Institute for Language, Logic and Information

ON THE PROOF OF SOLOVAY'S THEOREM

Dick de Jongh
Marc Jumelet
Franco Montagna

ITLI Prepublication Series
for Mathematical Logic and Foundations ML-89-04

University of Amsterdam

Instituut voor Taal, Logica en Informatie
Institute for Language, Logic and Information

Faculteit der Wiskunde en Informatica
(Department of Mathematics and Computer Science)
Plantage Muidergracht 24
1018TV Amsterdam

Faculteit der Wijsbegeerte
(Department of Philosophy)
Nieuwe Doelenstraat 15
1012CP Amsterdam

ON THE PROOF OF SOLOVAY'S THEOREM

Dick de Jongh
Department of Mathematics and Computer Science, University of Amsterdam

Marc Jumelet
Franco Montagna
Dipartmento di Matematica, Universitá di Siena

Received July 1989

The ITLI Prepublication Series

1986

86-01	The Institute of Language, Logic and Information
86-02 Peter van Emde Boas	A Semantical Model for Integration and Modularization of Rules
86-03 Johan van Benthem	Categorial Grammar and Lambda Calculus
86-04 Reinhard Muskens	A Relational Formulation of the Theory of Types
86-05 Kenneth A. Bowen, Dick de Jongh	Some Complete Logics for Branched Time, Part I Well-founded Time, Forward looking Operators
86-06 Johan van Benthem	Logical Syntax

1987

87-01 Jeroen Groenendijk, Martin Stokhof	Type shifting Rules and the Semantics of Interrogatives
87-02 Renate Bartsch	Frame Representations and Discourse Representations
87-03 Jan Willem Klop, Roel de Vrijer	Unique Normal Forms for Lambda Calculus with Surjective Pairing
87-04 Johan van Benthem	Polyadic quantifiers
87-05 Víctor Sánchez Valencia	Traditional Logicians and de Morgan's Example
87-06 Eleonore Oversteegen	Temporal Adverbials in the Two Track Theory of Time
87-07 Johan van Benthem	Categorial Grammar and Type Theory
87-08 Renate Bartsch	The Construction of Properties under Perspectives
87-09 Herman Hendriks	Type Change in Semantics: The Scope of Quantification and Coordination

1988

Logic, Semantics and Philosophy of Language:

LP-88-01 Michiel van Lambalgen	Algorithmic Information Theory
LP-88-02 Yde Venema	Expressiveness and Completeness of an Interval Tense Logic
LP-88-03	Year Report 1987
LP-88-04 Reinhard Muskens	Going partial in Montague Grammar
LP-88-05 Johan van Benthem	Logical Constants across Varying Types
LP-88-06 Johan van Benthem	Semantic Parallels in Natural Language and Computation
LP-88-07 Renate Bartsch	Tenses, Aspects, and their Scopes in Discourse
LP-88-08 Jeroen Groenendijk, Martin Stokhof	Context and Information in Dynamic Semantics
LP-88-09 Theo M.V. Janssen	A mathematical model for the CAT framework of Eurotra
LP-88-10 Anneke Kleppe	A Blissymbolics Translation Program

Mathematical Logic and Foundations:

ML-88-01 Jaap van Oosten	Lifschitz' Realizabiility
ML-88-02 M.D.G. Swaen	The Arithmetical Fragment of Martin Löf's Type Theories with weak Σ-elimination
ML-88-03 Dick de Jongh, Frank Veltman	Provability Logics for Relative Interpretability
ML-88-04 A.S. Troelstra	On the Early History of Intuitionistic Logic
ML-88-05 A.S. Troelstra	Remarks on Intuitionism and the Philosophy of Mathematics

Computation and Complexity Theory:

CT-88-01 Ming Li, Paul M.B.Vitanyi	Two Decades of Applied Kolmogorov Complexity
CT-88-02 Michiel H.M. Smid	General Lower Bounds for the Partitioning of Range Trees
CT-88-03 Michiel H.M. Smid, Mark H. Overmars Leen Torenvliet, Peter van Emde Boas	Maintaining Multiple Representations of Dynamic Data Structures
CT-88-04 Dick de Jongh, Lex Hendriks Gerard R. Renardel de Lavalette	Computations in Fragments of Intuitionistic Propositional Logic
CT-88-05 Peter van Emde Boas	Machine Models and Simulations (revised version)
CT-88-06 Michiel H.M. Smid	A Data Structure for the Union-find Problem having good Single-Operation Complexity
CT-88-07 Johan van Benthem	Time, Logic and Computation
CT-88-08 Michiel H.M. Smid, Mark H. Overmars Leen Torenvliet, Peter van Emde Boas	Multiple Representations of Dynamic Data Structures
CT-88-09 Theo M.V. Janssen	Towards a Universal Parsing Algorithm for Functional Grammar
CT-88-10 Edith Spaan, Leen Torenvliet Peter van Emde Boas	Nondeterminism, Fairness and a Fundamental Analogy
CT-88-11 Sieger van Denneheuvel Peter van Emde Boas	Towards implementing RL

Other prepublications:

X-88-01 Marc Jumelet	On Solovay's Completeness Theorem

1989

Logic, Semantics and Philosophy of Language:

LP-89-01 Johan van Benthem	The Fine-Structure of Categorial Semantics

Mathematical Logic and Foundations:

ML-89-01 Dick de Jongh, Albert Visser	Explicit Fixed Points for Interpretability Logic
ML-89-02 Roel de Vrijer	Extending the Lambda Calculus with Surjective Pairing is conservative
ML-89-03 Dick de Jongh, Franco Montagna	Rosser Orderings and Free Variables
ML-89-04 Dick de Jongh, Marc Jumelet Franco Montagna	On the Proof of Solovay's Theorem

Computation and Complexity Theory:

CT-89-01 Michiel H.M. Smid	Dynamic Deferred Data Structures
CT-89-02 Peter van Emde Boas	Machine Models and Simulations
CT-89-03 Ming Li, Herman Neuféglise Leen Torenvliet, Peter van Emde Boas	On Space efficient Solutions
CT-89-04 Harry Buhrman, Leen Torenvliet	A Comparison of Reductions on Nondeterministic Space
CT-89-05 Pieter H. Hartel, Michiel H.M. Smid Leen Torenvliet, Willem G. Vree	A Parallel Functional Implementation of Range Queries
CT-89-06 H.W. Lenstra, Jr.	Finding isomorphisms between finite fields

Other prepublications:

X-89-01 Marianne Kalsbeek	An Orey Sentence for Predicative Arithmetic
X-89-02 G. Wagemakers	New Foundations. a Survey of Quine's Set Theory
X-89-03 A.S. Troelstra	Index of the Heyting Nachlass

Introduction.

Solovay's arithmetical completeness theorem states that Löb's logic L (PRL in Smoryński[85]) is the modal logic of provability in PA and that the closure of L under reflection, $\Box A \rightarrow A$, and modus ponens is the provability logic of PA in the standard model. For any sentence φ such that $L \nvdash \varphi$, Solovay defines, using the formalized recursion theorem, a recursive function from which an interpretation $(\ )^*$ is obtained such that $PA \nvdash \varphi^*$. The proofs of the essential properties of this function, as well as the formalization of the recursion theorem, employ, prima facie, Σ_1-induction.

In this article[1] we take another look at Solovay's proof of his completeness theorem for the modal logic L with respect to arithmetical interpretations. An at first sight dominant feature in Solovay's proof is his use of the formalized recursion theorem. The use of the recursion theorem in this proof and others like it is not really necessary, but can be replaced by applications of Gödel's diagonalization lemma (mostly in the form including free variables). Using the recursion theorem makes his procedure somewhat easier to follow intuitively, but it adds to the mystery of the proof, and makes it harder to judge exactly which principles are used. Since one of our purposes is to investigate in how far one can weaken the arithmetical system and still have Solovay's completeness result, it is important to us to do without it. The concrete additional benefits of the proof of the arithmetic completeness of L given in section 2 are:

(1) it mainly uses modal properties of arithmetic as well as self-reference and is, therefore, closer to the spirit of modal logic;

(2) the modal properties used, i.e. these of Guaspari-Solovay's R plus diagonalization are valid in weak fragments of PA; they hold, for instance in any extension of $I\Delta_0$ which proves Σ_1-completeness, so they hold e.g. in $I\Delta_0 + EXP$, but not in $I\Delta_0 + \Omega_1$ (cf. Ver-

[1] Part of this article is a reworked version of the first chapter of the master's thesis of the second author (Jumelet [88])

brugge [88]). Consequently, the present proof allows us to extend Solovay's theorem to a large class of fragments of PA. The result concerning the provability logic of true formulae for these fragments falls under the scope of this proof as well.
The fixed point formulas used in the completeness proof for L are then, in section 3 of this paper, slightly modified to obtain a Δ_0-formula describing the behaviour of Solovay's function. This formula is used to introduce, by means of the diagonalization lemma again, standard proof predicates provably equivalent to the usual one, yielding the arithmetical completeness of Guaspari and Solovay's system R with respect to extensions of $I\Delta_0 + EXP$.

1. Preliminaries.

1.1. Definition. The language $L_\Box$ of propositional modal logic is defined as follows:
$L_\Box := \{\bot, \rightarrow,), (, \Box\} \cup P$, where P is some set of propositional letters, $\bot$ a propositional constant (*falsum*), $\rightarrow$ a binary connective (*material implication*) and $\Box$ a modal operator. The class of well-formed formulae $SEN_{L_\Box}$ of $L_\Box$ is the smallest class such that:

$P \subseteq SEN_{L_\Box}$,
$\bot \in SEN_{L_\Box}$,
$\varphi, \psi \in SEN_{L_\Box} \Rightarrow (\varphi \rightarrow \psi) \in SEN_{L_\Box}$,
and $\varphi \in SEN_{L_\Box} \Rightarrow \Box\varphi \in SEN_{L_\Box}$.

Boolean connectives $\vee$, $\wedge$, $\neg$, $\leftrightarrow$, as well as $\Diamond$ will be used as abbreviations with their standard meaning.

1.2. Definition. A semantics for modal formulae is developed by means of so-called *Kripke-models*. A *model* M *for* $L_\Box$ is a triple $\langle M, R, \Vdash \rangle$, where M is a non-empty set, R a binary relation on M and $\Vdash$ some subset of $M \times P$. $F = \langle M, R \rangle$ is called the *frame* of the model. The forcing relation is uniquely extended to all modal formulae χ in the following manner (writing $x \Vdash \chi$ for $\langle x, \chi \rangle \in \Vdash$ and $x \nVdash \chi$ for $\langle x, \chi \rangle \notin \Vdash$):

for all $x \in M$:
for $\chi = \varphi \rightarrow \psi$: $x \Vdash \chi$ iff $x \nVdash \varphi$ or $x \Vdash \psi$,
for $\chi = \Box\varphi$: $x \Vdash \chi$ iff for all $y \in M$ such that xRy: $y \Vdash \varphi$,
and, finally, $x \nVdash \bot$.

1.3. Definition. The modal system that primarily concerns us here, is the so-called modal *provability logic* L. This system is defined as the smallest set of modal formulae containing:

all tautologies of propositional logic;
all expressions of the forms
$\Box\varphi \rightarrow \Box\Box\varphi$, $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$, or $\Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi$,
which is closed under the following two rules of inference:
$\vdash \varphi \Rightarrow \vdash \Box\varphi$ (*necessitation*);
$\vdash \varphi \rightarrow \psi$ and $\vdash \varphi \Rightarrow \vdash \psi$.

The axiom $\Box\varphi \rightarrow \Box\Box\varphi$ is put on the list rather to stress its importance than its indispensability, since it can actually be derived from the other axioms and rules. The next result is of essential interest to us here.

1.4. Theorem. φ is not a theorem of L if and only if a model $\mathrm{M} := \langle M, R, \Vdash \rangle$ exists such that:

(i) M is finite, say $M = \{1, \ldots, n\}$;
(ii) R is a transitive and conversely well-founded relation on M, (i.e.: $\forall x,y,z \in M (xRy \wedge yRz \rightarrow xRz)$ and no infinite ascending chain $x_0 R x_1 R x_2 \ldots$ of elements of M exists);
(iii) for all $j \in M$, if $1 < j \leqslant n$, then $1Rj$;
(iv) $1 \Vdash \neg\varphi$.

This theorem is known as the *modal completeness theorem for* L with respect to the finite, transitive and conversely well-founded frames. For its proof one may consult e.g. Smoryński[85].

1.5. Interpretations. Let from now on T be a Σ_1-sound arithmetical theory proving the three Löb conditions (and hence Löb's theorem) and satisfying formalized Σ_1-completeness, i.e.:

$T \vdash \exists p\, \mathrm{proof}_T(p, \ulcorner A \rightarrow B \urcorner) \rightarrow (\exists p\, \mathrm{proof}_T(p, \ulcorner A \urcorner) \rightarrow \exists q\, \mathrm{proof}_T(q, \ulcorner B \urcorner))$;
$T \vdash A \rightarrow \exists p\, \mathrm{proof}_T(p, \ulcorner A \urcorner)$, for all $A \in \Sigma_1$;

An *interpretation* of a set of modal formulae is a function $(\ )^*$ that assigns a sentence φ^* in the language of T to each modal expression φ and obeys the following criteria:

$(\bot)^* = 0=1$;
$(\varphi\to\psi)^* = \varphi^*\to\psi^*$;
$(\Box\varphi)^* = \exists p\, \mathrm{proof}_T(p, \ulcorner\varphi^*\urcorner)$.

It is obvious that, once $(\ )^*$ has been defined for each propositional variable in the modal language used, the translation of the entire set of formulae is completely determined.

1.6. Solovay's first Completeness Theorem (Solovay[76]). This theorem is formulated as follows:

> Let φ be any modal expression. Then: $\vdash_L\varphi$ if and only if $T\vdash\varphi^*$ for every interpretation $(\ )^*$ of the modal language used which satisfies the clauses of the preceding paragraph.

The implication from the left to the right is of no concern to us here. The proof is simple, due to the fact that T is closed under the axioms and rules of L whenever the provability predicate is substituted for the modal operator $\Box$. The arithmetical versions of the rules and axioms of L are exactly the three Löb conditions and Löb's theorem which are fulfilled in T. The conditions imposed upon the interpretation function will do the rest. The implication in the other direction will be treated in section 2.

2. A modification of Solovay's completeness proof.

The original proof of the completeness theorem is based on the idea that a certain class of Kripke-models can be embedded in arithmetic. We have already seen that any modal expression φ which is not derivable from the axioms of L gives rise to some finite counter model falsifying φ. The embedding of such a model into arithmetic was carried out by Solovay by defining, with the aid of the recursion theorem, a recursive function h which paces through the model in a very particular way. Intuitively speaking, one can describe the Solovay function as follows. As its values it

takes only numbers denoting the nodes of the Kripke-model in question. The next value can only be the same as the previous one or one which is accessible from it by way of the relation R of the model. Thus it is clear that this function eventually reaches a limit. This limit is used to specify the next value, each time, in the following manner: for each argument the function takes the same value m as the previous one, unless the argument codes a proof in T of the fact that, for a certain number n, R-accessible from m, the limit of the function is not equal to n. In the latter case the function takes this value n.
To be able to be more precise we now first give some notation.

2.1. Definition. Let $F=\langle M,R\rangle$ be a finite, transitive and conversely well-founded frame. $M=\{1,\ldots,n\}$ and for all j, if $1<j\leqslant n$, then $1Rj$. A new root 0 is added to M, i.e., for all $j\in M$, $0Rj$.
We will use the following abbreviations:

$i\underline{R}j$ for $i=j\vee iRj$;

$i\circ j$ for $\neg i\underline{R}j\wedge\neg j\underline{R}i$.

The function h is represented by a formula Hxy. We write $\ell=i$ for $\exists x\forall y\geqslant xHyi$, i.e." the limit of h is i".

More formally, the function h, given by the formula Hxy, is defined as follows, using the formalized recursion theorem:

$h(0)=0$

$h(n+1)=h(n)$ unless

$h(n)Rm$ and $\mathrm{proof}(n,\ulcorner\neg\ell=m\urcorner)$ in which case

$h(n+1)=m$.

If the theory T is strong enough to allow definition by primitive recursion, the use of the recursion theorem can immediately be circumvented as follows. Let nonlim(u,v) be the function that, for each u and u, if u is the code of a formula Hxy, gives the code of $\neg\exists x\forall y\geqslant xHyv$. One can then define h'(u,n), dependent on the extra variable u, simply by primitive recursion:

$h'(u,0)=0$

$h'(u,n+1)=h(u,n)$ unless

$h(u,n)Rm$ and $\mathrm{proof}(n,\mathrm{nonlim}(u,m))$ in which case

$h'(u,n+1)=m$.

If h'(u, x) is defined by H'uxy, then Hxy with properties as required can be found with the aid of the diagonalization lemma:

$\vdash Hxy \leftrightarrow H'(\ulcorner Hxy \urcorner, x, y)$

However, we do not want to have to depend on our theory to be strong enough to have primitive recursion available: in essence this still requires Σ_1-induction and it turns out that with definitions like the one given above there is no necessity for this. For the definition of h(n+1) we only have to look at numbers $\leqslant n$ and the proofs of negations of limit assertions about h which they code.
Let us first consider the case of defining h only as a partial function at those arguments where relevant negations of limit assertions are actually proved. Then we can see that h(n+1)=m iff
(1) n+1 proves the negation of the limit assertion with respect to m,
(2) no such proof concerning a number m' with m R m' (or m=m') is coded by a number $\leqslant n$ [otherwise, h should have "passed" m already],
(3) if any such proof is coded by a number $n' \leqslant n$ for an m' incomparable to m with respect to R, then there has to be an even smaller number $n'' \leqslant n'$ that codes such a proof for a number m" R m (or m"=m) incomparable to m' [otherwise h should have taken a direction from which it could no longer reach m; in other words, any proof that could possibly "side-track" h from its way to m, has to have been preceded by a proof that makes it harmless, by side-tracking it].
More formally a partial function can be thus defined as H_p, slightly changing the definition of $\ell = y$ to
$\exists x (H_p xy \wedge \forall x' \geqslant x \neg \exists y' \leqslant n\, H_p x'y')$:

$$
\begin{aligned}
H_p xy \leftrightarrow\ & (x=0 \wedge y=0) \vee \\
& (\mathrm{Proof}(x, \ulcorner \neg \ell = \dot{y} \urcorner) \wedge \\
& \qquad \neg \exists x'' < x \exists y'' \leqslant n (y \underline{R} y'' \wedge \mathrm{Proof}(x'', \ulcorner \neg \ell = \dot{y}'' \urcorner) \wedge \\
& \forall x'' < x \forall y'' \leqslant n (y'' \circ y \wedge \mathrm{Proof}(x'', \ulcorner \neg \ell = \dot{y}'' \urcorner) \rightarrow \\
& \qquad \exists x''' < x'' \exists y''' (y''' \circ y'' \wedge y''' \underline{R}\, y \wedge \mathrm{Proof}(x''', \ulcorner \neg \ell = \dot{y}''' \urcorner)
\end{aligned}
$$

Hxy can then be obtained from $H_p xy$ as follows:

$Hxy \leftrightarrow \exists x' \leqslant x(H_p x'y \wedge \forall x''(x' < x'' \leqslant x \rightarrow \neg\exists y' \leqslant n H_p x''y'))$

This method of giving these definitions applies quite generally, and we will use it in section 3, but for the Solovay proof for L it can be further simplified. The proof involves only the mutual relations between a finite number of limit assertions, and we can more directly define corresponding sentences, using nothing but the desired connection between these sentences. More precisely, we may replace each expression "$\ell = i$" we come across in the original proof, by a single sentence λ_i, the definition of which is an exact imitation of the conditions which lead to $\ell = i$. It is important to notice that these conditions can all be spelled out in the form of finite conjunctions, claiming the existence or non-existence and order of succession of certain proofs, namely proofs of expressions of the form $\neg \ell = j$. But within proof predicates only codes of these expressions occur. It turns out to be possible for that reason to define each λ_i by means of a fixed point equation, containing only codes of these λ_j's. It will be demonstrated below, that, in doing so, the alternative sentences satisfy the same lemmas Solovay proved for the original ones. This makes them equally suitable to perform as a basis for arithmetical interpretations of the modal logic.

The n-ary fixed point theorem produces a set of sentences $\lambda_0,\ldots,\lambda_n$ in the language of T, which satisfy the following requirements:

$$T \vdash \lambda_1 \leftrightarrow \Box\neg\lambda_1 \wedge \bigwedge_{1Ri} \neg\Box\neg\lambda_j;$$

for all i such that $1 < i \leqslant n$:

$$T \vdash \lambda_i \leftrightarrow \Box\neg\lambda_i \wedge \bigwedge_{iRj} \neg\Box\neg\lambda_j \wedge \bigwedge_{i^\circ j} \bigvee_{\substack{kRi \\ k^\circ j}} (\Box\neg\lambda_k \prec \Box\neg\lambda_j).$$

Here "$\Box A \prec \Box B$" is the usual notation for:

$$\text{"}\exists p[\mathrm{proof}_T(p,\ulcorner A \urcorner) \wedge \neg\exists q \leqslant p\, \mathrm{proof}_T(q,\ulcorner B \urcorner)]\text{"}.$$

Finally, we define:

$$\lambda_0 := \neg \bigvee_{1 \leqslant i \leqslant n} \lambda_i .$$

2.2. Lemma. The set of sentences $\{\lambda_0, \ldots, \lambda_n\}$ of T defined as above has the following properties:

(1) $T \vdash \bigvee\!\!\!\bigvee_{0 \leqslant i \leqslant n} \lambda_i$.

(2) $\mathbb{N} \models \lambda_0$.

(3) For all i such that $0 \leqslant i \leqslant n$, $T + \lambda_i$ is consistent.

(4) $T \vdash \lambda_i \rightarrow \bigwedge\!\!\!\bigwedge_{iRj} \neg\Box\neg\lambda_j$ for all $i \geqslant 0$.

(5) $T \vdash \lambda_i \rightarrow \bigwedge\!\!\!\bigwedge_{\neg iRj} \Box\neg\lambda_j$ for all $i > 0$.

This lemma represents the heart of Solovay's proof. If we replace each expression of the form λ_i by $\ell = i$, we get the original lemma (cf. Solovay[76], lemma 4.1).

For reasons of economy, it is useful to prove lemma 2.2 within a more general framework. This will show us exactly which properties of our theory are used to prove it. We take for this purpose a modified version of R^-, the modal system of Guaspari and Solovay (cf. Guaspari and Solovay[79]). We first recall that R^- is an extension of L in which the class of well-formed formulae is extended by the so-called witness comparison formulae, viz. those of the forms $\Box A \prec \Box B$ and $\Box A \preccurlyeq \Box B$.

2.3. Axioms of R^-. R^- is axiomatized by adding to L the following axiom schemata (cf. de Jongh[87]):

$A \rightarrow \Box A$ for all boxed and witness comparison formulae. It is to be noted, that, since R^- is an extension of L, the same schema applies to the closure of this class under conjunctions and disjunctions, the so-called Σ-formulae; this gives us the so-called Σ-*completeness* axiom;

the *order axioms* (for all $\Box$-formulae A, B, C):

(O1) $A \rightarrow A \preccurlyeq B \vee B \preccurlyeq A$;
(O2) $A \preccurlyeq B \rightarrow A$;
(O3) $A \preccurlyeq B \wedge B \preccurlyeq C \rightarrow A \preccurlyeq C$;
(O4) $A \prec B \leftrightarrow A \preccurlyeq B \wedge \neg B \preccurlyeq A$.

We extend R^- as follows: for any frame $F=\langle M,R\rangle$, which is finite, transitive and conversely well-founded, with $M=\{1,\dots,n\}$ and $1Ri$ for all i such that $1<i\leqslant n$, let R_F^- be defined by the addition of the following axioms to R^- (we assume the language to contain propositional constants $L_0,\dots,L_n$, and we write $\boxdot A$ for $A\wedge\Box A$):

$$\boxdot(L_1\leftrightarrow\Box\neg L_1\wedge\bigwedge_{1Ri}\neg\Box\neg L_i);$$

for each i such that $1<i\leqslant n$:

$$\boxdot(L_i\leftrightarrow\Box\neg L_i\wedge\bigwedge_{iRj}\neg\Box\neg L_j\wedge\bigwedge_{i\circ j}\bigvee_{\substack{kRi\\ k\circ j}}(\Box\neg L_k\prec\Box\neg L_j));$$

$$\boxdot(L_0\leftrightarrow\neg\bigvee_{1\leqslant i\leqslant n}L_i).$$

These axioms will be referred to as the *limit axioms*. In addition, we let R_F^- contain

$$\boxdot(\neg(\Box\neg L_i\preccurlyeq\Box\neg L_j\wedge\Box\neg L_j\preccurlyeq\Box\neg L_i))$$

for all i,j such that $0\leqslant i,j\leqslant n$ and $i\neq j$, as so-called *proof apartness axioms*. In the next two paragraphs we will mention some properties of R_F^- that will be needed for the proof of lemma 2.2.

In the following discussion the frame F is to be thought of as fixed.

2.4. Theorem (*Soundness* of R_F^-). An interpretation $(\ )^+$ of sentences in the language of R_F^- into the language of arithmetic is called F-*sound* if and only if $(\ )^+$ fulfils the criteria cited for $(\ )^*$ in 1.5 and, in addition:

for all formulae φ, ψ:
$(\Box\varphi\preccurlyeq\Box\psi)^+=\exists p\,[\mathrm{proof}_T(p,\ulcorner\varphi^+\urcorner)\wedge\neg\exists q<p\ \mathrm{proof}_T(q,\ulcorner\psi^+\urcorner)]$;
$(\Box\varphi\prec\Box\psi)^+=\exists p\,[\mathrm{proof}_T(p,\ulcorner\varphi^+\urcorner)\wedge\neg\exists q\leqslant p\ \mathrm{proof}_T(q,\ulcorner\psi^+\urcorner)]$;
for all i such that $0\leqslant i\leqslant n$:
$L_i^+=\lambda_i$ (as defined above).

For all F-sound interpretations $(\ )^+$ of sentences in the language of R_F^- and any φ in that language, $R_F^-\vdash\varphi\Rightarrow T\vdash\varphi^+$.

The proof is straightforward by induction on the length of proof in R_F^-, since T is closed under the same rules and axioms we have at our disposal in R_F^-, provided $(\ )^+$ is F-sound. We will use this theorem extensively in the proof of lemma 2.2.

A Kripke-model for R^- is a finite, tree-ordered Kripke-model $\langle X, U, \Vdash \rangle$ for L in which witness-comparison formulae are treated as if they were atomic formulae and in which the following requirements are fulfilled:

(1) *persistency* of $\prec$ and $\preccurlyeq$:

if $s \Vdash A \preccurlyeq B$ and $s U s'$, then $s' \Vdash A \preccurlyeq B$,

and likewise for $\prec$, viz.:

if $s \Vdash A \preccurlyeq B$ and $s U s'$, then $s' \Vdash A \prec B$;

(2) each instance of the order axioms is satisfied at each node.

The completeness theorem for R^- is stated as follows: $R^- \vdash \varphi$ iff φ is valid on all finite, tree-ordered Kripke-models for R^-.
In the case of R^-_F this theorem implies:

2.5. Theorem (*completeness* of R^-_F).
If $R^-_F \nvdash \varphi$, then a finite, tree-ordered Kripke-model for R^- exists, in which all limit axioms and proof apartness axioms are forced at each node, and on which φ is falsified.
Proof. This result is a consequence of the completeness theorem for R^-, because we have :

$R^-_F \vdash \varphi \Leftrightarrow R^- \vdash \theta \rightarrow \varphi$,

where θ is the finite conjunction of limit axioms and proof apartness axioms listed in the definition of R^-_F.
The implication from the right to the left is easily proved. The other direction is shown by induction on the length of the proof in R^-_F. To obtain the desired result, we should check whether any proof of a formula φ in R^-_F can be transformed into a proof of $\theta \rightarrow \varphi$ in R^-. This can cause no difficulty, since any axiom of R^-_F is either an axiom of R^- or a consequence of θ, and, if the last rule applied in a proof in R^-_F of some formula ψ has been the necessitation rule, then we can use $\theta \rightarrow \Box \theta$ which is a theorem of R^-. ⊠

A simple proof of the completeness theorem for R^- can be found in De Jongh [87].

Now we are ready to commence the proof of lemma 2.2.

Proof of lemma 2.2.

Fix a finite, transitive and conversely well-founded frame $F=\langle M,R\rangle$, with $M=\{1,\dots n)$ and $1 R i$ for all i such that $1 < i \leqslant n$. Let $\lambda_0,\dots,\lambda_n$ and R_F^- be as defined above. We first show:

(a) $R_F^- \vdash L_0 \leftrightarrow \bigwedge\limits_{1 \leqslant i \leqslant n} \neg\Box\neg L_i$.

As the implication from the right to the left is obvious, we will concentrate on the opposite direction. Suppose the contrary to be the case. We will derive a contradiction as follows. By theorem 2.5 we would have a finite, tree-ordered Kripke-model $\langle X, U, \Vdash\rangle$ for R^- with the limit and proof apartness axioms forced everywhere in the model and with some bottom node k_0 such that

$$k_0 \Vdash L_0 \wedge \bigvee\limits_{1 \leqslant i \leqslant n} \Box\neg L_i .$$

We must have:

$$k_0 \Vdash \Box\neg L_{i_1} \wedge \dots \wedge \Box\neg L_{i_k} \wedge \bigwedge\limits_{j \notin \{i_1 \dots i_k\}} \neg\Box\neg L_j,$$

for some k such that $1 \leqslant k \leqslant n$.

As any instance of the order axioms and the proof apartness axioms is forced at k_0, we can stipulate, without loss of generality, that at k_0 the following is forced:

$$\Box\neg L_{i_1} \prec \Box\neg L_{i_2} \wedge \dots \wedge \Box\neg L_{i_{k-1}} \prec \Box\neg L_{i_k} .$$

At this point, we can construct a subset $\{m_1, \dots, m_l\}$ of the set of indices $\{1, \dots, k\}$ as follows:

$m_1 := 1$;
$m_{h+1} := m$ if m is the smallest index number in $\{1,\dots k\}$ such $i_{m_h} R\, i_m$ and $k_0 \Vdash \Box\neg L_{i_{m_h}} \prec \Box\neg L_{i_m}$. If no such m exists, set $l=h$ and $m_{h+1} = m_h$.

It will be understood that this construction comes to an end, because the set $\{1, \dots, k\}$ is finite. By means of a finite induction procedure we will now prove the following: for all p such that $1 \leqslant p \leqslant l$:

$$k_0 \Vdash \bigwedge\limits_{j \circ m_p} \bigvee\limits_{\substack{k \circ j \\ k \underline{R} m_p}} (\Box\neg L_k \prec \Box\neg L_j).$$

The case of $p=1$ is trivial, since $i_{m_1} = i_1$.

Induction step: suppose

$$k_0 \Vdash \bigwedge_{j \circ m_p} \bigvee_{\substack{k \circ j \\ k \underline{R} m_p}} (\Box\neg L_k \prec \Box\neg L_j)).$$

Now let j be such, that $j \circ m_{p+1}$. There are two possibilities: either $j \circ m_p$ as well, or not. In the first case we obtain

$$k_0 \Vdash \bigvee_{\substack{k \circ j \\ k \underline{R} m_{p+1}}} (\Box\neg L_k \prec \Box\neg L_j))$$

by the induction hypothesis, for $k\underline{R}m_p$ implies $k\underline{R}m_{p+1}$. In the latter case $m_p R j$ must hold. But the definition of m_{p+1} implies: $k_0 \Vdash \Box\neg L_{i_{m_{p+1}}} \prec \Box\neg L_j$ whence $k_0 \Vdash \bigvee_{\substack{k \circ j \\ k \underline{R} m_{p+1}}} (\Box\neg L_k \prec \Box\neg L_j)$ follows by propositional logic.

This completes the induction procedure. Since i_{m_1} has no R-successors in $\{i_1, \ldots, i_k\}$, we can conclude:

$$k_0 \Vdash \Box\neg L_{i_{m_1}} \wedge \bigwedge_{i_{m_1} R j} \neg\Box\neg L_j \wedge \bigwedge_{j \circ i_{m_1}} \bigvee_{\substack{k \circ j \\ i_{m_1} \underline{R} j}} (\Box\neg L_k \prec \Box\neg L_j) .$$

But this implies $k_0 \Vdash L_{i_{m_1}}$ contradicting $k_0 \Vdash L_0$. The proof is hereby completed.

(b) If $1 \leqslant i \leqslant n$, then $R_F^- \vdash L_i \rightarrow \bigwedge_{iRj} \neg\Box\neg L_j$. This is immediate from the definition of R_F^- .

Combining (a) and (b) we get lemma 2.2(4) by soundness.

(c) R_F^- contains all tautologies of propositional logic, so we have $R_F^- \vdash L_0 \vee \neg L_0$ from which $R_F^- \vdash \bigvee_{0 \leqslant i \leqslant n} L_i$ readily follows. Employing soundness, this accounts for of lemma 2.2(1).

As all theorems of T hold in the standard model, we must have $\mathbb{N} \models \lambda_i$ for some i such that $0 \leqslant i \leqslant n$. But it must be the case that $\mathbb{N} \models \lambda_0$, since for any $i \neq 0$ we would have $T \vdash \neg\lambda_i$ in case λ_i were true. Combining this with of lemma 2.2(4), we obtain $\mathbb{N} \models \bigwedge_{0 \leqslant i \leqslant n} \neg\Box\neg\lambda_j$. This settles lemma 2.2(2) and (3).

(d) If $0 < i \leqslant n$, then $R_F^- \vdash L_i \rightarrow \Box\neg L_0$.

By (a) we have $R_F^- \vdash \Box\neg L_i \rightarrow \neg L_0$. Applying the necessitation rule we infer: $R_F^- \vdash \Box\Box\neg L_i \rightarrow \Box\neg L_0$. As $\Box\neg L_i$ is a boxed formula, $\Box\neg L_i \rightarrow \Box\Box\neg L_i$ is a theorem of R_F^-. This completes the proof, as $R_F^- \vdash L_i \rightarrow \Box\neg L_i$ is a direct consequence of the definition of R_F^-.

(e) If $0 < i \leqslant n$ and iRj, then $R_F^- \vdash L_j \rightarrow \Box\neg L_i$.
If iRj is the case, we have $R_F^- \vdash \Box\neg L_j \rightarrow \neg L_i$ by the limit axiom that defines L_i. Arguing as in (d) we obtain the desired result.

(f) If $0 < i \leqslant n$ and $0 < j \leqslant n$ and $i \circ j$, then $R_F^- \vdash L_i \rightarrow \Box\neg L_j$.
Fix i and j such that $i \circ j$. By the definition of R_F^- we have:

$$R_F^- \vdash L_i \rightarrow \bigwedge_{i \circ j'} \bigvee_{\substack{kBi \\ k \circ j'}} (\Box\neg L_k \prec \Box\neg L_{j'}).$$

More specifically, we obtain:

$$R_F^- \vdash L_i \rightarrow \bigwedge_{\substack{i \circ j' \\ j'Bj}} \bigvee_{\substack{kBi \\ k \circ j'}} (\Box\neg L_k \prec \Box\neg L_{j'}).$$

As the order axioms and proof apartness axioms imply that the $\Box\neg L_k$'s in this formula are linearly ordered by $\prec$ (compare the proof of (a)), there must be a smallest one; in other words:

$$R_F^- \vdash \bigwedge_{\substack{i \circ j' \\ j'Bj}} \bigvee_{\substack{kBi \\ k \circ j'}} (\Box\neg L_k \prec \Box\neg L_{j'}) \rightarrow \bigvee_{\substack{k \circ j \\ kBi}} \bigwedge_{\substack{j'Bj \\ j' \circ i}} (\Box\neg L_k \prec \Box\neg L_{j'}).$$

But the consequent in the last formula is a Σ-expression implying $\neg L_j$, so: $R_F^- \vdash \bigwedge_{\substack{i \circ j' \\ j'Bj}} \bigvee_{\substack{kBi \\ k \circ j'}} (\Box\neg L_k \prec \Box\neg L_{j'}) \rightarrow \Box\neg L_j$.

(g) Putting (d), (e) and (f) together, we obtain:

$$R_F^- \vdash L_i \rightarrow \Box(\neg L_0 \wedge \bigwedge_{i \circ j} \neg L_j \wedge \bigwedge_{jRi} \neg L_j) \text{ for all } i \text{ such that } 0 < i \leqslant n.$$

Applying soundness, this settles lemma 2.2(5). ⊠

Let $M = \langle M, R, \Vdash \rangle$ be a finite, transitive and conversely well-founded model with $M = \{1, \ldots, n\}$ and for all i if $1 < i \leqslant n$, then $1 R i$. As usual, we expand M by adding an extra node 0 to it and defining $0 \Vdash$ as equivalent to $1 \Vdash$ for all propositional letters. In the manner indicated above we obtain sentences $\lambda_0, \ldots, \lambda_n$ satisfying lemma 2.2. We define an interpretation $(\ )^*$ by setting for all $p \in P$:

$p^* := \bigvee_{i \Vdash p} \lambda_i$. If there is no i such that $i \Vdash p$, then set:
$p^* :=$ "$0=1$"

The following lemma provides the necessary last step towards the completeness theorem:

2.7 Lemma: for all modal expressions φ, if $1 \leqslant i \leqslant n$, then:

$i \Vdash \varphi \Rightarrow T \vdash \lambda_i \rightarrow \varphi^*$ and

$i \nVdash \varphi \Rightarrow T \vdash \lambda_i \rightarrow \neg \varphi^*$.

The proof is exactly the same as Solovay's original one, with each expression of the form $\ell = i$ replaced by λ_i, so we will not give it here. Some attention however should be paid to the way clause (5) of lemma 2.2, in the form $\lambda_i \rightarrow \Box \bigvee_{iRj} \lambda_j$, is used, when $i \Vdash \varphi \Rightarrow T \vdash \lambda_i \rightarrow (\varphi)^*$ is proved by induction. In fact, it is at this point that full formalized Σ_1-completeness is used. This completes our explanation concerning the adaptation of the proof of Solovay's result.

3. Completeness of R.

In this section, we deal with the arithmetical completeness of Guaspari and Solovay's logic R with respect to arithmetical interpretations in $I\Delta_0 + EXP$ or in any given Σ_1-sound RE-extension of it. To formulate our result correctly, let us start with the following definitions (as usual, T denotes an arbitrary Σ_1-sound RE-extension of $I\Delta_0 + EXP$).

3.1. Definition. A *standard proof predicate* for T is a Σ_1-formula Th(v) numerating the set of theorems of T and such that for any two sentences α, β, $T \vdash Th(\ulcorner \alpha \urcorner) \wedge Th(\ulcorner \alpha \rightarrow \beta \urcorner) \rightarrow Th(\ulcorner \beta \urcorner)$ and $T \vdash Th(\ulcorner \alpha \urcorner) \rightarrow Th(\ulcorner Th(\ulcorner \alpha \urcorner) \urcorner)$

In our proof, we shall make use of a standard proof predicate which, in addition, is provably equivalent to the usual one.

3.2. Definition. Let Th(v) be a standard proof predicate for T. An *arithmetical interpretation based on* Th(v) is a mapping * from R formulas into arithmetical sentences satisfying the following conditions: $\bot^* \equiv 0=1$, $T^* \equiv 0=0$; * commutes with the logical connectives and witness comparisons, and $\Box A^* \equiv Th(\ulcorner A^* \urcorner)$.

We are now ready to state the main theorem of this section.

3.3. Theorem. Let A be any formula of R. The following are equivalent:

(i) $R \vdash A$.

(ii) For any standard proof predicate Th(v) and for any interpretation * based on it, $T \vdash A^*$.

(iii) For any standard proof predicate Th(v) provably equivalent to the usual one and for each interpretation based on it, $T \vdash A^*$.

Proof. That (i) implies (ii) is easy, and that (ii) implies (iii) is trivial. So, let us prove that (iii) implies (i). Suppose $R \nvdash A$. By a result of Guaspari and Solovay (cf. [79]) there is a model $M = \langle \{1,\dots,n\}, R, \Vdash \rangle$ of R^- with root 1 and a node i of M such that $i \nVdash A$; moreover, the model can be taken to be *A-sound*, i.e. we can assume that $1 \Vdash \Box B \to B$ for any subformula $\Box B$ of A. Add a new node 0, stipulate that $0 R i$ for $i = 1,\dots,n$, and give 0 the same forcing as 1 w.r.t. the subformulas of A. That this is possible is guaranteed by the fact that the model is A-sound.

Let S denote the set of $\Box$-subformulas of A, K denote the cardinality of S plus one. For $i = 0,\dots,n$ and for $\Box C, \Box D \in S$ define: $\Box C \equiv_i \Box D$ iff $i \Vdash \Box C \preccurlyeq \Box D$ and $i \Vdash \Box D \preccurlyeq \Box C$;
$\Box C \prec_i \Box D$ iff $i \Vdash \Box C \prec \Box D$. Furthermore, let $E_{i_1},\dots,E_{i_{h_i}}$ be the equivalence classes w.r.t. $\equiv_i$ enumerated according to $\prec_i$ (i.e. if $\Box C \in E_{i_j}$, $\Box D \in E_{i_h}$, and $j < h$, then $\Box C \prec_i \Box D$). Notice that, for $i = 0,\dots,n$, $h_i < K$.

We add some more notation:
$proof(v, \ulcorner p \urcorner) :\equiv proof_T(v, \ulcorner p \urcorner)$;
$\Box \ulcorner p \urcorner :\equiv \exists v\, proof(v, \ulcorner p \urcorner)$;
$\Box_{\leqslant x} \ulcorner p \urcorner :\equiv \exists v \leqslant x\, proof(v, \ulcorner p \urcorner)$;
$\Box_{\leqslant x} \ulcorner p \urcorner \preccurlyeq \Box_{\leqslant x} \ulcorner q \urcorner :\equiv \exists v \leqslant x [proof(v, \ulcorner p \urcorner) \wedge \forall u < v\, \neg proof(u, \ulcorner q \urcorner)]$;
$\Diamond_{\leqslant x} \ulcorner p \urcorner :\equiv \neg \Box_{\leqslant x} \ulcorner \neg p \urcorner$;
$\vdash \;:\equiv T\vdash$.

3.4. Definition. A formula A is *stable* iff
$\vdash \exists x (\forall y \geqslant x\, Ay \vee \forall y \geqslant x\, \neg Ay)$.

3.5. Lemma.

(1) Each Boolean combination of stable formulas in the same free variable x is a stable formula.

(2) $\Box_{\leqslant x}\ulcorner p\urcorner$, $\Diamond_{\leqslant x}\ulcorner p\urcorner$, $\Box_{\leqslant x}\ulcorner p\urcorner \preccurlyeq \Box_{\leqslant x}\ulcorner q\urcorner$ are stable.

(3) if $L(A_1(x),\ldots,A_n(x))$ is a lattice combination of stable formulas $A_1(x),\ldots,A_n(x)$, and if $L_i \equiv \exists y\forall x\geqslant y\, A_i(x)$, then:
$\vdash \exists y\forall x\geqslant y\, L(A_1(x),\ldots,A_n(x)) \leftrightarrow L(L_1,\ldots,L_n)$.

Proof. (1) and (2) are trivial, and (3) is proved by induction on the complexity of L. The step corresponding to $\wedge$ is trivial; the step corresponding to $\vee$ is proved by means of (1) and the induction hypothesis.

Next let the free variable formulas $H_i(x)$ for $1\leqslant i\leqslant n$ be defined, by self-reference, in such a way that:

$$\vdash H_i(x) \leftrightarrow \Box_{\leqslant x}\neg L_i \wedge \bigwedge_{iRj} \Diamond_{\leqslant x} L_j \wedge \bigwedge_{i\circ j}\ \bigvee_{\substack{k\underline{R}i\\ k\circ j}} (\Box_{\leqslant x}\neg L_k \prec \Box_{\leqslant x}\neg L_j);$$

where $i\underline{R}j$ and $i\circ j$ are defined as in 2.1 and $L_i :\equiv \exists y\forall x\geqslant y\, H_i(x)$.
Also, let $H_0(x) :\equiv \bigwedge_{i\neq 0} \neg H_i(x)$.

By lemma 3.2, $H_i(x)$, $i=0,\ldots,n$, are stable. Therefore, by the same lemma, clause (3):

$$\vdash L_0 \leftrightarrow \bigwedge_{i\neq 0} \neg L_i$$
$$\vdash L_i \leftrightarrow \Box\neg L_i \wedge \bigwedge_{iRj} \neg\Box\neg L_j \wedge \bigwedge_{i\circ j}\ \bigvee_{\substack{k\underline{R}i\\ k\circ j}} (\Box\neg L_k \prec \Box\neg L_j);$$

(of course we use: $\vdash\Box\ulcorner p\urcorner \leftrightarrow \exists x\Box_{\leqslant x}\ulcorner p\urcorner$, $\vdash\Diamond\ulcorner p\urcorner \leftrightarrow \forall x\Diamond_{\leqslant x}\ulcorner p\urcorner$, $\vdash\Box\ulcorner p\urcorner \preccurlyeq \Box\ulcorner q\urcorner \leftrightarrow \exists x(\Box_{\leqslant x}\ulcorner p\urcorner \preccurlyeq \Box_{\leqslant x}\ulcorner q\urcorner))$.
As in lemma 2.2, we can now deduce:

(1) $\vdash \bigvee_{0\leqslant i\leqslant n} L_i$.

(2) $\mathbb{N} \models L_0$.

(3) For all i such that $0\leqslant i\leqslant n$, $T+L_i$ is consistent.

(4) $\vdash L_i \rightarrow \bigwedge_{iRj} \neg\Box\neg L_j$ for all $i\geqslant 0$.

(5) $\vdash L_i \rightarrow \bigwedge_{\neg iRj} \Box\neg L_j$ for all $i>0$. ⊠

3.6. Lemma.

(1) If $i\neq j$, then $\vdash H(x,i) \rightarrow \neg H(x,j)$;

(2) $\vdash H(x,i) \rightarrow \bigvee_{i\underline{R}j} H(x+y,j)$.

Proof.

(1) Suppose $i \neq j$. If $i R j$, then $\vdash H(x,i) \to \Diamond_{\leqslant x} L_j$ and $\vdash H(x,i) \to \Box_{\leqslant x} \neg L_j$. The reasoning in the case $j R i$ is symmetric. If $i \circ j$, one can formalize the following argument: assume $H(x,i)$ and $H(x,j)$. Then, for each h incomparable with i, there is a k such that $k \circ h$, $k \underline{R} i$ and $\Box_{\leqslant x} \neg L_k \prec \Box_{\leqslant x} \neg L_h$. Moreover, a similar condition holds with j in place of i. Since $i \circ j$, $H(x,i)$ implies that there is an h_1 such that $h_1 \underline{R} i$, $h_1 \circ j$ and $\Box_{\leqslant x} \neg L_{h_1} \prec \Box_{\leqslant x} \neg L_j$. Using $H(x,j)$, we get, since $h_1 \circ j$, an $h_2 \underline{R} j$, $h_2 \circ h_1$ such that $\Box_{\leqslant x} \neg L_{h_2} \prec \Box_{\leqslant x} \neg L_{h_1}$. Thus, $h_1, \ldots, h_{n+1}$ are obtained such that $\Box_{\leqslant x} \neg L_{h_{n+1}} \prec \Box_{\leqslant x} \neg L_{h_n} \prec \ldots \prec \Box_{\leqslant x} \neg L_{h_1}$. The proof apartness condition implies that the h_i's are mutually distinct. This is impossible as X has cardinality n.

(2) Induction on y (notice that the formula $H(x,i) \to \bigvee_{i \underline{R} j} H(x+y,j)$ is Δ_0). Assume $H(x+y,j)$, where $i \underline{R} j$. Clearly, if $\neg \Box_{\leqslant x+y+1} \neg L_h$ for any h such that $j R h$, then $H(x+y+1,j)$ and we are done; otherwise, let h be such that $\Box_{\leqslant x+y+1} \neg L_h$ and $j R h$. Note that:

(a) $\text{proof}(x+y+1, \neg L_h) \wedge \forall u < x+y+1 \neg \text{proof}(x+y+1, \neg L_h)$, since otherwise we would have $\neg H(x+y,j)$;

(b) if hRk, then $\neg \Box_{\leqslant x+y+1} \neg L_k$, otherwise, since $\text{proof}(x+y+1, \neg L_h)$, and, consequently $\text{proof}(x+y+1, \neg L_k)$, we would get $\Box_{\leqslant x+y} \neg L_k$ and $\neg H(x+y,j)$;

(c) if $m \circ h$, then, either $j R m$, in which case $\neg \Box_{\leqslant x+y} \neg L_m$, $\neg \Box_{\leqslant x+y+1} \neg L_m$, and, since, by (a), $\Box_{\leqslant x+y+1} \neg L_h$, we can conclude $\Box_{\leqslant x+y+1} \neg L_h \prec \Box_{\leqslant x+y+1} \neg L_m$,
or $j \circ m$, in which case there exists l such that $l \circ m$, $l \underline{R} j$ (whence $l \underline{R} h$) and $\Box_{\leqslant x+y} \neg L_h \prec \Box_{\leqslant x+y} \neg L_m$ (whence $\Box_{\leqslant x+y+1} \neg L_l \prec \Box_{\leqslant x+y+1} \neg L_m$). In any case, if $m \circ h$, there exists l (possibly $l=h$) such that $l \underline{R} h$, $l \circ m$, whence $\Box_{\leqslant x+y+1} \neg L_l \prec \Box_{\leqslant x+y+1} \neg L_m$. Conclusion: $H(x+y+1,h)$. This completes our proof. ⊠

3.7. Corollary.

$\vdash H(x,i) \wedge y > x \to \bigvee_{i \underline{R} j} H(y,j)$.

We now introduce a standard proof predicate $\Box' \ulcorner p \urcorner \equiv \exists x\, \text{proof}'(x, \ulcorner p \urcorner)$, such that $T \nvdash A^*$, where $*$ is the interpretation based on $\Box'$ given by:

$$p_i^* := (i \equiv i) \wedge \bigvee\!\!\!\bigvee_{j \Vdash p_i} L_j$$

Roughly speaking, $\Box'$ proves p at stage Kx (K the cardinality of S plus one) iff proof(x,p) and, for all $\Box B \in S$, $p \neq B^*$, and proves p at stage Kx+y ($0 < y < K$) iff $\exists i \leqslant n[H(x,i) \wedge y \leqslant h_i \wedge \exists \Box B \in E_{i_y}(p = B^*)]$. So, if H(x, i) and $E_{i_y} \equiv \{\Box B_1, \ldots, \Box B_s\}$, then $\Box'$ proves $B_1^*, \ldots, B_s^*$ at stage Kx+y. Of course, the definition of $\Box'$ depends on the interpretation * which in fact is based on it. This circularity is avoided as usual by means of the diagonalization lemma. We will now present the formal definition of proof':

3.8. Definition.

Let, by self-reference, the formula proof'(x,p) be such that:

$$\begin{aligned} \vdash \text{proof}'(x,p) \leftrightarrow\ & \exists y \leqslant x(Ky = x \wedge \text{proof}(y,p) \wedge \forall \Box B \in S(\neg x = B^*)) \vee \\ & \exists i \leqslant n \exists y < x \exists z(0 < z < K \wedge x = Ky + z \wedge H(y,i) \wedge \\ & z \leqslant h_i \wedge \exists \Box B \in E_z(p \equiv B^*)) \end{aligned}$$

where * denotes the interpretation based on $\exists x$proof'(x,p) given by: $p_i^* := (i \equiv i) \wedge \bigvee\!\!\!\bigvee_{j \Vdash p_i} L_j$.

Notice that proof' is provably Δ_0. To prove Theorem 3.3, it is sufficient to show (cf. Smoryński[85] or Guaspari-Solovay[79]) the following lemmas:

3.9. Lemma. If B is a subformula of A, then for all $i \leqslant n$:

$$\vdash L_i \rightarrow (B^* \leftrightarrow i \Vdash B).$$

3.10. Lemma. If $\Box B$ is a subformula of A, then for all $i \leqslant n$:

$$\vdash L_i \rightarrow (\Box B^* \leftrightarrow i \Vdash \Box B).$$

3.11. Lemma. $\vdash \forall x(\Box x \leftrightarrow \Box' x)$.

Proof of lemma 3.9. By induction on the complexity of B; the proof works as in Guaspari-Solovay[79]. The only problem is that we have to be careful with the use of induction. But, even if we want to allow only Δ_0-induction, there can be no problem, since both H(x,y) and proof' are Δ_0. Anyway, the propositional cases and the Boolean cases are trivial.

$\Box$-case: if $\Box B$ is a subformula of A, then B^* can be proved only at a stage of the form $Kx+z$, where $0<z<K$. This happens iff $H(x,i)$ and $i \Vdash \Box B$; so: $\vdash (L_i \wedge i \Vdash \Box B) \rightarrow (\Box B)^*$.
Next, suppose $L_i \wedge i \Vdash \neg\Box B$. Lemma 3.6 and corollary 3.7 and the definition of $H(x,0)$ ensure that, provably in T, $H(x,y)$ defines the graph of a weakly monotonic function from $\mathbb{N}$ to $\{0,\dots,n\}$. So, $L_i \wedge i \Vdash \neg\Box B$ implies that, for all x:

$$H(x,j) \rightarrow j\underline{R}i$$
$$\rightarrow j \nVdash \Box B.$$

So, B^* is never proved by proof'.
Steps $\prec, \preccurlyeq$. Suppose $L_i \wedge i \Vdash \Box B \prec \Box C$; then there is a least x such that: $\exists j \leqslant n[H(x,j) \wedge j \Vdash \Box B]$ (we have applied the least number principle to the Δ_0-formula $\exists j \leqslant n[H(x,j) \wedge j \Vdash \Box B]$). Note that, by lemma 3.6, this j is unique and $j\underline{R}i$, and therefore, by Σ-persistency, $j \Vdash \Box B \prec \Box C$.
If $u<x$, then: $H(u,h) \rightarrow h \nVdash \Box C$ (otherwise, $h R j$, $h \Vdash \Box C \prec \Box B$, $j \Vdash \Box C \prec \Box B$). So, C^* is not proved by proof' at any stage $\leqslant Kx$. Since $j \Vdash \Box B \prec \Box C$, we get either $j \Vdash \neg\Box C$ or $\Box B \in E_{j_r}$, $\Box C \in E_{j_s}$ where $r<s$. It follows, that B^* is proved at stage $Kx+r$, and either C^* is proved at stage $Kx+s$, $s>r$ or $\neg\Box_{\leqslant K(x+1)}C^*$. In both cases, $(\Box B)^* \prec (\Box C)^*$.
The case $L_i \wedge i \Vdash \Box B \preccurlyeq \Box C$ is treated similarly.
If $L_i \wedge i \nVdash \Box B \prec \Box C$, then either $i \nVdash \Box B$ and $\neg(\Box B)^*$ by the $\Box$-step, or it is the case that $i \Vdash \Box C \preccurlyeq \Box B$, in which case $(\Box C \preccurlyeq \Box B)^*$, whence $\neg(\Box B \prec \Box C)^*$ follows.
The case $L_i \wedge i \nVdash \Box B \preccurlyeq \Box C$ is treated similarly.

Proof of lemma 3.10. By conditions (1),...,(5) of lemma 2.2 and by lemma 3.9, we are in a position to repeat the proof of the analogous lemma in Guaspari-Solovay[79].

Proof of lemma 3.11. Follows from lemma 3.9 and 3.10 as in Guaspari-Solovay[79].

This completes the proof of Theorem 3.3.

3.12. Remark. Σ_1-completeness is used only in the proof of Solovay's lemma, i.e; the the proof of lemma 2.2. It follows that if, for some Σ_1-sound theory $T \supseteq I\Delta_0$ we can get sentences L_i $i=0,\ldots,n$ satisfying (1),...,(5), we can embed finite R-models in T. This does not necessarily imply that we have arithmetical completeness for T, as R need not be arithmetically sound with respect to the interpretations in T.

Bibliography:

Guaspari, D. and R.M. Solovay, 'Rosser Sentences', *Annals of Mathematical Logic* 16, pp. 81-99, 1979.

de Jongh, D.H.J., 'A Simplification of a Completeness Proof of Guaspari and Solovay', *Studia Logica* 46, pp. 187-192, 1987.

Jumelet, M., On Solovay's Theorem, *Master's Thesis*, Department of Mathematics and Computer Science, University of Amsterdam, 1988.

Smoryński, C., *Self-Reference and Modal Logic*, Springer, New York, 1985.

Solovay, R.M., 'Provability interpretations of modal logic', *Israel Journal of Mathematics*, vol.25, pp. 287-304, 1976.

Verbrugge, L.C., 'Does Solovay's Completeness Theorem extend to Bounded Arithmetic?', *Master's Thesis*, Department of Mathematics and Computer Science, University of Amsterdam, 1988.