10. The method of trees for GL.
11. The incompleteness of the system $K + \{\Box(A \leftrightarrow \Box A) \rightarrow \Box A\}$, a simplest possible incomplete modal logic.
12. The system Grz(egorczyk), which extends S4, and its completeness under the interpretation of $\Box$ as meaning "true and provable".
13. Modal logics for three set-theoretical interpretations of $\Box$, under which it is read as "provable in ZF", "true in all transitive models", and "true in all models V_κ, κ inaccessible".
14. The analytical completeness of GL (for provability) and GLS (for truth) with respect to ordinary provability and, more interestingly, provability under unrestricted use of the ω-rule. (Analysis is second-order arithmetic.)
15. The arithmetical completeness of GLB and GLSB.
16. The fixed point theorem for GLB, a normal form theorem for letterless sentences of GLB, and a short discussion of the "analytical" completeness of GLB and GLSB with respect to ordinary provability in analysis and provability in analysis under unrestricted use of the ω-rule.
17. The set of always provable formulas of quantified modal logic and the set of always true formulas are as undecidable as it is possible, a priori, for them to be: Π^0_2-complete and Π^0_1-complete in the set of Gödel numbers of true sentences of arithmetic.
18. The results of Chapter 17 are extended to the case in which modal formulas contain only one one-place predicate letter and nested boxes are forbidden.

1

GL and other systems of propositional modal logic

We are going to investigate a system of propositional modal logic, which we call 'GL', for Gödel and Löb.[1] GL is also sometimes called *provability logic*, but the term is also used to mean modal logic, as applied to the study of provability. By studying GL, we can learn new and interesting facts about *provability* and *consistency*, concepts studied by Gödel in "On formally undecidable propositions of *Principia Mathematica* and related systems I", and about the phenomenon of self-reference.

Like the systems T (sometimes called 'M'), S4, B, and S5, which are four of the best-known systems of modal logic, GL is a *normal* system of propositional modal logic. That is to say, the theorems of GL contain all tautologies of the propositional calculus (including, of course, those that contain the special symbols of modal logic); contain all distribution axioms, i.e., all sentences of the form $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$; and are closed under the rules of modus ponens, substitution, and necessitation, according to which $\Box A$ is a theorem provided that A is. Nor does GL differ from those other systems in the syntax of its sentences: exactly the same sequences of symbols count as well-formed sentences in all five systems.

GL differs greatly from T, S4, B, and S5, however, with respect to basic questions of theoremhood. All sentences $\Box(\Box A \rightarrow A) \rightarrow \Box A$ are axioms of GL. In particular, then, $\Box(\Box p \rightarrow p) \rightarrow \Box p$ and $\Box(\Box(p \wedge \neg p) \rightarrow (p \wedge \neg p)) \rightarrow \Box(p \wedge \neg p)$ are axioms of GL. The other axioms of GL are the tautologies and distribution axioms; its rules of inference are, like those of the other systems, just modus ponens and necessitation.

It follows that either GL is inconsistent or some sentence $\Box A \rightarrow A$ is not a theorem of GL or some sentence $\Box(\Box A \rightarrow A)$ is not a theorem of GL. For if $\Box(\Box A \rightarrow A) \rightarrow \Box A$, $\Box(\Box A \rightarrow A)$, and $\Box A \rightarrow A$ are always theorems of G, then for any sentence A whatsoever, e.g. $(p \wedge \neg p)$, two applications of modus ponens show A to be a theorem of GL, and GL is inconsistent.

It will turn out that GL is perfectly consistent; we shall see quite soon that neither $\Box p \to p$ nor its substitution instance $\Box(p \wedge \neg p) \to (p \wedge \neg p)$ is a theorem of GL and, later, that $\Box(\Box p \to p)$ is also not a theorem.

In order to contrast GL with its better-known relatives, we shall take a general look at systems of propositional modal logic. Much of the material in this chapter may be quite familiar, but it will be important to reverify certain elementary facts in order to establish that they hold in the absence of $\Box p \to p$, which we shall be living without in most of the rest of this book. The material of this chapter will be of a purely syntactic or "proof-theoretical" character. We take up the semantics of modal logic in Chapter 4.

We begin our general look at modal logic by defining the notion of a sentence of propositional modal logic, or "modal sentence" or "sentence" for short.

Modal sentences. Fix a countably infinite sequence of distinct objects, of which the first five are $\bot$, $\to$, $\Box$, (and) and the others are the sentence letters; 'p', 'q', ... will be used as variables over sentence letters. Modal sentences will be certain finite sequences of these objects. We shall use 'A', 'B', ... as variables over modal sentences. Here is the inductive definition of *modal sentence*:

(1) $\bot$ is a modal sentence;
(2) each sentence letter is a modal sentence;
(3) if A and B are modal sentences, so is $(A \to B)$; and
(4) if A is a modal sentence, so is $\Box(A)$.

[We shall very often write: $(A \to B)$ and: $\Box(A)$ as: $A \to B$ and: $\Box A$.]

Sentences that do not contain sentence letters are *letterless*. For example, $\bot$, $\Box\bot$, and $\Box\bot \to \bot$ are letterless sentences.

Since a handy, perfectly general, and non-arbitrary way to say that a system is consistent is simply to say that $\bot$ is not one of its theorems, taking the 0-ary propositional connective $\bot$ to be one of our primitive symbols provides a direct way to represent in the notation of modal logic many interesting propositions expressible in the language of arithmetic concerning consistency and provability. Thus, e.g., the letterless sentence $\neg\Box\bot$ will turn out to represent the proposition that arithmetic is consistent; $\Box\neg\Box\bot$, the proposition that the consistency of arithmetic is provable in arithmetic; and $\neg\Box\bot \to \neg\Box\neg\Box\bot$, the second incompleteness theorem of Gödel.

Of course, with the aid of $\bot$ and $\to$, all connectives of ordinary propositional logic are definable: $\neg p$ may be defined as $(p \to \bot)$, and as is well known, all propositional connectives are definable from $\neg$ and $\to$.

$\wedge$ (and), $\vee$ (or), and $\leftrightarrow$ (iff) are defined in any one of the usual ways. The 0-ary propositional connective $\top$ has the definition $\bot \to \bot$. $\Diamond A$ is defined as $\neg\Box\neg A$, i.e., as $\Box(A \to \bot) \to \bot$.

The inductive definition of *subsentence* of A runs: A is a subsentence of A; if $B \to C$ is a subsentence of A, so is B and so is C; and if $\Box B$ is a subsentence of A, so is B. A sentence letter p *occurs*, or *is contained*, in a sentence A if it is a subsentence of A.

We shall take a system of propositional modal logic to be a set of sentences, the axioms of the system, together with a set of relations on the set of sentences, called the rules of inference of the system. As usual, a proof in a system is a finite sequence of sentences, each of which is either an axiom of the system or deducible from earlier sentences in the sequence by one of the rules of inference of the system. (B is said to be deducible from $A_1, \ldots, A_n$ by the rule of inference R if $\langle A_1, \ldots, A_n, B \rangle$ is in R.) A proof $A, B, \ldots, Z$ is a proof *of* Z, and a sentence is called a theorem of, or provable in, the system if there is a proof of it in the system. We write: $\mathrm{L} \vdash A$ to mean that A is a theorem of the system L.

A set of sentences is said to be *closed* under a rule of inference if it contains all sentences deducible by the rule from members of the set.

Modus ponens is the relation containing all triples $\langle (A \to B), A, B \rangle$.

Necessitation is the relation containing all pairs $\langle A, \Box A \rangle$.

Let F be a sentence. The result $(F_p(A))$ – $F_p(A)$ for short, or even $F(A)$, if the identity of p is clear from context – of substituting A for p in F may be inductively defined as follows:

If $F = p$, then $F_p(A)$ is A;
if F is a sentence letter $q \neq p$, then $F_p(A)$ is q;
if $F = \bot$, then $F_p(A)$ is $\bot$;
$(F \to G)_p(A) = (F_p(A) \to G_p(A))$; and
$\Box(F)_p(A) = \Box(F_p(A))$.

Thus $F_p(A)$ is the result of substituting an occurrence of A for each occurrence of p in F.

A sentence $F_p(A)$ is called a *substitution instance of F*.

Substitution is the relation containing all pairs $\langle F, F_p(A)\rangle$.

Simultaneous substitution. Let $p_1, \ldots, p_n$ be a list of distinct sentence letters, $F, A_1, \ldots, A_n$ a list of sentences. We define the simultaneous substitution $F_{p_1,\ldots,p_n}(A_1, \ldots, A_n)$ analogously:

If $F = p_i$ $(1 \leqslant i \leqslant n)$, then $F_{p_1,\ldots,p_n}(A_1, \ldots, A_n)$ is A_i;
if F is a sentence letter $q \neq p_1, \ldots, p_n$, then $F_{p_1,\ldots,p_n}(A_1, \ldots, A_n)$ is q;
the other cases are as in the previous definition.

Note that $F_p(A)_q(B)$ need not be identical with $F_{p,q}(A, B)$. For example, let $F = (p \wedge q)$, $A = (p \vee q)$, $B = (p \rightarrow q)$. Then $F_p(A) = ((p \vee q) \wedge q)$, and $F_p(A)_q(B) = ((p \vee (p \rightarrow q)) \wedge (p \rightarrow q))$. But $F_{p,q}(A, B) = ((p \vee q) \wedge (p \rightarrow q))$. However, a set of sentences that is closed under (ordinary) substitution is closed under simultaneous substitution. For let $q_1, \ldots, q_n$ be a list of distinct *new* sentence letters, i.e., sentence letters none of which is identical with any of $p_1, \ldots, p_n$ and that occur nowhere in $F, A_1, \ldots, A_n$. Then $F_{p_1,\ldots,p_n}(A_1, \ldots, A_n)$ is identical with

$$F_{p_1}(q_1)_{p_2}(q_2)\ldots_{p_n}(q_n)_{q_1}(A_1)_{q_2}(A_2)\ldots_{q_n}(A_n)$$

any so any set containing F and closed under substitution will contain $F_{p_1}(q_1)$, $F_{p_1}(q_1)_{p_2}(q_2), \ldots$, and $F_{p_1,\ldots,p_n}(A_1, \ldots, A_n)$

A *distribution axiom* is a sentence of the form

$(\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B))$, i.e., a sentence that *is*
$(\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B))$, for some sentences A, B.

A system is called *normal* if the set of its theorems contains all tautologies and all distribution axioms and is closed under modus ponens, necessitation, and substitution. (According to Kripke's original definition,[2] the axioms of a normal system had also to include all sentences $\Box A \rightarrow A$. The definition we have given, which does not impose this further requirement, is now the standard one, however.)

We now present seven systems of modal logic. In each system, all

tautologies and all distribution axioms are axioms and the rules of inference are just modus ponens and necessitation.

The system K, which is named after Kripke, has no other axioms.
The other axioms of the system K4 are the sentences $\Box A \rightarrow \Box\Box A$.
The other axioms of the system T are the sentences $\Box A \rightarrow A$.
The other axioms of the system S4 are the sentences $\Box A \rightarrow A$ and $\Box A \rightarrow \Box\Box A$.
The other axioms of the system B are the sentences $\Box A \rightarrow A$ and $A \rightarrow \Box\Diamond A$.[3]
The other axioms of the system S5 are the sentences $\Box A \rightarrow A$ and $\Diamond A \rightarrow \Box\Diamond A$.
The other axioms of the system GL are the sentences $\Box(\Box A \rightarrow A) \rightarrow \Box A$.

A system L′ *extends* a system L if every theorem of L is a theorem of L′. If we write '⊇' and '⊆' to mean "extends" and "is extended by", then it is evident that we have:

$$\begin{array}{ccccc} & & \mathrm{GL} & & \\ & & \cup\!\shortmid & & \\ & & \mathrm{K} \subseteq & \mathrm{K4} & \\ & & \cap\!\shortmid & \cap\!\shortmid & \\ \mathrm{S5} & \supseteq & \mathrm{T} \subseteq & \mathrm{S4} & \\ & & \cap\!\shortmid & & \\ & & \mathrm{B} & & \end{array}$$

By the end of the chapter we shall have shown that in fact:

$$\begin{array}{ccccc} \mathrm{K} & \subseteq & \mathrm{K4} & \subseteq & \mathrm{GL} \\ \cap\!\shortmid & & \cap\!\shortmid & & \\ \mathrm{T} & \subseteq & \mathrm{S4} & & \\ \cap\!\shortmid & & \cap\!\shortmid & & \\ \mathrm{B} & \subseteq & \mathrm{S5} & & \end{array}$$

But our first task will be to verify that these systems are normal. To see that they are, it is necessary only to verify that any substitution instance of a theorem is itself a theorem. Thus suppose that $F^1, \ldots, F^n$ is a proof in one of the systems – call it L. We want to see that $F^1_p(A), \ldots, F^n_p(A)$ is also a proof in L. But it is clear that it *is* a proof, since if F^i is an axiom of L, so is its substitution instance $F^i_p(A)$, and if F^i is immediately deducible from F^j and F^k by modus ponens or from F^j by necessitation, then the same goes for $F^i_p(A)$, $F^j_p(A)$, and $F^k_p(A)$, by the definitions of $(F \to G)_p(A)$ and $\Box(F)_p(A)$. Thus if F^n has a proof in L, so does its substitution instance $F^n_p(A)$.

Normal systems are also closed under truth-functional consequence, for if B follows truth-functionally from the theorems $A_1, \ldots, A_n$ of a normal system, then the tautology $A_1 \to (\cdots \to (A_n \to B) \cdots)$ is also a theorem of the system, and therefore so is B, which can be inferred from these theorems by n applications of modus ponens.

Until further notice, assume that L is a normal system.

Theorem 1. *Suppose* $L \vdash A \to B$. *Then* $L \vdash \Box A \to \Box B$.

Proof. Applying necessitation gives us that $L \vdash \Box(A \to B)$. Since $L \vdash \Box(A \to B) \to (\Box A \to \Box B)$, $L \vdash \Box A \to \Box B$, by modus ponens. ⊣

Theorem 2. *Suppose* $L \vdash A \leftrightarrow B$. *Then* $L \vdash \Box A \leftrightarrow \Box B$.

Proof. By truth-functional logic, $L \vdash A \to B$ and $L \vdash B \to A$. By Theorem 1, $L \vdash \Box A \to \Box B$ and $L \vdash \Box B \to \Box A$. The conclusion follows truth-functionally from these. ⊣

Theorem 3. $L \vdash \Box(A \wedge B) \leftrightarrow (\Box A \wedge \Box B)$.

Proof. We have $L \vdash (A \wedge B) \to A$ and $L \vdash (A \wedge B) \to B$, whence by Theorem 1,

(1) $L \vdash \Box(A \wedge B) \to \Box A$ and

(2) $L \vdash \Box(A \wedge B) \to \Box B$.

We also have $L \vdash A \to (B \to (A \wedge B))$, whence by Theorem 1,

(3) $L \vdash \Box A \to \Box(B \to (A \wedge B))$, and

(4) $L \vdash \Box(B \to (A \wedge B)) \to (\Box B \to \Box(A \wedge B))$ (distribution).

The theorem follows truth-functionally from (1), (2), (3), and (4). ⊣

Theorem 4. $L \vdash \Box(A_1 \wedge \cdots \wedge A_n) \leftrightarrow (\Box A_1 \wedge \cdots \wedge \Box A_n)$.

Proof. The theorem holds if $n = 0$, for the empty conjunction is identified with $\top$, and $L \vdash \Box \top$. The theorem is trivial if $n = 1$ and has just been proved if $n = 2$. If $n > 2$, then

$$\begin{aligned} L \vdash \Box(A_1 \wedge A_2 \wedge \cdots \wedge A_n) &\leftrightarrow \Box(A_1 \wedge (A_2 \wedge \cdots \wedge A_n)) \\ &\leftrightarrow \Box A_1 \wedge \Box(A_2 \wedge \cdots \wedge A_n) \\ &\leftrightarrow (\Box A_1 \wedge \Box A_2 \cdots \wedge \Box A_n) \end{aligned}$$

The first of these equivalences holds by Theorem 2, the second by Theorem 3, and the third by the induction hypothesis. ⊣

We write: $A \leftrightarrow B, \leftrightarrow C$, etc. to mean: $(A \leftrightarrow B) \wedge (B \leftrightarrow C)$, etc.

Theorem 5. *Suppose* $L \vdash A_1 \wedge \cdots \wedge A_n \to B$. *Then* $L \vdash \Box A_1 \wedge \cdots \wedge \Box A_n \to \Box B$.

Proof. By the supposition and Theorem 1, $L \vdash \Box(A_1 \wedge \cdots \wedge A_n) \to \Box B$. The conclusion then follows by Theorem 4. ⊣

Theorem 6. *Suppose* $L \vdash A \to B$. *Then* $L \vdash \Diamond A \to \Diamond B$.

Proof. Truth-functionally, we have
$L \vdash \neg B \to \neg A$, whence
$L \vdash \Box \neg B \to \Box \neg A$ by Theorem 1, and then truth-functionally
$L \vdash \neg \Box \neg A \to \neg \Box \neg B$, i.e., $L \vdash \Diamond A \to \Diamond B$. ⊣

Theorem 7. *Suppose* $L \vdash A \leftrightarrow B$. *Then* $L \vdash \Diamond A \leftrightarrow \Diamond B$.

Proof. The theorem follows from Theorem 6 via truth-functional logic and definitions. ⊣

Theorem 8. $L \vdash \Diamond A \wedge \Box B \to \Diamond(A \wedge B)$.

Proof. By the definition of $\Diamond$, it is enough to show that
$L \vdash \Box \neg (A \wedge B) \wedge \Box B \to \Box \neg A$. But this is clear, since
$L \vdash \Box \neg (A \wedge B) \to \Box(B \to \neg A)$. ⊣

Henceforth we shall refer to the facts stated in Theorems 1–8, together with obvious consequences of these, as *normality*.

The first substitution theorem. *Suppose* $L \vdash A \leftrightarrow B$. *Then* $L \vdash F_p(A) \leftrightarrow F_p(B)$.

Proof. Induction on the complexity of F. If $F = p$, the sentence asserted in the conclusion to be a theorem of L is just $A \leftrightarrow B$; if $F = q$, it is $q \leftrightarrow q$, and if $F = \bot$, it is $\bot \leftrightarrow \bot$, both theorems of L. If $F = (G \to H)$ and the conclusion of the theorem holds for G and H, then it holds for F by propositional logic and the definition of substitution. Finally, if $F = \Box(G)$ and $L \vdash G_p(A) \leftrightarrow G_p(B)$, then by Theorem 2,
$L \vdash \Box(G_p(A)) \leftrightarrow \Box(G_p(B))$, i.e.,
$L \vdash \Box(G)_p(A) \leftrightarrow \Box(G)_p(B)$, i.e.,
$L \vdash F_p(A) \leftrightarrow F_p(B)$. ⊣

Definition. For any modal sentence A, $\boxdot A$ is the sentence $(\Box A \wedge A)$.

The definition has a point since $\Box A \to A$ is not, in general, a theorem of K, K4, or GL. The notation $\boxdot$ is most useful when one is considering K4 or one of its extensions, e.g., GL.

Theorem 9. $\mathrm{K4} \vdash \boxdot \Box A \leftrightarrow \Box A, \leftrightarrow \Box \boxdot A$;
$\mathrm{K4} \vdash \boxdot A \leftrightarrow \boxdot \boxdot A$.

Proof. $\mathrm{K4} \vdash \Box A \to \Box \Box A$, and so by normality we have
$\mathrm{K4} \vdash (\Box \Box A \wedge \Box A) \leftrightarrow \Box A, \leftrightarrow \Box(\Box A \wedge A)$. That
$\mathrm{K4} \vdash \boxdot A \leftrightarrow \boxdot \boxdot A$ is proved similarly. $\dashv$

Theorem 10. *Suppose* L *extends* K4 *and* $\mathrm{L} \vdash \boxdot A \to B$. *Then* $\mathrm{L} \vdash \Box A \to \Box B$ *and* $\mathrm{L} \vdash \boxdot A \to \boxdot B$.

Proof. We have $\mathrm{L} \vdash \Box \boxdot A \to \Box B$, whence by Theorem 9, $\mathrm{L} \vdash \Box A \to \Box B$, and then by the definition of $\boxdot$, $\mathrm{L} \vdash \boxdot A \to \boxdot B$. $\dashv$

The second substitution theorem. $\mathrm{K4} \vdash \boxdot(A \leftrightarrow B) \to (F_p(A) \leftrightarrow F_p(B))$.

Proof. The proof is a formalization in K4 of the first substitution theorem and proceeds by induction on the complexity of F. If F is $p, q\ (\neq p)$, or $\bot$, then the sentence asserted to be a theorem of K4 is the tautology $\boxdot(A \leftrightarrow B) \to (A \leftrightarrow B)$, the tautology $\boxdot(A \leftrightarrow B) \to (q \leftrightarrow q)$, or the tautology $\boxdot(A \leftrightarrow B) \to (\bot \leftrightarrow \bot)$, respectively. If $F = (G \to H)$ and the theorem holds for G and H, then, truth-functionally it holds for F. Finally suppose that $F = \Box(G)$ and
$\mathrm{K4} \vdash \boxdot(A \leftrightarrow B) \to (G_p(A) \leftrightarrow G_p(B))$. Then
$\mathrm{K4} \vdash \Box \boxdot(A \leftrightarrow B) \to \Box(G_p(A) \leftrightarrow G_p(B))$, whence
$\mathrm{K4} \vdash \Box \boxdot(A \leftrightarrow B) \to (\Box(G_p(A)) \leftrightarrow \Box(G_p(B)))$, and then by the definition of substitution,
$\mathrm{K4} \vdash \Box \boxdot(A \leftrightarrow B) \to (\Box(G)_p(A) \leftrightarrow \Box(G)_p(B))$, i.e.,
$\mathrm{K4} \vdash \Box \boxdot(A \leftrightarrow B) \to (F_p(A) \leftrightarrow F_p(B))$. By Theorem 9,
$\mathrm{K4} \vdash \boxdot(A \leftrightarrow B) \to \Box \boxdot(A \leftrightarrow B)$, and we are done. $\dashv$

Corollary. $\mathrm{K4} \vdash \Box(A \leftrightarrow B) \to \Box(F_p(A) \leftrightarrow F_p(B))$;
$\mathrm{K4} \vdash \boxdot(A \leftrightarrow B) \to \boxdot(F_p(A) \leftrightarrow F_p(B))$.

Proof. By the theorem and Theorem 10. $\dashv$

The next theorem is a somewhat surprising result about K4.[4]

Theorem 11. $\mathrm{K4} \vdash \Box \Diamond \Box \Diamond A \leftrightarrow \Box \Diamond A$.

Proof. We begin by observing that $\mathrm{K} \vdash \Box(\Diamond B \wedge \Box C \to \Diamond D)$ whenever
$\mathrm{K} \vdash \Box(B \wedge C \to D)$, for then $\mathrm{K} \vdash \Box(C \wedge \neg D \to \neg B)$,
$\mathrm{K} \vdash \Box \Box(C \wedge \neg D \to \neg B)$,
$\mathrm{K} \vdash \Box(\Box C \wedge \Box \neg D \to \Box \neg B)$, whence
$\mathrm{K} \vdash \Box(\Diamond B \wedge \Box C \to \Diamond D)$. Similarly, $\mathrm{K} \vdash \Box(\Box B \wedge \Diamond C \to \Diamond D)$ whenever
$\mathrm{K} \vdash \Box(B \wedge C \to D)$.

Since, evidently,
$\mathrm{K} \vdash \Box(A \wedge \Diamond A \to \Diamond A)$, we have
$\mathrm{K} \vdash \Box(\Diamond A \wedge \Box \Diamond A \to \Diamond \Diamond A)$,
$\mathrm{K} \vdash \Box(\Box \Diamond A \wedge \Diamond \Box \Diamond A \to \Diamond \Diamond \Diamond A)$, and
$\mathrm{K} \vdash \Box(\Diamond \Box \Diamond A \wedge \Box \Diamond \Box \Diamond A \to \Diamond \Diamond \Diamond \Diamond A)$. But
$\mathrm{K4} \vdash \Diamond \Diamond \Diamond \Diamond A \to \Diamond A$, whence
$\mathrm{K4} \vdash \Box(\Diamond \Diamond \Diamond \Diamond A \to \Diamond A)$, and so
$\mathrm{K4} \vdash \Box(\Diamond \Box \Diamond A \wedge \Box \Diamond \Box \Diamond A \to \Diamond A)$ and
$\mathrm{K4} \vdash \Box \Diamond \Box \Diamond A \wedge \Box \Box \Diamond \Box \Diamond A \to \Box \Diamond A$. But
$\mathrm{K4} \vdash \Box \Diamond \Box \Diamond A \to \Box \Box \Diamond \Box \Diamond A$. Thus
$\mathrm{K4} \vdash \Box \Diamond \Box \Diamond A \to \Box \Diamond A$.

Conversely,
$\mathrm{K} \vdash \Diamond A \wedge \Box \Box \Diamond A \to \Diamond(A \wedge \Box \Diamond A)$, and so
$\mathrm{K} \vdash \Diamond A \wedge \Box \Box \Diamond A \to \Diamond \Box \Diamond A$, whence
$\mathrm{K} \vdash \Box(\Diamond A \wedge \Box \Box \Diamond A) \to \Box \Diamond \Box \Diamond A$. But
$\mathrm{K4} \vdash \Box \Diamond A \to \Box \Box \Box \Diamond A$,
$\mathrm{K4} \vdash \Box \Diamond A \to \Box \Diamond A \wedge \Box \Box \Box \Diamond A$, and so
$\mathrm{K4} \vdash \Box \Diamond A \to \Box(\Diamond A \wedge \Box \Box \Diamond A)$. Thus
$\mathrm{K4} \vdash \Box \Diamond A \to \Box \Diamond \Box \Diamond A$. $\dashv$

We emphasize that no use of $\Box p \to p$ has been made thus far; the two substitution theorems and their corollary are results about K4 and hence about all extensions of K4.

Theorem 12. $\mathrm{T} \vdash A \to \Diamond A$; $\mathrm{T} \vdash \Box A \to \Diamond A$.

Proof. $\mathrm{T} \vdash \Box \neg A \to \neg A$; contraposing, we obtain $\mathrm{T} \vdash A \to \Diamond A$. Since $\mathrm{T} \vdash \Box A \to A$, $\mathrm{T} \vdash \Box A \to \Diamond A$ also. $\dashv$

Theorem 13. $\mathrm{S4} \vdash \Diamond \Diamond A \to \Diamond A$.

Proof. By contraposition, from $\mathrm{S4} \vdash \Box \neg A \to \Box \Box \neg A$. $\dashv$

Theorem 14. $\mathrm{S4} \vdash \Box A \leftrightarrow \Box \Box A$; $\Diamond A \leftrightarrow \Diamond \Diamond A$.

Theorem 15. S4$\vdash$

$$\begin{array}{ccccc} \Box p & \rightarrow & \Box\Diamond\Box p & \rightarrow & \Diamond\Box p \\ \Big\downarrow & & \downarrow & & \downarrow \\ & & \Box\Diamond p & \rightarrow & \Diamond\Box\Diamond p \\ \Big\downarrow & & & & \downarrow \\ p & & \longrightarrow & & \Diamond p \end{array}$$

A *modality* is a sequence of $\Box$*s* and $\neg$*s*. It follows from Theorems 11, 14, and 15 that there are at most 14 inequivalent modalities σ in S4, i.e., at most 14 inequivalent sentences of the form σp, namely the 7 mentioned in Theorem 15 and their negations. The completeness theorem for S4 given in Chapter 5 will enable us to see that these 14 modalities are in fact inequivalent. The completeness theorems for B and GL also found there can be used to show that no two of the modalities [empty], $\Box, \Box\Box, \ldots$ are equivalent in either of those logics.

We now examine S5. We first show that S5 has an alternative axiomatization. Let S5* be the system of modal logic whose axioms are all the sentences that are either axioms of S4 or B and whose rules of inference are modus ponens and necessitation.

Theorem 16. S5*$\vdash A$ *iff* S5$\vdash A$.

Proof. It is enough to show that for every A, S5$\vdash \Box A \rightarrow \Box\Box A$, S5$\vdash A \rightarrow \Box\Diamond A$, and S5*$\vdash \Diamond A \rightarrow \Box\Diamond A$.
S5$\vdash \Box A \rightarrow \Box\Box A$: Since S5 extends *T*,
S5$\vdash \Box A \rightarrow \Diamond\Box A$; also
S5$\vdash \Diamond\Box A \rightarrow \Box\Diamond\Box A$ (because S5$\vdash \Diamond B \rightarrow \Box\Diamond B$), and therefore
S5$\vdash \Box A \rightarrow \Box\Diamond\Box A$. But also
S5$\vdash \Diamond\Box A \rightarrow \Box A$ (because S5$\vdash \Diamond\neg A \rightarrow \Box\Diamond\neg A$), whence by normality
S5$\vdash \Box\Diamond\Box A \rightarrow \Box\Box A$. Thus
S5$\vdash \Box A \rightarrow \Box\Box A$.
S5$\vdash A \rightarrow \Box\Diamond A$: This is immediate from
S5$\vdash A \rightarrow \Diamond A$ and S5$\vdash \Diamond A \rightarrow \Box\Diamond A$. Finally,
S5*$\vdash \Diamond A \rightarrow \Box\Diamond A$: For since
S5*$\vdash \Diamond\Diamond A \rightarrow \Diamond A$ (S5* extends S4), by normality,
S5*$\vdash \Box\Diamond\Diamond A \rightarrow \Box\Diamond A$. But also
S5*$\vdash \Diamond A \rightarrow \Box\Diamond\Diamond A$ (S5* extends B), and so we have what we want. $\dashv$

Theorem 17. S5$\vdash (\Diamond\Diamond A \leftrightarrow \Diamond A) \wedge (\Box\Diamond A \leftrightarrow \Diamond A) \wedge (\Box\Box A \leftrightarrow \Box A) \wedge (\Diamond\Box A \leftrightarrow \Box A)$.

According to Theorem 17, if σ is a string containing a positive number of $\Box$s and $\Diamond$s ending in $\Box$ or in $\Diamond$ but not $\neg$, then σp is equivalent to $\Box p$ or to $\Diamond p$, respectively. Thus there are at most six inequivalent modalities in S5: $\Box$, [empty], $\Diamond$, and their negations. The completeness theorem for S5 given in Chapter 5 will enable us to see that no two of these six modalities are in fact equivalent in S5.

We shall now show that $\Box p \rightarrow p$ is not a theorem of GL and that GL is consistent: Define A^* by $\bot^* = \bot$, $p^* = p$ (for all sentence letters p), $(A \rightarrow B)^* = (A^* \rightarrow B^*)$, and $\Box(A)^* = \top$. (Then A^* is the result of taking $\Box$ to be a *verum* operator in A.) If A is a tautology, so is A^*; if A is a distribution axiom, then A^* is $\top \rightarrow (\top \rightarrow \top)$; and if A is a sentence $\Box(\Box B \rightarrow B) \rightarrow \Box B$, then $A^* = \top \rightarrow \top$. Moreover if A^* and $(A \rightarrow B)^*$ are tautologies, so is B^*, and if A^* is a tautology, then so is $\Box(A)^* = \top$. Thus if A is a theorem of GL, A^* is a tautology. But $(\Box p \rightarrow p)^* = (\top \rightarrow p)$, which is not a tautology. Thus $\Box p \rightarrow p$ is not a theorem of GL, hence not one of K4 or K.

Similarly, $\Box(\Box p \rightarrow p) \rightarrow \Box p$ is not a theorem of S5, hence not one of B, S4, T, K4, or K. Define $\bot^*$, p^*, and $(A \rightarrow B)^*$ as before, but now let $\Box(A)^* = A^*$. (A^* is now the result of taking $\Box$ to be *decoration* in A.) Again if A is a theorem of S5, A^* is a tautology. But $(\Box(\Box p \rightarrow p) \rightarrow \Box p)^*$ is now $((p \rightarrow p) \rightarrow p)$, which is not a tautology. Therefore $(\Box(\Box p \rightarrow p) \rightarrow \Box p)$ is not a theorem of S5.

GL and T are thus consistent normal systems of modal logic, but there is no consistent normal system that extends both of them.

A remarkable fact about GL, the proof of which was independently discovered by de Jongh, Kripke, and Sambin, is that $\Box p \rightarrow \Box\Box p$ is a theorem of GL and thus that for all sentences A, $\Box A \rightarrow \Box\Box A$ is a theorem of GL. ("Had" $\Box p \rightarrow \Box\Box p$ not been a theorem of GL, we should have been interested in the smallest normal extension of GL in which it was one!) In practice, sentences $\Box A \rightarrow \Box\Box A$ are treated rather as if they were axioms of GL.

Theorem 18. GL$\vdash \Box A \rightarrow \Box\Box A$.

Proof. Truth-functionally, we have
GL$\vdash A \rightarrow ((\Box\Box A \wedge \Box A) \rightarrow (\Box A \wedge A))$, whence by normality,
GL$\vdash A \rightarrow (\Box(\Box A \wedge A) \rightarrow (\Box A \wedge A))$. By normality again,
GL$\vdash \Box A \rightarrow \Box(\Box(\Box A \wedge A) \rightarrow (\Box A \wedge A))$. But where $B = (\Box A \wedge A)$, $\Box(\Box B \rightarrow B) \rightarrow \Box B$ is an axiom of GL, i.e.,

$GL\vdash \Box(\Box(\Box A \wedge A)\rightarrow(\Box A \wedge A))\rightarrow \Box(\Box A \wedge A)$. Truth-functionally, $GL\vdash \Box A \rightarrow \Box(\Box A \wedge A)$. But by normality, $GL\vdash \Box(\Box A \wedge A)\rightarrow \Box\Box A$. From these last two, we have $GL\vdash \Box A \rightarrow \Box\Box A$. ⊣

It follows that GL extends K4; it is worth mentioning that the substitution theorems therefore hold when 'K4' is replaced by 'GL'.

Theorem 19. $GL\vdash \Box(\Box A \rightarrow A)\leftrightarrow \Box A, \leftrightarrow \Box(\Box A \wedge A)$.

Proof. Immediate by normality and Theorem 18. ⊣

Theorem 20. *If* $GL\vdash(\Box A_1 \wedge A_1 \wedge \cdots \wedge \Box A_n \wedge A_n \wedge \Box B)\rightarrow B$, *then* $GL\vdash(\Box A_1 \wedge \cdots \wedge \Box A_n)\rightarrow \Box B$.

Proof. Suppose that
$GL\vdash(\Box A_1 \wedge A_1 \wedge \cdots \wedge \Box A_n \wedge A_n \wedge \Box B)\rightarrow B$. Then
$GL\vdash \Box A_1 \wedge A_1 \wedge \cdots \wedge \Box A_n \wedge A_n \rightarrow(\Box B\rightarrow B)$. By normality,
$GL\vdash \Box(\Box A_1 \wedge A_1)\wedge \cdots \wedge \Box(\Box A_n \wedge A_n)\rightarrow \Box(\Box B\rightarrow B)$. By both equivalences of Theorem 19,
$GL\vdash(\Box A_1 \wedge \cdots \wedge \Box A_n)\rightarrow \Box B$. ⊣

Theorem 21. $GL\vdash \Box\bot \leftrightarrow \Box\Diamond p$.

Proof. $GL\vdash \bot \rightarrow \Diamond p$. Thus by normality,
$GL\vdash \Box\bot \rightarrow \Box\Diamond p$. Conversely,
$GL\vdash \Diamond p\rightarrow \Diamond\top$, and by the definition of $\Diamond$,
$GL\vdash \Diamond\top \rightarrow(\Box\bot \rightarrow \bot)$. Thus
$GL\vdash \Diamond p\rightarrow(\Box\bot \rightarrow \bot)$, and by normality,
$GL\vdash \Box\Diamond p\rightarrow \Box(\Box\bot \rightarrow \bot)$. Since
$GL\vdash \Box(\Box\bot \rightarrow \bot)\rightarrow \Box\bot$, we also have that
$GL\vdash \Box\Diamond p\rightarrow \Box\bot$. ⊣

Theorem 22. $GL\vdash \Box\Diamond\bot \rightarrow \Box\bot$.

Proof. Substitute $\bot$ for p in Theorem 21 and weaken. ⊣

In Chapter 3 we shall see how Theorem 21 can be regarded as telling us that (PA) asserts of each sentence S that PA is inconsistent if and only if it is provable (in PA) that S is consistent (with PA). Theorem 22, we shall also see there, will similarly tell us that the second incompleteness theorem is a theorem of PA.

Our proof that $\Box p\rightarrow p$ is not a theorem of GL cannot be used to show that $p\rightarrow \Box\Diamond p$ and $\Diamond p\rightarrow \Box\Diamond p$ are not theorems of GL. In Chapter 3 we shall see that $\Box\bot$ is not a theorem of GL. It

follows from Theorem 21 that $\top\rightarrow \Box\Diamond\top$ and $\Diamond\top \rightarrow \Box\Diamond\top$ are both equivalent to $\Box\Diamond\top$. Thus neither is provable in GL, and therefore $p\rightarrow \Box\Diamond p$ and $\Diamond p\rightarrow \Box\Diamond p$ are also unprovable in GL.

The proof of the next theorem formalizes the argument used in the proof of Löb's theorem. As we shall see in Chapter 3, the theorem may be used in a variant proof of a basic fact about GL: every theorem of GL is provable in PA under every translation.

Theorem 23. $K4\vdash \Box(q\leftrightarrow(\Box q\rightarrow p))\rightarrow(\Box(\Box p\rightarrow p)\rightarrow \Box p)$.

Proof.
(1) $K4\vdash \Box(q\leftrightarrow(\Box q\rightarrow p))\rightarrow(\Box q\rightarrow \Box(\Box q\rightarrow p))$, since K4 is normal.
(2) $K4\vdash \Box(\Box q\rightarrow p)\rightarrow(\Box\Box q\rightarrow \Box p)$ – a distribution axiom.
(3) $K4\vdash \Box q\rightarrow \Box\Box q$.
(4) $K4\vdash \Box(q\leftrightarrow(\Box q\rightarrow p))\rightarrow(\Box q\rightarrow \Box p)$ – (4) follows truth-functionally from (1), (2), and (3).
(5) $K4\vdash \Box\Box(q\leftrightarrow(\Box q\rightarrow p))\rightarrow \Box(\Box q\rightarrow \Box p)$ – (5) follows from (4) by normality.
(6) $K4\vdash \Box(q\leftrightarrow(\Box q\rightarrow p))\rightarrow \Box\Box(q\leftrightarrow(\Box q\rightarrow p))$ – (6) is of the form $\Box A\rightarrow \Box\Box A$.
(7) $K4\vdash \Box(\Box p\rightarrow p)\rightarrow(\Box(\Box q\rightarrow \Box p)\rightarrow \Box(\Box q\rightarrow p))$, by normality.
(8) $K4\vdash \Box(q\leftrightarrow(\Box q\rightarrow p))\rightarrow(\Box(\Box q\rightarrow p)\rightarrow \Box q)$, by normality. ⊣

Theorem 23 then follows truth-functionally from (6), (5), (7), (8), and (4).

Theorem 24
(a) $GL\vdash \Box(p\leftrightarrow\neg\Box p)\leftrightarrow \Box(p\leftrightarrow\neg\Box\bot)$,
(b) $GL\vdash \Box(p\leftrightarrow \Box p)\leftrightarrow \Box(p\leftrightarrow \top)$,
(c) $GL\vdash \Box(p\leftrightarrow \Box\neg p)\leftrightarrow \Box(p\leftrightarrow \Box\bot)$, *and*
(d) $GL\vdash \Box(p\leftrightarrow\neg\Box\neg p)\leftrightarrow \Box(p\leftrightarrow \bot)$.

Proof. (a) $K4\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow \Box(p\rightarrow\neg\Box p)$. Since
$K4\vdash \Box(p\rightarrow\neg\Box p)\rightarrow \Box\Box(p\rightarrow\neg\Box p)$,
$K4\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow \Box(\Box p\rightarrow \Box\neg\Box p)$ by normality. But
$K4\vdash \Box p\rightarrow \Box\Box p$ and
$K4\vdash \Box\Box p \wedge \Box\neg\Box p\rightarrow \Box\bot$. Thus
$K4\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow \Box(\Box p\rightarrow \Box\bot)$. Since
$K4\vdash \Box\bot \rightarrow \Box p$,

K4$\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow\Box(\Box p\leftrightarrow\Box\bot)$, and so
K4$\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow\Box(\neg\Box p\leftrightarrow\neg\Box\bot)$. But
K4$\vdash \Box(p\leftrightarrow\neg\Box p)\wedge\Box(\neg\Box p\leftrightarrow\neg\Box\bot)\rightarrow\Box(p\leftrightarrow\neg\Box\bot)$. Thus
K4$\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow\Box(p\leftrightarrow\neg\Box\bot)$, whence
GL$\vdash \Box(p\leftrightarrow\neg\Box p)\rightarrow\Box(p\leftrightarrow\neg\Box\bot)$.
Conversely, by Theorem 21 (with $\bot$ for p),
GL$\vdash \Box(p\leftrightarrow\neg\Box\bot)\rightarrow\Box\Box(p\leftrightarrow\neg\Box p)$, and so
GL$\vdash \Box(p\leftrightarrow\neg\Box\bot)\rightarrow\Box(\Box p\leftrightarrow\Box\neg\Box p)$. By Theorem 21 (with $\neg p$ for p)
GL$\vdash \Box\neg\Box p\leftrightarrow\Box\bot$. Thus
GL$\vdash \Box(p\leftrightarrow\neg\Box\bot)\rightarrow\Box(\Box p\leftrightarrow\Box\bot)$ and
GL$\vdash \Box(p\leftrightarrow\neg\Box\bot)\rightarrow\Box(\neg\Box\bot\leftrightarrow\neg\Box p)$. Since
GL$\vdash \Box(p\leftrightarrow\neg\Box\bot)\wedge\Box(\neg\Box\bot\leftrightarrow\neg\Box p)\rightarrow\Box(p\leftrightarrow\neg\Box p)$,
GL$\vdash \Box(p\leftrightarrow\neg\Box\bot)\rightarrow\Box(p\leftrightarrow\neg\Box p)$.
(b) Since GL$\vdash \top\leftrightarrow\Box\top$,
GL$\vdash \Box(p\leftrightarrow\Box p)\rightarrow\Box(\Box p\rightarrow p), \rightarrow\Box p, \rightarrow\Box(p\leftrightarrow\top), \rightarrow\Box p,$
$\rightarrow(\Box p\wedge\Box\Box p), \rightarrow\Box(p\wedge\Box p), \rightarrow\Box(p\leftrightarrow\Box p)$.
Substituting $\neg p$ for p in (a) yields
GL$\vdash \Box(\neg p\leftrightarrow\neg\Box\neg p)\leftrightarrow\Box(\neg p\leftrightarrow\neg\Box\bot)$. Simplifying, we obtain
GL$\vdash \Box(p\leftrightarrow\Box\neg p)\leftrightarrow\Box(p\leftrightarrow\Box\bot)$, i.e., (c).
We can obtain (d) by similarly substituting $\neg p$ for p in (b). ⊣

As we shall see, Theorem 24 will tell us that it is a theorem of PA that a sentence S is equivalent (in PA) to the assertion that S is unprovable/provable/disprovable/consistent if and only if S is respectively equivalent to the assertion that PA is consistent/that $0=0$/that PA is inconsistent/that $0=1$. Many other interesting facts about PA can be learned from a study of GL.

2

Peano arithmetic

Peano arithmetic (*PA*, or *arithmetic*, for short) is classical first-order arithmetic with induction. The aim of this chapter is to define the concepts mentioned in, and describe the proofs of, five important theorems about Bew(x), the standard "provability" or "theoremhood" predicate of PA:

(i) If $\vdash S$, then $\vdash$Bew($\ulcorner S\urcorner$),
(ii) $\vdash$Bew($\ulcorner(S\rightarrow T)\urcorner$)$\rightarrow$(Bew($\ulcorner S\urcorner$)$\rightarrow$Bew($\ulcorner T\urcorner$)),
(iii) $\vdash$Bew($\ulcorner S\urcorner$)$\rightarrow$Bew($\ulcorner$Bew($\ulcorner S\urcorner$)$\urcorner$),
(iv) Bew($\ulcorner S\urcorner$) is a Σ sentence, and
(v) if S is a Σ sentence, then $\vdash S\rightarrow$Bew($\ulcorner S\urcorner$)

(for all sentences S, T of Peano arithmetic).

'$\vdash$' is, as usual, the sign for theoremhood; in this chapter we write '$\vdash S$' to mean that S is a theorem of PA. $\ulcorner S\urcorner$ is the numeral in PA for the Gödel number of sentence S, that is, if n is the Gödel number of S, then $\ulcorner S\urcorner$ is **0** preceded by n occurrences of the successor sign **s**. Bew($\ulcorner S\urcorner$) is therefore the result of substituting $\ulcorner S\urcorner$ for the variable x in Bew(x), and (iii) immediately follows from (iv) and (v). Bew($\ulcorner S\urcorner$) may be regarded as a sentence asserting that S is a theorem of PA. Σ sentences (often called Σ_1 sentences) are, roughly speaking, sentences constructed from atomic formulas and negations of atomic formulas by means of conjunction, disjunction, existential quantification, and bounded universal quantification ("for all x less than y"), but not negation or universal quantification. A precise definition is given below.

Notice the distinction between 'Bew(x)' and '$\vdash$'. 'Bew(x)' denotes a certain formula of the language of PA and thus Bew(x) *is* that formula; it is a formula that is true of (the Gödel numbers of) those formulas of PA that are provable in PA. '$\vdash$', on the other hand, is a (pre-posed) predicate of *our* language (logicians' English, a mixture of English, mathematical terminology, and symbolism) and has the