ALBERT VISSER

THE PROVABILITY LOGICS OF RECURSIVELY ENUMERABLE THEORIES EXTENDING PEANO ARITHMETIC AT ARBITRARY THEORIES EXTENDING PEANO ARITHMETIC

1. INTRODUCTION

Solovay's proof of the arithmetical completeness theorem for provability logic shows more than is stated in the theorem. The idea is roughly this. Let φ be a modal propositional formula. Suppose φ is not derivable in L, i.e., Löb's Logic (also known as G, for 'Gödel'). There is a finite, transitive, irreflexive Kripke model K such that $K \nvDash \varphi$. Solovay provides a specific arithmetical interpretation connected with K of the atoms of φ to show that not all arithmetical interpretations of φ are derivable in PA, Peano Arithmetic. It turns out that Solovay's interpretation does more: it connects the interpretation of φ with sentences of the form $\Box_{PA}^{k}\bot$, reflecting the way φ is connected with sentences of the form $\Box^{k}\bot$ *in the specific model K*. We employ this fact to state a lemma that captures the content of the proof better.

Consequences:

> We fully characterize the provability logics of RE theories T extending PA at PA and at T.
>
> We show that 'nearly all' provability logics of RE theories T extending PA at arbitrary U extending PA are between Löb's logic L and Solovay's logic S.

At the end of the paper we take a brief look at what goes on between L and S. We give examples of new phenomena there.

We only give definitions and facts where our way of doing things differs from the literature. For further details see [2] or [7].

2. DEFINITIONS AND ELEMENTARY FACTS

For technical convenience we do not employ finite, transitive, irreflexive Kripke models but tail models.

Journal of Philosophical Logic **13** (1984) 97–113. 0022-3611/84/0131–0097$01.70

2.1. DEFINITION. A *tail model* K is a triple $\langle \omega, \prec, f \rangle$ where

ω is the set of natural numbers.

$\prec$ is a binary, transitive, irreflexive relation on ω.

f: $\omega \to$ the power set of $\mathscr{P}$, where $\mathscr{P}$ is the set of propositional atoms of $\mathscr{L}_\Box$, the language of propositional modal logic.

if $m \neq 0$ then $0 \prec m$.

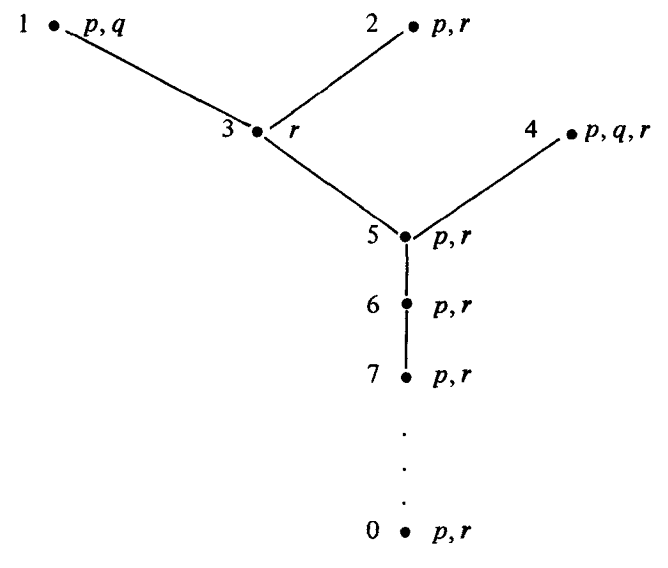
if $m \neq 0$ and $m \prec n$ then $m > n$. Here $>$ is the usual order on ω.

for some $N \neq 0$:
for every $n \geqslant N$ and every $m \neq 0$ if $n > m$ then $n \prec m$,
for every $n \geqslant N$, $f(n) = f(N)$,
$f(0) = f(N)$.

Such an N will be called a *tail element*.

$\models_K$ is defined as usual, where we take $f(n)$ as the set of p_i such that $n \models_K p_i$.

Example with $N = 5$:



Clearly every finite transitive irreflexive Kripke model is the top part of some tail model modulo isomorphism and vice versa: proper top parts of tail models are finite, transitive, irreflexive Kripke models.

2.2. TAIL LEMMA. For every tail model K:

$$0 \models_K \varphi \text{ iff for some } M, \text{ for all } n \geqslant M, n \models_K \varphi,$$

$$0 \not\models_K \varphi \text{ iff for some } M, \text{ for all } n \geqslant M, n \not\models_K \varphi.$$

Proof. Induction on φ. ∎

2.3. CONVENTION. For any set of formulas Y we will write $L + Y$ for the closure under modus ponens of the theorems of L plus Y. Thus:

$$S = L + \{(\Box\varphi \to \varphi) \mid \varphi \in \mathscr{L}_\Box\}.$$

2.4. COMPLETENESS THEOREMS FOR TAIL MODELS

(i) $L \vdash \varphi$ iff for all tail models K, for all n: $n \models_K \varphi$.

(ii) $L \vdash \varphi$ iff for all tail models K, for all tail elements N: $N \models_K \varphi$.

(iii) $S \vdash \varphi$ iff for all tail models K: $0 \models_K \varphi$.

Proof. (i) and (ii) are trivial consequences of the corresponding completeness theorems for finite, transitive, and irreflexive Kripke models.

(iii) "$\Rightarrow$". 0 satisfies the theorems of L because K is upwards well-founded. Closure under modus ponens is trivial. Suppose $0 \models_K \Box\varphi$, then for all $n \succ 0$, $n \models_K \varphi$. Hence by the tail lemma $0 \models_K \varphi$.

"$\Leftarrow$". Suppose $S \nvdash \varphi$, then certainly $L \nvdash \bigwedge (\Box\psi_i \to \psi_i) \to \varphi$, where the ψ_i are those subformulas of φ that have a box in front of them. Hence there are a tail model K and a tail element N such that $N \models_K \bigwedge (\Box\psi_i \to \psi_i)$ and $N \not\models_K \varphi$. That $0 \not\models_K \varphi$ follows easily from:

Claim. For all subformulas χ of φ:

$$\text{if } N \models_K \chi \text{ then for all } m \preccurlyeq N, m \models_K \chi,$$

$$\text{if } N \not\models_K \chi \text{ then for all } m \preccurlyeq N, m \not\models_K \chi.$$

Proof of claim. By induction on χ. The cases of atoms and propositional connectives are trivial. In case $\chi = \Box\rho$ we have: if $N \models_K \Box\rho$ then, because ρ is a subformula of φ with a box in front, $N \models_K \rho$. By Induction Hypothesis for all $m \preccurlyeq N$, $m \models_K \rho$. On the other hand for all $m \succ N$, $m \models_K \rho$. So for all k, $k \models_K \rho$. Conclude for $m \preccurlyeq N$: $m \models_K \Box\rho$. The second case is trivial. ∎

We will be interested in closed sentences, i.e., sentences in which no variables p_i occur and *degrees of falsity*.

2.5. DEFINITION

$$\Box^0 \bot := \bot$$
$$\Box^{k+1} \bot := \Box(\Box^k \bot)$$
$$\Box^\omega \bot := (\neg\bot).$$

Similarly we write $\Box_T^k\bot$ for the interpretation of $\Box^k\bot$ in T and $\Box_T^\omega\bot$ for $(0 = 0)$.

Note that our notation $\Box^\omega$ is perhaps a bit misleading for there is a perfectly natural interpretation in arithmetic for that expression different from ours.

2.6. DEFINITION. Consider a tail model K. Define the depth $d(n)$ of a node n as follows:

$$d(n) := 1 + \sup\{d(m) \mid n \prec m\}.$$

Note that if n is a top element of K then $d(n) = 1$ and that $d(0) = \omega$.

Moreover:

$$d(n) = \alpha \text{ iff } n \models \Box^\alpha\bot \text{ and } n \not\models \Box^\beta\bot \text{ for all } \beta < \alpha.$$

2.7. DEFINITION. A set X of formulas of $\mathscr{L}_\Box$ is *standard closed* if there is one element of the form $\Box^\alpha\bot$ in X ($\alpha \in \omega \cup \{\omega\}$) and all other elements are of the form $(\Box^{k+1}\bot \to \Box^k\bot)$ where $k + 1 < \alpha$.

2.8. NORMAL FORM THEOREM FOR SETS OF CLOSED FORMULAS. For any set Y of closed formulas of $\mathscr{L}_\Box$ there is a unique standard closed X such that $L + Y = L + X$.

Proof. Entirely routine. ∎

We turn to arithmetical interpretations.

2.9. DEFINITION. (i) Let $\mathscr{S}_{\mathrm{PA}}$ be the set of sentences of the language of Arithmetic. A function $g\colon \mathscr{P} \to \mathscr{S}_{\mathrm{PA}}$ will be called an *a-assignment*.

We define for a-assignments g and RE theories T extending PA in the language of PA:

$$\langle p_i \rangle(g, T) := g(p_i).$$

$\langle \cdot \rangle(g, T)$ commutes with the propositional constants.

$$\langle \Box\varphi \rangle(g, T) := \Box_T(\langle\varphi\rangle(g, T)).$$

As far as this paper is concerned we could as well have chosen to consider g from $\mathscr{P}$ to the *formulas* of Arithmetic under the appropriate conventions for handling free variables within the range of $\Box_T$. It could very well be that for further research this latter choice is the more natural.

(ii) Let T be an RE theory in the language of PA extending PA, let U be an arbitrary theory in the language of PA extending PA. Let Γ_0, Γ_1 be sets of formulas of $\mathscr{L}_\Box$. Define:

$\Gamma_0; \Gamma_1 \models \varphi(U, T)$ iff for all a-assignments f:

$$U + \{\langle\psi\rangle(g, T) \mid \psi \in \Gamma_0,\ g\ a\text{-assignment}\} + \{\langle\chi\rangle(f, T) \mid \chi \in \Gamma_1\} \vdash \langle\varphi\rangle(f, T).$$

$\Gamma_0; \models \varphi(U, T)$ iff $\Gamma_0; \emptyset \models \varphi(U, T)$.

$\Gamma_1 \models \varphi(U, T)$ iff $\emptyset; \Gamma_1 \models \varphi(U, T)$.

$\models \varphi(U, T)$ iff $\emptyset; \emptyset \models \varphi(U, T)$.

$L(U, T) := \{\varphi \mid\ \models \varphi(U, T)\}$. ($L(U, T)$ is the provability logic of T at U.)

Our definition of $\models$ is of a subtlety not really needed in the paper. Still it is nice to have the general notion around and to see how certain results can be formulated in terms of it.

2.10. THEOREM. $\Gamma \models \varphi(\mathrm{PA}, \mathrm{PA})$ iff $L + \Gamma \vdash \varphi$.

Proof. This is just a version of Solovay's theorem. To show that Γ may be infinite one uses the uniformized version of Solovay's theorem as proved by Artyomov, Montagna and Boolos (see, e.g., [1, 3–5]). ∎

2.11. COROLLARY (Artyomov). Let X be a set of closed formulae of $\mathscr{L}_\Box$ then:

$$X; \models \varphi \text{ iff } L + X \vdash \varphi.$$

Proof. Clearly $X; \models \varphi$ iff $X \models \varphi$. Use 2.10. ∎

3. A SHARPENED VERSION OF SOLOVAY'S THEOREM

Consider an RE theory T extending PA in the language of PA and a tail model K. We assume that the proof predicate $\mathrm{Proof}_T(x, y)$ satisfies:

$$\vdash_{\mathrm{PA}} \mathrm{Proof}_T(x, y) \wedge \mathrm{Proof}_T(x, z) \to y = z,$$
$$\vdash_{\mathrm{PA}} \neg \mathrm{Proof}_T(\mathbf{0}, y).$$

Define for g: $\omega \to \omega$: $\lim g = s$ iff for some m, $g(m) = s$ and for all p, n if $g(p) = s$ and $n > p$ then $g(n) = s$.

We can define or paraphrase in the language of arithmetic a term '$l(a)$' for '$\lim g$' where g is a total recursive function with index a.

By the recursion theorem we find a recursive function h with index e such that:

$$h(0) := 0,$$

$$h(k+1) := \begin{cases} n & \text{if for some } n \succ h(k), \mathrm{Proof}_T(k+1, \ulcorner l(\mathbf{e}) \neq \mathbf{n} \urcorner) \\ h(k) & \text{otherwise.} \end{cases}$$

Write 'l' for '$l(\mathbf{e})$'. We have:

$$\vdash_{\mathrm{PA}} \text{“}h \text{ is weakly monotonic in } \prec\text{”},$$
$$\vdash_{\mathrm{PA}} \text{“}l \text{ exists”}.$$

Because PA shows the existence of l scope problems are irrelevant.

Define:

$$[\varphi](K, T) := \begin{cases} \bigvee\{l = \mathbf{i} \mid i \models_K \varphi\} \\ \quad \text{if for only finitely many } i,\ i \models_K \varphi, \\ \bigwedge\{l \neq \mathbf{i} \mid i \not\models_K \varphi\} \\ \quad \text{if for only finitely many } i,\ i \not\models_K \varphi. \end{cases}$$

(We set $\bigvee\emptyset := (\bot)$, $\bigwedge\emptyset := (\mathbf{0} = \mathbf{0})$). By the tail lemma $[\varphi](K, T)$ is always defined.

Let $g_{K,T}(p_i) := [p_i](K, T)$ and $\langle\varphi\rangle(K, T) := \langle\varphi\rangle(g_{K,T}, T)$. We have:

3.1. MAIN LEMMA

$$\text{PA} \vdash \langle\varphi\rangle(K, T) \leftrightarrow [\varphi](K, T).$$

Proof. By induction on φ. The only interesting case is: $\varphi = \Box\psi$.

(a) In case there are only finitely many i s.t. $i \nVdash_K \Box\psi$ by the tail lemma: $n \models_K \Box\psi$ for all n and hence $n \models_K \psi$ for all n. By Induction Hypothesis:

$$\vdash_{\text{PA}} \langle\psi\rangle(K, T) \leftrightarrow (\mathbf{0} = \mathbf{0}). \quad \text{Hence}$$

$$\vdash_{\text{PA}} \Box_T(\langle\psi\rangle(K, T)), \quad \text{or}$$

$$\vdash_{\text{PA}} \langle\Box\psi\rangle(K, T) \leftrightarrow [\Box\psi](K, T).$$

(b) Suppose there are only finitely many i s.t. $i \models_K \Box\psi$. Let $j_0, \ldots, j_s$ be all j such that $j \models_K \Box\psi$, $j \nvDash_K \psi$. Note that if $i \nvDash_K \Box\psi$, then there is a k such that $i \prec j_k$. Clearly by Induction Hypothesis and the fact that T extends PA it is sufficient to show:

$$\vdash_{\text{PA}} \Box_T[\psi](K, T) \leftrightarrow [\Box\psi](K, T).$$

Argue in PA.

"$\rightarrow$".

Suppose $\Box_T[\psi](K, T)$. We have: $\Box_T(l \neq \mathbf{j}_k)$ by the definition of [ ] and the fact that $j_k \nvDash_K \psi$. Suppose Proof $_T(p+1, \ulcorner l \neq \mathbf{j}_k \urcorner)$ and $h(p) = y$. In case $y \prec j_k$: $h(p+1) = j_k$, so certainly not: $l \prec j_k$. In case not $y \prec j_k$ clearly not $l \prec j_k$. So not $l \prec j_k$. Conclude: $\bigwedge\{\text{not } l \prec j_k \mid k = 0, \ldots, s\}$. So by elementary reasoning: $\bigvee\{l = i \mid i \models_K \Box\psi\}$.

"$\leftarrow$".

Suppose $l = i$ for an $i \models_K \Box\psi$. By the definition of h and the fact that $i \neq 0$ (for $0 \nvDash_K \Box\psi$ *ex hypothesi*) we find: $\Box_T\, l \neq \mathbf{i}$ ("How else could h move up to i!") Moreover $\exists x\ hx = i$ so by Σ-completeness: $\Box_T \exists x\ hx = \mathbf{i}$. Combining: $\Box_T\, l \succ \mathbf{i}$. Hence $\Box_T \bigvee\{l = \mathbf{j} \mid j \models_K \psi\}$. ∎

3.2. DEFINITION. Let T be an RE theory extending PA in the language of PA, let U be an arbitrary theory extending PA in the language of PA.

Define:

$$\alpha(U, T) := \text{the smallest } \alpha \in \omega \cup \{\omega\} \text{ such that } U \vdash \Box_T^{\alpha}\bot.$$

$$X(U, T) := \{\Box^{k+1}\bot \to \Box^k\bot \mid k+1 < \alpha \quad \text{and}$$

$$U \vdash \Box_T^{k+1}\bot \to \Box_T^k\bot\} \cup \{\Box^{\alpha}\bot\}, \text{ where } \alpha = \alpha(U, T).$$

Note that $X(U, T)$ is standard closed.

3.3. CLOSED PART LEMMA. Let T, U be as in 3.2. We have: if $X(U, T)$ is finite then $L(U, T) = L + X(U, T)$.

Proof. Set $X := X(U, T)$, $\alpha := \alpha(U, T)$. Trivially $L + X \subseteq L(U, T)$. Suppose X is finite. Suppose further $\varphi \in L(U, T)$ and $L + X \nvdash \varphi$. We have $L \nvdash \bigwedge X \to \varphi$, so there is a tail model K and a tail element N such that $N \models_K \bigwedge X$ and $N \not\models_K \varphi$. Clearly $(\Box^{d(N)}\bot \to \Box^{d(N)-1}\bot) \notin X$. Moreover $d(N) \leqslant \alpha$. For any m: $m \models_K \varphi \to (\Box^{d(N)}\bot \to \Box^{d(N)-1}\bot)$, since:

if $d(m) < d(N)$: $m \models_K \Box^{d(N)-1}\bot$.

if $d(m) = d(N)$ then $N = m$, N being a tail element, so $m \models_K \neg\varphi$.

if $d(m) > d(N)$ $m \models_K \neg\Box^{d(N)}\bot$.

So by our main lemma:

$$\mathrm{PA} \vdash \langle\varphi \to (\Box^{d(N)}\bot \to \Box^{d(N)-1}\bot)\rangle(K, T).$$

Or

$$\mathrm{PA} \vdash \langle\varphi\rangle(K, T) \to (\Box_T^{d(N)}\bot \to \Box_T^{d(N)-1}\bot).$$

Ex hypothesi $U \vdash \langle\varphi\rangle(K, T)$ and U extends PA, so

$$U \vdash \Box_T^{d(N)}\bot \to \Box_T^{d(N)-1}\bot.$$

Now $(\Box^{d(N)}\bot \to \Box^{d(N)-1}\bot) \notin X$ so not $d(N) < \alpha$. We saw before: $d(N) \leqslant \alpha$. Conclude: $d(N) = \alpha$. But then $U \vdash \Box_T^{d(N)-1}\bot$, contradicting the minimality of α. ∎

3.4. THEOREM. Let U, T be as before and $\alpha := \alpha(U, T)$. Suppose $U \subseteq T$, we have:

$$L(U, T) = L + \Box^{\alpha}\bot.$$

Proof. If $\Box^{k+1}\bot \to \Box^k\bot$ were in $X(U, T)$ we would have:

$$U \vdash \Box_T^{k+1}\bot \to \Box_T^k\bot, \quad \text{so}$$
$$T \vdash \Box_T^{k+1}\bot \to \Box_T^k\bot, \quad \text{hence}$$
$$\mathrm{PA} \vdash \Box_T(\Box_T^{k+1}\bot \to \Box_T^k\bot), \quad \text{so by Löb's Theorem:}$$
$$\mathrm{PA} \vdash \Box_T^{k+1}\bot, \quad \text{so}$$
$$U \vdash \Box_T^{k+1}\bot, \quad \text{conclude:}$$
$$U \vdash \Box_T^k\bot.$$

This contradicts the definition of $X(U, T)$. Hence only $\Box^\alpha\bot$ is in $X(U, T)$. So $X(U, T)$ is finite and $L(U, T) = L + \Box^\alpha\bot$. ∎

3.5. CONSEQUENCE

(i) $L(\mathrm{PA}, \mathrm{PA}) = L.$

(ii) $L(\mathrm{PA}, T) = L + \Box^{\alpha(\mathrm{PA}, T)}\bot = L + \Box^{1+\alpha(T, T)}\bot.$

(iii) $L(T, T) = L + \Box^{\alpha(T, T)}\bot.$

3.6. THEOREM. Let T, U be as in 3.2. Suppose $L(U, T) \nsubseteq S$, then $\alpha(U, T) \in \omega$, and so $L(U, T) = L + X(U, T)$.

Proof. Let $\varphi \in L(U, T)$, $\varphi \notin S$. There is a tail model K with $0 \nvDash_K \varphi$. By the tail lemma for some $d \in \omega$ and all m: $m \vDash_K \varphi \to \Box^d\bot$. By our main lemma: $U \vdash \langle\varphi\rangle(K, T) \to \Box_T^d\bot$. Conclude $U \vdash \Box_T^d\bot$. ∎

3.7. CONSEQUENCE. $L(\mathrm{Th}(\mathbb{N}), \mathrm{PA}) = L(\mathrm{PA} + \mathrm{RFN}(\mathrm{PA}), \mathrm{PA}) = S$.

3.8. TWO EXAMPLES. (i) Let T, U be as before. Suppose $(\Box(p \vee q) \to (\Box p \vee \Box q)) \in L(U, T)$. Clearly $L(U, T) \nsubseteq S$. Inspection of the proof shows: $\Box\bot \in L(U, T)$ or $L(U, T) = L + \Box\bot$. Of course this can also be seen more directly by substituting the Σ_1^0 Rosser Sentence R – satisfying $\mathrm{PA} \vdash R \leftrightarrow (\Box_T \neg R < \Box_T R)$ – for p and $\neg R$ for q.

On the other hand:

$$L + \{\Box(\varphi \vee \psi) \to (\Box\varphi \vee \Box\psi) \mid \varphi, \psi \in \mathscr{L}_\Box\} \vdash \Box\Box\bot$$

as can be seen by substituting $\Box\bot$ for φ and $\neg\Box\bot$ for ψ, but:

$$L + \{\Box(\varphi \vee \psi) \rightarrow (\Box\varphi \vee \Box\psi) \mid \varphi, \psi \in \mathscr{L}_\Box\} \nvdash \Box\bot.$$

This can easily be shown by considering a linear tail model: the two highest nodes will satisfy $\Box(\varphi \vee \psi) \rightarrow (\Box\varphi \vee \Box\psi)$.

Facts like the above are often more naturally formulated in terms of $\models$; for example we have:

$$\Box(p \vee q) \rightarrow (\Box p \vee \Box q);\ \models \Box\bot\,(\mathrm{PA}, \mathrm{PA}).$$

$$\{\Box(\varphi \vee \psi) \rightarrow (\Box\varphi \vee \Box\psi) \mid \varphi, \psi \in \mathscr{L}_\Box\} \models \Box\Box\bot\,(\mathrm{PA}, \mathrm{PA}).$$

$$\{\Box(\varphi \vee \psi) \rightarrow (\Box\varphi \vee \Box\psi) \mid \varphi, \psi \in \mathscr{L}_\Box\} \not\models \Box\bot\,(\mathrm{PA}, \mathrm{PA}).$$

(ii) Let R be the Σ^0_1 Rosser Sentence as in (i). Then $L(\mathrm{PA} + \neg R, \mathrm{PA}) = L$.

Proof. It is clearly sufficient to show that for no k: $\mathrm{PA} + \neg R \vdash \Box^{k+1}_{\mathrm{PA}}\bot \rightarrow \Box^k_{\mathrm{PA}}\bot$, $\neg R$ being true.

Suppose $\mathrm{PA} \vdash \neg R \rightarrow (\Box^{k+1}_{\mathrm{PA}}\bot \rightarrow \Box^k_{\mathrm{PA}}\bot)$.

In case $k = 0$: $\mathrm{PA} \vdash \Box_{\mathrm{PA}}\bot \rightarrow R$, so $\mathrm{PA} \vdash \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\bot \rightarrow \Box_{\mathrm{PA}}R$, hence

$\mathrm{PA} \vdash \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\bot \rightarrow \Box_{\mathrm{PA}}\bot$. *Quod non.*

In case $k > 0$, we have from $\mathrm{PA} \vdash \neg\Box^k_{\mathrm{PA}}\bot \rightarrow \neg R$:

$\mathrm{PA} \vdash \neg\Box^k_{\mathrm{PA}}\bot \rightarrow (\Box^{k+1}_{\mathrm{PA}}\bot \rightarrow \Box^k_{\mathrm{PA}}\bot)$, hence

$\mathrm{PA} \vdash \Box^{k+1}_{\mathrm{PA}}\bot \rightarrow \Box^k_{\mathrm{PA}}\bot$. *Quod non.* ∎

3.9. REMARK. Using 3.3 and 3.6 we can state weak local versions of Solovay's theorem in terms of $\models$, for example:

(i) $L \nvdash \varphi \Rightarrow$ for some $k \in \omega$ $\Box\varphi$; $\models \Box^k\bot\,(\mathrm{PA}, \mathrm{PA})$,

(ii) $S \nvdash \varphi \Rightarrow$ for some $k \in \omega$ φ; $\models \Box^k\bot\,(\mathrm{PA}, \mathrm{PA})$.

4. BETWEEN L AND S

What $L(U, T)$'s are there between L and S? We show that the question is not trivial by treating some examples, proving that the most obvious attempts at axiomatization are not complete.

Clearly $\{\varphi \mid \psi; \models \varphi(U', T)\}$ is an $L(U, T)$, and we will use this fact without mentioning it.

Define:

$$D := \{\varphi \mid (\Box(\Box p \vee \Box q) \to (\Box\Box p \vee \Box\Box q)); \models \varphi(\mathrm{PA}, \mathrm{PA})\}.$$

$$E := \{\varphi \mid \Box\Box p \to \Box p; \models \varphi(\mathrm{PA}, \mathrm{PA})\}.$$

$$L_D := L + \{\Box(\Box\varphi \vee \Box\psi) \to (\Box\Box\varphi \vee \Box\Box\psi) \mid \varphi, \psi \in \mathscr{L}_\Box\}.$$

$$L_E := L + \{\Box\Box\varphi \to \Box\varphi \mid \varphi \in \mathscr{L}_\Box\}.$$

The following charming theorem is a useful tool.

4.1. THEOREM (Goldfarb). Let T be an RE theory extending PA in the language of PA. We have:

(i) For every Σ_1^0 sentence A there is a Σ_1^0 sentence S such that:

$$\mathrm{PA} \vdash A \vee \Box_T\bot \leftrightarrow \Box_T S.$$

(ii) For all sentences B and C there is a Σ_1^0 sentence S such that:

$$\mathrm{PA} \vdash \Box_T B \vee \Box_T C \leftrightarrow \Box_T S.$$

Proof. (i) By the arithmetical fixed point theorem pick S such that:

$$\mathrm{PA} \vdash S \leftrightarrow A \leqslant \Box_T S.$$

We have in PA:

$$\dfrac{\dfrac{\Box_T S}{A \leqslant \Box_T S \vee \Box_T S < A} \qquad \dfrac{\dfrac{[A \leqslant \Box_T S]^{\textcircled{1}}}{A}}{A \vee \Box_T\bot} \qquad \dfrac{\dfrac{\dfrac{\dfrac{[\Box_T S < A]^{\textcircled{1}}}{\Box_T(\Box_T S < A)}}{\Box_T \neg S} \quad \Box_T S}{\Box_T\bot}}{A \vee \Box_T\bot}}{A \vee \Box_T\bot}\,1,$$

$$\dfrac{A \vee \Box_T\bot \qquad \dfrac{[\Box_T\bot]^{\textcircled{3}}}{\Box_T S} \qquad \dfrac{\dfrac{[A]^{\textcircled{3}}}{A \leqslant \Box_T S \vee \Box_T S < A} \quad \dfrac{\dfrac{[A \leqslant \Box_T S]^{\textcircled{2}}}{S}}{\Box_T S} \quad \dfrac{[\Box_T S < A]^{\textcircled{2}}}{\Box_T S}}{\Box_T S}\,2}{\Box_T S}\,3$$

(ii) $\Box_T B \vee \Box_T C$ is Σ_1^0 so by *i* for some Σ_1^0 sentence S:

$$\text{PA} \vdash \Box_T B \vee \Box_T C \vee \Box_T \bot \leftrightarrow \Box_T S.$$

Hence $\text{PA} \vdash \Box_T B \vee \Box_T C \leftrightarrow \Box_T S.$ ∎

Note that we could as well have proved 4.1 for formulas under the appropriate conventions for handling free variables.

4.2. FACT. We have:

$$\left(\Box\left(\bigvee_{i=0}^{k+1} \Box p_i\right) \rightarrow \bigvee_{i=0}^{k+1} \Box p_i\right) \in E \quad (k \in \omega).$$

Proof. Let $B_0, \ldots, B_{k+1}$ be any sentences of Arithmetic. By 4.1 there is an S such that:

$$\text{PA} \vdash \bigvee_{i=0}^{k+1} \Box_{\text{PA}} B_i \leftrightarrow \Box_{\text{PA}} S.$$

Hence

$$\begin{aligned} \text{PA} + \Box_{\text{PA}}\Box_{\text{PA}} S \rightarrow \Box_{\text{PA}} S \vdash \Box_{\text{PA}} \bigvee_{i=0}^{k+1} \Box_{\text{PA}} B_i &\rightarrow \Box_{\text{PA}}\Box_{\text{PA}} S \\ &\rightarrow \Box_{\text{PA}} S \\ &\rightarrow \bigvee_{i=0}^{k+1} \Box_{\text{PA}} B_i \end{aligned}$$

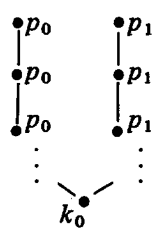
So certainly:

$$(\Box\Box p \rightarrow \Box p); \models \left(\Box \bigvee_{i=0}^{k+1} \Box p_i \rightarrow \bigvee_{i=0}^{k+1} \Box p_i\right)(\text{PA}, \text{PA}).$$ ∎

4.3. FACT

$$L_E \nvdash \Box(\Box p_0 \vee \Box p_1) \rightarrow (\Box p_0 \vee \Box p_1).$$

Proof. Consider the transitive irreflexive Kripke model K:



As is easily seen

$k_0 \models_K \Box\Box\varphi \rightarrow \Box\varphi$

and hence $k_0 \models_K L_E$.

Moreover $k_0 \models_K \Box(\Box p_0 \vee \Box p_1)$,

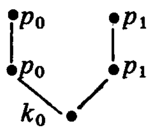
but $k_0 \not\models_K (\Box p_0 \vee \Box p_1)$. ∎

4.4. FACT

$$L \subsetneq D \subsetneq E \subsetneq S.$$

Proof

$L \subsetneq D$: Inclusion is trivial. Consider the irreflexive transitive Kripke model K:



Clearly $k_0 \models_K L$ and

$k_0 \models_K \Box(\Box p_0 \vee \Box p_1)$ but

$k_0 \not\models_K \Box\Box p_0 \vee \Box\Box p_1$.

$D \subsetneq E$: Inclusion follows by 4.2. We claim $\Box\Box\bot \to \Box\bot \notin D$. For:

$$\mathrm{PA} + \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\bot \vdash \Box_{\mathrm{PA}}(\Box_{\mathrm{PA}}A \vee \Box_{\mathrm{PA}}B) \to (\Box_{\mathrm{PA}}\Box_{\mathrm{PA}}A \vee \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}B).$$

So if $\Box\Box\bot \to \Box\bot$ were in D, it would follow that:

$$\mathrm{PA} + \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\bot \vdash \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\bot \to \Box_{\mathrm{PA}}\bot.$$

Hence $\mathrm{PA} \vdash \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\bot \to \Box_{\mathrm{PA}}\bot$. *Quod non.*

$E \subsetneq S$: Inclusion is trivial. We claim: $\Box\bot \to \bot \notin E$. For:

$$\mathrm{PA} + \Box_{\mathrm{PA}}\bot \vdash \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}A \to \Box_{\mathrm{PA}}A.$$

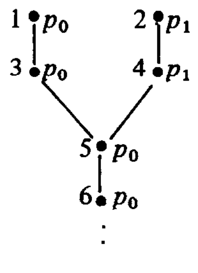
Hence if $\Box\bot \to \bot$ were in E:

$$\mathrm{PA} + \Box_{\mathrm{PA}}\bot \vdash \Box_{\mathrm{PA}}\bot \to \bot, \quad \text{so}$$

$$\mathrm{PA} \vdash \Box_{\mathrm{PA}}\bot \to \bot. \quad \textit{Quod non.}$$ ∎

4.5. FACT. $\Box\Box\Box\bot \to \Box\Box\bot \in D$.

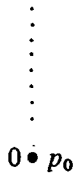
Proof. Consider the tail model K:



We have: for every n:

$$n \models_K (\Box(\Box p_0 \vee \Box p_1) \to \Box\Box p_0 \vee \Box\Box p_1) \to (\Box\Box\Box\bot \to \Box\Box\bot).$$

So by our main lemma we can find A and B such that:



$$\mathrm{PA} \vdash (\Box_{\mathrm{PA}}(\Box_{\mathrm{PA}} A \vee \Box_{\mathrm{PA}} B) \to \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} A \vee \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} B) \to (\Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\Box_{\mathrm{PA}} \bot \to \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} \bot).$$

Hence

$$\Box(\Box p_0 \vee \Box p_1) \to \Box\Box p_0 \vee \Box\Box p_1 ; \models \Box\Box\Box\bot \to \Box\Box\bot\,(\mathrm{PA}, \mathrm{PA}). \quad \blacksquare$$

4.6. FACT. $L_D \nvdash \Box\Box\Box\bot \to \Box\Box\bot$.

Proof. Any linear, finite, transitive, irreflexive Kripke model satisfies L_D. ■

4.7. FACT

$$\left(\Box \bigvee_{i=0}^{k} \Box p_i \to \bigvee_{i=0}^{k} \Box\Box p_i\right) \in D.$$

Proof. Induction on k. The cases $k = 0, 1$ are trivial. Suppose $k > 1$. For any sentences $A_1, \ldots, A_k$ we can find by 4.1 a Σ_1^0 S such that $\mathrm{PA} \vdash \bigvee_{i=1}^{k} \Box_{\mathrm{PA}} A_i \leftrightarrow \Box_{\mathrm{PA}} S$. Hence

$$\mathrm{PA} + \Box_{\mathrm{PA}}(\Box_{\mathrm{PA}} A_0 \vee \Box_{\mathrm{PA}} S) \to \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} A_0 \vee \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} S$$

$$+ \Box_{\mathrm{PA}} \bigvee_{i=1}^{k} \Box_{\mathrm{PA}} A_i \to \bigvee_{i=1}^{k} \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} A_i \vdash$$

$$\Box_{\mathrm{PA}} \bigvee_{i=0}^{k} \Box_{\mathrm{PA}} A_i \to \bigvee_{i=0}^{k} \Box_{\mathrm{PA}}\Box_{\mathrm{PA}} A_i .$$

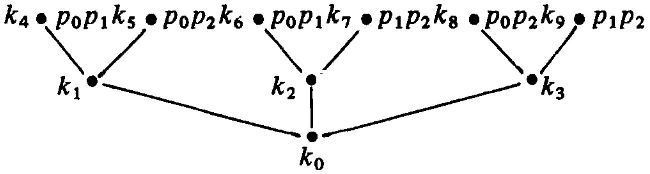
By IH we may conclude:

$$\Box(\Box p_0 \vee \Box p_1) \to \Box\Box p_0 \vee \Box\Box p_1 ; \models \Box \bigvee_{i=0}^{k} \Box p_i \to \bigvee_{i=0}^{k} \Box\Box p_i (\mathrm{PA}, \mathrm{PA}). \quad \blacksquare$$

4.8. FACT

$$L_D \nvdash \Box \bigvee_{i=0}^{2} \Box p_i \to \bigvee_{i=0}^{2} \Box\Box p_i .$$

Proof. Consider the transitive irreflexive Kripke model K:



Clearly k_0 satisfies L. Suppose $k_0 \models_K \Box(\Box\varphi \vee \Box\psi)$. Then one of $\Box\varphi$, $\Box\psi$ is true at at least two of k_1, k_2, k_3, e.g., $\Box\varphi$ is true at k_1, k_2. But in that case φ is true at k_4, k_5, k_6, k_7. Moreover k_5 and k_8, k_7 and k_9 satisfy the same formulas so φ is true at k_8, k_9. Hence $k_0 \models_K \Box\Box\varphi$.

Clearly: $k_0 \not\models_K \Box(\Box p_0 \vee \Box p_1 \vee \Box p_2) \rightarrow (\Box\Box p_0 \vee \Box\Box p_1 \vee \Box\Box p_2)$. ∎

4.9. FACT

$$\left(\Box\Box\Box \bigvee_{i=0}^{k} \Box p_i \rightarrow \Box\Box \bigvee_{i=0}^{k} \Box p_i\right) \in D.$$

Proof. We show: For any $A \in \Sigma_1^0$-sentences:

$$\mathrm{PA} + \{\Box_{\mathrm{PA}}(\Box_{\mathrm{PA}}B \vee \Box_{\mathrm{PA}}C) \rightarrow (\Box_{\mathrm{PA}}\Box_{\mathrm{PA}}B \vee \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}C) \mid B, C \in \mathscr{S}_{\mathrm{PA}}\} \vdash \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}\Box_{\mathrm{PA}}A \rightarrow \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}A.$$

Our claimed result then follows by noting that for any

$$B_0, \ldots, B_k \in \mathscr{S}_{\mathrm{PA}}: \left(\bigvee_{i=0}^{k} \Box_{\mathrm{PA}}B_i\right) \in \Sigma_1^0.$$

Step 1. For any RE theory T extending PA in the language of PA we have:

$$\Box(\Box p_0 \vee \Box p_1) \rightarrow (\Box\Box p_0 \vee \Box\Box p_1) \models \Box\Box\Box\bot \rightarrow \Box\Box\bot\ (\mathrm{PA}, T).$$

This follows by more or less copying the proof of 4.5.

Step 2. Let A be a Σ_1^0-sentence; $B_1, C_1 \in \mathscr{S}_{\mathrm{PA}}$. We have:

$$\mathrm{PA} + \{\Box_{\mathrm{PA}}(\Box_{\mathrm{PA}}B \vee \Box_{\mathrm{PA}}C) \rightarrow (\Box_{\mathrm{PA}}\Box_{\mathrm{PA}}B \vee \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}C) \mid B, C \in \mathscr{S}_{\mathrm{PA}}\} \vdash \Box_{\mathrm{PA}+\neg A}(\Box_{\mathrm{PA}+\neg A}B_1 \vee \Box_{\mathrm{PA}+\neg A}C_1) \rightarrow (\Box_{\mathrm{PA}+\neg A}\Box_{\mathrm{PA}+\neg A}B_1 \vee \Box_{\mathrm{PA}+\neg A}\Box_{\mathrm{PA}+\neg A}C_1).$$

Reason in

$\mathrm{PA} + \{\Box_{\mathrm{PA}}(\Box_{\mathrm{PA}}B \vee \Box_{\mathrm{PA}}C) \rightarrow (\Box_{\mathrm{PA}}\Box_{\mathrm{PA}}B \vee \Box_{\mathrm{PA}}\Box_{\mathrm{PA}}C) \mid B, C \in \mathscr{S}_{\mathrm{PA}}\}$.

Suppose $\Box_{PA+\neg A}(\Box_{PA+\neg A}B_1 \vee \Box_{PA+\neg A}C_1)$. We have:

$\Box_{PA}(A \vee \Box_{PA}(\neg A \to B_1) \vee \Box_{PA}(\neg A \to C_1))$. Because $A \in \Sigma_1^0$ we have $\Box_{PA}(A \to \Box_{PA}A)$, so $\Box_{PA}(A \to \Box_{PA}(\neg A \to B_1))$ and $\Box_{PA}(A \to \Box_{PA}(\neg A \to C_1))$.

Hence we find:

$$\Box_{PA}(\Box_{PA}(\neg A \to B_1) \vee \Box_{PA}(\neg A \to C_1)).$$

So:

$$\Box_{PA}\Box_{PA}(\neg A \to B_1) \vee \Box_{PA}\Box_{PA}(\neg A \to C_1).$$

Conclude:

$$\Box_{PA+\neg A}\Box_{PA+\neg A}B_1 \vee \Box_{PA+\neg A}\Box_{PA+\neg A}C_1.$$

Step 3. For $k > 0$: $PA \vdash \Box_{PA}^k A \leftrightarrow \Box_{PA+\neg A}^k \bot$, where A is a Σ_1^0-sentence. By induction on k. The induction step is:

$$\begin{aligned} PA \vdash \Box_{PA}^k A &\leftrightarrow \Box_{PA}\Box_{PA+\neg A}^{k-1}\bot \\ &\leftrightarrow \Box_{PA}(A \vee \Box_{PA+\neg A}^{k-1}\bot) \\ &\leftrightarrow \Box_{PA+\neg A}\Box_{PA+\neg A}^{k-1}\bot. \end{aligned}$$

The middle step is because: $PA \vdash \Box_{PA}(A \to \Box_{PA+\neg A}^{k-1}\bot)$.

Step 4. Take T of Step 1: $PA + \neg A$, where A is a Σ_1^0-sentence. We have:

$PA + \{\Box_{PA}(\Box_{PA}B \vee \Box_{PA}C) \to (\Box_{PA}\Box_{PA}B \vee \Box_{PA}\Box_{PA}C) \mid B, C \in \mathscr{S}_{PA}\}$

$\vdash PA + \{\Box_{PA+\neg A}(\Box_{PA+\neg A}B \vee \Box_{PA+\neg A}C) \to$

$(\Box_{PA+\neg A}\Box_{PA+\neg A}B \vee \Box_{PA+\neg A}\Box_{PA+\neg A}C) \mid B, C \in \mathscr{S}_{PA}\}$

$\vdash \Box_{PA+\neg A}\Box_{PA+\neg A}\Box_{PA+\neg A}\bot \to \Box_{PA+\neg A}\Box_{PA+\neg A}\bot$

$\vdash \Box_{PA}\Box_{PA}\Box_{PA}A \to \Box_{PA}\Box_{PA}A.$ ∎

4.9 is not surprising for those familiar with Friedman's proof that RE theories extending Heyting's Arithmetic satisfying the Disjunction Property also satisfy the Existence Property or with applications of Shepherdson's Fixed Points. Our argument in 4.5 is closely related to uses of Friedman's and Shepherdson's Fixed Points. For a survey of various applications of Fixed Points see [6].

REFERENCES

[1] S. N. Artyomov, 'Arithmetically complete modal theories', (Russian), *Semiotics and Information Science*, No. 14 (Russian) (1980), pp. 115–133, Akad. Nauk SSSR, Vsesojuz. Inst. Naučn. i. Tehn. Informacii, Moscow.
[2] G. Boolos, *The Unprovability of Consistency*, Cambridge University Press (1979).
[3] G. Boolos, 'Extremely undecidable sentences', *JSL* 47 (1982), 191–196.
[4] F. Montagna, 'On the diagonabizable algebra of Peano Arithmetic', *Bollettino Unione Matematica Italiana* (5) 16B (1979), 795–812.
[5] F. Montagna, 'Relatively precomplete numerations and arithmetic', *Journal of Philosophical Logic* 11, No. 4 (1982) 419–430.
[6] C. Smoryński, 'Calculating self-referential statements', *Fund. Math.* 109 (1980), 189–210.
[7] R. M. Solovay, 'Provability interpretations of modal logic', *Israel J. Math.* 25 (1976), 287–304.

Department of Philosophy,
University of Utrecht,
Utrecht,
The Netherlands.