{s}_n) \rightarrow \mathbf{P}(\mathbf{t}_1, \ldots \mathbf{t}_n)))$

for any n ≥ 1, and 2*n* terms $\mathbf{s}_1, \ldots \mathbf{s}_n, \mathbf{t}_1, \ldots \mathbf{t}_n$ and *n*-place predicate symbol **P**

As before, any *instance* of an Axiom Schema is a **logical axiom**

Formal Axiomatic Proofs

We then add the 7 **non-logical** axioms of $\boldsymbol{Q}$ to the **set of axioms**.

- Thus the formal theory $\boldsymbol{Q}$ is the <u>deductive</u> <u>closure</u> in the language $\boldsymbol{L}$ of the axiomatic system that we have just defined.
- Hence if sentence $\boldsymbol{S}$ of $\boldsymbol{L}$ is a <u>theorem</u> of $\boldsymbol{Q}$

  (written $\vdash_{\boldsymbol{Q}} \boldsymbol{S}$ and implying that $\boldsymbol{S} \in \boldsymbol{Q}$),

  then there is a finite sequence of formulas,

  $$\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_k$$

  where $\boldsymbol{F}_k$ is the sentence $\boldsymbol{S}$, and each $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots$, in the sequence is either an axiom of **FOL=**,

  or an axiom of $\boldsymbol{Q}$, or is obtained from two previous formulas in the sequence $\boldsymbol{F}_n$ and $\boldsymbol{F}_m$ using MP.

Arithmetizing Proofs

- The set of (Gödel numbers of) axioms of **FOL=** plus axioms of ***Q*** is **recursive** and hence is definable in ***Q***

Furthermore, it's a **recursive** matter to determine whether a given sentence follows from 2 other sentences via MP:

Arithmetizing Proofs

E.g., consider the sequence ...., $(\boldsymbol{A} \rightarrow \boldsymbol{B})$, ..., $\boldsymbol{A}$, ..., $\boldsymbol{B}$

$\boldsymbol{gn}[\ (\ ] = 1$, $\quad \boldsymbol{gn}[\ )\ ] = 2$, $\quad \boldsymbol{gn}[\ \rightarrow\ ] = 39999$

Suppose $\boldsymbol{gn}[\ \boldsymbol{A}\ ] = n$ and $\boldsymbol{gn}[\ \boldsymbol{B}\ ] = k$,

then $\boldsymbol{gn}[\ (\boldsymbol{A} \rightarrow \boldsymbol{B})\ ] = 1n39999k2$

- So if sentence with Gödel number k follows in a proof sequence by MP,

then there had to be two previous entries in the sequence with Gödel numbers n and $1n39999k2$.

- Since there are only finitely many previous entries, in principle it's an effective matter to determine whether or not this is the case.

Arithmetizing Proofs

- Let the Gödel number of a proof be the Gödel number of the **total expression** consisting of the sentences of the proof sequence separated by commas, where ***gn***[,] = 29.

  As an example, consider a proof sequence of the form

  $\boldsymbol{A}$, ($\boldsymbol{A} \rightarrow \boldsymbol{B}$), $\boldsymbol{B}$ [suppose $\boldsymbol{A}$ and ($\boldsymbol{A} \rightarrow \boldsymbol{B}$) are axioms]

  Since $\boldsymbol{B}$ is the last formula in the sequence, this is a proof of $\boldsymbol{B}$.

  Again, suppose ***gn***[$\boldsymbol{A}$] = n and ***gn***[$\boldsymbol{B}$] = k

  Then the Gödel number of the proof of $\boldsymbol{B}$ is $n291n39999k229k$

  Let this number $n291n39999k229k$ = $\boldsymbol{j}$

  Then $\boldsymbol{j}$ is the Gödel number of a **proof**

  of the sentence with Gödel number k.

2-Place Proof Relation

- The technical relation **proof** is specified such that:

  proof = { < j, k > : j is the Gödel number of a proof of the sentence with Gödel number k}.

  proof is a recursive relation and hence is definable in ***Q***.

- Let the formula ***Pr*** (x,y) define the relation **proof** in ***Q***.

  So if < j, k > $\in$ **proof** then $\vdash_{Q}$ ***Pr*** (**j**, **k**) and

  if < j, k > $\notin$ **proof** then $\vdash_{Q} \neg$ ***Pr*** (**j**, **k**)

1-Place Proof Predicate

- Now take the formula ***Pr***(x,y) and bind the free variable x with an existential quantifier to get $\exists x$***Pr*** (x,y).
- This formula has only the variable y free, and we will abbreviate

$\exists x$***Pr*** (x,y) as ***Prov*** (y).

Thus ***Prov*** (y) 'asserts', **in the theory *Q***,

that there is a proof in ***Q*** of the sentence with Gödel number y,

and hence that this sentence is a **theorem** of ***Q*** (**!**)

Prov (y) has 3 essential features that will be used to characterize the general notion of a **provability predicate**:

1-Place Proof Predicate

For all sentences $\boldsymbol{A}$, $\boldsymbol{B}$ in the language of $\boldsymbol{Q}$

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

This property follows directly from the fact that $\boldsymbol{Pr}(x,y)$ <u>defines</u> the relation **proof** in $\boldsymbol{Q}$.

For suppose $\vdash_{\boldsymbol{Q}} \boldsymbol{A}$. Then there is a proof in $\boldsymbol{Q}$ of $\boldsymbol{A}$.

Let the Gödel number of the proof be m.

Then $< m, \boldsymbol{gn}[\, \boldsymbol{A} \,] > \ \in$ **proof** and so $\vdash_{\boldsymbol{Q}} \boldsymbol{Pr}(\mathbf{m}, \ulcorner \boldsymbol{A} \urcorner)$

hence $\vdash_{\boldsymbol{Q}} \exists x \boldsymbol{Pr}(x, \ulcorner \boldsymbol{A} \urcorner)$,

i.e. $\vdash \boldsymbol{Prov}(\ulcorner \boldsymbol{A} \urcorner)$

1-Place Proof Predicate

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

This property follows directly from the fact that a proof of $\boldsymbol{B}$ can be obtained from a proof of $\boldsymbol{A} \rightarrow \boldsymbol{B}$ and a proof of $\boldsymbol{A}$ by writing one after the other.

In turn, this can be formalized in $\boldsymbol{Q}$.

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

This property is the **formalization** of (i) in the object language

Thus $\vdash \boldsymbol{A}$ only if $\vdash \mathit{Prov}(\ulcorner \boldsymbol{A} \urcorner)$

/ | | | \

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

1-Place Proof Predicate

Prov (y) has the additional characteristic of ‘correctness’:

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

This property also follows directly from the fact that ***Pr*** (x,y) <u>defines</u> the relation **proof** in ***Q***.

if $\vdash$ ***Prov*** ($\ulcorner \boldsymbol{A} \urcorner$), i.e. $\vdash \exists x$***Pr*** (x, $\ulcorner \boldsymbol{A} \urcorner$),

then this is a true statement in arithmetic, so there is some number m such that m is the Gödel number of a proof of $\boldsymbol{A}$.

thus $\vdash$ $\boldsymbol{A}$

Proof of Gödel's First Incompleteness Theorem

- The diagonal lemma, (i) and (iv) are sufficient to now prove **Gödel's First Incompleteness Theorem**:

  if formal arithmetic is **consistent**, then it is **incomplete**.

- **proof**: Since $\neg$ ***Prov*** (y) is a formula with only the variable y free, it follows by the diagonal lemma that there is a sentence ***S*** such that

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

  Assume $\vdash$ ***S*** .

  Then $\vdash \neg$ ***Prov*** $(\ulcorner \boldsymbol{S} \urcorner)$ by $(*''')$

  and $\vdash$ ***Prov*** $(\ulcorner \boldsymbol{S} \urcorner)$ by reductio hypothesis and (i)

  and arithmetic is inconsistent.

Proof of Gödel's First Incompleteness Theorem

Assume $\vdash \neg \boldsymbol{S}$.

By (*´´´) alone we get $\vdash \neg \boldsymbol{Prov}(\ulcorner \boldsymbol{S} \urcorner) \rightarrow \boldsymbol{S}$

and contraposition and double neg. elim. on this yield

$\vdash \neg \boldsymbol{S} \rightarrow \boldsymbol{Prov}(\ulcorner \boldsymbol{S} \urcorner)$.

With the reductio hypothesis $\vdash \neg \boldsymbol{S}$, MP, and distribution of $\vdash$ over the conditional we get $\vdash \boldsymbol{Prov}(\ulcorner \boldsymbol{S} \urcorner)$,

and (iv) yields $\vdash \boldsymbol{S}$.

So $\vdash \boldsymbol{S}$ and $\vdash \neg \boldsymbol{S}$ and arithmetic is **inconsistent**.

Thus ***if*** formal arithmetic is consistent, ***then***

neither $\vdash \boldsymbol{S}$ *nor* $\vdash \neg \boldsymbol{S}$, and formal arithmetic is **incomplete** ■

Consequences for Hilbert's Program

- A <u>fundamental</u> <u>wedge</u> has thereby been driven between **truth** in the intended model and **provability** in a formal system.
- So Gödel's First Incompleteness Theorem is generally taken to refute one of the basic tenets of Hilbert's Program

  by establishing that **not** all of the **true** statements in elementary arithmetic can be **proved** in an axiomatizable formal theory.
- All statements in the language of arithmetic are either true or false in the *intended model*, but neither ***S*** nor ¬ ***S*** is provable.
- Hence even elementary arithmetic cannot be reduced to

  "… an inventory of provable formulas"

Consequences for Hilbert's Program

- So it would appear that the method of axiomatization and finitary proof is inherently too weak to capture all mathematical truths.
- And this type of 'foundation' thereby seems to be rendered inadequate, *in principle*.
- However, the areas in which such 'undecidable' sentences (i.e. where neither $\boldsymbol{A}$ nor $\neg\boldsymbol{A}$ is provable) actually arise appears to be quite narrow and specialized.
- No arithmetical conjecture or problem that has occurred to mathematicians in contexts, <u>outside</u> of logic and the foundations of mathematics,

has ever been proved to be undecidable.

Consequences for Hilbert’s Program

- But there has been work to find statements closer to ‘ordinary mathematics’ than Gödel sentences that *are* undecidable in PA.

  e.g. **The Paris-Harrington Theorem** (1976)

Consequences for Philosophy of Mind?

- Since $\boldsymbol{S} \leftrightarrow \neg\, \boldsymbol{Prov}\,(\ulcorner \boldsymbol{S} \urcorner)$, the ‘Gödel sentence’ ***S*** can be interpreted as ‘asserting its own unprovability’,

  and if arithmetic is consistent then ***S*** *is* unprovable, hence **true.**
- The human mind seems able to intuitively grasp the truth of the Gödel sentence,

  even though the sentence does not follow as a consequence of the finitary deductive system.
- Does this show that the human mind cannot be reduced to a finitary deductive system?
- And given the relationship between computation and finitary deductive systems, does this show that

  the Computational Theory of Mind is false??

Consequences for Philosophy of Mind?

- Both Lucas, and more recently Penrose, have put forward arguments to this effect.
- But are such arguments convincing?
- Gödel's theorem establishes that if the formal theory of arithmetic is consistent, then ***S*** is not provable.
- Thus in order for the human mind to 'know' that ***S*** is **not provable** and hence **true**,

  the human mind must first 'know' that arithmetic is consistent.
- ***Q*** is quite a simple theory,

  so suppose we grant the claim that we *know* that
  Q is consistent and that ***S*** is **unprovable** and thus **true**.

Consequences for Philosophy of Mind?

- Indeed, *whenever* we know a formal theory **T** (in which **diag** is representable) to be consistent, we also know the truth of a corresponding sentence ***S*** that is not provable in **T**.
- Does this show that the human mind has a special power and can thereby outperform any given formal system?
- **No** –
- While it *does* follow that we know the conditional
- 'If the formal system **T** is consistent then the corresponding sentence ***S*** is true'

this is <u>not</u> equivalent to the claim

- 'If the formal system **T** is consistent then *we know* that the corresponding sentence ***S*** is true'

Consequences for Philosophy of Mind?

- Thus consider cases where the formal system is so complex that we have no idea whether or not it is consistent.
- Then we also have no idea whether or not the corresponding sentence ***S*** is true**!**
- So what strictly follows from the claim that *we* know the simple system ***Q*** is consistent, and hence that the original Gödel sentence *is* true**?**
- Only that, *if* some version of CTM is the case, *then* the formal system on which the human mind operates isn't ***Q***!
- But the possibility still remains open that it could be some other, much more sophisticated and elaborate formal system (of which we are unaware).

Inexhaustability

- While these standard attempts to derive anti-CTM consequences from Gödel's Theorem are not successful,

  Gödel himself put forward a potential argument based on the seeming 'inexhaustability' of human mathematical knowledge:

  it may not be possible to specify **any one** formal system which completely exhausts <u>all</u> of our mathematical knowledge.
- If this is true, then perhaps human mathematical knowledge cannot ultimately be captured in terms of

  computations and finitary deductive mechanisms**?**

Proving Gödel's First Theorem the Fast Way

- We have stated and proved Gödel's First Incompleteness Theorem in the 'classic' manner,

by constructing a sentence ***S*** that 'asserts its own unprovability'.

- However, since the time of Gödel's original Theorem, more streamlined methods have been devised for establishing a theoretically equivalent result:
- A formal theory **T** is **axiomatizable** iff there is a **decidable** subset of **T** whose logical consequences are the theorems of **T**.
- So any **decidable** theory **is** **axiomatizable** (why?),

but not every **axiomatizable** theory is **decidable** (why?).

Proving Gödel's First Theorem the Fast Way

- As above, a formal theory **T** is **complete** iff
  for all sentences ***S*** in the language of **T**, either $\vdash_{\mathbf{T}}$ ***S*** or $\vdash_{\mathbf{T}} \neg$ ***S***.
- Furthermore,
  (**i**) any **axiomatizable**, **complete** theory is **decidable**
  (Theorem 5, B&J p. 177).
- Recall the previous result that
  (**ii**) **no consistent** extension of ***Q*** is **decidable**.
- It follows as an immediate consequence of (**i**) and (**ii**) that:
  Gödel's First Incompleteness Theorem (reformulated)
  There is **no consistent**, **complete**, and **axiomatizable** extension of ***Q*** ■