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. Syntax 2
6.2. Structures 6
6.3. Satisfaction 9
6.4. Entailment 14
6.5. Soundness of NJ 19
6.6. The Law of the Excluded Middle 25
6.7. Completeness of NK 28
6.8. Classical and Intuitionistic Logic 38
History 42
Exercises 47

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?

Definitions answer the first two questions. A theorem answers the third, though the first rules of proof we try turn out to fall short. Section 6.1 fixes the language in which sentences about a system are written. Section 6.2 makes a system into a mathematical object, a *structure*, and Section 6.3 defines when a sentence is true in one. 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, those of Gentzen's natural deduction system NJ, and shows that they never lead from true premisses to a false conclusion. Section 6.6 finds that they miss some sentences that follow, all for want of a single law, and adds it: the law of the excluded middle, which gives the system NK. Section 6.7 proves NK 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.

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.

6.1. **Syntax.** The finance director's claim is about one company, but the rules of the personnel system are not: they are meant to hold of every table the system can produce, of five people or five hundred, and each such table must be able to read them. So the claim cannot be written in terms of Thorne and Finch. It needs a vocabulary that any table of the right shape can interpret in its own way: a name for the chief executive, whoever she is, a symbol for *the boss of*, symbols for *is an engineer* and *earns more than*. Such a vocabulary is a *signature*, and it is the one new thing in this section. Given a signature, we name things (*terms*), we say things about the things named (*atomic formulas*), and we combine what we have said (*formulas*). That is the whole language, and its definitions follow one straight after another.

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.

To speak of a thing without naming it (*anyone*, *something*), the language has, besides the symbols of the signature, a countably infinite set Var of *variables*, x, y, z, x_1, x_2 and so on, disjoint from the symbols of every signature.

Definition 6.2 (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.

A term names something. A relation symbol applied to terms says something about the things named. That is the smallest statement the language

can make, an *atomic* formula; $\bot$ 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 6.3 (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 $2 + 3 \times 4$.

That is the language. The next two examples give the two signatures that the rest of this chapter uses most.

Example 6.4 (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.

The terms ceo, $\mathsf{boss}(\mathsf{ceo})$ and $\mathsf{boss}(\mathsf{boss}(\mathsf{ceo}))$ name the chief executive, her boss, and her boss's boss, whoever those turn out to be; $\mathsf{boss}(x)$ names the boss of whoever 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: it is an atomic formula, "the chief executive is an engineer", and so is $\mathsf{Richer}(\mathsf{boss}(x), x)$, "the boss of x earns more than x". Figure 6.2 gives four sentences in the company signature, each with its English. The first is the finance director's claim and the second is the claim with Thorne's row allowed for. Nothing in them mentions Thorne: they are written in the signature, and any company that can read the signature can read them.

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 6.2. Four sentences and their English.

Example 6.5 (the counting numbers)**.** The counting numbers $0, 1, 2, \ldots$ are spoken of in 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 symbol s is read "the successor of", the number after. The terms 0, $s(0)$, $s(s(0))$, ... are the *numerals*, one name for each number, and $x + s(0)$ names the number after whatever x is. The formula $x < s(x)$ says that x is less than its successor, the sentence $\forall x\, \exists y\, (x < y)$ that every number has a bigger one, and $\neg\exists x\, (x < 0)$ that nothing is less than zero.

Remark 6.6 (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".

Remark 6.7 ("every" goes with $\to$, "some" with $\wedge$)*.* 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 third and fourth sentences of Figure 6.2 follow the rule. 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. The rule of thumb is reliable but not infallible.

Two more notions concern the variables. A formula with variables does not yet say anything definite. The formula $\mathsf{Richer}(\mathsf{boss}(x), x)$ is true or false according to who x is, whereas the sentence $\forall x\, \mathsf{Richer}(\mathsf{boss}(x), x)$ makes one claim about the whole domain: everyone earns less than their boss. The first

notion separates the variables whose values must be supplied from those that a quantifier has already taken charge of.

The second notion turns a claim about anybody into a claim about somebody in particular. From "everyone earns less than their boss" anyone concludes that the chief executive earns less than hers, and the rule (∀E) of NJ makes the same move on the page: it puts a term in place of the variable, so that $\mathsf{Richer}(\mathsf{boss}(x), x)$ becomes $\mathsf{Richer}(\mathsf{boss}(\mathsf{ceo}), \mathsf{ceo})$. This is *substitution*, an operation on strings, and it has 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. A last definition says when that cannot happen.

Definition 6.8 (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:

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

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:

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

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

Definition 6.10 (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:

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

Definition 6.11 (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 $y \in \mathrm{var}(t)$.

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

Example 6.12 (the company, cont'd: variables)**.** 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 x is free in it, 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.

If φ is $x \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(x), x)$, then

$$\varphi[\mathsf{boss}(\mathsf{ceo})/x] \;\text{ is }\; \mathsf{boss}(\mathsf{ceo}) \neq \mathsf{ceo} \to \mathsf{Richer}(\mathsf{boss}(\mathsf{boss}(\mathsf{ceo})), \mathsf{boss}(\mathsf{ceo})) :$$

the claim made of x is now made of the chief executive's boss. In $\mathsf{Eng}(x) \to \forall x\,\mathsf{Mgr}(x)$, putting ceo in for x gives $\mathsf{Eng}(\mathsf{ceo}) \to \forall x\,\mathsf{Mgr}(x)$, by the fourth line of the definition: only the free occurrence is replaced.

Capture is the one way in which substitution can go wrong. Let ψ 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 term is not substitutable, since the free occurrence of x lies inside $\exists y\,\mathsf{Richer}(y, x)$ and $y \in \mathrm{var}(\mathsf{boss}(y))$.

The sentences of Figure 6.2 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 company: it is written in a signature, and every company that can read the signature reads it in its own way. To say when a sentence is true, we must first say what it is true *of*.

6.2. **Structures.** What a sentence is true *of* has two parts: what the symbols of the signature are taken to mean, and among what things. Engineers use the word *system* for almost anything with parts, and we begin from what the word covers rather than from a definition. Figure 6.3 lists five systems.

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 6.3. 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. The relations, the functions and the named things are exactly what a signature has symbols for, and a structure is a system seen through a signature: it says which things there are and what each symbol means among them.

Definition 6.13 (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]

[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.

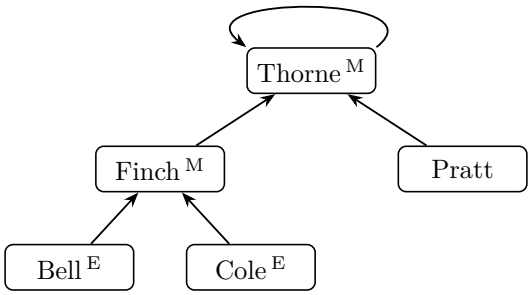

FIGURE 6.4. The company $\mathfrak{C}$.

Example 6.14 (the company, cont'd: Tuesday's table as a structure)**.** Tuesday's table is the Σ_{C}-structure $\mathfrak{C} = (\mathbb{D}_{\mathrm{C}}, I_{\mathrm{C}})$ given by

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

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

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 6.4 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 6.3 have an infinite domain.

[2]See Chapter 8.

Example 6.15 (the counting numbers, cont'd: the structure)**.** The counting numbers are the Σ_{A}-structure $\mathfrak{N} = (\mathbb{N}, I_{\mathrm{N}})$, 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 6.16 (the other three systems)**.** The other three systems of Figure 6.3 are structures in the same way.

- The server network has a domain of four machines and one two-place relation.
- 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 6.17 (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 the empty structure out.

Remark 6.18 (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 6.19 (the company, cont'd: a near miss)**.** Suppose the table left Thorne's boss blank, since the chief executive has none. Leaving it blank is sensible, 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 they chose Thorne herself, which is why Harriet's first query returned a row.

6.3. **Satisfaction.** A structure says what the symbols of a formula mean, and we already know what the logical words mean in English: 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 definition of truth writes that meaning down, one kind of formula at a time. 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.

One thing is still missing. A structure does not say what the variables stand for, and a formula with free variables, such as $\mathsf{Richer}(\mathsf{boss}(x), x)$, is true or false only once they have been given values. So three definitions

are needed, one after another: what the variables stand for, an *assignment*; what a term names, its *value*; and when a formula is true, *satisfaction*.

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

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

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.

Definition 6.22 (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* 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 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 formula ψ itself is left as it is: the clause changes the assignment, and puts no term into ψ. The structure could not do this job: it 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.

[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.

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

FIGURE 6.5. Satisfaction.

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.

The same clauses cover the simplest formulas of all, those built from propositional letters with $\bot$ and the connectives, and there they reduce to something familiar. A propositional 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 $[\![p]\!]^{\mathfrak{M}}$ is either empty or all of $\mathbb{D}^0$, and the atomic clause makes p false or true accordingly. The assignment plays no part, since p has no terms. The clauses for $\bot$ and the connectives then compute the truth of a formula from the truth of its parts.

So for a formula built from letters, whether $\mathfrak{M}, s \models \varphi$ depends only on which letters $\mathfrak{M}$ makes true. For such formulas a structure amounts to a choice of true or false for each letter, a *valuation*. Propositional logic is the part of first-order logic in which every relation symbol has no places, and its structures are, in effect, the rows of a truth table.

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 the variables 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 6.23 (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 is how a proof passes from a claim about anybody to the same claim about somebody in particular. Satisfaction makes the same passage in its own way: the clause for $\forall$ lets x stand for each element in turn, by changing the assignment. Proofs are to be judged by truth, so the two passages had better agree: putting the term t in for x should come to the same as letting x stand for whatever t names. The next lemma says that they agree, provided no quantifier of φ captures a variable of t.

Lemma 6.24 (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 6.12 shows a case).

The proof, by induction on φ, is long and unilluminating, though not difficult, and we omit it, as most textbooks do. 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*.

Remark 6.25 (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, 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.

Remark 6.26 (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".

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.

- 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.

Example 6.27 (the company, cont'd: truth in Tuesday's table)**.** Work in $\mathfrak{C}$, under an assignment s. If $s(x) =$ Cole, the term $\mathsf{boss}(\mathsf{boss}(x))$ has the value

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

The closed term $\mathsf{boss}(\mathsf{boss}(\mathsf{ceo}))$ also has the value Thorne, whatever s is, since the first line of the definition of value is never reached.

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, the sentence is true in $\mathfrak{C}$ under s exactly when, 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, and no row consulted s: each replaced whatever s gave x by d. 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 that puzzles readers: 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.

A sentence with two quantifiers needs a witness inside each row. 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 e whose boss is Thorne and who is not Thorne; Finch will do.
- For Finch, Bell will do.

So this sentence is true in $\mathfrak{C}$ too. The evidence has a characteristic shape: for the universal quantifier we examined every row, and for each existential quantifier we produced one witness.

Example 6.28 (propositional logic)**.** Take two letters, p and q. There are four valuations, one for each row below, and the clauses of Figure 6.5 compute the column of each connective from the columns of its parts. Writing T and F for true and false:

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \to q$
T	T	F	T	T	T
T	F	F	F	T	F
F	T	T	F	T	T
F	F	T	F	F	T

The column for $\neg p$ comes from the clause for $\to$ with $\bot$, which is false in every row, as consequent. The column for $p \to q$ is false in one row only, the row in which p is true and q false.

A longer formula is evaluated the same way, part by part. For "if p implies q, and q is false, then p is false":

p	q	$p \to q$	$\neg q$	$(p \to q) \wedge \neg q$	$\neg p$	$((p \to q) \wedge \neg q) \to \neg p$
T	T	T	F	F	F	T
T	F	F	T	F	F	T
F	T	T	F	F	T	T
F	F	T	T	T	T	T

The last column is true in every row. Since every structure, under every assignment, gives p and q the truth values of one of the four rows, the formula is true in every structure under every assignment, whatever else its signature contains.

The formula $(p \to q) \to (q \to p)$, by contrast, is false in the third row: $p \to q$ is true there and $q \to p$ false. A structure that makes p false and q true falsifies it, and four rows were enough to find one. With n letters there are 2^n rows to check.

6.4. **Entailment.** Satisfaction answers a question about one structure and one assignment: is this formula true here? Often, though, 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. So two definitions are needed: one for the

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.

requirements and the systems that meet them, and one for what follows from the requirements.

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 6.29 (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.

Definition 6.30 (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.

The two possible answers to "does Γ entail φ?" call for different kinds of evidence, and the two kinds are not on an equal footing. To show that an entailment *fails*, one structure suffices: a structure and an assignment that satisfy Γ and make φ false. That is a certificate, which anyone can check by evaluating formulas in it. To show that an entailment *holds*, one needs an argument, "let $\mathfrak{M}$ be a model of the theory and d an element", which can be convincing but cannot be checked by running anything: it speaks of every

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

model of the theory, of every size, including the infinite ones, and no loop runs over all of those. The last two examples of the section each show an entailment and a non-entailment.

Remark 6.31 (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.

Remark 6.32 (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.
- 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.

Remark 6.33 (the conditional of science). The word "implies" is ambiguous in English, and English forms conditionals of several kinds. The clause for $\to$ fixes one of them: the *law-like* indicative conditional, the "if" with which mathematics and science state their laws. Newton's first law is one: if a body is in motion, then it stays in motion unless an external force acts on it. Does the law hold of a body at rest? Of course it does. A law is a claim about everything, and a body at rest is a case it does not apply to; the law does not become false because it does not apply. It is broken only by a case it does apply to, a body in motion that stops with no force acting on it.

Read this way, a conditional is false only when its antecedent is true and its consequent false, which is the clause for $\to$. Followed through, the reading arrives at two validities that look odd at first sight:

$$\neg\varphi \to (\varphi \to \psi) \qquad \text{and} \qquad \psi \to (\varphi \to \psi).$$

A case in which the antecedent fails does not break a law, whatever the consequent says, and neither does a case in which the consequent holds, whatever the antecedent says. So "if $2 + 2 = 5$, then $1 + 1 = 3$" is true, and "if n is even, then n^2 is even" holds of every n, odd ones included.

This is not the conditional of conversation, which often promises a connection between its two halves, and still less the subjunctive or counterfactual, "if I had left earlier, I would have caught the bus", whose antecedent is known to be false and which is nevertheless a substantial claim, about what would have followed. No clause that decides $\varphi \to \psi$ from the truth values of φ and ψ alone could capture those, and the definition of satisfaction needs

exactly that: it decides an implication by deciding its two halves.[6] The History tells how the clause came to be chosen, and why the two validities above have been called the paradoxes of material implication.

Remark 6.34 (the equality axioms). The symbol $=$ is a two-place relation symbol like any other, and a structure may interpret it as any set of pairs. A system that compares names relies on two facts about identity, though: everything is equal to itself, and equals may be put for equals. Written as sentences of a signature Σ that contains $=$, these facts are the *equality axioms* of Σ:

(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*, and every normal structure is a model of 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.

Remark 6.35 (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 6.36 (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.

[6] That the indicative conditional of English is, after all, the material one is defended by Williamson (2020): we assess a conditional by supposing its antecedent and judging its consequent, a reliable but fallible habit, and the habit, not the truth condition, is what makes "if $2 + 2 = 5$, then $1 + 1 = 3$" sound wrong.

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), from the formula $\mathsf{Richer}(u, w)$ with u and v for x and y, says that whoever equals someone who earns more than w also earns more than w:

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

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 is a model of 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.

An entailment. Sharpened a little, the finance director's question is whether the theory guarantees that nobody but the chief executive is her own boss:

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

It does. Let $\mathfrak{M}$ be a model of Θ, and let d be an element of its domain with $(d, [\![\mathsf{ceo}]\!]^{\mathfrak{M}}) \notin [\![=]\!]^{\mathfrak{M}}$, so that $x \neq \mathsf{ceo}$ is true under $s[d/x]$. Suppose that $([\![\mathsf{boss}]\!]^{\mathfrak{M}}(d), d) \in [\![=]\!]^{\mathfrak{M}}$. Then

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

which contradicts R_1.

So $([\![\mathsf{boss}]\!]^{\mathfrak{M}}(d), d) \notin [\![=]\!]^{\mathfrak{M}}$: the formula $\mathsf{boss}(x) \neq x$ is true under $s[d/x]$, and so is the implication. For an element d with $(d, [\![\mathsf{ceo}]\!]^{\mathfrak{M}}) \in [\![=]\!]^{\mathfrak{M}}$ the implication is true by its clause. Since $\mathfrak{M}$ and d were arbitrary, the entailment holds.

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? It does not. 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 one-person company is a model of Θ in which the sentence is false, and so the entailment fails.

The two answers have the two shapes the section described: the entailment needed an argument about an arbitrary model, and the non-entailment needed only one small structure, which anyone can check.

Example 6.38 (the counting numbers, cont'd: a theory of order)**.** In the signature of arithmetic, let $\Theta_<$ be the theory with the two sentences

$$\mathrm{O}_1 : \ \forall x\, \neg(x < x),$$
$$\mathrm{O}_2 : \ \forall x\, \forall y\, \forall z\, (x < y \wedge y < z \to x < z).$$

Nothing is less than itself, and "less than" passes along chains. The counting numbers are a model of $\Theta_<$: no number is less than itself, and if $l < m$ and $m < n$ then $l < n$.

An entailment. The theory entails that "less than" never runs both ways:

$$\Theta_< \ \models \ \forall x\, \forall y\, (x < y \to \neg(y < x)).$$

Let $\mathfrak{M}$ be a model of $\Theta_<$, and d and e elements with $(d, e) \in [\![<]\!]^{\mathfrak{M}}$. Suppose that $(e, d) \in [\![<]\!]^{\mathfrak{M}}$ as well. Then $(d, d) \in [\![<]\!]^{\mathfrak{M}}$, by O_2 with d, e, d for x, y, z, and that contradicts O_1. So $(e, d) \notin [\![<]\!]^{\mathfrak{M}}$, the implication is true for d and e, and since $\mathfrak{M}$, d and e were arbitrary, the entailment holds.

A non-entailment. The sentence $\forall x\, \exists y\, (x < y)$, "every number has a bigger one", is true in the counting numbers, but $\Theta_<$ does not entail it. Take the structure with one element a, in which 0 names a, the function symbols s, $+$ and $\times$ send everything to a, and $<$ is empty. Then O_1 holds, since nothing is less than anything; O_2 holds, since its antecedent is never true; and $\forall x\, \exists y\, (x < y)$ is false at a.

So being true in the counting numbers is one thing, and following from a theory that the counting numbers satisfy is another: $\mathfrak{N}$ satisfies many sentences that $\Theta_<$ leaves open. Whether some theory settles every truth about $\mathfrak{N}$ is the question of Chapter 8.

6.5. **Soundness 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, each step of which a program can check, showing that the commitments made by Γ yield φ. Part 1 supplied a system of such proofs, 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, and $\mathrm{OH}(\mathcal{D})$ for the set of open hypotheses of a derivation $\mathcal{D}$. A derivation is the certificate we asked for: a finite object, each step of which a program can check against the rules.

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. 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 a derivation. This section proves the first.

Induction on derivations. Soundness is a claim about every derivation, and there are infinitely many derivations, of every size. A claim about every counting number is proved by induction: show it for 0, show that it passes from each n to $n + 1$, and it holds for every number, since every number is reached from 0 by adding 1 again and again. Derivations are built in the same way, from leaves by rules, and the same method applies to them.

A derivation of NJ is either a single leaf, which is a hypothesis, or a rule applied to one, two or three smaller derivations, its *immediate subderivations*. Drawn as in Part 1, each smaller derivation is a triangle, with its name inside and its conclusion at the point, and the rule is a bar beneath them. The derivation that ($\to$E) builds from $\mathcal{D}_1$ and $\mathcal{D}_2$, for instance, is

$$\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2 \\ \varphi \to \psi & \varphi\end{array}}{\psi}\ {\to}\mathrm{E}$$

Every derivation is reached by starting from leaves and adding one rule at a time. So to show that some property holds of *every* derivation, it is enough to show two things:

- *the base:* the property holds of every single leaf;
- *the steps:* for each rule, if the property holds of the immediate subderivations, then it holds of the derivation that the rule builds from them.

The supposition in a step, that the property holds of the smaller derivations, is the *induction hypothesis*. The base gives the property for the leaves; the steps then give it for every derivation with one rule, then for every derivation with two, and so on, so it holds for all of them. The relevance lemma was proved in the same way, by induction on formulas, with the atomic formulas as the base and one step for each connective and quantifier.

Which property? It cannot be "the conclusion is true": a derivation reasons from suppositions, and a supposition may be false, and so may everything drawn from it. It cannot speak of a fixed set Γ either, since the open hypotheses change as a derivation grows: ($\to$I) discharges one, so a subderivation may have an open hypothesis that the whole derivation lacks. The property must speak of each derivation's own open hypotheses. For a derivation $\mathcal{D}$ with conclusion φ:

> (S) for every structure $\mathfrak{M}$ and every assignment s into $\mathfrak{M}$,
> if every formula of $\mathrm{OH}(\mathcal{D})$ is true in $\mathfrak{M}$ under s, then so is φ.

In words: the conclusion of a derivation can be false only where one of its open hypotheses is.

Theorem 6.39 (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 show (S) for every derivation of NJ, by induction on derivations. The theorem follows from it: if $\Gamma \vdash_{\mathrm{NJ}} \varphi$, some derivation $\mathcal{D}$ has conclusion φ and $\mathrm{OH}(\mathcal{D}) \subseteq \Gamma$; if $\mathfrak{M}, s \models \Gamma$, then every formula of $\mathrm{OH}(\mathcal{D})$ is true in $\mathfrak{M}$ under s, and (S) makes φ true.

In each step below, the derivation $\mathcal{D}$ is displayed with its immediate subderivations as triangles. The induction hypothesis is that (S) holds for the subderivations. We fix a structure $\mathfrak{M}$ and an assignment s, suppose that every formula of $\mathrm{OH}(\mathcal{D})$ is true in $\mathfrak{M}$ under s, and show that the conclusion of $\mathcal{D}$ is.

For the connectives, how the formulas φ, ψ and χ are built does not matter. All that is used is whether each is true or false under s, and what the clause for the connective says, so in those steps the reader may think of φ, ψ and χ as propositional letters, each simply true or false. The quantifier rules are different: they look inside a formula, and they need the relevance and substitution lemmas.

The base. The derivation is a single leaf φ, whose only open hypothesis is φ itself, and φ is true by supposition.

The steps for the connectives.

- *($\to$I).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}[\varphi]\\ \mathcal{D}'\\ \psi\end{array}}{\varphi \to \psi}\to\mathrm{I}$$

  and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}')$ with φ removed. If φ is false under s, then $\varphi \to \psi$ is true by the clause for $\to$, whether or not ψ is. If φ is true, then every formula of $\mathrm{OH}(\mathcal{D}')$ is true, since each is φ or in $\mathrm{OH}(\mathcal{D})$; so ψ is true by the induction hypothesis for $\mathcal{D}'$, and $\varphi \to \psi$ is true by the clause for $\to$.
- *($\to$E).* The derivation $\mathcal{D}$ is the one displayed before the theorem, and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}_1) \cup \mathrm{OH}(\mathcal{D}_2)$. Every formula of both sets is true under s, so by the induction hypothesis $\varphi \to \psi$ and φ are true. By the clause for $\to$, φ is false or ψ is true; φ is true, so ψ is.
- *($\wedge$I).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{cc}\mathcal{D}_1 & \mathcal{D}_2\\ \varphi & \psi\end{array}}{\varphi \wedge \psi}\wedge\mathrm{I}$$

  and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}_1) \cup \mathrm{OH}(\mathcal{D}_2)$. By the induction hypothesis φ and ψ are true, and by the clause for $\wedge$ so is $\varphi \wedge \psi$.
- *($\wedge$E).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \varphi \wedge \psi\end{array}}{\varphi}\wedge\mathrm{E}$$

  or the same with ψ beneath, and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}')$. By the induction hypothesis $\varphi \wedge \psi$ is true, and by the clause for $\wedge$ both φ and ψ are.

- *($\vee I$).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \varphi\end{array}}{\varphi \vee \psi}\vee\mathrm{I}$$

  or the same with ψ in the triangle, and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}')$. By the induction hypothesis φ is true, and by the clause for $\vee$ so is $\varphi \vee \psi$.
- *($\vee E$).* The derivation $\mathcal{D}$ is

$$\dfrac{\varphi \vee \psi \qquad \begin{array}{c}[\varphi]\\ \mathcal{D}_1\\ \chi\end{array} \qquad \begin{array}{c}[\psi]\\ \mathcal{D}_2\\ \chi\end{array}}{\chi}\vee\mathrm{E}$$

  and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}_0)$, together with $\mathrm{OH}(\mathcal{D}_1)$ without φ and $\mathrm{OH}(\mathcal{D}_2)$ without ψ. The formulas of $\mathrm{OH}(\mathcal{D}_0)$ are true, so $\varphi \vee \psi$ is true by the induction hypothesis, and by the clause for $\vee$, φ is true or ψ is. If φ is, every formula of $\mathrm{OH}(\mathcal{D}_1)$ is true, since each is φ or in $\mathrm{OH}(\mathcal{D})$, and χ is true by the induction hypothesis for $\mathcal{D}_1$. If ψ is, the same argument applies to $\mathcal{D}_2$.
- *($\bot E$).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \bot\end{array}}{\varphi}\bot\mathrm{E}$$

  and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}')$. By the induction hypothesis $\bot$ is true under s, which its clause forbids. So no $\mathfrak{M}, s$ makes every formula of $\mathrm{OH}(\mathcal{D})$ true, and (S) holds for $\mathcal{D}$ because there is no case to check.

The steps for the quantifiers.

- *($\forall E$).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \forall x\, \varphi\end{array}}{\varphi[t/x]}\forall\mathrm{E}$$

  with t substitutable for x in φ, and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}')$. By the induction hypothesis $\forall x\, \varphi$ is true under s, so by the clause for $\forall$ the formula φ is true under $s[d/x]$ for every element d, in particular for $d = \mathrm{Val}_s(t)$, and by the substitution lemma $\varphi[t/x]$ is true under s.
- *($\exists I$).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \varphi[t/x]\end{array}}{\exists x\, \varphi}\exists\mathrm{I}$$

  with t substitutable for x in φ, and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}')$. By the induction hypothesis $\varphi[t/x]$ is true under s, by the substitution lemma φ is

true under $s[\mathrm{Val}_s(t)/x]$, and by the clause for $\exists$, with $d = \mathrm{Val}_s(t)$, the formula $\exists x\,\varphi$ is true under s.

- *($\forall I$).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \varphi[y/x]\end{array}}{\forall x\,\varphi}\forall\mathrm{I}$$

where y is free in no formula of $\mathrm{OH}(\mathcal{D}')$, which is $\mathrm{OH}(\mathcal{D})$, and not free in $\forall x\,\varphi$. We must show that φ is true under $s[d/x]$ for every element d. Let d be an element and $s' = s[d/y]$. Since y is free in no formula of $\mathrm{OH}(\mathcal{D}')$,

the formulas of $\mathrm{OH}(\mathcal{D}')$ are true under s'	by relevance,
$\varphi[y/x]$ 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 kept the open hypotheses true when the value of y was changed to d. Had y been free in one of them, the change might have made it false, and the induction hypothesis would have told us nothing about s'.

- *($\exists E$).* The derivation $\mathcal{D}$ is

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \exists x\,\varphi\end{array}\qquad\begin{array}{c}[\varphi[y/x]]\\ \mathcal{D}_2\\ \chi\end{array}}{\chi}\exists\mathrm{E}$$

where y is free neither in χ, nor in $\exists x\,\varphi$, nor in the formulas of $\mathrm{OH}(\mathcal{D}_2)$ other than $\varphi[y/x]$; and $\mathrm{OH}(\mathcal{D})$ is $\mathrm{OH}(\mathcal{D}_1)$ together with $\mathrm{OH}(\mathcal{D}_2)$ without $\varphi[y/x]$. By the induction hypothesis for $\mathcal{D}_1$, $\exists x\,\varphi$ is true, so there is an element d with φ true under $s[d/x]$. Let $s' = s[d/y]$. Then

$\varphi[y/x]$ is true under s'	by relevance and substitution,
the rest of $\mathrm{OH}(\mathcal{D}_2)$ is true under s'	by relevance,
χ is true under s'	by the induction hypothesis,
χ is true under s	by relevance.

The first line uses that y is not free in $\exists x\,\varphi$, unless it is x; the second that y is free in no other open hypothesis; the last that y is not free in χ.

This completes the induction, and with it the proof. □

Four rules carried all the weight. Two discharge a supposition, ($\to$I) and ($\vee$E), and there the open hypotheses of the whole are fewer than those of the parts. Two have a side condition on a variable, ($\forall$I) and ($\exists$E), and there the relevance and substitution lemmas were needed. The other seven follow from the clause for their connective or quantifier, two of them with

the substitution lemma as well. Exercise 6.13 tries the method on a rule that is not sound, to see where it breaks.

Example 6.40 (the company, cont'd: a derivation)**.** The entailment of Example 6.37 has a derivation:

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

It is the argument of that example written as a tree. 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:

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

The ($\forall$I) step is legal because, once both suppositions are discharged, the only open hypotheses are the sentences R_1, R_2 and $\mathrm{E}_{\mathsf{Richer}}$. Compare the tree with the argument of Example 6.37: "let d be an element" is the ($\forall$I)

at the root, "suppose that" is each (→I), and "by R_2" is a (∀E) followed by a (→E). By soundness the tree is a second proof of the entailment, and one that a program can check.

6.6. **The Law of the Excluded Middle.** Is NJ also complete? It is not.

Theorem 6.41 (incompleteness of NJ)**.** *There are a set of formulas* Γ *and a formula* φ *such that* $\Gamma \models \varphi$ *but* $\Gamma \nvdash_{\mathrm{NJ}} \varphi$*.*

The proof uses a property of NJ proved in Part 1: a derivation of a disjunction from no hypotheses can be straightened into one that ends with (∨I), and so it says which disjunct holds.

Fact 6.42 (disjunction property, from Part 1)*.* If $\vdash_{\mathrm{NJ}} \psi \vee \chi$, then $\vdash_{\mathrm{NJ}} \psi$ or $\vdash_{\mathrm{NJ}} \chi$.

Proof of the theorem. Take Γ empty, and φ the sentence $p \vee \neg p$ for a propositional letter p.

The sentence is valid. In every structure, under every assignment, p is true or false, and when it is false, $\neg p$ is true.

It has no derivation in NJ. Suppose that $\vdash_{\mathrm{NJ}} p \vee \neg p$. By the disjunction property (Fact 6.42), 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 some structure makes it true and another makes it 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 not confined to excluded middle: the next example meets it again.

Remark 6.43 (two senses of incompleteness)*.* The word *incompleteness* in the theorem 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 6.44 (the company, cont'd: one negation too many)**.** The finance director could have put his question the other way round: not "nobody but the chief executive is her own boss" but "whoever is their own boss is the chief executive",

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

The two sentences are true in the same structures, each being the other's contrapositive, and Θ entails the second as it entails the first. Let $\mathfrak{M}$ be a model of Θ and d an element with $(\llbracket\mathsf{boss}\rrbracket^{\mathfrak{M}}(d), d) \in \llbracket=\rrbracket^{\mathfrak{M}}$. Then:

- *either* d *is the chief executive or not*;

- if so, there is nothing to show;
- if not, the argument of Example 6.37 gives a contradiction, so d is the chief executive after all.

Can NJ derive it?

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

The rules do not provide the *either ... or not*. They do provide the following:

- suppose $\mathsf{boss}(x) = x$ and $x \neq \mathsf{ceo}$;
- then, as in Example 6.40, absurdity;
- hence $\neg(x \neq \mathsf{ceo})$, which is $\neg\neg(x = \mathsf{ceo})$.

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

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 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:

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

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).

---

[7] Showing that no derivation exists needs a different kind of structure, described in the section on classical and intuitionistic logic.

[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?

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

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

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

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

Definition 6.45 (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 6.46 (the company, cont'd: a derivation in NK)**.** The sentence of Example 6.44 is derivable in NK: $\Theta \vdash_{\text{NK}} \forall x\,(\mathsf{boss}(x) = x \to x = \mathsf{ceo})$. The derivation is the argument of that example written as a tree, and most of it is already built: the piece $\mathcal{E}$ of Example 6.40 derives $\bot$ from $x \neq \mathsf{ceo}$ and $\mathsf{boss}(x) = x$, with R_1, R_2 and $\text{E}_{\mathsf{Richer}}$. The new step splits on $x = \mathsf{ceo}$:

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

The left case of the ($\vee$E) needs no work. The right case is the old derivation of $\bot$, followed by ($\bot$E) to reach $x = \mathsf{ceo}$, the conclusion both cases of the split must share. The ($\forall$I) step is legal for the same reason as before: only R_1, R_2 and $\text{E}_{\mathsf{Richer}}$ remain open.

Theorem 6.47 (soundness of NK)**.** *For every set* Γ *of formulas and every formula* φ*, if* $\Gamma \vdash_{\text{NK}} \varphi$ *then* $\Gamma \models \varphi$*.*

Proof. The soundness proof for NJ goes through with one more case, the rule (LEM): its conclusion $\varphi \vee \neg\varphi$ rests on no open hypotheses, so it 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. □

[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.

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 6.48 (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.

How might one prove such a thing? The theorem is an implication, and its antecedent is the hard side to use. $\Gamma \models \varphi$ says that φ is true in every model of Γ, and a proof would have to start from that and arrive at a derivation, a finite object that the antecedent gives no hint of. The standard move when an antecedent is hard to use is to prove the contrapositive instead:

$$\text{if } \Gamma \nvdash_{\mathrm{NK}} \varphi \text{ then } \Gamma \not\models \varphi.$$

The antecedent is now a failure to derive, which seems to carry no information at all, but the conclusion has become easier to reach. $\Gamma \not\models \varphi$ asks only for *one* structure, a model of Γ in which φ is false, and a single structure is something we can hope to build.

Both sides of the contrapositive can be rephrased in terms of one set, $\Gamma \cup \{\neg\varphi\}$. The conclusion says that some structure satisfies every member of Γ and falsifies φ; by the clause for negation, a structure falsifies φ exactly when it satisfies $\neg\varphi$, so the conclusion says that $\Gamma \cup \{\neg\varphi\}$ has a model. The antecedent says that φ has no derivation from Γ, and it too is a fact about $\Gamma \cup \{\neg\varphi\}$: that $\bot$ has no derivation from it. For if the set derived $\bot$, ($\to$I) would turn that derivation into one of $\neg\neg\varphi$ from Γ, and (DN) into one of φ. A set from which $\bot$ cannot be derived has a name.

Definition 6.49 (consistency)**.** A consistent set is one from which the rules cannot derive absurdity.

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

Keep the two notions 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 already gives one direction: a set with a model is consistent, since $\bot$ is true in no structure and a derivation of it would make $\bot$ true in the model. We record the two rephrasings as a lemma.

Lemma 6.50 (two reformulations)**.** *Let Γ be a set of sentences and φ a sentence.*

[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.

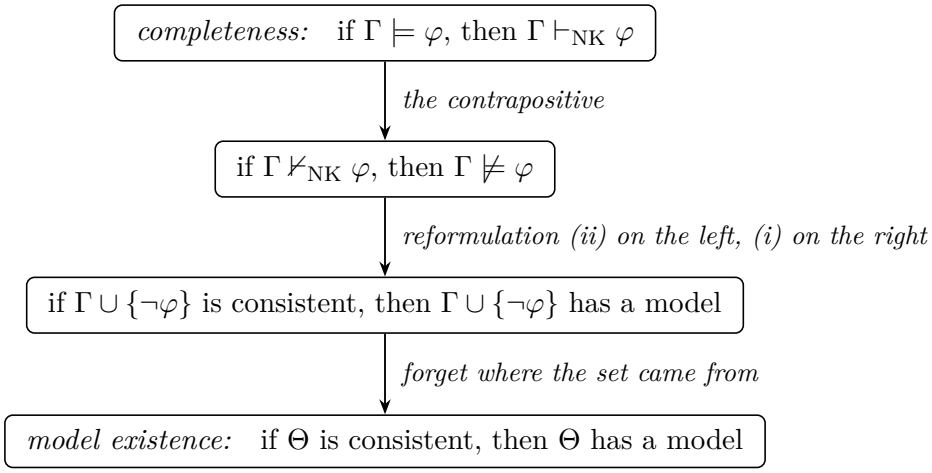

FIGURE 6.8. From completeness to model existence.

(i) $\Gamma \not\models \varphi$ *if and only if* $\Gamma \cup \{\neg\varphi\}$ *has a model.*
(ii) $\Gamma \nvdash_{\mathrm{NK}} \varphi$ *if and only if* $\Gamma \cup \{\neg\varphi\}$ *is consistent.*

Proof.
- *(i)* A structure is a model of Γ in which φ is false exactly when it is a model of Γ and of $\neg\varphi$, by the clause for negation.
- *(ii)* If $\Gamma \cup \{\neg\varphi\}$ derived $\bot$, then Γ would derive $\neg\neg\varphi$ by ($\to$I), and φ by (DN). Conversely, if Γ derived φ, then $\Gamma \cup \{\neg\varphi\}$ would derive $\bot$ by ($\to$E).

□

Put together, the contrapositive reads: *if* $\Gamma \cup \{\neg\varphi\}$ *is consistent, then it has a model.* The set $\Gamma \cup \{\neg\varphi\}$ is just some set of sentences, and nothing about how it arose will be of any use in building its model. So we may as well drop the Γ and the φ and prove the statement for every set of sentences. Figure 6.8 records the chain of rephrasings.

Lemma 6.51 (model existence)**.** *Let* Σ *be a countable signature. Every consistent set of* Σ*-sentences has a model.*

In words: a theory that does not contradict itself describes something. The lemma is the whole of the completeness theorem in another form, and the rest of the section proves it. Completeness follows from it at once.

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

$\Gamma \cup \{\neg\varphi\}$ is consistent	by reformulation (ii),
$\Gamma \cup \{\neg\varphi\}$ has a model	by the model existence lemma,
$\Gamma \not\models \varphi$	by reformulation (i).

This is the contrapositive of the theorem. □

A theory that behaves like a model. We are handed a consistent set of sentences Θ and must produce a structure in which every one of them is true.

$\bot \notin \Theta$		
$\psi \wedge \chi \in \Theta$	iff	$\psi \in \Theta$ and $\chi \in \Theta$
$\psi \vee \chi \in \Theta$	iff	$\psi \in \Theta$ or $\chi \in \Theta$
$\psi \to \chi \in \Theta$	iff	$\psi \notin \Theta$ or $\chi \in \Theta$
$\forall x\, \psi \in \Theta$	iff	$\psi[t/x] \in \Theta$ for every closed term t
$\exists x\, \psi \in \Theta$	iff	$\psi[t/x] \in \Theta$ for some closed term t

FIGURE 6.9. The clauses of a model-like theory.

The only material in sight is Θ itself, so we build the structure out of what Θ says. Its elements will be the names that the signature provides, the closed terms, and a relation will hold of some names exactly when Θ says that it does. A constant c will name the element c, the term $\mathsf{boss}(c)$ the element $\mathsf{boss}(c)$, and so on: the structure is made of its own names. This is the boldest idea in the proof, and the rest of the section makes it work.

For which theories can it work? Look at the set of sentences true in a structure in which every element has a name. Satisfaction says how that set behaves: it contains $\psi \wedge \chi$ exactly when it contains both conjuncts, $\psi \vee \chi$ exactly when it contains at least one disjunct, $\forall x\, \psi$ exactly when it contains every instance $\psi[t/x]$, and so on through the clauses. A theory that behaved in the same way would carry all the information that a structure carries, and nothing more would be needed to read a structure off it. So we make the clauses into a definition.

Definition 6.52 (model-like theory)**.** A model-like theory treats the connectives and quantifiers as satisfaction does, with membership in place of truth and closed terms in place of elements.

Formally, a set Θ of Σ-sentences is *model-like* if the clauses of Figure 6.9 hold for all Σ-sentences ψ and χ, and, in the two quantifier clauses, for every Σ-formula ψ whose only free variable is x.

Set Figure 6.9 beside Figure 6.5: it is the definition of satisfaction with "$\in \Theta$" in place of "$\mathfrak{M}, s \models$", and closed terms in place of elements of the domain. In words, a model-like theory is the list of truths of a structure, kept by someone who has the list and nothing else. Since it behaves like a model, we may as well make it into one.

Definition 6.53 (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 Σ be a signature with at least one constant symbol, and Θ a set of Σ-sentences. The *term structure* of Θ is the Σ-structure $\mathfrak{M}_\Theta = (\mathbb{D}_\Theta, I_\Theta)$ in which

- $\mathbb{D}_\Theta$ 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,

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

The domain is not empty, since the signature has a constant symbol. By induction on terms, every closed term t names itself: $\mathrm{Val}_s(t) = t$ under every assignment s. So every element of the domain has a name, and by the substitution lemma $\mathfrak{M}_\Theta \models \forall x\, \psi$ holds exactly when $\mathfrak{M}_\Theta \models \psi[t/x]$ for every closed term t, and $\mathfrak{M}_\Theta \models \exists x\, \psi$ exactly when $\mathfrak{M}_\Theta \models \psi[t/x]$ for some closed term t. The difficulty of an element without a name cannot arise, because the elements *are* the names.

Lemma 6.54 (truth lemma)**.** *Let Σ have at least one constant symbol and Θ be a model-like set of Σ-sentences. For every Σ-sentence ψ, $\mathfrak{M}_\Theta \models \psi$ if and only if $\psi \in \Theta$.*

In words: in the structure made of names, a sentence is true exactly when Θ 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.) In each case a clause of satisfaction meets the matching clause of Figure 6.9.

- *Atomic sentences.* Since every closed term names itself, $\mathfrak{M}_\Theta \models R(t_1, \ldots, t_n)$ exactly when $(t_1, \ldots, t_n) \in [\![R]\!]^{\mathfrak{M}_\Theta}$, that is, by the definition of $[\![R]\!]^{\mathfrak{M}_\Theta}$, exactly when $R(t_1, \ldots, t_n) \in \Theta$.
- $\bot$. Neither side holds: $\bot$ is true nowhere, and $\bot \notin \Theta$.
- $\psi \vee \chi$.

	$\mathfrak{M}_\Theta \models \psi \vee \chi$	
iff	$\mathfrak{M}_\Theta \models \psi$ or $\mathfrak{M}_\Theta \models \chi$	by the clause of satisfaction for $\vee$
iff	$\psi \in \Theta$ or $\chi \in \Theta$	by the induction hypothesis
iff	$\psi \vee \chi \in \Theta$	by the clause for $\vee$ of Figure 6.9.

- $\psi \wedge \chi$ *and* $\psi \to \chi$. The same, with the clauses for $\wedge$ and $\to$.
- $\forall x\, \psi$.

	$\mathfrak{M}_\Theta \models \forall x\, \psi$	
iff	$\mathfrak{M}_\Theta \models \psi[t/x]$ for every closed t	since every element is a name
iff	$\psi[t/x] \in \Theta$ for every closed t	by the induction hypothesis
iff	$\forall x\, \psi \in \Theta$	by the clause for $\forall$ of Figure 6.9.

- $\exists x\, \psi$. The same, with "some" for "every".

□

Corollary 6.55 (a model-like theory has a model)**.** *Over a signature with at least one constant symbol, every model-like set of sentences Θ has a model, its term structure $\mathfrak{M}_\Theta$.*

Proof. By the truth lemma, every member of Θ is true in $\mathfrak{M}_\Theta$. □

What a consistent theory already has. The question is now how to turn a given consistent theory into a model-like one. Some of the work is free. Call a set of sentences Θ *closed under derivation* if every sentence that NK derives from Θ is in Θ. Closing a consistent theory costs nothing: add everything it derives, and the result is still consistent, since whatever the enlarged set derives, the original set already derived. A consistent theory closed under derivation already satisfies half of Figure 6.9, and each half-clause comes from a single rule.

Lemma 6.56 (what closure gives)**.** *Let Θ be consistent and closed under derivation. Then, for all sentences ψ and χ and every formula ψ whose only free variable is x:*

(i) $\bot \notin \Theta$*;*
(ii) $\psi \wedge \chi \in \Theta$ *if and only if* $\psi \in \Theta$ *and* $\chi \in \Theta$*;*
(iii) if $\psi \in \Theta$ *or* $\chi \in \Theta$*, then* $\psi \vee \chi \in \Theta$*;*
(iv) if $\psi \to \chi \in \Theta$ *and* $\psi \in \Theta$*, then* $\chi \in \Theta$*; and if* $\chi \in \Theta$*, then* $\psi \to \chi \in \Theta$*;*
(v) if $\forall x\, \psi \in \Theta$*, then* $\psi[t/x] \in \Theta$ *for every closed term* t*;*
(vi) if $\psi[t/x] \in \Theta$ *for some closed term* t*, then* $\exists x\, \psi \in \Theta$*.*

Proof. Part (i) is consistency, since a leaf $\bot$ would be a derivation of $\bot$. In each other part, the rule named derives the sentence that the part says is in Θ from the sentences it is said to follow from, and closure puts it in Θ:

- *(ii)* ($\wedge$E) from left to right, ($\wedge$I) from right to left;
- *(iii)* ($\vee$I);
- *(iv)* ($\to$E), and ($\to$I) discharging nothing;
- *(v)* ($\forall$E);
- *(vi)* ($\exists$I).

□

What is missing is the other half of four clauses:

- $\vee$*:* from $\psi \vee \chi \in \Theta$ to $\psi \in \Theta$ or $\chi \in \Theta$;
- $\to$*:* from $\psi \notin \Theta$ to $\psi \to \chi \in \Theta$;
- $\exists$*:* from $\exists x\, \psi \in \Theta$ to some instance $\psi[t/x] \in \Theta$;
- $\forall$*:* from every instance $\psi[t/x] \in \Theta$ to $\forall x\, \psi \in \Theta$.

These four halves come down to two pieces of information that a theory can lack and a structure cannot: which disjunct, and which witness. A consistent theory can contain $p \vee q$ without p or q, and $\exists x\, P(x)$ without any $P(t)$. The first two halves need a theory that decides every sentence, and the Lindenbaum construction supplies one. The last two need named witnesses, and the Henkin construction supplies them.

Deciding every sentence: the Lindenbaum construction.

Definition 6.57 (maximal consistency)**.** A maximal consistent set is a consistent set that takes a side on every sentence.

Formally, a set Δ of Σ-sentences is *maximal consistent* if

- it is consistent, and
- for every Σ-sentence ψ, either $\psi \in \Delta$ or $\neg\psi \in \Delta$.

Lemma 6.58 (Lindenbaum)**.** *Let Σ be countable. Every consistent set Θ of Σ-sentences is contained in a maximal consistent set of Σ-sentences, which we write $\mathrm{L}(\Theta)$.*

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 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 $\Delta_0 = \Theta$ 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 $\mathrm{L}(\Theta) = \bigcup_{n\in\mathbb{N}} \Delta_n$.

- Every sentence or its negation belongs to $\mathrm{L}(\Theta)$, since ψ_n was decided at step n.
- Each Δ_n is consistent, by induction on n. The set $\Delta_0 = \Theta$ is consistent by hypothesis. If Δ_n is consistent, so is Δ_{n+1}: 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 $\bot$.
- So $\mathrm{L}(\Theta)$ is consistent: a derivation of $\bot$ from it uses finitely many hypotheses, and they all 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 the question is to be answered; it needs only that the question has an answer. So the proof shows that $\mathrm{L}(\Theta)$ exists without constructing it. Nor is the result unique, since another order of the list can give another maximal consistent set; $\mathrm{L}(\Theta)$ names the result for one list, fixed once and for all, and the proof needs only one.

A set that takes a side on every sentence supplies the two propositional clauses.

Lemma 6.59 (maximal consistent sets)**.** *Let Δ be maximal consistent. Then*

(i) Δ is closed under derivation;
(ii) if $\psi \vee \chi \in \Delta$, then $\psi \in \Delta$ or $\chi \in \Delta$;
(iii) if $\psi \notin \Delta$, then $\psi \to \chi \in \Delta$.

In words: a maximal consistent set satisfies every clause of Figure 6.9 except, perhaps, the two for the quantifiers.

Proof. In each part, suppose the conclusion fails and derive $\bot$ from Δ, against consistency.

- *(i)* If Δ derives ψ but $\psi \notin \Delta$, then $\neg\psi \in \Delta$, and ($\to$E) derives $\bot$.

- *(ii)* If $\psi \vee \chi \in \Delta$ and neither disjunct is, then $\neg\psi$ and $\neg\chi$ are, and ($\vee$E) derives $\bot$ from the three.
- *(iii)* If $\psi \notin \Delta$, then $\neg\psi \in \Delta$. From a supposition ψ, the rule ($\to$E) gives $\bot$ and ($\bot$E) gives χ, so ($\to$I) derives $\psi \to \chi$ from Δ, and it is in Δ by (i).

□

Together with the lemma on what closure gives, this lemma leaves only the quantifier clauses. They can fail in a maximal consistent set, because the closed terms may not name enough elements: the set may contain $\exists x\, \psi$ and the negation of every instance of ψ (Example 6.68 at the end of the section shows a case).

Naming the witnesses: the Henkin construction. The remedy is to add names: a new constant symbol for every existential sentence, and a sentence that makes it the witness, $\exists x\, \psi \to \psi[c/x]$. Provided that each name is new, the enlarged set stays consistent. The lemma that says so rests on a fact about derivations: a constant about which nothing is assumed is as good as a variable.

Lemma 6.60 (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 6.61 (Henkin)**.** *Let* Σ *be countable and* Θ *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 the set

$$\mathrm{H}(\Theta) \coloneqq \Theta \cup \{\theta_0, \theta_1, \theta_2, \ldots\}$$

of Σ^+*-sentences is consistent.*

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

Proof. First, Θ remains consistent when derivations may use the new constants: replacing each new constant in a derivation of $\bot$ from Θ by a variable that occurs nowhere in it yields a derivation of $\bot$ from Θ over Σ, since no rule mentions a constant. A derivation of $\bot$ uses finitely many hypotheses, so it is enough to show that every $\Theta_n = \Theta \cup \{\theta_0, \ldots, \theta_{n-1}\}$ is consistent. We do so by induction on n. The set $\Theta_0 = \Theta$ is consistent, as just shown; it remains to show that if Θ_n is consistent, then so is $\Theta_n \cup \{\theta_n\}$, which is Θ_{n+1}.

Suppose that $\Theta_n \cup \{\theta_n\}$ is inconsistent. Then Θ_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 Θ_n nor in ψ_n. So Θ_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 Θ_n is inconsistent, contrary to assumption. □

Remark 6.62 (why the witnesses must be new). Two small sets show what the proviso 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 would derive $P(c)$ and $\neg P(c)$. A name already in use carries information, and a witness must carry none.

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) in another guise: the witness must be a name about which nothing has been assumed.

Once the witnesses are named, a maximal consistent set supplies the two quantifier clauses.

Lemma 6.63 (the quantifier clauses)**.** *Let Δ be a maximal consistent set of Σ^+-sentences that contains $\mathrm{H}(\Theta)$. Then, for every Σ^+-formula ψ whose only free variable is x:*

(i) if $\exists x\,\psi \in \Delta$, then $\psi[t/x] \in \Delta$ for some closed term t;

(ii) if $\psi[t/x] \in \Delta$ for every closed term t, then $\forall x\,\psi \in \Delta$.

Proof. Recall that Δ is closed under derivation, by the lemma on maximal consistent sets.

- *(i)* The sentence $\exists x\,\psi$ is some $\exists x_n\,\psi_n$ of the list, and $\theta_n \in \mathrm{H}(\Theta) \subseteq \Delta$. By ($\to$E), Δ derives $\psi_n[d_n/x_n]$, which is in Δ by closure, and d_n is a closed term.
- *(ii)* Suppose that every instance $\psi[t/x]$ is in Δ but $\forall x\,\psi$ is not. Then

$\neg\forall x\,\psi \in \Delta$	since Δ takes a side,
$\exists x\,\neg\psi \in \Delta$	by Exercise 6.10 and closure,
$\neg\psi[t/x] \in \Delta$ for some closed t	by (i),

  and with $\psi[t/x] \in \Delta$ the rule ($\to$E) derives $\bot$, against consistency.

□

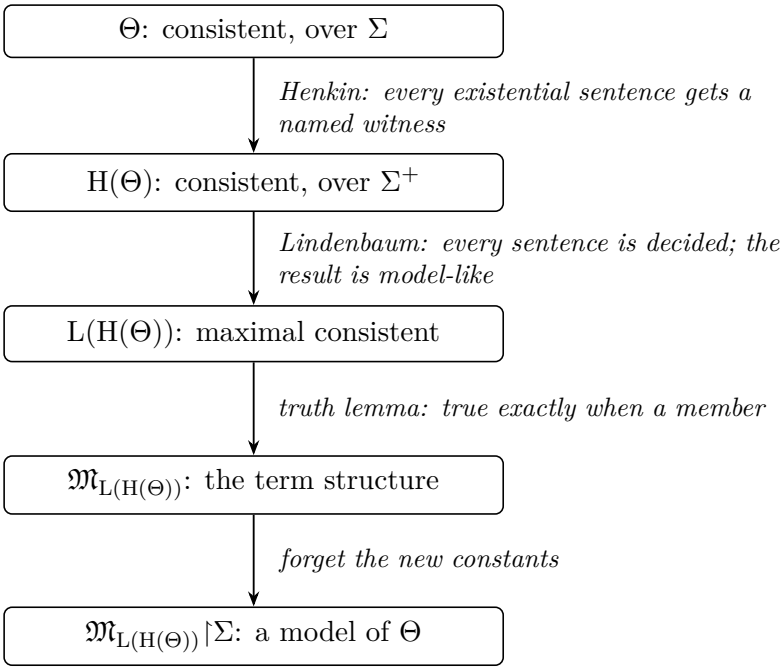

FIGURE 6.10. From a consistent theory to a model.

Putting it together. Henkin's construction names the witnesses, Lindenbaum's then decides every sentence, and the three lemmas just proved say that the result is model-like. Figure 6.10 draws the route.

Proof of the model existence lemma. Let Θ be a consistent set of Σ-sentences, with Σ countable.

1. $\mathrm{H}(\Theta)$ is a consistent set of Σ^+-sentences, by the Henkin lemma, and Σ^+ is countable, having gained countably many constants.
2. $\Delta = \mathrm{L}(\mathrm{H}(\Theta))$ is a maximal consistent set of Σ^+-sentences containing $\mathrm{H}(\Theta)$, by Lindenbaum's lemma.
3. Δ is model-like over Σ^+. It is consistent, and closed under derivation by the lemma on maximal consistent sets, so the lemma on what closure gives supplies half of every clause of Figure 6.9. The lemma on maximal consistent sets supplies the missing halves for $\vee$ and $\to$, and the lemma on the quantifier clauses those for $\exists$ and $\forall$.
4. The term structure $\mathfrak{M}_\Delta$ exists, since Σ^+ has constant symbols, and by the truth lemma a Σ^+-sentence is true in it exactly when it belongs to Δ. In particular every member of Θ is true in $\mathfrak{M}_\Delta$, since $\Theta \subseteq \mathrm{H}(\Theta) \subseteq \Delta$.
5. The new constants occur in no sentence of Θ, 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. The clauses of satisfaction for a Σ-sentence look only at the symbols of Σ, so the same Σ-sentences are true in it, and it is a model of Θ.

□

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 6.64 (where excluded middle was used). The law entered the proof in two places, and nowhere else: in reformulation (ii), through (DN), and in the lemma on the quantifier clauses, in the passage from $\neg\forall x\,\psi$ to $\exists x\,\neg\psi$, a relative of the principle $\neg\forall x\,\neg\varphi \to \exists x\,\varphi$ that NJ could not derive. At the first of these the argument already breaks down for NJ. With Γ empty and φ the sentence $p \vee \neg p$, reformulation (ii) fails for NJ: the system does not derive $p \vee \neg p$, yet $\{\neg(p \vee \neg p)\}$ is inconsistent, by the tree that derives $\neg\neg(p \vee \neg p)$.

Remark 6.65 (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*). For if Γ had no model, it would be inconsistent, by the model existence lemma; a derivation of $\bot$ from Γ uses only finitely many hypotheses, so some finite subset of Γ would be inconsistent, and by soundness it would have no model. 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 tell structures apart less well than one might hope.

Remark 6.66 (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]

Example 6.67 (the Lindenbaum 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 would not: q and $q \to r$ give r, which clashes with $\neg r$, so $\neg q$ goes in.

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

- Adding r meets $\neg r$, so $\neg r$ goes in again.

A structure that makes p true and q and r false, as the three decisions say, does make $p \vee q$ and $q \to r$ true and r false. The decisions were taken, one letter at a time, from what the hypotheses could not rule out, and they show that the entailment fails.

The first step asked whether $\{p \vee q, q \to r, \neg r, p\}$ is consistent. We answered yes because this very structure makes all of it true; the construction itself only asks the question and does not say how to answer it.

Example 6.68 (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 $\mathfrak{C}$ is maximal consistent, since it has a model and takes a side on every sentence. It contains $\exists x\, \mathsf{Eng}(x)$, and also $\neg\mathsf{Eng}(t)$ for every closed term t, so the clause for $\exists$ of Figure 6.9 fails in it: the set is not model-like. Its term structure would contain no engineer. This is the company's version of the remark on unnamed elements.

Example 6.69 (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 remark on normal structures shows that looking only at such structures loses nothing.

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 between derivability and entailment, 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 add derivations, and the new derivations cost NK two properties that NJ has in Part 1. One is the disjunction property (Fact 6.42), used in the proof that NJ is incomplete. The other, also proved in Part 1, says that a proof of "there is an x such that φ" from no hypotheses names a term for x.

Fact 6.70 (existence property, from Part 1). If $\vdash_{\mathrm{NJ}} \exists x\, \varphi$, then $\vdash_{\mathrm{NJ}} \varphi[t/x]$ for some term t.

Remark 6.71 (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]

[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, one drinking and one not, 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 mathematicians refuse to accept the argument.

Example 6.72 (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 ($\to$I) step discharges nothing, since $\forall y\, \mathsf{Dr}(y)$ was reached without supposing $\mathsf{Dr}(x)$:

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

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

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

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 instance of (LEM) is true, and the derivation does not say.

Proposition 6.73 (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 6.72 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*, see the loss of witnesses as the heart of the matter. 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 each connective its meaning by *proof* conditions, which say 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 6.44 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, the argument concludes, 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 still open; 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 mathematicians read 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, which let a user add the law as an axiom or, in Lean, through the axiom of choice.[22] Figure 6.11 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 6.11. Two logics for one language.

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

Frege built his language 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. The section on classical and intuitionistic logic describes the logic that Brouwer's answer produced. Hilbert's is told here.

[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).

Hilbert proposed to secure classical mathematics by treating its proofs as finite strings of symbols and proving their consistency 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 does not argue that Brouwer was wrong, only that Brouwer's 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.

The clause for the conditional is older than the rest of the definition. Philo of Megara, around 300 BC, held that a conditional is false only when it leads from a truth to a falsehood, against his teacher Diodorus Cronus, who asked that it never could; the dispute is known through Sextus Empiricus. Frege's conditional is Philo's, and Russell called it *material implication*. Its consequences have been called paradoxes ever since (see the remark on the conditional of science): on the material reading a true sentence is implied by anything, $\psi \to (\varphi \to \psi)$, a false one implies anything, $\neg\varphi \to (\varphi \to \psi)$, and of any two sentences one implies the other, $(\varphi \to \psi) \vee (\psi \to \varphi)$. All three are valid, though "implies" in English seems to promise a connection between the two sentences. C. I. Lewis took the paradoxes for a defect and in 1912 proposed a *strict* implication, on which φ implies ψ when it is impossible that φ hold and ψ fail; the relevance logics of the 1950s to 1970s went further and asked that the antecedent actually be used in reaching the consequent.[28]

[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.

[28] Lewis (1912); Anderson and Belnap (1975); Read (1988) is a readable account.

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 6.12. Classical logic and its semantics: who did what.

Defenders of the material reading answer that the paradoxes mix up what a conditional says with what makes it worth saying, and that the conditional of a law, which holds even of the cases it does not apply to, is exactly the material one; Williamson (2020) is a recent defence of this kind.

Tarski gave this picture a precise form. His definition of truth, published in 1933, is our definition of satisfaction, and his definition of consequence, presented in 1935 at the Paris congress of the Unity of Science movement, is our definition of entailment. 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 completeness 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 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 6.12 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 for each kind of formula. 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. 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. Soundness says that in a correct derivation no line is false while all its open hypotheses are true; find the first line of this derivation that is false while 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. (a) Someone proposes a rule ($\vee\mathrm{E}'$) that concludes φ from $\varphi \vee \psi$, with no case analysis. Try to write its case in the soundness proof, say where the attempt breaks, and give a structure that shows the rule unsound.

(b) The derived step (DN) concludes φ from $\neg\neg\varphi$. Write the case it would have in the soundness proof if it were a rule. Which clause of satisfaction does the case use, and why does the same case not go through for a rule concluding φ from $\neg\varphi$?

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) Write out the case $\exists x\, \psi$ of the truth lemma.

(b) Let Δ be the set of sentences true in a structure $\mathfrak{M}$ in which every element is named by a closed term. Show that Δ is model-like. Where is the hypothesis on names 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 Henkin lemma does the second added sentence break?

References.

Alan Ross Anderson and Nuel D. Belnap, Jr. *Entailment: The Logic of Relevance and Necessity*, volume I. Princeton University Press, Princeton, NJ, 1975.

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.

C. I. Lewis. Implication and the algebra of logic. *Mind*, 21(84):522–531, 1912.

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.

Stephen Read. *Relevant Logic: A Philosophical Examination of Inference*. Basil Blackwell, Oxford, 1988.

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.

Timothy Williamson. *Suppose and Tell: The Semantics and Heuristics of Conditionals*. Oxford University Press, Oxford, 2020.