COMP0009 Logic

Lecture notes · 2026–27

Part 2: Automated reasoning

Chapter 6

Model Theory

Alexander V. Gheorghiu

Department of Computer Science, University College London

This chapter is issued on its own from the complete lecture notes for COMP0009 Logic. Its section, theorem, exercise and equation numbers are those of the complete notes, so that the two can be read side by side; references to other chapters and parts point into the complete notes.

Contents

6. Model Theory 1
6.1. Structures 2
6.2. Syntax 5
6.3. Satisfaction 9
6.4. Entailment 13
6.5. Soundness and Incompleteness of NJ 17
6.6. Law of the Excluded Middle 21
6.7. Completeness of NK 23
6.8. Classical and Intuitionistic Logic 30
History 34
Exercises 39

Employee	Boss	Engineer?	Manager?	Salary
Thorne	Thorne	no	yes	200
Finch	Thorne	no	yes	120
Pratt	Thorne	no	no	95
Bell	Finch	yes	no	90
Cole	Finch	yes	no	85

FIGURE 6.1. The personnel table on Tuesday.

6. Model Theory

Suppose you look after the platform at a company of five people, and you want, more than anything, never to be surprised by your own system. On Tuesday the finance director asks you to confirm something everyone takes for granted: that everybody earns less than their boss. The personnel table, Figure 6.1, is short enough to print.

You don't check the claim row by row. Like anyone who has written SQL, you ask the database for the rows that break it:

```
SELECT e.name, b.name AS boss
FROM   staff e JOIN staff b ON e.boss = b.name
WHERE  NOT (b.salary > e.salary);
```

One row comes back: `Thorne | Thorne`. The chief executive is her own boss, and she doesn't earn more than herself. You add `AND e.name <> 'Thorne'` to the query, it returns nothing, and you reply that the claim holds.

The finance director answers with one line: "Can that ever stop being true?" This is a different question. The table changes continually as people are hired, promoted, or given rises. He does not want to know whether the claim is true now, but whether it is true of every table the personnel system can produce given the HR policy.

The story contains two ideas, and we should keep them apart from the start. One is that of a sentence being *true here*, of this table on this Tuesday. The other is that of a sentence being *true everywhere the constraints allow*, of every table that obeys them. The first is a matter of looking, and the query engine settles it in a fraction of a second. The second is a claim about infinitely many tables, most of which nobody will see, and no amount of looking settles it.

This chapter is about the two ideas and the connection between them. It asks three questions:

- What does it mean for a sentence to be true of a system?
- What does it mean for a sentence to follow from the requirements on a system, so that it is true of every system that meets them?
- Is there a finite way to show that it does?

The first two questions are answered by definitions, and the third by a theorem, though the first rules of proof we try turn out to fall short. Section 6.1 makes a system into a mathematical object, a *structure*. Section 6.2 gives

System	Things	Relations	Functions
The company	five people	is an engineer; is a manager; earns more than	the boss of
Counting numbers	$0, 1, 2, \ldots$	is less than	successor, $+$, $\times$
Server network	four machines	is linked to	none
A program	its states	is an error state	the next state
Truth values	true and false	none	and, or, not

FIGURE 6.2. Five systems.

a language for talking about structures, and Section 6.3 defines when a sentence of that language is true in a structure. Section 6.4 defines what it is for a sentence to follow from others: to be true in every structure in which they are true. Section 6.5 tries the proofs we already have, Gentzen's natural deduction system NJ: they never lead from true premisses to a false conclusion, but they miss some sentences that follow, all for want of a single law. Adding that law, the law of the excluded middle, gives the system NK (Section 6.6), and Section 6.7 proves it complete: a sentence follows from others exactly when it has a derivation from them in NK. Section 6.8 counts what the extra law costs.

6.1. **Structures.** Before the word *true* can mean anything, we must say what a sentence is true *of*. Engineers use the word *system* for almost anything with parts, and we begin from what the word covers rather than from a definition. Figure 6.2 lists five systems.

The five have nothing in common as subject matter, yet each comes in the same three pieces:

- a collection of things;
- some relations among the things, each true of some tuples of things and false of others;
- some functions, each taking things and returning a thing.

Some systems also have a few things with names of their own: a company has *the* chief executive, the numbers have *the* starting number 0, and a program has *the* initial state.

Definition 6.1 (signature)**.** A signature is the vocabulary that sentences about one kind of system may use: names for particular things, symbols for functions and symbols for relations, each with its number of places.

Formally, a *signature* is a quadruple $\Sigma = (\mathcal{C}, \mathcal{F}, \mathcal{R}, \mathrm{ar})$ in which

- $\mathcal{C}$, $\mathcal{F}$ and $\mathcal{R}$ are pairwise disjoint sets, whose elements are called *constant symbols*, *function symbols* and *relation symbols* respectively;
- $\mathrm{ar} : \mathcal{F} \cup \mathcal{R} \to \mathbb{N}$ is a function, the *arity*, with $\mathrm{ar}(f) \geq 1$ for every $f \in \mathcal{F}$.

A relation symbol of arity 0 is a *propositional letter*. The signature is *countable* if the set $\mathcal{C} \cup \mathcal{F} \cup \mathcal{R}$ is countable.

The arity says how many inputs each function or relation symbol takes. Given only the symbol f, nobody could tell whether it names a function of one variable, of two, or of a million.

Definition 6.2 (structure)**.** A structure for a signature says which things there are and what each symbol of the signature means.

Formally, let $\Sigma = (\mathcal{C}, \mathcal{F}, \mathcal{R}, \mathrm{ar})$ be a signature. A Σ-*structure* is a pair $\mathfrak{M} = (\mathbb{D}, I)$ in which

- $\mathbb{D}$ is a *non-empty* set, the *domain* of $\mathfrak{M}$;
- I is a function with domain $\mathcal{C} \cup \mathcal{F} \cup \mathcal{R}$, the *interpretation*, such that
  1. $I(c) \in \mathbb{D}$ for every $c \in \mathcal{C}$;
  2. $I(f)$ is a function from $\mathbb{D}^n$ to $\mathbb{D}$ for every $f \in \mathcal{F}$ with $\mathrm{ar}(f) = n$;
  3. $I(R) \subseteq \mathbb{D}^n$ for every $R \in \mathcal{R}$ with $\mathrm{ar}(R) = n$.

Here $\mathbb{D}^n$ is the set of n-tuples of elements of $\mathbb{D}$; in particular $\mathbb{D}^0 = \{()\}$, whose one element is the empty tuple. We write $[\![c]\!]^{\mathfrak{M}}$, $[\![f]\!]^{\mathfrak{M}}$ and $[\![R]\!]^{\mathfrak{M}}$ for $I(c)$, $I(f)$ and $I(R)$.

A relation with n places is a table with n columns and one row for each tuple of which it holds, so a structure is a database: one table for each relation symbol, one lookup table for each function symbol, and a few named rows.[1]

Example 6.3 (the company)**.** Harriet Ashworth looks after the platform at the company of the opening, and the personnel table of Figure 6.1 is hers. Its vocabulary is the *company signature*

$$\Sigma_{\mathrm{C}} = (\{\mathsf{ceo}\},\ \{\mathsf{boss}\},\ \{=, \mathsf{Eng}, \mathsf{Mgr}, \mathsf{Richer}\},\ \mathrm{ar})$$

with $\mathrm{ar}(\mathsf{boss}) = \mathrm{ar}(\mathsf{Eng}) = \mathrm{ar}(\mathsf{Mgr}) = 1$ and $\mathrm{ar}(=) = \mathrm{ar}(\mathsf{Richer}) = 2$. The formula $\mathsf{Richer}(x, y)$ is read "x earns more than y". The symbol $=$ is a two-place relation symbol like the others, written between its arguments as usual.

Tuesday's table is the Σ_{C}-structure $\mathfrak{C} = (\mathbb{D}_{\mathrm{C}}, I_{\mathrm{C}})$ given by

$$\begin{aligned}
\mathbb{D}_{\mathrm{C}} &= \{\mathrm{Thorne}, \mathrm{Finch}, \mathrm{Pratt}, \mathrm{Bell}, \mathrm{Cole}\},\\
[\![\mathsf{ceo}]\!]^{\mathfrak{C}} &= \mathrm{Thorne},\\
[\![\mathsf{boss}]\!]^{\mathfrak{C}} &= \{\mathrm{Thorne} \mapsto \mathrm{Thorne},\ \mathrm{Finch} \mapsto \mathrm{Thorne},\ \mathrm{Pratt} \mapsto \mathrm{Thorne},\\
&\qquad \mathrm{Bell} \mapsto \mathrm{Finch},\ \mathrm{Cole} \mapsto \mathrm{Finch}\},\\
[\![\mathsf{Eng}]\!]^{\mathfrak{C}} &= \{\mathrm{Bell}, \mathrm{Cole}\},\\
[\![\mathsf{Mgr}]\!]^{\mathfrak{C}} &= \{\mathrm{Thorne}, \mathrm{Finch}\},\\
[\![=]\!]^{\mathfrak{C}} &= \{(d, d) : d \in \mathbb{D}_{\mathrm{C}}\},\\
[\![\mathsf{Richer}]\!]^{\mathfrak{C}} &= \{(d, e) \in \mathbb{D}_{\mathrm{C}}^2 : \mathit{sal}(d) > \mathit{sal}(e)\},
\end{aligned}$$

where $\mathit{sal} : \mathbb{D}_{\mathrm{C}} \to \mathbb{N}$ is the salary column of Figure 6.1.

[1] The resemblance is no accident. Codd's relational model of data (Codd, 1970), on which SQL rests, treats a database as a finite structure, and its query languages are first-order logic in other notation.

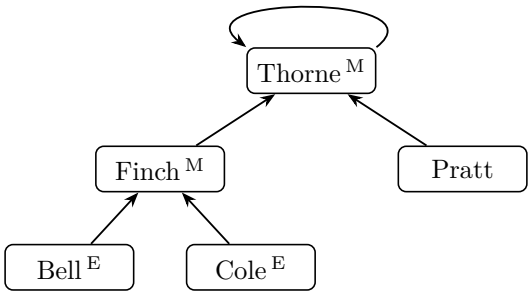


FIGURE 6.3. The company $\mathfrak{C}$.

The structure interprets $=$ as identity, the relation of each person to herself. Nothing in the definition of a structure demands that, since $=$ may be interpreted as any set of pairs, but every structure for the company signature that we describe interprets it as identity.

The salaries are all different, so exactly half of the twenty ordered pairs of different people are in $[\![\mathsf{Richer}]\!]^{\mathfrak{C}}$: ten pairs, from (Thorne, Finch) down to (Bell, Cole). The function *sal* is not part of the structure. The signature has no symbol for a salary, so the structure keeps only what the symbols can ask about: who earns more than whom.

Figure 6.3 draws the structure, with an arrow from each person to their boss; M marks a manager and E an engineer. Thorne's loop is the row that Harriet's first query returned.

A structure with a finite domain can be listed, table by table, as a personnel table can. A structure with an infinite domain can be described but not printed, and how to describe one exactly is a delicate matter.[2] The counting numbers of Figure 6.2 are of that kind.

Example 6.4 (the counting numbers)**.** The counting numbers are a structure for the *signature of arithmetic*

$$\Sigma_{\mathrm{A}} = (\{0\},\ \{s, +, \times\},\ \{<\},\ \mathrm{ar})$$

with $\mathrm{ar}(s) = 1$ and $\mathrm{ar}(+) = \mathrm{ar}(\times) = \mathrm{ar}(<) = 2$.

The structure is $\mathfrak{N} = (\mathbb{N}, I_{\mathrm{N}})$, traditionally called the *standard model of arithmetic*, where $\mathbb{N} = \{0, 1, 2, \ldots\}$ and

- $[\![0]\!]^{\mathfrak{N}}$ is the number zero,
- $[\![s]\!]^{\mathfrak{N}}(n) = n + 1$,
- $[\![+]\!]^{\mathfrak{N}}$ and $[\![\times]\!]^{\mathfrak{N}}$ are addition and multiplication,
- $[\![<]\!]^{\mathfrak{N}} = \{(m, n) \in \mathbb{N}^2 : m \text{ is less than } n\}$.

Example 6.5 (the other three systems)**.** The other three systems of Figure 6.2 are structures in the same way.

- The server network has a domain of four machines and one two-place relation.

[2] See Chapter 8.

- The program has a domain of states, a constant for the initial state, a one-place relation for the error states and a one-place function for the next state.
- The truth values form a domain of two elements with three functions and no relations.

Remark 6.6 (non-empty domains). Almost every textbook requires the domain to be non-empty, and so do we. The requirement looks fussy, but it decides the truth of sentences such as "if everything is a manager, then something is a manager". In an empty structure everything is a manager, since there is nothing that is not, and yet nothing is a manager, so the sentence would be false. The requirement rules this out.

Remark 6.7 (total functions). The functions of a structure are *total*: a function symbol is interpreted as a function that gives an element of the domain for every input from the domain. That is what makes every term name something: the boss of anybody's boss is always somebody. A real system need not be so obliging, and where a value may be missing, a relation such as *reports to* can stand in for the function *the boss of*.

Example 6.8 (the company, cont'd: a near miss)**.** Suppose the table left Thorne's boss blank, since the chief executive has none. That is a sensible way to run a company, but the result is not a structure: $[\![\mathsf{boss}]\!]^{\mathfrak{C}}$ would have a hole at Thorne, and a lookup table with a hole in it is not a function. Somebody had to choose a value for Thorne, and choosing Thorne herself is the reason Harriet's first query returned a row.

6.2. **Syntax.** A structure can be asked questions, and this section builds the language to ask them in. We start from a signature $\Sigma = (\mathcal{C}, \mathcal{F}, \mathcal{R}, \mathrm{ar})$ and a countably infinite set Var of *variables*, x, y, z, x_1, x_2 and so on, disjoint from the symbols of Σ. Constants and variables name elements of the domain, and function symbols build new names from old ones.

Definition 6.9 (terms)**.** The Σ-*terms* are given by the grammar

$$t \ ::= \ x \ \mid \ c \ \mid \ f(t_1, \ldots, t_n),$$

where x is a variable, c a constant symbol of Σ, and f a function symbol of Σ with $\mathrm{ar}(f) = n$. We write Term_Σ for the set of Σ-terms.

Example 6.10 (the company, cont'd: terms)**.** In the company signature, ceo, $\mathsf{boss}(\mathsf{ceo})$, $\mathsf{boss}(\mathsf{boss}(\mathsf{ceo}))$ and $\mathsf{boss}(x)$ are terms. In $\mathfrak{C}$ the first three all name Thorne, and $\mathsf{boss}(x)$ names somebody only once we know who x is. The string $\mathsf{boss}(\mathsf{ceo}, \mathsf{ceo})$ is not a term, since $\mathrm{ar}(\mathsf{boss}) = 1$, and neither is $\mathsf{Eng}(\mathsf{ceo})$, which says something rather than naming something. In arithmetic, $s(s(0))$ and $x + s(0)$ are terms.

Having named things, we want to say something about them. A relation applied to terms is the smallest such statement, and it is called an *atomic* formula: $\mathsf{Richer}(\mathsf{boss}(x), x)$ says that the boss of x earns more than x, and $\bot$ says something false outright. Atomic formulas are combined into complex ones by the logical symbols. We use Greek letters φ, ψ and χ for formulas, and Latin letters p and q for propositional letters.

Definition 6.11 (formulas)**.** The Σ-*formulas* are given by the grammar

$$\varphi, \psi \ ::= \ \bot \ \mid \ R(t_1, \ldots, t_n) \ \mid \ (\varphi \to \psi) \ \mid \ (\varphi \wedge \psi) \ \mid \ (\varphi \vee \psi) \ \mid \ \forall x\, \varphi \ \mid \ \exists x\, \varphi,$$

where R is a relation symbol of Σ with $\mathrm{ar}(R) = n$, the t_i are Σ-terms, and x is a variable. We write Form_Σ for the set of Σ-formulas.

The formulas $\bot$ and $R(t_1, \ldots, t_n)$ are *atomic*. For a propositional letter p the atomic formula $p()$ is written p. Three more symbols are abbreviations:

- $\neg\varphi$ is $\varphi \to \bot$;
- $\varphi \leftrightarrow \psi$ is $(\varphi \to \psi) \wedge (\psi \to \varphi)$;
- when Σ contains $=$, the formula $t_1 \neq t_2$ is $\neg(t_1 = t_2)$.

Familiar two-place symbols are written between their arguments, as usual: $x + y < z$ is the atomic formula $<(+(x, y), z)$, and $t_1 = t_2$ is $=(t_1, t_2)$. Brackets are kept down by the usual conventions:

- The outermost pair is dropped.
- $\wedge$ and $\vee$ take precedence over $\to$, which groups to the right.
- $\neg$ and the quantifiers $\forall$ and $\exists$ take precedence over everything else.

A symbol that takes precedence applies first, to as little of the formula as it can, just as $\times$ applies before $+$ in $2 + 3 \times 4$.

Example 6.12 (reading the conventions)**.** With propositional letters p, q and r:

- $\neg p \wedge q$ is $(\neg p) \wedge q$, not $\neg(p \wedge q)$;
- $p \wedge q \to r$ is $((p \wedge q) \to r)$;
- $p \to q \to r$ is $(p \to (q \to r))$.

A quantifier, too, covers as little as it can: in the company signature, $\forall x\, \mathsf{Eng}(x) \to \mathsf{Mgr}(x)$ is $(\forall x\, \mathsf{Eng}(x)) \to \mathsf{Mgr}(x)$, in which the quantifier covers only $\mathsf{Eng}(x)$. A quantifier meant to cover more is followed by brackets, as in $\forall x\, (\mathsf{Eng}(x) \to \mathsf{Mgr}(\mathsf{boss}(x)))$, "every engineer reports to a manager".

A formula with variables does not yet say anything definite. $\mathsf{Richer}(\mathsf{boss}(x), x)$ is true or false according to who x is, whereas $\forall x\, \mathsf{Richer}(\mathsf{boss}(x), x)$ makes one claim about the whole domain: everyone earns less than their boss. The next definition separates the variables whose values must be supplied from those that a quantifier has already taken charge of.

Definition 6.13 (free variables)**.** The free variables of a term or formula are those whose values must be supplied before it names or says anything.

Formally, the set $\mathrm{var}(t)$ of *variables* of a term t is defined by recursion on the shape of t:

$$\mathrm{var}(t) \coloneqq \begin{cases} \{x\} & \text{if } t \text{ is a variable } x, \\ \varnothing & \text{if } t \text{ is a constant } c, \\ \mathrm{var}(t_1) \cup \cdots \cup \mathrm{var}(t_n) & \text{if } t \text{ is } f(t_1, \ldots, t_n). \end{cases}$$

Formula	English
$\forall x\, \mathsf{Richer}(\mathsf{boss}(x), x)$	everyone earns less than their boss
$\forall x\, (x \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(x), x))$	everyone except the chief executive earns less than their boss
$\forall x\, (\mathsf{Eng}(x) \to \mathsf{Mgr}(\mathsf{boss}(x)))$	every engineer reports to a manager
$\exists x\, (\mathsf{Mgr}(x) \wedge \mathsf{boss}(x) \neq x)$	some manager has a boss other than themselves

FIGURE 6.4. Four sentences and their English.

The set $\mathrm{FV}(\varphi)$ of *free variables* of a formula φ is defined by recursion on the shape of φ, with $\circ$ any of $\to$, $\wedge$ and $\vee$, and Q either quantifier:

$$\mathrm{FV}(\varphi) \coloneqq \begin{cases} \varnothing & \text{if } \varphi \text{ is } \bot, \\ \mathrm{var}(t_1) \cup \cdots \cup \mathrm{var}(t_n) & \text{if } \varphi \text{ is } R(t_1, \ldots, t_n), \\ \mathrm{FV}(\psi) \cup \mathrm{FV}(\chi) & \text{if } \varphi \text{ is } \psi \circ \chi, \\ \mathrm{FV}(\psi) \setminus \{x\} & \text{if } \varphi \text{ is } Qx\, \psi. \end{cases}$$

An occurrence of x in φ is *bound* if it lies inside a part of φ of the form $\forall x\, \psi$ or $\exists x\, \psi$, and *free* otherwise; the quantifier *binds* the occurrences of x in ψ.

Definition 6.14 (closed term; sentence)**.** A term t is *closed* if $\mathrm{var}(t) = \varnothing$, and a formula φ is a *sentence* if $\mathrm{FV}(\varphi) = \varnothing$.

Example 6.15 (free and bound)**.** In the company signature, ceo and $\mathsf{boss}(\mathsf{boss}(\mathsf{ceo}))$ are closed terms, and $\mathsf{boss}(x)$ is not. Of the two formulas

$$\mathsf{Richer}(\mathsf{boss}(x), x) \qquad \text{and} \qquad \forall x\, \mathsf{Richer}(\mathsf{boss}(x), x),$$

the first is not a sentence, since its one free variable is x, and the second is, since the last line of the definition of FV removes x.

In the formula

$$\mathsf{Eng}(x) \to \forall x\, \mathsf{Mgr}(x)$$

the first occurrence of x is free and the other two are bound, so the formula is not a sentence.

Using the language is mostly translation, in both directions, and one choice causes most mistakes: which connective joins a quantifier to the condition that restricts it. "Every P ..." restricts a universal quantifier, and the restriction goes in with $\to$: for every x, *if* x is a P, then .... "Some P ..." restricts an existential quantifier, and the restriction goes in with $\wedge$: some x is a P *and* .... The rule of thumb, universal with $\to$ and existential with $\wedge$, is reliable but not infallible.

Example 6.16 (the company, cont'd: four sentences in English)**.** Figure 6.4 gives four sentences in the company signature, each with its English.

The third and fourth sentences follow the rule of thumb. Written with $\wedge$, the third becomes $\forall x\, (\mathsf{Eng}(x) \wedge \mathsf{Mgr}(\mathsf{boss}(x)))$, which says that everyone is an engineer, false of any firm that employs somebody else. Written with $\to$,

the fourth becomes $\exists x\,(\mathsf{Mgr}(x) \to \mathsf{boss}(x) \neq x)$, which anyone who is not a manager makes true.

Proofs pass from a claim about anybody to the same claim about somebody in particular, as ($\forall$E) does, and the other way round, as ($\exists$I) does. The syntactic tool is *substitution*: put a term in place of x, so that $\mathsf{Richer}(\mathsf{boss}(x), x)$ becomes $\mathsf{Richer}(\mathsf{boss}(\mathsf{ceo}), \mathsf{ceo})$, which says it of the chief executive. Substitution is an operation on strings, and truth is not defined through it (Remark 6.24 says why). It comes with a danger. The term put in may have variables of its own, and a quantifier in the formula can capture them, so that the result says something quite different. The definition of substitution is therefore followed by a second, of the condition that rules capture out.

Definition 6.17 (substitution)**.** Substitution puts a term in place of the free occurrences of a variable.

Formally, let t be a term and x a variable. For a term u, the term $u[t/x]$ is defined by recursion on the shape of u:

$$u[t/x] \coloneqq \begin{cases} t & \text{if } u \text{ is } x, \\ u & \text{if } u \text{ is a variable } y \neq x \text{ or a constant}, \\ f(u_1[t/x], \ldots, u_n[t/x]) & \text{if } u \text{ is } f(u_1, \ldots, u_n). \end{cases}$$

For a formula φ, the formula $\varphi[t/x]$ is defined by recursion on the shape of φ, with $\circ$ any of $\to$, $\wedge$ and $\vee$, and Q either quantifier:

$$\varphi[t/x] \coloneqq \begin{cases} \bot & \text{if } \varphi \text{ is } \bot, \\ R(u_1[t/x], \ldots, u_n[t/x]) & \text{if } \varphi \text{ is } R(u_1, \ldots, u_n), \\ \psi[t/x] \circ \chi[t/x] & \text{if } \varphi \text{ is } \psi \circ \chi, \\ Qx\,\psi & \text{if } \varphi \text{ is } Qx\,\psi, \\ Qy\,\psi[t/x] & \text{if } \varphi \text{ is } Qy\,\psi \text{ with } y \neq x. \end{cases}$$

Definition 6.18 (substitutable)**.** A term is substitutable for a variable when putting it in place of the variable lets no quantifier capture a variable of the term.

Formally, the term t is *substitutable for x in φ* if no free occurrence of x in φ lies inside a part of φ of the form $\forall y\,\psi$ or $\exists y\,\psi$ with $y \in \mathrm{var}(t)$.

Example 6.19 (the company, cont'd: substitution and capture)**.** If φ is $x \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(x), x)$, then $\varphi[\mathsf{boss}(\mathsf{ceo})/x]$ is $\mathsf{boss}(\mathsf{ceo}) \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(\mathsf{boss}(\mathsf{ceo})), \mathsf{boss}(\mathsf{ceo}))$: the claim made of x is now made of the chief executive's boss.

Substitutability rules out the one way in which substitution can go wrong. Let φ be $\exists y\,\mathsf{Richer}(y, x)$, "someone earns more than x", and let t be $\mathsf{boss}(y)$. The recursion gives $\exists y\,\mathsf{Richer}(y, \mathsf{boss}(y))$, which says that someone earns more than their own boss, a different claim altogether: the y of the term has been *captured* by the quantifier. The condition forbids it, since the free occurrence of x lies inside $\exists y\,\mathsf{Richer}(y, x)$ and $y \in \mathrm{var}(\mathsf{boss}(y))$.

We write $\varphi[t/x]$ only when t is substitutable for x in φ. A closed term always is, having no variables to capture.

The sentences of Figure 6.4 look as if they speak about a particular firm, and we have read them so. But nothing in this section says how a string of symbols speaks about anything, and so nothing here decides whether $\forall x\, \mathsf{Richer}(\mathsf{boss}(x), x)$ is true. Nothing could, since the string mentions no structure: every structure for its signature reads it. The reading we have been doing by intuition needs a definition, and the next section gives it.

6.3. **Satisfaction.** A structure says what the symbols of a formula mean. It does not say what the variables stand for, and a formula with free variables is true or false only once something does.

Definition 6.20 (assignment)**.** An assignment says what each variable stands for.

Formally, let $\mathfrak{M}$ be a structure with domain $\mathbb{D}$. An *assignment* into $\mathfrak{M}$ is a function $s : \mathrm{Var} \to \mathbb{D}$. For $x \in \mathrm{Var}$ and $d \in \mathbb{D}$, the assignment $s[d/x]$, "s with d for x", is defined by

$$s[d/x](y) \coloneqq \begin{cases} d & \text{if } y \text{ is } x, \\ s(y) & \text{if } y \neq x. \end{cases}$$

Together, a structure and an assignment settle what every term names and whether every formula is true.

Definition 6.21 (value of a term)**.** The value of a term is the element it names, once a structure and an assignment are fixed.

Formally, let $\mathfrak{M}$ be a structure with domain $\mathbb{D}$, and s an assignment into $\mathfrak{M}$. The *value* $\mathrm{Val}_s^{\mathfrak{M}}(t) \in \mathbb{D}$ of a term t in $\mathfrak{M}$ under s is defined by recursion on the shape of t:

$$\mathrm{Val}_s^{\mathfrak{M}}(t) \coloneqq \begin{cases} s(x) & \text{if } t \text{ is a variable } x, \\ [\![c]\!]^{\mathfrak{M}} & \text{if } t \text{ is a constant } c, \\ [\![f]\!]^{\mathfrak{M}}(\mathrm{Val}_s^{\mathfrak{M}}(t_1), \ldots, \mathrm{Val}_s^{\mathfrak{M}}(t_n)) & \text{if } t \text{ is } f(t_1, \ldots, t_n). \end{cases}$$

We write $\mathrm{Val}_s(t)$ when the structure is clear.

Example 6.22 (the company, cont'd: the boss's boss)**.** In $\mathfrak{C}$, under an assignment s with $s(x) = \mathrm{Cole}$, the term $\mathsf{boss}(\mathsf{boss}(x))$ has the value

$$\mathrm{Val}_s(\mathsf{boss}(\mathsf{boss}(x))) = [\![\mathsf{boss}]\!]^{\mathfrak{C}}([\![\mathsf{boss}]\!]^{\mathfrak{C}}(\mathrm{Cole})) = [\![\mathsf{boss}]\!]^{\mathfrak{C}}(\mathrm{Finch}) = \mathrm{Thorne}.$$

The closed term $\mathsf{boss}(\mathsf{boss}(\mathsf{ceo}))$ also has the value Thorne, and under every assignment.

We already know what the logical words mean in English, and the definition writes that meaning down, one kind of formula at a time: a conjunction is true when both conjuncts are, a disjunction when at least one disjunct is, and "for every x" when what follows holds whichever element x stands for. The structure enters only at the atomic formulas: $R(t_1, \ldots, t_n)$ is true when the things the terms name stand in the relation that R names.

Definition 6.23 (satisfaction)**.** Satisfaction says when a formula is true in a structure under an assignment.

Formally, let $\mathfrak{M}$ be a structure with domain $\mathbb{D}$, and s an assignment into $\mathfrak{M}$. The relation $\mathfrak{M}, s \models \varphi$, read "$\varphi$ is *true in* $\mathfrak{M}$ *under* s" or "$\mathfrak{M}, s$ *satisfies*

$$
\begin{array}{lll}
\mathfrak{M}, s \models R(t_1, \ldots, t_n) & \text{iff} & (\mathrm{Val}_s(t_1), \ldots, \mathrm{Val}_s(t_n)) \in [\![R]\!]^{\mathfrak{M}} \\
\mathfrak{M}, s \not\models \bot & & \\
\mathfrak{M}, s \models \psi \wedge \chi & \text{iff} & \mathfrak{M}, s \models \psi \text{ and } \mathfrak{M}, s \models \chi \\
\mathfrak{M}, s \models \psi \vee \chi & \text{iff} & \mathfrak{M}, s \models \psi \text{ or } \mathfrak{M}, s \models \chi \\
\mathfrak{M}, s \models \psi \to \chi & \text{iff} & \mathfrak{M}, s \not\models \psi \text{ or } \mathfrak{M}, s \models \chi \\
\mathfrak{M}, s \models \forall x\, \psi & \text{iff} & \mathfrak{M}, s[d/x] \models \psi \text{ for every } d \in \mathbb{D} \\
\mathfrak{M}, s \models \exists x\, \psi & \text{iff} & \mathfrak{M}, s[d/x] \models \psi \text{ for some } d \in \mathbb{D}
\end{array}
$$

FIGURE 6.5. Satisfaction.

φ", is defined by recursion on the shape of φ by the clauses of Figure 6.5, one for each kind of formula. There $\mathfrak{M}, s \not\models \varphi$ means that $\mathfrak{M}, s \models \varphi$ does not hold, and "or" is inclusive.[3]

The clause for $\bot$ says that it is true nowhere. Since $\neg\varphi$ is $\varphi \to \bot$, the clause for $\to$ yields one for negation: $\mathfrak{M}, s \models \neg\varphi$ if and only if $\mathfrak{M}, s \not\models \varphi$.

The clause for $\to$ is a deliberate choice. It ignores any connection of content between the two halves, so that "if $2+2=5$ then $1+1=3$" is true, because its antecedent is false. Other readings of "if", which ask for such a connection, have been studied, and none would make the truth of $\varphi \to \psi$ depend only on the truth of φ and of ψ. This one does, and that is what lets the definition decide an implication by deciding its two halves.

The quantifier clauses are where the assignment earns its keep. To decide whether $\forall x\, \psi$ is true, we must look at ψ with x standing for each element of the domain in turn, while every other variable keeps the value it had. The assignment $s[d/x]$ is exactly that: s with x pointed at d and nothing else changed. Whatever s gave x before does not matter, since the quantifier has taken charge of x.[4] The job falls to the assignment because a structure fixes once and for all what each constant symbol names, while a variable can stand for one element now and for another a moment later.

Remark 6.24 (unnamed elements). One might try to define truth by substitution instead: $\forall x\, \psi$ is true when $\psi[t/x]$ is true for every closed term t, and $\exists x\, \psi$ when it is true for some. That would be wrong, because a domain can contain elements that no closed term names. Take the real numbers as a structure for a signature with a constant symbol 0, a one-place function symbol s and a two-place relation symbol $<$, read as zero, adding one and less than. Every closed term, 0, $s(0)$, $s(s(0))$ and so on, names a natural number,

[3] This is Tarski's definition of truth, from Tarski (1935), first published in Polish in 1933. Its point, for Tarski, was that truth for a formal language can be defined without circularity: the definition of "true" for $\forall x\, \varphi$ uses "true" only for φ. It was written for mathematicians and philosophers, and forty years later it was the evaluator inside every database query engine. Tarski was born Alfred Teitelbaum and took the name Tarski in 1923. In August 1939 he left for a congress at Harvard, on the last ship to sail from Poland for the United States before the German invasion; from 1942 he taught at Berkeley (Feferman and Feferman, 2004).

[4] Just as a local variable in a program hides an outer one of the same name.

d	$x \neq \mathsf{ceo}$?	$[\![\mathsf{boss}]\!]^{\mathfrak{C}}(d)$	$\mathsf{Richer}(\mathsf{boss}(x), x)$?	the implication
Thorne	no	Thorne	no ($200 \not> 200$)	true (antecedent false)
Finch	yes	Thorne	yes ($200 > 120$)	true
Pratt	yes	Thorne	yes ($200 > 95$)	true
Bell	yes	Finch	yes ($120 > 90$)	true
Cole	yes	Finch	yes ($120 > 85$)	true

FIGURE 6.6. One row for each element of the domain.

and no natural number lies strictly between 0 and 1, yet $\exists x\,(0 < x \wedge x < s(0))$ is true. An assignment reaches every element of the domain, named or not.

Example 6.25 (the company, cont'd: evaluating a universal sentence)**.** Consider the sentence

$$\forall x\,(x \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(x), x)).$$

Since $\mathfrak{C}$ interprets $=$ as identity, the atomic clause makes an equation $t_1 = t_2$ true in it exactly when the two terms name the same person.

By the clause for the universal quantifier, it is true in $\mathfrak{C}$ if, for each of the five people d, the formula $x \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(x), x)$ is true under $s[d/x]$. Figure 6.6 has one row for each person. Every row is true, so the sentence is true. Without the antecedent the Thorne row becomes false, so $\forall x\,\mathsf{Richer}(\mathsf{boss}(x), x)$ is false in $\mathfrak{C}$, with Thorne as the counterexample. These are Harriet's two queries again, now as two applications of a definition.

The Thorne row is the one people query: why is the implication *true* there, when its consequent is false? Read the sentence as a rule of the company: if you are not the chief executive, you earn less than your boss. A rule of that form is broken only by someone it applies to, and it does not apply to Thorne, so Thorne cannot break it. The clause for $\to$ says exactly this: an implication fails only when its antecedent is true and its consequent false.

Example 6.26 (the company, cont'd: evaluating a $\forall\exists$-sentence)**.** The sentence

$$\forall x\,(\mathsf{Mgr}(x) \to \exists y\,(\mathsf{boss}(y) = x \wedge y \neq x))$$

says that every manager manages someone other than themselves.

- For the three people who are not managers, the implication is true at once.
- For Thorne, the clause for the existential quantifier asks for one d whose boss is Thorne and who is not Thorne; Finch will do.
- For Finch, Bell will do.

So the sentence is true in $\mathfrak{C}$.

The evidence has a characteristic shape: for the universal quantifier we examined every row, and for each existential quantifier we produced one witness.

On a finite structure the definition is also an algorithm: to decide whether a sentence is true, unfold it clause by clause. Each quantifier clause runs

through the domain, so a sentence with k nested quantifiers, on a domain of n elements, needs about n^k cases. On an infinite domain, such as the counting numbers, the quantifier clauses ask about infinitely many elements. The definition still makes sense, since it says what truth *is*, but it no longer tells us how to find it out.

The same sentence can be true in one structure and false in another: truth belongs to the pair of a sentence and a structure, not to the sentence alone.

Example 6.27 (one sentence, several structures)**.** Take the signature with one two-place relation symbol R, and the sentence $\forall x\, \exists y\, R(x, y)$, "everything is R-related to something".

- In the company, with R read as *earns more than*, the sentence is false, since Cole earns more than nobody.
- In the counting numbers, with R read as $<$, it is true, with $n+1$ as the witness for n.
- In a network of three servers linked in a ring it is true.
- In three servers linked one way in a line, a to b and b to c, it is false at c.

The one string receives two verdicts in four structures.

The same clauses cover the simplest formulas of all, those built from propositional letters, and there they reduce to something familiar.

Proposition 6.28 (propositional logic as a special case)**.** *Let φ be a formula built from propositional letters, $\bot$ and the connectives, $\mathfrak{M}$ a structure and s an assignment into $\mathfrak{M}$. Whether $\mathfrak{M}, s \models \varphi$ depends only on which letters $\mathfrak{M}$ makes true.*

Proof. A letter p is a relation symbol with no places, so $[\![p]\!]^{\mathfrak{M}}$ is a subset of $\mathbb{D}^0 = \{()\}$. That set has one element, the empty tuple, so the subset is either empty or everything, and the atomic clause makes p false or true accordingly, under every assignment. The clauses for $\bot$ and the connectives compute the truth of φ from that of its parts. □

So for formulas built from letters a structure amounts to a choice of true or false for each letter, a *valuation*: a row of a truth table. With two letters there are four rows, and the clauses for the connectives compute the familiar tables. For $p \to q$, writing T and F for true and false:

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

Propositional logic is the part of first-order logic in which every relation symbol has no places.

The definition asks for an assignment even when the formula is a sentence, yet a sentence has no free variables, and its truth ought not to depend on what they stand for. If it did, "true in this structure" would mean nothing until somebody chose an assignment, and a question such as whether

everybody earns less than their boss would have no answer. The next lemma removes the worry.

Lemma 6.29 (relevance)**.** *Let s and s' be assignments into $\mathfrak{M}$ with $s(x) = s'(x)$ for every $x \in \mathrm{FV}(\varphi)$. Then $\mathfrak{M}, s \models \varphi$ if and only if $\mathfrak{M}, s' \models \varphi$.*

In words: a formula cares only about the variables it leaves free.

Proof sketch.

- *Atomic formulas.* The value of a term depends only on what the assignment does to its variables.
- *Connectives.* The clauses for the connectives pass the claim straight down.
- $\forall x\, \psi$. The set $\mathrm{FV}(\psi)$ is included in $\mathrm{FV}(\forall x\, \psi) \cup \{x\}$, and $s[d/x]$ and $s'[d/x]$ agree on that set; so, by the induction hypothesis, ψ is true under the one exactly when it is true under the other, for every d.
- $\exists x\, \psi$. The existential case is the same.

Exercise 6.14 asks for the details. □

In particular, the truth of a sentence in $\mathfrak{M}$ does not depend on the assignment. We write $\mathfrak{M} \models \varphi$, read "$\varphi$ is true in the structure $\mathfrak{M}$", when $\mathfrak{M}, s \models \varphi$ for one assignment s, and hence for all.

Substitution was introduced to pass from a claim about anybody to the same claim about somebody in particular. The passage is safe only if $\varphi[t/x]$ says about the thing t names exactly what φ says about x, and the next lemma shows that it does, provided nothing is captured.

Lemma 6.30 (substitution)**.** *If t is substitutable for x in φ, then $\mathfrak{M}, s \models \varphi[t/x]$ if and only if $\mathfrak{M}, s[\mathrm{Val}_s(t)/x] \models \varphi$.*

In words: putting the term t in for x has the same effect as letting x stand for whatever t names. Without the proviso the lemma fails: when a quantifier captures a variable of t, the formula $\varphi[t/x]$ no longer says of the thing t names what φ says of x (Example 6.19 shows a case).

The proof, by induction on φ, is long and unilluminating, and we omit it, as most textbooks do; but the omission is of length, not of difficulty. The case of terms, on which the rest is built, is Exercise 6.15, which also asks where substitutability is used.

One feature of the definition looks too obvious to mention, yet its consequences are subtle and contentious. The clause for negation makes $\neg\varphi$ true in $\mathfrak{M}$ under s exactly when φ is not, so exactly one of φ and $\neg\varphi$ is true, and by the clause for disjunction $\varphi \vee \neg\varphi$ is true in every structure under every assignment, whatever φ says. This property of the definition, that every formula has exactly one of two truth values in a structure under an assignment, is called *bivalence*.

6.4. **Entailment.** Satisfaction answers a question about one structure and one assignment: is this formula true here? Often we want something else. We have no particular system in mind, only requirements that every acceptable system must meet, and we want to know what holds in every system

that meets them: whether everybody earns less than their boss in every table a personnel system could ever accept, say.

A set of sentences that says what any acceptable system must satisfy, written down before any particular system is in view, is called a *theory*.

Definition 6.31 (theory; model)**.** A theory is a set of sentences, and its models are the structures in which all its sentences are true.

Formally, a Σ-*theory* is a set of Σ-sentences. A Σ-structure $\mathfrak{M}$ is a *model of* a Σ-theory Θ, or *satisfies* Θ, written $\mathfrak{M} \models \Theta$, if $\mathfrak{M} \models \psi$ for every $\psi \in \Theta$.

Entailment is defined for sets of formulas, whose members may have free variables and are then true or false only under an assignment. For such a set Γ we write $\mathfrak{M}, s \models \Gamma$ when $\mathfrak{M}, s \models \psi$ for every $\psi \in \Gamma$, and say that $\mathfrak{M}, s$ satisfies Γ. When Γ is a theory the assignment makes no difference, by the relevance lemma, and $\mathfrak{M}, s \models \Gamma$ says just that $\mathfrak{M}$ is a model of Γ.

Remark 6.32 (structures and models)*.* A structure is a mathematical object, and a model is always a model *of* something: the same structure may be a model of one theory and not of another. A structure with an assignment gets no name of its own; we write $\mathfrak{M}, s$ and read $\mathfrak{M}, s \models \varphi$ as "φ is true in $\mathfrak{M}$ under s".

English also uses *model* in the opposite direction. A physicist who calls a set of equations a model of nature means that the equations describe the world. Here the world is the model, and the sentences describing it are the theory.

Definition 6.33 (entailment)**.** A set of formulas entails a formula if the formula is true wherever all the members of the set are, and a formula entailed by the empty set is valid.

Formally, let Γ be a set of Σ-formulas and φ a Σ-formula. Then Γ *entails* φ, written $\Gamma \models \varphi$, if $\mathfrak{M}, s \models \varphi$ for every Σ-structure $\mathfrak{M}$ and every assignment s into $\mathfrak{M}$ with $\mathfrak{M}, s \models \Gamma$.

The formula φ is *valid*, written $\models\varphi$, if $\varnothing \models \varphi$, that is, if $\mathfrak{M}, s \models \varphi$ for every Σ-structure $\mathfrak{M}$ and every assignment s into $\mathfrak{M}$.[5]

When Θ is a theory and φ a sentence, $\Theta \models \varphi$ says that φ is true in every model of Θ. Validity is entailment by the empty theory, of which every structure is a model, and it is rarely what we want. A claim such as "nobody but the person at the top is their own boss" is not valid, and nobody expects it to be: it holds, if it holds, because of the theory a system satisfies. The three notions differ in what they examine, as Figure 6.7 shows.

Remark 6.34 (one symbol, three jobs)*.* The symbol $\models$ now does three jobs, and what stands on its left says which.

- With a structure and an assignment on the left, $\mathfrak{M}, s \models \varphi$ says that φ is true in $\mathfrak{M}$ under s, and $\mathfrak{M}, s \models \Gamma$ that every member of Γ is.
- With a structure alone, $\mathfrak{M} \models \varphi$ says that the sentence φ is true in it, and $\mathfrak{M} \models \Theta$ that $\mathfrak{M}$ is a model of the theory Θ.

[5]This is Tarski's definition of consequence (Tarski, 1936); the History tells how it answered Carnap's.

Notion	Means	Examined
$\mathfrak{M}, s \models \varphi$	φ is true in $\mathfrak{M}$ under s	one $\mathfrak{M}, s$
$\Gamma \models \varphi$	$\mathfrak{M}, s \models \varphi$ for every $\mathfrak{M}, s$ with $\mathfrak{M}, s \models \Gamma$	those satisfying Γ
$\models \varphi$	$\mathfrak{M}, s \models \varphi$ for every $\mathfrak{M}, s$	all

FIGURE 6.7. Three semantic notions.

- With a set of formulas on the left (a list such as ψ, χ abbreviates the set), or nothing, $\Gamma \models \varphi$ and $\models \varphi$ speak of every structure and every assignment at once.

So $\mathfrak{M} \models \varphi$ is a fact about one structure, and $\{\psi\} \models \varphi$ is a fact about all of them.

Example 6.35 (the company, cont'd: valid, and not valid)**.** The sentence $\forall x\, \mathsf{Eng}(x) \to \mathsf{Eng}(\mathsf{ceo})$ is valid: in every structure, if every element is in $[\![\mathsf{Eng}]\!]^{\mathfrak{M}}$, then so is $[\![\mathsf{ceo}]\!]^{\mathfrak{M}}$.

The sentence

$$\forall x\, \mathsf{Richer}(\mathsf{boss}(x), x)$$

is not valid, since $\mathfrak{C}$ falsifies it, but it is true in some structures: in the counting numbers, with ceo read as 0, boss as s, Richer as $>$, and Eng and Mgr as anything at all, everyone earns less than their boss. It is true in no finite company whose Richer comes from salaries, since the best-paid person cannot earn less than her boss.

A sentence true in $\mathfrak{C}$ need not be valid, however natural it sounds. The sentence $\exists x\, \mathsf{Eng}(x)$ is true in $\mathfrak{C}$ and false in every company without engineers.

One relation symbol needs a theory of its own. A structure may interpret $=$ as any set of pairs, but a system that compares names relies on two facts about identity: everything is equal to itself, and equals may be put for equals. Written as sentences, these facts are the equality axioms.

Definition 6.36 (equality axioms)**.** The equality axioms of a signature say, in its own symbols, that $=$ behaves as identity.

Formally, let Σ be a signature containing the two-place relation symbol $=$. The *equality axioms* of Σ are the Σ-sentences

(Eq1) $\forall x\, (x = x)$;

(Eq2) $\forall z_1 \cdots \forall z_k\, (x = y \to (\varphi \to \varphi[y/x]))$, for every Σ-formula φ and all variables x and y such that y is substitutable for x in φ, where $z_1, \ldots, z_k$ are the free variables of $x = y \to (\varphi \to \varphi[y/x])$.

We write Eq_Σ for the set of them.

The list does not mention symmetry or transitivity, and need not: because (Eq2) allows every formula φ, both follow from (Eq1) and (Eq2) (Exercise 6.8).

A structure that interprets $=$ as identity is *normal*. Every normal structure satisfies Eq_Σ. The axiom (Eq1) holds because everything is identical to itself, and each instance of (Eq2) holds by the substitution lemma, since

under an assignment that gives x and y the same value, φ and $\varphi[y/x]$ have the same truth value.

Example 6.37 (the company, cont'd: the theory of the personnel system)**.** Harriet's personnel system enforces, among others, two rules that can be written in the company signature:

$$\mathrm{R}_1 : \ \forall x \, \neg \mathsf{Richer}(x, x),$$
$$\mathrm{R}_2 : \ \forall x \, (x \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(x), x)).$$

Nobody earns more than themselves, and everyone but the chief executive earns less than their boss.

The system also relies, without saying so, on the equality axioms $\mathrm{Eq}_{\Sigma_\mathrm{C}}$ of the company signature, which its query engine uses whenever it compares two names. One instance of (Eq2) says that whoever equals someone who earns more than w also earns more than w:

$$\mathrm{E}_{\mathsf{Richer}} : \ \forall u \, \forall v \, \forall w \, (u = v \to (\mathsf{Richer}(u, w) \to \mathsf{Richer}(v, w))),$$

obtained from the formula $\mathsf{Richer}(u, w)$ with u and v for x and y.

Let Θ be the set consisting of R_1, R_2 and $\mathrm{Eq}_{\Sigma_\mathrm{C}}$: the theory of the personnel system, in the company signature. A normal structure satisfies Θ exactly when both rules hold in it, that is, exactly when it is a table that the personnel system would accept as a possible state of the company. An entailment $\Theta \models \varphi$ says that φ holds in every state the system can be in, which is what the finance director asked.

Remark 6.38 (equality as a logical symbol). Many presentations make $=$ a logical symbol instead of a relation symbol, with two rules of its own: (=I) concludes $t = t$ from nothing, and (=E) concludes $\varphi[t_2/x]$ from $t_1 = t_2$ and $\varphi[t_1/x]$. The two treatments derive the same formulas from the same hypotheses. Each equality axiom is derivable with the rules, using ($\to$I) and ($\forall$I), and each use of a rule can be replaced by an instance of (Eq1) or (Eq2) followed by ($\forall$E) and ($\to$E). Keeping $=$ as a relation symbol leaves the logic, and every proof about it, unchanged.

Example 6.39 (the company, cont'd: an entailment)**.** Sharpened a little, the finance director's question is whether the theory guarantees that only the chief executive is her own boss:

$$\Theta \ \models \ \forall x \, (\mathsf{boss}(x) = x \to x = \mathsf{ceo}).$$

It does.

Let $\mathfrak{M}$ be a model of Θ, and let d be an element of its domain with $(\llbracket \mathsf{boss} \rrbracket^{\mathfrak{M}}(d), d) \in \llbracket = \rrbracket^{\mathfrak{M}}$, and suppose, for a contradiction, that $(d, \llbracket \mathsf{ceo} \rrbracket^{\mathfrak{M}}) \notin \llbracket = \rrbracket^{\mathfrak{M}}$. Then

$(\llbracket \mathsf{boss} \rrbracket^{\mathfrak{M}}(d), d) \in \llbracket \mathsf{Richer} \rrbracket^{\mathfrak{M}}$	by R_2,
$(d, d) \in \llbracket \mathsf{Richer} \rrbracket^{\mathfrak{M}}$	by $\mathrm{E}_{\mathsf{Richer}}$, with $\llbracket \mathsf{boss} \rrbracket^{\mathfrak{M}}(d)$, d, d for u, v, w,

which contradicts R_1.

Hence $(d, \llbracket \mathsf{ceo} \rrbracket^{\mathfrak{M}}) \in \llbracket = \rrbracket^{\mathfrak{M}}$, and since $\mathfrak{M}$ and d were arbitrary, the entailment holds.

Example 6.40 (the company, cont'd: a non-entailment)**.** Does the theory guarantee that no engineer is a manager, that is, does $\Theta \models \forall x\,(\mathsf{Eng}(x) \to \neg\mathsf{Mgr}(x))$ hold? To show that an entailment *fails*, one structure suffices: a model of Θ in which the sentence is false.

Take the normal structure with domain {Ada}, in which

- ceo names Ada;
- boss sends Ada to Ada;
- Eng and Mgr are both {Ada};
- Richer is empty.

Then:

- R_1 holds because Richer is empty;
- R_2 holds because the only person is the chief executive, so its antecedent is never true;
- the equality axioms hold because the structure is normal;
- Ada is an engineer who is a manager.

The entailment therefore fails. The one-person company is a model of Θ in which the sentence is false, and so it is the counter-model that shows the failure.

The two kinds of answer are not on an equal footing. A non-entailment comes with a certificate, a structure, which anyone can check by evaluating sentences in it. An entailment comes with an argument, "let $\mathfrak{M}$ be a model of the theory and d an element", which is convincing but cannot be checked by running anything: it speaks of every model of the theory, of every size, including the infinite ones, and no loop runs over all of those (Examples 6.39 and 6.40 show one of each).

6.5. **Soundness and Incompleteness of** NJ**.** How can we certify that $\Gamma \models \varphi$? Listing the structures that satisfy Γ and checking each one is out of the question: there are infinitely many, and most of them are infinite. What we want instead is a *proof*, a finite object showing that the commitments made by Γ yield φ, each step of which a program can check. Part 1 supplied one, Gentzen's NJ. We write $\Gamma \vdash_{\mathrm{NJ}} \varphi$ when some derivation in NJ has conclusion φ and all its open hypotheses in Γ.

Two notions of consequence are now available, $\Gamma \models \varphi$ and $\Gamma \vdash_{\mathrm{NJ}} \varphi$. The first quantifies over all structures and assignments, the second asks for one finite object, and they can be related in two ways:

- *Soundness*: if $\Gamma \vdash_{\mathrm{NJ}} \varphi$, then $\Gamma \models \varphi$.
- *Completeness*: if $\Gamma \models \varphi$, then $\Gamma \vdash_{\mathrm{NJ}} \varphi$.

Soundness says that derivations are safe: none leads from true hypotheses to a false conclusion. Completeness says that derivations are enough: whatever follows has one.

Example 6.41 (the company, cont'd: a derivation)**.** The following derivation shows that *nobody but the chief executive is her own boss*:

$$\Theta \;\vdash_{\mathrm{NJ}}\; \forall x\,(x \neq \mathsf{ceo} \to \mathsf{boss}(x) \neq x).$$

In words: let x be anyone, and suppose that x is not the chief executive and that x is her own boss.

- By R_2, the boss of x earns more than x.
- Since the boss of x *is* x, the equality axiom $\mathrm{E}_{\mathsf{Richer}}$ gives that x earns more than x.
- By R_1 that is absurd.

Hence x is not her own boss.

In the tree, a leaf labelled R_1, R_2 or $\mathrm{E}_{\mathsf{Richer}}$ is that hypothesis, and $\neq$ is short for an implication into $\bot$. To save width we write t for the term $\mathsf{boss}(x)$, and since the tree is still too wide for the page, we give it in four pieces. The piece $\mathcal{D}_1$ instantiates $\mathrm{E}_{\mathsf{Richer}}$ three times and uses the hypothesis $t = x$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\mathrm{E}_{\mathsf{Richer}}}{\forall v\,\forall w\,(t = v \to (\mathsf{Richer}(t,w) \to \mathsf{Richer}(v,w)))}\,\forall\mathrm{E}}{\forall w\,(t = x \to (\mathsf{Richer}(t,w) \to \mathsf{Richer}(x,w)))}\,\forall\mathrm{E}}{t = x \to (\mathsf{Richer}(t,x) \to \mathsf{Richer}(x,x))}\,\forall\mathrm{E} \qquad t = x}{\mathsf{Richer}(t,x) \to \mathsf{Richer}(x,x)}\,{\to}\mathrm{E}$$

The piece $\mathcal{D}_2$ uses R_2 and the hypothesis $x \neq \mathsf{ceo}$:

$$\dfrac{\dfrac{\mathrm{R}_2}{x \neq \mathsf{ceo} \to \mathsf{Richer}(t,x)}\,\forall\mathrm{E} \qquad x \neq \mathsf{ceo}}{\mathsf{Richer}(t,x)}\,{\to}\mathrm{E}$$

The piece $\mathcal{E}$ combines the two and meets R_1:

$$\dfrac{\dfrac{\mathrm{R}_1}{\neg\mathsf{Richer}(x,x)}\,\forall\mathrm{E} \qquad \dfrac{\mathcal{D}_1 \qquad\qquad \mathcal{D}_2}{\mathsf{Richer}(t,x) \to \mathsf{Richer}(x,x) \qquad \mathsf{Richer}(t,x)}\,{\to}\mathrm{E}}{\bot}\,{\to}\mathrm{E}$$

Its open hypotheses are R_1, R_2, $\mathrm{E}_{\mathsf{Richer}}$, $x \neq \mathsf{ceo}$ and $t = x$. The last piece discharges the final two and generalises:

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{cc}[x \neq \mathsf{ceo}]^1 & [t = x]^2\end{array}}{\vdots\,\mathcal{E}}}{\bot}\,{\to}\mathrm{I},2}{t \neq x}\,{\to}\mathrm{I},1}{x \neq \mathsf{ceo} \to t \neq x}\,\forall\mathrm{I}}{\forall x\,(x \neq \mathsf{ceo} \to \mathsf{boss}(x) \neq x)}$$

The ($\forall$I) step is legal because, once both suppositions are discharged, the only open hypotheses are the sentences R_1, R_2 and $\mathrm{E}_{\mathsf{Richer}}$.

The sentence the tree proves is not quite the finance director's, $\forall x\,(\mathsf{boss}(x) = x \to x = \mathsf{ceo})$. The two are true in the same structures, but they are different strings, and the tree proves only the first.

A derivation is the certificate we asked for: a finite object, each step of which a program can check against the rules.

Theorem 6.42 (soundness of NJ)**.** *For every set* Γ *of formulas and every formula* φ*, if* $\Gamma \vdash_{\mathrm{NJ}} \varphi$ *then* $\Gamma \models \varphi$*.*

Proof. We prove a stronger statement about every derivation $\mathcal{D}$, by induction on its construction: *in every structure $\mathfrak{M}$ under every assignment s, if all the open hypotheses of $\mathcal{D}$ are true, then so is its conclusion.* The theorem is the case of a derivation whose open hypotheses lie in Γ.

There is one case for each rule, and each case asks the same question: if the premisses of the rule are true wherever their open hypotheses are, is the conclusion true wherever *its* open hypotheses are? Fix a structure $\mathfrak{M}$ and an assignment s, and suppose that the open hypotheses of $\mathcal{D}$ are all true in $\mathfrak{M}$ under s.

- *A hypothesis.* The derivation is a single leaf φ, whose open hypothesis is φ itself, true by supposition.
- *($\to$I).* The derivation $\mathcal{D}$ ends with $\varphi \to \psi$, inferred from a derivation of ψ whose open hypotheses are those of $\mathcal{D}$ and possibly φ. If φ is false under s, then $\varphi \to \psi$ is true by its clause. If φ is true, then every open hypothesis above the step is true, so ψ is true by the induction hypothesis, and so is $\varphi \to \psi$.
- *($\vee$E).* The open hypotheses of the derivation of $\varphi \vee \psi$ are among those of $\mathcal{D}$, so $\varphi \vee \psi$ is true, and hence φ or ψ is true. If φ is, every open hypothesis of the derivation of χ from $[\varphi]$ is true, and so is χ. If ψ is, the same argument applies to the other derivation.
- *($\forall$I).* The derivation $\mathcal{D}$ ends with $\forall x\, \varphi$, inferred from a derivation $\mathcal{D}'$ of $\varphi[y/x]$, where y is free in no open hypothesis and not free in $\forall x\, \varphi$. We must show that φ is true under $s[d/x]$ for every d. Let d be an element of the domain and $s' = s[d/y]$. The open hypotheses do not mention y, so

they remain true under s'	by the relevance lemma,
$\varphi[y/x]$ is true under s'	by the induction hypothesis,
φ is true under $s'[d/x]$	by the substitution lemma.

  That assignment differs from $s[d/x]$ at most at y, and y is not free in φ unless it is x itself, so by relevance φ is true under $s[d/x]$.

  The side condition was used to keep the open hypotheses true when y was moved to d. Had they said something about y, moving y might have made them false, and the induction hypothesis would have said nothing.
- *($\exists$E).* The derivation of $\exists x\, \varphi$ has its open hypotheses among those of $\mathcal{D}$, so $\exists x\, \varphi$ is true, and there is an element d with φ true under $s[d/x]$. Let $s' = s[d/y]$. The other open hypotheses above χ do not mention y, and y is not free in χ, so

$\varphi[y/x]$ is true under s'	as in the previous case,
the other open hypotheses are true under s'	by relevance,
χ is true under s'	by the induction hypothesis,
χ is true under s	by relevance.

  The first step uses that y is not free in $\exists x\, \varphi$, unless it is x.

The remaining rules, ($\to$E), ($\wedge$I), ($\wedge$E), ($\vee$I), ($\bot$E), ($\forall$E) and ($\exists$I), discharge nothing and have no side condition on a variable, and each is checked by reading its clause in the definition of satisfaction, with the substitution lemma for ($\forall$E) and ($\exists$I). Exercise 6.13 treats two of them; the others are similar. □

Remark 6.43 (truth along a derivation). Soundness does not promise that every line of a derivation is true. A derivation reasons from suppositions, and a supposition may be false. The promise is that a line can be false only if one of its open hypotheses is. Listing the lines with their open hypotheses and their truth values in a structure under an assignment makes the promise visible, and it exposes a broken rule: the bad step is the first line that is false while all its open hypotheses are true.

Example 6.44 (reasoning from a supposition)**.** From $p \vee q$ and $\neg p$ the rule ($\vee$E) derives q:

$$\dfrac{p \vee q \qquad \dfrac{\dfrac{\neg p \qquad [p]^1}{\bot}\,{\to}\mathrm{E}}{q}\,{\bot}\mathrm{E} \qquad [q]^1}{q}\,{\vee}\mathrm{E},1$$

Take a structure that makes both hypotheses true, so that p is false and q true. The branch on the left passes through two false lines, the supposition p and the line $\bot$, and each has a false open hypothesis, the supposition p itself. Nothing has gone wrong. That branch reasons about a case this structure rules out, and ($\vee$E) takes from it only the conclusion q, which the other branch also reaches.

Is NJ also complete? It is not. To show that a sentence has no derivation in NJ we use two properties of the rules.

Fact 6.45 (disjunction and existence properties).

- If $\vdash_{\mathrm{NJ}} \varphi \vee \psi$, then $\vdash_{\mathrm{NJ}} \varphi$ or $\vdash_{\mathrm{NJ}} \psi$.
- If $\vdash_{\mathrm{NJ}} \exists x\, \varphi$, then $\vdash_{\mathrm{NJ}} \varphi[t/x]$ for some term t.

In words: a proof in NJ, from no hypotheses, of "φ or ψ" says which of the two holds, and a proof of "there is an x such that φ" names a term for x.

The reason fits in three sentences. Every derivation can be rewritten, step by step, into one with no *detours*, that is, no introduction rule followed at once by the elimination rule for the same symbol. A derivation with no detours and no open hypotheses ends with an introduction rule. For $\varphi \vee \psi$ that rule is ($\vee$I), and the tree above it derives φ or derives ψ; for $\exists x\, \varphi$ it is ($\exists$I), and the tree above it derives $\varphi[t/x]$ for a particular term t.[6]

Theorem 6.46 (incompleteness of NJ)**.** *For a propositional letter p, the sentence $p \vee \neg p$ is valid, but $\nvdash_{\mathrm{NJ}} p \vee \neg p$. Hence there are sentences φ with $\models \varphi$ and $\nvdash_{\mathrm{NJ}} \varphi$, and the converse of soundness fails for* NJ*.*

Proof. The sentence is valid: in every structure under every assignment p is true or false, and in the second case $\neg p$ is true. Now suppose that $\vdash_{\mathrm{NJ}} p \vee \neg p$.

[6] Prawitz (1965) is the classic source.

By the disjunction property, NJ derives p or derives $\neg p$, and by soundness the disjunct it derives is valid. Neither is:

- in a structure that makes p true, $\neg p$ is false;
- in one that makes p false, p is false.

□

The proof used soundness to pass from a derivation to validity, and two structures to take validity away. It used nothing about p except that a structure can make it true and another false, so the same proof works for every sentence φ of which that holds: NJ does not derive $\varphi \vee \neg\varphi$.

The sentence $\varphi \vee \neg\varphi$ is the *law of the excluded middle*: every structure obeys it, and NJ lacks it. The rules come close, since they derive $\neg\neg(\varphi\vee\neg\varphi)$, but no rule leads from $\neg\neg\varphi$ to φ. The gap is no isolated freak, and the next example meets it again.

Remark 6.47 (two senses of incompleteness). The word *incompleteness* in the title of this section describes a proof system: a system is complete if it derives everything that is entailed, and incomplete if some entailment has no derivation. Chapter 8 uses it of theories instead: a theory is incomplete when some truth about its subject matter cannot be derived from it.

Example 6.48 (the company, cont'd: one negation too many)**.** The finance director's sentence is entailed by Θ (Example 6.39). Can NJ derive it?

$$\Theta \;\vdash_{\mathrm{NJ}}\; \forall x\,(\mathsf{boss}(x) = x \to x = \mathsf{ceo})\;?$$

The argument of Example 6.39 went:

- *either x is the chief executive or not*;
- if so, done;
- if not, absurdity.

The second half is the derivation of Example 6.41. The rules do not provide the first half, the *either . . . or not*. They do provide the following:

- suppose $x \neq \mathsf{ceo}$;
- then, as in Example 6.41, absurdity;
- hence $\neg(x \neq \mathsf{ceo})$, which is $\neg\neg(x = \mathsf{ceo})$.

That is one negation too many. In every structure $\neg\neg(x = \mathsf{ceo})$ is true exactly when $x = \mathsf{ceo}$, but no rule of NJ makes the last step, and no cleverer derivation finds a way round.[7]

6.6. **Law of the Excluded Middle.** The gap in NJ is a principle that every structure obeys: of a claim and its denial, one is true, and there is no third option in the middle. This is the *law of the excluded middle* (LEM), and it is more than two thousand years old.[8] In a structure it holds because

[7] Showing that no derivation exists needs a different kind of structure, which Section 6.8 describes.

[8] Aristotle argues for it in Book Γ of the *Metaphysics*. In chapter 9 of *On Interpretation* he also raises the first recorded doubt about bivalence: is "there will be a sea battle tomorrow" already true or false today?

the clause for negation makes exactly one of φ and $\neg\varphi$ true (bivalence); in a proof system it becomes a rule with no premisses, for every formula φ:

$$\frac{}{\varphi \vee \neg\varphi}\ \text{LEM}$$

The law is not the only principle that NJ lacks. Three more fall into the gap, each a schema with an instance for every choice of formulas:

$$\neg\neg\varphi \to \varphi,$$
$$((\varphi \to \psi) \to \varphi) \to \varphi,$$
$$\neg\forall x\, \neg\varphi \to \exists x\, \varphi.$$

- The first, *double negation*, says that what is not false is true;
- the second, *Peirce's law*,[9] says much the same without negation;
- the third says that if not everything lacks the property φ, then something has it.

Every instance of each is valid, as a check against the clauses shows, but NJ does not derive every instance: already $\neg\neg p \to p$, $((p \to q) \to p) \to p$ and $\neg\forall x\, \neg P(x) \to \exists x\, P(x)$, for propositional letters p, q and a relation symbol P, have no derivation (Exercise 6.10).

With the law added, all three become derivable. Here is double negation:

$$\dfrac{\dfrac{}{\varphi \vee \neg\varphi}\ \text{LEM} \qquad \begin{array}{c}[\varphi]^2\end{array} \qquad \dfrac{\dfrac{[\neg\neg\varphi]^1 \qquad [\neg\varphi]^2}{\bot}\ \to\text{E}}{\varphi}\ \bot\text{E}}{\dfrac{\varphi}{\neg\neg\varphi \to \varphi}\ \to\text{I},1}\ \vee\text{E},2$$

We write (DN) for this derivation used as a step. The system with the law added is *classical logic*.

Definition 6.49 (NK)**.** The system NK is NJ with the law of the excluded middle added as a rule with no premisses.

Formally, the derivations of NK are built by the rules of NJ together with the rule (LEM).

Every derivation of NJ is a derivation of NK, so $\Gamma \vdash_{\text{NJ}} \varphi$ implies $\Gamma \vdash_{\text{NK}} \varphi$. The letters are J for *intuitionistisch* and K for *klassisch*.[10] One rule with no premisses looks like a small addition. The next example shows what it buys.

Example 6.50 (the company, cont'd: a derivation in NK)**.** The finance director's sentence is derivable in NK: $\Theta \vdash_{\text{NK}} \forall x\, (\mathsf{boss}(x) = x \to x = \mathsf{ceo})$. The derivation is the argument of Example 6.39 written as a tree, and most

[9] From Peirce (1885). Peirce, working at the same time as Frege and independently of him, also introduced quantifiers.

[10] Gentzen named the two systems so in the paper that introduced them, Gentzen (1934). The usual explanation of the J is that German typography of the time treated the capitals I and J as one letter. Gentzen's NK adds excluded middle as an axiom; a rule with no premisses, as here, comes to the same thing.

of it is already built: the piece $\mathcal{E}$ of Example 6.41 derives $\bot$ from $x \neq \mathsf{ceo}$ and $\mathsf{boss}(x) = x$, with R_1, R_2 and $\mathrm{E}_{\mathsf{Richer}}$. The new step splits on $x = \mathsf{ceo}$:

$$\dfrac{\dfrac{\dfrac{}{x = \mathsf{ceo} \vee x \neq \mathsf{ceo}}\,\mathrm{LEM} \qquad \dfrac{\begin{array}{c}[x \neq \mathsf{ceo}]^2 \\ [\mathsf{boss}(x) = x]^1 \\ \mathcal{E} \\ \bot\end{array}}{x = \mathsf{ceo}}\,\bot\mathrm{E} \qquad [x = \mathsf{ceo}]^2}{x = \mathsf{ceo}}\,\vee\mathrm{E},\,2}{\mathsf{boss}(x) = x \to x = \mathsf{ceo}}\,\to\mathrm{I},\,1}{\forall x\,(\mathsf{boss}(x) = x \to x = \mathsf{ceo})}\,\forall\mathrm{I}$$

The left case of the ($\vee$E) needs no work. The right case is the old derivation, followed by a step out of the contradiction, a step that the split makes worth taking. The ($\forall$I) step is legal for the same reason as before: only R_1, R_2 and $\mathrm{E}_{\mathsf{Richer}}$ remain open.

Theorem 6.51 (soundness of NK)**.** *For every set* Γ *of formulas and every formula* φ*, if* $\Gamma \vdash_{\mathrm{NK}} \varphi$ *then* $\Gamma \models \varphi$*.*

Proof. The soundness proof for NJ goes through with one more case: a leaf $\varphi \vee \neg\varphi$, with no open hypotheses, must be true in every structure under every assignment. By bivalence it is: the clause for negation makes one of φ and $\neg\varphi$ true. □

So far, so good: NK is sound, and it reaches sentences that NJ missed. But is one more rule enough? Soundness is half of the connection between $\vdash_{\mathrm{NK}}$ and $\models$. The other half asks whether NK derives everything that follows, or whether further laws are missing.

6.7. Completeness of NK.

Soundness puts derivability inside entailment: whatever NK derives from Γ is true in every model of Γ. For NJ the converse failed, and the law of the excluded middle was the missing step. For NK the converse holds.

Theorem 6.52 (completeness[11])**.** *Let* Σ *be a countable signature,* Γ *a set of* Σ*-sentences and* φ *a* Σ*-sentence. If* $\Gamma \models \varphi$*, then* $\Gamma \vdash_{\mathrm{NK}} \varphi$*.*

In words: whatever is true in every model of a theory can be derived from it in NK.

To show that $\Gamma \models \varphi$ implies $\Gamma \vdash_{\mathrm{NK}} \varphi$, we prove the contrapositive: if $\Gamma \nvdash_{\mathrm{NK}} \varphi$, then some model of Γ makes φ false. We are handed a failure to derive and must build a structure from it. Two definitions name what the construction aims at.

Definition 6.53 (consistency; maximal consistency)**.** A consistent set is one from which the rules cannot derive absurdity, and a maximal consistent set is a consistent set that takes a side on every sentence.

Formally, a set Γ of Σ-sentences is *consistent* if $\Gamma \nvdash_{\mathrm{NK}} \bot$, and *inconsistent* otherwise. It is *maximal consistent* if

- it is consistent, and

[11] Gödel (1930) proved it for a Hilbert-style system. The proof given here is Henkin (1949)'s, and the History tells how both were found.

- for every Σ-sentence ψ, either $\psi \in \Gamma$ or $\neg\psi \in \Gamma$.

Keep the two words apart. *Consistent* is about derivations: no derivation leads from the set to $\bot$. Having a model is about structures: some structure makes the whole set true.

Soundness gives one direction at once, since a set with a model cannot derive $\bot$, which is true in no structure. The proof below gives the other, and in this vocabulary that is the whole of the completeness theorem: *every consistent set of sentences has a model.* (Applied to $\Gamma \cup \{\neg\varphi\}$, this gives the theorem as stated, by Step 1 below.) The proof has five steps.

1. Show that $\Gamma \cup \{\neg\varphi\}$ is consistent.
2. Add new constants, and sentences that make each of them a witness for an existential sentence.
3. Extend the result to a maximal consistent set Δ.
4. Build a structure from the closed terms of the language, so that its elements are pieces of syntax.
5. Prove that a sentence is true in this structure exactly when it belongs to Δ. Since Γ and $\neg\varphi$ are in Δ, the structure is the one we want.

Steps 2 and 4 are needed only for the quantifiers, so we first carry out the propositional case in full and then extend it.

Step 1.

Lemma 6.54 (refuting the negation)**.** *If $\Gamma \nvdash_{\mathrm{NK}} \varphi$, then $\Gamma \cup \{\neg\varphi\}$ is consistent.*

Proof. If $\Gamma \cup \{\neg\varphi\}$ derived $\bot$, then Γ would derive $\neg\neg\varphi$ by ($\to$I), and φ by (DN). □

Excluded middle enters the propositional proof here and nowhere else, through (DN), and here the proof breaks down for NJ. With Γ empty and φ the sentence $p \vee \neg p$, the lemma fails for NJ: the system does not derive $p \vee \neg p$, yet $\{\neg(p \vee \neg p)\}$ is inconsistent, by the tree that derived $\neg\neg(p \vee \neg p)$.

Steps 3 to 5, propositional case.

Lemma 6.55 (Lindenbaum)**.** *Over a countable signature, every consistent set of sentences is contained in a maximal consistent set.*

In words: a set of sentences that does not contradict itself can be enlarged until it takes a side on every question, and still not contradict itself.

Proof. List all the sentences of the signature as $\psi_0, \psi_1, \psi_2, \ldots$; they can be listed, because they are finite strings over a countable alphabet. Decide each sentence in turn: let $\Delta_0 = \Gamma$ and

$$\Delta_{n+1} \coloneqq \begin{cases} \Delta_n \cup \{\psi_n\} & \text{if that set is consistent,} \\ \Delta_n \cup \{\neg\psi_n\} & \text{otherwise.} \end{cases}$$

Let $\Delta = \bigcup_{n\in\mathbb{N}} \Delta_n$.

- Every sentence or its negation belongs to Δ.

- Each Δ_n is consistent: if both $\Delta_n \cup \{\psi_n\}$ and $\Delta_n \cup \{\neg\psi_n\}$ derived $\bot$, then by ($\to$I) the set Δ_n would derive both $\neg\psi_n$ and $\neg\neg\psi_n$, and hence $\bot$.
- Finally, Δ is consistent, because a derivation of $\bot$ from Δ uses finitely many hypotheses, all of which lie in some Δ_n.

□

The proof is not an algorithm. At each step it asks whether a set of sentences is consistent, and nothing in it says how that question is to be answered; it needs only that the question has an answer. The proof shows that Δ exists without constructing it. Nor is Δ unique: another order of the list can give another maximal consistent set. The lemma promises *a* maximal consistent set, not a particular one, and the completeness proof needs only one.

The next lemma shows, one connective at a time, that a maximal consistent set behaves like the set of truths of a single structure.

Lemma 6.56 (maximal consistent sets)**.** *Let Δ be maximal consistent. For all sentences ψ and χ:*

(i) if $\Delta \vdash_{\mathrm{NK}} \psi$, then $\psi \in \Delta$; in particular, $\bot \notin \Delta$;
(ii) $\psi \wedge \chi \in \Delta$ if and only if $\psi \in \Delta$ and $\chi \in \Delta$;
(iii) $\psi \vee \chi \in \Delta$ if and only if $\psi \in \Delta$ or $\chi \in \Delta$;
(iv) $\psi \to \chi \in \Delta$ if and only if $\psi \notin \Delta$ or $\chi \in \Delta$.

Proof.
- *(i)* If Δ derives ψ but $\psi \notin \Delta$, then $\neg\psi \in \Delta$, and ($\to$E) derives $\bot$ from Δ. And $\bot \notin \Delta$, since Δ is consistent.
- *(iii)* If $\psi \vee \chi \in \Delta$ and neither ψ nor χ is, then $\neg\psi$ and $\neg\chi$ are, and ($\vee$E) derives $\bot$ from the three. The converse follows from ($\vee$I) and (i).
- *(iv)* If $\psi \to \chi$ and ψ are in Δ, then Δ derives χ by ($\to$E), so $\chi \in \Delta$ by (i). Conversely:
  - if $\chi \in \Delta$, then Δ derives $\psi \to \chi$ by ($\to$I), discharging nothing;
  - if $\psi \notin \Delta$, then $\neg\psi \in \Delta$, from which, supposing ψ, ($\to$E) gives $\bot$ and ($\bot$E) gives χ, so Δ again derives $\psi \to \chi$.

  In either case (i) puts $\psi \to \chi$ in Δ.

Part (ii) is Exercise 6.16. □

For a signature of propositional letters a structure amounts to a valuation, and the obvious one serves: make a letter true exactly when it belongs to Δ.

Lemma 6.57 (truth lemma, propositional)**.** *Let Δ be maximal consistent, and let v make a letter p true exactly when $p \in \Delta$. For every formula ψ built from letters, $\bot$ and the connectives, ψ is true under v if and only if $\psi \in \Delta$.*

Proof. By induction on ψ.

- *A letter.* The claim is the definition of v.
- $\bot$. Neither side holds, by part (i) of the previous lemma.

- $\psi \vee \chi$.

	the formula $\psi \vee \chi$ is true under v	
$\iff$	ψ or χ is	
$\iff$	$\psi \in \Delta$ or $\chi \in \Delta$	by the induction hypothesis
$\iff$	$\psi \vee \chi \in \Delta$	by (iii) above.

- $\psi \wedge \chi$ *and* $\psi \to \chi$. The cases of $\wedge$ and $\to$ are the same, with parts (ii) and (iv) of the lemma.

□

This completes the propositional case. If $\Gamma \nvdash_{\mathrm{NK}} \varphi$, then

$\Gamma \cup \{\neg\varphi\}$ is consistent	(Step 1),
it is contained in a maximal consistent set Δ	(Lindenbaum's lemma),
the valuation v of Δ makes all of Δ true	(the truth lemma),

in particular every member of Γ and $\neg\varphi$. So the structure given by v satisfies Γ and makes φ false, and $\Gamma \not\models \varphi$.

Example 6.58 (Lindenbaum's construction on three letters)**.** Let $\Gamma = \{p \vee q,\ q \to r\}$ and $\varphi = r$, and ask whether Γ entails r. Start the construction from $\Delta_0 = \{p \vee q,\ q \to r,\ \neg r\}$, which is consistent, and suppose that the list of sentences begins p, q, r.

- Adding p leaves the set consistent, so p goes in.
- Adding q does not, since q and $q \to r$ give r and $\neg r$ is present, so $\neg q$ goes in.
- Adding r meets $\neg r$, so $\neg r$ goes in again.

The valuation read off makes p true and q and r false, and it does make $p \vee q$ and $q \to r$ true and r false. The valuation was assembled, one letter at a time, from what the hypotheses could not rule out, and it shows that the entailment fails.

The first step asked whether $\{p \vee q, q \to r, \neg r, p\}$ is consistent, and we answered yes because this very valuation makes all of it true: the construction itself only asks.

Step 2: Henkin witnesses. Two things go wrong when the language has quantifiers, and the two new steps repair them. The first is that nothing yet says what the elements of the structure should be; the propositional case needed none. The answer is the boldest idea in the proof: the elements are the closed terms themselves. A constant c names the element c, the term $\mathsf{boss}(c)$ names the element $\mathsf{boss}(c)$, and so on, so that the structure is made of its own names.

The second problem is that a maximal consistent set may contain $\exists x\, \psi$ without any instance $\psi[t/x]$: the closed terms may not name enough elements.

Example 6.59 (the company, cont'd: an engineer with no name)**.** In $\mathfrak{C}$ every closed term names Thorne, since ceo is Thorne and Thorne is her own boss. The set of sentences true in $\mathfrak{C}$, which is maximal consistent, therefore

contains $\exists x\, \mathsf{Eng}(x)$ and also $\neg \mathsf{Eng}(t)$ for every closed term t. A structure built from the closed terms would contain no engineer, and the truth lemma would fail. This is the company's version of Remark 6.24.

We must ensure that there are enough names and that every existential sentence in Δ has one of them as a witness.

Definition 6.60 (witness property)**.** A set of sentences has the witness property if every existential sentence in it comes with a named instance.

Formally, a set Δ of Σ-sentences has the *witness property* if, for every sentence $\exists x\, \psi \in \Delta$, there is a constant symbol c of Σ with $\psi[c/x] \in \Delta$.

Lemma 6.61 (generalisation on a new constant)**.** *If $\Gamma \vdash_{\mathrm{NK}} \psi[c/x]$, where the constant c occurs neither in Γ nor in ψ, then $\Gamma \vdash_{\mathrm{NK}} \forall x\, \psi$.*

Proof sketch. In a derivation of $\psi[c/x]$, replace c everywhere by a variable y that occurs nowhere in the derivation. No rule mentions c, so the result is still a derivation, now of $\psi[y/x]$ from hypotheses in Γ, which do not mention y. One step of ($\forall$I) completes it. □

Lemma 6.62 (Henkin witnesses)**.** *Let Σ be countable and Γ a consistent set of Σ-sentences.*

- *Let Σ^+ be Σ with new constant symbols $c_0, c_1, c_2, \ldots$ added.*
- *List the existential Σ^+-sentences as $\exists x_0\, \psi_0, \exists x_1\, \psi_1, \exists x_2\, \psi_2, \ldots$.*
- *For each n let θ_n be the sentence $\exists x_n\, \psi_n \to \psi_n[d_n/x_n]$, where d_n is the first of the new constants that occurs neither in ψ_n nor in any of $\theta_0, \ldots, \theta_{n-1}$.*

Then $\Gamma \cup \{\theta_0, \theta_1, \ldots\}$ is consistent.

Proof. First, Γ remains consistent when derivations may use the new constants: replacing each new constant in a derivation of $\bot$ from Γ by a variable that occurs nowhere in it yields a derivation of $\bot$ from Γ over Σ, since no rule mentions a constant. Since a derivation of $\bot$ uses finitely many hypotheses, it now suffices to show that if $\Gamma_n = \Gamma \cup \{\theta_0, \ldots, \theta_{n-1}\}$ is consistent, then so is $\Gamma_n \cup \{\theta_n\}$.

Suppose not. Then Γ_n derives $\neg\theta_n$, the denial of an implication, and from the denial of an implication $\chi_1 \to \chi_2$ the rules of NJ alone derive $\neg\chi_2$ and $\neg\neg\chi_1$ (Exercise 6.12). The constant d_n occurs neither in Γ_n nor in ψ_n. So Γ_n derives

$\neg\psi_n[d_n/x_n]$ and $\neg\neg\exists x_n\, \psi_n$	from $\neg\theta_n$,
$\forall x_n\, \neg\psi_n$	by generalisation on a new constant,
$\neg\exists x_n\, \psi_n$	by ($\exists$E) and ($\forall$E),
$\bot$	together with $\neg\neg\exists x_n\, \psi_n$.

Hence Γ_n is inconsistent, contrary to assumption. No use of excluded middle was needed. □

In words: every existential sentence may be given a named witness without risk of contradiction, provided that each name is new.

Remark 6.63 (why the witnesses must be new). The proviso is not fussiness, and two small sets show what it prevents. Let P and Q be one-place relation

symbols and c a constant symbol. The set $\{\exists x\, P(x),\ \neg P(c)\}$ is consistent: every structure with an element of P that c does not name satisfies it. Had the witness for $\exists x\, P(x)$ been the old constant c, the sentence added would be $\exists x\, P(x) \to P(c)$, and with it the set derives $P(c)$ and $\neg P(c)$. A name already in use carries information, and the witness may not have it.

Nor may two existential sentences share a new name. The set

$$\{\exists x\, P(x),\ \exists x\, Q(x),\ \neg\exists x\, (P(x) \wedge Q(x))\}$$

is consistent, since every structure with an element of P and a different element of Q, and nothing in both, satisfies it; give both existential sentences the witness c_0, and the set derives $P(c_0) \wedge Q(c_0)$, and so a contradiction. Exercise 6.17 asks for the details.

The condition is the one on the eigenvariable of ($\exists$E), that it be a name about which nothing has been assumed, in another guise.

Now apply Lindenbaum's lemma, over Σ^+, to $\Gamma \cup \{\neg\varphi\} \cup \{\theta_0, \theta_1, \ldots\}$. The result is a maximal consistent set Δ of Σ^+-sentences with the witness property. Indeed, if $\exists x\, \psi \in \Delta$, then $\exists x\, \psi$ is some $\exists x_n\, \psi_n$, the sentence θ_n is in Δ, and part (iv) of the lemma on maximal consistent sets puts $\psi_n[d_n/x_n]$ in Δ.

Steps 3 to 5, first-order case.

Definition 6.64 (term structure)**.** The term structure of a set of sentences is made of the closed terms, related as the set says they are.

Formally, let Δ be a maximal consistent set of Σ^+-sentences with the witness property. Its *term structure* is the Σ^+-structure $\mathfrak{M}_\Delta = (\mathbb{D}_\Delta, I_\Delta)$ in which

- $\mathbb{D}_\Delta$ is the set of closed Σ^+-terms, and
- for each constant symbol c, each function symbol f of arity n, and each relation symbol R of arity n of Σ^+,

$$\begin{aligned} [\![c]\!]^{\mathfrak{M}_\Delta} &= c, \\ [\![f]\!]^{\mathfrak{M}_\Delta}(t_1, \ldots, t_n) &= f(t_1, \ldots, t_n), \\ [\![R]\!]^{\mathfrak{M}_\Delta} &= \{(t_1, \ldots, t_n) \in \mathbb{D}_\Delta^n : R(t_1, \ldots, t_n) \in \Delta\}. \end{aligned}$$

The domain $\mathbb{D}_\Delta$ is not empty, since the new constants are closed terms. By induction on terms, every closed term t names itself: $\mathrm{Val}_s(t) = t$ under every assignment s. Every element of the domain therefore has a name, and by the substitution lemma $\mathfrak{M}_\Delta \models \forall x\, \psi$ holds exactly when $\mathfrak{M}_\Delta \models \psi[t/x]$ for every closed term t, and $\mathfrak{M}_\Delta \models \exists x\, \psi$ exactly when $\mathfrak{M}_\Delta \models \psi[t/x]$ for some closed term t.

Lemma 6.65 (truth lemma)**.** *For every Σ^+-sentence ψ, $\mathfrak{M}_\Delta \models \psi$ if and only if $\psi \in \Delta$.*

In words: in the structure made of names, a sentence is true exactly when Δ says it is.

Proof. By induction on the number of connectives and quantifiers in ψ. The induction is not on subformulas: $\psi[t/x]$ is not a subformula of $\forall x\, \psi$, but it has one quantifier fewer.

- *Atomic sentences.* The claim is the definition of $[\![R]\!]^{\mathfrak{M}_\Delta}$.
- *Connectives.* Each follows from the induction hypothesis and the matching part of the lemma on maximal consistent sets: $\psi \vee \chi$ is true exactly when ψ or χ is, so exactly when $\psi \in \Delta$ or $\chi \in \Delta$, so, by part (iii), exactly when $\psi \vee \chi \in \Delta$; the cases of $\wedge$ and $\to$ use parts (ii) and (iv) in the same way.
- $\exists x\, \psi$. The existential case is Exercise 6.16.
- $\forall x\, \psi$. If $\forall x\, \psi \in \Delta$, then every instance $\psi[t/x]$

is derivable from Δ	by ($\forall$E)
hence belongs to Δ	
hence is true	by the induction hypothesis;

  so $\forall x\, \psi$ is true.

  Conversely, suppose that $\forall x\, \psi \notin \Delta$. Then

$\neg\forall x\, \psi \in \Delta$	
$\exists x\, \neg\psi \in \Delta$	derivable in NK (Exercise 6.10)
$\neg\psi[c/x] \in \Delta$ for some c	by the witness property
$\psi[c/x] \notin \Delta$	
$\psi[c/x]$ is false	by the induction hypothesis
$\forall x\, \psi$ is false.	

□

Excluded middle has now been used a second time, in the passage from $\neg\forall x\, \psi$ to $\exists x\, \neg\psi$, a relative of the quantifier principle $\neg\forall x\, \neg\varphi \to \exists x\, \varphi$ that NJ could not derive.

Proof of the completeness theorem. Suppose that $\Gamma \nvdash_{\mathrm{NK}} \varphi$.

- Then $\Gamma \cup \{\neg\varphi\}$ is consistent (Step 1).
- Add the witnesses (Step 2) and extend to a maximal consistent set Δ with the witness property (Step 3).
- Its term structure (Step 4) makes every member of Δ true (Step 5), in particular every member of Γ, and $\neg\varphi$.

The new constants occur in neither Γ nor φ, so they can be forgotten: the structure $(\mathbb{D}_\Delta, I_\Delta{\restriction}(\mathcal{C} \cup \mathcal{F} \cup \mathcal{R}))$, the restriction of I_Δ to the symbols of Σ, is a Σ-structure in which the same Σ-sentences are true. It satisfies Γ and makes φ false, so $\Gamma \not\models \varphi$. □

The proof has a slightly impudent quality. Asked to show that a sentence true in every structure has a proof, we turned the sentences themselves into a structure made of words. Nothing in the definition of a structure forbids this, since a domain may be any non-empty set, and a set of strings is a set like any other.

Remark 6.66 (compactness and the Löwenheim–Skolem theorem). The proof yields two further results. If every finite subset of Γ has a model, so does Γ (*compactness*), since a derivation uses only finitely many hypotheses. And the structure the proof builds is made of terms, of which there are only

countably many, so a set of sentences with a model has one whose domain can be listed as $0, 1, 2, \ldots$ (the *Löwenheim–Skolem theorem*). Both say that first-order sentences discriminate less than one might hope.

Example 6.67 (the company, cont'd: the finance director's two answers)**.** Take Γ to be the theory Θ of the personnel system. By soundness and completeness, "true in every model of Θ" and "derivable from Θ in NK" are one relation, so the finance director's question has two answers, and they are the same answer. The tables the system can produce interpret $=$ as identity, and the next remark shows that looking only at such structures loses nothing.

Remark 6.68 (normal structures)*.* Let Θ be a theory that contains the equality axioms of its signature. The term structure the proof builds need not be normal, since Δ may contain $c = d$ for two different constants c and d. Because Θ contains the equality axioms, the defect can be repaired. Write $t \approx t'$ when $(t = t') \in \Delta$. The axioms (Eq1) and (Eq2) make $\approx$ an equivalence relation (Exercise 6.8) that is compatible with the symbols: if $t_i \approx t'_i$ for each i, then

- $f(t_1, \ldots, t_n) \approx f(t'_1, \ldots, t'_n)$;
- $R(t_1, \ldots, t_n) \in \Delta$ exactly when $R(t'_1, \ldots, t'_n) \in \Delta$.

The $\approx$-classes of closed terms therefore form a normal structure in which exactly the same sentences are true. Hence if a sentence φ is true in every normal model of Θ, then $\Theta \vdash_{\mathrm{NK}} \varphi$. The equality axioms say everything about identity that a derivation can use, so looking only at normal structures loses nothing.[12]

6.8. **Classical and Intuitionistic Logic.** We now have two logics for one language. The system NJ is the logic of Part 1, whose proofs are programs. The system NK is complete for the structures of this chapter. They differ by a single rule, and the difference decides what a proof contains.

The law of the excluded middle closes the gap, and that is its gain: entailment, which quantifies over every structure, becomes derivability, which asks for a finite certificate. With the law in hand an argument by cases on any claim is available, whether or not anyone can tell which case holds. Adding a rule can only let more through, and what it lets through costs NK the two properties of Fact 6.45.

Remark 6.69 (the drinker principle)*.* Take a signature with one one-place relation symbol Dr, think of the domain as the customers in a pub (there is at least one, since domains are not empty), and read $\mathsf{Dr}(x)$ as "x is drinking". The sentence

$$\exists x\,(\mathsf{Dr}(x) \to \forall y\,\mathsf{Dr}(y))$$

says that there is someone in the pub such that, if *they* are drinking, everyone is.[13]

[12] Open Logic Project (2026, §23.7) gives the details for any set of sentences that contains the equality axioms.

[13] The name is from Smullyan (1978), who called it the *drinking principle*. Smullyan, a logician and a prolific inventor of logic puzzles, also performed as a professional magician.

The principle sounds like a claim about a remarkable person, and so it is often suspected of being false. It is valid, by a two-line argument. Either everyone is drinking or someone is not. If everyone is, pick anyone: the consequent is true, so the implication is. If someone is not, pick that person: the antecedent is false, so the implication is true. There is no remarkable person, only a choice made differently in the two cases. The first case is where the non-empty domain is used; in an empty pub there would be nobody to pick, and the sentence would be false.

The system NJ does not derive this sentence, and this time the existence property is the reason. If it did, it would derive $\mathsf{Dr}(t) \to \forall y\, \mathsf{Dr}(y)$ for some term t. The signature has no constants and no function symbols, so t is a variable, say z. But $\mathsf{Dr}(z) \to \forall y\, \mathsf{Dr}(y)$ is false in a pub of two people under an assignment that sends z to the one who is drinking, so by soundness it has no derivation. Asked for *someone*, the rules of NJ insist on being told who. The argument above never says who: it says "this one if everyone is drinking, that one if not", without knowing which case holds.

In a pub of five that is no defect, since one can go and look. Replace "drinking" by a property of numbers that nobody knows how to check, and the pub by all the counting numbers, and the same argument proves that a certain number exists without indicating how to find it, even in principle. That is why some people refuse to accept it.

Example 6.70 (a derivation of the drinker principle in NK)**.** The drinker principle, which we write δ, is derivable in NK: $\vdash_{\mathrm{NK}} \exists x\, (\mathsf{Dr}(x) \to \forall y\, \mathsf{Dr}(y))$. The derivation splits on whether someone is not drinking. If someone is not, call them z; then z is the witness, because the antecedent fails for z. This is the derivation $\mathcal{F}_1$:

$$\dfrac{[\exists y\, \neg \mathsf{Dr}(y)]^1 \qquad \dfrac{\dfrac{\dfrac{[\neg \mathsf{Dr}(z)]^2 \qquad [\mathsf{Dr}(z)]^3}{\bot}\,{\to}\mathrm{E}}{\forall y\, \mathsf{Dr}(y)}\,{\bot}\mathrm{E}}{\mathsf{Dr}(z) \to \forall y\, \mathsf{Dr}(y)}\,{\to}\mathrm{I},3}{\delta}\,\exists \mathrm{I}}{\delta}\,\exists \mathrm{E},2$$

If nobody is not drinking, then everyone is, and anyone at all, say x, is the witness. This is the derivation $\mathcal{F}_2$, in which the $(\to\mathrm{I})$ step discharges nothing, since $\forall y\, \mathsf{Dr}(y)$ was reached without supposing $\mathsf{Dr}(x)$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[\neg\exists y\, \neg \mathsf{Dr}(y)]^1 \qquad \dfrac{[\neg \mathsf{Dr}(y)]^4}{\exists y\, \neg \mathsf{Dr}(y)}\,\exists \mathrm{I}}{\bot}\,{\to}\mathrm{E}}{\neg\neg \mathsf{Dr}(y)}\,{\to}\mathrm{I},4}{\mathsf{Dr}(y)}\,\mathrm{DN}}{\forall y\, \mathsf{Dr}(y)}\,\forall \mathrm{I}}{\mathsf{Dr}(x) \to \forall y\, \mathsf{Dr}(y)}\,{\to}\mathrm{I}}{\delta}\,\exists \mathrm{I}$$

One step of (∨E) on an instance of (LEM) joins the two cases:

$$\dfrac{\dfrac{}{\exists y\,\neg\mathsf{Dr}(y) \vee \neg\exists y\,\neg\mathsf{Dr}(y)}\ \text{LEM} \qquad \begin{array}{cc} [\exists y\,\neg\mathsf{Dr}(y)]^1 & [\neg\exists y\,\neg\mathsf{Dr}(y)]^1 \\ \mathcal{F}_1 & \mathcal{F}_2 \\ \delta & \delta \end{array}}{\delta}\ \vee\text{E}, 1$$

In $\mathcal{F}_1$ the witness is z, the person who is not drinking, and the (∃E) step is legal because z is free neither in δ, nor in $\exists y\,\neg\mathsf{Dr}(y)$, nor in any other open hypothesis above δ. In $\mathcal{F}_2$ the witness is x, anyone at all. The (∀I) step is legal because y is not free in the only open hypothesis, $\neg\exists y\,\neg\mathsf{Dr}(y)$. The (∃I) step uses a variable as the witness, and that is where the non-empty domain is paid for.

Which case supplies the conclusion depends on which side of the (LEM) is true, and the derivation does not say.

Proposition 6.71 (loss of the disjunction and existence properties)**.** *The system* NK *has neither the disjunction property nor the existence property.*

Proof.

- *Disjunction property.* The system NK derives $p \vee \neg p$ in one step, and it derives neither disjunct, since neither is valid and NK is sound.
- *Existence property.* Take a signature whose only symbol is the one-place relation symbol Dr, so that every term is a variable. The system NK derives $\exists x\,(\mathsf{Dr}(x) \to \forall y\,\mathsf{Dr}(y))$: split by (LEM) on $\exists y\,\neg\mathsf{Dr}(y)$. If someone is not drinking, that person is a witness, since the antecedent fails for them. If not, everyone is drinking, by (DN), and every variable is a witness (Example 6.70 gives the tree). Consider the structure with two elements, a and b, in which Dr holds of a only, and an assignment that sends every variable to a. For every term t the formula $\mathsf{Dr}(t)$ is true and $\forall y\,\mathsf{Dr}(y)$ is false, so $\mathsf{Dr}(t) \to \forall y\,\mathsf{Dr}(y)$ is false, and by soundness NK derives no such formula.

□

In NJ a proof of an existential sentence contains a witness, and the rules were designed so that it would: every derivation of $\exists x\,\varphi$ from no hypotheses can be straightened into one that ends by exhibiting a term t. A proof in NK can conclude $\exists x\,\varphi$ by showing that both answers to some question yield an x, without saying which answer is the true one. It certifies that something exists but no longer shows which thing.

Over a finite structure the price is nothing: whether φ or $\neg\varphi$ holds there is settled by running through the domain, so an argument by cases on φ can always be completed by looking. The price is felt only over domains that nobody can run through, such as the counting numbers. Whether it is worth paying depends on what makes a sentence true, a question that the mathematics of this chapter can sharpen but not settle.

The mathematicians who refuse the law, the *intuitionists*, take the loss of witnesses to be the point. The law was abstracted, they argue, from reasoning about finite collections, where going through the list settles every

question, and it was carried over unexamined to infinite collections, which nobody can go through.[14]

Goldbach's conjecture says that every even number greater than 2 is the sum of two primes.[15] Written with symbols for zero, successor, addition and multiplication, it is a sentence about the counting numbers, and by the definitions of this chapter it is true or false in the structure they form. To assert "Goldbach's conjecture or its negation" is to claim, the intuitionists say, that the question is settled, or at least can be, when nobody has any idea how to settle it. For them mathematics is a construction carried out in the mind, and "true" can only mean "constructed".[16]

Their logic gives the connectives meaning by *proof* conditions, what counts as a proof of a sentence, rather than by truth conditions.

- A proof of $\varphi \wedge \psi$ is a proof of each conjunct;
- a proof of $\varphi \to \psi$ is a method that turns every proof of φ into a proof of ψ;
- a proof of $\varphi \vee \psi$ is a proof of one disjunct *together with which one*;
- a proof of $\exists x\, \varphi$ is a particular x with a proof of φ for it.

On this reading the disjunction and existence properties of NJ are not accidents of the rules, waiting to be patched; they are what the connectives mean, and the rules of NJ write them down.[17]

Intuitionistic logic also has a semantics of its own: structures that grow in stages, which model a mathematician whose knowledge grows over time, and for which NJ is sound and complete.[18] Each logic therefore has a semantics that it fits exactly, and the question becomes which semantics describes what mathematicians are doing.

---

[14]Brouwer (1907) and Brouwer (1908), both in Dutch; the second is titled *The unreliability of the logical principles*. In the years that followed, Brouwer proved the theorems that made his name in topology, among them his fixed-point theorem, by exactly the kind of argument his philosophy rejected: the proof shows that a fixed point exists without showing how to find one.

[15]Goldbach proposed a form of it in a letter to Euler in 1742, and Euler restated it in this form. It has been checked by computer for every even number up to 4×10^{18}, which settles nothing.

[16]The dispute was bitter. In 1921 Hermann Weyl, Hilbert's former student, announced his conversion with the sentence "Brouwer, that is the revolution" (*und Brouwer – das ist die Revolution*) (Weyl, 1921); English translations of this paper, and of the other contributions to the debate, are collected in Mancosu (1998). Weyl later drifted back towards Hilbert. In 1928 Hilbert had Brouwer removed from the editorial board of the *Mathematische Annalen*, the leading journal of the day. Einstein, also on the board, refused to take sides and called the affair a *Frosch-Mäusekrieg*, a war of the frogs and the mice, after the ancient parody of Homer (van Dalen, 1990).

[17]The rules were first written down by Kolmogorov (1925), for a fragment, and by Heyting (1930), in full; the explanation of the connectives is named after Brouwer, Heyting and Kolmogorov. Kolmogorov's version (Kolmogoroff, 1932) reads formulas as problems: $\varphi \to \psi$ is the problem of solving ψ given a solution of φ. Gentzen's NJ of 1934 is this logic in the form used here.

[18]Kripke (1965). A structure with two stages, in which $=$ grows from one stage to the next, makes the sentence of Example 6.48 fail, and that is how one shows that NJ cannot derive it. Kripke had published a completeness theorem for modal logic, by similar means, at the age of eighteen (Kripke, 1959).

The two logics are equally safe from contradiction: every theorem of classical arithmetic can be translated into a theorem of intuitionistic arithmetic, with $\neg\neg$ inserted in the right places, so that if intuitionistic arithmetic is consistent, so is classical arithmetic.[19] What separates the logics is meaning: what a sentence asserts, and what counts as a proof of it.

The strongest modern case for the intuitionist answer is an argument about language rather than about mathematics.[20] The meaning of a sentence is fixed by how it is used, and a speaker's grasp of that meaning must show itself in what the speaker can do: above all, in recognising a proof of the sentence when one is presented. But some truth conditions could obtain without anyone being able to recognise that they do, and no speaker could display a grasp of those. For mathematics, then, meaning is given by proof conditions, and the right logic is intuitionistic. Realists reply that we understand more than we can verify, and the debate is alive; Button and Walsh (2018) is a guide to its current state. Mathematics itself has also moved on: a great deal of analysis can be developed constructively without the philosophy, and classical mathematics reads that work as mathematics that happens to supply witnesses.[21]

The proof conditions came back, half a century on, as the reason the proofs of NJ can be read as programs. A proof of a disjunction is a value with a tag on it, so a proof of the law of the excluded middle would be a program that, given any claim at all, returns the right tag, and there is no such program. For this reason the proof assistants used to verify software today, such as Lean and Rocq, are built on intuitionistic type theories, and they let a user add the law as an axiom or, in Lean, through the axiom of choice.[22] Figure 6.8 sets the two logics side by side.

The rest of this part of the course is classical, and that is a decision rather than a default. Its question is what holds in *every* structure, a question about truth, and the theorem that answers it, soundness and completeness together, is a theorem about NK. It fails for NJ, which is sound for the structures of this chapter but not complete for them. From here on we therefore work in a different logic from the one in which proofs are programs, and we have paid for completeness with the witnesses.

History. The logic of this chapter was built over half a century of argument about what mathematics may assume. Its language is, in modern dress, the one Gottlob Frege introduced in his *Begriffsschrift* of 1879.[23] It was the first formal language with quantifiers, and it came with the first system of proof precise enough that checking a proof needs no understanding of what

[19] Gödel (1933). Gentzen found the same translation independently, and withdrew his paper when Gödel's appeared.

[20] Dummett (1975); the book-length version is Dummett (2000).

[21] Bishop (1967).

[22] The separation is less clean than it looks. Griffin (1990) found that classical proofs are programs too, provided the programs may jump: the type of the control operator `call/cc` of Scheme is Peirce's law. What such a program returns as a witness may later be taken back.

[23] Frege (1879). Frege's notation was two-dimensional, a branching diagram on the page, and almost nobody adopted it. The notation that survived descends instead from Peirce's, by way of Schröder and Peano.

	NJ (intuitionistic)	NK (classical)
Rules	natural deduction as in Part 1	those rules plus the law of the excluded middle
Meaning of a connective	proof conditions	truth conditions
Semantics it fits exactly	structures that grow in stages	the structures of this chapter
Complete for the structures of this chapter	no	yes
Disjunction and existence properties	yes	no
Proofs as programs	yes	only with jumps

FIGURE 6.8. Two logics for one language.

it is about, only a comparison of shapes. Every proof checker since descends from it.

Frege built it for a purpose. The nineteenth century had rebuilt the calculus on the arithmetic of the real numbers, and the real numbers on sets of rational numbers. Frege set out to finish the job by showing that arithmetic itself is logic: that the truths about numbers follow from laws of thought alone, so that mathematics rests on the firmest ground there is. This programme is *logicism*, and Frege carried it out in the two volumes of his *Grundgesetze der Arithmetik* (Frege, 1893).

In June 1902, with the second volume at the printer, a letter arrived from Bertrand Russell. One of Frege's basic laws, Russell pointed out, allows a set of all sets that are not members of themselves; that set is a member of itself exactly when it is not, so the laws contradict one another.[24]

Russell's paradox was one of several, and together they raised the question of what makes mathematics safe. Russell's own answer was the theory of types, the ancestor of the type systems of programming languages. Two others bear on this chapter: David Hilbert's, which kept classical mathematics whole and set out to prove it safe, and L. E. J. Brouwer's, which rejected part of it, beginning with the law of the excluded middle. Section 6.8 describes the logic that Brouwer's answer produced. Hilbert's is told here.

Hilbert's answer was to keep classical mathematics and secure it. Its proofs, he proposed, should be treated as finite strings of symbols, and

[24]Frege added an afterword to the second volume, which begins: "Einem wissenschaftlichen Schriftsteller kann kaum etwas Unerwünschteres begegnen, als dass ihm nach Vollendung einer Arbeit eine der Grundlagen seines Baues erschüttert wird": hardly anything less welcome can happen to a scientific writer than to have one of the foundations of his building shaken after the work is finished. Sixty years later Russell wrote that he knew of nothing to compare with Frege's dedication to truth: with his life's work collapsing, Frege "responded with intellectual pleasure clearly submerging any feelings of personal disappointment". Both letters are in van Heijenoort (1967, pp. 124–128).

their consistency proved by reasoning about those strings with methods so elementary that even Brouwer could not object. This is *Hilbert's programme*, whose three questions are the rows of the scoreboard at the end of each chapter.

His ambition was large: "Aus dem Paradies, das Cantor uns geschaffen, soll uns niemand vertreiben können" (no one shall drive us out of the paradise that Cantor created for us).[25] A year later he put the same thought more bluntly: "Taking the principle of excluded middle from the mathematician would be the same, say, as proscribing the telescope to the astronomer or to the boxer the use of his fists."[26]

He meant that classical mathematics was too valuable to give up, and most mathematicians since have agreed with him in practice, whatever their view of the philosophy. Hilbert's reply is not an argument that Brouwer was wrong. It is an argument that his being right would cost a great deal.

Kurt Gödel answered one of Hilbert's questions in his doctoral thesis of 1929, written when he was twenty-three: the completeness theorem of this chapter, proved for a Hilbert-style system rather than for NK. Two years later he showed that the consistency proof Hilbert wanted cannot be given, and the last chapter of this part of the course explains why.

The definition of satisfaction fits Frege's view of truth. A sentence is about some things, and the things are there, determinate, whether or not anyone is looking. The sentence is true or false according to how they are, and it has its truth value before anyone finds out which value it is, and whether or not anyone ever does.[27]

Bivalence is no discovery about this picture. It was built into the definition of satisfaction when the clause for negation was written, and it says, in a single line, that every sentence has in every structure exactly one of two truth values.

Tarski gave this picture a precise form. His definition of truth, published in 1933, is the one of Section 6.3, and his definition of consequence, presented in 1935 at the Paris congress of the Unity of Science movement, is the one of Section 6.4. The second answered Carnap, who had tried in *Logische Syntax der Sprache* (Carnap, 1934) to define consequence by the shapes of sentences alone, without mentioning truth. Tarski used Gödel's incompleteness theorem (Chapter 8) to show that no definition by finitely many rules of inference can do the job, and objected that Carnap's own definition depends on special features of his language. Carnap, who had met Tarski in Vienna and Warsaw in 1930, became the most energetic convert to semantics and built his *Introduction to Semantics* (Carnap, 1942) on Tarski's work. Others in the Vienna Circle, Otto Neurath above all, suspected that talk of truth was metaphysics returning by the back door.

With these definitions Gödel's theorem can be read as showing that, on this picture, the classical system is exactly right. On this view the law of the excluded middle is not a rule that we chose but a law of truth, and

[25] Hilbert (1926).

[26] Hilbert (1928), quoted from the English translation in van Heijenoort (1967, p. 476).

[27] Frege put it most directly in "The Thought", Frege (1918), where a thought is true or false independently of whether anyone grasps it.

Who	When	What
Frege	1879	quantifiers; proofs checked by their shape
Frege; Russell	1893–1902	logicism; the paradox in its laws
Hilbert	1920s	a programme to prove mathematics safe
Gödel	1929	completeness
Gödel	1931	incompleteness
Tarski	1933; 1936	definitions of truth and of consequence
Gentzen	1934	NJ and NK
Henkin	1949	the proof of completeness used here

FIGURE 6.9. Classical logic and its semantics: who did what.

a proof system without it fails to see some truths. Gentzen's NK of 1934 gave the proofs the form used here, and the proof of completeness in this chapter is Henkin's of 1949. He found it by accident, as a graduate student at Princeton, while trying to prove something else, and he tells the story in Henkin (1996); it is now the proof in every textbook. Figure 6.9 collects the dates.

Summary. A system is modelled as a *structure*: a non-empty set of things, with a meaning for each symbol of a signature. A sentence is *true* or false in a structure by a definition with one clause per symbol. Its two quantifier clauses loop over the domain, so on a finite structure the definition is an algorithm.

A theory *entails* a sentence when the sentence is true in every model of the theory; for a theory that describes a system, this means "true of every system that meets it". The derivations of NJ are finite certificates, and they are *sound*: they never lead from truths to a falsehood. They do not reach every sentence that is entailed, and the whole of the gap is one law, excluded middle.

Adding excluded middle gives NK, which is sound and *complete*, so a theory entails a sentence exactly when the sentence has a derivation from it. The price is that a proof of an existential sentence no longer says what exists.

HILBERT'S PROGRAMME — STATUS.

Item	Status after this chapter
1. Consistency	• **holds**: NK derives no $\bot$, because $\bot$ is true in no structure and NK is sound
2. Completeness	• **holds**: NK derives every sentence true in every model of its hypotheses; NJ does not
3. Decision procedure	○ **open**: not yet asked. Truth in one finite structure can be computed; truth in *every* structure is another matter

• holds ○ open, or settled in part × fails

The three questions are Hilbert's, asked of each logic in turn. For the logic of this chapter the first two are answered, and the third has not yet been posed.

What remains. The system NK derives everything a theory entails, but finding an NK derivation of a hard fact is a separate problem. At every step there is an instance of excluded middle for every formula one could imagine and an unbounded choice of terms for each quantifier, with no hint of which choice leads anywhere. We have a proof system that reaches every truth and no method for searching it.

But the completeness proof contains an idea we have not used. It built a counter-model out of a failure to derive: if there is no derivation, there is a structure in which the claim fails. Instead of hunting for the proof, then, could we hunt for the world?

Exercises. The exercises run from practice to proof.

- Exercises 6.1 to 6.6 use the definitions on particular sentences and structures;
- Exercises 6.7 to 6.12 ask for derivations and for the arguments behind the claims about NJ and NK;
- Exercises 6.13 to 6.17 fill in the proofs that the text left to the reader.

Exercise 6.1. Let Σ have a constant symbol c, a one-place function symbol f, one-place relation symbols P and Q, and two-place relation symbols R and $=$. Let $\mathfrak{A}$ be the Σ-structure with domain $\{0, 1, 2, 3, 4\}$ in which

- c names 0;
- f sends 0 to 0 and every other n to $n - 1$;
- P is $\{0, 2, 4\}$ and Q is $\{1, 3\}$;
- R is *greater than*, and $=$ is identity.

Read $P(x)$ as "x is even", $Q(x)$ as "x is odd", $R(x, y)$ as "x is greater than y" and $f(x)$ as "the predecessor of x". Translate the following into Σ, and say whether each is true in $\mathfrak{A}$:

(a) every even number is greater than every odd number;

(b) only 0 is the predecessor of its own predecessor.

Translate the following into English, and say whether each is true in $\mathfrak{A}$:

(c) $\forall x\,(P(x) \to \exists y\,(Q(y) \wedge f(y) = x))$;

(d) $\exists x\,\forall y\,(x \neq y \to R(x, y))$.

Exercise 6.2. Let Σ have a constant symbol c, a one-place function symbol f, one-place relation symbols P and Q, and a two-place relation symbol R. Let $\mathfrak{B}$ be the Σ-structure with domain $\{1, 2, 3, 4, 5\}$ in which c names 1, R is *less than*, and f, P and Q are given by the table

n	1	2	3	4	5
$f(n)$	1	1	1	2	2
in P?	no	no	no	yes	yes
in Q?	yes	yes	no	no	no

Evaluate each of the following in $\mathfrak{B}$ by a table with one row for each value of the outermost variable:

(a) $\forall x\,(P(x) \to Q(f(x)))$;

(b) $\forall x\,(P(x) \vee Q(x))$;

(c) $\exists x\,R(x, f(x))$.

Exercise 6.3. Let Σ have a constant symbol c, a one-place function symbol f and the two-place relation symbol $=$. Let $\mathfrak{A}$ and $\mathfrak{B}$ be the Σ-structures in which $=$ is identity, c names 1, and

- $\mathfrak{A}$ has domain $\{1, 2, 3, 4, 5\}$, and f sends 1, 2 and 3 to 1 and sends 4 and 5 to 2;
- $\mathfrak{B}$ has domain $\{1, 2, 3, 4\}$, and f sends 1 and 2 to 1 and sends 3 and 4 to 2.

Find a Σ-sentence that is true in $\mathfrak{A}$ and false in $\mathfrak{B}$, and check both verdicts against the definition of satisfaction.

Exercise 6.4. Let Σ have a constant symbol c, a one-place function symbol f, a one-place relation symbol P and a two-place relation symbol R, and let φ be the formula

$$\forall y\,(R(y,x) \to P(y)).$$

Apply the recursion in the definition of substitution to put c, $f(z)$ and $f(y)$ in turn for x in φ, and say which of the three terms are substitutable for x in φ.

Exercise 6.5. Let Σ have a constant symbol c, a one-place function symbol f, one-place relation symbols P and Q, and the two-place relation symbol $=$. For each of the following, find a counter-model as small as you can make it, check it against the definition of satisfaction, and say why no smaller structure will do.

(a) $\exists x\,P(x) \models P(c)$;
(b) $\forall x\,(P(x) \vee Q(x)) \models \forall x\,P(x) \vee \forall x\,Q(x)$;
(c) $Q(c),\ \forall x\,(Q(x) \to \exists y\,(f(y) = x \wedge y \neq x)) \models \exists x\,(Q(x) \wedge x \neq c)$, in structures that interpret $=$ as identity.

Exercise 6.6. Let Σ have a constant symbol c, a one-place function symbol f, and two-place relation symbols R and $=$, and let Θ be the theory consisting of the two sentences

$$\mathrm{T}_1:\ \forall x\,\neg R(x,x),$$
$$\mathrm{T}_2:\ \forall x\,(x \neq c \to R(f(x),x)).$$

For each sentence, decide whether Θ entails it. If it does, give a derivation in NK; if not, give a counter-model in which $=$ is identity.

(a) $\neg R(c,c)$;
(b) $\forall x\,(R(f(x),x) \vee x = c)$;
(c) $\forall x\,\neg R(x,f(x))$;
(d) $\exists x\,R(f(x),x)$.

Exercise 6.7. Give a derivation of Peirce's law, $((\varphi \to \psi) \to \varphi) \to \varphi$, in NK, for all formulas φ and ψ.

Exercise 6.8. Let Σ be a signature containing the two-place relation symbol $=$, and Eq_Σ its equality axioms (Definition 6.36). Show that

$$\mathrm{Eq}_\Sigma \vdash_{\mathrm{NJ}} \forall x\,\forall y\,(x = y \to y = x)$$

and

$$\mathrm{Eq}_\Sigma \vdash_{\mathrm{NJ}} \forall x\,\forall y\,\forall z\,(x = y \to (y = z \to x = z)).$$

Exercise 6.9. Let P be a one-place relation symbol. Someone offers a derivation witnessing $\exists x\,P(x) \vdash_{\mathrm{NJ}} \forall y\,P(y)$, as follows: from $\exists x\,P(x)$, and the derivation of $P(y)$ from the supposition $[P(y)]$ (a single leaf), conclude $P(y)$ by ($\exists$E); then conclude $\forall y\,P(y)$ by ($\forall$I). Take the structure with domain $\{1,2\}$ in which P is $\{1\}$, and an assignment sending y to 2. List the formulas of the derivation with their open hypotheses and truth values, as in Remark 6.43, find the first line that is false while all its open hypotheses are true, and say which side condition it breaks.

Exercise 6.10. (a) Show that if NJ derived every instance of Peirce's law, $((\varphi \to \psi) \to \varphi) \to \varphi$, it would derive $p \vee \neg p$.

(b) Show the same for every instance of $\neg\forall x\, \neg\psi \to \exists x\, \psi$.

(c) Show that $\neg\forall x\, \psi \vdash_{\mathrm{NK}} \exists x\, \neg\psi$ for every formula ψ.

Exercise 6.11. In a signature with two constants a and b and a one-place relation symbol Dr, show that NK derives $\exists x\, (\mathsf{Dr}(x) \to \forall y\, \mathsf{Dr}(y))$ but derives $\mathsf{Dr}(t) \to \forall y\, \mathsf{Dr}(y)$ for no term t. Why does it not help that there are now closed terms to choose from?

Exercise 6.12. Show that $\neg(\varphi \to \psi) \vdash_{\mathrm{NJ}} \neg\psi$ and $\neg(\varphi \to \psi) \vdash_{\mathrm{NJ}} \neg\neg\varphi$, and that $\neg(\varphi \to \psi) \vdash_{\mathrm{NK}} \varphi$. Does NJ derive the last?

Exercise 6.13. Write out the cases ($\wedge$E) and ($\exists$I) of the soundness proof for NJ. Which lemma does the second case need?

Exercise 6.14. Prove the relevance lemma: first for terms (if s and s' agree on $\mathrm{var}(t)$, then $\mathrm{Val}_s(t) = \mathrm{Val}_{s'}(t)$), and then for formulas, by induction.

Exercise 6.15. (a) Prove, by induction on the term u, that if t is a term then

$$\mathrm{Val}_s(u[t/x]) = \mathrm{Val}_{s[\mathrm{Val}_s(t)/x]}(u).$$

(b) Deduce from (a) the atomic case of the substitution lemma.

(c) Where would the induction for formulas use the hypothesis that t is substitutable for x?

Exercise 6.16. (a) Prove part (ii) of the lemma on maximal consistent sets.

(b) Prove the existential case of the truth lemma, and say where the witness property is used.

Exercise 6.17. Let P and Q be one-place relation symbols and c_0 a constant symbol, and let Γ be $\{\exists x\, P(x),\ \exists x\, Q(x),\ \neg\exists x\, (P(x) \wedge Q(x))\}$.

(a) Show that Γ is consistent.

(b) Show that Γ, together with $\exists x\, P(x) \to P(c_0)$ and $\exists x\, Q(x) \to Q(c_0)$, derives $\bot$ in NJ.

(c) Which condition of the lemma on Henkin witnesses does the second added sentence break?

References.

Errett Bishop. *Foundations of Constructive Analysis*. McGraw-Hill, New York, 1967.

L. E. J. Brouwer. *Over de Grondslagen der Wiskunde*. PhD thesis, Universiteit van Amsterdam, 1907.

L. E. J. Brouwer. De onbetrouwbaarheid der logische principes. *Tijdschrift voor Wijsbegeerte*, 2:152–158, 1908. English translation, "The unreliability of the logical principles", in Brouwer, *Collected Works*, vol. 1, North-Holland, 1975.

Tim Button and Sean Walsh. *Philosophy and Model Theory*. Oxford University Press, Oxford, 2018.

Rudolf Carnap. *Logische Syntax der Sprache*. Springer, Wien, 1934.

Rudolf Carnap. *Introduction to Semantics*. Harvard University Press, Cambridge, MA, 1942.

E. F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.

Michael Dummett. The philosophical basis of intuitionistic logic. In H. E. Rose and J. C. Shepherdson, editors, *Logic Colloquium '73*, pages 5–40. North-Holland, Amsterdam, 1975.

Michael Dummett. *Elements of Intuitionism*. Oxford University Press, Oxford, second edition, 2000.

Anita Burdman Feferman and Solomon Feferman. *Alfred Tarski: Life and Logic*. Cambridge University Press, Cambridge, 2004.

Gottlob Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Louis Nebert, Halle, 1879. English translation by S. Bauer-Mengelberg in J. van Heijenoort (ed.), *From Frege to Gödel*, Harvard University Press, 1967, pp. 1–82.

Gottlob Frege. *Grundgesetze der Arithmetik*. Hermann Pohle, Jena, 1893. Volume I 1893, volume II 1903.

Gottlob Frege. Der Gedanke. Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus*, 1:58–77, 1918. English translation, "The thought: a logical inquiry", *Mind* 65 (1956), 289–311.

Gerhard Gentzen. Untersuchungen über das logische Schließen I, II. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934. Appeared 1934–35; English translation in *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo, North-Holland, 1969.

Kurt Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, 37:349–360, 1930. Doctoral dissertation, Vienna, 1929.

Kurt Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums*, 4:34–38, 1933.

Timothy G. Griffin. A formulae-as-types notion of control. In *Proceedings of the 17th ACM Symposium on Principles of Programming Languages (POPL)*, pages 47–58, 1990.

Leon Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, 14(3):159–166, 1949.

Leon Henkin. The discovery of my completeness proofs. *Bulletin of Symbolic Logic*, 2(2):127–158, 1996.

Arend Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische*

Klasse, pages 42–56, 1930.

David Hilbert. Über das Unendliche. *Mathematische Annalen*, 95:161–190, 1926. English translation in van Heijenoort (1967), pp. 367–392.

David Hilbert. Die Grundlagen der Mathematik. *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, 6:65–85, 1928. Lecture of 1927. English translation, "The foundations of mathematics", in van Heijenoort (ed.), *From Frege to Gödel*, 1967, 464–479.

A. Kolmogoroff. Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift*, 35:58–65, 1932.

A. N. Kolmogorov. O principe tertium non datur. *Matematičeskij Sbornik*, 32: 646–667, 1925. English translation in van Heijenoort (1967), pp. 414–437.

Saul A. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24(1):1–14, 1959.

Saul A. Kripke. Semantical analysis of intuitionistic logic I. In J. N. Crossley and M. A. E. Dummett, editors, *Formal Systems and Recursive Functions*, pages 92–130. North-Holland, Amsterdam, 1965.

Paolo Mancosu, editor. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*. Oxford University Press, New York, 1998.

Open Logic Project. The Open Logic Text. `https://openlogicproject.org`, 2026. Release 9620cc7, 12 July 2026. Licensed CC BY 4.0.

Charles S. Peirce. On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics*, 7(2):180–202, 1885.

Dag Prawitz. *Natural Deduction: A Proof-Theoretical Study*. Almqvist & Wiksell, Stockholm, 1965.

Raymond M. Smullyan. *What Is the Name of This Book?* Prentice-Hall, Englewood Cliffs, NJ, 1978.

Alfred Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1:261–405, 1935. First published in Polish, 1933. English translation in Tarski, *Logic, Semantics, Metamathematics*, Oxford, 1956.

Alfred Tarski. Über den Begriff der logischen Folgerung. In *Actes du Congrès International de Philosophie Scientifique, Paris 1935*, volume VII, pages 1–11. Hermann, Paris, 1936.

Dirk van Dalen. The war of the frogs and the mice, or the crisis of the Mathematische Annalen. *The Mathematical Intelligencer*, 12(4):17–31, 1990.

Jean van Heijenoort, editor. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, Cambridge, MA, 1967.

Hermann Weyl. Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift*, 10:39–79, 1921.