Notre Dame Journal of Formal Logic
Volume 46, Number 2, 2005

The Closed Fragment of the Interpretability Logic of PRA with a Constant for $\mathrm{I}\Sigma_1$

Joost J. Joosten

Abstract In this paper we carry out a comparative study of $\mathrm{I}\Sigma_1$ and PRA. We will in a sense fully determine what these theories have to say about each other in terms of provability and interpretability. Our study will result in two arithmetically complete modal logics with simple universal models.

1 Introduction

In this paper we provide a modal logic that can decide on simple questions involving provability and interpretability over PRA and $\mathrm{I}\Sigma_1$. One should think of questions such as $\mathrm{I}\Sigma_1 \vdash^? \mathsf{Con}(\mathrm{PRA})$, $\mathrm{PRA} + \mathsf{Con}(\mathrm{PRA}) \vdash^? \mathrm{I}\Sigma_1$, $\mathrm{PRA} + \mathsf{Con}(\mathrm{PRA}) \rhd^? \mathrm{PRA} + \mathsf{Con}(\mathrm{I}\Sigma_1) + \neg\mathrm{I}\Sigma_1$, $\mathrm{I}\Sigma_1 \rhd^? \mathrm{PRA} + \mathsf{Con}(\mathrm{PRA})$, $\mathrm{I}\Sigma_1 + \mathsf{Con}(\mathrm{I}\Sigma_1) \rhd^? \mathrm{PRA} + \mathsf{Con}(\mathsf{Con}(\mathrm{PRA}))$, and so on. As we shall see, some quite interesting questions can be formulated in the logics we give.

In Section 3 we shall first compute the closed fragment of the provability logic of PRA with a constant for $\mathrm{I}\Sigma_1$. The full provability logic of PRA with a constant for $\mathrm{I}\Sigma_1$ actually has already been determined in Beklemishev [1]. We give an elementary proof here so that we can extend it when computing the closed fragment of the interpretability logic of PRA with a constant for $\mathrm{I}\Sigma_1$ in Section 4.

1.1 Interpretations Interpretations in the form we will consider them have been around for quite a while in common mathematical practice. A good example is the interpretation of non-Euclidean geometry in Euclidean geometry. As a metamathematical tool, interpretations were first introduced in full generality in Tarski et al. [26] where they were used to show relative consistency and undecidability of theories.

The notion of interpretability that we will study is essentially the same as in [26]. Thus, an interpretation of a theory T in a theory S is nothing more than a structure preserving translation of formulas of T to formulas of S such that the translation

Received June 20, 2004; accepted October 11, 2004; printed May 25, 2005
2000 Mathematics Subject Classification: Primary, 03F45, 03F25, 03F30
Keywords: interpretability logic, provability logic

of any theorem of T is provable in S. In case such a translation exists we say that S interprets T or that T is interpretable in S and write $S \rhd T$. As in [26] we are interested in relative interpretability. This means that in S we have a domain function $\delta(x)$ to which all our quantifiers are restricted/relativized. A precise and formal definition of relative interpretability can be found in, for example, Japaridze and de Jongh [11] or Visser [30]. In these references and especially in Visser [28] the formalization of interpretability is studied. This gives rise to interpretability logics with a binary modal operator $\rhd$ for formalized interpretability.

1.2 Collection Many of the interesting properties of interpretability are only provable in the presence of the Σ_1-collection principle $\mathsf{B}\Sigma_1$. Our base theory PRA lacks $\mathsf{B}\Sigma_1$ and thus, for example,

$$(\mathrm{PRA} \cup \{\alpha\}) \rhd (\mathrm{PRA} \cup \{\beta\}) \to (\mathsf{Con}(\mathrm{PRA} \cup \{\alpha\}) \to \mathsf{Con}(\mathrm{PRA} \cup \{\beta\}))$$

is not provable in PRA by the standard argument. And it is actually an open question if this is provable at all in PRA. We will thus talk rather of *smooth interpretability* as introduced in [28]. This notion of interpretability can be seen as the notion where the needed collection has been built in by defining it accordingly. When we speak of interpretablility we will in this paper always mean the smooth version.

In the presence of $\mathsf{B}\Sigma_1$ the two versions of interpretability coincide. Moreover, for finitely axiomatizable theories T we have that interpretability and smooth interpretability in U coincide. We refer the reader to Hájek and Pudlák [8] and Buss [5] for the arithmetical principles and theories that we use in this paper.

1.3 Interpretability logics Just as in the case of provability logics we have that a modal sentence $A \rhd B$ is a valid principle for a theory T if for any arithmetical realization $*$ holds $T \vdash (T \cup \{A^*\}) \rhd (T \cup \{B^*\})$. Often $T + A^*$ will be written instead of $T \cup \{A^*\}$. Sometimes we will write $A^* \rhd_T B^*$ for $(T + A^*) \rhd (T + B^*)$. We will denote both the modal operator and the formalized notion of smooth interpretability by the same symbol $\rhd$ but this will hardly lead to any confusion.

As the definition of interpretability invokes that of provability it does not come as a surprise that interpretability and provability logics are closely related. As a matter of fact, provability logics are literally included in the interpretability logics.

Definition 1.1 The logic **IL** is the smallest set of formulas being closed under the rules of necessitation and of modus ponens that contains all tautological formulas and all instantiations of the following axiom schemata.

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	$(A \rhd B) \wedge (B \rhd C) \to A \rhd C$
J3	$(A \rhd C) \wedge (B \rhd C) \to A \vee B \rhd C$
J4	$A \rhd B \to (\Diamond A \to \Diamond B)$
J5	$\Diamond A \rhd A$

The interpretability logic **IL** is a sort of basic interpretability logic. All other interpretability logics we consider shall be extensions of **IL** by further principles. Principles we shall consider in this paper are among the following.

$$\begin{array}{lcl}
\mathsf{F} & := & (A \rhd \Diamond A) \rightarrow \Box\neg A \\
\mathsf{W} & := & (A \rhd B) \rightarrow (A \rhd (B \wedge \Box\neg A)) \\
\mathsf{M} & := & (A \rhd B) \rightarrow ((A \wedge \Box C) \rhd (B \wedge \Box C)) \\
\mathsf{P} & := & (A \rhd B) \rightarrow \Box(A \rhd B)
\end{array}$$

If X is a set of axiom schemata we will denote by $\mathbf{IL}\mathsf{X}$ the logic that arises by adding the axiom schemata in X to **IL**. Thus, $\mathbf{IL}\mathsf{X}$ is the smallest set of formulas being closed under the rules of modus ponens and necessitation and containing all tautologies and all instantiations of the axiom schemata of **IL** ($\mathsf{L1}-\mathsf{J5}$) and of the axiom schemata of X.

The interpretability logic for essentially reflexive theories has been proved to be $\mathbf{IL}\mathsf{M}$, independently in Berarducci [3] and Shavrukov [20]. Also the situation is known for finitely axiomatized theories in which case the logic is $\mathbf{IL}\mathsf{P}$ (Visser [27]).

No interpretability logic is known for a theory that is neither essentially reflexive nor finitely axiomatizable. PRA is such a theory. Thus we find it interesting to investigate the interpretability logic of this theory. More insight into the interpretability logic of PRA, from now on **IL**(PRA), can also shed some light on the question what interpretability principles hold in any reasonable theory as studied in Joosten and Visser [13].

In this paper we constrain ourselves to the closed fragment of **IL**(PRA), that is, modal formulas without propositional variables. It is shown in Hájek and Švejdar [9] that for any interpretability logic extending $\mathbf{IL}\mathsf{F}$, the interpretability closed fragment coincides with the provability closed fragment. It is easily seen that **IL**(PRA) indeed does extend $\mathbf{IL}\mathsf{F}$.

1.4 A comparison to other papers We have chosen to add an extra constant to our closed fragment that denotes the sentence axiomatizing $\mathrm{I}\Sigma_1$. By writing $\mathrm{I}\Sigma_1$ we will refer both to the finitely axiomatizable theory and to the finite axiom axiomatizing it. We can thus study what these theories have to say about each other's provability and interpretability behavior.

In this respect our enterprise is rather akin to a certain part of Beklemishev's paper [1] on the classification of bimodal logics. As an example he gives the provability logic (not just the closed fragment) of PRA with a constant for $\mathrm{I}\Sigma_1$. The closed fragment of this logic is just the logic **PGL** which we present in Section 3. We have chosen to give explicit proofs for the correctness and completeness of **PGL** again, so that we can easily extend them to the situation where interpretability is added to the vocabulary in Section 4.

This paper also is reminiscent of Visser's paper on exponentiation [29]. In that paper the closed fragment of the interpretability logic of the arithmetical theory Ω is presented. (The theory Ω is also known as $\mathrm{I}\Delta_0 + \Omega_1$.) The modal language is enriched with an additional constant exp. The arithmetical translation of this constant is the Π_2-formula stating the totality of the exponential function.

A fundamental difference between Visser's [29] and our paper is that although $\mathrm{I}\Sigma_1$ is a proper extension of PRA, no new recursive functions are proved to be total, as $\mathrm{I}\Sigma_1$ is a Π_2-conservative extension of PRA. In this sense the gap between PRA and $\mathrm{I}\Sigma_1$ is smaller than the gap between Ω and $\Omega + \mathsf{exp}$. This difference is also manifested already in the corresponding logics when we just constrain ourselves to provability. For example, we have that

$$\mathrm{PRA} + \mathsf{Con}(\mathrm{PRA}) \vdash \mathsf{Con}(\mathrm{I}\Sigma_1),$$

whereas

$$\Omega + \mathsf{Con}(\Omega) \nvdash \mathsf{Con}(\Omega + \mathsf{exp}).$$

Actually even $\Omega+\mathsf{exp}+\mathsf{Con}(\Omega)$ does not prove $\mathsf{Con}(\Omega+\mathsf{exp})$. It does hold however that $\Omega+\mathsf{Con}(\mathsf{Con}(\Omega)) \vdash \mathsf{Con}(\Omega+\mathsf{exp})$ and there are more similarities. We have that $\mathsf{Con}(\mathrm{PRA})$ is not provable in $\mathrm{I}\Sigma_1$. Similarly, $\mathsf{Con}(\Omega)$ is not provable in $\Omega + \mathsf{exp}$. In turn, $\mathrm{I}\Sigma_1$ is not provable in PRA together with any iteration of consistency statements and the same holds for exp and Ω.[1]

The interpretability logics have similarities and differences too. For example, we have that $\mathrm{PRA} \rhd \mathrm{PRA} + \neg\mathrm{I}\Sigma_1$ and $\Omega \rhd \Omega + \neg\mathsf{exp}$. Also $\mathrm{PRA} + \mathsf{Con}(\mathrm{PRA}) \rhd \mathrm{I}\Sigma_1$ and $\Omega + \mathsf{Con}(\Omega) \rhd \Omega + \mathsf{exp}$. On the other hand, $\mathrm{I}\Sigma_1 \not\rhd \mathrm{PRA} + \mathsf{Con}(\mathrm{PRA})$ whereas $\Omega + \mathsf{exp} \rhd \Omega + \mathsf{Con}(\Omega)$. However, we do have that $\mathrm{I}\Sigma_1 \rhd \Omega + \mathsf{Con}(\mathrm{PRA})$. We have that $\mathrm{I}\Sigma_1 \not\rhd \mathrm{PRA} + \mathsf{Con}(\mathrm{PRA})$ but PRA itself cannot see this. PRA can only see that $\mathrm{I}\Sigma_1 \rhd \mathrm{PRA} + \mathsf{Con}(\mathrm{PRA}) \to \neg\mathsf{Con}(\mathrm{PRA})$.

The differences between the pairs of theories is probably best reflected by the corresponding universal models. The interested reader is advised to compare the universal models from this paper to the ones from [29].

2 Preliminaries

In this section we describe the central notions that we shall study in this paper. Also we agree on some notational conventions.

2.1 Arithmetics The base theory in this enterprise is PRA which is a system of arithmetic that goes by many different formulations. We will briefly mention these formulations here and then stick to one of them. In a rudimentary form, PRA was first introduced in Skolem [22]. The emergence of PRA is best understood in the light of Hilbert's program and finitism (see Tait [25]) or instrumentalism as Ignjatovic calls it in [10].

Since Π_1-sentences or open formulas played a prominent role in Hilbert's program, the first versions of PRA were formulated in a quasi-equational setting without quantifiers but with a symbol for every primitive recursive function. (See, for example, Goodstein [7] or Schwartz [18] and [19].)

Other formulations are in the full language of predicate logic and also contain a function symbol for every primitive recursive function. The amount of induction can either be for Δ_0-formulas or for open formulas. Both choices yield the same set of theorems. This definition of PRA has, for example, been used in Smoryński [23].[2]

In this paper we will associate to each arithmetical theory T in a uniform way a proof predicate $\Box_T$ as is done in Feferman [6]. Thus, we will also have the obvious properties of this predicate like $\Box_{T+\varphi}\psi \leftrightarrow \Box_T(\varphi \to \psi)$ available in any theory of some reasonable minimal strength. We will also extensively make use of reflection principles.

For a theory T and a class of formulas Γ we define the uniform reflection principle for Γ over T to be a set of formulas in the following way:

$$\mathrm{RFN}_\Gamma(T) := \{\forall x\ (\Box_T\gamma(\dot{x}) \to \gamma(x)) \mid \gamma \in \Gamma\}.$$

This set of formulas is often equivalent to a single formula also denoted by $\mathrm{RFN}_\Gamma(T)$. For ordinals $\alpha \leq \omega$ we define $(T)^\Gamma_0 := T$, $(T)^\Gamma_{\alpha+1} := (T)^\Gamma_\alpha + \mathrm{RFN}_\Gamma((T)^\Gamma_\alpha)$, and $(T)^\Gamma_\omega := \cup_{\beta<\omega}(T)^\Gamma_\beta$. This can be extended to transfinite ordinals provided an elementary system of ordinal notation is given. If Γ is just the class of Π_n formulas we write $(T)^n_\alpha$ instead of $(T)^{\Pi_n}_\alpha$.

For some purposes it is not convenient that these definitions of PRA are in a language properly extending the language of PA. An alternative way to define PRA is as follows. We can define PRA to be EA + Σ_1−IR which is formulated in the language of PA and is obtained by adding to EA the induction rule for Σ_1 formulas. Thus, for $\sigma \in \Sigma_1$, the Σ_1 induction rule allows you to conclude $\forall x\ \sigma(x)$ from $\sigma(0)$ and $\forall x\ (\sigma(x) \to \sigma(x+1))$. The theory EA is just $\mathrm{I}\Delta_0 + \mathsf{exp}$. It is folklore that PRA and Σ_1−IR are in a sense the same theory. The theories EA + Σ_n−IR are defined likewise and we denote them by $\mathrm{I}\Sigma^R_n$.

In Beklemishev [2] it is shown (for $n \geq 1$) that $\mathrm{I}\Sigma^R_n$ can be axiomatized by reflection principles in the following sense, $\mathrm{I}\Sigma^R_n = (\mathrm{EA})^{n+1}_\omega$ (as sets of theorems). All the above definitions of PRA give rise to the same theory and these equivalences are all provable in PRA itself. In our approach we will take $(\mathrm{EA})^2_\omega$ to be the definition of PRA. It turns out that this is a very convenient formulation for us. It is also nice that this is an axiomatic formulation in the language of PA.

Moreover, we will fix an enumeration of the axioms of PRA. It is known that EA is finitely axiomatizable. Since we have partial truth definitions and we are talking global reflection we have that $\{\forall x\ (\Box_{\mathrm{EA}}\pi(\dot{x}) \to \pi(x)) \mid \pi \in \Pi_2\}$ can be expressed by a single sentence $\mathrm{RFN}_{\Pi_2}(\mathrm{EA})$. Likewise we see that $(\mathrm{EA})^2_\alpha$ can be expressed by a single sentence for any $\alpha < \omega$. In our enumeration of PRA, the ith axiom will be $(\mathrm{EA})^2_i$.

By taking this definition of PRA we get almost for free that every extension of PRA with a Σ_2 sentence σ is reflexive. For, reason in PRA + σ and suppose $\Box_{\mathrm{PRA}\restriction n+\sigma}\bot$. Then $\Box_{\mathrm{PRA}\restriction n}\neg\sigma$, and as $\neg\sigma$ is Π_2 we get $\neg\sigma$ by Π_2-reflection. But this contradicts σ whence $\neg\Box_{\mathrm{PRA}\restriction n+\sigma}\bot$.

2.2 Reading conventions When writing modal formulas we will omit superfluous brackets. These omissions do not bring the unique readability of formulas to danger due to our binding conventions. The strongest binding connectives are negation and the modalities $\Box$ and $\Diamond$. The connectives $\vee$ and $\wedge$ bind less strongly but still more strongly than the $\rhd$ modality which in its turn binds more strongly than $\to$. We will also omit outer brackets. Thus, $A \rhd B \to A \wedge \Box\neg C \rhd B \wedge \Box\neg C$ is short for $((A \rhd B) \to ((A \wedge \Box(\neg C)) \rhd (B \wedge \Box(\neg C))))$. Often we will use $A \rhd B \rhd C$ as short for $(A \rhd B) \wedge (B \rhd C)$ and we do the same for implication.

3 The Closed Fragment of the Provability Logic of PRA with a Constant for $\mathrm{I}\Sigma_1$.

In this section we will calculate the closed fragment of the provability logic of PRA with a constant for $\mathrm{I}\Sigma_1$ and call it **PGL**. We shall prove it sound and complete with respect to its arithmetical reading. Also we will give a universal model for **PGL**.

3.1 The logic PGL Inductively we define F, the formulas of **PGL**.

$$F := \ \bot \mid \top \mid \mathsf{S} \mid F \wedge F \mid F \vee F \mid F \to F \mid \neg F \mid \Box F.$$

The symbol S is a constant in our language just as $\bot$ is a constant. There are no propositional variables. As always we will use $\Diamond A$ as an abbreviation for $\neg\Box\neg A$. We define $\Box^0\bot := \bot$ and $\Box^{n+1}\bot := \Box(\Box^n\bot)$. We also define $\Box^\gamma\bot$ to be $\top$ for limit ordinals γ.

Throughout this section we shall reserve $B, B_0, B_1, \ldots$ to denote Boolean combinations of formulas of the form $\Box^n\bot$ with $n \in \omega + 1$.

Definition 3.1 (The logic PGL) The formulas of the logic **PGL** are given by F. The logic **PGL** is the smallest normal extension of **GL** in this language that contains the following two axiom schemes.

$$\begin{array}{ll} \mathsf{S}_1: & \Box(\mathsf{S} \to B) \to \Box B \\ \mathsf{S}_2: & \Box(\neg\mathsf{S} \to B) \to \Box B \end{array}$$

It is good to emphasize that **PGL** is a variable free logic. By our notational convention both in S_1 and in S_2 the B is a Boolean combination of formulas of the form $\Box^n\bot$ with $n \in \omega + 1$. Immediate consequences of S_1 and S_2 are that both $\Diamond(\mathsf{S} \wedge B)$ and $\Diamond(\neg\mathsf{S} \wedge B)$ are equivalent in **PGL** to $\Diamond B$.

Every sentence in F can also be seen as an arithmetical statement as follows: we translate S to the canonical sentence $\mathrm{I}\Sigma_1$ (the single sentence axiomatizing the theory $\mathrm{I}\Sigma_1$), $\bot$ to, for example, 0=1, and $\top$ to 1=1. As usual we inductively extend this translation to what is sometimes called an arithmetical interpretation by taking for the translation of $\Box$ the canonical proof predicate for PRA.

If there is no chance of confusion we will use the same letter to indicate both a formal sentence of **PGL** and the arithmetical statement expressed by it. With this convention we can formulate the main theorem of this subsection.

Theorem 3.2 *For all sentences $A \in F$ we have*

$$\mathrm{PRA} \vdash A \Leftrightarrow \mathbf{PGL} \vdash A.$$

Proof The implication '$\Leftarrow$' is proved in Subsection 3.2 in Corollary 3.3 and Lemma 3.4. The other direction is proved in Subsection 3.3, in Lemma 3.5. $\Box$

3.2 Arithmetical soundness of PGL To see the arithmetical soundness of **PGL**, we should check only the validity of S_1 and S_2. Axiom S_1 can be seen as a direct consequence of the formalization of Parsons' theorem (Parsons [15], [16]). As is pointed out, for example, in the first proof of Joosten [12], the proof of Parsons' theorem essentially relies on cut elimination. The proof can thus be formalized as soon as the totality of the superexponential function is provable.

Corollary 3.3 $\mathrm{PRA} \vdash \Box_{\mathrm{PRA}}(\mathrm{I}\Sigma_1 \to B) \to \Box_{\mathrm{PRA}} B$ *for $B \in \Pi_2$ and thus certainly whenever B is as in* S_1.

Lemma 3.4 $\mathrm{EA} \vdash \forall^{\Pi_3} B\ (\Box_{\mathrm{PRA}}(\neg\mathrm{I}\Sigma_1 \to B) \to \Box_{\mathrm{PRA}} B)$.

Proof It is well known that $\mathrm{I}\Sigma_n \vdash \mathrm{RFN}_{\Pi_{n+2}}(\mathrm{EA})$. (See, for example, Leivant [14] or [8].) Consequently, the formalization of $\mathrm{I}\Sigma_1 \vdash \mathrm{RFN}_{\Pi_3}(\mathrm{EA})$ is a true Σ_1-sentence and thus provable in EA. As $\mathrm{EA} \vdash \Box_{\mathrm{I}\Sigma_1}(\mathrm{RFN}_{\Pi_3}(\mathrm{EA}))$ we also have

$$(*) \qquad \mathrm{EA} \vdash \Box_{\mathrm{EA}}(\mathrm{I}\Sigma_1 \to \mathrm{RFN}_{\Pi_3}(\mathrm{EA})).$$

Now we reason in EA, fix some $B\in\Pi_3$, and assume $\Box_{\mathrm{PRA}}(\neg \mathrm{I}\Sigma_1 \to B)$. We get

$$\begin{array}{rll}
\Box_{\mathrm{PRA}}(\neg \mathrm{I}\Sigma_1 \to B) & \to & \\
\Box_{\mathrm{PRA}}(\neg B \to \mathrm{I}\Sigma_1) & \to & \\
\exists \pi\in\Pi_2\ \Box_{\mathrm{EA}}(\neg B \wedge \pi \to \mathrm{I}\Sigma_1) & \to & \text{by } (*) \\
\exists \pi\in\Pi_2\ \Box_{\mathrm{EA}}(\neg B \wedge \pi \to \mathrm{RFN}_{\Pi_3}(\mathrm{EA})) & \to & \text{as } B \vee \neg\pi \in \Pi_3 \\
\exists \pi\in\Pi_2\ \Box_{\mathrm{EA}}(\neg B \wedge \pi \to (\Box_{\mathrm{EA}}(B \vee \neg\pi) \to B \vee \neg\pi)) & & (**)
\end{array}$$

But, by simple propositional logic, we also have

$$\Box_{\mathrm{EA}}(\neg(\neg B \wedge \pi) \to (\Box_{\mathrm{EA}}(B \vee \neg\pi) \to B \vee \neg\pi))$$

which combined with $(**)$ yields $\Box_{\mathrm{EA}}(\Box_{\mathrm{EA}}(B \vee \neg\pi) \to (B \vee \neg\pi))$. By Löb's axiom we get $\Box_{\mathrm{EA}}(B \vee \neg\pi)$ which is the same as $\Box_{\mathrm{EA}}(\pi \to B)$. Thus certainly we have $\Box_{\mathrm{PRA}}B$, as π was just a part of PRA. □

We note that Lemma 3.4 actually holds for a wider class of formulas than just Boolean combinations of $\Box^{\alpha}\bot$ formulas. For example, $\neg(A \rhd B)$ is always Π_3. One can also isolate a set of sentences that is always Π_2 in PRA. (See, for example, [30].) When we study the logic **PIL** it will become clear why we only need to include these low-complexity instantiations of the above arithmetical facts in our axiomatic systems: in the closed fragment we have simple normal forms.

3.3 Arithmetical completeness of PGL

Lemma 3.5 *For all A in F we have that if* $\mathrm{PRA} \vdash A$*, then* $\mathbf{PGL} \vdash A$*.*

Proof The completeness of **PGL** actually boils down to an exercise in normal forms in modal logic. The only arithmetical ingredients are the soundness of **PGL**, the fact that $\mathrm{PRA} \vdash \Box A$ whenever $\mathrm{PRA} \vdash A$, and the fact that $\mathrm{PRA} \nvdash \Box^{\alpha}\bot$ for $\alpha \in \omega$.

In Lemma 3.7 we will show that $\Box A$ is always equivalent in **PGL** to $\Box^{\alpha}\bot$ for some $\alpha \in \omega{+}1$. Then, in Lemma 3.8, we show that if $\mathbf{PGL} \vdash \Box A$ then $\mathbf{PGL} \vdash A$. So, if $\mathbf{PGL} \nvdash A$ then $\mathbf{PGL} \nvdash \Box A$. As $\mathbf{PGL} \vdash \Box A \leftrightarrow \Box^{\alpha}\bot$ for some $\alpha \in \omega$ (not $\omega{+}1$ as we assumed $\mathbf{PGL} \nvdash \Box A$!) and **PGL** is sound we also have $\mathrm{PRA} \vdash \Box A \leftrightarrow \Box^{\alpha}\bot$. Hence $\mathrm{PRA} \nvdash \Box A$ and also $\mathrm{PRA} \nvdash A$. □

We work out the exercise in modal normal forms. Although this is already carried out in the literature (see, e.g., Boolos [4] or [29]) we repeat it here to obtain some subsidiary information which we shall need later on.

Recall that we will in this subsection reserve the letters B, B_0, $B_1, \ldots$ for Boolean combinations of $\Box^{\alpha}\bot$-formulas. Thus a sentence B can be written in conjunctive normal form, that is,

$$\textstyle\bigwedge\!\!\bigwedge_i (\bigvee\!\!\bigvee_j \neg\Box^{a_{ij}}\bot \vee \bigvee\!\!\bigvee_k \Box^{b_{ik}}\bot).$$

Each conjunct $\bigvee\!\!\bigvee_j \neg\Box^{a_{ij}}\bot \vee \bigvee\!\!\bigvee_k \Box^{b_{ik}}\bot$ can be written as $\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot$ where $\alpha_i := \min(\{a_{ij}\})$ and $\beta_i := \max(\{b_{ik}\})$.

By convention the empty conjunction is just $\top$ and the empty disjunction is just $\bot$. In order to have this convention in concordance with our normal forms we define $\min(\varnothing)=\omega$ and $\max(\varnothing)=0$. In $\bigwedge\!\!\bigwedge_i (\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$ we can leave out the conjuncts whenever $\alpha_i \leq \beta_i$, for in that case, $\mathbf{PGL} \vdash \Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot$.

So, if we say that some formula B is in conjunctive normal form we will in the sequel assume that B is written as $\bigwedge\!\!\bigwedge_i (\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$ with $\alpha_i > \beta_i$. The empty conjunction gives $\top$ and if we take $\alpha_0 = \omega > 0 = \beta_0$, we get with one conjunct just $\bot$.

Lemma 3.6 *If a formula B can be written in the form $\bigwedge\!\!\!\bigwedge_i(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$ with $\alpha_i > \beta_i$, then we have that* **PGL** $\vdash \Box B \leftrightarrow \Box^{\beta+1}\bot$ *where* $\beta = \min(\{\beta_i\})$.

Proof The proof is actually carried out in **GL**. We have that

$$\Box B \leftrightarrow \Box(\bigwedge\!\!\!\bigwedge_i(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)) \leftrightarrow \bigwedge\!\!\!\bigwedge_i \Box(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot).$$

We will see that $\Box(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$ is equivalent to $\Box^{\beta_i+1}\bot$.

So we assume $\Box B$. As $\beta_i < \alpha_i$ we know that $\beta_i + 1 \leq \alpha_i$ and thus $\Box^{\beta_i+1}\bot \to \Box^{\alpha_i}\bot$. Now $\Box(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot) \to \Box(\Box^{\beta_i+1}\bot \to \Box^{\beta_i}\bot)$. One application of L_3 yields $\Box(\Box^{\beta_i}\bot)$, that is, $\Box^{\beta_i+1}\bot$.

On the other hand, we easily see that $\Box(\Box^{\beta_i}\bot) \to \Box(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$, hence we have shown the equivalence. Finally we remark that $(\bigwedge\!\!\!\bigwedge_i \Box^{\beta_i+1}\bot) \leftrightarrow \Box^{\beta+1}\bot$ where $\beta = \min(\{\beta_i\})$. $\Box$

Lemma 3.7 *For any formula A in F we have that A is equivalent in* **PGL** *to a Boolean combination of formulas of the form $\mathtt{S}$ or $\Box^{\beta}\bot$. If, on top of that, A is of the form $\Box C$, then A is equivalent in* **PGL** *to $\Box^{\alpha}\bot$, for some $\alpha \in \omega + 1$.*

Proof By induction on the complexity of formulas in F. The base cases are trivial. The only interesting case in the induction is where we consider the case that $A = \Box C$. Note that C, by induction being a Boolean combination of $\Box^{\alpha}\bot$ formulas and $\mathtt{S}$, can be written as $(\mathtt{S} \to B_0) \wedge (\neg\mathtt{S} \to B_1)$. So, by Lemma 3.6, we have that, for suitable indices β, β', β'',

$$\begin{array}{cc}
\Box C & \leftrightarrow \\
\Box((\mathtt{S} \to B_0) \wedge (\neg\mathtt{S} \to B_1)) & \leftrightarrow \\
\Box(\mathtt{S} \to B_0) \wedge \Box(\neg\mathtt{S} \to B_1) & \leftrightarrow \\
\Box B_0 \wedge \Box B_1 & \leftrightarrow \\
\Box^{\beta'+1}\bot \wedge \Box^{\beta''+1}\bot & \leftrightarrow \\
\Box^{\beta}\bot. &
\end{array}$$

$\Box$

Lemma 3.8 *If* **PGL** $\vdash \Box A$, *then* **PGL** $\vdash A$.

Proof By Lemma 3.7, we can write A as a Boolean combination of formulas of the form $\mathtt{S}$ or $\Box^{\beta}\bot$. Thus let $A \leftrightarrow (\mathtt{S} \to B_0) \wedge (\neg\mathtt{S} \to B_1)$ with B_0 and B_1 in conjunctive normal form and assume $\vdash \Box A$. For appropriate indices $\alpha_i > \beta_i$ and $\alpha'_j > \beta'_j$ we have $B_0 = \bigwedge\!\!\!\bigwedge_i(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$ and $B_1 = \bigwedge\!\!\!\bigwedge_j(\Box^{\alpha'_j}\bot \to \Box^{\beta'_j}\bot)$. Using S_1, S_2 and Lemma 3.6, we get that $\Box A \leftrightarrow \Box^{\beta+1}\bot$ with $\beta = \min(\{\beta_i, \beta'_j\})$. By assumption $\beta = \omega$, thus all the β_i and β'_j were ω and hence $\vdash A$. $\Box$

3.4 Modal semantics for PGL, decidability In this subsection we will provide a modal semantics for **PGL**. Actually we will give a model $\mathcal{M}$ as depicted in Figure 1 on the next page which in some sense displays all there is to know about closed sentences with a constant for $\mathrm{I}\Sigma_1$ in **PGL**.

Definition 3.9 We define the model $\mathcal{M}$ as follows: $\mathcal{M} := \langle M, R, \Vdash\rangle$. Here $M := \{\langle n, i\rangle \mid n \in \omega,\ i \in \{0, 1\}\}$ and $\langle n, i\rangle R\langle m, j\rangle \Leftrightarrow m < n$. Furthermore, $\langle n, i\rangle \Vdash \mathtt{S} \Leftrightarrow i = 1$.

Theorem 3.10 $\forall \mathbf{m}\ \mathcal{M}, \mathbf{m} \Vdash A \Leftrightarrow$ **PGL** $\vdash A$.

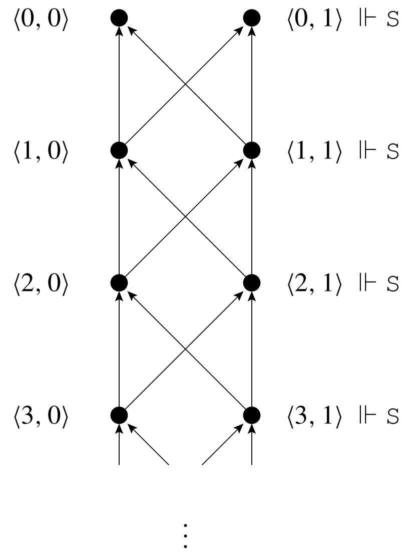

Figure 1 The model $\mathcal{M}$

Proof ($\Leftarrow$) This direction is obtained by induction on the complexity of proofs in **PGL**. As $\mathcal{M}$ is a transitive and upward well-founded model, it is indeed a model of all instantiations of the axioms L_1, L_2, and L_3. Thus consider S_1.

So suppose at some world $\mathbf{m}$ $(= \langle m, i\rangle)$, we have that $\langle m, i\rangle \Vdash \Box(\mathtt{S} \to B)$. Then $\langle n, 1\rangle \Vdash B$ for $n < m$. Recall that B does not contain $\mathtt{S}$. It is well known that the forcing of B depends solely on the depth of the world, so we also have $\langle n, 0\rangle \Vdash B$. Thus $\mathbf{m}R\mathbf{n}$ yields $\mathbf{n} \Vdash B$. Consequently, $\mathbf{m} \Vdash \Box B$, which gives us the validity of S_1.

The S_2-case is treated completely similarly. It is also clear that this direction of the theorem remains valid under applications of both modus ponens and the necessitation rule.

($\Rightarrow$) Suppose **PGL** $\nvdash A$. By Lemma 3.8, **PGL** $\nvdash \Box A$, thus **PGL** $\vdash \Box A \leftrightarrow \Box^{\alpha}\bot$ for a certain $\alpha \in \omega$. By the first part of this proof we may conclude that $\mathbf{m} \Vdash \Box A \leftrightarrow \Box^{\alpha}\bot$ for any $\mathbf{m}$. As $\langle \alpha, i\rangle \nVdash \Box^{\alpha}\bot$, we automatically get $\langle \alpha, i\rangle \nVdash \Box A$. So, for some $\langle \beta, j\rangle$ with $\langle \alpha, i\rangle R \langle \beta, j\rangle$, we have $\langle \beta, j\rangle \Vdash \neg A$ showing the "nonvalidity" of A. $\Box$

The set of theorems of **PGL** is clearly recursively enumerable. If a formula is not provable in **PGL**, then by Theorem 3.10, in some node of the model $\mathcal{M}$, it is refuted. Thus the theoremhood of **PGL** is actually decidable.

4 Closed Fragment of Interpretability Logic of PRA with a Constant for $\mathrm{I}\Sigma_1$

In this section we calculate the closed fragment of the interpretability logic of PRA with a constant for $\mathrm{I}\Sigma_1$ and call it **PIL**. We shall give two different arithmetical soundness proofs. In one of these proofs we need that $\mathrm{I}\Sigma_1$ proves the consistency of PRA on a definable cut. This itself will also be proven in a more general theorem.

The logic **PIL** contains **PGL** as a sublogic, and also the universal model for **PIL** that we shall give in this section is an extension of the model we defined in Subsection 3.4. We conclude this section by characterizing the always true sentences of our language I.

4.1 The logic PIL Inductively we define I, the formulas of **PIL**.

$$I := \quad \bot \mid \top \mid \mathtt{S} \mid I \wedge I \mid I \vee I \mid I \to I \mid \neg I \mid \Box I \mid I \rhd I.$$

Again the constants of the language are $\bot$, $\top$, and $\mathtt{S}$, and we will reserve the symbols, $B, B_0, B_1, \ldots$ to denote Boolean combinations of $\Box^\alpha \bot$ formulas. We will write $C \equiv D$ as short for $(C \rhd D) \wedge (D \rhd C)$ and we say that they are equi-interpretable.

Definition 4.1 (The logic PIL) The formulas of the logic **PIL** are given by I. The logic **PIL** is the smallest normal extension of **ILW** in this language that contains the following four axiom schemes.

$$\begin{array}{ll} \mathsf{S}_1: & \Box(\mathtt{S} \to B) \to \Box B \\ \mathsf{S}_2: & \Box(\neg\mathtt{S} \to B) \to \Box B \\ \mathsf{S}_3: & \neg\mathtt{S} \wedge B \equiv B \\ \mathsf{S}_4: & (B \rhd \mathtt{S} \wedge B) \to \Box\neg B \end{array}$$

It is good to stress that **PIL** is a variable free logic too. As the interpretability logic **ILW** is a part of **PIL** we have access to all known reasoning in **IL** and **ILW**. In this section, unless mentioned otherwise, $\vdash$ refers to provability in **PIL**.

Fact 4.2

1. $\vdash \Box A \leftrightarrow \neg A \rhd \bot$;
2. $\vdash \Box^{\alpha+1}\bot \to \Diamond^\beta \top \rhd A \quad$ if $\alpha \leq \beta$;
3. $\vdash A \equiv A \vee \Diamond A$;
4. $\vdash A \rhd \Diamond A \to \Box\neg A$.

As an example we prove (2). We reason in **PIL** and use our notational conventions. It is sufficient to prove the case when $\alpha = \beta$. Thus,

$$\Box^{\alpha+1}\bot \to \Box(\Box^\alpha\bot) \to \Box(\neg A \to \Box^\alpha\bot) \to \Box(\Diamond^\alpha\top \to A) \to \Diamond^\alpha\top \rhd A.$$

Fact (4) is Feferman's principle and can be seen as a "coordinate-free" version of Gödel's second incompleteness theorem. It follows immediately from W realizing that $A \rhd \bot$ is by (1) nothing but $\Box\neg A$.

Again we can see any sentence in I as an arithmetical statement translating $\rhd$ as the intended arithmetization of smooth interpretability over PRA and $\Box$ as an arithmetization of provability in PRA and propagating this inductively along the structure of the formulas as usual. With this convention we can formulate the arithmetical completeness theorem for **PIL**.

Theorem 4.3 *For all sentences $A \in I$ we have* PRA $\vdash A \Leftrightarrow$ **PIL** $\vdash A$.

Proof The implication "$\Leftarrow$" is proved in the next subsection in Lemma 4.4 and Lemma 4.5. The other direction is proved in Subsection 4.4, in Lemma 4.10. $\Box$

4.2 Arithmetical soundness of PIL In [28] it has been shown that **IL**W is sound for any reasonably formulated theory extending $\mathrm{I}\Delta_0+\Omega_1$. So to check for soundness of **PIL** with respect to PRA we only need to see that all translations of S_3 and S_4 are provable in PRA.

We shall give two soundness proofs for S_3 and S_4. The first proof, consisting of Lemmas 4.4 and 4.5, uses finite approximations of theories. The second proof makes use of reflection principles and definable cuts.

Lemma 4.4 $\mathrm{PRA} \vdash B \rhd_{\mathrm{PRA}} B \wedge \neg \mathrm{I}\Sigma_1$ *for* $B{\in}\Sigma_2$*, so certainly for* B *as in* S_3.

Proof We want to show inside PRA that $\mathrm{PRA} + B \rhd \mathrm{PRA} + B + \neg\mathrm{I}\Sigma_1$. For reflexive theories U, we know that interpretability in U can be characterized in terms of provability and consistency. This characterization is known as the Orey-Hájek characterization of interpretability and reads as follows.

$$\vdash U \rhd V \leftrightarrow \forall n\ \Box_U \mathsf{Con}(V{\restriction}n).$$

From [28] it follows that for reflexive U, the Orey-Hájek characterization is actually a characterization of smooth interpretability. To prove our lemma, we need to see inside PRA that $\mathrm{PRA} + B \rhd \mathrm{PRA} + B + \neg\mathrm{I}\Sigma_1$. As we know (inside PRA) every finite Σ_2-extension of PRA is reflexive; we are by the Orey-Hájek characterization done if we can prove[3]

$$\mathrm{PRA} \vdash \forall n\ \Box_{\mathrm{PRA}+B}(\Diamond_{\mathrm{PRA}[n]+B+\neg\mathrm{I}\Sigma_1}\top). \tag{1}$$

We will set out to prove that

(i) $\mathrm{EA} \vdash \forall n\ \Box_{\mathrm{PRA}+B}(\Box_{\mathrm{PRA}[n]+B+\neg\mathrm{I}\Sigma_1}\bot \to \Box_{\mathrm{PRA}[n]+B}\bot)$,
(ii) $\mathrm{EA} \vdash \forall n\ \Box_{\mathrm{PRA}+B}(\Box_{\mathrm{PRA}[n]+B}\bot \to \bot)$,

from which (1) immediately follows.

The proof of (i) is just a slight modification of the proof of Lemma 3.4. We reason in EA and fix some n:

$$\begin{array}{lll}
\Box_{\mathrm{PRA}+B} & (& \Box_{\mathrm{PRA}[n]+B+\neg\mathrm{I}\Sigma_1}\bot \\
& \to & \Box_{\mathrm{PRA}[n]+B}\mathrm{I}\Sigma_1 \\
& \to & \Box_{\mathrm{PRA}[n]+B}\mathrm{RFN}_{\Pi_3}(\mathrm{EA}) \\
& \to & \Box_{\mathrm{EA}}(\mathrm{PRA}[n] \wedge B \to \mathrm{RFN}_{\Pi_3}(\mathrm{EA})) \\
& \to & \Box_{\mathrm{EA}}(\mathrm{PRA}[n] \wedge B \to (\Box_{\mathrm{EA}}\neg(\mathrm{PRA}[n] \wedge B) \to \neg(\mathrm{PRA}[n] \wedge B))) \\
& \to & \Box_{\mathrm{EA}}(\Box_{\mathrm{EA}}\neg(\mathrm{PRA}[n] \wedge B) \to \neg(\mathrm{PRA}[n] \wedge B)) \\
& \to & \Box_{\mathrm{EA}}\neg(\mathrm{PRA}[n] \wedge B) \\
& \to & \Box_{\mathrm{EA}}(\mathrm{PRA}[n] \to \neg B) \\
& \to & \Box_{\mathrm{PRA}[n]}\neg B \\
& \to & \Box_{\mathrm{PRA}[n]+B}\bot\ ).
\end{array}$$

The proof of (ii) is just a formalization of the fact that every finite Σ_2-extension of PRA is reflexive. So again we reason in EA. Recall that we have $\mathrm{PRA}[n]{=}(\mathrm{EA})^2_n$ in our axiomatization of PRA. Thus, by definition, $\Box_{\mathrm{PRA}[n+1]}(\Box_{\mathrm{PRA}[n]}\pi \to \pi)$ for $\pi{\in}\Pi_2$. Consequently, for our $\neg B{\in}\Pi_2$, we get $\Box_{\mathrm{PRA}[n+1]}(\Box_{\mathrm{PRA}[n]}\neg B \to \neg B)$.

Obviously we also have $\Box_{\mathrm{PRA}[n+1]+B}B$. Combining, we get a proof of (ii):

$$\begin{array}{lll}
\Box_{\mathrm{PRA}[n+1]+B} & (& \Box_{\mathrm{PRA}[n]+B}\bot \\
& \to & \Box_{\mathrm{PRA}[n]}\neg B \\
& \to & \neg B \\
& \to & \bot\ ).
\end{array}$$

□

Lemma 4.5 $\mathrm{PRA} \vdash B \rhd_{\mathrm{PRA}} B \wedge \mathrm{I}\Sigma_1 \to \Box_{\mathrm{PRA}} \neg B$ *for* $B \in \Sigma_2$*, so certainly for* B *as in* S_4.

Proof The theory $\mathrm{PRA} + B + \mathrm{I}\Sigma_1$ is, verifiably in PRA, equivalent to the finitely axiomatizable theory $\mathrm{I}\Sigma_1 + B$. Now we will reason in PRA.

We suppose that $\mathrm{PRA}+B \rhd \mathrm{PRA}+B+\mathrm{I}\Sigma_1$. As $\mathrm{PRA}+B+\mathrm{I}\Sigma_1$ is finitely axiomatizable we have that $\mathrm{PRA}[k]+B \rhd \mathrm{PRA}+B+\mathrm{I}\Sigma_1$ for some natural number k. $\mathrm{PRA}+B$ is reflexive as it is a finite Σ_2-extension of PRA and thus $\Box_{\mathrm{PRA}+B}\mathsf{Con}(\mathrm{PRA}[k]+B)$. So, certainly $\Box_{\mathrm{PRA}+B+\mathrm{I}\Sigma_1}\mathsf{Con}(\mathrm{PRA}[k]+B)$ and thus,

$$\mathrm{PRA} + B + \mathrm{I}\Sigma_1 \rhd \mathrm{PRA}[k] + B + \mathsf{Con}(\mathrm{PRA}[k] + B).$$

Consequently,

$$\mathrm{PRA}[k] + B \rhd \mathrm{PRA}[k] + B + \mathsf{Con}(\mathrm{PRA}[k] + B),$$

and by Feferman's principle we get that $\Box_{\mathrm{PRA}[k]+B}\bot$. Thus $\Box_{\mathrm{PRA}+B}\bot$ and also $\Box_{\mathrm{PRA}}(B \to \bot)$, that is, $\Box_{\mathrm{PRA}}\neg B$. $\Box$

Lemma 4.5 certainly proves the correctness of axiom scheme S_4. The proof also yields the following insights.

Corollary 4.6 *A consistent reflexive theory* U *does not interpret any finitely axiomatized theory extending it. In particular* PRA *does not interpret* $\mathrm{I}\Sigma_1$.

Corollary 4.7 $\mathrm{PRA} + \neg\mathrm{I}\Sigma_1$ *is not finitely axiomatizable.*

In the next subsection we shall give alternative proofs of Lemmas 4.4 and 4.5. A central ingredient is that $\mathrm{I}\Sigma_1$ proves the consistency of PRA on a definable cut.

4.3 $\mathrm{I}\Sigma_1$ proves the consistency of PRA on a cut

Theorem 4.8 *For each* $n \in \omega$ *with* $n \geq 1$*, there exists some* $\mathrm{I}\Sigma_n$*-cut* J_n *such that for all* Σ_{n+1}*-sentences* σ, $\mathrm{I}\Sigma_n + \sigma \vdash \mathsf{Con}^{J_n}(\mathrm{I}\Sigma_n^R + \sigma)$.

Proof From [2] it is known that $\mathrm{I}\Sigma_n^R \equiv (\mathrm{EA})_\omega^{n+1}$. Let ϵ be the arithmetical sentence axiomatizing EA. We fix the following axiomatization $\{i_m^n\}_{m\in\omega}$ of $\mathrm{I}\Sigma_n^R$:

$$i_0^n := \epsilon,$$
$$i_{m+1}^n := i_m^n \wedge \forall^{\Pi_{n+1}}\pi \ (\Box_{i_m^n}\pi \to \mathsf{True}_{\Pi_{n+1}}(\pi)).$$

The map that sends m to the code of i_m^n is clearly primitive recursive. We will assume that the context makes clear if we are talking about the formula or its code when writing i_m^n. Similarly for other formulas. An $\mathrm{I}\Sigma_n$-cut J_n is defined in the following way:

$$J_n'(x) := \forall\, y {\leq} x \ \mathsf{True}_{\Pi_{n+1}}(i_y^n).$$

We will now see that J_n' defines an initial segment in $\mathrm{I}\Sigma_n$. Clearly $\mathrm{I}\Sigma_n \vdash J_n'(0)$. It remains to show that $\mathrm{I}\Sigma_n \vdash J_n'(m) \to J_n'(m{+}1)$.

So we reason in $\mathrm{I}\Sigma_n$ and assume $J_n'(m)$. We need to show that $\mathsf{True}_{\Pi_{n+1}}(i_{m+1}^n)$, that is,

$$\mathsf{True}_{\Pi_{n+1}}(i_m^n \wedge \forall^{\Pi_{n+1}}\pi \ (\Box_{i_m^n}\pi \to \mathsf{True}_{\Pi_{n+1}}(\pi))).$$

Our assumption gives us $\mathsf{True}_{\Pi_{n+1}}(i_m^n)$; thus we need to show

$$\mathsf{True}_{\Pi_{n+1}}(\forall^{\Pi_{n+1}}\pi \ (\Box_{i_m^n}\pi \to \mathsf{True}_{\Pi_{n+1}}(\pi)))$$

or equivalently

$$\forall^{\Pi_{n+1}}\pi \ (\Box_{i_m^n}\pi \to \mathsf{True}_{\Pi_{n+1}}(\pi)).$$

The latter is equivalent to

$$\forall^{\Pi_{n+1}}\pi\ \Box_{\mathrm{EA}}(\mathsf{True}_{\Pi_{n+1}}(i_m^n) \to \mathsf{True}_{\Pi_{n+1}}(\pi)) \to \mathsf{True}_{\Pi_{n+1}}(\pi). \tag{2}$$

But as $\mathsf{True}_{\Pi_{n+1}}(i_m^n) \to \mathsf{True}_{\Pi_{n+1}}(\pi) \in \Pi_{n+2}$, and as $\mathrm{I}\Sigma_n \equiv \mathrm{RFN}_{\Pi_{n+2}}(\mathrm{EA})$, we get that

$$\forall^{\Pi_{n+1}}\pi\ \Box_{\mathrm{EA}}(\mathsf{True}_{\Pi_{n+1}}(i_m^n) \to \mathsf{True}_{\Pi_{n+1}}(\pi)) \to (\mathsf{True}_{\Pi_{n+1}}(i_m^n) \to \mathsf{True}_{\Pi_{n+1}}(\pi)).$$

We again use our assumption $\mathsf{True}_{\Pi_{n+1}}(i_m^n)$ to obtain (2). Thus indeed, $J_n'(x)$ defines an initial segment. By well-known techniques, J_n' can be shortened to a definable cut.

To finish the proof, we reason in $\mathrm{I}\Sigma_n + \sigma$ and suppose $\Box^{J_n}_{\mathrm{I}\Sigma_n^R+\sigma}\bot$. Thus for some $m \in J_n$ we have $\Box_{i_m^n \wedge \sigma}\bot$, whence also $\Box_{i_m^n}\neg\sigma$. Now $m \in J_n$, so also $m+1 \in J_n$, and thus $\mathsf{True}_{\Pi_{n+1}}(i_m^n \wedge \forall^{\Pi_{n+1}}\pi\ (\Box_{i_m^n}\pi \to \mathsf{True}_{\Pi_{n+1}}(\pi)))$.

As $\forall^{\Pi_{n+1}}\pi\ (\Box_{i_m^n}\pi \to \mathsf{True}_{\Pi_{n+1}}(\pi))$ is a standard Π_{n+1}-formula (with possibly nonstandard parameters) we see that we have the required Π_{n+1}-reflection whence $\Box_{i_m^n}\neg\sigma$ yields us $\neg\sigma$. This contradicts with σ. Thus we get $\mathsf{Con}^{J_n}(\mathrm{I}\Sigma_n^R + \sigma)$. □

Corollary 4.9 *There exists an $\mathrm{I}\Sigma_1$-cut J such that for any Σ_2 sentence σ we have $\mathrm{I}\Sigma_1 + \sigma \vdash \mathsf{Con}^J(\mathrm{PRA} + \sigma)$.*

Proof Immediate from Theorem 4.8 as $\mathrm{PRA} = \mathrm{I}\Sigma_1^R$. □

Ignjatovic has shown in his dissertation [10] that $\mathrm{I}\Sigma_1$ proves the consistency of PRA on a cut. He used this result to show that the length of PRA-proofs can be roughly superexponentially larger than the length of the corresponding $\mathrm{I}\Sigma_1$ proofs.

His reasoning was based on Pudlák [17]. Pudlák showed in this paper by model-theoretic means that GB proves the consistency of ZF on a cut. The cut that Ignjatovic exposes is actually an RCA_0-cut. (See, for example, Simpson [21] for a definition of RCA_0.)

We now give alternative proofs of Lemmas 4.4 and 4.5.

Second Proof of Lemma 4.4 We consider $B \in \Sigma_2$ and want to show in EA that $\mathrm{PRA} + B \rhd \mathrm{PRA} + B + \neg\mathrm{I}\Sigma_1$. We fix the $\mathrm{I}\Sigma_1$-cut J as given by Corollary 4.9 and reason in EA. Clearly,

$$\mathrm{PRA} + B \rhd (\mathrm{PRA} + B + (\mathrm{I}\Sigma_1 \vee \neg\mathrm{I}\Sigma_1)).$$

So we are done if we can show that $\mathrm{PRA} + B + \mathrm{I}\Sigma_1 \rhd \mathrm{PRA} + B + \neg\mathrm{I}\Sigma_1$. By Corollary 4.9 we get that $\Box_{\mathrm{I}\Sigma_1+B}\mathsf{Con}^J(\mathrm{PRA}+B)$.

Using this cut J to relativize the identity translation, we find an interpretation that witnesses $\mathrm{I}\Sigma_1 + B \rhd S^1_2 + \Diamond_{\mathrm{PRA}}B$. It is well known that Buss's S^1_2 is finitely axiomatizable (see, e.g., [8], V, 4.36), whence also $S^1_2 + \Diamond_{\mathrm{PRA}}B$ is finitely axiomatizable. Thus, interpretability and smooth interpretability are in this case the same. We now get

$$\begin{array}{lll}
\mathrm{I}\Sigma_1 + B & \rhd & \\
S^1_2 + \Diamond_{\mathrm{PRA}}B & \rhd & \text{by W} \\
S^1_2 + \Diamond_{\mathrm{PRA}}B + \Box_{\mathrm{I}\Sigma_1+B}\bot & \rhd & \\
S^1_2 + \Diamond_{\mathrm{PRA}}B + \Box_{\mathrm{PRA}}(B \to \neg\mathrm{I}\Sigma_1) & \rhd & \\
S^1_2 + \Diamond_{\mathrm{PRA}}(B + \neg\mathrm{I}\Sigma_1) & \rhd & \\
\mathrm{PRA} + B + \neg\mathrm{I}\Sigma_1. & &
\end{array}$$

□

Second Proof of Lemma 4.5 We have $B \in \Sigma_2$ and assume in EA that $\mathrm{PRA} + B \rhd \mathrm{PRA} + B + \mathrm{I}\Sigma_1$. We have already seen in the above proof that $\mathrm{PRA} + B + \mathrm{I}\Sigma_1 \rhd S^1_2 + \Diamond_{\mathrm{PRA}} B$.

Thus, by transitivity $\mathrm{PRA} + B \rhd S^1_2 + \Diamond_{\mathrm{PRA}} B$, and

$$\begin{array}{ll} \mathrm{PRA} + B & \rhd \quad \text{by W} \\ S^1_2 + \Diamond_{\mathrm{PRA}} B + \Box_{\mathrm{PRA}+B}\bot & \rhd \\ \bot. & \end{array}$$

This is the same as $\Box_{\mathrm{PRA}+B}\bot$, that is, $\Box_{\mathrm{PRA}}\neg B$. $\Box$

4.4 Arithmetical completeness of PIL This subsection is mainly dedicated to proving the next lemma.

Lemma 4.10 *For all A in I we have that if* $\mathrm{PRA} \vdash A$ *then* $\mathbf{PIL} \vdash A$.

Proof The reasoning is completely analogous to that in the proof of Lemma 3.5. We thus need to prove a Lemma 4.17 stating that for any formula A in I we have that $\Box A$ is equivalent over **PIL** to a formula of the form $\Box^\alpha\bot$, and a Lemma 4.18 which tells us that $\mathbf{PIL} \vdash A$ whenever $\mathbf{PIL} \vdash \Box A$. $\Box$

In a series of rather technical lemmas we will work up to the required lemmata. It is good to recall that in this paper B will always denote some Boolean combination of formulas of the form $\Box^\alpha\bot$.

Lemma 4.11 $\mathbf{PIL} \vdash \mathtt{S} \wedge B \equiv (\mathtt{S} \wedge \Diamond^\beta\top) \vee \Diamond^{\beta+1}\top$ *for some* $\beta \in \omega + 1$.

Proof $\mathtt{S} \wedge B \equiv (\mathtt{S} \wedge B) \vee \Diamond(\mathtt{S} \wedge B) \equiv \neg(\neg(\mathtt{S} \wedge B) \wedge \Box\neg(\mathtt{S} \wedge B))$, but

$$\neg(\mathtt{S} \wedge B) \wedge \Box\neg(\mathtt{S} \wedge B) \leftrightarrow (\mathtt{S} \to \neg B) \wedge \Box(\mathtt{S} \to \neg B) \leftrightarrow (\mathtt{S} \to \neg B) \wedge \Box\neg B.$$

Now we consider a conjunctive normal form of $\neg B$. Thus, $\neg B$ is equivalent to $\bigwedge\!\!\!\bigwedge_i (\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$ for certain $\alpha_i > \beta_i$ (possibly none). So, by Lemma 3.6, $\Box\neg B \leftrightarrow \bigwedge\!\!\!\bigwedge_i \Box^{\beta_i+1}\bot \leftrightarrow \Box^{\beta+1}\bot$ for $\beta = \min(\{\beta_i\})$. So,

$$\begin{array}{ll} (\mathtt{S} \to \neg B) \wedge \Box\neg B & \leftrightarrow \\ (\mathtt{S} \to \neg B) \wedge \Box^{\beta+1}\bot & \leftrightarrow \\ (\mathtt{S} \to \neg B) \wedge (\mathtt{S} \to \Box^{\beta+1}\bot) \wedge \Box^{\beta+1}\bot & \leftrightarrow \\ (\mathtt{S} \to (\bigwedge\!\!\!\bigwedge_i (\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot) \wedge \Box^{\beta+1}\bot)) \wedge \Box^{\beta+1}\bot. & \end{array} \tag{3}$$

As $\alpha_i > \beta_i \geq \beta$ we have $\beta + 1 \leq \alpha_i$, whence $\Box^{\beta+1}\bot \to \Box^{\alpha_i}\bot$. Thus,

$$\bigwedge\!\!\!\bigwedge_i (\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot) \wedge \Box^{\beta+1}\bot \leftrightarrow \bigwedge\!\!\!\bigwedge_i \Box^{\beta_i}\bot \leftrightarrow \Box^\beta\bot,$$

and (3) reduces to $(\mathtt{S} \to \Box^\beta\bot) \wedge \Box^{\beta+1}\bot$. Consequently,

$$\begin{array}{ll} (\mathtt{S} \wedge B) \vee \Diamond(\mathtt{S} \wedge B) & \leftrightarrow \\ \neg(\neg(\mathtt{S} \wedge B) \wedge \Box\neg(\mathtt{S} \wedge B)) & \leftrightarrow \\ \neg((\mathtt{S} \to \Box^\beta\bot) \wedge \Box^{\beta+1}\bot) & \leftrightarrow \\ (\mathtt{S} \wedge \Diamond^\beta\top) \vee \Diamond^{\beta+1}\top. & \end{array}$$

$\Box$

By a proof similar to that of Lemma 4.11 we get the following lemma.

Lemma 4.12 $\mathbf{PIL} \vdash B \equiv \Diamond^{\gamma'}\top$ *for certain* $\gamma' \in \omega + 1$.

In **PIL** we have a substitution lemma in the sense that $\vdash F(C) \leftrightarrow F(D)$ whenever $\vdash C \leftrightarrow D$. We do not have a substitution lemma for equi-interpretable formulas[4] but we do have a restricted form of it.

Lemma 4.13 *If (provably in* **PIL***) $C \equiv C'$, $D \equiv D'$, $E \equiv E'$, and $F \equiv F'$, then* **PIL** $\vdash C \vee D \rhd E \vee F \leftrightarrow C' \vee D' \rhd E' \vee F'$.

We reason in **PIL**. Suppose that $C \vee D \rhd E \vee F$. We have for any G that $C' \vee D' \rhd G \leftrightarrow (C' \rhd G) \wedge (D' \rhd G)$. As $C' \rhd C \rhd (C \vee D)$ and $D' \rhd D \rhd (C \vee D)$ we have that $C' \vee D' \rhd C \vee D$. Likewise we obtain $E \vee F \rhd E' \vee F'$ thus $C' \vee D' \rhd C \vee D \rhd E \vee F \rhd E' \vee F'$. The other direction is completely analogous.

Lemma 4.14 $\mathtt{S} \wedge \Diamond^{\alpha}\top \rhd (\mathtt{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top$ *is provably equivalent in* **PIL** *to*

$$\begin{cases} \Box^{\omega}\bot & \text{if } \alpha \geq \min(\{\beta, \gamma\}) \\ \Box^{\alpha+1}\bot & \text{if } \alpha < \beta, \gamma. \end{cases}$$

Proof The case when $\alpha \geq \min(\{\beta, \gamma\})$ is trivial as $\Diamond^{\alpha}\top \rightarrow \Diamond^{\delta}\top$ whenever $\alpha \geq \delta$. So we consider the case when $\neg(\alpha \geq \min(\{\beta, \gamma\}))$, that is, $\alpha < \beta, \gamma$. Then we have $\Diamond^{\beta}\top \rhd \Diamond^{\alpha+1}\top \rhd \Diamond(\Diamond^{\alpha}\top) \rhd \Diamond(\mathtt{S} \wedge \Diamond^{\alpha}\top)$ and likewise for $\Diamond^{\gamma}\top$ in place of $\Diamond^{\beta}\top$. Thus, together with our assumption, we get $\mathtt{S} \wedge \Diamond^{\alpha}\top \rhd (\mathtt{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top \rhd \Diamond(\mathtt{S} \wedge \Diamond^{\alpha}\top)$. By Feferman's principle we get $\Box\neg(\mathtt{S} \wedge \Diamond^{\alpha}\top)$, whence $\Box^{\alpha+1}\bot$. The implication in the other direction is immediate by Fact 4.2. $\Box$

Lemma 4.15 $\Diamond^{\alpha}\top \rhd (\mathtt{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top$ *is provably equivalent in* **PIL** *to*

$$\begin{cases} \Box^{\omega}\bot & \text{if } \alpha \geq \min(\{\beta+1, \gamma\}) \\ \Box^{\alpha+1}\bot & \text{if } \alpha < \beta+1, \gamma. \end{cases}$$

Proof The proof is completely analogous to that of Lemma 4.14 with the sole exception in the case that $\alpha = \beta < \gamma$. In this case

$$\Diamond^{\gamma}\top \rhd \Diamond^{\alpha+1}\top \rhd \Diamond(\Diamond^{\alpha}\top) \rhd \Diamond(\mathtt{S} \wedge \Diamond^{\alpha}\top) \rhd \mathtt{S} \wedge \Diamond^{\alpha}\top$$

and thus $(\mathtt{S} \wedge \Diamond^{\alpha}\top) \vee \Diamond^{\gamma}\top \rhd \mathtt{S} \wedge \Diamond^{\alpha}\top$. An application of S_4 yields the desired result, that is, $\Box^{\alpha+1}\bot$.

In case $\alpha \geq \beta + 1$ it is useful to realize that

$$\Diamond^{\alpha}\top \rhd \Diamond^{\beta+1}\top \rhd \Diamond(\Diamond^{\beta}\top) \rhd \Diamond(\mathtt{S} \wedge \Diamond^{\beta}\top) \rhd \mathtt{S} \wedge \Diamond^{\beta}\top.$$

$\Box$

Lemma 4.16 *If C and D are both Boolean combinations of $\mathtt{S}$ and sentences of the form $\Box^{\gamma}\bot$ then we have that* **PIL** $\vdash (C \rhd D) \leftrightarrow \Box^{\delta}\bot$ *for some $\delta \in \omega + 1$.*

Proof So let C and D meet the requirements of the lemma and reason in **PIL**. We get that

$$C \rhd D \leftrightarrow (\mathtt{S} \wedge B_0) \vee (\neg\mathtt{S} \wedge B_1) \rhd (\mathtt{S} \wedge B_2) \vee (\neg\mathtt{S} \wedge B_3)$$

for some B_0, B_1, B_2, and B_3. The right-hand side of this bi-implication is equivalent to

$(*)$ $((\mathtt{S} \wedge B_0) \rhd (\mathtt{S} \wedge B_2) \vee (\neg\mathtt{S} \wedge B_3)) \wedge ((\neg\mathtt{S} \wedge B_1) \rhd (\mathtt{S} \wedge B_2) \vee (\neg\mathtt{S} \wedge B_3))$.

We will show that each conjunct of $(*)$ is equivalent to a formula of the form $\Box^{\epsilon}\bot$. Starting with the left conjunct we get, by repeatedly applying Lemma 4.13, that

$$
\begin{array}{lll}
\mathsf{S} \wedge B_0 \rhd (\mathsf{S} \wedge B_2) \vee (\neg\mathsf{S} \wedge B_3) & \leftrightarrow & \text{Lemma 4.11} \\
(\mathsf{S} \wedge \Diamond^{\alpha}\top) \vee \Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge B_2) \vee (\neg\mathsf{S} \wedge B_3) & \leftrightarrow & \mathsf{S}_3 \\
(\mathsf{S} \wedge \Diamond^{\alpha}\top) \vee \Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge B_2) \vee B_3 & \leftrightarrow & \text{Lemma 4.12} \\
(\mathsf{S} \wedge \Diamond^{\alpha}\top) \vee \Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge B_2) \vee \Diamond^{\gamma'}\top & \leftrightarrow & \text{Lemma 4.11} \\
(\mathsf{S} \wedge \Diamond^{\alpha}\top) \vee \Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\beta+1}\top \vee \Diamond^{\gamma'}\top & \leftrightarrow & \\
(\mathsf{S} \wedge \Diamond^{\alpha}\top) \vee \Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top & \leftrightarrow & \\
(\mathsf{S} \wedge \Diamond^{\alpha}\top \rhd (\mathsf{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top) \quad \wedge & & \\
(\Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top) & \leftrightarrow & \text{Lemma 4.14} \\
\Box^{\mu}\bot \wedge (\Diamond^{\alpha+1}\top \rhd (\mathsf{S} \wedge \Diamond^{\beta}\top) \vee \Diamond^{\gamma}\top) & \leftrightarrow & \text{Lemma 4.15} \\
\Box^{\mu}\bot \wedge \Box^{\lambda}\bot & \leftrightarrow & \\
\Box^{\delta}\bot & &
\end{array}
$$

for suitable indices $\alpha, \beta, \ldots$. For the right conjunct of $(*)$ we get a similar reasoning. □

Lemma 4.16 is the only new ingredient needed to prove the next two lemmas in complete analogy to their counterparts 3.7 and 3.8 in **PGL**.

Lemma 4.17 *For any formula A in I we have that A is equivalent in* **PIL** *to a Boolean combination of formulas of the form* S *or* $\Box^{\beta}\bot$. *If, on top of that, A is of the form $\Box C$, then A is equivalent in* **PIL** *to* $\Box^{\alpha}\bot$ *for some* $\alpha \in \omega + 1$.

Lemma 4.18 *For all A in I we have that* **PIL** $\vdash A$ *whenever* **PIL** $\vdash \Box A$.

4.5 Modal semantics for PIL, decidability As in the case of **PGL**, we shall define a universal model for the logic **PIL**. We shall use the well-known notion of Veltman semantics for interpretability logic. A Veltman model is a pair $\langle M, S\rangle$. Here M is just a **GL**-model. The S is a ternary relation on M. We shall write S as a set of indexed binary relations. On Veltman models, for all x, the S_x is a binary relation on all the worlds that lie above (w.r.t. the R-relation) x. It is reflexive and transitive and extends R on the domain on which it is defined. The forcing of formulas is extended to interpretability by the following clause:

$$x \Vdash A \rhd B \Leftrightarrow \forall y\ (xRy \Vdash A \Rightarrow \exists z\ (yS_x z \Vdash B)).$$

Definition 4.19 (Universal model for PIL) The model $\mathcal{N} = \langle M, R, \{S_m\}_{m\in M}, \Vdash\rangle$ is obtained from the model $\mathcal{M} = \langle M, R, \Vdash\rangle$ as defined in Definition 3.9 as follows. We define $\langle m, 1\rangle S_{\mathbf{n}} \langle m, 0\rangle$ for $\mathbf{n}R\langle m, 1\rangle$ and close off so as to have the $S_{\mathbf{n}}$ relations reflexive and transitive and containing R, the amount it should.

Theorem 4.20 $\forall \mathbf{n}\ \ \mathcal{N}, \mathbf{n} \Vdash A \Leftrightarrow$ **PIL** $\vdash A$.

Proof The proof is completely analogous to that of Theorem 3.10. We only need to check that all the instantiations of S_3 and S_4 hold in all the nodes of $\mathcal{N}$.

We first show that S_3 holds at any point $\mathbf{n}$. So, for any B, consider any point $\langle m, i\rangle$ such that $\mathbf{n}R\langle m, i\rangle \Vdash B$. As $\langle m, i\rangle S_{\mathbf{n}} \langle m, 0\rangle$, we see that $\mathbf{n} \Vdash B \rhd B \wedge \neg\mathsf{S}$.

To see that any instantiation of S_4 holds at any world $\mathbf{n}$ we reason as follows. If $\mathbf{n} \Vdash \Diamond B$ we can pick the minimal $m \in \omega$ such that $(m, 0) \Vdash B$. It is clear that no $S_{\mathbf{n}}$-transition goes to a world where $B \wedge \mathsf{S}$ holds, hence $\mathbf{n} \Vdash \neg(B \rhd B \wedge \mathsf{S})$. □

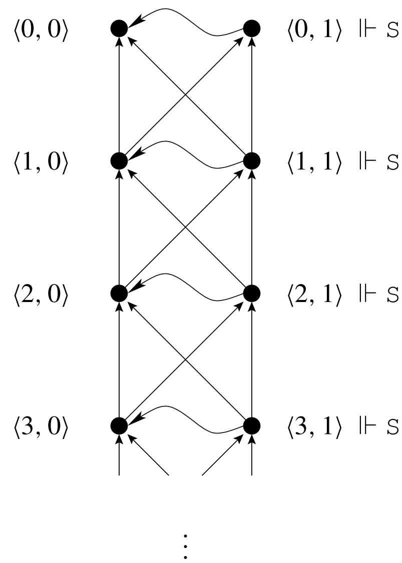

Figure 2 The (simplified) model $\mathcal{N}$

The modal semantics gives us the decidability of the logic **PIL**. In our case it is very easy to obtain a so-called simplified Veltman model. This is a model $\langle M, R, S, \Vdash \rangle$ where S now is a binary relation. Accordingly we define

$$x \Vdash A \rhd B \Leftrightarrow \forall y\ (xRy \Vdash A \Rightarrow \exists z\ (ySz \Vdash B)).$$

Our model $\mathcal{N}$ is transformed into a simplified Veltman model by defining $\mathbf{n}S\mathbf{m} \Leftrightarrow \exists \mathbf{k}\ \mathbf{n}S_{\mathbf{k}}\mathbf{m}$. A perspicuous picture is readily drawn. The S-relation is depicted with a wavy arrow.

4.6 Adding reflection Just as always, if we want to go from all provable statements to all true statements, we have only to add reflection. As we are in the closed fragment and as we have good normal forms, this reflection will amount to iterated consistency statements.

The logics **PGLS** and **PILS** are defined as follows. The axioms of **PGLS** (respectively, **PILS**) are all the theorems of **PGL** (respectively, **PILS**) together with S and $\{\Diamond^{\alpha}\top \mid \alpha \in \omega\}$. Its sole rule of inference is modus ponens.

Theorem 4.21 $\mathbf{PGLS} \vdash A \Leftrightarrow \mathbb{N} \models A$.

Proof By induction on the length of $\mathbf{PGLS} \vdash A$ we see that $\mathbf{PGLS} \vdash A \Rightarrow \mathbb{N} \models A$. To see the converse, we reason as follows. Consider $A \in F$ such that $\mathbb{N} \models A$. By Lemma 3.7 we can find an A' which is a Boolean combination of S and $\Diamond^{\alpha}\top$ ($\alpha \in \omega{+}1$) such that $\mathbf{PGL} \vdash A \leftrightarrow A'$. Thus $\mathrm{PRA} \vdash A \leftrightarrow A'$ and also $\mathbb{N} \models A \leftrightarrow A'$. Consequently $\mathbb{N} \models A'$.

Moreover, as A' is a Boolean combination of S and $\Diamond^{\alpha}\top$ ($\alpha \in \omega + 1$), for some $m \in \omega$, $\mathsf{S} \wedge \bigwedge_{i=1}^{m} \Diamond^{i}\top \rightarrow A'$ is a propositional logical tautology whence A' is provable in **PGLS**. Also $\mathbf{PGLS} \vdash A \leftrightarrow A'$ whence $\mathbf{PGLS} \vdash A$. □

Clearly the theorems of **PGLS** are recursively enumerable. As **PGLS** is a complete logic in the sense that it either refutes a formula or proves it, we see that theoremhood of **PGLS** is actually decidable.

Theorem 4.22 **PILS** $\vdash A \Leftrightarrow \mathbb{N} \models A$.

Proof As the proof of Theorem 4.21. □

Clearly, **PILS** is a decidable logic too.

Notes

1. It is well known that $\mathrm{I}\Sigma_1 \equiv \mathrm{RFN}_{\Pi_3}(\mathrm{EA})$ and that $\mathrm{I}\Sigma_1$ is not contained in any Σ_3-extension of EA. Consistency statements are all Π_1-sentences. For the case of Ω and exp reason as follows. Take any nonstandard model of true arithmetic together with the set $\{2^c > \omega_1^k(c) \mid k \in \omega\}$. Take the smallest set containing c being closed under the ω_1 function. Consider the initial segment generated by this set. This initial segment is a model of Ω and of all true Π_1 sentences but clearly not closed under exp.

2. Confusingly enough Smoryński later defines in [24] a version of PRA which is equivalent to $\mathrm{I}\Sigma_1$.

3. PRA[n] will denote the conjunction of the first n axioms of PRA. Here “first n axioms” refers to the order fixed in Subsection 2.1.

4. We have that $\neg \mathsf{S} \equiv \top$. If the substitution lemma were to hold for equi-interpretable formulas then $\mathsf{S} \equiv \neg(\neg \mathsf{S}) \equiv \bot$ which will turn out not to be the case.

References

[1] Beklemishev, L. D., “Bimodal logics for extensions of arithmetical theories,” *The Journal of Symbolic Logic*, vol. 61 (1996), pp. 91–124. Zbl 0858.03024. MR 97c:03051. 127, 129

[2] Beklemishev, L. D., “Induction rules, reflection principles, and provably recursive functions,” *Annals of Pure and Applied Logic*, vol. 85 (1997), pp. 193–242. Zbl 0882.03055. MR 98h:03072. 131, 138

[3] Berarducci, A., “The interpretability logic of Peano arithmetic,” *The Journal of Symbolic Logic*, vol. 55 (1990), pp. 1059–89. Zbl 0725.03037. MR 92f:03066. 129

[4] Boolos, G., *The Logic of Provability*, Cambridge University Press, Cambridge, 1993. Zbl 0891.03004. MR 95c:03038. 133

[5] Buss, S. R., “First-order proof theory of arithmetic,” pp. 79–147 in *Handbook of Proof Theory*, vol. 137 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1998. Zbl 0911.03029. MR 2000b:03208. 128

[6] Feferman, S., “Arithmetization of metamathematics in a general setting,” *Fundamenta Mathematicae*, vol. 49 (1960), pp. 35–92. Zbl 0095.24301. 130

[7] Goodstein, R. L., *Recursive Number Theory: A Development of Recursive Arithmetic in a Logic-free Equation Calculus*, North-Holland Publishing Company, Amsterdam, 1957. Zbl 0077.01401. MR 21:1272. 130

[8] Hájek, P., and P. Pudlák, *Metamathematics of First-Order Arithmetic*, Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1993. Zbl 0781.03047. MR 94d:03001. 128, 132, 139

[9] Hájek, P., and V. Švejdar, "A note on the normal form of closed formulas of interpretability logic," *Studia Logica*, vol. 50 (1991), pp. 25–28. Zbl 0728.03015. MR 93i:03023. 129

[10] Ignjatovic, A. D., *Fragments of First and Second Order Arithmetic and Length of Proofs*, Ph.D. thesis, University of California, Berkeley, 1990. 130, 139

[11] Japaridze, G., and D. de Jongh, "The logic of provability," pp. 475–546 in *Handbook of Proof Theory*, vol. 137 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1998. Zbl 0915.03019. MR 2000a:03107. 128

[12] Joosten, J. J., "Two Proofs of Parsons' Theorem," Logic Group Preprint Series 127, University of Utrecht, 2002. 132

[13] Joosten, J. J., and A. Visser, "The interpretability logic of all reasonable arithmetical theories. The new conjecture," *Erkenntnis*, vol. 53 (2000), pp. 3–26. Zbl 0974.03049. MR 2002h:03129. 129

[14] Leivant, D., "The optimality of induction as an axiomatization of arithmetic," *The Journal of Symbolic Logic*, vol. 48 (1983), pp. 182–84. Zbl 0515.03018. MR 84f:03051. 132

[15] Parsons, C., "On a number theoretic choice schema and its relation to induction," pp. 459–73 in *Intuitionism and Proof Theory (Proceedings of the Conference, Buffalo, NY 1968)*, North-Holland, Amsterdam, 1970. Zbl 0202.01202. MR 43:6050. 132

[16] Parsons, C., "On n-quantifier induction," *The Journal of Symbolic Logic*, vol. 37 (1972), pp. 466–82. Zbl 0264.02027. MR 48:3712. 132

[17] Pudlák, P., "On the length of proofs of finitistic consistency statements in first order theories," pp. 165–96 in *Logic Colloquium '84 (Manchester, 1984)*, edited by J. B. Paris et al., vol. 120 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1986. Zbl 0619.03037. MR 87k:03072. 139

[18] Schwartz, D. G., "A free-variable theory of primitive recursive arithmetic," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 33 (1987), pp. 147–57. Zbl 0624.03042. MR 88g:03083. 130

[19] Schwartz, D. G., "On the equivalence between logic-free and logic-bearing systems of primitive recursive arithmetic," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 33 (1987), pp. 245–53. Zbl 0618.03027. MR 88i:03096. 130

[20] Shavrukov, V., "The logic of relative interpretability over Peano arithmetic (in Russian)," Technical Report 5, Steklov Mathematical Institute, Moscow, 1988. 129

[21] Simpson, S. G., *Subsystems of Second Order Arithmetic*, Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1999. Zbl 0909.03048. MR 2001i:03126. 139

[22] Skolem, T., "The foundations of elementary arithmetic established by means of the recursive mode of thought, without the use of apparent variables ranging over infinite domains," pp. 302–33 in *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*, edited by J. van Heijenoort, Harvard University Press, Cambridge, 2002. Reprint of the 1967 original. MR 2003a:03008. 130

[23] Smoryński, C., "The incompleteness theorems," pp. 821–65 in *Handbook of Mathematical Logic*, edited by J. Barwise, vol. 90 of *Studies in Logic and the Foundations of Mathematics*, North-Holland Publishing Co., Amsterdam, 1977. Zbl 0528.03001. MR 56:15351. 130

[24] Smoryński, C., *Self-Reference and Modal Logic*, Springer-Verlag, New York, 1985. Zbl 0596.03001. MR 88d:03001. 144

[25] Tait, W., "Finitism," *Journal of Philosophy*, vol. 78 (1981), pp. 524–46. 130

[26] Tarski, A., *Undecidable Theories*, Studies in Logic and the Foundations of Mathematics. In collaboration with A. Mostowski and R. M. Robinson. North-Holland Publishing Company, Amsterdam, 1953. Zbl 0187.27702. MR 15,384h. 127, 128

[27] Visser, A., "Interpretability logic," pp. 175–209 in *Mathematical Logic*, edited by P. P. Petkov, Plenum, New York, 1990. Zbl 0793.03064. MR 93k:03022. 129

[28] Visser, A., "The formalization of interpretability," *Studia Logica*, vol. 50 (1991), pp. 81–105. Zbl 0744.03023. MR 93f:03009. 128, 137

[29] Visser, A., "An inside view of EXP; or, The closed fragment of the provability logic of $\mathrm{I}\Delta_0 + \Omega_1$ with a propositional constant for EXP," *The Journal of Symbolic Logic*, vol. 57 (1992), pp. 131–65. Zbl 0785.03008. MR 93c:03027. 129, 130, 133

[30] Visser, A., "An overview of interpretability logic," pp. 307–59 in *Advances in Modal Logic, vol. 1*, edited by M. Kracht, M. de Rijke, and H. Wansing, vol. 87 of *CSLI Lecture Notes*, CSLI Publications, Stanford, 1998. Selected papers from the 1st Conference (AiML'96) held at the Free University of Berlin, Berlin, October 1996. Zbl 0915.03020. MR MR1688529. 128, 133

Acknowledgments

I would like to thank Lev D. Beklemishev, Volodya Yu. Shavrukov, and Albert Visser for fruitful discussions. A very informative referee report pointed out some inaccuracies in an earlier version of this paper. The report also made many good suggestions to improve the readability.

Department of Philosophy
Heidelberglaan 8
3584 CS Utrecht
THE NETHERLANDS
jjoosten@phil.uu.nl