

Interpretability over Peano arithmetic

Claes Strannegård*
Department of Philosophy
Utrecht University
`claes.strannegard@phil.ruu.nl`

Abstract

We investigate the modal logic of interpretability over Peano arithmetic (PA). Our main result is an extension of the arithmetical completeness theorem for the interpretability logic ILM^ω . This extension concerns recursively enumerable sets of formulas of interpretability logic (rather than single formulas). As corollaries we obtain a uniform arithmetical completeness theorem for the interpretability logic ILM and a theorem answering a question of Orey from 1961. All these results also hold for Zermelo-Fraenkel set theory (ZF).

1 Introduction

Provability logic

Provability logic is concerned with the investigation of various metamathematical relations with the aid of modal logic. For instance, provability logic has been used to investigate different notions of provability, interpretability, conservativity, and tolerance.

The fundamental connection between modal logic and metamathematics is provided by certain functions called realizations. These are primitive recursive functions that translate modal formulas into arithmetical sentences, in a way that preserves logical form and maps modalities on formalizations of metamathematical relations.

The so-called arithmetical completeness theorems (ACTs) have played an extremely important role in provability logic ever since the first ones were showed by Solovay in 1976 [Sol76]. These theorems connect some notion of realization with a system of modal logic \mathbf{L} and may be phrased as follows:

“Let A be a modal formula. Then every realization translates A into a true (provable) arithmetical sentence iff A is provable in \mathbf{L} .”

Thus, ACTs tell us exactly what metamathematical principles of a certain linguistic class are true (provable). Furthermore, in cases where the modal logic \mathbf{L} is decidable, the ACTs imply decidability theorems for the corresponding fragments of metamathematics.

*I would like to thank Per Lindström, Volodya Shavrukov, Daniel Vallström and Albert Visser for valuable comments and the Swedish STINT foundation for financial support.

When it comes to certain infinitary aspects of metamathematics, however, standard ACTs are of little help. To get at those aspects, the perspective must be shifted from single modal formulas to infinite sets of modal formulas. Then one may ask, for instance, what infinite sets of modal formulas are realizable in the sense that there is a realization mapping all the formulas of the set on true (provable) arithmetical sentences.

So far, only a few scattered results seem to exist in this field of (infinitary) provability logic. Among those are the uniform ACTs (see [Smo85]) and some theorems of Shavrukov on Magari algebras [Sha93]. This author's Ph. D. thesis [Str96] was also devoted to the subject.

A survey of provability logic can be found in [JdJ98].

Interpretability logic

As a preliminary, let us start by outlining the definition of interpretability. Roughly, the theory S *interprets* the theory T if there is a natural way of translating the language of S into the language of T in such a way that the translations of all the axioms of T become provable in S . We write $S \geq T$ if this is the case. For a precise definition, see e.g. [Vis97].

A derived notion is that of (relative) interpretability over a base theory T . Let α and β be arithmetical sentences. We say that α *interprets* β *over* T if $T + \alpha \geq T + \beta$.

Interpretability logic is the branch of provability logic that studies interpretability. The subject was introduced by Visser [Vis90], who also showed two ACTs for the relation of interpretability over finitely axiomatized theories. In the same vein, Berarducci [Ber90] and Shavrukov [Sha88] independently showed two ACTs for the relation of interpretability over theories like PA and ZF .

An overview of interpretability logic, including ample motivation for studying the subject, can be found in [Vis97].

In the present paper, we obtain extensions of the above-mentioned ACTs by Berarducci and Shavrukov. Our extensions concern r. e. sets of modal formulas (rather than single formulas). The paper is organized as follows. In section 2 we state our main theorem and in section 3 we prove it. Section 4 contains four easy corollaries to the main theorem.

2 Main theorem

We begin this section by introducing the terminology needed to state the main theorem.

- The *formulas of interpretability logic* are constructed from propositional variables using the binary modality \triangleright , the unary modality \Box , the connective \rightarrow , and the propositional constant \perp . The symbols \neg , \vee , \wedge , and \leftrightarrow are introduced as abbreviations in the usual way. We write \top for $\neg\perp$ and \Diamond for $\neg\Box\neg$.
- We use the letter p for propositional variables and A, B, C, D for formulas of interpretability logic. The letters $\mathcal{E}, \mathcal{F}, \mathcal{G}$ are reserved for sets of formulas of interpretability logic.

- The set $S(A)$ of *subformulas* of A is defined as expected. Let $S(\mathcal{E})$ be the set of subformulas of formulas of \mathcal{E} .
- The rules of the interpretability logic ILM are modus ponens and necessitation (from A infer $\Box A$), and its axioms are as follows:

All propositional tautologies,
 $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$,
 $\Box A \rightarrow \Box \Box A$,
 $\Box(\Box A \rightarrow A) \rightarrow \Box A$,
 $\Box(A \rightarrow B) \rightarrow A \triangleright B$,
 $(A \triangleright B \wedge B \triangleright C) \rightarrow A \triangleright C$,
 $(A \triangleright C \wedge B \triangleright C) \rightarrow (A \vee B) \triangleright C$,
 $A \triangleright B \rightarrow (\Diamond A \rightarrow \Diamond B)$,
 $\Diamond A \triangleright A$,
 $A \triangleright B \rightarrow (A \wedge \Box C \triangleright B \wedge \Box C)$.

- The logic ILM^ω has the same set of formulas as ILM . Its axioms are the theorems of ILM together with the formulas $\Box A \rightarrow A$. Its only rule is modus ponens.
- \mathcal{E} is *well-specified* iff $A \in S(\mathcal{E}) \Rightarrow A \in \mathcal{E}$ or $\neg A \in \mathcal{E}$, for every A whose main symbol is \triangleright or \Box .
- The *formulas of arithmetic* are built up as usual from the non-logical symbols 0 , succ , $+$, \times . We use the letters α and β for arithmetical sentences.
- Let $\alpha \triangleright \beta$ and $\Box \alpha$ be natural formalizations of the assertions “ $\text{PA} + \alpha$ interprets $\text{PA} + \beta$ ”, and “ PA proves α ”, respectively. Note that we use the symbols \Box and \triangleright both in modal and arithmetical contexts. This ambiguity should be unproblematic because of our other notational conventions.
- A *realization* is a recursive function i from modal formulas to arithmetical sentences such that $i(A \triangleright B) = i(A) \triangleright i(B)$, $i(\Box A) = \Box i(A)$, $i(\perp) = \perp$, and $i(A \rightarrow B) = i(A) \rightarrow i(B)$. Note that realizations are uniquely determined by their values on propositional variables.
- \mathcal{E} is *realizable* iff there is a realization i such that $i(A)$ is true, for every $A \in \mathcal{E}$.

Example 2.1 The following sets are realizable: $\{\neg \Box p, \neg \Box \neg p\}$ (by Gödel’s first incompleteness theorem), $\{\neg \Box \Diamond \top\}$ (by Gödel’s second incompleteness theorem), $\{\top \triangleright p, \top \triangleright \neg p\}$ (by a result of Orey), $\{\neg(\top \triangleright p), \neg(\top \triangleright \neg p)\}$ (by a result of Montague).

Now we may state our main theorem:

Theorem 2.2 Suppose \mathcal{E} is r. e. and well-specified. Then \mathcal{E} is realizable iff \mathcal{E} is consistent with ILM^ω .

Example 2.3 Let us give a quick example of how this theorem can be applied. Let $\mathcal{E} := \{\neg(p_i \triangleright p_j) : i \neq j\}$. Then \mathcal{E} is clearly r. e. and well-specified. By an easy modal logical argument based on Lemma 3.1(iii), \mathcal{E} can be shown to be consistent with ILM^ω . Hence \mathcal{E} is realizable.

3 Proof of the main theorem

To prove the main theorem we need some preliminary definitions and results.

Modal preliminaries

- A *model* is a tuple (W, R, S, \Vdash, r) , where (W, R, \Vdash, r) is a finite Kripke structure and S is a three-place relation on W , with the following properties. The relation R is irreflexive and transitive and rRa holds for all $a \in W$ such that $a \neq r$. We write $bS_a c$ if $(a, b, c) \in S$. Moreover, S_a is a reflexive transitive relation between nodes that lie above a in the sense of R . Finally, the following conditions are met.

$$V: aRbRc \Rightarrow bS_a c,$$

$$M: bS_a cRd \Rightarrow bRd.$$

We use the letters a, b, c, d for nodes in general and r for roots. The letter k is reserved for models.

- The *forcing relation* \Vdash is extended to cover arbitrary formulas of interpretability logic using the following key clause:

$$a \Vdash A \triangleright B \Leftrightarrow \forall b[(aRb \wedge b \Vdash A) \Rightarrow \exists c(bS_a c \wedge c \Vdash B)].$$

- Let $k \Vdash A$ mean that the root of k forces A .
- k is *A-reflexive* if $k \Vdash \Box B \rightarrow B$, for every $\Box B \in S(A)$, and furthermore, $k \Vdash \Box \neg B \rightarrow \neg B$, for every $B \triangleright C \in S(A)$.
- Let $k = (W, R, S, \Vdash, r)$ be a model. The following conditions define the *derived model* $k' = (W', R', S', \Vdash', r')$ of k (up to isomorphism):

- $W' = W \cup \{r'\}$, for some $r' \notin W$
- $R' = R \cup \{(r', a) : a \in W\}$
- $S'_{r'} = \{(r, r)\} \cup R \cup S_r$
- $S'_a = S_a$ if $a \neq r'$
- $\Vdash' = \Vdash \cup \{(r', p) : r \Vdash p\}$

- Let $\vdash A$ mean that A is a theorem of ILM^ω and let $\vdash A$ mean that A is a theorem of ILM . When \mathcal{G} is finite, we take $\mathcal{G} \vdash A$ to mean that $\vdash \bigwedge \mathcal{G} \rightarrow A$.

Now let us fix a natural way of encoding nodes and models. We assume that 0 is not the code of any node of any model. We also assume that the code of each node belonging to a model k permits us to recover the model k in a primitive recursive fashion.

Lemma 3.1 (de Jongh, Veltman) The following is provable in PA :

- (i) A is consistent with ILM iff A is forced by some model.
- (ii) If k is an A -reflexive model forcing A , then so is k' .

(iii) A is consistent with ILM^ω iff A is forced by some A -reflexive model.

(iv) $\vdash A$ iff $\vdash \Box A$.

(v) The relations $\vdash A$ and $\vdash A$ are primitive recursive.

Proof.

(i) See [dJV90].

(ii) Argue in PA: “Suppose k is an A -reflexive model forcing A . We show that $k \Vdash B$ iff $k' \Vdash B$, by induction on $B \in S(A)$.

- The base step and the Boolean cases are easy.
- $B = C \triangleright D$. Note that k' forces $\neg(C \triangleright D)$ if k does. Conversely, suppose k' forces $\neg(C \triangleright D)$ and that k does not. Then, by considering k' , it is not hard to see that k forces C and $\Box \neg D$. But k is A -reflexive and therefore $k \Vdash \Box \neg C \rightarrow \neg C$. So we can conclude that k forces $\Diamond C$ and $C \triangleright D$. Consequently $k \Vdash \Diamond D$. But that, we know, is false.
- $B = \Box C$. This is similar to the previous case.

This completes the induction. It follows that k' forces A and also that k' is A -reflexive.”

(iii) This follows from (i) and (ii) by a standard argument based on the pigeon-hole principle.

(iv) This follows from (i), (ii), and the same kind of pigeon-hole argument again.

(v) This follows from (i), (iii), and some simple combinatorial considerations.

□

In view of Lemma 3.1(iii) we make the following definition:

- When \mathcal{G} is finite, we say that the model k *witnesses* that $\mathcal{G} \not\models A$ if k is a $(\bigwedge \mathcal{G} \wedge \neg A)$ -reflexive model forcing $\bigwedge \mathcal{G} \wedge \neg A$.

Arithmetical preliminaries

- We use the dot-notation of [Smo85]. Thus, if $\phi(x)$ is a formula with one free variable x , then $\phi(\dot{x})$ is a term with one free variable x .
- The theory $\text{I}\Sigma_s$ is PA with induction restricted to Σ_s formulas.
- Let $\Box_s(\alpha)$ be a natural formalization of “ $\text{I}\Sigma_s$ proves α ” and let $\Diamond_s(\alpha)$ be an abbreviation of $\neg \Box_s(\neg \alpha)$.

The following characterization of interpretability over PA will be used several times.

Lemma 3.2 (Orey-Hájek) PA proves: $\forall \alpha, \beta [\alpha \triangleright \beta \leftrightarrow \forall s \Box(\alpha \rightarrow \Diamond_s(\beta))]$.

Proof. See page 1065 in [Ber90].

□

We also need the following reflection-principle:

Lemma 3.3 Fix $\phi(x)$. PA proves: $\forall x, s \Box(\Box_s(\phi(\dot{x})) \rightarrow \phi(\dot{x}))$.

Proof. See page 1064 in [Ber90]. □

Next we introduce the theory ACA_0 . This step is not strictly necessary for our present purposes, but working in ACA_0 instead of PA will make some of our proofs easier.

- ACA_0 is a theory in a fragment of second-order logic. It has two sorts of variables: number variables x, y, \dots and set variables X, Y, \dots . Its vocabulary contains that of PA and in addition the membership relation-symbol \in . Its axioms are as follows:
 - all the axioms of PA except the induction axioms
 - induction: $\forall X(0 \in X \wedge \forall x(x \in X \rightarrow x + 1 \in X) \rightarrow \forall x(x \in X))$
 - comprehension for arithmetical sets: $\exists X \forall x(x \in X \leftrightarrow \phi(x))$, for every formula $\phi(x)$ in which no second-order variables are quantified
 - extensionality: $\forall x(x \in X \leftrightarrow x \in Y) \rightarrow X = Y$.

ACA_0 is discussed in [Ber90] and [HP91]. ACA_0 is a conservative extension of PA. First-order models can be defined in ACA_0 together with their satisfaction predicates. In particular, models of PA may be discussed in ACA_0 and, as explained in [Ber90] and [Zam92], ACA_0 can differentiate between (relatively) standard and non-standard elements of such models. Below we will carry out a certain number of (model theoretical) arguments in ACA_0 . We will make use of the fact that Henkin’s proof of Gödel’s completeness theorem can be formalized in ACA_0 . Here is a second characterization of interpretability over PA.

Lemma 3.4 ACA_0 proves: For every α and β , $\alpha \triangleright \beta$ iff every model of $\text{PA} + \alpha$ has an end-extension which is a model of $\text{PA} + \beta$.

Proof. See page 1067 in [Ber90]. □

In addition to the lemmas mentioned here, we will make (implicit) use of provable Σ_1 -completeness, the recursion theorem, and the definability of primitive recursive functions in arithmetic. Now we leave the modal and arithmetical preliminaries behind and embark on the proof of the main theorem. This proof was profoundly inspired by two proofs by Zambella: the proof in [Zam94] of Shavrukov’s embeddability theorem for Magari algebras and the proof in [Zam92] of Berarducci’s and Shavrukov’s ACT for ILM. For convenience, let us restate the main theorem:

“Suppose \mathcal{E} is r. e. and well-specified. Then \mathcal{E} is realizable iff \mathcal{E} is consistent with ILM^ω .”

Proof

The implication \Rightarrow follows from the fact that all realizable sets are consistent with ILM^ω . See e.g. [Ber90].

To show the implication \Leftarrow , suppose that \mathcal{E} is recursively enumerable, well-specified, and consistent with ILM^ω . We must show that \mathcal{E} is realizable, but for practical reasons we show a slightly stronger statement. Let \mathcal{F} be the closure of \mathcal{E} under ILM^ω . We actually show that \mathcal{F} is realizable. Since $\mathcal{E} \subseteq \mathcal{F}$, this is sufficient. Our main strategy for showing that \mathcal{F} is realizable is to invoke the following claim.

Claim 3.5 Suppose that there is a recursive function i mapping modal formulas on arithmetical sentences so that

- $i(A)$ is true, for every $A \in \mathcal{F}$
- for every A and B , PA proves:
 - $i(A \triangleright B) \leftrightarrow [i(A) \triangleright i(B)]$
 - $i(\Box A) \leftrightarrow \Box(i(A))$
 - $i(A \rightarrow B) \leftrightarrow [i(A) \rightarrow i(B)]$
 - $i(\perp) \leftrightarrow \perp$.

Then \mathcal{F} is realizable.

Proof. Let i' be the unique realization satisfying $i'(p) = i(p)$, for every propositional variable p . Then an easy induction on formulas shows that PA proves $i(A) \leftrightarrow i'(A)$, for every A . But $i(A)$ is true, for every $A \in \mathcal{F}$. Hence, so is $i'(A)$. Consequently \mathcal{F} is realizable. \square

We will eventually define a function i satisfying the premises of Claim 3.5. The definition of i consists of one modal part (a maximal consistent superset of \mathcal{F}) and one arithmetical part (a Solovay function).

First we set out to arithmetize the set \mathcal{F} . This we do with some care, since we want certain properties of the arithmetization to be provable in PA. To that end we begin by introducing an auxiliary modal concept.

- \mathcal{G} is *disjunctive* if the following holds for every n and every choice of A_i and B_i , for $i \leq n$.
 - $\mathcal{G} \not\vdash \perp$,
 - $\mathcal{G} \vdash \bigvee \{A_i \triangleright B_i : i \leq n\} \Rightarrow \exists i \leq n (\mathcal{G} \vdash A_i \triangleright B_i)$.

Claim 3.6 PA proves: if \mathcal{G} is finite and well-specified, then \mathcal{G} is disjunctive.

Proof. We want to prove a connection between a local syntactical property (well-specified) and a global property that involves quantification over infinitely many formulas (disjunctive).

Argue in PA: “Suppose \mathcal{G} is finite and well-specified, but not disjunctive. Then we may assume that $\mathcal{G} \vdash \bigvee \{A_i \triangleright B_i : i \leq n\}$ and $\mathcal{G} \not\vdash A_i \triangleright B_i$, for every $i \leq n$. By Lemma 3.1(iii) there are models k_i witnessing that $\mathcal{G} \not\vdash A_i \triangleright B_i$, for every $i \leq n$.

Let k be the model obtained by first taking the disjoint union of all the models k_i , then identifying their roots, and finally letting the root of the resulting model inherit its forcing-relation from k_0 . Note that k forces the negation of $\bigvee\{A_i \triangleright B_i : i \leq n\}$.

Now let us use induction on formulas to show that $k_0 \Vdash A$ iff $k \Vdash A$, for every $A \in S(\mathcal{G})$.

- The basis and the Boolean cases are easy.
- $A = B \triangleright C$. First note that $k_0 \Vdash \neg(B \triangleright C) \Rightarrow k \Vdash \neg(B \triangleright C)$. Conversely, suppose that $k \Vdash \neg(B \triangleright C)$. Then some k_i forces $\neg(B \triangleright C)$. Consequently $B \triangleright C \notin \mathcal{G}$ and since \mathcal{G} is well-specified, $\neg(B \triangleright C) \in \mathcal{G}$. Hence k_0 forces $\neg(B \triangleright C)$.
- $A = \Box B$. This case is similar to the previous one.

This completes the induction. Hence k_0 and k agree on $S(\mathcal{G})$. But k_0 is a $\bigwedge \mathcal{G}$ -reflexive model forcing all the formulas of \mathcal{G} . Hence, so is k .

Thus by Lemma 3.1(ii), starting with k we may iterate the derived model operator until we get a model witnessing that $\mathcal{G} \not\models \bigvee\{A_i \triangleright B_i : i \leq n\}$. Contradiction. \square

Now we are ready to arithmetize \mathcal{F} .

Claim 3.7 There is a primitive recursive function Q such that, for each n , $Q(n)$ is finite set of modal formulas and

- (i) $A \in \mathcal{F} \Leftrightarrow \exists n(Q(n) \vdash A)$,
- (ii) PA proves: every $Q(n)$ is disjunctive,
- (iii) PA proves: $\forall n(Q(n) \subseteq Q(n+1))$.

Proof. Since \mathcal{E} is r. e. and well-specified by assumption, there is a recursive function Q' such that $\mathcal{E} = \bigcup Q'(n)$, $Q'(0) = \emptyset$, $Q'(n) \subseteq Q'(n+1)$, where $Q'(n)$ is well-specified, for every n . By using a standard trick if necessary, we may even assume that Q' is primitive recursive. Since \mathcal{E} is consistent with ILM^ω , we also have $Q'(n) \not\models \perp$, for every n . Define Q as follows:

Stage 0.
$Q(0) := Q'(0)$.
Stage $n+1$.
1. $Q'(n+1) \not\models \perp$, $Q'(n+1)$ is well-specified, and $Q(n) \subseteq Q'(n+1)$.
$Q(n+1) := Q'(n+1)$.
2. Otherwise.
$Q(n+1) := Q(n)$.

First note that Q is primitive recursive by Lemma 3.1(v).

- (i) Clearly, $Q(n) = Q'(n)$ for every n . Hence $\mathcal{E} = \bigcup Q(n)$ and so (i) follows.
- (ii) An easy induction in PA shows that every $Q(n)$ is well-specified. Hence by Claim 3.6, every $Q(n)$ is disjunctive.
- (iii) This can also be shown by an easy induction in PA.

\square

In view of Claim 3.7(i) we may henceforth consider $A \in \mathcal{F}$ to be an abbreviation of $\exists n(Q(n) \vdash A)$. Thus, by Lemma 3.1(v), $A \in \mathcal{F}$ is a Σ_1 sentence.

Our next goal is to define a formula Z encoding a maximal consistent superset of \mathcal{F} .

- The modal formula A is *conservative* if for all n and all sequences A_0, \dots, A_n and B_0, \dots, B_n we have

$$(A \rightarrow \bigvee \{A_i \triangleright B_i : i \leq n\}) \in \mathcal{F} \Rightarrow \exists i \leq n (A_i \triangleright B_i \in \mathcal{F}).$$

- Let C_1, C_2, \dots be an effective enumeration of all the formulas of interpretability logic. We will construct formulas $\{Z_m : m \in \omega\}$ as follows. Let $Z_0 = \top$. Define

$$Z_{m+1} = \begin{cases} Z_m \wedge C_{m+1} & \text{if this sentence is conservative} \\ Z_m \wedge \neg C_{m+1} & \text{otherwise.} \end{cases}$$

Let $Z(A)$ be the sentence $\exists m(\vdash Z_m \rightarrow A)$.

Claim 3.8 Fix A and B . Then PA proves the following properties of Z :

- (Z1) $\neg Z(\perp)$,
- (Z2) $Z(A \rightarrow B) \leftrightarrow [Z(A) \rightarrow Z(B)]$,
- (Z3) $A \in \mathcal{F} \rightarrow Z(A)$,
- (Z4) $Z(A \triangleright B) \rightarrow A \triangleright B \in \mathcal{F}$.

Proof. Argue in PA: “First note that $Z_0 = \top$ is conservative, since every $Q(n)$ is disjunctive by Claim 3.7(ii). Suppose there is a smallest number m such that Z_{m+1} is not conservative. Then neither $Z_m \wedge C_{m+1}$ nor $Z_m \wedge \neg C_{m+1}$ is conservative. From this it is straight-forward to see that Z_m is also not conservative. This is a contradiction, and thus we may conclude that all the Z_m are conservative.

- (Z1) Suppose $Z(\perp)$. Then $\vdash Z_m \rightarrow \perp$, for some m . Hence $\vdash Z_m \rightarrow \top \triangleright \perp$. Since \mathcal{F} is closed under ILM^ω , $Z_m \rightarrow \top \triangleright \perp \in \mathcal{F}$. But Z_m is conservative, so $\top \triangleright \perp \in \mathcal{F}$. A short proof in ILM^ω then shows that $\perp \in \mathcal{F}$. Contradiction.
- (Z2) This is clear, since Z is maximal when viewed as a set.
- (Z3) Suppose $A \in \mathcal{F}$. If $\neg Z(A)$, then $Z(\neg A)$ follows from (Z1) and (Z2). Hence $\vdash Z_m \rightarrow \neg A$, for some m . Since Z_m is conservative, it follows that $\neg A \in \mathcal{F}$. Consequently $\perp \in \mathcal{F}$. Contradiction.
- (Z4) Suppose $Z(A \triangleright B)$. Then $\vdash Z_m \rightarrow A \triangleright B$, for some m . Therefore, $(Z_m \rightarrow A \triangleright B) \in \mathcal{F}$. But Z_m is conservative, so $A \triangleright B \in \mathcal{F}$.”

□

This concludes the modal part of the construction. Next we will proceed with the arithmetical part and define a Solovay function f using the recursion theorem. To that end we first introduce some abbreviations that refer to the gödelnumber of f , and then we define f itself.

- Let $\lambda_0 := \forall n(f(n) = 0)$, $\lambda_a :=$ “the limit of f is the node a ”. Recall that no nodes are encoded by 0. Thus $a \neq 0$ holds automatically.
- Let $\lambda_A :=$ “the limit of f is a node that forces A (in its uniquely determined model)” and $\lambda_{a,b}(n) := \forall m \geq n(f(m) = a \vee f(m) = b)$.
- Let $\text{rank}(a, n)$ be the smallest number s such that $I\Sigma_s$ proves $\neg\lambda_a$ with a proof of length $\leq n$, if such a proof exists. Otherwise $\text{rank}(a, n)$ is undefined.
- When $m \leq n$, let $f(m) = \dots = f(n)$ be short for

$$\forall n'(m \leq n' \leq n \rightarrow f(m) = f(n')).$$

- Define f as follows:

Stage 0. $f(0) := 0$.
Stage $n + 1$.

1. $f(n) = 0$, n is a proof in PA of $\lambda_A \rightarrow \Diamond_1(\lambda_B)$ and some minimal model k with root r witnesses that $Q(n) \not\vdash A \triangleright B$. Then in k some node $a \neq r$ forces A and lacks S_r -connections to nodes that force B .
 $f(n + 1) :=$ the smallest such a .
2. $f(n) \neq 0$.
 1. For some a and $m \leq n$,
 - (i) n is a proof in PA of $\neg\lambda_{f(n),a}(m)$,
 - (ii) $f(m) = \dots = f(n)$,
 - (iii) $f(n)Ra$. $f(n + 1) := a$.
 2. Case 2.1 does not apply and there is a node a and some $i < n$ such that
 - (i) $\text{rank}(a, n) \leq i < \text{rank}(f(n), n)$,
 - (ii) $f(i) \neq 0$,
 - (iii) $f(n)S_{f(i)}a$. $f(n + 1) :=$ the smallest such a .
 3. Cases 2.1 and 2.2 do not apply and there is a node a and some $i < n$ such that
 - (i) $\text{rank}(a, n) \leq i < \text{rank}(f(n), n)$,
 - (ii) $f(i) = 0$,
 - (iii) $f(n)S_r a$, where r is the root of the unique model associated with $f(n)$. $f(n + 1) :=$ the smallest such a .
3. Otherwise. $f(n + 1) := f(n)$.

Remark 3.9 Here are a few remarks about the definition of f :

1. By the recursion theorem, f is partially recursive. So, by inspecting the definition of f , we may conclude that f is primitive recursive.
2. We will soon see that f does not move in the standard model. The reason is that f can only make finitely many moves and every move presupposes that (provably in PA) more moves are to come.
3. Note that if $f(n+1) = f(n)$, then Case 3 applies at $n+1$. Also note that whenever Case 2.2 or 2.3 applies, all ranks involved are defined.

Now we are ready to bring the modal and arithmetical parts together.

- Let $i(A) := (\lambda_0 \wedge Z(A)) \vee \lambda_A$.

Remark 3.10 Before proving that i satisfies the premises of Claim 3.5, let us give some motivation for this definition of i .

- Given that λ_0 is true, (Z3) ensures that $i(A)$ is true, for every $A \in \mathcal{F}$.
- We also want i to commute with the symbols \triangleright , \rightarrow , and \perp up to provable equivalence in PA. When λ_0 is false, we proceed roughly as in [Zam92]. When λ_0 is true, we proceed as follows:
 - The Boolean connectives commute, since Z , regarded as a set, is maximal consistent by (Z1) and (Z2).
 - Suppose $i(A \triangleright B)$. Then $Z(A \triangleright B)$. To get a contradiction we assume $\neg(i(A) \triangleright i(B))$. Then, to get a concrete situation to work with, we apply Lemma 3.2 (losing no generality). Following this strategy we have to work inside a \Box -environment and furthermore, some useful consequence of the non- Σ_1 sentence $Z(A \triangleright B)$ should be transferable into that \Box -environment. Here we see the benefit of (Z4) (and ultimately the assumption that \mathcal{E} is well-specified), which permits us to conclude that $A \triangleright B \in \mathcal{F}$. To be able to move on from this point towards the desired contradiction, Case 2.3 was introduced into the definition of f .
 - Suppose $i(A) \triangleright i(B)$. Then we want to have $i(A \triangleright B)$, which is equivalent to $Z(A \triangleright B)$, and by (Z4), to $A \triangleright B \in \mathcal{F}$. So we have to avoid the situation $i(A) \triangleright i(B) \wedge (A \triangleright B \notin \mathcal{F})$. This we do by ensuring that f moves whenever such a situation arises. But guaranteeing this in the most straight-forward way would make f non-recursive. This would in turn prevent us from transferring values of f into \Box -environments. Therefore we let f move in Case 2.1 already when a decidable tendency in this direction arises. Lemma 3.2 is used to formulate this tendency. Note that the situation that must be avoided is described by a Π_2 sentence (again by Lemma 3.2), consisting of one arithmetical part and one modal part, none of which is Σ_1 . This feature distinguishes the present construction from its raw-models.

Claim 3.11 PA proves: f has a limit.

Proof. Argue in PA: “If f moves at all, then it moves into or inside some finite model k . Clearly, if h is the R-height of k , then f cannot take more than h

consecutive R -steps (steps according to 2.1). But by an easy argument that involves M , f cannot take more than h R -steps in total. Also, every time f takes an S_a -step (a step according to 2.2), or an S_r -step (a step according to 2.3), it moves to nodes of strictly lower rank at the relevant stages. Moreover, the rank of a fixed node never increases at later stages. Hence f eventually reaches a limit.” \square

Claim 3.12 PA proves: For every m , $f(m) \neq 0 \rightarrow \Box(f \text{ moves after stage } m)$.

Proof. Argue in PA: “Suppose $f(m) \neq 0$. Then f moves in some model k with root r . By the least number principle, fix $n < m$ so that

$$f(n) \neq f(n+1) = \dots = f(m).$$

We then have the following possibilities:

- Case 1 applies at $n+1$. Then there is a minimal node a and formulas A and B as in Case 1 of the definition of f . Note that $f(m) = a$. We will show that $\Box(\neg\lambda_a)$, which is more than we need. We already know that $\Box(\lambda_a \rightarrow \Diamond_1(\lambda_B))$ holds. Therefore it suffices to show that $\Box(\Box_1(\neg\lambda_B))$. Argue in PA:

“Let us show that $\Box_1(\neg\lambda_B)$. Argue in $I\Sigma_1$:

“We will show that $\neg\lambda_B$. Note that $f(n+1) = a$ and also that (in k) a lacks S_r -connections to nodes that force B . Thus it is enough to show that $n'+1 \geq n+1$ implies $aS_rf(n'+1)$, by Σ_1 induction on n' .

- Base step. Suppose $n'+1 = n+1$. Then, since S_r is reflexive, $aS_rf(n'+1)$ holds.
- Induction step. Suppose $n'+1 > n+1$.
 - * Case 2.1 applies at $n'+1$. By induction we have $aS_rf(n')$. We also have $f(n')Rf(n'+1)$ and therefore $f(n')S_rf(n'+1)$ holds by V. Hence, by the transitivity of S_r , we have $aS_rf(n'+1)$.
 - * Case 2.2 applies at $n'+1$. Then, for some $i < n'$, the step from $f(n')$ to $f(n'+1)$ is an $S_{f(i)}$ -step. From this we draw two conclusions. First, that $f(i)Rf(n'+1)$ holds. Since it is easy to see that $f(i) \neq r$, we also have $rRf(i)$. Thus, V yields $f(i)S_rf(n'+1)$. Secondly, since $f(i) \neq 0$ and $f(n) = 0$, we may conclude that $i \geq n+1$. Hence $aS_rf(i)$ holds by the induction hypothesis. All in all we have $aS_rf(n'+1)$.
 - * Case 2.3 applies at $n'+1$. Then the induction hypothesis yields $aS_rf(n')$. But $f(n')S_rf(n'+1)$. So we have $aS_rf(n'+1)$.
 - * Case 3 applies at $n'+1$. Then, by the induction hypothesis, we have $aS_rf(n'+1)$.”

So we have just seen that $\Box_1(\neg\lambda_B)$.”

It follows that $\Box(\Box_1(\neg\lambda_B))$, as desired.

- Case 2.1 applies at $n + 1$. Then $\Box(\neg\lambda_{f(n),a}(n'))$, for some $n' \leq n$ such that $f(n') = \dots = f(n)$. Let us show that, provably in PA, f moves after stage m . Argue in PA:

“We know that $\neg\lambda_{f(n),a}(n')$. Also, $f(n') = \dots = f(n)$ and $f(n + 1) = \dots = f(m) = a$. Together these facts imply that $f(m') \neq a$ for some $m' > m$.”

- Case 2.2 or 2.3 applies at $n + 1$. By the definition of rank, this clearly presupposes that $\Box(\neg\lambda_a)$, which, as already mentioned, is more than sufficient.”

□

Claim 3.13 f never leaves 0.

Proof. This follows from Claims 3.11 and 3.12. □

Claim 3.14 $A \in \mathcal{F} \Rightarrow i(A)$ is true.

Proof. Suppose $A \in \mathcal{F}$. Then $Z(A)$ is true, by (Z3) and the truth of the theory PA. But by Claim 3.13, λ_0 is also true. Consequently $i(A)$ is true. □

Let X be a model of PA. Since f is a function in X , the sequence $f(0), f(1), \dots$ is defined in X . By Claim 3.11 it converges. Whenever f moves from one node to another it moves by an R -step, an S_a -step or an S_r -step. The *orbit* generated by f in X is the above sequence, but with commas and consecutive repetitions omitted and with symbols for the relations that are responsible for the transitions inserted. For example, $0aRbS_r cS_a b$ could be an orbit.

The following is a strengthening of Claim 3.12, the proof of which invokes Claim 3.12.

Claim 3.15 PA proves: $\forall a[f \text{ visits } a \neq 0 \rightarrow \Box(\bigvee\{\lambda_b : aRb\})]$.

Proof. Argue in ACA_0 : “Suppose $f(m) = a \neq 0$. Then f moves in some model k , with root r , say. By Gödel’s completeness theorem, it suffices to show that $\bigvee\{\lambda_b : aRb\}$ holds in an arbitrary model X of PA. Taking advantage of our view of X from the outside, we first note that m is standard in X . Let b be the limit of f in X . We show that aRb holds:

Let Ω be the orbit of f in X . By Claim 3.12, f will change its value in X at some stage $> m$. Let b' be the immediate predecessor of the final occurrence of b in Ω . Then $a = b'$ is a possibility. Keeping that in mind, Ω may be written

$$0 \dots \dots \dots a \dots \dots \dots b' ? b,$$

where $?$ is an unknown relation-symbol, $b' \neq b$, and the exhibited a occurs at stage m .

Suppose the last step in Ω occurs at $n + 1$. Thus $f(n) = b'$ and $f(n + 1) = b$ hold in X . We have the following possibilities in X .

- Case 2.1 applies at $n+1$. Then $b'Rb$. Hence an easy induction that utilizes M yields aRb .
- Case 2.2 applies at $n+1$. Then there is a number $i < n$ such that

- (i) $\text{rank}(b, n) \leq i < \text{rank}(b', n)$,
- (ii) $f(i) \neq 0$,
- (iii) $b'S_{f(i)}b$.

Since b is the limit of f in X , Lemma 3.3 implies that $\text{rank}(b, n)$ is non-standard in X . Thus, by (i), so is the number i . However, m is standard in X , as already mentioned. Hence, in Ω , $f(m) = a$ precedes $f(i)$ or equals $f(i)$. On the other hand, $f(i)Rb$ holds by (iii). Thus we have the relations $a \cdots f(i)Rb$, where the dotted line is as in Ω . Hence we have aRb by the transitivity of R and by M .

- Case 2.3 applies at $n+1$. Then there is a number $i < n$ such that

- (i) $\text{rank}(b, n) \leq i < \text{rank}(b', n)$,
- (ii) $f(i) = 0$,
- (iii) $b'S_r b$.

By Lemma 3.3, $\text{rank}(b, n)$ is non-standard in X . Thus, by (i), so is the number i . But according to (ii), f has not left 0 at i . On the other hand, recall that m is standard in X and $f(m) \neq 0$. This is a contradiction. Hence Case 2.3 does not apply at $n+1$.

□

In particular, by Claim 3.15, PA proves: for every a , if f visits a , then $\Box(\neg\lambda_a)$. Hence PA proves that the rank of any node that f visits eventually becomes defined.

Claim 3.16 Fix A and B . Then PA proves:

- $i(A \triangleright B) \leftrightarrow [i(A) \triangleright i(B)]$
- $i(\Box A) \leftrightarrow \Box(i(A))$
- $i(A \rightarrow B) \leftrightarrow [i(A) \rightarrow i(B)]$
- $i(\perp) \leftrightarrow \perp$.

Proof. We prove this claim by combining several subclaims.

Subclaim 1 $\text{PA} + \lambda_0 \vdash i(A \triangleright B) \rightarrow i(A) \triangleright i(B)$.

Proof. The formula $\Box\neg A \rightarrow \neg A$ is an axiom of ILM^ω . Since \mathcal{F} is closed under ILM^ω , this implies that $A \rightarrow \Diamond A \in \mathcal{F}$. Argue in ACA_0 : “Suppose λ_0 and that the implication is false. Then, to begin with, we have $i(A \triangleright B)$ and thus $Z(A \triangleright B)$ and hence $A \triangleright B \in \mathcal{F}$, by property (Z4). Furthermore, since $i(A) \triangleright i(B)$ does not hold, Lemma 3.2 guarantees that there is a number s and a model X of PA in which $i(A)$ and $\Box_s(\neg i(B))$ hold. Note that $f(s) = 0$ and that s is standard in X .

- f moves in X . Then f moves in X within some model k with root r , say. Let a be the limit of f in X . Since $i(A)$ holds in X , we know that $a \Vdash A$ and it is not hard to see that $r \Vdash A \triangleright B$. Thus there is a minimal node b such that $aS_r b$ and $b \Vdash B$. Since f visits a in X , the formalization of Claim 3.15 guarantees that the rank of a eventually gets defined in X . Also, in X we have $\Box_s(\neg i(B))$ and thus $\Box_s(\neg \lambda_B)$. But $b \Vdash B$, so in X we have $\text{rank}(b, n) \leq s$ when n is large enough. Hence we can fix n so that the following holds in X : $\text{rank}(a, n)$ is defined, $\text{rank}(b, n) \leq s$ and f never leaves a after stage n . In particular, Cases 2.1 and 2.2 never apply in X after stage n . Note that $\text{rank}(a, n)$ is non-standard in X by Lemma 3.3. Thus all in all (again taking advantage of our view on X from the outside) we can conclude that the conditions $f(s) = 0$ and $\text{rank}(b, n) \leq s < \text{rank}(a, n)$ hold in X . Since we also have $aS_r b$, it follows that, in X , Case 2.3 applies at $n+1$ and $f(n+1) = b$. Hence $f(n+1) \neq a$ holds in X . Contradiction.
- f does not move in X . Then $Z(A)$ holds in X , since $i(A)$ does. Now we will use (Z1)–(Z3) several times and argue in X :

“Since $\Box \neg A \rightarrow \neg A$ is an axiom of ILM^ω , we have $A \rightarrow \Diamond A \in \mathcal{F}$ and therefore $Z(A \rightarrow \Diamond A)$. Since we also have $Z(A)$, it follows that $Z(\Diamond A)$. Moreover, $A \triangleright B \in \mathcal{F}$ and hence $Z(A \triangleright B)$. Since $(A \triangleright B) \rightarrow (\Diamond A \rightarrow \Diamond B)$ is an axiom of ILM , it follows that $Z(\Diamond B)$ holds. But then $\Box \neg B \notin \mathcal{F}$, in other words $B \triangleright \perp \notin \mathcal{F}$.”

So in X we have $B \triangleright \perp \notin \mathcal{F}$. On the other hand $\Box_s(\neg i(B))$ holds in X . So in particular $\Box(\neg i(B))$ and thereby $\Box(\lambda_B \rightarrow \Diamond_1(\lambda_\perp))$ hold in X . Hence f eventually moves in X . Contradiction.”

◇

Subclaim 2 $\text{PA} + \neg \lambda_0 \vdash i(A \triangleright B) \rightarrow i(A) \triangleright i(B)$.

Proof. Argue in ACA_0 : “Suppose that $\neg \lambda_0$ holds and the implication is false. Let a be the limit of f . Then $a \Vdash A \triangleright B$, since $i(A \triangleright B)$ holds. Furthermore $i(A) \triangleright i(B)$ does not hold. Hence by Lemma 3.4 there is a number s and a model X of PA in which λ_A and $\Box_s(\neg \lambda_B)$ hold. Then s is standard in X and we may clearly assume that $f(s) = a$.

Let b be the limit of f in X . Let m be so large that, in X , f assumes the value b at all stages $\geq m$. In particular, Case 2.1 never applies in X after stage m .

By Claim 3.15 we have aRb . But then, since $a \Vdash A \triangleright B$ and $b \Vdash A$, there is a minimal node c that forces B such that $bS_a c$. We know that $c \Vdash B$ and $\Box_s(\neg \lambda_B)$ hold in X . Hence, in X , $\text{rank}(c, n) \leq s$ when n is large enough. Now fix $n \geq m$ so that $\text{rank}(c, n) \leq s$ holds in X . By Lemma 3.3 $\text{rank}(b, n)$ is non-standard in X . However s is standard in X . Thus in X we have $\text{rank}(c, n) \leq s < \text{rank}(b, n)$. Since we also have $bS_a c$, it follows that, in X , Case 2.2 applies at $n+1$ and $f(n+1) = c$. Consequently $f(n+1) \neq b$ holds in X . Contradiction.” ◇

Subclaim 3 $\text{PA} + \lambda_0 \vdash i(A) \triangleright i(B) \rightarrow i(A \triangleright B)$.

Proof. Argue in **PA**: “Suppose that λ_0 holds and the implication is false. Then, to begin with, we have $\neg i(A \triangleright B)$ i.e. $\neg Z(A \triangleright B)$ and hence $A \triangleright B \notin \mathcal{F}$ by (Z3). Furthermore we have $i(A) \triangleright i(B)$. Hence Lemma 3.2 yields

$$\Box[i(A) \rightarrow \Diamond_1(i(B))]. \quad (1)$$

Note that by provable Σ_1 -completeness we have $\Box(\neg\lambda_0 \rightarrow \Box_1\neg\lambda_0)$. Let us show that $\Box(\lambda_A \rightarrow \Diamond_1(\lambda_B))$. Argue in **PA**:

“Suppose λ_A . Then $i(A)$ holds and thus we have $\Diamond_1(i(B))$ by (1). Moreover, λ_A implies $\neg\lambda_0$ and hence $\Box_1(\neg\lambda_0)$. Thus propositional logic yields $\Diamond_1(\lambda_B)$.”

Hence Case 1 eventually applies. So not λ_0 . Contradiction.” \diamond

Subclaim 4 $\text{PA} + \neg\lambda_0 \vdash i(A) \triangleright i(B) \rightarrow i(A \triangleright B)$.

Proof. Argue in **ACA**₀: “Suppose $\neg\lambda_0$ and that the implication is false. Then, to begin with, we have $i(A) \triangleright i(B)$. By Lemma 3.4 this implies that

$$\text{every model } X \models \text{PA} + i(A) \text{ has an end-extension } Y \models \text{PA} + i(B). \quad (2)$$

Furthermore we have $\neg i(A \triangleright B)$. Let a be the limit of f . Then $a \Vdash \neg(A \triangleright B)$. Hence there is a node b that forces A , satisfies aRb and lacks S_a -connections to nodes that force B . At some stage $n' + 1$, f will have reached a and Case 2.1 will not apply afterwards. Since aRb holds, this implies that $\Diamond(\lambda_{a,b}(n'))$. So there is a model X of **PA** in which $\lambda_{a,b}(n')$ holds. Also, since f visits a , Claim 3.15 yields $\Box(\neg\lambda_a)$. So in particular $\neg\lambda_a$ holds in X . Hence, in X , at some stage, f moves directly from a to b and then f remains on b forever.

Also, since $b \Vdash A$, λ_A and consequently $i(A)$ hold in X . Thus, by (2), X has an end-extension Y which is a model of **PA** in which $i(B)$ holds. Let us note in passing that f assumes the value a at the standard stage $n' + 1$ in Y . Now, let c be the limit of f in Y . Then $c \Vdash B$. Hence, by the choice of b , we cannot have bS_ac . Thus, to obtain the desired contradiction, it suffices to show that bS_ac :

If $b = c$, then we are done. So we may assume that $b \neq c$. Let Ω be the orbit of f in Y . Since Y is an end-extension of X , Ω looks as follows:

$$0 \cdots a?b \cdots \cdots \cdots c,$$

where $?$ is an unknown relation symbol (although we know that aRb holds). Clearly, if no relation-symbol other than S_a appears between b and c , then we are done. Thus we may assume that some node between b and c is reached by a non- S_a step. Let c' be the last of those. For later use note that $c' = c$ is a possibility and that $c'S_ac$ holds. Furthermore we may assume that $b \neq c'$ since otherwise we are done again. Then, remembering that $c' = c$ is a possibility, we may write Ω as

$$0 \cdots a?b \cdots \cdots \cdots Rc' \cdots c,$$

or as

$$0 \cdots a?b \cdots \cdots \cdots S_dc' \cdots c,$$

for some $d \neq a$.

- In the first case, we have bRc' , by M (applied zero times or more). Thus $bS_a c'$ holds by V . Also recall that $c'S_a c$. Hence the transitivity of S_a yields $bS_a c$, as desired.
- Let us consider the second case. Now we will make use of the conditions (i) of Cases 2.2 and 2.3, where stages are compared to ranks. Suppose f moves permanently to c at stage $n + 1$ in Y . Let us look closer at the step $\cdots S_d c'$ appearing in Ω . Suppose it occurs at stage $m + 1$, so that $f(m + 1) = c'$. Then Case 2.2 or 2.3 applies at stage $m + 1$. Hence there is a number i associated with the move at $m + 1$ such that $\text{rank}(c', m) \leq i$. Moreover, note that all the moves between c' and c , if any, are by Case 2.2, since a is not a root. But moves by this case always lead to nodes of strictly lower rank. Also, recall that the rank of a certain node cannot increase at later stages. Thus we have

$$\text{rank}(c, n) \leq \text{rank}(c', m) \leq i.$$

But by Lemma 3.3, $\text{rank}(c, n)$ is non-standard in Y . Hence so is i .

We have already noted that f visits a at a standard stage in Y . So in particular, f leaves 0 at a standard stage in Y . But i is non-standard in Y . Hence $f(i) \neq 0$. Consequently Case 2.3 does not apply at $m + 1$. So Case 2.2 applies at $m + 1$. In particular we have $f(i) = d$. Thus, in Y , f visits a at a standard stage and d at a non-standard stage.

Above we wrote Ω as

$$0 \cdots a?b \cdots S_d c' \cdots c.$$

At this point we also know that d occurs strictly between the displayed occurrences of a and S_d . Thus, remembering that $b = d$ and $c' = c$ are possibilities, we may write Ω as

$$0 \cdots a?b \cdots d \cdots S_d c' \cdots c.$$

(Here the point in Zambella's idea of using expressions of the kind $\lambda_{a,b}(n)$, rather than λ_a in the definition of f can be seen: the possibility that d lies strictly between a and b is excluded [Zam92].) Now, since $\cdots S_d c'$ is a step in Ω , we know that dRc' . Also aRb holds. Thus the following relations hold in k :

$$aRb \cdots dRc' \cdots c,$$

where the dotted lines are as in Ω . Thus, since $b = d$, or alternatively, by M , we have the relations

$$aRbRc' \cdots c.$$

Hence V yields $bS_a c'$. Also recall that $c'S_a c$. Thus we have $bS_a c$, as desired."

◇

Subclaim 5 $\text{PA} + \lambda_0 \vdash i(A \rightarrow B) \leftrightarrow [i(A) \rightarrow i(B)]$.

Proof. This is clear since $\text{PA} \vdash Z(A \rightarrow B) \leftrightarrow [Z(A) \rightarrow Z(B)]$, by (Z2). \diamond

Subclaim 6 $\text{PA} + \neg\lambda_0 \vdash i(A \rightarrow B) \leftrightarrow [i(A) \rightarrow i(B)]$.

Proof. This is clear since $\text{PA} + \neg\lambda_0$ proves that λ_A and $\lambda_{A \rightarrow B}$ imply λ_B , and also that $\lambda_A \rightarrow \lambda_B$ implies $\lambda_{A \rightarrow B}$. \diamond

Subclaim 7 $\text{PA} + \lambda_0 \vdash \neg i(\perp)$.

Proof. This is clear since $\text{PA} \vdash \neg Z(\perp)$ holds by (Z1). \diamond

Subclaim 8 $\text{PA} + \neg\lambda_0 \vdash \neg i(\perp)$.

Proof. This is clear since $\text{PA} \vdash \lambda_\perp \leftrightarrow \perp$. \diamond

Subclaim 9 $\text{PA} \vdash i(\Box A) \leftrightarrow \Box(i(A))$.

Proof. Using what we have established at this point, together with the readily verifiable fact that $\Box A \leftrightarrow (\neg A) \triangleright \perp$ is a theorem of ILM and its arithmetical counterpart $\Box \alpha \leftrightarrow (\neg \alpha) \triangleright \perp$ is a theorem of PA , the subclaim follows. \diamond

Putting all the subclaims together, Claim 3.16 follows. \blacksquare

Finally, by Claim 3.14 and Claim 3.16, we may use Claim 3.5 to conclude that \mathcal{F} is realizable. Since $\mathcal{E} \subseteq \mathcal{F}$, it follows that \mathcal{E} is also realizable. This concludes the proof of the main theorem.

4 Corollaries

Now we can show some easy corollaries of the main theorem. The following two results were obtained independently by Berarducci [Ber90] and Shavrukov [Sha88].

Corollary 4.1 (Berarducci, Shavrukov) A is consistent with ILM^ω iff $\{A\}$ is realizable.

Proof. Suppose A is consistent with ILM^ω . Then by Lemma 3.1(iii), A has an A -reflexive model k . Let $\mathcal{E} = \{B \in S(A) : k \Vdash B\}$. By Lemma 3.1, \mathcal{E} is consistent with ILM^ω . So by the main theorem, \mathcal{E} is realizable. Hence $\{A\}$ is also realizable. The converse implication is not hard, see e.g. [Ber90]. \square

Corollary 4.2 (Berarducci, Shavrukov) $\vdash A$ iff $\text{PA} \vdash i(A)$, for every realization i .

Proof. This proof was pointed out to me by Per Lindström. Suppose $\not\vdash A$. Then, by Lemma 3.1(iv), $\not\vdash \Box A$. Hence $\neg\Box A$ is consistent with ILM^ω . So by Corollary 4.1, $\{\neg\Box A\}$ is realizable. Hence $\text{PA} \not\vdash i(A)$, for some realization i . For the converse implication, see e.g. [Ber90]. \square

We can also prove a uniform version of Corollary 4.2:

Corollary 4.3 There is a realization i such that for every A , $\vdash A$ iff $\text{PA} \vdash i(A)$.

Proof. Let \mathcal{E} be the union of the sets $\{A \triangleright B : \vdash A \triangleright B\}$, $\{\Box A : \vdash \Box A\}$, $\{\neg(A \triangleright B) : \not\vdash A \triangleright B\}$, and $\{\neg\Box A : \not\vdash \Box A\}$. Then \mathcal{E} is r. e. and well-specified. It is also easy to check that it is consistent with ILM^ω . So by the main theorem, some realization i witnesses that \mathcal{E} is realizable. Using Lemma 3.1(iv), note that $\vdash A \Rightarrow \vdash \Box A \Rightarrow \Box A \in \mathcal{E} \Rightarrow \text{PA} \vdash i(A)$ and $\not\vdash A \Rightarrow \not\vdash \Box A \Rightarrow \neg\Box A \in \mathcal{E} \Rightarrow \text{PA} \not\vdash i(A)$. \square

For the last corollary, we need a definition:

- Let an *Orey set* be a set \mathcal{E} such that for all A , if $A \in \mathcal{E}$, then A is either of the form $B \triangleright C$ or of the form $\neg(B \triangleright C)$, for some formulas of propositional logic B and C .

In [Ore61], Orey asks a question that could be phrased thus in our present terminology: “What (finite) Orey sets are realizable?”. We can answer this question as follows:

Corollary 4.4 Suppose \mathcal{E} is an r. e. Orey set. Then \mathcal{E} is realizable iff it is consistent with ILM^ω .

Proof. This follows from the main theorem, since Orey sets are well-specified. \square

Remark 4.5 It can be shown that finite Orey sets \mathcal{G} that are consistent with ILM^ω have $\bigwedge \mathcal{G}$ -reflexive models of a particularly simple kind. Those models consist of a root r and one single layer of nodes above r . Moreover, the nodes above r may be partitioned into groups inside of which the relation S_r is total, and outside of which S_r is empty.

Remark 4.6 It is straight-forward to verify that all the above results also hold with ZF in the place of PA (and the Gödel-Bernay set theory GB in the place of ACA_0).

References

- [Ber90] A. Berarducci. The interpretability logic of Peano arithmetic. *The Journal of Symbolic Logic*, 55:1059–1089, 1990.
- [dJV90] D. de Jongh and F. Veltman. Provability logics for relative interpretability. In [Pet90], pages 31–42, 1990.
- [HP91] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic*. Perspectives in Mathematical Logic. Springer, Berlin, 1991.
- [JdJ98] G. Japaridze and D. de Jongh. The logic of provability. In S. Buss, editor, *Handbook of Proof Theory*. Elsevier, 1998.
- [Ore61] S. Orey. Relative interpretations. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 7:146–153, 1961.

- [Pet90] P. Petkov, editor. *Mathematical logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria*. Plenum Press, Boston, 1990.
- [Sha88] V. Shavrukov. The logic of relative interpretability over Peano arithmetic (in Russian). Technical Report 5, Stekhlov Mathematical Institute, Moscow, 1988.
- [Sha93] V. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes mathematicae (Rozprawy matematyczne)*, CCCXXIII, 1993.
- [Smo85] C. Smoryński. *Self-Reference and Modal Logic*. Universitext. Springer, New York, 1985.
- [Sol76] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 25:287–304, 1976.
- [Str96] C. Strannegård. *Arithmetical realizations of modal formulas*. PhD thesis, University of Göteborg, 1996.
- [Vis90] A. Visser. Interpretability logic. In [Pet90], pages 175–209, 1990.
- [Vis97] A. Visser. An overview of interpretability logic. Technical Report LGPS 174, Department of Philosophy, Utrecht University, 1997.
- [Zam92] D. Zambella. On the proofs of arithmetical completeness of interpretability logic. *The Notre Dame Journal of Formal Logic*, 32:542–551, 1992.
- [Zam94] D. Zambella. Shavrukov’s theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + \text{EXP}$. *The Notre Dame Journal of Formal Logic*, 35:147–157, 1994.