ISRAEL JOURNAL OF MATHEMATICS, Vol. 25, 1976

PROVABILITY INTERPRETATIONS OF MODAL LOGIC

BY
ROBERT M. SOLOVAY

ABSTRACT
We consider interpretations of modal logic in Peano arithmetic ($\boldsymbol{P}$) determined by an assignment of a sentence v^* of $\boldsymbol{P}$ to each propositional variable v. We put $(\perp)^* =$ "$\mathbf{0} = \mathbf{1}$", $(\chi \to \psi)^* =$ "$\chi^* \to \psi^*$" and let $(\Box\chi)^*$ be a formalization of "χ^* is a theorem of $\boldsymbol{P}$". We say that a modal formula, χ, is *valid* if χ^* is a theorem of $\boldsymbol{P}$ in each such interpretation. We provide an axiomitization of the class of valid formulae and prove that this class is recursive.

§1. Introduction

1.1. The language, $\boldsymbol{M}$, of propositional modal logic, has an infinite stock of propositional variables, $v_0, v_1, \cdots$, a propositional constant, $\perp$, (denoting falsehood), the binary propositional connective, $\to$, (material implication), and the modal operator $\Box$. (The standard interpretation of "$\Box\chi$" is "χ is necessarily true". In this paper "$\Box\chi$" is some variant of "χ is provable".)

We define the other Boolean and modal connectives in terms of these in some usual way. In particular, $\Diamond\chi$ (χ is posssible) is $\neg\Box\neg\ \chi$.

A well-formed formula of $\boldsymbol{M}$ will be referred to as a modal formula or simply as a formula if no confusion is likely.

1.2. $\boldsymbol{P}$ is the usual formalization of Peano arithmetic. (Cf. [3, §19].) We shall assume that the language of $\boldsymbol{P}$ is enriched with a description operator so that we may introduce defined terms freely. (Cf. [3, §74].) We let $\boldsymbol{n}$, the numeral for n, be $S^n\mathbf{0}$. If χ is a formula of $\boldsymbol{P}$, 'χ' denotes the numeral of the Gödel number of χ. Bew (x) is the formula that expresses "x is the Gödel number of a theorem of $\boldsymbol{P}$".

1.3. An interpretation of $\boldsymbol{M}$ in $\boldsymbol{P}$ is a function that assigns to each formula χ of $\boldsymbol{M}$ a sentence, χ^* of $\boldsymbol{P}$, and which satisfies the following requirements:

1) $(\perp)^* =$ "$\mathbf{0} = \mathbf{1}$";

2) $(\chi \to \psi)^* =$ "$\chi^* \to \psi^*$";

3) $(\Box \chi)^* =$ "Bew('χ^*')".

Evidently each map of the set of variables of M into sentences of P has a unique prolongation to an interpretation of M in P.

1.4. A modal formula χ is P-valid if, in every interpretation, χ^* is a theorem of P. Our goal is to characterize the set of P-valid formulae. We will show that they are precisely the theorems of a certain system for modal logic, G. (The "G" is for "Gödel". The key axiom for G is an expression of Gödel's second incompleteness theorem.)

Here is how this paper is organized. In §2, we describe the system G. It will be evident that each theorem of G is P-valid. In §3, we work out a Kripke-style semantical analysis for G and prove that G has the finite model property. (It will then follow that the set of theorems of G is recursive.) In §4, we prove a technical lemma which, roughly speaking, allows us to imbed a finite Kripke model of G within Peano arithmetic. Our characterization of P-valid formulae follows readily.

Let ω be the set of non-negative integers. We say that a formula χ is ω-valid if, in each interpretation, χ^* is true in the standard model, $\langle \omega; +, \cdot \rangle$. In §5, we give a characterization of the set of ω-valid formulae, and show that this set is recursive.

Finally, in §6, we state without proof various further results on G and on other provability interpretations.

1.5. My interest in this problem was stimulated by a recent announcement of Boolos [1]. He formulates the system G (he calls it L), and proves that a formula *without free variables* is P-valid iff it is a theorem of G. (This settles problem 35 of [2].) Aside from Boolos' work, my main source for inspiration was the work of Kripke on the semantics of modal logic [4].

The work presented in §2 and §3 was known to researchers in this area, e.g., Kripke, Boolos, deJongh; they were also aware of the natural conjecture that the theorems of G are precisely the P-valid formulae. Thus the real contribution this paper makes is in §4. Nevertheless, for expository purposes, I have decided to include the material of §§2–3.

§2. The system G

2.1. In the following presentation, χ and ψ are metavariables ranging over modal formulae. G is the smallest collection of formulae containing the following axiom schemata and closed under the following rules of inference:

A0. All tautologies are axioms.
A1. $\Box(\chi \to \psi) . \to . (\Box\chi \to \Box\psi)$.
A2. $\Box\chi . \to . \Box\Box\chi$.
A3. $\Box(\Box\chi \to \chi) . \to . \Box\chi$.
R1. If $\vdash \chi \to \psi$ and $\vdash \chi$, then $\vdash \psi$.
R2. If $\vdash \chi$, then $\vdash \Box\chi$.

2.2. Remarks.

1) deJongh has shown that A2 can be derived from the remaining axioms and rules.

2) Note that, in contrast to R2, $\chi \to \Box\chi$ is *not* a theorem schemata of G. (This may easily be seen using Kripke models. Cf. §3.)

3) Any substitution instance of a theorem of G is again a theorem of G.

4) We use $\vdash$, decorated perhaps with subscripts, to indicate provability in a formal system.

2.3. Lemma. *Every theorem of G is $\boldsymbol{P}$-valid.*

Proof. The proof will be by induction on the number of axioms and rules of inference used in a G-proof of ψ. The discussion will split into cases according to the last rule or axiom schema cited in the proof of ψ.

The cases of A0 and R1 are evident. It is easy to show in $\boldsymbol{P}$ that the theorems of $\boldsymbol{P}$ are closed under modus ponens, which handles A1.

Bew (x) is a Σ_1^0 formula, i.e., it is provably equivalent in $\boldsymbol{P}$ to a formula of the form $(\exists y)R(x, y)$ with R (the standard formalization of) a primitive recursive predicate. It is a standard fact about Σ_1^0 sentences (owing ultimately to the "numeralwise representability" of primitive recursive formulae such as R), that if a Σ_1^0 sentence is true, it is provable. (Cf. [3, §49, theorem 27].) This shows that our lemma is true in the case the last step in the G-proof is an instance of R2. Moreover, the proof of the "standard fact" can be formalized in $\boldsymbol{P}$, handling the case of A2. (Cf. [3, p. 244, remark 1].)

It remains to handle the most interesting axiom, A3. We argue in $\boldsymbol{P}$. We are given a sentence χ such that $\vdash_P \mathrm{Bew}(`\chi') \to \chi$. We must show that $\vdash_P \chi$. (This is due to Löb [5]. Löb's theorem is essentially just the second Gödel incompleteness theorem as we shall see in a moment.) We have:

i) $\vdash_P \mathrm{Bew}(`\chi') \to \chi$.
ii) $\vdash_P \neg\chi \to \neg\mathrm{Bew}(`\chi')$.
iii) $\vdash_P \neg\chi \to \mathrm{Con}(\boldsymbol{P} + `\neg\chi')$.

Whence, $P+$ '$\neg\chi$' proves its own consistency. By the second Gödel incompleteness theorem [3, theorem 30] the system $P+$ '$\neg\chi$' is inconsistent, i.e.,

iv) $\vdash_P \chi$.

This completes the discussion of case A3. The lemma is now proved.

2.4. Lemma. 1) *If* $G\vdash\chi\to\psi$, *then* $G\vdash\Box\chi.\to.\Box\psi$.

2) $G\vdash\Box(\chi\wedge\psi).\leftrightarrow.\Box\chi\wedge\Box\psi$.

Proof. 1) Say $G\vdash\chi\to\psi$. By R2, $G\vdash\Box(\chi\to\psi)$. By A1, $G\vdash\Box(\chi\to\psi).\to.\Box\chi\to\Box\psi$. By R1, $G\vdash\Box\chi\to\Box\psi$.

2) We shall use the phrase "by propositional logic" to indicate uses of A0 and R1. We have:

i) $\vdash\Box(\chi\wedge\psi).\to.\ \Box\chi,\ \Box(\chi\wedge\psi)\to.\ \Box\psi$ (by (1) of this lemma);

ii) $\vdash\Box(\chi\wedge\psi).\to.\ \Box\chi\wedge\Box\psi$ (by (i), propositional logic);

iii) $\vdash\Box\chi\to\Box(\psi\to(\chi\wedge\psi))$ (by (1));

iv) $\vdash\Box(\psi\to(\chi\wedge\psi)).\to.\ \Box\psi\to\Box(\chi\wedge\psi)$ (by A1);

From (ii), (iii) and (iv), by propositional logic, (2) of the lemma follows.

§3. The semantics of G

3.1. Let X be a set and $>$ a binary relation on X. (We do not assume, for the moment, that $>$ is transitive.) We write "$<$" for the converse relation: $x_1<x_2$ iff $x_2>x_1$. Similarly, for x_1, x_2 in X, $x_1\leqslant x_2$ iff $x_2\geqslant x_1$ iff $[x_1<x_2$ or $x_1=x_2]$.

3.2. We begin by recalling the Kripke semantics for modal logic. Our terminology is slightly different from that of Kripke [4].

A frame is a pair $\langle X;>\rangle$ where X is a non-empty set and $>$ is a binary relation on X. The intuition is that X is the set of possible worlds. $x_1>x_2$ if the world x_2 is accessible to x_1.

(The following example is a good one to contemplate briefly. X is the set of consistent recursively axiomitized extensions of Peano arithmetic. $T_1>T_2$ iff $T_1\vdash\mathrm{Con}(T_2)$. To make this precise, it is necessary to view a theory as provided with a fixed recursive enumeration of its axioms. Otherwise, Con (T) would not be well-defined.

Note that in this example, we never have $T>T$ (by the second Gödel incompleteness theorem).)

3.3. Let V be the set of propositional variables. As usual, $2=\{0,1\}$ is taken as the set of truth values (with 0 being falsehood and 1 being truth). In the usual way 2 is a Boolean algebra.

A model (over the frame $\langle X;>\rangle$) is a map $e: V\times X\to 2$.

We prolong e to a map $e : M \times X \to 2$. We define $e(\chi, x)$ by induction on the number of logical connectives in χ:

1) $e(\perp, x) = 0$.

2) $e(\chi \to \psi, x) = 1$ if $e(\chi, x) = 0$ or $e(\psi, x) = 1$;
$e(\chi \to \psi, x) = 0$ if $e(\chi, x) = 1$ and $e(\psi, x) = 0$.

3) $e(\Box \chi, x) = 1$ iff $(\forall y)\ (y < x \to e(\chi, y) = 1)$;
otherwise $e(\Box \chi, x) = 0$.

We sometimes say "χ holds at x" (with respect to the model $\langle X; >, e\rangle$) to mean $e(\chi, x) = 1$.

The intuition behind (1) and (2) of the definition is evident. (3) says that $\Box \chi$ holds at x iff χ holds in all worlds possible relative to x.

We say that χ is *valid* in the frame $\langle X; >\rangle$ if for every $e : V \times X \to 2$ and every $x \in X$, we have $e(\chi, x) = 1$.

3.4. Our next goal is to characterize those frames in which the theorems of G are valid. We shall prove that they are precisely those frames $\langle X; >\rangle$ such that $<$ is transitive and well-founded.

Let S be a non-empty collection of frames. Let T be the set of formulae valid in each frame of S. It is easy to check that T is closed under the rules of inference R1 and R2 and contains all instances of the axiom schemata A0 and A1.

Lemma. *The frame $\langle X; >\rangle$ satisfies* A2 *iff $<$ is transitive.*

Proof. ($\Leftarrow$): We assume $<$ is transitive and show that A2 is valid. Let $e : V \times X \to 2$ give a model on the frame $\langle X; >\rangle$. Let $x \in X$, and χ a formula such that $\Box \chi$ holds at x. We must show that $\Box^2 \chi$ holds at x.

Suppose not. Then for some $y < x$, $\Box \chi$ fails at y. But then χ fails at some $z < y$. Since $<$ is transitive, $z < x$. But then $\Box \chi$ fails at x, since χ fails at z. This contradicts our assumption that $\Box \chi$ holds at x. The upshot is that $\Box^2 \chi$ must hold at x, as desired.

($\Rightarrow$): Suppose A2 holds in $\langle X; >\rangle$, and that $z < y$, and $y < x$. We show that $z < x$.

Let then v be some variable. Define $e : V \times X \to 2$ so that $e(v, w) = 1$ iff $w < x$. Then $e(\Box v, x) = 1$. Since A2 holds in $\langle X; >\rangle$, we have $e(\Box^2 v, x) = 1$. Whence we have $e(\Box v, y) = 1$ and then $e(v, z) = 1$. So $z < x$.

3.5. The binary relation $<$ on X is *well-founded* iff every non-empty subset Y of X contains a $<$-minimal element, y (i.e., for no z *in* Y is $z < y$).

Theorem. *Let $\langle X; >\rangle$ be a frame. Then the following are equivalent:*

1) *All the theorems of G are valid in $\langle X; > \rangle$.*

2) *$<$ is transitive and well-founded.*

PROOF. (1) $\Rightarrow$ (2): Suppose that all the theorems of G are valid in the frame $\langle X; > \rangle$. We know by Lemma 3.4 that $<$ is transitive. We must show that it is well-founded. Suppose to the contrary that Y is a non-empty subset of X with no $<$-minimal element. We shall derive a contradiction.

Let v be a propositional variable. Define $e : V \times X \to 2$ so that $e(v, x) = 0$ iff $x \in Y$. It will suffice to show that $\Box(\Box v \to v) . \to . \Box v$ is false at any point of Y. (Since A3 is valid in the frame $\langle X; > \rangle$, this will yield a contradiction.)

Since Y has no $<$-minimal element, $\Box v$ is false at each point of Y. It follows first that $\Box v \to v$ holds at each point in X. (The hypothesis is false at points in Y, and the conclusion is true at points not in Y.) But then $\Box(\Box v \to v)$ is true at all points in X, so $\Box(\Box v \to v) \to \Box v$ is false at each y in Y.

(2) $\Rightarrow$ (1): By Lemma 3.4 and the remarks that precede it, the set of formulae valid in $\langle X; > \rangle$ contains the instances of A0, A1 and A2 and is closed under R1 and R2. It suffices to verify that the instances of A3 are valid in $\langle X; > \rangle$.

Suppose that $e : V \times X \to 2$ gives a model, that $x \in X$, χ is a formula, and that

$$\Box(\Box \chi \to \chi) . \to . \Box \chi$$

is false at x. We derive a contradiction.

We must have $\Box(\Box \chi \to \chi)$ true at x and $\Box \chi$ false at x. Let

$$Y = \{y \in X : y < x \quad \text{and} \quad \chi \quad \text{is false at } y\}.$$

Y is non-empty since $\Box \chi$ is false at x. Let y be $<$-minimal in Y.

If $z < y$, then by transitivity of $<$, $z < x$. Since y is $<$-minimal in Y. $z \notin Y$. So χ must hold at z. Since $z < y$ was arbitrary, $\Box \chi$ holds at y.

Since $\Box(\Box \chi \to \chi)$ holds at x, $\Box \chi \to \chi$ holds at y. Whence, since $\Box \chi$ holds at y, we have χ holds at y. But this contradicts $y \in Y$.

3.6. We define for each modal formula, χ, a number $l(\chi)$ as follows:

1) If v is a variable, $l(v) = 1$;
2) $l(\perp) = 1$;
3) $l(\chi \to \psi) = l(\chi) + l(\psi) + 1$;
4) $l(\Box \chi) = l(\chi) + 1$.

Note that any reasonable encoding of χ as a binary string will have length at least $l(\chi)$.

We define the notion of subformula in an evident way; we arrange the definition so that χ is a subformula of itself. One checks easily by induction on $l(\chi)$ that χ has at most $l(\chi)$ subformulae.

The following theorem is the main result of §3.

THEOREM. *Let χ be a modal formula which is not a theorem of G. Then there is a model $\langle X; >, e\rangle$ and an $x_0 \in X$ such that:*

1) $e(\chi, x_0) = 0$.

2) $<$ *is transitive and well-founded. (So by Theorem 3.5, G is valid in $\langle X; >\rangle$.)*

3) *If $x \in X$, $x \preccurlyeq x_0$.*

4) *X is finite. In fact X has at most $2^{l(x)}$ elements.*

3.7. Before commencing the proof of Theorem 3.6, we reap some corollaries. Our first corollary is a completeness theorem for G.

COROLLARY 1. *$G \vdash \chi$ iff χ is valid in every (finite) frame $\langle X; >\rangle$ in which $<$ is transitive and well-founded.*

COROLLARY 2. *The set of theorems of G is recursive.*

PROOF. We prove, in fact, that if the formulae of M are encoded as binary strings one can check for G-theoremhood in at most $2^{2^{cn}}$ Turing machine steps, where n is the length of the encoding of χ. (Here c is some constant independent of χ.)

By (4) of Theorem 3.6 it is enough to look at frames of size at most 2^n. It is easy to see that there are at most $2^{2^{3n}}$ such frames.

All that matters about the function $e: V \times X \to 2$ is what it does on pairs $\langle v, x\rangle$ with v occurring in χ. There are at most n such v's, so each frame underlies at most $2^{2^{2n}}$ essentially different models. To check if χ holds at each point of a model of size 2^n takes $2^{c'n}$ steps for some c'. Putting these estimates together, we can test for validity of χ in all models of size at most 2^n in at most $2^{2^{cn}}$ steps.

REMARK. By exploiting the ideas behind the proof of Theorem 3.6, we can construct an algorithm for G-theoremhood that takes at most 2^{cn} steps (for some suitable constant c).

3.8. We now commence the proof of Theorem 3.6. We fix a formula χ that is not a theorem of G. Let E be the set of subformulae of χ. Let A be the subset of E consisting of all propositional variables or formulae with principal connective $\Box$ that occur in χ. (We think of A as the "atomic formulae" of E.)

Each map $s: A \to 2$ has a canonical prolongation (which we again denote by

s) to a map $s : E \to 2$. This is because each formula in E is a Boolean combination of formulae in A.

To each truth assignment $s : A \to 2$ we associate a modal formula $\Phi(s)$ as follows. Say $A = \{\psi_1, \cdots, \psi_r\}$ then $\Phi(s) = \psi_1' \wedge \cdots \wedge \psi_r'$ where $\psi_i' = \psi_i$ if $s(\psi_i) = 1$ and $\psi_i' = \neg \psi_i$ if $s(\psi_i) = 0$. (Intuitively, $\Phi(s)$ expresses that s is correct.) Note that if $s(\chi) = 1$, $\Phi(s) \to \chi$ is a tautology; if $s(x) = 0$, $\Phi(s) \to \neg \chi$ is a tautology.

Let X_1 be $\{s : s : A \to 2$ and $\neg \Phi(s)$ is not a theorem of $G\}$. Note that for some $s \in X_1$, $s(\chi) = 0$. (Otherwise, for each $s \in X_1$, $\Phi(s) \to \chi$ is a tautology. But then χ would be a theorem of G.) Fix $s_0 \in X_1$ such that $s_0(\chi) = 0$.

3.9. If $s \in X_1$, define the rank of s to be the number of formulae in A of the form $\Box \psi$ such that $s(\Box \psi) = 0$.

We now define a binary relation $<_1$ on X_1. Say that $s <_1 s'$ iff:

1) rank $(s) < \operatorname{rank}(s')$;

2) If $s'(\Box \psi) = 1$, then $s(\Box \psi) = 1$ and $s(\psi) = 1$.

It is readily checked that $<_1$ is well-founded and transitive.

Let X be $\{s \in X_1 : s \leqslant_1 s_0\}$. Define a relation $<$ on X by putting $s < s'$ iff $s <_1 s'$ (for $s, s' \in X$). Evidently, $<$ is transitive and well-founded. Since A has at most $l(\chi)$ elements, X has at most $2^{l(x)}$ elements. Clearly (3) of Theorem 3.6 holds (with s_0 in the role of x_0). It remains to prove that $e(\chi, s_0) = 0$.

3.10. Lemma. *Let $s \in X$. Let $\Box \psi \in A$ such that $s(\Box \psi) = 0$. Then there is an $s' \in X$ with $s' < s$ and $s'(\psi) = 0$.*

Proof. Let $\psi_1, \cdots, \psi_k$ be those formulae in E such that $\Box \psi_i \in A$ and $s(\Box \psi_i) = 1$. We shall prove that there is a truth assignment $s' \in X_1$ such that

a) $s'(\psi_i) = s'(\Box \psi_i) = 1$;

b) $s'(\psi) = 0$, $s'(\Box \psi) = 1$.

Note that since $s(\Box \psi) = 0$, (a) and (b) imply that rank $(s') < \operatorname{rank}(s)$. Whence, by (a), $s' <_1 s$. Hence, $s' <_1 s_0$. So $s' \in X$, and the lemma will follow.

Our plan now is to assume there is no $s' \in X_1$ satisfying (a) and (b), and show that $\Phi(s)$ is refutable in G, contradicting $s \in X_1$.

Let $\theta = \psi_1 \wedge \cdots \wedge \psi_k$. By Lemma 2.4 (2), G proves $\Box \theta \leftrightarrow (\Box \psi_1 \wedge \cdots \wedge \Box \psi_k)$. Thus if for every s' satisfying (a) and (b), $\Phi(s)$ is refutable in G,

i) $G \vdash \Box \theta \wedge \theta . \to . (\Box \psi \to \psi)$.

By Lemma 2.4 (1),

ii) $G \vdash \Box(\Box \theta \wedge \theta) . \to . \Box(\Box \psi \to \psi)$.

By A3,

iii) $G \vdash \Box(\Box \psi \to \psi) \to \Box \psi$.

By A2 and Lemma 2.4 (2),

iv) $G \vdash \Box\theta \rightarrow \Box(\Box\theta \wedge \theta)$.

By (ii), (iii), and (iv),

v) $G \vdash \Box\theta \rightarrow \Box\psi$.

But (v) entails that $\Phi(s)$ is refutable in G, contradicting $s \in X_1$.

3.11. We now define $e : V \times X \rightarrow 2$. If $v \in E$, put $e(v, x) = x(v)$. Otherwise, put $e(v, x) = 1$.

LEMMA. *Let $\psi \in E$, and $x \in X$. Then $e(\psi, x) = x(\psi)$.*

PROOF. We proceed by induction on $l(\chi)$. (It is crucial for the proof that all subformulae of ψ lie in E if $\psi \in E$.) The case when ψ is a variable is true by the definition of e, and the cases when $\psi = \perp$ or ψ has principal connective $\rightarrow$ are evident. Suppose now that ψ is $\Box\theta$. We consider two subcases.

Case 1. $x(\psi) = 1$.

We have to show that if $y < x$ then $e(\theta, y) = 1$. But since $y < x$, $y(\theta) = 1$. And by inductive hypothesis, $e(\theta, y) = y(\theta)$.

Case 2. $x(\psi) = 0$.

We must prove the existence of a $y < x$ with $e(\theta, y) = 0$. We apply Lemma 3.10 getting a $y < x$ such that $y(\theta) = 0$. But by inductive hypothesis, $e(\theta, y) = y(\theta)$.

3.12. The fact that $e(\chi, s_0) = 0$ is now immediate from Lemma 3.11 since $s_0(\chi) = 0$. Theorem 3.6 is proved.

From now on, if we use the symbol $<$ for a binary relation, *it is understood that $<$ is transitive and well-founded.*

Note that "$x < x$" is always false. (Otherwise, $\{x\}$ would have no $<$-least member.)

Finally, (3) of Theorem 3.6 says that x_0 is the topmost member of X. It is worth remarking that there is at most one topmost member of X. For if x_1 is another such, we have $x_0 < x_1$ and $x_1 < x_0$. Whence $x_0 < x_0$ contradicting our previous remark.

§4. Imbedding Kripke models in *P*

4.1. Our goal in this section is to prove that every ***P***-valid formula is a theorem of G. The main technical lemma, whose statement follows in a moment, may be viewed, roughly speaking, as imbedding a finite frame in which G is valid "into Peano arithmetic".

We shall be considering the following situation. $<$ is a transitive well-founded relation on $\{1, \cdots, n\}$. If $1 < j \leqq n$, then $j < 1$.

Recall that we are assuming $\boldsymbol{P}$ is formalized so that we can create terms with a description operator.

LEMMA. *There is a term l of $\boldsymbol{P}$ such that:*

1) $\boldsymbol{P} \vdash \mathbf{0} \leqq l \leqq \boldsymbol{n}$.

2) *In the standard model of $\boldsymbol{P}$, $l = 0$.*

3) *If $0 \leqq i \leqq n$, "$\boldsymbol{P} + {}$'$l = \boldsymbol{i}$'" is consistent.*

For $1 \leqq i \leqq n$, let $S_i = \{j : j < i\}$. Let $S_0 = \{1, \cdots, n\}$.

4) *Let $0 \leqq i \leqq n$. Let $j \in S_i$. Then*
$\boldsymbol{P} \vdash$ *If $l = \boldsymbol{i}$, then "$\boldsymbol{P} + {}$'$l = \boldsymbol{j}$'" is consistent.*

5) *Let $0 < i \leqq n$. Let $j \notin S_i$. Then*
$\boldsymbol{P} \vdash$ *If $l = \boldsymbol{i}$, then $\boldsymbol{P} \vdash$ "$l \neq \boldsymbol{j}$".*

4.2. Our proof of Lemma 4.1 will be based on the following sort of construction. We will define a primitive recursive function $h : \omega \to \{0, \cdots, n\}$. We will have $h(0) = 0$. Moreover, if $h(m) = i$, then either $h(m+1) = i$ or $h(m+1) \in S_i$. Since $<$ is well-founded, h is eventually constant. We let l denote the eventual value of h. (If $\lim_{m \to \infty} h(m)$ does not exist, we set $l = n + 1$.)

Our definition of h will be in terms of a Gödel number e for h. The apparent circularity is handled, using the recursion theorem, in the usual way. (Cf. [3, theorem XXVII, §66, p. 352].)

4.3. We now give the formal definition of h. Let e be a Gödel number for h. Let l be a term which defines the following number: 1) If e is the Gödel number of a recursive function, h say, and $\lim_{m \to \infty} h(m)$ exists, then l is this limit. Otherwise, $l = n + 1$.

The usual Gödel numbering of proofs of Peano arithmetic has the property (which we will use subsequently) that each theorem has infinitely many proofs, but each proof is the proof of exactly one theorem.

Let $h(0) = 0$. If $h(m) = i$, we put $h(m+1) = i$ unless for some $j \in S_i$, m is the Gödel number of a proof in $\boldsymbol{P}$ of $l \neq \boldsymbol{j}$. In that case set $h(m+1) = j$. This completes the definition of h.

4.4. The following arguments about the construction can, unless otherwise noted, be formalized in $\boldsymbol{P}$:

a) $h(m)$ is defined for all values of m and is $\leqq n$.

b) If $h(m) = j$, then for all $m' \geqq m$, $h(m') \in \{j\} \cup S_j$.

This is proved by induction on m' using the transitivity of $<$.

c) There is an m and a j such that for all $m' \geqq m$, $h(m') = j$.

The following proof of (c) is easily formalized in $\boldsymbol{P}$. Let $F(i)$ = cardinality of S_i. Then if $j \in S_i$, $F(j) < F(i)$. For some m, $F(h(m))$ takes on the minimum value of $F \circ h$. But then for all $m' \geqq m$, we must, by (b), have $h(m') = h(m)$.

So l denotes the eventual value of $h(m)$, for m sufficiently large, provided by (c). Part (1) of Lemma 4.1 is now clear.

d) If $l = i$, and $j \in S_i$, then $\boldsymbol{P} + $ "$\boldsymbol{l} = \boldsymbol{j}$" is consistent.

We suppose that (d) is false and derive a contradiction. Pick m so that for all $m' \geqq m$, $h(m') = i$. If (d) is false, $\boldsymbol{P} \vdash l \neq \boldsymbol{j}$. Let $m' \geqq m$ be some stage at which $l \neq \boldsymbol{j}$ is proved. By our choice of m, $h(m') = h(m'+1) = i$. On the other hand, inspection of the definition of h shows that if $h(m') = i$, $h(m'+1) = j$. But $j \neq i$, since $j \in S_i$, and we have our contradiction. By formalizing this argument in $\boldsymbol{P}$, we get (4) of Lemma 4.1.

e) If $l = i$, and $j \notin \{i\} \cup S_i$, then P proves $l \neq \boldsymbol{j}$.

If $l = i$, then for some m, $h(m) = i$. Since $\boldsymbol{P}$ proves all true Σ_1^0 sentences, $\boldsymbol{P}$ proves $h(\boldsymbol{m}) = \boldsymbol{i}$. By formalizing the proof of (b), $\boldsymbol{P}$ proves that if $h(\boldsymbol{m}) = \boldsymbol{i}$, $l \in \{\boldsymbol{i}\} \cup S_i$. Finally, $\boldsymbol{P}$ proves $\boldsymbol{j} \notin \{\boldsymbol{i}\} \cup S_i$. (e) is now clear.

f) If $l = i$, and $i > 0$, then P proves $l \neq \boldsymbol{i}$.

Let m be the least integer such that $h(m+1) = i$. By the definition of h, m is a proof of $l \neq \boldsymbol{i}$.

By formalizing the proofs of (e) and (f) within $\boldsymbol{P}$, we get (5) of Lemma 4.1.

4.5. The following arguments cannot be completely formalized in $\boldsymbol{P}$.

g) $l = 0$.

If $l = i > 0$, then by (f), $\boldsymbol{P}$ proves "$l \neq \boldsymbol{i}$". But all the theorems of $\boldsymbol{P}$ hold in the standard model. Thus to escape contradiction, we must have $l = 0$.

h) For $0 \leqq i \leqq n$, $\boldsymbol{P} + $ "$l = \boldsymbol{i}$" is consistent.

For $i = 0$, the standard model, by (g), is a model of $\boldsymbol{P} + $ "$l = \boldsymbol{i}$". The cases when $i > 0$ follow from (g) and (d).

Since (g) and (h) are precisely (2) and (3) of Lemma 4.1, the lemma is now proved.

4.6. The following result is the main theorem of this paper.

THEOREM. *A modal formula χ is $\boldsymbol{P}$-valid iff χ is a theorem of G.*

PROOF. By Lemma 2.3 every theorem of G is P-valid. Let now χ be a modal formula which is not a theorem of G. We shall find an interpretation of $\boldsymbol{M}$ in $\boldsymbol{P}$ such that χ^* is not a theorem of $\boldsymbol{P}$. We shall show in fact that we can arrange that for each propositional variable v, v^* is (provably equivalent to) a Boolean combination of Σ_1^0 sentences.

4.7. We first apply Theorem 3.6 to χ. We may clearly assume that $X = \langle 1, \cdots, n \rangle$ for some n, and that $x_0 = 1$. Thus Theorem 3.6 says that there is a model $\langle X; \succ, e \rangle$ such that

1) $X = \{1, \cdots, n\}$;
2) $e(\chi, 1) = 0$;
3) For $1 < j \leqq n$, $j \prec 1$;
4) $\prec$ is transitive and well-founded.

4.8. We now invoke Lemma 4.1 getting a term l.

For the purposes of §5, it will be convenient to set $e(v,0) = e(v,1)$.

We now define the interpretation we use to show that χ is not $\boldsymbol{P}$-valid. We set

$$v^* = \bigvee \{ l = \boldsymbol{i} : 0 \leqq i \leqq n \quad \text{and} \quad e(v, i) = 1 \}.$$

Here the right hand side is a finite disjunction. If the right hand side is the empty disjunction, put $v^* =$ "$\mathbf{0} = \mathbf{1}$".

LEMMA. *Let ψ be a modal formula. Let $1 \leqq i \leqq n$. Then*

1) *If $e(\psi, i) = 1$,*

$$\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow \psi^*.$$

2) *If $e(\psi, i) = 0$,*

$$\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow \neg \psi^*.$$

PROOF. The proof is by induction on $l(\psi)$. For v a variable, this is clear from the definition of v^*. The cases when $\psi = \perp$ and when ψ has principal connective $\rightarrow$ are straighforward.

Now suppose that ψ is $\Box \theta$. We consider two cases:

Case 1. $e(\Box \theta, i) = 1$.

By (1) and (5) of Lemma 4.1, if $i > 0$,

i) $\boldsymbol{P} \vdash$ If $l = \boldsymbol{i}$, then $\boldsymbol{P} \vdash l \in S_i$.

Since $e(\Box \theta, i) = 1$, $e(\theta, j) = 1$ for all $j \in S_i$. Whence, by induction hypothesis,

ii) $\boldsymbol{P} \vdash l = \boldsymbol{j} \rightarrow \theta^*$, $\quad (j \in S_i)$.

Whence,

iii) $\boldsymbol{P} \vdash l \in S_i \rightarrow \theta^*$.

By (i) and (iii),

iv) $\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow$ "$\boldsymbol{P} \vdash \theta^*$",

i.e., $\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow (\Box \theta)^*$.

Case 2. $e(\Box\,\theta, i) = 0$.

Then for some $j \in S_i$, we have $e(\theta, j) = 0$. But then by induction hypothesis,

i) $\boldsymbol{P} \vdash l = \boldsymbol{j} \rightarrow \neg\,\theta^*$.

By Lemma 4.1 (4),

ii) $\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow \mathrm{Con}(\boldsymbol{P} + \text{“}l = \boldsymbol{j}\text{”})$.

By (i) and (ii),

iii) $\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow \mathrm{Con}(\boldsymbol{P} + \neg\,\theta^*)$,

i.e., $\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow \neg(\Box\,\theta)^*$.

4.9. The proof of Theorem 4.6 is now immediate. Since $e(\chi, 1) = 0$, $\boldsymbol{P} \vdash l = \boldsymbol{1} \rightarrow \neg\,\chi^*$. By Lemma 4.1 (3), $\boldsymbol{P} + \text{“}l = \boldsymbol{1}\text{”}$ is consistent. *A fortiori*, $\boldsymbol{P} + \neg\,\chi^*$ is consistent, i.e., χ^* is not a theorem of $\boldsymbol{P}$. So χ is not $\boldsymbol{P}$-valid.

It remains to see that in our construction, each v^* is provably equivalent to a Boolean combination of Σ_1^0 sentences. It evidently suffices to see that $l = \boldsymbol{i}$ is a Boolean combination of Σ_1^0 sentences. But $l = i$ iff 1) for some m, $h(m) = i$ and it is not the case that 2) for some m, $h(m) \in S_i$. Since (1) and (2) are clearly Σ_1^0 sentences, our claim follows.

4.10. The following rather technical result will be needed in §5. We keep the context of §§4.7–4.8.

LEMMA. *Let χ be a formula. We suppose that for each subformula of χ of the form $\Box\,\theta$, $e(\Box\,\theta \rightarrow \theta, 1) = 1$. Then if ψ is a subformula of χ:*

1) *If $e(\psi, 1) = 1$,*

$$\boldsymbol{P} \vdash l = \boldsymbol{0} \rightarrow \psi^*.$$

2) *If $e(\psi, 1) = 0$,*

$$\boldsymbol{P} \vdash l = \boldsymbol{0} \rightarrow \neg\,\psi^*.$$

PROOF. The proof is by induction on $l(\psi)$. The case when ψ is a variable, v, is clear from our definition of v^*. (This is why we arranged previously that $e(v, 0) = e(v, 1)$.)

As usual, the only problematical case is when ψ has the form $\Box\,\theta$.

Case 1. $e(\Box\,\theta, 1) = 1$.

If $1 < i \leqq n$, $i \in S_1$ so $e(\theta, i) = 1$. Also by the hypothesis of the lemma, $e(\theta, 1) = 1$. It follows by Lemma 4.8 and our induction hypothesis, that if $0 \leqq i \leqq n$, $\boldsymbol{P} \vdash l = \boldsymbol{i} \rightarrow \theta^*$. By Lemma 4.1 (1), $\boldsymbol{P} \vdash \boldsymbol{0} \leqq l \leqq \boldsymbol{n}$. Whence $\boldsymbol{P} \vdash \theta^*$. So $(\Box\,\theta)^* = \mathrm{Bew}(\text{‘}\theta^*\text{’})$ is a true Σ_1^0 sentence. Hence, $\boldsymbol{P} \vdash (\Box\,\theta)^*$. *A fortiori*, $\boldsymbol{P} \vdash l = \boldsymbol{0} \rightarrow (\Box\,\theta)^*$.

Case 2. $e(\Box\theta, 1) = 0$.

Then for some j with $1 < j \leqq n$, $e(\theta, j) = 0$. By Lemma 4.8, $\boldsymbol{P} \vdash l = \boldsymbol{j} \rightarrow \neg\theta^*$. By Lemma 4.1 (4), $\boldsymbol{P} \vdash l = \boldsymbol{0} \rightarrow \mathrm{Con}(\boldsymbol{P} + `l = \boldsymbol{j}\text{'})$. So $\boldsymbol{P} \vdash l = \boldsymbol{0} . \rightarrow . \mathrm{Con}(\boldsymbol{P} + \neg\theta^*)$, i.e., $\boldsymbol{P} \vdash l = \boldsymbol{0} \rightarrow \neg(\Box\theta)^*$.

§5. The case of truth

5.1. We begin by introducing a new system of modal logic G'. G' will have two axiom schemata and one rule of inference:

A4: All theorems of G.

A5: $\Box\chi \rightarrow \chi$.

R1: If $\vdash \chi$ and $\vdash \chi \rightarrow \psi$, then $\vdash \psi$.

The main result of this section is:

THEOREM. *A formula χ is ω-valid iff χ is a theorem of G'.*

5.2. Theorem 5.1 will follow easily from what we have already proved. Before showing this, we note the following corollary:

COROLLARY 5.2. *G' is not closed under* R2. *(Recall that* R2 *says: From $\vdash \chi$ infer $\vdash \Box\chi$.)*

PROOF. We suppose that G' is closed under R2 and derive a contradiction. By R2 and A5, $G' \vdash \Box(\Box\bot \rightarrow \bot)$. By A3, $G' \vdash \Box(\Box\bot \rightarrow \bot) \rightarrow \Box\bot$. By R1, $G' \vdash \Box\bot$. By Theorem 5.1, $\Box\bot$ is ω-valid, i.e., $\boldsymbol{P}$ is inconsistent, which is absurd.

5.3. LEMMA. *Every theorem of G' is ω-valid.*

PROOF Clearly the set of ω-valid formulae is closed under R1. By Lemma 2.3, it contains all instances of A4. Since the theorems of $\boldsymbol{P}$ hold in the standard model of arithmetic, all instances of A5 are ω-valid. The lemma follows:

5.4. LEMMA. *Let χ be a modal formula Let $\Box\psi_1, \cdots, \Box\psi_r$ be all the subformulae of χ with principal connective $\Box$. Then if*

$$(\alpha) \qquad [(\Box\psi_1 \rightarrow \psi_1) \wedge \cdots \wedge (\Box\psi_r \rightarrow \psi_r)] \rightarrow \chi$$

is not a theorem of G, then there is an interpretation of $\boldsymbol{M}$ in $\boldsymbol{P}$ such that χ^* is false in the standard model.

PROOF. We apply Theorem 3.6 to (α). We get a model $\langle X; \succ, e\rangle$ such that (1) through (4) of §4.7 hold and, in addition,

$$e(\Box\psi_i \rightarrow \psi_i, 1) = 1 \qquad (1 \leqq i \leqq r).$$

By Lemma 4.10,

$$\boldsymbol{P} \vdash l = \mathbf{0} \rightarrow \neg \chi^* .$$

By Lemma 4.1 (2), $l = \mathbf{0}$ is true. Whence χ^* is false, as desired.

5.5. Theorem 5.1 follows readily. If (α) is a theorem of G, χ is a theorem of G'. Thus if χ is not a theorem of G', χ is not ω-valid. This completes the proof of Theorem 5.1.

It is clear from Lemmas 5.3 and 5.4 that χ is a theorem of G' iff (α) is a theorem of G. Since G is decidable, so is G'.

§6. Other results

6.1. In this section, we present some further results, without proofs. We first discuss what properties of $\boldsymbol{P}$ are needed in the proof of the main theorem. Then we present some results to the effect that there is no simple normal form for G-equivalence classes of modal formulae. Finally, we discuss other notions of "provability", such as holding in all transitive models. It turns out that for some of these other notions of provability, other modal systems than G come into play.

6.2. Our results about $\boldsymbol{P}$-validity adapt without essential change to a theory $\boldsymbol{T}$ satisfying the following conditions:

1) $\boldsymbol{T}$ is recursively axiomitizable;
2) $\boldsymbol{P}$ is relatively interpretable in $\boldsymbol{T}$;
3) $\boldsymbol{T}$ is Σ_2^0-sound.

(I.e., if χ is a Σ_2^0 sentence of $\boldsymbol{P}$, and χ^* is the sentence of $\boldsymbol{T}$ that is the interpretation of χ, then if $T \vdash \chi^*$, then χ is true.) Condition (3) follows from the more familiar condition that $\boldsymbol{T}$ is ω-consistent.

6.3. We define an equivalence relation on the set of modal formulae containing no propositional variable other than v_0 by putting $\chi_0 \sim \chi_1$ iff $G \vdash \chi_0 \leftrightarrow \chi_1$. The set of equivalence classes forms a Boolean algebra, B_1.

B_1 contains, as a subalgebra, the equivalence classes of formulae that contain no propositional variables. Call this subalgebra B_0. Then the results of Boolos imply the following facts about B_0:

1) The Stone space of all homomorphisms of B_0 is countable, with exactly one non-isolated point.

2) Every element of B_0 is a Boolean combination of elements of the form $\Box^r \perp$. (Here $\Box^r$ is the r-fold iterate of $\Box$.)

Before stating the situation for B_1, we define a sequence, H_n, of subalgebras of B_1. H_0 is the subalgebra generated by v_0. H_{n+1} is the subalgebra generated by elements of the form $\Box^r \psi$ where $\psi \in H_n$. Evidently B_1 is the union of the sequence H_n of subalgebras.

We interpret the following results as saying that B_1 is much more complicated than B_0.

1*) The Stone space of B_1 has power $2^{\aleph_0}$.

2*) For no n is $B_1 = H_n$.

(By (2) $B_0 \equiv H_1$.)

6.4. We now consider various notions of interpretation of $\boldsymbol{M}$ in **ZFC** (Zermelo–Frenkel set theory including the axiom of choice). Our notions will differ in their treatment of $\Box$.

a) The analog of the definition we used in defining $\boldsymbol{P}$-validity is: $(\Box\chi)^* = \chi^*$ is a theorem of **ZFC**. Equivalently, $(\Box\chi)^* = : \chi^*$ holds in all models of **ZFC**. The remarks of §6.1 handle this case.

b) The second notion we consider is obtained by taking $(\Box\chi)^* = : \chi^*$ holds in all ω-models of **ZFC**. Equivalently, $(\Box\chi)^* = : \chi^*$ is provable in **ZFC** + ω-rule.

In order to investigate this notion, we assume: **ZFC** has a countable transitive model.

The appropriate modal logic is again the system G. In proving the analogue of Lemma 4.1 one uses the well-ordering of proofs in ω-logic of order type ω_1 in place of the usual ω ordering of proofs in first order logic.

6.5. The next interpretation we consider is: $(\Box\chi)^* = : \chi^*$ holds in all transitive models of **ZFC**. The relevant assumption is: **ZFC** has an uncountable transitive model.

Before stating our results, we review the facts on transitive models that underlie them.

Let M be a transitive model of **ZFC**. Then the set of ordinals of M form an initial segment of the ordinals, say λ_M. (As usual, each ordinal is taken to be the set of all smaller ordinals.)

Let χ be a sentence of the language of set theory. We let $\lambda(\chi) = \min\{\lambda_M : M \models \mathbf{ZFC} + \chi\}$. If $\mathbf{ZFC} + \chi$ has no transitive models, $\lambda(\chi) = \infty$.

Let M be a transitive model of **ZFC**. Then $M \models$ "'$\mathbf{ZFC} + \chi$' has a transitive model" iff $\lambda(\chi) < \lambda_M$. In that case,

$$\lambda(\chi)^M = \lambda(\chi).$$

6.6. The set of modal formulae valid with respect to the notion and interpretation presented in §6.5 is precisely the set of theorems of the following modal system H.

H contains all the axiom schemata and rules of inference of G and in addition the following axiom schemata.

A6. $$\Box(\chi \to \Diamond \psi) . \vee . \Box(\psi \to \Diamond \chi) . \vee . \Box(\Diamond \chi \leftrightarrow \Diamond \psi).$$

Our proof yields that H has the finite model property. Indeed a sufficient stock of frames to yield counter-models to all the non-theorems of H is obtained by considering finite sets X equipped with a map h of X onto some finite integer n. We put $x_1 > x_2$ iff $h(x_1) > h(x_2)$.

A careful analysis of the proof of the finite model property yields a decision procedure for H that halts in at most $2^{cn \log n}$ steps (where n is the length of the formula being checked for H-theoremhood and c is some suitable absolute constant).

On can similarly discuss the class of interpretations where $(\Box \chi)^*$ is: For every inaccessible cardinal κ, $R(\kappa) \models \chi^*$.

Here the relevant system is obtained by adding to H the schema:

A7. $$\Diamond \chi \wedge \Box(\Diamond \chi \leftrightarrow \Diamond \psi) . \to . \Box[\chi \wedge \Box \sim \chi . \to \psi].$$

An adequate class of Kripke models is the set of finite linearly ordered sets. This can be proved in **ZFC** + "There are infinitely many inaccessible cardinals".

Bibliographical remark

Professors Smorynski and Magari have helped me in understanding in more detail the previous work done on the problems considered in this paper. The following remarks are my summary of their letters.

Friedman's 35th problem has also been solved by van Bentham (a student of deJongh) in his doctoral dissertation, and by Professors Bernardi and Montagna, of Sienna, in a paper entitled "Solution of Problem 35 of Harvey Friedman".

The Sienna group (consisting of Professors Magari, Montagna and Bernardi) had also shown that axion A2 of G is redundant, and that the set of theorems of G is recursive.

References

1. George Boolos, *Friedman's 35th problem has an affirmative solution*, Abstract *75T–E66, Notices Amer. Math. Soc. **22** (1975), A–646.

2. Harvey Friedman, *One hundred and two problems in mathematical logic*, J. Symbolic Logic **40** (1975), 113–129.

3. S. C. Kleene, *Introduction to Metamathematics*, Van Nostrand, New York, 1952.

4. Saul Kripke, *Semantical analysis of modal logic I*, Z. Math. Logik Grundlagen Math. **9** (1963), 67–96.

5. M. H. Löb, *Solution of a problem of Leon Henkin*, J. Symbolic Logic **20** (1955), 115–118.

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598 USA

Present address:

Department of Mathematics
California Institute of Technology
Pasadena, California 91125 USA