Notre Dame Journal of Formal Logic
Volume 31, Number 2, Spring 1990

The Modal Logic of Pure Provability

SAMUEL R. BUSS*

Abstract We introduce a propositional modal logic PP of "pure" provability in arbitrary theories (propositional or first-order) where the $\Box$ operator means "*provable in all extensions*". This modal logic has been considered in another guise by Kripke. An axiomatization and a decision procedure are given and the $\Box\Diamond$ subtheory is characterized.

1 Introduction This paper discusses a modal logic PP of *pure* provability; that is to say, of provability in arbitrary theories (propositional or first-order). The modal formula $\Box\phi$ is intended to mean "ϕ is provable in all possible extensions of the present theory"; the subtleties arise in the interpretation of iterated modalities. The modal theory we are studying is very different from the provability interpretation of Solovay [6]; we allow theories which are much weaker than Peano Arithmetic and which may be unable to formalize metamathematics. Our theory PP was briefly mentioned by Kripke [3]; however, Kripke's aims were somewhat different from ours and he did not explore PP in depth. The aim and motivation of this paper is to give a modal theory of provability; Kripke's program (fulfilled by Solovay) was to give a provability interpretation of modal logic.

It turns out that the modal theory PP of pure provability is somewhat pathological; most notably, PP is not closed under the rule of substitution of arbitrary wff's for propositional variables. On the other hand, PP does satisfy all consequences of S4 and McKinsey's axiom (of S4.1). Furthermore, we provide

*Supported in part by NSF postdoctoral fellowship DMS-8511465 and NSF Grant DMS-8701828.

I wish to thank Michael Kremer for pointing out some errors in an earlier version of this paper and suggesting some corrections.

Received April 7, 1988; revised July 8, 1988 and September 12, 1988

an axiomatization for PP and prove the relevant completeness theorem. We also give a simple characterization of the $\Box\Diamond$-fragment of PP.

A restricted form of the Grzegorczyk axiom Grz is valid for PP although the usual full Grz axiom is not. The Grzegorczyk axiom has already been studied in connection with modal logics with provability interpretations by Boolos [1] and Goldblatt [2]; however, we know of no direct connections of this with our work.

McDermott [4] and McDermott and Doyle [5] defined a modal logic intended for nonmonotonic logic with an operator M meaning 'is consistent'. In spite of the linguistic similarity between our definition of $\Diamond$ and their definition of M, the resulting mathematical systems appear to be radically different. It would be interesting to investigate whether our modal logic of pure provability can be utilized to describe nonmonotonic logics.

2 Semantics of pure provability logic We define the modal theory PP by defining the truth of a modal formula ϕ with respect to a theory T. The theory T is a propositional theory with, say, binary connectives $\wedge$, $\vee$, and $\rightarrow$, unary connective $\neg$, and propositional variables $p, q, r, \ldots$. The modal formula ϕ may involve these connectives plus the modal necessitation operator $\Box$. The symbol $\Diamond$ is an abbreviation for $\neg\Box\neg$.

Definition We define ϕ *is true for* T *under the pure provability interpretation*, denoted $T \vDash_{\mathrm{PP}} \phi$, inductively on the number of occurrences of the modal operator $\Box$ in ϕ:

(1) If ϕ is $\Box$-free then $T \vDash_{\mathrm{PP}} \phi$ if and only if $T \vdash \phi$ (i.e., ϕ is a consequence of T);
(2) If ϕ is $\Box\psi$ then $T \vDash_{\mathrm{PP}} \phi$ if and only if for every consistent extension S of T, $S \vDash_{\mathrm{PP}} \psi$;
(3) Otherwise, let $\Box\psi_1, \ldots, \Box\psi_k$ be the maximal subformulas of ϕ which have outermost connective $\Box$. Let ϕ^* be obtained from ϕ by replacing each $\Box\psi_j$ by the tautology $p \vee \neg p$ if $T \vDash_{\mathrm{PP}} \Box\psi_j$ and by $p \wedge \neg p$ if $T \nvDash_{\mathrm{PP}} \Box\psi_j$. Now ϕ^* is $\Box$-free and we define $T \vDash_{\mathrm{PP}} \phi$ to hold if and only if $T \vDash_{\mathrm{PP}} \phi^*$ (i.e., if and only if $T \vdash \phi^*$).

Definition The *pure provability theory* PP is the theory consisting of the modal formulas which are true for every consistent propositional theory under the pure provability interpretation. A formula is said to be PP-*valid* if it is in PP.

It is obvious from the definition that PP contains every tautology. PP is also closed under modus ponens; to prove this suppose that ϕ and $\phi \rightarrow \psi$ are PP-valid. Then for an arbitrary propositional theory T, define ϕ^* and ψ^* as in case (3) above. By assumption, $T \vdash \phi^*$ and $T \vdash \phi^* \rightarrow \psi^*$ so, by modus ponens, $T \vdash \psi^*$. Hence $T \vDash_{\mathrm{PP}} \psi$. Since T was arbitrary, ψ is PP-valid.

However, PP is not closed under substitution. For example, it is easy to see that $\Box\Diamond p \rightarrow \Box p$ is in PP and yet $\Box\Diamond(q \rightarrow \Box q) \rightarrow \Box(q \rightarrow \Box q)$ is not. To prove the latter assertion note that $T \vDash_{\mathrm{PP}} (q \rightarrow \Box q)$ if and only if either $T \vdash q$ or $T \vdash \neg q$; hence $\Box\Diamond(q \rightarrow \Box q)$ is PP-valid. On the other hand, $\Box(q \rightarrow \Box q)$ is not PP-valid.

Next we give a second characterization of PP by giving a Kripke modal structure M in which the valid modal formulas comprise PP. The modal structure M consists of all "worlds" (T, τ) where T is a consistent propositional theory and τ is a truth assignment to the propositional variables such that $\tau \models T$, or in words, τ is a model of T. The reachability relation R between worlds is defined thus: (T, τ) is reachable from (S, σ) if and only if T is an extension of S (denoted by $(S, \sigma)R(T, \tau)$). We write $(T, \tau) \models_M \phi$, ϕ a modal formula, to denote ϕ being true in the world (T, τ) in the modal structure M. The definition is by induction on the complexity of ϕ; for ϕ atomic, $(T, \tau) \models_M \phi$ means that $\tau(\phi) =$ True, and the usual definitions for the propositional and modal connectives define $(T, \tau) \models_M \phi$ for more complicated ϕ. In particular, $(T, \tau) \models_M \Box\phi$ holds if and only if $(S, \sigma) \models_M \phi$ for all consistent $S \supseteq T$ and all truth assignments σ satisfying S. We say that ϕ is M-valid if $(T, \tau) \models_M \phi$ for all worlds (T, τ) of M.

Theorem 1 *A modal formula is* PP*-valid if and only if it is M-valid.*

Lemma 1 *For any wff ϕ and any consistent T, ϕ is true for T if and only if $(T, \tau) \models_M \phi$ for all $\tau \models T$.*

Proof: Theorem 1 follows immediately from Lemma 1 so we need only prove the lemma. We proceed by induction on the complexity of ϕ.

Case (1): If ϕ is $\Box$-free then the claim is an immediate consequence of the completeness theorem for propositional logic.

Case (2): If ϕ is $\Box\psi$ then ϕ is true for T if and only if ψ is true for every $S \supseteq T$. By the induction hypothesis this is equivalent to $(S, \sigma) \models_M \psi$ for all $S \supseteq T$ and $\sigma \models S$. And that is equivalent to $(T, \tau) \models_M \phi$ for all $\tau \models T$.

Case (3): If ϕ is $\psi \wedge \chi$ then ϕ is true for T if and only if both ψ and χ are. By the induction hypothesis, the latter holds if and only if $(T, \tau) \models_M \psi$ and $(T, \tau) \models_M \chi$ for all $\tau \models T$. And this is equivalent to $(T, \tau) \models_M \phi$ for all $\tau \models T$.

Case (4): The case for disjunction and negation is slightly more complicated: we first consider the case where ϕ is of the form

$$\Box\psi_1 \vee \Box\psi_2 \vee \ldots \vee \Box\psi_k \vee \neg\Box\chi_1 \vee \neg\Box\chi_2 \vee \ldots \vee \neg\Box\chi_m \vee \gamma$$

where γ is $\Box$-free. By the induction hypothesis, we have that $\Box\psi_i$ being true for T is equivalent to $(T, \tau) \models_M \Box\psi_i$ for all $\tau \models T$, and similarly for $\neg\Box\chi_i$ and for γ. Thus if ϕ is true for T then one of $\Box\psi_i$, $\neg\Box\chi_j$, and γ is true for T and hence $(T, \tau) \models_M \phi$ for all $\tau \models T$. For the converse, suppose that $(T, \tau) \models_M \phi$ for all $\tau \models T$. Because of the definition of the reachability relation R, if there is *some* $\tau \models T$ such that $(T, \tau) \models_M \Box\psi_i$ or $(T, \tau) \models_M \neg\Box\chi_i$ then for *every* $\tau \models T$, $(T, \tau) \models_M \Box\psi_i$ or $(T, \tau) \models_M \neg\Box\chi_i$, respectively. Otherwise, $(T, \tau) \models_M \gamma$ for every $\tau \models T$. In any case the induction hypothesis applied to $\Box\psi_i$, $\neg\Box\chi_i$, or γ shows that ϕ is true for T.

Case (5): When ϕ is a Boolean combination of formulas we can put ϕ in conjunctive normal form, use Case (4) to show that each conjunct satisfies Theorem 1, and then use Case (3) to show that ϕ satisfies Theorem 1.

Theorem 1 helps to justify our definition of the pure provability logic PP: although some objections could be raised to our definition of PP, the fact that PP is the set of M-valid formulas shows that PP has a natural semantics and is a reasonable theory. Kripke [3] defined a structure similar to M except that his worlds are pairs $(T, \boldsymbol{A})$, where T is a first-order theory extending Peano arithmetic and $\boldsymbol{A}$ is a structure such that $\boldsymbol{A} \models T$. The modal theory of provability in Kripke's structure is easily seen to be equivalent to the modal theory of our structure M.

3 Axioms for pure provability logic We have already noted that PP contains all tautologies and is closed under modus ponens. It is also easy to see that PP includes all theorems of S4. This is because of Theorem 1 and the fact that the reachability relation R is reflexive and transitive. In addition, as Kripke [3] noted, PP includes the McKinsey S4.1 axiom $\Diamond(\Box\phi \vee \Box\neg\phi)$ for all wff's ϕ and hence PP includes S4.1. To prove this, since every theory can be extended to a complete theory, it suffices to show that if T is a complete theory and ϕ is a wff then $T \vdash_{\mathrm{PP}} \Box\phi$ or $T \vdash_{\mathrm{PP}} \Box\neg\phi$. Since T is complete, there is a unique truth assignment τ such that $\tau \models T$ and hence in the modal structure described above the only world reachable from (T, τ) is (T, τ) itself. Thus if $(T, \tau) \models_M \phi$ then $(T, \tau) \models_M \Box\phi$ and, by Lemma 1, $T \vdash_{\mathrm{PP}} \Box\phi$. Similarly, if $(T, \tau) \models_M \neg\phi$ then $T \vdash_{\mathrm{PP}} \Box\neg\phi$. Thus either $T \vdash_{\mathrm{PP}} \Box\phi$ or $T \vdash_{\mathrm{PP}} \Box\neg\phi$.

We have established that PP contains the theory S4.1; however, PP is not closed under the rule of substitution of arbitrary wff's for propositional variables and hence cannot be equal to S4.1. Indeed, PP cannot be axiomatized by an axiom scheme closed under substitution. Nonetheless, we can axiomatize PP. We let Greek letters $\phi, \psi, \chi, \ldots$ range over arbitrary modal wff's, capital Roman letters $A, B, C, \ldots$ range over $\Box$-free wff's and lower case Roman letters $p, q, r, \ldots$ range over propositional variables.

Definition Let S4 + X be the modal theory made up of the S4 axioms, the S4 rules of inference of modus ponens and necessitation, and the additional axioms (denoted by X):

$$\Box(\Box A \to \Box B_1 \vee \cdots \vee \Box B_k \vee C) \leftrightarrow \Box(A \to B_1) \vee \cdots \vee \Box(A \to B_k) \vee \Box(A \to C)$$

where $A, B_1, \ldots, B_k, C$ are $\Box$-free formulas and $k \geq 0$.

A useful special case of the X axioms is when A is a tautology and the axioms reduce to

$$\Box(\Box B_1 \vee \cdots \vee \Box B_k \vee C) \leftrightarrow \Box B_1 \vee \cdots \vee \Box B_k \vee \Box C.$$

We now claim that the X axioms are PP-valid. Actually, the right-to-left implication of the X axioms is a consequence of S4; thus the X axioms could equivalently be stated as an implication instead of an equivalence. To show that an X axiom is PP-valid it will suffice to show that it is M-valid. We assume for simplicity that $k = 1$; now suppose that $\tau \models T$ and $(T, \tau) \models_M \Box(\Box A \to \Box B \vee C)$. By the definition of the reachability relation in the modal structure M and by Lemma 1 this is equivalent to $T \vdash_{\mathrm{PP}} \Box(\Box A \to \Box B \vee C)$. This is equivalent to the condition that if $S \supseteq T$ and $S \vdash A$ then either $S \vdash B$ or $S \vdash C$. This is fur-

ther equivalent to $T \vdash A \to B$ or $T \vdash A \to C$; which is the same as $\Box(A \to B) \vee \Box(A \to C)$ being true for T. Again by Lemma 1, this is equivalent to $(T, \tau) \models_M \Box(A \to B) \vee \Box(A \to C)$ for all (or some) $\tau \models T$.

We have established that S4 + X $\subseteq$ PP; we show below that equality holds. First we prove a lemma which is interesting in its own right—it states that the $\Box$-operator does not need to be iterated.

Lemma 2 *Every formula is equivalent in* S4 + X *to a formula with no nested modal operators. Every formula of the form* $\Box\phi$ *is equivalent to a positive Boolean combination of formulas of the form* $\Box A$ *with* A $\Box$*-free.*

Proof: The second assertion of the lemma implies the first; we prove the second assertion by induction on the complexity of ϕ. The base case where ϕ is $\Box$-free is trivial since ϕ is already in the desired form. For the induction step we can, without loss of generality, assume that ϕ is in conjunctive normal form, $\phi = \phi_1 \wedge \cdots \wedge \phi_n$. By S4, $\Box\phi$ is equivalent to $\Box\phi_1 \wedge \cdots \wedge \Box\phi_n$. By the induction hypothesis and the fact that ϕ is in conjunctive normal form, each ϕ_i is equivalent to a formula of the form

$$\Box A_1 \wedge \cdots \wedge \Box A_r \to \Box B_1 \vee \cdots \vee \Box B_s \vee C$$

where each A_j, B_j, C is $\Box$-free. This is S4-equivalent to

$$\Box A \to \Box B_1 \vee \cdots \vee \Box B_s \vee C$$

where A is $A_1 \wedge \cdots \wedge A_r$. So by axiom X, $\Box\phi_i$ is equivalent to $\Box(A \to B_1) \vee \cdots \vee \Box(A \to B_s) \vee \Box(A \to C)$. Each $\Box(A \to B_i)$ occurs positively and the lemma is proved.

Theorem 2 PP = S4 + X.

Proof: Since PP $\supseteq$ S4 + X it suffices to show that every PP-valid formula ϕ is a consequence of S4 + X. By Lemma 2, $\Box\phi$ is (S4 + X)-equivalent and hence PP-equivalent to a formula of the form

$$\psi = \bigwedge_i \bigvee_j \Box B_{i,j}$$

with each $B_{i,j}$ $\Box$-free. (Note that we are using the fact that $\Box\phi$ is equivalent to a *positive* Boolean combination of the formulas $\Box B_{i,j}$.) If any $\Box B_{i,j}$ is PP-valid then $B_{i,j}$ is a tautology and hence S4 $\vdash \Box B_{i,j}$. Thus, since $\Box\phi$ is PP-valid, ψ is a consequence of S4 and thus ϕ is a consequence of S4 + X.

Theorem 2 gives an axiomatization for PP. Note also that the proofs of Lemma 2 and Theorem 2 give a decision procedure for PP.

Our axiomatization for PP is not a finite scheme since the description of the X axioms included a variable k. The author does not know if there is a finite set of axiom schemes for PP.

4 The Grzegorczyk axiom The Grzegorczyk axiom Grz is $\Box(\Box(\phi \to \Box\phi) \to \phi) \to \phi$. To understand what this means, take the contrapositive and let ψ be $\neg\phi$. The Grzegorczyk axiom becomes

$$\psi \to \Diamond(\psi \wedge \Box(\neg\psi \to \Box\neg\psi)). \tag{1}$$

In essence this says: "If ψ is true in some world then there is some reachable world W where ψ is true such that in any world reachable from W if ψ is false then ψ is necessarily false." One way to interpret this is as a discreteness property of the reachability relation on worlds, in that it says that there is a point where ψ last changes from being true to being false.

In the modal logic of pure provability the Grzegorczyk axiom is not valid for arbitrary wff's; indeed, we show below that Grz is not PP-valid when ϕ is $p \to \Box p$. However, when ϕ is restricted to be $\Box$-free then the Grzegorczyk axioms *are* valid. This follows immediately from the X axioms as follows: We say that two wff's ψ_1 and ψ_2 are PP-equivalent, written as $\psi_1 \equiv \psi_2$, if and only if $\psi_1 \leftrightarrow \psi_2$ is PP-valid. Now if A is $\Box$-free then $\Box(A \to \Box A)$ is PP-equivalent to $\Box(A \to A)$, which is PP-valid. Thus the Grzegorczyk axiom $\Box(\Box(A \to \Box A) \to A) \to A$ is PP-equivalent to $\Box A \to A$. Since the latter formula is an axiom of S4, the Grzegorczyk axiom with $\phi = A$ a $\Box$-free formula is valid.

Next we show that the Grzegorczyk axiom is not valid when ϕ is the wff $p \to \Box p$. Indeed in this case the Grzegorczyk axiom is equivalent to ϕ itself as the following chain of PP-equivalences shows:

$$\begin{aligned}
&\Box(\Box((p \to \Box p) \to \Box(p \to \Box p)) \to (p \to \Box p)) \to (p \to \Box p)\\
&\quad\equiv \Box(\Box((p \to \Box p) \to \Box(\Box p \vee \neg p)) \to (p \to \Box p)) \to (p \to \Box p)\\
&\quad\equiv \Box(\Box((p \to \Box p) \to \Box p \vee \Box\neg p) \to (p \to \Box p)) \to (p \to \Box p)\\
&\quad\equiv \Box(\Box(p \vee \Box p \vee \Box\neg p) \to (p \to \Box p)) \to (p \to \Box p)\\
&\quad\equiv \Box((\Box p \vee \Box p \vee \Box\neg p) \to (p \to \Box p)) \to (p \to \Box p)\\
&\quad\equiv \Box(\Box\neg p \to \neg p \vee \Box p) \to (p \to \Box p)\\
&\quad\equiv \Box(\neg p \to \neg p) \vee \Box(\neg p \to p) \to (p \to \Box p)\\
&\quad\equiv \Box(\neg p \to p) \to (p \to \Box p)\\
&\quad\equiv \Box p \to (p \to \Box p)\\
&\quad\equiv p \to \Box p.
\end{aligned}$$

The above equivalences are obtained by alternately using the X axioms and replacing a subformula by a tautologically equivalent subformula; the third from last equivalence follows because $\Box(\neg p \to \neg p)$ is S4-valid. However, $p \to \Box p$ is valid for a theory T if and only if either T proves p or T disproves p; hence $p \to \Box p$ and the Grzegorczyk axiom for $p \to \Box p$ are not PP-valid.

5 The $\Box\Diamond$-fragment of PP We next show that the $\Box\Diamond$-fragment of PP has a very simple characterization. Let $\phi^{-\Box}$ be the formula obtained by removing all modal operators from ϕ while leaving the propositional connectives intact.

Theorem 3 *$\Box\Diamond\phi$ is* PP*-valid if and only if $\phi^{-\Box}$ is a tautology.*

Proof: Note that if T is complete then $T \vDash_{\mathrm{PP}} A \leftrightarrow \Box A$ for all $\Box$-free A. (Actually $T \vDash_{\mathrm{PP}} \phi \leftrightarrow \Box\phi$ for all ϕ.) Hence the modal operators in ϕ can be eliminated one at a time from ϕ to get that $T \vDash_{\mathrm{PP}} \phi^{-\Box} \leftrightarrow \phi$ for all complete T.

So if $\phi^{-\Box}$ is a tautology ϕ is true for every complete theory, and thus $\Box\Diamond\phi$ is PP-valid since every consistent theory can be extended to a complete theory. Conversely, if $\Box\Diamond\phi$ is PP-valid then ϕ is true for every complete theory and hence every complete theory $T \vdash \phi^{-\Box}$, i.e., $\phi^{-\Box}$ is a tautology.

REFERENCES

[1] Boolos, G., "On systems of modal logic with provability interpretations," *Theoria*, vol. 46 (1980), pp. 7–18.

[2] Goldblatt, R., "Arithmetical necessity, provability and intuitionistic logic," *Theoria*, vol. 44 (1978), pp. 38–46.

[3] Kripke, S., "Semantical considerations on modal logic," *Acta Philosophica Fennica*, vol. 16 (1963), pp. 83–94.

[4] McDermott, D., "Nonmonotonic logic II: Nonmonotonic modal theories," *Journal of the Association for Computing Machinery*, vol. 29 (1982), pp. 33–57.

[5] McDermott, D. and J. Doyle, "Nonmonotonic logic I," *Artificial Intelligence*, vol. 13 (1980), pp. 41–72.

[6] Solovay, R., "Provability interpretations of modal logic," *Israel Journal of Mathematics*, vol. 25 (1976), pp. 287–304.

Department of Mathematics
University of California, San Diego
La Jolla, California 92093