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

So a theory **T** is decidable iff its set of theorems is a recursive set.

- Connection between the 2 notions:

  if a set of expressions Θ is decidable then its characteristic function is recursive and hence is representable in $\boldsymbol{Q}$,

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

  This is because if the characteristic function f_Θ of Θ is recursive and $\boldsymbol{A}_{f\Theta}(x,y)$ represents f_Θ in $\boldsymbol{Q}$,

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

Definability and Decidability

- **Lemma**: if **T** is a consistent extension of ***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 ***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

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

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). 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 $\vdash_{\mathbf{T}}$ $\boldsymbol{G}$ **and** $\vdash_{\mathbf{T}} \neg\, \boldsymbol{G}$, rendering **T** inconsistent, contrary to initial hypothesis. Conclusion: *there can be no such* $\boldsymbol{C}(y)$ □

- Bigger conclusion: **no consistent** extension of $\boldsymbol{Q}$ is **decidable**.

Undecidability of FOL (from a different angle)

- **Church's Theorem**: FOL is undecidable.
- **proof**: we have just established that $\boldsymbol{Q}$ 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 $\boldsymbol{Q}$.

Then a sentence $\boldsymbol{S}$ is a theorem of $\boldsymbol{Q}$ iff the conditional

$\Phi \rightarrow \boldsymbol{S}$ is a theorem of FOL.

In other words

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

Hence (intuitively) *if* FOL were decidable *then* so would $\boldsymbol{Q}$ be.

Undecidability of FOL (from a different angle)

- 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 $(\Phi \rightarrow \boldsymbol{S})$

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

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

  But $\{n: f(n) \in \Theta\}$ is the set of Gödel numbers of theorems of $\boldsymbol{Q}$, which has just been shown not to be decidable.

- Thus Θ is not recursive and FOL is not decidable □

Indefinability of Arithmetical Truth

- **Tarski's Theorem**: the set of Gödel numbers of true sentences in 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

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

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

So *if* ***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*

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

for all sentences ***S*** of ***L***,

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

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 ***L***.

Indefinability of ‘True-in-L’ in L

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

$(*'') \quad \vdash_{L} \boldsymbol{G} \leftrightarrow \neg\, \boldsymbol{Tr}(\ulcorner \boldsymbol{G} \urcorner)$ 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 **axiomatizable** iff there is a **decidable** subset of **T** whose logical consequences are the theorems of **T**.
- 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 **<u>in</u>complete** 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 <u>neither</u> ***S*** nor ¬ ***S*** is provable

  if the formal theory of arithmetic is <u>consistent</u>.

  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 premises $\boldsymbol{B}_1, \dots, \boldsymbol{B}_n$

is a finite sequence of formulas,

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

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

and where each $\boldsymbol{F}_1, \boldsymbol{F}_2, \dots,$ 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

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

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 axiomatizable extensions),

  where FOL can be formalized in terms of 7 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) $\quad \mathbf{A \rightarrow (B \rightarrow A)}$

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

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

Any *instance* of a schema is a **logical axiom**

Rule of Inference: Modus Ponens

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

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

Formal Axiomatic Proofs

Here is an axiomatic demonstration for

$$\vdash P \rightarrow P \quad \text{(so no premises involved)}$$

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 an axiomatic system for FOL we then add the 7 (non-logical) axioms of $\boldsymbol{Q}$. Hence if sentence $\boldsymbol{C}$ is a theorem of $\boldsymbol{Q}$, 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 $\boldsymbol{C}$, 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 $\boldsymbol{F}_{\boldsymbol{n}}$ and $\boldsymbol{F}_{\boldsymbol{m}}$ using MP.
- The set of (Gödel numbers of) axioms of FOL plus axioms of $\boldsymbol{Q}$ is **recursive**

Arithmetizing Proofs

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

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

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

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 <u>finitely</u> <u>many</u> previous entries, it's an effective matter to check.

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]

Its Gödel number is $n291n39999k229k$

Let $n291n39999k229k = j$

Then j is the Gödel number of a proof of the sentence with Gödel number k.

2-Place Proof Relation

- In general, the 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 > ∈ **proof** then $\vdash_{Q}$ ***Pr*** (**j**, **k**) and

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

1-Place Proof Predicate

Thus ***Prov*** (y) ‘asserts that’ there exists 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 proof predicate:

For all sentences ***A***, ***B*** in the language of ***Q***

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

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

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

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

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

Proof of Gödel's First Incompleteness Theorem

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

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

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

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

Assume $\vdash \boldsymbol{S}$. Then $\vdash \neg \textbf{\textit{Prov}}\,(\ulcorner \boldsymbol{S} \urcorner)$ by $(*''')$,

and $\vdash \textbf{\textit{Prov}}\,(\ulcorner \boldsymbol{S} \urcorner)$ 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** □

- A fundamental wedge has thereby been driven between **truth** in the intended model and **provability** in a formal system.

Consequences for Hilbert's Program

- 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"
- So it would appear that the method of axiomatization and finitary proof is inherently too weak to capture mathematics, and this type of 'foundation' is rendered inadequate.

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??

Proving Gödel's First Theorem the Fast Way

- 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?).
- 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**.

Proving Gödel's First Theorem the Fast Way

- Hence 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*** □