Institute for Language, Logic and Information

ON THE PROOFS OF ARITHMETICAL COMPLETENESS FOR INTERPRETABILITY LOGIC

Domenico Zambella

ITLI Prepublication Series
for Mathematical Logic and Foundations ML-91-03

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 PROOFS OF ARITHMETICAL COMPLETENESS FOR INTERPRETABILITY LOGIC

Domenico Zambella
Department of Mathematics and Computer Science
University of Amsterdam

ITLI Prepublications
for Mathematical Logic and Foundations
ISSN 0924-2090

Received February 1991

On the proofs of arithmetical completeness for interpretability logic.

§0. Introduction. Visser [Vis1] introduced the binary modal logic IL (interpretability logic) and its extensions ILM (interpretability logic with Montagna's axiom) and ILP (interpretability logic with a persistent relation in its models) to describe the interpretability logic of PA and the interpretability logic of any sufficiently strong theory T which is finitely axiomatizable and Σ_1 sound. The modal completeness of IL, ILP and ILM was provided by de Jongh and Veltman [dJV] using so called Veltman models. These are a very natural generalization of Kripke models. Visser [Vis2] obtained the arithmetical completeness for ILP and more recently, Berarducci [Ber] and Shavrukov [Sha] have shown ILM to be complete for arithmetical interpretation over PA . All these proofs of arithmetical completeness do not directly use the Veltman models. Using a bisimulation Visser [Vis2] showed ILP to be modal complete with respect to his so called Friedman models and then used these to prove arithmetical completeness. Berarducci and Shravrukov also used a bisimulation due to Visser [Vis1] showing that ILM is modal complete with respect to the so called simplified models to prove arithmetical completeness. The use of simplified models in proving arithmetical completeness for ILM adds an additional complication due to the fact that in general these cannot be taken to be finite.

Our aim is to provide simpler and more natural proofs of arithmetical completeness for ILP and ILM . For both we shall use the original Veltman models. As all proofs of arithmetical completeness known so far, ours are based on the ideas exposed in the pioneering work of Solovay [Sol] and made explicit in [dJJM].

The organization of this paper is the following: in the next section we recall to the reader the axioms of ILM and ILP and the corresponding classes of Veltman frames. We shall not give any details. We refer the reader to the literature (see e.g. [Vis1], [dJV] and [Ber]) both for details and comments as well as for the proofs of soundness of the axoims. In section 2 we present a general technique inspired by Solovay 's work to obtain arithmetical completeness for theories containing IL, provided that we already have modal completeness w.r.t. a certain class of finite frames. The common preparatory work of section 2 is used in the last two sections for the two arithmetical completeness proofs.

I would like to thank Albert Visser for correcting and simplifying some of my arguments, Dick de Jongh and Rineke Verbrugge for their continuous and patient help.

§1. Interpretability logics. The language of the logic of interpretability contains (atomic) propositional letters $p_0, p_1, \ldots$, logical connectives $\rightarrow$, $\neg$ and a binary modal operator $\cdot \rhd \cdot$. All other connectives, as $\wedge, \vee$ and $\leftrightarrow$ are defined in the usual way. We use $\bot$ for falsum and $\top$ for true. The unary modal operator $\Box \cdot$ is defined as $\cdot \rhd \bot$. The axiom of IL are:

(L0) All tautologies of the propositional calculus.

(L1) $\Box(A \to B) \to (\Box A \to \Box B)$.

(L2) $\Box A \to \Box\Box A$.

(L3) $\Box(\Box A \to A) \to \Box A$.

(J1) $\Box(A \to B) \to A \rhd B$.

(J2) $(A \rhd B \wedge B \rhd C) \to A \rhd B$.

(J3) $A \rhd B \to (\Diamond A \to \Diamond B)$.

(J4) $\Diamond A \rhd A$.

The deduction rules of IL are modus ponens and necessitation The following two other axioms are the characteristic axioms of ILP and ILM.

(P) $A \rhd B \to (A \wedge \Box C \rhd B \wedge \Box C)$.

(M) $A \rhd B \to \Box(A \rhd B)$.

A *Veltman frame* is a triple $<W,S,R>$ where W is a set called *universe* , R and S are respectively a binary and a ternary relation on W. The elements of W are called *nodes* . We shall write xRy for $<x,y> \in R$ and yS_xz for $<x,y,z> \in S$. It is further required that R is transitive and conversely well founded and that for every $x \in W$, S_x is a reflexive and transitive relation on $\{y \mid xRy\} \subseteq W$. Moreover for every $x,y,z \in W$, xRyRz implies yS_xz.

A *Veltman model* is a Veltman frame together with a *forcing* relation $\Vdash$ between elements of W and the formulas of IL commuting with the logical connectives and satisfying the following:

$$x \Vdash \Box A \text{ iff } \forall y\, (xRy \Longrightarrow y \Vdash A),$$

$$x \Vdash A \rhd B \text{ iff } \forall y\, [(xRy \,\&\, y \Vdash A) \Longrightarrow (\exists z\, yS_xz \,\&\, z \Vdash B)].$$

As usual we shall improperly use the same letter W both for the model, the frame and the underlying universe. If W is a frame we write $W \models A$ iff for all forcing relations on W and all nodes of W, $x \Vdash A$.

We shall consider two other possible properties of Veltman frames:

P: If xS_wy then xS_zy for every z such that wRzRx.

M: If xS_wyRz then xRz.

We call W a *P-Veltman model* (resp. *M-Veltman model*) if the underlaying frame satisfies **P** (resp.**M**).

The modal completeness of IL, ILP and ILM has been proved by de Jongh and Veltman. In particular, they proved the following three theorems:

(1) $IL \vdash A$ iff for every finite Veltman frame W, $W \models A$.

(2) $ILP \vdash A$ iff for every finite P-Veltman frame W, $W \models A$.

(3) $ILM \vdash A$ iff for every finite M-Veltman frame W, $W \models A$.

§2. A Solovay style strategy. We want to find a general strategy for proving the arithmetical completeness of the interpretability logic for various arithmetical theories. Let T be a theory in the language of the arithmetic which is Σ_1 sound and Σ_1 complete and enough strong to formalize syntax. Given two arithmetical sentences α and β we shall write $\alpha \rhd \beta$ to mean the arithmetical formalization of the statement: *"$T+\alpha$ interprets $T+\beta$"* . It will be always clear from the context to which theory T we refer. We will use Latin letters for modal formulas and Greek letters for arithmetical formulas so that no confusion will arise from the fact that we are using the same symbols $\rhd$ and $\Box$ both for the modal and for the arithmetical operators.

An *interpretation* is a mapping ι from modal formulas to sentences of the language of the arithmetic such that:

(1) $\iota(A \rightarrow B) = \iota(A) \rightarrow \iota(B)$

(2) $\iota(\neg A) = \neg\iota(A)$

(3) $\iota(A \rhd B) = \iota(A) \rhd \iota(B)$

Let us write IL(T) for the set of modal formulas which are provable in T for every interpretation ι, i.e. $IL(T)=\{A \mid \forall\iota\ T \vdash \iota(A)\}$. Let ILX be a modal theory in the language of IL containing IL. We say that ILX is *arithmetically sound* for T if for every modal formula A if $ILX \vdash A$, then for every interpretation ι, $T \vdash \iota(A)$, i.e. if $IL(T) \supseteq ILX$. We say that ILX is *arithmetically complete* for T if the reverse inclusion also holds, i.e. whenever A is not a theorem of ILX then there is an interpretation ι such that $\iota(A)$ is not provable in T.

Claim. Let us suppose there is a class of finite Veltman frames X with respect to which we have modal completeness for the theory ILX. Let us suppose also that $IL(T) \supseteq IL$. If for any frame $W \in X$, there is a set $\{\lambda_x \mid x \in W\}$ of arithmetical sentences such that (o)-(iv) below are satisfied, then $IL(T) \subseteq ILX$.

(o) for every $x,y \in W$ if $x \neq y$ then $T \vdash \neg(\lambda_x \wedge \lambda_y)$

(i) for every $x \in W$, $T+\lambda_x$ is consistent.

(ii) for every $x \in W$, $T \vdash \lambda_x \rightarrow \Box \bigvee_{xRy} \lambda_y$.

(iii) for every $x,y,z \in W$ such that yS_xz, $T \vdash \lambda_x \rightarrow \lambda_y \rhd \lambda_z$

(iv) for every $x,y \in W$ such that xRy, $T \vdash \lambda_x \rightarrow \neg(\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z)$

Proof of the claim. We assume $ILX \nvdash C$ and define and interpretation ι such that $T \nvdash \iota(C)$. By the modal completeness there is a finite model W with frame in X such that $W \nVdash C$. Let $\{\lambda_x \mid x \in W\}$ be a set of arithmetical sentences satisfying conditions (o)–(iv). Let ι the interpretation which maps the atomic proposition p occurring in C to $\iota(p) := \bigvee\{\lambda_x \mid x \Vdash p\}$. We shall show by induction on the complexity of the modal formula A that for every $x \in W$:

(a) $x \Vdash A \Rightarrow T \vdash \lambda_x \rightarrow \iota(A)$

(b) $x \nVdash A \Rightarrow T \vdash \lambda_x \rightarrow \neg\iota(A)$.

This will suffice to prove the arithmetical completeness, because if $W \nVdash C$ then for some forcing relation on W and some $x \in W$, $x \nVdash C$, from which then by (b), $T \vdash \lambda_x \rightarrow \neg \iota(C)$. By (i), λ_x is consistent with T, as is therefore $\neg\iota(C)$. Hence $T \nvdash \iota(C)$.

It remains only to prove (a) and (b) by induction on the complexity of the formula A. By condition (o) it is clear that (a) and (b) hold for atomic sentences. The inductive step for $\rightarrow$ and $\neg$ are straightforward, so let us consider just the inductive steps for $\rhd$.

Let us prove first (a). Assume $x \Vdash A \rhd B$. Then for every y such that xRy, if $y \Vdash A$, there is a node z such that $yS_xz \Vdash B$. By the induction hypothesis we can write: for every y such that xRy, if $y \Vdash A$, there is a node z such that yS_xz and $T \vdash \lambda_z \rightarrow \iota(B)$. Using (iii) and Σ_1 completeness and the soundness of IL (i.e. making few steps of reasoning in IL) we get $T \vdash \lambda_x \rightarrow \bigwedge_{xRy \Vdash A} (\lambda_y \rhd \iota(B))$ and finally $T \vdash \lambda_x \rightarrow (\bigvee_{xRy \Vdash A} \lambda_y \rhd \iota(B))$. On the other hand, by (ii) and using the induction hypothesis (b) we obtain $T \vdash \iota(A) \rightarrow \neg \bigvee_{y \nVdash A} \lambda_y$, from which, since we assumed $T \vdash \lambda_x \rightarrow \Box \bigvee_{xRy} \lambda_y$, we get $T \vdash \lambda_x \rightarrow \Box(\iota(A) \rightarrow \bigvee_{xRy \Vdash A} \lambda_y)$. Again by the soundness of IL, $T \vdash \lambda_x \rightarrow \iota(A) \rhd \bigvee_{xRy \Vdash A} \lambda_y$. Thus the proof of (a) follows.

We prove now (b). Assume $x \nVdash A \rhd B$. Then there is a y such that xRy and $y \Vdash A$ and for every node z such that yS_xz, $z \nVdash B$. Thus, for some y such that xRy we have: $y \Vdash A \wedge \bigwedge_{yS_xz} z \nVdash B$. By the inductive hypotheses we have $T \vdash \lambda_y \rightarrow \iota(A)$ and $T \vdash \bigvee_{yS_xz} \lambda_z \rightarrow \neg\iota(B)$. By Σ_1 completeness we have $T \vdash \Box[\lambda_y \rightarrow \iota(A)]$ and $T \vdash \Box[\iota(B) \rightarrow \neg \bigvee_{yS_xz} \lambda_z]$, from which by the soundness of IL we get $T \vdash \lambda_y \rhd \iota(A)$ and $T \vdash \iota(B) \rhd \neg \bigvee_{yS_xz} \lambda_z$. Reason in T and assume λ_x. Assume for a contradiction that $\iota(A) \rhd \iota(B)$. By the soundness of IL we would have $\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z$, so from (iv) we obtain the desired contradiction. This completes the proof of the claim.

We conclude this section by remarking that conditions (o)-(iv) are not in general necessary, we believe that with a little additional work one can obtain more general, sufficient and necessary, conditions as is done in [BV] for the case of provability logic.

§3. The interpretability logic of finitely axiomatizable theories. In this section T may be any finitely axiomatizable Σ_1 sound theory extending $I\Delta_0$+SUPEXP. The main property which distinguishes interpretability over these theories is that the interpretability predicate in T is Σ_1 from which the soundness of the modal axiom P follows immediately. In T it is possible to characterize interpretability as follows. Let Δ_{EXP} be tableaux provability in $I\Delta_0$+EXP, Δ tableaux provably in T and $\nabla = \neg \Delta \neg$, i.e. the tableaux consistency in T. According to the Friedman-Visser characterization [Vis2], α interprets β iff $\Delta_{EXP}(\nabla\alpha \rightarrow \nabla\beta)$.

We want to prove that IL(T)=ILP. We leave, as usual, the proof of soundness to the reader and we shall prove only IL(T)$\subseteq$ILP. We shall find sentences (o)-(iv) as in the previous section. The method is as in Solovay [Sol]. We define a function F using the fixed point theorem and let the λ_x be some limit statements concerning F.

Assume for convenience W has been given as a finite set of nonzero natural numbers. We shall use the symbols x,y and z only for elements of W. Let λ_x be the sentence $\lim_n F(n)=x$ and $\lambda_0:=\forall n F(n)=0$. Together with the function F we will define also an auxiliary function G which will aid us in book keeping. The function G will always "follow" the function F, i.e. if for some n, F(n)=x then G(n)=F(m) for some m≤n. Speaking informally, G(n)≠F(n) will warn us of the fact that there is no proof of code less then n of $\neg\lambda_{F(n)}$. This has to be considered as a "dangerous signal" since we would like in the end to have $\lambda_x \to \Box\neg\lambda_x$. When such a situation occurs then only "safe" moves are allowed, i.e. F as well as G will move only to a node y for which there is a proof of $\neg\lambda_y$.

The definition of F and G is the following:

(a) F(0)=G(0)=0. If F(n)=0 and for some $x\in W$, n witnesses $\Delta\neg\lambda_x$, then F(n+1)=G(n+1)=x.

(b) If F(n)=G(n)=$x\in W$ and for some node y such that xRy, n witnesses $\Delta_{EXP}(\nabla\lambda_y \to \nabla\neg\bigvee_{yS_xz}\lambda_z)$, then F(n+1)=y and G(n+1)=G(n).

(c) If F(n)=y and G(n)=x, for some z, yS_xz and n witnesses $\Delta\neg\lambda_z$, then F(n+1)=G(n+1)=z.

(d) In all other cases F(n+1)=F(n) and G(n+1)=G(n).

Let μ_x be the sentence $\lim_n G(n)=x$. We shall eventually prove that the two functions have the same limit, i.e. $\mu_x \leftrightarrow \lambda_x$, but for proving this we need the cut elimination theorem. The formalization of the cut elimination theorem is provable in T since T contains SUPEXP but is surely not provable in EXP. To carry on with our proof we need to know what $I\Delta_0$+EXP proves about the functions F and G, hence the following:

Lemma 1. $I\Delta_0$+EXP proves the following:

.1 For every $w\in W$, $\mu_w \to \Delta\bigvee_{wRx}\lambda_x$.

.2 For every $w,x\in W$, if $x\neq w$ then $\mu_w \wedge \lambda_x \to \Delta\bigvee_{xS_wy}\lambda_y$.

.3 For every $w\in W$, $\mu_w \wedge \lambda_w \to \nabla\lambda_y$.

.4 For every $x,y,w\in W$, if xS_wy then $\mu_w \wedge \lambda_x \to \nabla\lambda_y$.

Proof. Directly from the definition of F, $I\Delta_0$+EXP proves that if, for some n, G(n)=w then after stage n the function F remains either in w or in the upper cone above w. Thus the limit of F is either w or is some node above w. If G(n)=w then by provable Σ_1 completeness, $\Delta_{EXP}(G(n)=w)$ and a fortiori $\Delta(G(n)=w)$. The proof of (.1) follows by combining all this with the fact that G(n)=w implies $\Delta\neg\lambda_w$. To prove (.2) assume that for some x≠w we have $\mu_w \wedge \lambda_x$. Then for some n $\Delta_{EXP}(G(n)=w \wedge F(n)=x)$. Again, observing the definition of the functions F and G, it is easy to argue that whenever $G(n)=w \wedge F(n)=x$ for some w≠x, the function F never leaves the set of nodes which are in S_w relation with x. This gives (.2); (.3) is immediate and (.4) becomes obvious by inspection of case (b) in the definition of F.

For the following lemma we need that the formula $(\nabla\alpha\wedge\alpha\rhd\beta)\rightarrow\nabla\beta)$ is provable in T. It is easy to chek that T (or even $I\Delta_0$+EXP) proves $(\Diamond\alpha\wedge\alpha\rhd\beta)\rightarrow\Diamond\beta)$, and since in T the formalization of the cut elimination theorem is provable, we can substitute tableaux consistency with normal consistency, so also the former formula is derivable in T. We can prove the following:

Lemma 2. For every $x\in W$, $T\vdash\mu_x\leftrightarrow\lambda_x$.

Proof. Reason in T and assume for a contradiction that $\lambda_x\wedge\neg\mu_x$. Then for some wRx we have μ_w. This implies $\nabla\lambda_x$, for otherwise the function G would have jump to x. Since $x\neq w$ the last move of the function F has been from w to x using condition (b) and therefore $\lambda_x\rhd\neg\bigvee_{xS_wy}\lambda_y$. Bythe remak above we get immediately $\neg\Delta\bigvee_{xS_wy}\lambda_y$. From lemma 1.2 we get also $\Delta\bigvee_{xS_wy}\lambda_y$. Thus we have the desired contradiction.

Lemma 3. For every $x,y,z\in W$ such that yS_xz, $T\vdash\lambda_x\rightarrow\lambda_y\rhd\lambda_z$.

Proof. Reason in T and assume λ_x. We want to show that for every y,z such that yS_xz, $\lambda_y\rhd\lambda_z$, i.e. $\Delta_{EXP}(\nabla\lambda_y\rightarrow\nabla\lambda_z)$. By lemma 2 we have μ_x and by provable Σ_1 completeness we have that for some k, $\Delta_{EXP}(G(k)=x)$. Reason in $I\Delta_0$+EXP. Assume $\nabla\lambda_y$ and let w be the limit of the function G. Since G(k)=x, the limit w is either x or is above x. By lemma 1.1, from $\nabla\lambda_y$ we know that w has to be strictly below y. Thus either x=wRy or xRwRy and, by the characteristic property of the P-Veltman frames, from yS_xz we get yS_wz. Let u be the limit of F. If u=w from wRz and lemma 1.3 the lemma follows immediately. Otherwise by lemma 1.2 and $\nabla\lambda_y$ one has uS_wy. By the transitivity of S_w we obtain uS_wz and thus finally, by lemma 1.4, $\nabla\lambda_z$.

Lemma 4. For every $x\in W$, $T\vdash\lambda_x\rightarrow\Delta\bigvee_{xRy}\lambda_y$

Proof. Immediate by lemmas 1.1 and 2.

We can now easily check that the set of sentences $\{\lambda_x \mid x\in W\}$ satisfies (o)-(iv). In fact (o) is trivial, the proof of (i) is completely standard, (ii) derives from lemma 4 and the provability in T of the cut elimination theorem. Condition (iii) is lemma 2 and (iv) is obvious by the definition of F. This concludes the proof of the completeness theorem.

§4. The interpretability logic of PA. In this section we want to prove that IL(PA)=ILM. The main characteristic of the interpretability in Peano arithmetic is the Orey-Hajek characterization: let $\Box_k\beta$ be the formalization of the sentence *"there is a proof of β which uses only the first k axioms of PA"*, let $\Diamond_k\equiv\neg\Box_k\neg$, then it is provable in PA that α interprets β iff $\forall k\Box(\alpha\rightarrow\Diamond_k\beta)$. Another characteristic property of PA is that it proves full reflection for any of its finite subtheories, moreover this is formalizable in PA, namely: for every α, $PA \vdash \forall k \Box(\Box_k \alpha \rightarrow \alpha)$. These facts would be sufficient to carry out the following proof, but for sake of better readability we shall, following Berarducci, work in ACA_0 rather then in PA. The second order theory ACA_0 is a conservative extention of PA; in ACA_0 we can speak of models of PA and easy theorems of basic model theory are formalizable and provable in ACA_0. In particular in ACA_0 we have the following characterization of the interpretability over PA: *"PA+α interprets PA+β iff every model of PA+α has an end extension to a model of PA+β"*. In ACA_0 the *standard model* is the set $\{x \mid x=x\}$ with the obvious choice of operations, any other *nonstandard model* has an initial segment which is isomorfic to it. Numbers belonging to this initial segment are called as usual *standard numbers* . Full reflection translates in ACA_0 in the following manner: *"for every model Y of PA and every standard number k, $Y \models \Box_k \alpha \rightarrow \alpha$"*.

As in the previous section we shall prove only that $IL(PA) \subseteq ILM$, leaving the converse to the reader. The sentences which are meant to satisfy (o)-(iv) are defined as limits of a recursive function F exactly as in the previous proof . Define, as in [Ber] for every $x \in W$, rank(x,n):=*"the minimal k such that there is a witness $\leq n$ of $\Box_k \neg \lambda_x$"* . If k is a number, $x,y \in W$, xRy then we define the sentence $\alpha_{x,y}(k)$ as $\forall j \geq k[F(j)=x \vee F(j)=y]$[1]. Our definition of the function F resembles Berarducci's as far as it is concerned with the S–jumps but it differs in the R-jumps. Roughly speaking we allow the function F to make an R–jump if there is a proof that this will not be the last move. We assume for convenience that W has been coded as a finite set of nonzero natural numbers, we shall use the symbols w,x,y,...etc. only for elements of W.

(a) Let F(0)=0 and if F(n)=0 and for some $x \in W$, n witnesses $\Box \neg \lambda_x$, then F(n+1)=x.

(b) If F(n)=x and for some $y \in W$ and some k<n such that $\forall j \in [k,n]$ F(j)=x and xRy, n witnesses $\Box \neg \alpha_{x,y}(\mathbf{k})$ (here the bold **k** means the numeral of k), then F(n+1)=y.

(c) If F(n)=x and for some nodes y and z, such that xS_zy and $\exists i \leq n[\text{rank}(y,n) \leq i < \text{rank}(x,n) \wedge F(i)=z]$, then F(n+1)=y. (If this condition obtains for two different nodes, choose the one with minimal code.)

(d) In all the other cases F(n+1)=F(n).

Note that any two points in the orbit of F are connected by an S and/or R arrow. We shall write $Y \models ...x...y$ if, according to the model Y the function F goes from x to y (possibly in a nonstandard number of steps). We write $Y \models ...xRy...$ (resp. $Y \models ...xS_zy...$) if, in the model Y, F moves in one step from x to y and xRy (resp. xS_zy). If in a model Y the function F moves at stage n from x to y, then

---

[1] The reader might find the following alternative definition of $\alpha_{x,y}(k)$ more intuitive: $\exists p[\forall j \in [k,p]\ F(j)=x \wedge \forall j>p F(j)=y]$ This means *"from k on the function F remains in the node x until a stage p is reached at which it jumps to y and stays there forever"* .

we say F moves with an R-step (resp. with S-step) if at stage n condition (b) (resp. condition (c)) has been applied. If, at stage n, F moves from 0 to some node x, we say that F moves with an (a)-step.

Lemma 1. In PA it is provable that the function F has a limit.

Proof. This is not obvious since the S-relations are in general not well founded. It is clear that if h is the height of the frame the function cannot make more than h consecutive R-moves. By the property **M** of the M-frame F cannot make more than h R-moves, whether they are consecutive or not. Thus eventually F is allowed only to make S moves. If S would not have a limit we could construct a definable infinite decreasing sequence of ranks. This is provably false in PA.

We are eventually going to prove $\lambda_x \to \Box\neg\lambda_x$, but to achieve this goal we need to prove first a weaker form of it.

Lemma 2. For every $x\in W$ and for every $k\in\omega$, $PA\vdash F(k)=x \to \Box\exists j>\mathbf{k}\ F(j)\neq x$.

Proof. Assume $F(k)=x$. Reasoning in ACA_0 we claim that for model Y of PA, $Y\models\exists j>k\ F(j)\neq x$. If F moved to x with an (a)-step or with an S-step we would have $\Box\neg\lambda_x$ and then $Y\models\neg\lambda_x$ so our claim would hold trivially. So, assume that the last move of F has been an R-step, and that say at stage h, the function F moves from z to x. Then for some i<h such that $\forall j\in[i,h]\ F(j)=z$, h codes a proof of $\neg\alpha_{z,x}(i)$. So, $Y\models\exists j\geq i\,[F(j)\neq z \wedge F(j)\neq x]$. We have assumed $\forall j\in[i,k]\ [F(j)=z \vee F(x)]$, this is a Σ_1 statement so, by provable Σ_1 completeness, it is true also in Y. Thus $Y\models\exists j>k\ F(j)\neq x$ and our claim is proved.

Lemma 3. For every $x\in W$, $PA\vdash\lambda_x\to\Box\bigvee_{xRy}\lambda_y$.

Proof. It is sufficient to prove that for every x and y, if $\neg xRy$ then $PA\vdash\lambda_x\to\Box\neg\lambda_y$. Reason in ACA_0 and assume for a contradiction that λ_x, $\Diamond\lambda_y$ and $\neg xRy$. Let k be the minimal number such that $\forall j>k\ F(j)=x$ and let Y be a model of λ_y. By provable Σ_1 completeness we have that $Y\models F(k+1)=x$. Now, in Y, let z be the last node that the function passes through before arriving to y. The last step must be an S-step otherwise zRy and by the **M** property of the M-Veltman frames we would have xRy. We shall picture the situation as $Y\models\ldots x\ldots zS_wy$ but we have to remember that z could be equal to x. (Anyhow, by the previuos lemma we can exclude that both z and y are equal to x.) By the definition of F we have that at some stage n, for some $i\leq n$, $rank(y,n)\leq i<rank(z,n)$ and $F(i)=w$. Since zS_wy and in particular wRy we have that $w\neq x$. By the reflection principle rank(y,n) has to be nonstandard in Y, and since we have chosen k standard, $rank(y,n)\geq k$. Thus also $i\geq k$ and so $Y\models\ldots F(k)\ldots.F(i)$ and therefore $Y\models\ldots x\ldots w\ldots zS_wy$. By the **M** property of the M-Veltman frames from wRy we get xRy. Contradiction.

Lemma 4. For every $x,y,z \in W$ such that yS_xz, $PA \vdash \lambda_x \rightarrow \lambda_y \rhd \lambda_z$.

Proof. Assume λ_x and yS_xz. We shall prove in ACA_0 that, for arbitrary large k, in any model Y of PA, $\lambda_y \rightarrow \Diamond_k \lambda_z$. Let k be such that F(k)=x. Suppose for a contradiction that there exists a model $Y \models \lambda_y \wedge \Box_k \neg \lambda_z$. Then for n large enough we have $Y \models rank(z,n) \leq k < n$. Suppose n is also large enough so that (in Y) F has already reached its limit. By the reflection principle rank(y,n) must be nonstandard in Y. Then $Y \models rank(z,n) \leq k < rank(y,n) \wedge F(k)=x$. So, $Y \models F(n+1)=z$ which contradicts the fact that F has already reached its limit.

Lemma 5. for every $x,y \in W$ such that xRy, $PA \vdash \lambda_x \rightarrow \neg(\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z)$.

Proof. Reason in ACA_0 and assume for a contradiction that λ_x and $\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z$. Then every model $Y \models \lambda_y$ has an end extension to a model of $\neg \bigvee_{yS_xz} \lambda_z$. Let Z be any end extension of such a model Y and let z such that $Z \models \lambda_z$. We shall obtain a contradiction by showing that yS_xz. For this purpose we have to choose the model Y a bit carefully. Let k be such that $\forall j \geq k\ F(j)=x$. Since xRy we have: $\Diamond \alpha_{x,y}(k)$ otherwise the function would jump from x to y contradicting λ_x. Then let $Y \models \forall j > k[F(j)=x \vee F(j)=y]$; from the latter, since we have assumed λ_x and therefore (by lemma 3) $Y \models \neg \lambda_x$, we can conclude that $Y \models \lambda_y$. Let $Y \subseteq_e Z \models \lambda_z$ and let n be the minimal number in Z such that such that $Z \models F(n+1)=z$. By provable Σ_1 completeness and since Σ_1 formulas are conserved by end extensions, we have $Z \models$ ...xRy.....z. Let w be the last node reached with an R step i.e. for some u, $Z \models$ ...xRy...uRw...z and between w and z only S steps occur. Then the rank of all the steps between w and z is larger than rank(z,n). By the reflection principle rank(z,n) is a nonstandard number in Z. If all the step between w and z are S_x steps, we are done, otherwise let S_t be the last non S_x step between w and z i.e. $Z \models$ ...xRy...uRw...$S_t v\ S_x$....$S_x z$. Let $i \geq rank(z,n)$, be such that F(i)=t. Since rank(z,n) is nonstandard in Z, t cannot occur in the orbit of F before x, so either t=y or $Z \models$ ...xRy...t...$S_t v S_x$....$S_x z$. In both cases one can conclude that yRv and hence yS_xz.

We can now easily check that the set of sentences $\{\lambda_x \mid x \in W\}$ satisfies (o)-(iv). In Fact (o) is trivial, the proof of (i) is completely standard, (ii) is lemma 3, (iii) is lemma 4 and (iv) is lemma 5. This concludes the proof of the completeness theorem.

References.

[Ber] A. Berarducci, The interpretability logic of Peano arithmetic. *J. Symb. Logic* **56**, 1059-1089 (1990).

[BV] A.Berarducci and R.Verbrugge, On the metamathematics of weak theories. *ITLI Prepublication Series*, **ML-91-02**, University of Amsterdam (1991), Department of Mathmatics and Computer Science, Plantage Muidergracht 24, 1018 TV Amsterdam.

[dJV] D. de Jongh and F. Veltman, Provability logic for relative interpretability. In: *Mathematical logic*, 31-42, edited by P.P.Petkov, Plenum Press, New York (1990).

[dJJM] D. de Jongh, M.Jumelet and F.Montagna, On the proofof Solovay's theorem. To appear in *Studia Logica* .

[Sha] V.Y.Shavrukov, The logic of relative interpretability over Peano arithmetic. (Preprint in Russian) Steklov Mathematical Institute (1988), Vavilova 42, 117966 Moscow GSP-1, Moscow, USSR.

[Sol] R.Solovay, The provability interpretation of modal logic. *Isr. J. Math.* **25**, 287-304, (1976).

[Vis1] A.Visser, Preliminary notes on interpretability logic. *Logic Group Preprint Series* , **29**, University of Utrecht (1988), Dept. of Philosophy, University of Utrecht, Heidelberglaan 2, 3584 CS Utrecht.

[Vis2] A. Visser, Interpretability logic. In: *Mathematical logic*, 175-209, edited by P.P.Petkov, Plenum Press, New York (1990).

The ITLI Prepublication Series

1990

Logic, Semantics and Philosophy of Language

LP-90-01 Jaap van der Does	A Generalized Quantifier Logic for Naked Infinitives
LP-90-02 Jeroen Groenendijk, Martin Stokhof	Dynamic Montague Grammar
LP-90-03 Renate Bartsch	Concept Formation and Concept Composition
LP-90-04 Aarne Ranta	Intuitionistic Categorial Grammar
LP-90-05 Patrick Blackburn	Nominal Tense Logic
LP-90-06 Gennaro Chierchia	The Variablity of Impersonal Subjects
LP-90-07 Gennaro Chierchia	Anaphora and Dynamic Logic
LP-90-08 Herman Hendriks	Flexible Montague Grammar
LP-90-09 Paul Dekker	The Scope of Negation in Discourse, towards a flexible dynamic Montague grammar
LP-90-10 Theo M.V. Janssen	Models for Discourse Markers
LP-90-11 Johan van Benthem	General Dynamics
LP-90-12 Serge Lapierre	A Functional Partial Semantics for Intensional Logic
LP-90-13 Zhisheng Huang	Logics for Belief Dependence
LP-90-14 Jeroen Groenendijk, Martin Stokhof	Two Theories of Dynamic Semantics
LP-90-15 Maarten de Rijke	The Modal Logic of Inequality
LP-90-16 Zhisheng Huang, Karen Kwast	Awareness, Negation and Logical Omniscience
LP-90-17 Paul Dekker	Existential Disclosure, Implicit Arguments in Dynamic Semantics

Mathematical Logic and Foundations

ML-90-01 Harold Schellinx	Isomorphisms and Non-Isomorphisms of Graph Models
ML-90-02 Jaap van Oosten	A Semantical Proof of De Jongh's Theorem
ML-90-03 Yde Venema	Relational Games
ML-90-04 Maarten de Rijke	Unary Interpretability Logic
ML-90-05 Domenico Zambella	Sequences with Simple Initial Segments
ML-90-06 Jaap van Oosten	Extension of Lifschitz' Realizability to Higher Order Arithmetic, and a Solution to a Problem of F. Richman
ML-90-07 Maarten de Rijke	A Note on the Interpretability Logic of Finitely Axiomatized Theories
ML-90-08 Harold Schellinx	Some Syntactical Observations on Linear Logic
ML-90-09 Dick de Jongh, Duccio Pianigiani	Solution of a Problem of David Guaspari
ML-90-10 Michiel van Lambalgen	Randomness in Set Theory
ML-90-11 Paul C. Gilmore	The Consistency of an Extended NaDSet

Computation and Complexity Theory

CT-90-01 John Tromp, Peter van Emde Boas	Associative Storage Modification Machines
CT-90-02 Sieger van Denneheuvel, Gerard R. Renardel de Lavalette	A Normal Form for PCSJ Expressions
CT-90-03 Ricard Gavaldà, Leen Torenvliet, Osamu Watanabe, José L. Balcázar	Generalized Kolmogorov Complexity in Relativized Separations
CT-90-04 Harry Buhrman, Edith Spaan, Leen Torenvliet	Bounded Reductions
CT-90-05 Sieger van Denneheuvel, Karen Kwast	Efficient Normalization of Database and Constraint Expressions
CT-90-06 Michiel Smid, Peter van Emde Boas	Dynamic Data Structures on Multiple Storage Media, a Tutorial
CT-90-07 Kees Doets	Greatest Fixed Points of Logic Programs
CT-90-08 Fred de Geus, Ernest Rotterdam, Sieger van Denneheuvel, Peter van Emde Boas	Physiological Modelling using RL
CT-90-09 Roel de Vrijer	Unique Normal Forms for Combinatory Logic with Parallel Conditional, a case study in conditional rewriting

Other Prepublications

X-90-01 A.S. Troelstra	Remarks on Intuitionism and the Philosophy of Mathematics, Revised Version
X-90-02 Maarten de Rijke	Some Chapters on Interpretability Logic
X-90-03 L.D. Beklemishev	On the Complexity of Arithmetical Interpretations of Modal Formulae
X-90-04	Annual Report 1989
X-90-05 Valentin Shehtman	Derived Sets in Euclidean Spaces and Modal Logic
X-90-06 Valentin Goranko, Solomon Passy	Using the Universal Modality: Gains and Questions
X-90-07 V.Yu. Shavrukov	The Lindenbaum Fixed Point Algebra is Undecidable
X-90-08 L.D. Beklemishev	Provability Logics for Natural Turing Progressions of Arithmetical Theories
X-90-09 V.Yu. Shavrukov	On Rosser's Provability Predicate
X-90-10 Sieger van Denneheuvel, Peter van Emde Boas	An Overview of the Rule Language RL/1
X-90-11 Alessandra Carbone	Provable Fixed points in $I\Delta_0+\Omega_1$, revised version
X-90-12 Maarten de Rijke	Bi-Unary Interpretability Logic
X-90-13 K.N. Ignatiev	Dzhaparidze's Polymodal Logic: Arithmetical Completeness, Fixed Point Property, Craig's Property
X-90-14 L.A. Chagrova	Undecidable Problems in Correspondence Theory
X-90-15 A.S. Troelstra	Lectures on Linear Logic

1991

Mathematical Logic and Foundations

ML-91-01 Yde Venema	Cylindric Modal Logic
ML-91-02 Alessandro Berarducci, Rineke Verbrugge	On the Metamathematics of Weak Theories
ML-91-03 Domenico Zambella	On the Proofs of Arithmetical Completeness for Interpretability Logic

Other Prepublications

X-91-01 Alexander Chagrov, Michael Zakharyaschev	The Disjunction Property of Intermediate Propositional Logics
X-91-02 Alexander Chagrov, Michael Zakharyaschev	On the Undecidability of the Disjunction Property of Intermediate Propositional Logics
X-91-03 V. Yu. Shavrukov	Subalgebras of Diagonizable Algebras of Theories containing Arithmetic
X-91-04 K.N. Ignatiev	Partial Conservativity and Modal Logics
X-91-05 Johan van Benthem	Temporal Logic