The Axiomatic Approach to Propositional Logic

Roadmap of this supplementary chapter

In Chapter 1, we defined propositional logic **semantically (语义地)**: we fixed two truth values $\{0, 1\}$, gave each connective a meaning by a truth table, and called a formula *valid* if it is true under every assignment. In this supplementary chapter, we look at the other side, which is how mathematicians formalize *proofs*: we define logic **syntactically (语法地)**, as a game of manipulating symbols by fixed rules, *without ever mentioning "true" or "false"*. The guiding question is:

Is everything provable true? Is everything true provable?

We will proceed in five steps:

1. **Warm-up:** what a formal system is, and why some questions about it can only be answered from outside (the MIU-puzzle, §1).
2. **Syntax:** the Hilbert system $\mathcal{H}$: axioms, Modus Ponens, derivations, and the difference between $\to$ and $\vdash$ (§2).
3. **The bridge:** soundness, consistency and completeness, which connect "provable" ($\vdash$) with "true" ($\models$) (§3).
4. **Playing the game:** formal derivations, the Deduction Theorem, and the familiar laws of logic rebuilt inside $\mathcal{H}$ (§4–§6).
5. **Closing the loop:** why $\mathcal{H}$ is complete, and what the axiomatic view teaches us beyond truth tables (§7–§8).

Exercises are in §9. *Prerequisite:* Chapter 1 up to proof by resolution.

1 A Toy Formal System: the MIU-System

▶ Before logic, let us play a game from D. Hofstadter's book *Gödel, Escher, Bach* (GEB). A **formal system (形式系统)** consists of an alphabet, some initial strings called *axioms*, and some *rules* for producing new strings from old ones.

Definition: The MIU-System

- **Symbols:** `M`, `I`, `U`. We work with finite strings over them, e.g., `MIUIU`.
- **Axiom:** `MI`.
- **Rules:** below, x and y stand for *arbitrary* (possibly empty) strings.

[R1] If $\vdash x\texttt{I}$, then $\vdash x\texttt{IU}$ (add U after a final I)
[R2] If $\vdash \texttt{M}x$, then $\vdash \texttt{M}xx$ (double everything after M)
[R3] If $\vdash x\texttt{III}y$, then $\vdash x\texttt{U}y$ (replace III by U)
[R4] If $\vdash x\texttt{UU}y$, then $\vdash xy$ (delete UU)

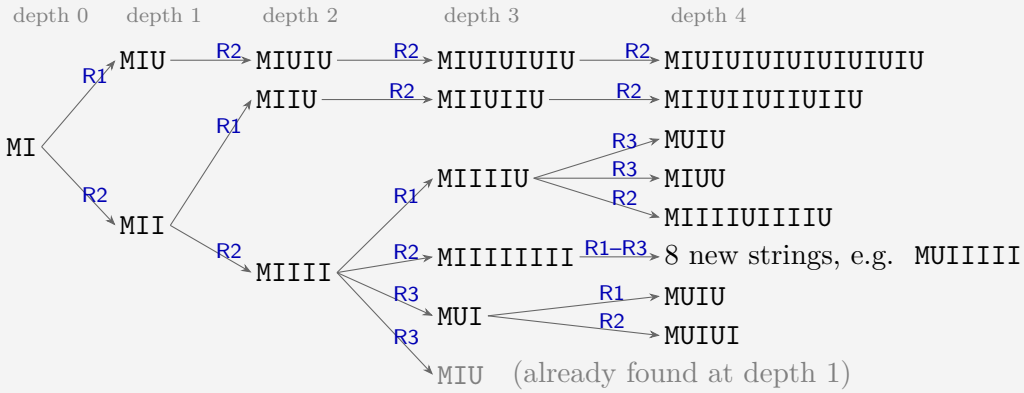
A string s is a **theorem (定理)**, written $\vdash s$, if it can be obtained from the axiom by finitely many applications of the rules.

- The symbols *mean nothing*. The rules only look at the *shape* of strings (they are *typographical*). For example, here is a **derivation (推导)** showing $\vdash$ MUIU:

$$\texttt{MI} \xrightarrow{\text{R2}} \texttt{MII} \xrightarrow{\text{R2}} \texttt{MIIII} \xrightarrow{\text{R1}} \texttt{MIIIIU} \xrightarrow{\text{R3}} \texttt{MUIU}.$$

  Note that a rule may apply in several ways: applying R3 to the *last* three I's of MIIIIU gives MIUU, and then R4 gives back MI.

- **The MU-puzzle.** *Is MU a theorem?* A natural attempt is to search for a derivation systematically. The following tree shows (part of) all strings derivable within four steps.



  The tree grows explosively and MU never shows up. But "not found yet" does not mean "not derivable": maybe MU needs 100 steps, going through very long strings (rules R3, R4 can shorten strings). Working *inside* the system, we can never conclude a negative answer.

- **Stepping outside the system: the 310-system.** Hofstadter's trick is to replace each symbol by a digit

$$\texttt{M} \mapsto 3, \qquad \texttt{I} \mapsto 1, \qquad \texttt{U} \mapsto 0,$$

  so that every string becomes a natural number written in decimal, e.g., MI $\mapsto$ 31, MIIU $\mapsto$ 3110 and MU $\mapsto$ 30. The rules R1–R4 now act on numbers instead of strings: we apply exactly the same operations to the digits. This is an **isomorphism (同构)**: the axiom MI becomes the "axiom number" 31, and MU is a theorem iff 30 can be produced from 31. So the MU-puzzle becomes a question about numbers, where we have tools that the MIU-system does not have.

Theorem: `MU` is NOT a theorem of the MIU-system

Every number produced from 31 is *not* divisible by 3. Since 30 is divisible by 3, `MU` is not a theorem.

Why: Recall that a number is divisible by 3 iff its digit sum is. We start from 31, whose digit sum is 4, not divisible by 3. Now look at what each rule does to the digit sum:

- R1 appends a 0, and R4 deletes 00: the digit sum does not change.
- R3 replaces 111 by 0: the digit sum decreases by 3, so divisibility by 3 does not change.
- R2 doubles everything after the leading 3: if that part has digit sum s, the total digit sum changes from $3+s$ to $3+2s$. If $3 \nmid s$, then also $3 \nmid 2s$.

So no rule can ever produce a multiple of 3 from a non-multiple, and 30 is never reached. Therefore, the number of `I`'s in a theorem is never a multiple of 3, whereas `MU` has none.

▶ **Three morals of the MU-puzzle**, which will guide the rest of these notes:

① **Inside vs. outside.** Deriving `MUIU` is a job *inside* the system (just apply rules). Showing that `MU` is *not* derivable requires reasoning *about* the system from outside. Such a statement is called a **metatheorem (元定理)**.

② **The invariant method.** To show something is not derivable, find a property that (i) every axiom has, and (ii) every rule preserves, but (iii) the target does not have. For the MIU-system, the property is "$3 \nmid \#\texttt{I}$". For propositional logic, the property will be *"being a tautology"*, and this is exactly the *soundness theorem* below.

③ **Interpretation.** The rules of MIU mean nothing, but by *interpreting* strings as numbers we could use arithmetic to reason about them. Similarly, the formulas of a logical system mean nothing, but by *interpreting* them via truth tables we can reason about provability. (Also, turning strings into numbers is exactly how Gödel made arithmetic "talk about itself", which led to his incompleteness theorem.)

▶ **Remark:** One can further show that the theorems of MIU are *exactly* the strings `M`x, where x consists of `I`'s and `U`'s and the number of `I`'s in x is not a multiple of 3 (a nice challenge). So for MIU, the "semantic" description (invariant) and the "syntactic" description (derivable) coincide. We will see the same phenomenon for logic: this is the *completeness theorem*.

2 The Hilbert System for Propositional Logic

▶ Now we build a formal system whose theorems are supposed to be the "logical truths". There are many choices; we use the classical **Hilbert-style system (希尔伯特系统)** $\mathcal{H}$ (due to Frege, Hilbert and Łukasiewicz).

Definition: The Hilbert System $\mathcal{H}$

- **Symbols:** variables $P, Q, R, P_1, P_2, \ldots$; connectives $\neg$ and $\to$; parentheses.
- **Formulas:** the WFFs built from variables using only $\neg$ and $\to$.
- **Axioms:** for arbitrary formulas α, β, γ, every formula of the following forms is an axiom:

  (A1) $\alpha \to (\beta \to \alpha)$

  (A2) $\big(\alpha \to (\beta \to \gamma)\big) \to \big((\alpha \to \beta) \to (\alpha \to \gamma)\big)$

  (A3) $(\neg\beta \to \neg\alpha) \to (\alpha \to \beta)$

- **Rule:** **Modus Ponens (MP, 分离规则)**: from α and $\alpha \to \beta$, infer β.

▶ Some remarks on the definition:

- **Axiom schemas (公理模式).** Here α, β, γ are *metavariables*, playing the same role as x, y in the MIU rules. Each of (A1)–(A3) is a template that stands for infinitely many axioms, called its *instances*. For example, taking $\alpha = P$ and $\beta = \neg Q \to R$ in (A1) gives the axiom $P \to ((\neg Q \to R) \to P)$.
- **Only $\neg$ and $\to$.** Other operators are viewed as *abbreviations*: $\alpha \vee \beta := \neg\alpha \to \beta$, $\alpha \wedge \beta := \neg(\alpha \to \neg\beta)$, $\alpha \leftrightarrow \beta := (\alpha \to \beta) \wedge (\beta \to \alpha)$. Since $\{\neg, \to\}$ is functionally complete, nothing is lost, and a small language makes the metatheory (proofs *about* $\mathcal{H}$) much shorter.
- **The comparison with MIU:**

	MIU-system	Hilbert system $\mathcal{H}$
objects	strings over `M`,`I`,`U`	formulas over variables, $\neg, \to, (,)$
axioms	one axiom `MI`	three axiom schemas (A1)–(A3)
rules	R1–R4	only MP
meaning?	none	none (!), until we interpret it

▶ In practice, we want to prove things from assumptions. This leads to the notion of **derivation from hypotheses**.

Definition: Derivation, Inference and Theorem

Let Γ be a set of formulas (the *hypotheses*, 前提). A **derivation (推导)** of β from Γ is a finite sequence of formulas $\varphi_1, \varphi_2, \ldots, \varphi_m = \beta$ such that each φ_i is

(i) an instance of (A1), (A2) or (A3); or

(ii) a hypothesis, i.e., $\varphi_i \in \Gamma$; or

(iii) obtained by MP from φ_j and $\varphi_k = (\varphi_j \to \varphi_i)$ for some $j, k < i$.

We write $\Gamma \vdash \beta$ ("β is *derivable/inferable* from Γ") if such a derivation exists. We also write $\alpha_1, \ldots, \alpha_n \vdash \beta$ for $\{\alpha_1, \ldots, \alpha_n\} \vdash \beta$. When $\Gamma = \emptyset$, we write $\vdash \beta$ and call β a **theorem (定理)** of $\mathcal{H}$.

- **Object language vs. metalanguage.** Be careful with the different arrows:
  - $\to$ is a symbol *inside* $\mathcal{H}$; $\alpha \to \beta$ is just a formula (a string).
  - $\vdash$ is *not* a symbol of $\mathcal{H}$; "$\Gamma \vdash \beta$" is a statement *about* $\mathcal{H}$ ("a derivation exists"), which is either correct or not, like "`MU` is a MIU-theorem".
  - Words like "if…then", "iff", $\Rightarrow$, and proof techniques like induction, belong to the *metalanguage*: the ordinary mathematics we use to talk about $\mathcal{H}$.

- The following structural properties follow directly from the definition of derivations (think about why):

Lemma: Structural Properties of $\vdash$

① (Reflexivity) If $\alpha \in \Gamma$, then $\Gamma \vdash \alpha$.
② (Monotonicity) If $\Gamma \subseteq \Delta$ and $\Gamma \vdash \alpha$, then $\Delta \vdash \alpha$.
③ (Closure under MP) If $\Gamma \vdash \alpha$ and $\Gamma \vdash \alpha \to \beta$, then $\Gamma \vdash \beta$.
④ (Cut) If $\Gamma \vdash \alpha$ and $\Gamma \cup \{\alpha\} \vdash \beta$, then $\Gamma \vdash \beta$.
⑤ (Finiteness) If $\Gamma \vdash \alpha$, then $\Gamma_0 \vdash \alpha$ for some *finite* $\Gamma_0 \subseteq \Gamma$.

Proof: For ③, concatenate the two derivations and append β (by MP). For ④, concatenate the derivation of α from Γ with the derivation of β from $\Gamma \cup \{\alpha\}$; every use of the hypothesis α in the second one is now justified by the first one. For ⑤, a derivation is finite, so it uses only finitely many hypotheses.

In particular, ④ says that once we have proved a theorem, we may freely use it as a line in later derivations (as a *lemma*). Moreover, because (A1)–(A3) are schemas, every theorem we prove below with letters A, B, C holds for *all* formulas substituted for A, B, C.

3 Soundness, Consistency and Completeness

- So far, $\mathcal{H}$ is just a symbol game. Now we bring back truth tables and compare. Recall the semantic notion from the lecture, extended to a set of hypotheses:

Definition: Semantic Consequence

We write $\Gamma \models \alpha$ ("Γ *semantically implies* α") if every assignment that makes all formulas in Γ true also makes α true. In particular, $\models \alpha$ means α is a tautology. (This is the "$\Rightarrow$" of the lecture.)

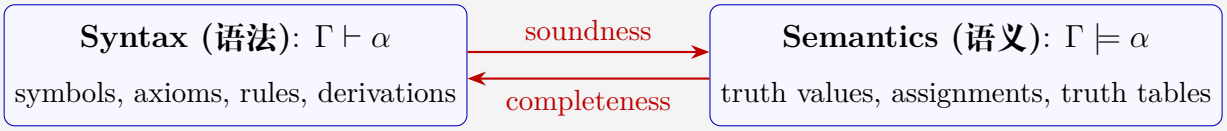
So we have two relations: $\vdash$ (*provable*, defined by symbols) and $\models$ (*true*, defined by meanings). The following three properties describe how a proof system relates to truth.

Definition: Soundness, Consistency, Completeness

A proof system (like $\mathcal{H}$) is said to be

- **sound (可靠的)** if everything provable is true: $\Gamma \vdash \alpha \Rightarrow \Gamma \models \alpha$;
- **consistent (一致的/无矛盾的)** if there is no formula α such that both $\vdash \alpha$ and $\vdash \neg\alpha$;
- **complete (完备的)** if everything true is provable: $\Gamma \models \alpha \Rightarrow \Gamma \vdash \alpha$.

The special cases with $\Gamma = \emptyset$ ($\vdash \alpha \Rightarrow \models \alpha$, and $\models \alpha \Rightarrow \vdash \alpha$) are called *weak* soundness/completeness.



- Note that consistency is a *purely syntactic* property: it does not mention truth at all. It is the minimal requirement for a proof system to be meaningful (we will see below that an inconsistent system proves *everything*). Soundness is exactly the "invariant method" of the MU-puzzle, with the invariant "being true".

Theorem: Soundness of $\mathcal{H}$

If $\Gamma \vdash \alpha$, then $\Gamma \models \alpha$. In particular, every theorem of $\mathcal{H}$ is a tautology.

Proof: *Step 1: every axiom is a tautology.* We argue by contradiction, trying to make each axiom false.

- (A1): $\alpha \to (\beta \to \alpha)$ is false only if $\alpha = 1$ and $\beta \to \alpha = 0$; the latter needs $\alpha = 0$. Contradiction.
- (A2): suppose it is false. Then $\alpha \to (\beta \to \gamma) = 1$ and $(\alpha \to \beta) \to (\alpha \to \gamma) = 0$, so $\alpha \to \beta = 1$ and $\alpha \to \gamma = 0$, i.e., $\alpha = 1, \gamma = 0$. Then $\beta = 1$ (from $\alpha \to \beta = 1$), so $\beta \to \gamma = 0$ and $\alpha \to (\beta \to \gamma) = 0$. Contradiction.
- (A3): suppose it is false. Then $\neg\beta \to \neg\alpha = 1$, $\alpha = 1$ and $\beta = 0$. So $\neg\beta = 1$, hence $\neg\alpha = 1$, i.e., $\alpha = 0$. Contradiction.

Step 2: MP preserves truth. Under any assignment, if α and $\alpha \to \beta$ are both true, then β is true (truth table of $\to$).
Step 3: induction. Fix an assignment I making all of Γ true, and let $\varphi_1, \ldots, \varphi_m$ be a derivation from Γ. By induction on i, each φ_i is true under I: axioms by Step 1,

hypotheses by the choice of I, and MP-lines by Step 2 and the induction hypothesis. Hence $\varphi_m = \alpha$ is true under I.

Corollary: Consistency of $\mathcal{H}$

$\mathcal{H}$ is consistent. For example, $\neg(P \to P)$ is not a theorem of $\mathcal{H}$.

Proof: If $\vdash \alpha$ and $\vdash \neg\alpha$, then by soundness both α and $\neg\alpha$ are tautologies, which is impossible. Also, $\neg(P \to P)$ is a contradiction, not a tautology, so it cannot be a theorem.

Compare this with the MU-puzzle: "$\nvdash \neg(P \to P)$" is proved exactly like "$\nvdash$ `MU`": by an invariant that every theorem has, but the target lacks.

- The converse direction is much harder and is the central result about $\mathcal{H}$. We state it now and sketch its proof in Section 7.

Theorem: Completeness of $\mathcal{H}$ (Post 1921; strong form)

If $\Gamma \models \alpha$, then $\Gamma \vdash \alpha$. In particular, every tautology is a theorem of $\mathcal{H}$.

Together: $\Gamma \vdash \alpha$ iff $\Gamma \models \alpha$. Provability and truth coincide.

4 Proving Theorems Inside $\mathcal{H}$

- Now let us actually play the game. Each line of a derivation must be justified. For an axiom, we state which schema is used and *what is substituted* for α, β, γ. For MP, we give the line numbers of φ and $\varphi \to \psi$. Below, A, B, C denote arbitrary formulas.

Theorem 1 (Identity): $\vdash A \to A$

(1)	$\big(A \to ((A \to A) \to A)\big) \to \big((A \to (A \to A)) \to (A \to A)\big)$	A2 [α=A, β=$A\to A$, γ=A]
(2)	$A \to ((A \to A) \to A)$	A1 [α=A, β=$A\to A$]
(3)	$(A \to (A \to A)) \to (A \to A)$	MP (2),(1)
(4)	$A \to (A \to A)$	A1 [α=A, β=A]
(5)	$A \to A$	MP (4),(3)

- Checking such a derivation is completely mechanical, but *finding* it requires some creativity. Here is how one can discover it by working *backwards*:
  - The only rule is MP, so $A \to A$ must come from some X and $X \to (A \to A)$.
  - The only axiom with a conclusion of the shape $\ldots \to (\alpha \to \gamma)$ is (A2). With $\alpha = \gamma = A$, (A2) says: from $A \to (\beta \to A)$ and $A \to \beta$ we get $A \to A$.

– $A \to (\beta \to A)$ is an instance of (A1) for *any* β. So we only need some β for which $A \to \beta$ is also easy: $\beta = A \to A$ works, since $A \to (A \to A)$ is an instance of (A1).

Five steps to prove the most obvious "truth" $A \to A$, which a truth table checks in two rows! This is the price of using so few axioms and rules.

Theorem 2 (Hypothetical Syllogism, 假言三段论): $A \to B,\ B \to C \vdash A \to C$

(1)	$B \to C$	hypothesis
(2)	$(B \to C) \to (A \to (B \to C))$	A1 [α=$B \to C$, β=A]
(3)	$A \to (B \to C)$	MP (1),(2)
(4)	$(A \to (B \to C)) \to ((A \to B) \to (A \to C))$	A2 [α=A, β=B, γ=C]
(5)	$(A \to B) \to (A \to C)$	MP (3),(4)
(6)	$A \to B$	hypothesis
(7)	$A \to C$	MP (6),(5)

▶ **Derived rule HS.** Combining Theorem 2 with the cut property, we get: *if* $\Gamma \vdash A \to B$ *and* $\Gamma \vdash B \to C$, *then* $\Gamma \vdash A \to C$. In later derivations, we write "HS $(i), (j)$" for this. It is only a shorthand: we could always replace it by the 7 lines above.

Theorem 3 (Double Negation Elimination): $\vdash \neg\neg A \to A$

(1)	$\neg\neg A \to (\neg\neg\neg\neg A \to \neg\neg A)$	A1 [α=$\neg\neg A$, β=$\neg\neg\neg\neg A$]
(2)	$(\neg\neg\neg\neg A \to \neg\neg A) \to (\neg A \to \neg\neg\neg A)$	A3 [α=$\neg A$, β=$\neg\neg\neg A$]
(3)	$\neg\neg A \to (\neg A \to \neg\neg\neg A)$	HS (1),(2)
(4)	$(\neg A \to \neg\neg\neg A) \to (\neg\neg A \to A)$	A3 [α=$\neg\neg A$, β=A]
(5)	$\neg\neg A \to (\neg\neg A \to A)$	HS (3),(4)
(6)	$\big(\neg\neg A \to (\neg\neg A \to A)\big) \to \big((\neg\neg A \to \neg\neg A) \to (\neg\neg A \to A)\big)$	A2 [α=β=$\neg\neg A$, γ=A]
(7)	$(\neg\neg A \to \neg\neg A) \to (\neg\neg A \to A)$	MP (5),(6)
(8)	$\neg\neg A \to \neg\neg A$	Theorem 1 [$A := \neg\neg A$]
(9)	$\neg\neg A \to A$	MP (8),(7)

▶ Some observations from Theorem 3:

– Line (2) is a funny instance of (A3): to use (A3) we need a formula of the form $\neg\beta \to \neg\alpha$, so we artificially *added* two more negations to $\neg\neg A$ in line (1).

– Lines (6)–(9) show a useful trick (*contraction*): from $X \to (X \to Y)$ we get $X \to Y$ using (A2) and Theorem 1.

– This is the first theorem that uses (A3), the only axiom about $\neg$. Indeed, one can show that $\neg\neg A \to A$ *cannot* be derived from (A1) and (A2) alone (see the exercises). Double negation elimination is where *classical* logic enters $\mathcal{H}$.

Writing derivations like this quickly becomes painful. What we need is a better way to *reason about* derivations.

5 Metatheorems: the Deduction Theorem

▶ A **metatheorem (元定理)** is a theorem *about* the system $\mathcal{H}$, proved by ordinary mathematical reasoning (like induction) outside of $\mathcal{H}$. We have already seen some: "`MU` is not a theorem of MIU", soundness, and the structural properties of $\vdash$. The most useful metatheorem for constructing derivations is the following one, which formalizes the familiar strategy "*to prove $A \to B$, assume A and prove B*".

Theorem: Deduction Theorem (演绎定理, Herbrand 1930)

For any set Γ and formulas α, β:

$$\Gamma \cup \{\alpha\} \vdash \beta \quad \text{iff} \quad \Gamma \vdash \alpha \to \beta.$$

Proof: "$\Leftarrow$" Take a derivation of $\alpha \to \beta$ from Γ, append α (a hypothesis) and then β (MP).

"$\Rightarrow$" Let $\varphi_1, \ldots, \varphi_m = \beta$ be a derivation from $\Gamma \cup \{\alpha\}$. We prove, by strong induction on i, that $\Gamma \vdash \alpha \to \varphi_i$ for every i. There are three cases for φ_i:

– *φ_i is an axiom or $\varphi_i \in \Gamma$.* Then $\Gamma \vdash \varphi_i$, and $\varphi_i \to (\alpha \to \varphi_i)$ is an instance of (A1), so by MP, $\Gamma \vdash \alpha \to \varphi_i$.

– $\varphi_i = \alpha$. Then $\alpha \to \varphi_i$ is $\alpha \to \alpha$, which is a theorem by Theorem 1.

– *φ_i is obtained by MP* from φ_j and $\varphi_k = \varphi_j \to \varphi_i$ with $j, k < i$. By the induction hypothesis, $\Gamma \vdash \alpha \to \varphi_j$ and $\Gamma \vdash \alpha \to (\varphi_j \to \varphi_i)$. Using the instance of (A2)

$$\big(\alpha \to (\varphi_j \to \varphi_i)\big) \to \big((\alpha \to \varphi_j) \to (\alpha \to \varphi_i)\big)$$

and MP twice, we get $\Gamma \vdash \alpha \to \varphi_i$.

Taking $i = m$ gives $\Gamma \vdash \alpha \to \beta$.

▶ **Remarks.**

– Look at which axioms the proof used: (A1) for axioms and hypotheses, Theorem 1 (which itself only uses (A1) and (A2)) for α itself, and (A2) for MP. So (A1) and (A2) are *exactly* what is needed to make the Deduction Theorem true! (A2) is literally "MP under an assumption α".

- The proof is *constructive*: it transforms a derivation of β from $\Gamma \cup \{\alpha\}$ with m lines into a derivation of $\alpha \to \beta$ from Γ with at most $5m$ lines.
- The "rule of conditional proof" used in the lecture is justified exactly by this theorem. The Deduction Theorem is not a new rule of $\mathcal{H}$; it only says that such a rule would not prove anything new (it is *admissible*).
- Do not confuse: the Deduction Theorem is about $\vdash$, while "$\alpha \Rightarrow \beta$ iff $\alpha \to \beta$ is a tautology" from the lecture is the corresponding fact about $\models$.

▶ **The Deduction Theorem in action.** Proofs become much shorter.

Example: Theorems 2 and 3 revisited

Theorem 2. We first show $A \to B,\ B \to C,\ A \vdash C$:

(1)	A	hypothesis
(2)	$A \to B$	hypothesis
(3)	B	MP (1),(2)
(4)	$B \to C$	hypothesis
(5)	C	MP (3),(4)

By the Deduction Theorem, $A \to B,\ B \to C \vdash A \to C$. Applying it twice more gives the theorem $\vdash (A \to B) \to ((B \to C) \to (A \to C))$.

Theorem 3. We first show $\neg\neg A \vdash A$:

(1)	$\neg\neg A$	hypothesis
(2)	$\neg\neg A \to (\neg\neg\neg\neg A \to \neg\neg A)$	A1 [α=$\neg\neg A$, β=$\neg\neg\neg\neg A$]
(3)	$\neg\neg\neg\neg A \to \neg\neg A$	MP (1),(2)
(4)	$(\neg\neg\neg\neg A \to \neg\neg A) \to (\neg A \to \neg\neg\neg A)$	A3 [α=$\neg A$, β=$\neg\neg\neg A$]
(5)	$\neg A \to \neg\neg\neg A$	MP (3),(4)
(6)	$(\neg A \to \neg\neg\neg A) \to (\neg\neg A \to A)$	A3 [α=$\neg\neg A$, β=A]
(7)	$\neg\neg A \to A$	MP (5),(6)
(8)	A	MP (1),(7)

By the Deduction Theorem, $\vdash \neg\neg A \to A$. No contraction trick is needed.

6 More Theorems: Negation and Contradiction

- With the Deduction Theorem, we can quickly build up the familiar laws of logic.

Theorem 4 (Ex Falso Quodlibet, 爆炸原理): $A,\ \neg A \vdash B$, **hence** $\vdash \neg A \to (A \to B)$

(1)	$\neg A$	hypothesis
(2)	$\neg A \to (\neg B \to \neg A)$	A1 $[\alpha{=}\neg A,\ \beta{=}\neg B]$
(3)	$\neg B \to \neg A$	MP (1),(2)
(4)	$(\neg B \to \neg A) \to (A \to B)$	A3 $[\alpha{=}A,\ \beta{=}B]$
(5)	$A \to B$	MP (3),(4)
(6)	A	hypothesis
(7)	B	MP (6),(5)

So $\neg A, A \vdash B$. Applying the Deduction Theorem twice gives $\vdash \neg A \to (A \to B)$.

- Theorem 4 says that "*from a contradiction, anything follows*". It explains why consistency is so important:

Corollary: Consistency Means Something Is Unprovable

$\mathcal{H}$ is inconsistent iff every formula is a theorem of $\mathcal{H}$.

Proof: "$\Leftarrow$" is clear (then both $\vdash P$ and $\vdash \neg P$). "$\Rightarrow$" If $\vdash \alpha$ and $\vdash \neg\alpha$, then for any β, by Theorem 4 and the cut property, $\vdash \beta$.

So an inconsistent system is useless: it "proves" every statement. And to show consistency, it suffices to exhibit *one* unprovable formula, just like `MU` for the MIU-system.

Theorem 5 (Double Negation Introduction): $\vdash A \to \neg\neg A$

(1)	$\neg\neg\neg A \to \neg A$	Theorem 3 $[A := \neg A]$
(2)	$(\neg\neg\neg A \to \neg A) \to (A \to \neg\neg A)$	A3 $[\alpha{=}A,\ \beta{=}\neg\neg A]$
(3)	$A \to \neg\neg A$	MP (1),(2)

Theorem 6 (Contraposition, 逆否): $\vdash (A \to B) \to (\neg B \to \neg A)$

We show $A \to B \vdash \neg B \to \neg A$ and then apply the Deduction Theorem.

(1)	$A \to B$	hypothesis
(2)	$\neg\neg A \to A$	Theorem 3
(3)	$\neg\neg A \to B$	HS (2),(1)
(4)	$B \to \neg\neg B$	Theorem 5 $[A := B]$
(5)	$\neg\neg A \to \neg\neg B$	HS (3),(4)
(6)	$(\neg\neg A \to \neg\neg B) \to (\neg B \to \neg A)$	A3 $[\alpha{=}\neg B,\ \beta{=}\neg A]$
(7)	$\neg B \to \neg A$	MP (5),(6)

As a consequence, we get **Modus Tollens**: $A \to B,\ \neg B \vdash \neg A$ (apply MP to line (7) and $\neg B$).

▶ Next, we justify "*proof by contradiction (反证法)*" from the lecture as a metatheorem of $\mathcal{H}$.

Theorem 7 (Reductio ad Absurdum, 归谬法)

If $\Gamma \cup \{\neg A\} \vdash B$ and $\Gamma \cup \{\neg A\} \vdash \neg B$, then $\Gamma \vdash A$.

Proof: By Theorem 4 (with $A := B$ and $B := \neg(A \to A)$) and the cut property, $\Gamma \cup \{\neg A\} \vdash \neg(A \to A)$. Then:

(1)	$\neg A \to \neg(A \to A)$	Deduction Theorem, from Γ
(2)	$(\neg A \to \neg(A \to A)) \to ((A \to A) \to A)$	A3 $[\alpha{=}A{\to}A,\ \beta{=}A]$
(3)	$(A \to A) \to A$	MP (1),(2)
(4)	$A \to A$	Theorem 1
(5)	A	MP (4),(3)

So $\Gamma \vdash A$.

Theorem 8 (Consequentia Mirabilis): $\vdash (\neg A \to A) \to A$

Let $\Gamma = \{\neg A \to A\}$. Then $\Gamma \cup \{\neg A\} \vdash A$ (by MP) and $\Gamma \cup \{\neg A\} \vdash \neg A$ (hypothesis). By Theorem 7, $\Gamma \vdash A$, i.e., $\neg A \to A \vdash A$. By the Deduction Theorem, $\vdash (\neg A \to A) \to A$.

▶ **Remark (the lecture's inference rules).** Recall the abbreviation $A \vee B := \neg A \to B$. Then, e.g., the rule "$\neg A \wedge (A \vee B) \Rightarrow B$" from the lecture becomes $\neg A,\ \neg A \to B \vdash B$, which is just MP. In the same way, all the basic inference formulas and equivalence laws of the lecture can be derived in $\mathcal{H}$ (either directly, or at once by the completeness theorem).

7 Why Is $\mathcal{H}$ Complete? (Proof Sketch)

▶ We sketch the proof of weak completeness ($\models \alpha \Rightarrow \vdash \alpha$) due to Kalmár (1935). The idea is beautiful: *simulate the truth table inside* $\mathcal{H}$. Besides Theorems 1–8, we need two more theorems of $\mathcal{H}$ (exercises):

$$\textbf{(L1)} \quad \vdash A \to (\neg B \to \neg(A \to B)), \qquad \textbf{(L2)} \quad \vdash (A \to B) \to ((\neg A \to B) \to B).$$

For an assignment I and a formula φ, define $\varphi^I = \varphi$ if φ is true under I, and $\varphi^I = \neg\varphi$ otherwise. (So φ^I is always true under I.)

Lemma (Kalmár): one row of the truth table = one derivation

Let α be a formula over $P_1, \ldots, P_n$ and I any assignment. Then $P_1^I, \ \ldots, \ P_n^I \vdash \alpha^I$.

Proof sketch: Induction on the structure of α; let $\Gamma_I = \{P_1^I, \ldots, P_n^I\}$.

- $\alpha = P_i$: then $\alpha^I = P_i^I \in \Gamma_I$.
- $\alpha = \neg\beta$: if β is false under I, then $\alpha^I = \neg\beta = \beta^I$, done by the induction hypothesis. If β is true, then $\alpha^I = \neg\neg\beta$, and we get it from $\beta^I = \beta$ by Theorem 5.
- $\alpha = \beta \to \gamma$: if β is false, then $\Gamma_I \vdash \neg\beta$ and we get $\beta \to \gamma$ by Theorem 4. If γ is true, then $\Gamma_I \vdash \gamma$ and we get $\beta \to \gamma$ by (A1). If β is true and γ is false, then $\Gamma_I \vdash \beta$, $\Gamma_I \vdash \neg\gamma$, and we get $\alpha^I = \neg(\beta \to \gamma)$ by (L1).

▶ **Proof of completeness.** Suppose $\models \alpha$. Then $\alpha^I = \alpha$ for *every* I, so by the lemma, $P_1^I, \ldots, P_n^I \vdash \alpha$ for all 2^n assignments. Take two assignments that differ only in P_n. They give $\Gamma', P_n \vdash \alpha$ and $\Gamma', \neg P_n \vdash \alpha$ with the same $\Gamma' = \{P_1^I, \ldots, P_{n-1}^I\}$. By the Deduction Theorem, $\Gamma' \vdash P_n \to \alpha$ and $\Gamma' \vdash \neg P_n \to \alpha$, and by (L2) and MP twice, $\Gamma' \vdash \alpha$. So we have eliminated P_n. Repeating this n times eliminates all hypotheses, and we get $\vdash \alpha$. □

▶ Notice that the proof literally *glues together the* 2^n *rows of the truth table* into one formal derivation. The resulting derivation is exponentially long, reflecting the exponential size of the truth table. (The strong version with an infinite Γ needs a different argument, e.g., via maximal consistent sets.)

8 Axioms vs. Truth Tables: What Is the Difference?

▶ We now have two completely different definitions of "logically valid":

	Semantic ($\models$, truth tables)	**Syntactic** ($\vdash$, axioms)
primitive notion	truth values $\{0,1\}$, meanings of operators	symbols, axioms, rules
"α is valid" means	true under all 2^n assignments	a finite derivation exists
to show valid	check all rows (mechanical, but 2^n work)	exhibit a derivation (needs creativity)
to show *not* valid	exhibit one falsifying row (easy)	hard from inside (like `MU`)
evidence	a whole table	a short, checkable certificate (a proof)
generalizes to	needs a new notion of "model"	just change axioms/rules

Soundness and completeness say that, for classical propositional logic, both define the *same* set of formulas. This is why the lecture could use truth tables everywhere, and why truth tables give an algorithm to decide provability in $\mathcal{H}$. But this coincidence is a *theorem*, not something obvious. Here are some reasons why the axiomatic view is still indispensable.

Insight 1: Truth tables are only one model

In the axiomatic view, **T** and **F** are not primitive. The equivalence laws of the lecture (commutative, distributive, De Morgan, …) define an algebraic structure called a *Boolean algebra*, and $\{0,1\}$ is just the smallest one. The power set 2^U with $\cup, \cap, ^c$ is another (you will see this in set theory!). A deep fact is that an identity holds in *all* Boolean algebras iff it holds in $\{0,1\}$. So the tiny two-row world of truth tables already captures everything, which is again a completeness phenomenon.

Insight 2: Different axioms, different logics

If we replace (A3) by weaker axioms, e.g., $\neg A \to (A \to B)$ and $(A \to B) \to ((A \to \neg B) \to \neg A)$, then $\neg\neg A \to A$ is no longer provable, and we get *intuitionistic logic*, in which "A is true" means "we have a construction/proof of A". There, $A \vee \neg A$ is not a theorem. Gödel (1932) showed that intuitionistic logic has *no* semantics by finite truth tables at all. Here the axioms come first, and a suitable semantics (Kripke models, Heyting algebras) had to be invented later. The same holds for modal logics, temporal logics (used in verification of programs), etc. Axioms are the flexible way to define a logic.

Insight 3: Beyond propositional logic, truth tables break down

In first-order logic (with $\forall, \exists$, see later), variables range over possibly infinite domains, so there is no finite table to check. Validity becomes *undecidable* (Church and Turing, 1936). Yet Gödel's completeness theorem (1929) says every valid first-order formula still has a finite proof. So proofs survive where truth tables die. For arithmetic it gets even worse: by Gödel's incompleteness theorem (1931), for any consistent, mechanically checkable axiom system for arithmetic there are *true* statements that are *not provable*, and the system cannot prove its own consistency. The gap between $\models$ and $\vdash$ becomes real, and its proof uses exactly the "strings as numbers" trick we saw in the 310-system.

Insight 4: Hilbert's program and the MU-puzzle

Around 1920, Hilbert proposed to formalize all of mathematics as a MIU-like symbol game and prove, by simple finitary reasoning *about* the game, that it can never derive a contradiction (i.e., that some "`MU`" is not derivable). For propositional logic this works, and our consistency proof is exactly of this type. For arithmetic, Gödel showed that it cannot work.

Insight 5: The cost of proofs and P vs. NP

A truth table has 2^n rows; a proof may be much shorter, and it is easy to check. Is there a proof system in which *every* tautology has a polynomial-size proof? Cook and Reckhow (1979) showed that this is equivalent to NP = coNP, a major open problem. Resolution (from the lecture) is one proof system of this kind, and it is known to need exponentially long proofs for some tautologies (e.g., the pigeonhole principle). Today, SAT solvers and proof assistants (Lean, Coq, Isabelle) are, at their core, machines for producing and checking syntactic derivations.

▶ **Summary.** Semantics tells us *what* is true; syntax tells us *how* we can come to know it by finite, checkable steps. Soundness guarantees that our methods of proof never lead us astray, and completeness guarantees that they are strong enough to reach every truth. For propositional logic we are lucky to have both; much of modern logic is about what happens when we are not.

9 Exercises

1. (MIU) Derive `MUIIU` in the MIU-system. Also prove that every MIU-theorem starts with `M` and contains no other `M`.
2. (MIU, challenge) Prove that $\texttt{MI}^k$ is a MIU-theorem for every k with $3 \nmid k$. (Hint: first get $\texttt{MI}^{2^j}$ and use $2^j \bmod 3$.) Conclude the exact characterization of MIU-theorems stated in Section 1.
3. Using the Deduction Theorem, prove in $\mathcal{H}$: (a) $\vdash (A \to (B \to C)) \to (B \to (A \to C))$; (b) $\vdash (A \to B) \to ((A \to \neg B) \to \neg A)$ (hint: Theorem 7); (c) $\vdash \neg(A \to B) \to A$.
4. Prove the two lemmas (L1) and (L2) used in the completeness proof.
5. (Independence of (A3)) Interpret $\to$ as usual, but interpret $\neg$ as the constant function $\neg x = 1$. Show that all instances of (A1), (A2) are true under every assignment in this "strange semantics" and MP preserves this property, but $\neg\neg P \to P$ is not. Conclude that $\neg\neg P \to P$ cannot be derived from (A1), (A2) and MP alone.
6. (Consistency without soundness) Add the propositional variable P itself as a fourth axiom (just one formula, not a schema). Is the new system still sound? Is it still consistent? (Hint: look at the assignment making P true.) What if we add the schema α (every formula is an axiom) instead?