



# Transductions in arithmetic



Albert Visser<sup>1</sup>

Department of Philosophy, Utrecht University, Janskerkhof 13, 3512BL Utrecht, The Netherlands

## ARTICLE INFO

### Article history:

Received 18 February 2014

Received in revised form 29 October 2015

Accepted 10 November 2015

Available online 18 November 2015

This paper is dedicated to the memory of Franco Montagna

### MSC:

03B25

03F25

03F30

03F45

### Keywords:

Interpretability

Provability Logic

Second Incompleteness Theorem

## ABSTRACT

In this paper we study a new relation between sentences: *transducibility*. The idea of transducibility is based on an analysis of Feferman's Theorem that the inconsistency of a theory  $U$  is interpretable over  $U$ . Transducibility is based on a converse of Feferman's Theorem: if a sentence is interpretable over a theory  $U$ , it is, in a sense that we will explain, an inconsistency statement for  $U$  over  $U$ .

We show that, for a wide class of theories  $U$ , transducibility coincides with interpretability over  $U$  and, for an even wider class, it coincides with  $\Pi_1$ -conservativity over  $U$ . Thus, transducibility provides a new way of looking at interpretability and  $\Pi_1$ -conservativity. On the other hand, we will show that transducibility admits variations that are distinct from interpretability and  $\Pi_1$ -conservativity.

We show that transducibility satisfies the interpretability logic ILM.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

In this paper we provide new characterizations of interpretability for essentially reflexive theories and of  $\Pi_1$ -conservativity for theories extending Elementary Arithmetic EA (aka  $\text{I}\Delta_0 + \text{Exp}$ ). These characterizations stand in the tradition of characterizations such as the Orey-Hájek Characterization and the Friedman Characterization, but they are of a different flavor. Our approach uncovers a connection between interpretability and  $\Pi_1$ -conservativity, on the one hand, and inconsistency statements of provability predicates satisfying the Hilbert–Bernays–Löb conditions, on the other.

We can view what is achieved in the paper from various other perspectives. The paper is a study of *role provability predicates* as is explained in Subsection 1.1. It provides a converse of a beautiful theorem due

E-mail address: [a.visser@uu.nl](mailto:a.visser@uu.nl).

<sup>1</sup> I am grateful to Rosalie Iemhoff and Thomas Müller for their comments and advice. I thank the anonymous referee for his helpful comments.

to Feferman as is explained in Subsection 1.2. It provides a rather general arithmetical semantics for the interpretability logic ILM. See Subsection 1.3.

The present paper is closely related to two of my other papers. The first is [57], which is to appear in the *Bulletin of Symbolic Logic*. The second is a somewhat more philosophical paper [59]. The present paper can be read independently of its two companions.

### 1.1. Role provability predicates

Syntactical approaches to modality come in two flavors. A first idea is to add a predicate or predicates to a language that has sufficient coding possibilities. Then, we stipulate that the predicate, considered as a predicate of sentences, satisfies a number of desired modal properties. An important question is which properties we can consistently (or also conservatively) demand of such predicates and whether we can define a Kripke style semantics for them. For examples of this approach, see e.g. [33,39,49,43,25,22,47].

A second approach is the modal study of predicates that are definable in theories with sufficient coding possibilities. This line of research usually zooms in on specific predicates like *provability* and *interpretability*. Provability Logic is a perfect example of this study. The classical papers in this field are [17,34,37,46]. For expository texts, see: [9,8,35,29,48,1,23]. There are many variations.

1. Over EA, also known as  $\text{ID}_0 + \text{Exp}$ , cutfree provability and ordinary provability are not equivalent. On the other hand they both validate Löb's Logic. See [51] and [32].
2. Over PA we can consider the predicates 'provable in PA with an oracle for  $\Pi_{n+1}$ -truth'. The logic of the hierarchy of such predicates is Japaridze's Logic GLP. See [28]. See also [8]. This logic was used by Lev Beklemishev to extract proof theoretic ordinals from its closed fragment. See e.g. [2,3,5,4].
3. We consider over the theory ZF, the predicate *truth in all transitive models of ZF*. This example was studied by Solovay in [46]. See also [8]. A closely related example is to consider truth in all  $V_\kappa$  where  $\kappa$  is inaccessible. See [8].
4. Let  $\text{PA}^2$  be the first-order version of second order arithmetic. We may consider the arithmetization of provability in  $\text{PA}^2$  with the  $\omega$ -rule. This predicate was studied e.g. in [8].
5. Per Lindström studied Parikh provability in his paper [36].
6. Sy Friedman, Michael Rathjen and Andreas Weiermann study *slow provability* for PA in their paper [14]. The modal behavior of slow provability predicates is currently studied by Fedor Pakhomov and Paula Henk.
7. Over EA, provability with an oracle for  $\Sigma_1$ -truth and ordinary provability are not equivalent. On the other hand they both validate Löb's Logic. See [60].
8. Graham Leach-Krouse studied an internal version of  $\Omega$ -validity over ZFC with the von Neumann interpretation.
9. A new kind of predicates called *supremum adapters* is studied by Paula Henk.

All predicates in the above list validate Löb's Logic. There are however other modally interesting predicates. The principal of the alternative unary predicates is the Feferman Predicate that was introduced in [12]. It was studied in [38,50,45]. Of a quite different kind is the binary predicate for interpretability over a given theory. This can be viewed as a generalization of ordinary provability. We refer the reader to e.g. [29,53,31,1,18]. An alternative arithmetical interpretation of interpretability logics is provided by various notions of conservativity. See [19]. In the present paper we will provide yet another interpretation: transducibility.

This paper is in the second tradition: we treat modal predicates as *objets trouvés*. They are already present in a given theory. On the other hand, we will not zoom in on specific predicates in the given theory, but we will be interested in the totality of predicates over the given theory satisfying such-and-such properties. The appropriate analogy is as follows. A predicate of a theory satisfying a given modal theory is like a model

of a theory, for example *group theory*. We will be interested in the relationships between these ‘models’. In the analogy: groups are models of group theory. One studies the relations between different groups and constructions on groups. In the same spirit we want to study certain transformations of predicates satisfying modal principles.

We will use quantification over provability predicates to define a constant binary predicate that is an interpretability analogue.

### 1.2. Generalizing Feferman’s Theorem and giving it a converse

We have the following theorem:

**Theorem 1.1** (Feferman). *Consider any theory  $U$  with a  $p$ -time decidable axiom set. Suppose  $N$  is an interpretation of Buss’ theory  $S_2^1$  in  $U$ , then there is an interpretation  $K$  of  $U + \text{incon}^N(U)$  in  $U$ .*

In the statement of the theorem, we assume that the theory  $U$  is given with a  $\Delta_1^b$ -formula representing its axiom set. We note that Feferman’s Theorem is a strengthening of the Second Incompleteness Theorem. The theory  $U$  not only fails to prove its own consistency as coded in any choice  $N$  for the natural numbers, no, it is positively able to produce uniformly internal models of itself in which we have its inconsistency coded in  $N$ . Feferman’s Theorem enables us to view the Second Incompleteness Theorem as a strength rather than as a weakness. Feferman proved this theorem in his classical paper [12]. In his retrospective paper [13], Feferman provides a historical discussion of a.o. the genesis of his theorem which is warmly recommended. We give a simple proof of Feferman’s Theorem in Section 3.

We show that Feferman’s result lifts to a class of predicates that we call HBL-predicates and, also, to an even wider class: the regular HBL-predicates. In this generalized form, the theorem admits a converse. Let us restrict ourselves for definiteness to PA. We find:  $\text{PA} + B$  is interpretable in PA iff  $B$  is of the form  $\Delta \perp$  for some HBL-predicate  $\Delta$  over PA. So, we have that the inconsistency of PA is interpretable over PA and that, conversely, a sentence is only interpretable over PA because it can be viewed as an inconsistency.

The generalization and extension of Feferman’s Theorem also has a version involving regular HBL-predicates. This will allow us, e.g. for extensions  $U$  of PA, to find a new characterization of the relation  $U + B$  is interpretable in  $U + A$ .

### 1.3. Semantics of interpretability logic

The idea of interpretability logic is very simple, given that we already know Provability Logic. It is the modal study of the predicate  $A$  interprets  $B$  over  $U$  or  $A \triangleright_U B$ , which means  $U + A$  interprets  $U + B$ . We refer the reader to the papers [29,53,31,1,18] for more information. The basic system of interpretability logic is IL which is Löb’s Logic plus the following principles.<sup>2</sup>

- IL1.  $\vdash \Box(\phi \rightarrow \psi) \rightarrow \phi \triangleright \psi$
- IL2.  $\vdash (\phi \triangleright \psi \wedge \psi \triangleright \chi) \rightarrow \phi \triangleright \chi$
- IL3.  $\vdash (\phi \triangleright \chi \wedge \psi \triangleright \chi) \rightarrow (\phi \vee \psi) \triangleright \chi$
- IL4.  $\vdash \phi \triangleright \psi \rightarrow (\Diamond \phi \rightarrow \Diamond \psi)$
- IL5.  $\vdash \Diamond \phi \triangleright \phi$

If we take an essentially reflexive theory like Peano Arithmetic (PA) as our basic theory, then we have the extra principle M.

<sup>2</sup> Usually, we have two operators  $\Box$  and  $\triangleright$ . However, the operator  $\Box$  can be defined by:  $\Box A := (\neg A) \triangleright \perp$ .

M.  $\vdash \phi \triangleright \psi \rightarrow (\phi \wedge \Box \chi) \triangleright (\psi \wedge \Box \chi)$

The arithmetical completeness of ILM for theories like PA was proven by Volodya Shavrukov in [44] and Alessandro Berarducci in [7]. Apart from relative interpretability the binary connective has a natural interpretation as conservativity. See e.g. [19,27,20,30]. Regular L-predicates provide a new interpretation of our modal interpretability logic.

#### 1.4. Basic notions and facts

In Appendix A we introduce the basic notions and facts needed to read the paper. The reader is also referred to the textbook [21]. At this point, we just fix a number of conventions and notations.

Theories are, in this paper, theories of first-order predicate logic, that have a finite relational signature and that are axiomatized by an axiom set that is represented by a  $\Delta_1^b$ -formula. We will pretend that a theory also has function symbols. Terms can be eliminated using the well-known term-unwinding algorithm.

We use modal notations as much as possible. For example,  $\Box_U A$  is  $\text{prov}_U(\ulcorner A \urcorner)$ . We use  $\Box_{U,x} A$  for restricted provability where the Gödel numbers of the axioms used in the proof are all below  $x$  and the complexity (= depth of quantifier alternations) is smaller than  $x$ . See Appendix A.4 for more information. We use  $A \triangleright_U B$  for interpretability over  $U$ .

## 2. Transduction

In this section we define regular L-predicates and present their connection to interpretability logic.

### 2.1. Regular L-predicates

Let  $U$  be any theory and let  $N$  be an interpretation of the Tarski–Mostowski–Robinson Arithmetic R in  $U$ . Let  $P$  be a predicate of the  $N$ -numbers. We write  $\Delta A$  for  $P(\ulcorner A \urcorner)$ , where  $\ulcorner \cdot \urcorner$  is some standard efficient Gödelnumbering used w.r.t.  $N$ . We use  $\vdash$  for  $\vdash_U$ .

A predicate  $\Delta$  is a *regular L-predicate* w.r.t.  $U, N$  if it satisfies the following principles. In our formulation, the variables  $A, B, \dots$  range over  $U$ -sentences.

- rL1.  $A \vdash B \Rightarrow \Delta A \vdash \Delta B$
- rL2.  $\Delta A, \Delta(A \rightarrow B) \vdash \Delta B$
- rL3.  $\Delta A \vdash \Delta \Delta A$

The name *regular* is taken from the regular modal logics described in [26].

**Remark 2.1.** The original Hilbert–Bernays conditions ([24], in the second edition: pp. 294, 295) were approximately, in modern notation:

- 1.  $A \vdash B \Rightarrow \Delta A \vdash \Delta B$
- 2.  $\vdash \Delta \neg Ax \rightarrow \Delta \neg A\bar{x}$
- 3.  $S \vdash \Delta S$ , where  $S$  is a  $\Sigma_1$ -formula

Here  $A$  and  $B$  range over formulas. Free variables have universal reading inside  $\Delta$ , so e.g.  $\Delta A$  stands for the provability of the universal closure of  $A$ . The negation in Condition 2 is connected with the precise form of Hilbert & Bernays' argument.

It is a curious fact is that our **rL1** was the true first Hilbert–Bernays condition (modulo the treatment of variables). It should be noted that Hilbert and Bernays assumed that  $\Delta$  was  $\Sigma_1$ , so their conditions were not really intended as fully abstract conditions.  $\square$

We note one alternative formulation for our axioms. Let  $\Gamma$  and  $\Theta$  range over finite sets of  $U$ -sentences. We write  $\Delta\Theta := \{\Delta C \mid C \in \Theta\}$ . A predicate  $\Delta$  is a regular L-predicate w.r.t.  $U, N$  iff it satisfies:

- $\Gamma, \Delta\Theta \vdash A \Rightarrow \Delta\Gamma, \Delta\Theta \vdash \Delta A$ , where we demand that  $\Gamma \cup \Theta \neq \emptyset$ .

A predicate  $\Delta$  is a *normal* L-predicate or simply an L-predicate if it is a regular L-predicate and if we have in addition that  $\vdash \Delta\top$ .

**Example 2.2.** Even over Robinson’s theory **R**, we have non-trivial examples of L-predicates. E.g., let  $A$  be the conjunction of the axioms of a finite axiomatization of  $S_2^1$ , or, rather, a version of  $S_2^1$  in the arithmetical language. Then, the predicate  $\Delta B := (A \rightarrow \Box_R B)$  is an L-predicate. Here  $\Box_R$  is a standard formalization of provability in **R**.  $\square$

Since we have the necessary fixed points in **R**, we can prove Löb’s Theorem.

**Theorem 2.3** (*Löb’s Theorem*). *Suppose  $U$  is a theory with natural numbers  $N$  satisfying **R**. Suppose that  $\Delta$  is a regular L-predicate. Then  $\Delta(\Delta A \rightarrow A) \vdash \Delta A$ .*

**Proof.** By the Gödel fixed Point Lemma, we find a  $B$  such that

$$(\dagger) \quad \vdash B \leftrightarrow (\Delta B \rightarrow A).$$

Since  $B, \Delta B \vdash A$ , we find  $\Delta B \vdash \Delta A$ . So, we have  $\Delta A \rightarrow A \vdash \Delta B \rightarrow A$  and, hence,  $\Delta A \rightarrow A \vdash B$ . Ergo, we have  $\Delta(\Delta A \rightarrow A) \vdash \Delta(\Delta B \rightarrow A)$  and  $\Delta(\Delta A \rightarrow A) \vdash \Delta B$ . By the usual reasoning, it follows that  $\Delta(\Delta A \rightarrow A) \vdash \Delta A$ .  $\square$

We could also have given a purely modal formulation of the insight contained in Löb’s theorem. The only role of **R** in the argumentation above is the fact that it supports the Fixed Point Lemma. If we have a purely modal language and a modal operator satisfying the regular L-property extended with constants and axioms for fixed points for guarded (aka modalized) propositional variables, then the above reasoning works.

## 2.2. Regular Löb’s Logic

Corresponding to the idea of a *regular L-predicate* we have the purely modal theory Regular Löb’s Logic or **rGL**. It has the following axioms.

$$\text{rGL1. } \phi \vdash \psi \Rightarrow \Delta\phi \vdash \Delta\psi$$

$$\text{rGL2. } \Delta\phi, \Delta(\phi \rightarrow \psi) \vdash \Delta\psi$$

$$\text{rGL3. } \Delta\phi \vdash \Delta\Delta\phi$$

$$\text{rGL4. } \Delta(\Delta\phi \rightarrow \phi) \vdash \Delta\phi$$

The following result is both obvious and useful.

**Theorem 2.4.**  $\text{GL} \vdash \phi$  iff  $\text{rGL} \vdash \Delta\top \rightarrow \phi$ .

**Proof.** From left to right is an induction of proof length. From right to left is trivial, since GL extends rGL.  $\square$

### 2.3. Transductions

We define  $A \blacktriangleright B$  as: for some regular  $L$ -predicate  $\Delta$ , we have  $A \vdash \Delta \top$  and  $\Delta \perp \vdash B$ .

We will call a triple  $\langle A, \Delta, B \rangle$  a *transduction* if  $\Delta$  is a regular  $L$ -predicate and  $A \vdash \Delta \top$  and  $\Delta \perp \vdash B$ . We write  $\Delta : B \blacktriangleleft A$  or  $\Delta : A \blacktriangleright B$  for:  $\langle A, \Delta, B \rangle$  is a transduction.

Two transductions  $\Delta : A \blacktriangleright B$  and  $\Delta' : A' \blacktriangleright B'$  are *equivalent* iff we have  $\vdash A \leftrightarrow A'$  and  $\vdash B \leftrightarrow B'$  and  $\vdash \Delta C \leftrightarrow \Delta' C$ , for all  $C$ .

Given any transduction from  $A$  to  $B$ , we can find a transduction from  $A$  to  $B$  that satisfies some further desirable properties. We will call such transductions *basic transductions*. A triple  $\langle A, \Delta, B \rangle$  is a *basic* transduction iff it satisfies:

- bD1.  $C \vdash D \Rightarrow \Delta C \vdash \Delta D$
- bD2.  $\Delta C, \Delta(C \rightarrow D) \vdash \Delta D$
- bD3.  $\vdash \Delta \top \leftrightarrow (A \vee B)$ .
- bD4.  $\vdash \Delta \perp \leftrightarrow B$ .
- bD5.  $\vdash \Delta \Delta \perp \leftrightarrow (A \vee B)$ .

It is easy to see that a basic transduction is a transduction. Suppose  $\Delta : A \blacktriangleright B$  is an arbitrary transduction. Consider  $\Psi \langle A, \Delta, B \rangle := \langle A, \Delta^*, B \rangle$ , where  $\Delta^*$  is defined as  $\Delta^* C := \vdash B \vee (A \wedge \Delta(B \rightarrow C))$ . We claim that  $\Delta^* : A \blacktriangleright B$  is a basic transduction. The verification is an easy exercise in modal logic. We treat as an example bD4. We have:

$$\begin{aligned}
 B &\vdash B \vee (A \wedge \Delta(B \rightarrow \perp)) \\
 &\vdash B \vee (A \wedge \Delta(\Delta \perp \rightarrow \perp)) \\
 &\vdash B \vee (A \wedge \Delta \perp) \\
 &\vdash B \vee (A \wedge B) \\
 &\vdash B
 \end{aligned}$$

We note that  $\Psi$  preserves equivalence of transductions. The operation  $\Psi$  is idempotent, since:  $\Psi \langle A, \Psi \langle A, \Delta, B \rangle, B \rangle = \langle A, \Delta^\circ, B \rangle$ , where:

$$\Delta^\circ C := B \vee (A \wedge (B \vee (A \wedge \Delta(B \rightarrow C)))).$$

We verify the validity of IL for  $\blacktriangleright$ . The verification is for the moment in the metalanguage. We want it to be verifiable in the  $U$  itself. We postpone discussion of the demands on  $U$  until after the proof. We remind the reader that the logic IL is defined as follows.

- IL1.  $\vdash \Box(\phi \rightarrow \psi) \rightarrow \phi \triangleright \psi$
- IL2.  $\vdash (\phi \triangleright \psi \wedge \psi \triangleright \chi) \rightarrow \phi \triangleright \chi$
- IL3.  $\vdash (\phi \triangleright \chi \wedge \psi \triangleright \chi) \rightarrow (\phi \vee \psi) \triangleright \chi$
- IL4.  $\vdash \phi \triangleright \psi \rightarrow (\Diamond \phi \rightarrow \Diamond \psi)$
- IL5.  $\vdash \Diamond \phi \triangleright \phi$

Each principle except **IL4** corresponds to an operation on transductions.<sup>3</sup> The operations as chosen by us all yield a basic transduction as output independent of whether the input transductions are basic. In our verifications we will use the fact that  $\Gamma, \Delta\Theta \vdash A \Rightarrow \Delta\Gamma, \Delta\Theta \vdash \Delta A$ , provided that  $\Gamma \cup \Theta$  is non-empty.

**IL1:** Suppose  $\vdash A \rightarrow B$ . We define  $\Phi_1(A, B) := \langle A, \Delta^* B \rangle$ , where  $\Delta^* C := \leftrightarrow B$ .

It is easy to verify that  $\Phi_1(A, B)$  is a basic transduction.

**IL2:** Suppose  $\Delta_0 : A \blacktriangleright B$  and  $\Delta_1 : B \blacktriangleright C$ . We define  $\Phi_2(A, \Delta_0, B, \Delta_1, C) := \langle A, \Delta^* C \rangle$ , where:

$$\Delta^* D := \leftrightarrow C \vee (A \wedge ((B \wedge \Delta_1(C \rightarrow D)) \vee (\neg B \wedge \Delta_0(B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D)))))))$$

We verify **BD1**. Suppose  $D \vdash E$ . It follows that  $(C \rightarrow D) \vdash (C \rightarrow E)$ . Ergo:

$$(a) \quad \Delta_1(C \rightarrow D) \vdash \Delta_1(C \rightarrow E).$$

Hence, also:

$$B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D))) \vdash B \rightarrow ((C \wedge E) \vee (\neg C \wedge \Delta_1(C \rightarrow E))).$$

It follows that:

$$(b) \quad \Delta_0(B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D)))) \vdash \Delta_0(B \rightarrow ((C \wedge E) \vee (\neg C \wedge \Delta_1(C \rightarrow E)))).$$

It is immediate from (a) and (b) that  $\Delta^* D \vdash \Delta^* E$ .

We verify **BD2**. We can easily derive: (a)  $\Delta_1(C \rightarrow D), \Delta_1(C \rightarrow (D \rightarrow E)) \vdash \Delta_1(E)$  and from this:

$$\begin{aligned} & B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D))), \\ & B \rightarrow ((C \wedge (D \rightarrow E)) \vee (\neg C \wedge \Delta_1(C \rightarrow (D \rightarrow E)))) \vdash \\ & B \rightarrow ((C \wedge E) \vee (\neg C \wedge \Delta_1(C \rightarrow E))). \end{aligned}$$

It follows that:

$$\begin{aligned} & (b) \quad \Delta_0(B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D)))), \\ & \Delta_0(B \rightarrow ((C \wedge (D \rightarrow E)) \vee (\neg C \wedge \Delta_1(C \rightarrow (D \rightarrow E))))) \vdash \\ & \Delta_0(B \rightarrow ((C \wedge E) \vee (\neg C \wedge \Delta_1(C \rightarrow E)))). \end{aligned}$$

We reason in  $U$ . Suppose (c)  $\Delta^* D$  and (d)  $\Delta^*(D \rightarrow E)$ . It follows that we have one of the exclusive cases  $C$  or  $\neg C \wedge A \wedge B$  or  $\neg C \wedge A \wedge \neg B$ . In case we have  $C$  we are immediately done. Suppose we have  $\neg C$  and  $A$  and  $B$ . In this case (c) gives us  $\Delta_1(C \rightarrow D)$  and (d) gives us  $\Delta_1(C \rightarrow (D \rightarrow E))$ . By (a), we find  $\Delta_1(C \rightarrow E)$ . Ergo  $\Delta^* E$ . Suppose we have  $\neg C$  and  $A$  and  $\neg B$ . In this case (c) and (d) give us  $\Delta_0(B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D))))$  and  $\Delta_0(B \rightarrow ((C \wedge (D \rightarrow E)) \vee (\neg C \wedge \Delta_1(C \rightarrow (D \rightarrow E)))))$ . By (b) we find the desired conclusion  $\Delta_0(B \rightarrow ((C \wedge E) \vee (\neg C \wedge \Delta_1(C \rightarrow E))))$ , and hence  $\Delta^* E$ .

<sup>3</sup> In hindsight it would have been more natural to have **IL4** as the last principle of the list. However, we do not want to diverge from the traditional order.

We treat **bd3,4,5**. First we have:

$$\begin{aligned}
 \vdash \Delta^* \top &\leftrightarrow C \vee (A \wedge ((B \wedge \Delta_1(C \rightarrow \top)) \vee \\
 &\quad (\neg B \wedge \Delta_0(B \rightarrow ((C \wedge \top) \vee (\neg C \wedge \Delta_1(C \rightarrow \top)))))) \\
 &\leftrightarrow C \vee (A \wedge ((B \wedge \top) \vee \\
 &\quad (\neg B \wedge \Delta_0(B \rightarrow (C \vee (\neg C \wedge \top)))))) \\
 &\leftrightarrow C \vee (A \wedge (B \vee (\neg B \wedge \top))) \\
 &\leftrightarrow C \vee A
 \end{aligned}$$

We note that since  $\Delta_0 \perp \vdash B$ , we have  $\neg B \vdash \neg \Delta_0 \perp$ , and, hence,  $\Delta_0 \neg B \vdash \Delta_0(\neg \Delta_0 \perp)$  and, so,  $\Delta_0 \neg B \vdash \Delta_0 \perp$ . Similarly,  $\Delta_1 \neg C \vdash \Delta_1 \perp$ . We have:

$$\begin{aligned}
 \vdash \Delta^* \perp &\leftrightarrow C \vee (A \wedge ((B \wedge \Delta_1 \neg C) \vee \\
 &\quad (\neg B \wedge \Delta_0(B \rightarrow ((C \wedge \perp) \vee (\neg C \wedge \Delta_1 \neg C)))))) \\
 &\leftrightarrow C \vee (A \wedge ((B \wedge \Delta_1 \perp) \vee (\neg B \wedge \Delta_0(B \rightarrow (\neg C \wedge \Delta_1 \perp))))) \\
 &\leftrightarrow C \vee (A \wedge ((B \wedge C) \vee (\neg B \wedge \Delta_0 \neg B))) \\
 &\leftrightarrow C \vee (A \wedge ((B \wedge C) \vee (\neg B \wedge \Delta_0 \perp))) \\
 &\leftrightarrow C \vee (A \wedge B \wedge C) \\
 &\leftrightarrow C
 \end{aligned}$$

Finally:

$$\begin{aligned}
 \vdash \Delta^* \Delta^* \perp &\leftrightarrow \Delta^* C \\
 &\leftrightarrow C \vee (A \wedge ((B \wedge \Delta_1(C \rightarrow C)) \vee \\
 &\quad (\neg B \wedge \Delta_0(B \rightarrow ((C \wedge C) \vee (\neg C \wedge \Delta_1(C \rightarrow C)))))) \\
 &\leftrightarrow C \vee (A \wedge ((B \wedge \top) \vee (\neg B \wedge \Delta_0(B \rightarrow (\top \vee (\neg C \wedge \Delta_1 \top)))))) \\
 &\leftrightarrow C \vee (A \wedge (B \vee (\neg B \wedge \top))) \\
 &\leftrightarrow C \vee A
 \end{aligned}$$

**IL3:** Suppose  $\Delta_0 : A \blacktriangleright C$  and  $\Delta_1 : B \blacktriangleright C$ . We define  $\Phi_3(A, \Delta_0, B, \Delta_1, C) := \langle A \vee B, \Delta^*, C \rangle$ , where:

$$\Delta^* D := C \vee (A \wedge \Delta_0(C \rightarrow D)) \vee (\neg A \wedge B \wedge \Delta_1(C \rightarrow D)).$$

All cases except **bd4** are like the cases of **IL2** but simpler. We treat **bd4**.

$$\begin{aligned}
 \vdash C &\rightarrow C \vee (A \wedge \Delta_0 \neg C) \vee (\neg A \wedge B \wedge \Delta_1 \neg C) \\
 &\rightarrow C \vee (A \wedge C) \vee (\neg A \wedge B \wedge C) \\
 &\rightarrow C
 \end{aligned}$$

**IL4:** Suppose  $\Delta : A \blacktriangleright B$  and  $\vdash \neg B$ . It follows that  $\Delta \top \vdash \Delta \neg B$ , and, hence,  $\Delta \top \vdash \Delta \neg \Delta \perp$  and, thus, that  $\Delta \top \vdash \Delta \perp$ . We may conclude that  $A \vdash B$  and, so,  $A \vdash \perp$ , i.e.,  $\vdash \neg A$ .

**IL5:** We prove a stronger fact, say **IL5<sup>+</sup>**. Suppose  $\Delta$  is a regular L-predicate and  $A \vdash \Delta \top$ . We show that  $(A \wedge \nabla B) \triangleright B$ . We define  $\Phi_4(A, B, \Delta) := \langle (A \wedge \nabla B), \Delta^*, B \rangle$ , where:



$$\Delta^*C : \leftrightarrow B \vee (A \wedge \nabla B \wedge \Delta(B \rightarrow C)).$$

We leave the simple verifications to the reader.

We will call  $\Phi_1, \Phi_2, \Phi_3, \Phi_4$ : *the  $\Phi$ -operations*. We will call a class of transductions closed under the  $\Phi$ -operations:  *$\Phi$ -closed*.

To truly obtain **IL**, we need verifiability of the above proofs in  $U$  itself w.r.t. some chosen  $N : S_2^1 \triangleleft U$ . Fortunately all the transformations in our verification are p-time, so we do not encounter a problem in internalizing the argument.

Clearly, every  $\Phi$ -closed class of transductions  $\mathcal{D}$  will satisfy **IL**. The basic transductions are an example of such a class. We will write  $\triangleright_{\mathcal{D}}$  for the transducibility relation obtained by only considering transductions from  $\mathcal{D}$ .

**Open Question 2.5.** One would hope that the transductions (or a closed subclass of the transductions) form a category, but it seems that our definitions do not yield the associativity of composition. Since neither the class of transductions nor the chosen operations on transductions are uniquely determined, there is still some hope that we can find the desired category. So we formulate the open question: *can we find a category of transductions?*  $\square$

#### 2.4. Regular HBL-predicates

In this subsection we introduce the class of regular HBL-predicates. We will show that transductions associated with these predicates are  $\Phi$ -closed.

We formulate our relevant notion of  $\exists\Sigma_1^b$ -completeness. Consider a theory  $U$  and an interpretation  $N : S_2^1 \triangleleft U$ . We define:

$\text{r-C } \Delta \top, S^N \vdash \Delta S^M$ , where  $S$  is a  $\exists\Sigma_1^b$ -sentence and  $M$  is any interpretation of  $S_2^1$  in  $U$ .  
 $\text{r-C}_0 \Delta \top, S^N \vdash \Delta S^N$ , where  $S$  is a  $\exists\Sigma_1^b$ -sentence.

Note that the definition assumes that we have  $U$  and  $N$  fixed in the background. We call a regular L-predicate that satisfies  $\text{r-C}$  w.r.t.  $U, N$ : *a regular HBL-predicate*. We call a regular L-predicate that satisfies  $\text{r-C}_0$  w.r.t.  $U, N$ : *a regular HBL<sub>0</sub>-predicate*. The name “HBL” stands for: Hilbert–Bernays–Löb. The reason for this choice is the fact that the third Hilbert–Bernays was verifiable  $\Sigma_1$ -completeness.

Here is a basic theorem about regular  $\exists\Sigma_1^b$ -completeness, connecting it with restricted provability.

**Theorem 2.6.** *Suppose that  $U$  is sequential.<sup>4</sup> Let  $N$  interpret  $S_2^1$  in  $U$ . Suppose that  $\Delta$  is a regular L-predicate for  $U, N$ . Then, the following are equivalent:*

- i.  $\Delta$  is a regular HBL-predicate.
- ii. For all  $U$ -sentences  $A$ , and for all  $n$ , we have  $\Delta \top \vdash \Box_n^N A \rightarrow \Delta A$ .

**Proof.** Suppose that  $U$  is sequential and  $N : S_2^1 \triangleleft U$ . Suppose that  $\Delta$  is a regular L-predicate for  $U, N$ .

(i)  $\Rightarrow$  (ii). Suppose  $\Delta$  is a regular HBL-predicate. Consider any sentence  $A$  and any number  $n$ . Since  $U$  is sequential, there is an interpretation  $M : S_2^1 \triangleleft U$ , such that  $U \vdash \Box_n^M A \rightarrow A$ . By  $\text{r-C}$ , we have  $\Delta \top \vdash \Box_n^N A \rightarrow \Delta \Box_n^M A$ . Since  $\vdash \Box_n^M A \rightarrow A$ , we have, by  $\text{rL1}$ , that  $\vdash \Delta \Box_n^M A \rightarrow \Delta A$ . It follows that  $\Delta \top \vdash \Box_n^N A \rightarrow \Delta A$ .

<sup>4</sup> See Appendix A.3 for a brief explanation of the notion of sequentiality.

(ii)  $\Rightarrow$  (i). Suppose that, for all  $U$ -sentences  $A$ , and for all  $n$ , we have  $\Delta \top \vdash \Box_n^N A \rightarrow \Delta A$ . Consider any  $\exists \Sigma_1^b$ -sentence  $S$  and any  $M : S_2^1 \triangleleft U$ . We have, for sufficiently large  $n$ ,  $\vdash S^N \rightarrow \Box_n^N S^M$  and  $\Delta \top \vdash \Box_n^N S^M \rightarrow \Delta S^M$ . Hence,  $\Delta \top \vdash S^N \rightarrow \Delta S^M$ .  $\square$

We remind the reader of our operations:

- $\Psi \langle A, \Delta, B \rangle := \langle A, \Delta^*, B \rangle$ , where  $\Delta^*$  is defined as:  
 $\Delta^* C :\leftrightarrow B \vee (A \wedge \Delta(B \rightarrow C))$ .
- $\Phi_1(A, B) := \langle A, \Delta^*, B \rangle$ , where  $\Delta^* C :\leftrightarrow B$ .
- Suppose  $\Delta_0 : A \blacktriangleright B$  and  $\Delta_1 : B \blacktriangleright C$ . We define  $\Phi_2(A, \Delta_0, B, \Delta_1, C) := \langle A, \Delta^*, C \rangle$ , where:  
 $\Delta^* D :\leftrightarrow C \vee (A \wedge ((B \wedge \Delta_1(C \rightarrow D)) \vee (\neg B \wedge \Delta_0(B \rightarrow ((C \wedge D) \vee (\neg C \wedge \Delta_1(C \rightarrow D)))))))$ .
- Suppose  $\Delta_0 : A \blacktriangleright C$  and  $\Delta_1 : B \blacktriangleright C$ . We define  $\Phi_3(A, \Delta_0, B, \Delta_1, C) := \langle A \vee B, \Delta^*, C \rangle$ , where:  
 $\Delta^* D :\leftrightarrow C \vee (A \wedge \Delta_0(C \rightarrow D)) \vee (\neg A \wedge B \wedge \Delta_1(C \rightarrow D))$ .
- Suppose  $\Delta : A \blacktriangleright D$ . Note that  $D$  is not necessarily  $B$ . We define  $\Phi_4(A, B, \Delta) := \langle (A \wedge \nabla B), \Delta^*, B \rangle$ , where:  
 $\Delta^* C :\leftrightarrow B \vee (A \wedge \nabla B \wedge \Delta(B \rightarrow C))$ .

If the predicates in the input of the operations are HBL (HBL<sub>0</sub>) for  $U, N$ , then so are the predicates in the output.

We treat the case of  $\Phi_2$  for HBL. Suppose  $\Delta_0 : A \blacktriangleright B$  and  $\Delta_1 : B \blacktriangleright C$ , where  $\Delta_0$  and  $\Delta_1$  are HBL. We have, for any  $M : S_2^1 \triangleleft U$ , that:  $A \vdash S^N \rightarrow \Delta_0 S^M$ , so *a fortiori*  $A \vdash S^N \rightarrow \Delta_0(B \rightarrow S^M)$ . Similarly,  $B \vdash S^N \rightarrow \Delta_1(C \rightarrow S^M)$  and, hence,  $A \vdash \Delta_0(B \rightarrow (S^N \rightarrow \Delta_1(C \rightarrow S^M)))$ . We also have  $A \vdash S^N \rightarrow (\Delta_0 S^N \wedge \Delta_0 S^M)$ . So:

$$A \vdash S^N \rightarrow \Delta_0(B \rightarrow ((C \wedge S^M) \vee (\neg C \wedge \Delta_1(C \rightarrow S^M)))).$$

From these facts the desired result is immediate.

We show that, if we restrict ourselves to transductions based on HBL (HBL<sub>0</sub>) predicates for  $U, N$ , we have:

**M** Suppose  $S$  is  $\exists \Sigma_1^b$ , then:  $A \blacktriangleright B \Rightarrow (A \wedge S^N) \blacktriangleright (B \wedge S^N)$ .

Suppose  $\Delta : A \blacktriangleright B$ . Let  $\Phi_5(A, \Delta, B) := \langle A \wedge S^N, \Delta^*, B \wedge S^N \rangle$ , where  $\Delta^* C :\leftrightarrow (S^N \wedge \Delta C)$ .

We leave the easy verification that  $\Delta^*$  is indeed a HBL (HBL<sub>0</sub>) predicate for  $U, N$  and that  $\Delta^* : (A \wedge S^N) \blacktriangleright (B \wedge S^N)$  to the reader. We note that, since  $\Box_U C$  is  $\exists \Sigma_1^b$ , the usual form of **M** follows:

$$A \blacktriangleright B \Rightarrow (A \wedge \Box^N C) \blacktriangleright (B \wedge \Box^N C).$$

### 3. Feferman's Theorem

In this section we present a simple proof of Feferman's Theorem. We remind the reader of the theorem.

**Feferman's Theorem.** Consider any theory  $U$  with a  $p$ -time decidable axiom set. Suppose  $N$  is an interpretation of Buss' theory  $S_2^1$  in  $U$ , then there is an interpretation  $K$  of  $U + \text{incon}^N(U)$  in  $U$ .

**Proof.** Consider any theory  $U$  with  $p$ -time decidable axiom set and an interpretation  $N : S_2^1 \triangleleft U$ . Clearly,  $\Diamond^N \top \vdash_U \Diamond^N \Box^N \perp$  and  $\Diamond^N \Box^N \perp \triangleright_U \Box^N \perp$ , by, respectively, the Second Incompleteness Theorem and the

Gödel–Hilbert–Bernays–Wang–Henkin–Feferman Theorem ([Theorem A.1](#) of the Appendix). By composition,  $\Diamond^N \top \triangleright_U \Box^N \perp$ . Suppose  $K$  witnesses that  $\Diamond^N \top \triangleright_U \Box^N \perp$ . We also have  $ID : \Box^N \perp \triangleright_U \Box^N \perp$ . Hence  $K \langle \Diamond^N \top \rangle ID : \top \triangleright_U \Box^N \perp$ . Here  $K \langle \Diamond^N \top \rangle ID$  is the disjunctive interpretation that ‘is’  $K$  if  $\Diamond^N \top$  and  $ID$  if not  $\Diamond^N \top$ . (See [Appendix A.2](#) for the definition of disjunctive interpretations.)  $\square$

The proof of Feferman’s Theorem presented here was given in [\[51\]](#). The same proof is reported in [\[13\]](#). Feferman learned the proof in conversation from Per Lindström, who discovered the proof independently.

#### 4. Transducibility, conservativity, interpretability

In this section, we prove that, for essentially reflexive theories, transducibility and interpretability coincide and we prove that for theories interpreting EA transducibility and  $\Pi_1$ -conservativity coincide. (The preceding formulation is still not fully precise. It will be refined below.)

##### 4.1. Interpretability over essentially reflexive theories

In this subsection, we show that HBL-transducibility and interpretability coincide for essentially reflexive theories. We first prove Interpretation Existence for HBL predicates.

A theory  $U$  is *locally sententially essentially reflexive* if, for every  $U$ -sentence  $A$  and for every  $n$ , there is an  $M : S_2^1 \triangleleft U$  such that  $U \vdash \Box_{U,n}^M A \rightarrow A$ . Here  $\Box_{U,n}$  is restricted provability as described in [Appendix A.4](#). As is well known, sequential theories are locally sententially essentially reflexive.

The theory  $U$  is *sententially essentially reflexive* if, there is a fixed  $N : S_2^1 \triangleleft U$  such that, for every  $U$ -sentence  $A$  and for every  $n$ , we have  $U \vdash \Box_{U,n}^N A \rightarrow A$ . We often make  $N$  part of the data and say, e.g.,  $U$  is *sententially essentially reflexive w.r.t.  $N$* .

The theory  $U$  is *essentially reflexive* if, there is a fixed  $N : S_2^1 \triangleleft U$  such that, for every  $U$ -formula  $A$  and, for every  $n$ , we have

$$U \vdash \forall \vec{x} \in \delta_N (\Box_{U,n}^N A\vec{x} \rightarrow A\vec{x}).$$

Here the occurrence of  $\vec{x}$  inside  $\Box_{U,n}$  is implemented as the substitution of the Gödel numbers of the binary  $N$ -numerals corresponding to the  $x_i$ , in the usual manner.

If a theory is essentially reflexive with respect to  $N$ , it satisfies full induction with respect to  $N$ . Conversely, if  $U$  satisfies full induction with respect to  $N$  and is *sequential*, then  $U$  is essentially reflexive with respect to  $N$ .

Examples of sequential and essentially reflexive theories are the familiar PA and ZF. For a worked out-example of a theory that is sententially essentially reflexive but not essentially reflexive, see [\[58\]](#).

**Theorem 4.1.** *Suppose that  $U$  is locally sententially essentially reflexive. Let the interpretation  $N$  provide natural numbers satisfying  $S_2^1$ . Suppose that  $\Delta$  is a regular HBL-predicate for  $U, N$ . Then,  $(\Delta \top \wedge \nabla A) \triangleright A$ .*

**Proof.** By [Theorem 2.6](#), we have, for every  $n$ , that  $\Delta \top \vdash \nabla A \rightarrow \Diamond_{U,n}^N A$ . Hence, by [Theorem A.3](#), we find that  $(\Delta \top \wedge \nabla A) \triangleright A$ .  $\square$

Next we prove Feferman’s Theorem w.r.t. HBL predicates.

**Theorem 4.2.** *Suppose that  $U$  is locally sententially essentially reflexive. Let our natural numbers be given by  $N : S_2^1 \triangleleft U$ . Suppose that  $\Delta$  is a regular HBL-predicate for  $U, N$ . Then,  $\Delta \top \triangleright \Delta \perp$ .*

**Proof.** First, we trivially have  $(\Delta \top \wedge \Delta \perp) \triangleright \Delta \perp$ . Secondly, we have, by Löb's Theorem,  $(\Delta \top \wedge \nabla \top) \vdash (\Delta \top \wedge \nabla \Delta \perp)$  and, by Theorem 4.1,  $(\Delta \top \wedge \nabla \Delta \perp) \triangleright \Delta \perp$ . By IL3 we are done.  $\square$

We now move to a result where we really need global reflexivity. This theorem is a strengthening of an earlier result in [59].

**Theorem 4.3.** *We work over a theory  $U$  and  $N : S_2^1 \triangleleft U$ . Suppose  $U$  is sententially essentially reflexive w.r.t.  $N$ . Then, over  $U, N$ , we have:  $A \triangleright B$  iff  $A \blacktriangleright_{\text{hbl}} B$ .*

**Proof.** Suppose  $A \triangleright B$ . It follows that, for every  $n$ ,  $A \vdash \Diamond_n^N B$ . We define:

$$\Delta C :\leftrightarrow B \vee (A \wedge \exists x (\Box_x^N (B \rightarrow C) \wedge \Diamond_x^N B)).$$

We show that  $\Delta : A \blacktriangleright B$  is a basic transduction. We have:

$$\begin{aligned} \vdash \Delta \perp &\leftrightarrow B \vee (A \wedge \exists x (\Box_x^N \neg B \wedge \Diamond_x^N B)) \\ &\leftrightarrow B \\ \vdash \Delta \top &\leftrightarrow B \vee (A \wedge \exists x (\Box_x^N \top \wedge \Diamond_x^N B)) \\ &\leftrightarrow A \vee B \\ \vdash \Delta \Delta \perp &\leftrightarrow \Delta B \\ &\leftrightarrow B \vee (A \wedge \exists x (\Box_x^N \top \wedge \Diamond_x^N B)) \\ &\leftrightarrow A \vee B \end{aligned}$$

We treat rL1. Suppose  $C \vdash D$ . Then, (a) for some  $n$ ,  $C \vdash_n D$ . We reason in  $U$ . Suppose (b)  $B \vee (A \wedge \exists x (\Box_x^N (B \rightarrow C) \wedge \Diamond_x^N B))$ . We want to prove (c)  $B \vee (A \wedge \exists x (\Box_x^N (B \rightarrow D) \wedge \Diamond_x^N B))$ . If  $B$  we are easily done. Suppose  $\neg B$ . We find (d)  $A \wedge \exists x (\Box_x^N (B \rightarrow C) \wedge \Diamond_x^N B)$ . It follows that we have  $A$  and hence  $\Diamond_n^N B$ . Thus, we may assume that for some  $a \geq n$ , (e)  $\Box_a^N (B \rightarrow C) \wedge \Diamond_a^N B$ . Combining this with (a), we find: (f)  $\Box_a^N (B \rightarrow D) \wedge \Diamond_a^N B$ . From this we easily find the desired (c).

Both rL2 and the  $\exists \Sigma_1^b$ -condition are easy.

The other direction is immediate from Theorem 4.2.  $\square$

The following two applications are taken from [59]. We remind the reader that every essentially reflexive theory  $U$  has Orey sentences. This means that, there is a sentence  $O$  such that  $\top \triangleright_U O$  and  $\top \triangleright_U \neg O$ . It follows from Theorem 4.3, that there are HBL-predicates  $\Delta_0$  and  $\Delta_1$  such that  $U \vdash \Delta_0 \perp \leftrightarrow \neg \Delta_1 \perp$ .

Both Per Lindström and Robert Solovay have shown that interpretability over an essentially reflexive theory is complete  $\Pi_1$ . Inspecting the proof of Theorem 4.3 we can see that we can reduce the question whether  $A \triangleright_U B$  to the question whether the specific predicate  $\Delta$  as constructed in the proof is a HBL-predicate. Hence the question whether a predicate is HBL is complete  $\Pi_2$ .

#### 4.2. $\Pi_1$ -conservativity

Suppose  $\Gamma$  is a set of arithmetical sentences. We define  $\Gamma$ -conservativity. Let  $N : S_2^1 \triangleleft U$  and  $M : S_2^1 \triangleleft V$ . Then:

- $(U, N) \triangleright_\Gamma (V, M)$  iff, for all  $\Gamma$ -sentences  $C$ , if  $V \vdash C^M$ , then  $U \vdash C^N$ .
- $A \triangleright_{U, N, \Gamma} B$  iff  $(U + A, N) \triangleright_\Gamma (U + B, N)$ .
- In case  $U, N$  are given in the background we will write  $A \triangleright_\Gamma B$  for  $A \triangleright_{U, N, \Gamma} B$ .

The logic of  $\Pi_1$ -conservativity was studied by Petr Hájek and Franco Montagna in two papers [19,20]. They proved the arithmetical completeness of  $\text{ILM}$  for extensions of  $\text{I}\Sigma_1$ . A careful analysis of precisely what principles are involved in the proof can be found in [6]. The basic system for which the proof works is  $\text{III}_1^- + \text{Exp}$ . In this section we prove that  $\triangleright_{\text{hbl}_0}$  coincides with  $\triangleright_{\Pi_1}$  for extensions of  $\text{EA}$ , aka  $\text{I}\Delta_0 + \text{Exp}$ .

**Theorem 4.4.** *Consider any theory  $U$  and any interpretation  $N : \mathbf{S}_2^1 \triangleleft U$ . Suppose  $A \triangleright_{\text{hbl}_0} B$  w.r.t.  $U, N$ . Then,  $A \triangleright_{\forall\Pi_1^b} B$  w.r.t.  $U, N$ .*

**Proof.** Let  $U$  and any  $N : \mathbf{S}_2^1 \triangleleft U$  be given. Suppose  $A \triangleright_{\text{hbl}_0} B$  w.r.t.  $U, N$ . Let  $\Delta$  be a  $\text{HBL}_0$  predicate for  $U, N$  and let  $P$  be a  $\forall\Pi_1^b$ -sentence and let  $S$  be the negation of  $P$ . We have:

$$\begin{aligned}
 B \vdash P^N &\Rightarrow \Delta \perp \vdash P^N \\
 &\Rightarrow S^N \vdash \neg \Delta \perp \\
 &\Rightarrow \Delta S^N \vdash \Delta \neg \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow \Delta S^N \vdash \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow S^N, \Delta \top \vdash \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow S^N, \Delta \top \vdash \perp \\
 &\Rightarrow \Delta \top \vdash P^N \\
 &\Rightarrow A \vdash P^N
 \end{aligned}$$

Hence  $B$  is  $\forall\Pi_1^b$ -conservative over  $A$  w.r.t.  $U, N$ .  $\square$

If we have the totality of exponentiation in  $N$ , then we can transform  $\forall\Pi_1^b$ -conservativity into  $\Pi_1$ -conservativity. For completeness sake we reproduce the simple argument.

**Theorem 4.5.** *Consider any theory  $U$  and any interpretation  $N : \text{EA} \triangleleft U$ . Suppose  $A \triangleright_{\text{hbl}_0} B$  w.r.t.  $U, N$ . Then  $A \triangleright_{\Pi_1} B$  w.r.t.  $U, N$ .*

**Proof.** Let  $U$  and any  $N : \text{EA} \triangleleft U$  be given. Suppose  $A \triangleright_{\text{hbl}_0} B$  w.r.t.  $U, N$ . Let  $\Delta$  be a  $\text{HBL}_0$  predicate for  $U, N$  and let  $P$  be a  $\Pi_1$ -sentence and let  $S$  be the negation of  $P$ . We have  $\text{EA} \vdash S \leftrightarrow S_0$ , for some  $S_0$  in  $\exists\Sigma_1^b$ . We have:

$$\begin{aligned}
 B \vdash P^N &\Rightarrow \Delta \perp \vdash P^N \\
 &\Rightarrow S^N \vdash \neg \Delta \perp \\
 &\Rightarrow S_0^N \vdash \neg \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow \Delta S_0^N \vdash \Delta \neg \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow \Delta S_0^N \vdash \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow S_0^N, \Delta \top \vdash \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow S^N, \Delta \top \vdash \Delta \perp \text{ and } S^N \vdash \neg \Delta \perp \\
 &\Rightarrow S^N, \Delta \top \vdash \perp \\
 &\Rightarrow \Delta \top \vdash P^N \\
 &\Rightarrow A \vdash P^N
 \end{aligned}$$

Hence  $B$  is  $\Pi_1$ -conservative over  $A$  w.r.t.  $U, N$ .  $\square$

It would be nice to prove a converse of [Theorem 4.4](#). However, we could not do it. In stead we prove a converse of [Theorem 4.5](#). To prove this converse we need to develop some machinery. Our strategy is to develop an analogue of restricted provability and then simply mimic the proofs we gave for the case of interpretability and transducibility.

**Open Question 4.6.** Do we have the converse of [Theorem 4.4](#)? I.o.w., consider any theory  $U$  and any  $N : S_2^1 \triangleleft U$ . Suppose  $A \triangleright_{\forall\Pi_1^b} B$  w.r.t.  $U, N$ . Do we have:  $A \blacktriangleright_{\text{hbl}_0} B$  w.r.t.  $U, N$ ?  $\square$

Let  $\phi$  be a formula of propositional logic. We define:  $\text{subst}_U(\phi)$  is the set of all  $\sigma : \text{FV}(\phi) \rightarrow \text{sent}_U$ . We write  $\text{taut}(\phi)$  for ‘ $\phi$  is a tautology’ and  $\Box_{\text{prop}}$  for provability in propositional logic.

**Lemma 4.7.** Suppose  $U$  is any theory and  $N : S_2^1 \triangleleft U$ . We have:

- i.  $\text{EA} \vdash \forall \phi (\text{taut}(\phi) \rightarrow \Box_{\text{prop}} \phi)$ ,
- ii.  $\text{EA} \vdash \forall \phi \forall \sigma \in \text{subst}_U(\phi) (\text{taut}(\phi) \rightarrow \Box_U \sigma(\phi))$ ,
- iii.  $\text{EA} \vdash \forall \phi (\neg \text{taut}(\phi) \rightarrow \Box_U \neg \text{taut}^N(\phi))$ ,
- iv.  $\text{EA} \vdash \forall \phi \forall \sigma \in \text{subst}_U(\phi) \Box_U (\text{taut}^N(\phi) \rightarrow \sigma(\phi))$ .

**Proof.** The proof of (i) is simply the formalization of the usual completeness proof of propositional logic. Item (ii) is a direct consequence of (i). Item (iii) is an instance of  $\Sigma_1$ -completeness. Item (iv) follows from (ii) and (iii).  $\square$

We define  $\text{sub}_0$  is follows:

- $\text{sub}_0(A) := \{A\}$  if  $A$  is of the form  $Qx B$ , for  $Q \in \{\forall, \exists\}$ , or  $P\vec{t}$ , where  $P\vec{t}$  is an atomic sentence.
- $\text{sub}_0(B \wedge C) := \text{sub}_0(B) \cup \text{sub}_0(C) \cup \{(B \wedge C)\}$ , and similarly for the other propositional connectives.

The set  $\text{at}_0(A)$  is the set of all  $B$  in  $\text{sub}_0(A)$  of the form  $Qx C$  or  $P\vec{t}$ , where  $P$  is atomic. We define the function  $\theta$  by  $\theta(A) := p_{r_{A\neg}}$ , if  $A$  is of the form  $Qx B$  or  $P\vec{t}$ , where  $P$  is atomic, and  $\theta$  commutes with the propositional connectives. Suppose  $\nu : p_{r_{A\neg}} \mapsto A$ . Then  $\nu(\theta(B)) = B$ . Let  $\text{taut}_U^*(A) := \text{taut}(\theta(A))$ . We have:

**Lemma 4.8.** Suppose  $U$  is any theory and  $N : S_2^1 \triangleleft U$ . We have:

$$\text{EA} \vdash \forall A \Box_U (\text{taut}_U^{*N}(A) \rightarrow A).$$

**Proof.** The lemma is immediate by [Lemma 4.7\(iv\)](#).  $\square$

We will employ a  $\Sigma_1$ -truth predicate in the definition of our restricted-provability-analogue. We note that we have  $S_2^1 \vdash \text{true}_{\Sigma_1}(S) \rightarrow S$  and  $\text{EA} \vdash S \rightarrow \text{true}_{\Sigma_1}(S)$ . See [\[21, V5\(b\)\]](#). Suppose  $\text{true}_{\Sigma_1}(x)$  is of the form  $\exists y \text{true}_0(y, x)$ , where  $\text{true}_0$  is  $\Delta_0$ . We write  $\text{true}_{\Sigma_1}^z(x)$  for:  $\exists y \leq z \text{true}_0(y, x)$ .

Let  $S^*(A)$  be the set of  $S$  in  $\Sigma_1$  such that  $S^N$  is in  $\text{at}_0(A)$ . Here we assume that all such formulas start with an existential quantifier. Let  $S^*(X) := \bigcup_{A \in X} S^*(A)$ . We write  $X^N$  for the set of  $B^N$  such that  $B$  is in  $X$ . We write  $\Box$  for  $\Box_U$  and  $\text{taut}^*$  for  $\text{taut}_U^*$ . Let  $Y_x := \{B \mid \exists p < x \text{proof}_U(p, B)\}$ . We define:

$$\blacksquare_x A := \exists S \subseteq S^*(Y_x \cup \{A\}) \exists z (\forall S \in \mathcal{S} \text{true}_{\Sigma_1}^z(S) \wedge \text{taut}^*(\bigwedge (Y_x \cup \mathcal{S}^N) \rightarrow A)).$$

The business with variable ‘ $z$ ’ is just a trick to avoid the use of  $\Sigma_1$ -collection. In case we do have  $\Sigma_1$ -collection in the ambient theory we can omit ‘ $z$ ’ from the definition. We collect the basic facts about  $\blacksquare_x$  in a lemma.

**Lemma 4.9.** Suppose  $U$  is any theory and  $N : S_2^1 \triangleleft U$ . The variable ‘ $S$ ’ ranges over  $\Sigma_1$ -sentences, that begin with an existential quantifier. We have:

- i.  $\blacksquare_x A$  is  $\Sigma_1$ .
- ii.  $\text{EA} \vdash \forall A (\Box A \rightarrow \exists x \blacksquare_x A)$ .
- iii.  $\text{EA} \vdash \forall A (\Box A \rightarrow \exists x \Box \blacksquare_x^N A)$ .
- iv.  $\text{EA} \vdash \forall S, x (\text{true}_{\Sigma_1}(S) \rightarrow \blacksquare_x S^N)$ .
- v.  $\text{EA} \vdash \forall x, A, B ((\blacksquare_x A \wedge \blacksquare_x (A \rightarrow B)) \rightarrow \blacksquare_x B)$ .
- vi.  $\text{EA} \vdash \forall x, A (\Box(\blacksquare_x^N A \rightarrow A))$ .
- vii.  $\text{EA} \vdash \forall A (\exists x \blacksquare_x A \rightarrow \Box A)$ .
- viii.  $\text{EA} \vdash \forall A (\exists x \blacksquare_x A \leftrightarrow \Box A)$ .

**Proof.** Items (i) and (ii) are trivial. Item (iii) follows from (i) and (ii) by  $\Sigma_1$ -completeness. (iv) is again trivial.

We address item (v). Reason in EA. Consider  $A$ ,  $B$  and  $x$ . Suppose  $\blacksquare_x A$  and  $\blacksquare_x (A \rightarrow B)$ . We have  $\mathcal{S}_0 \subseteq \mathcal{S}^*(A)$  and  $\mathcal{S}_1 \subseteq \mathcal{S}^*(A \rightarrow B)$  and  $z_0$  and  $z_1$  such that: all elements of  $\mathcal{S}_0$  are true witnessed below  $z_0$  and all elements of  $\mathcal{S}_1$  are true witnessed below  $z_1$  and  $\text{taut}^*(\bigwedge(Y_x \cup \mathcal{S}_0^N) \rightarrow A)$  and  $\text{taut}^*(\bigwedge(Y_x \cup \mathcal{S}_1^N) \rightarrow (A \rightarrow B))$ . Thus, clearly, all elements of  $\mathcal{S}_0 \cup \mathcal{S}_1$  are true witnessed below  $z := \max(z_0, z_1)$ . Moreover,  $\text{taut}^*(\bigwedge(Y_x \cup \mathcal{S}_0^N \cup \mathcal{S}_1^N) \rightarrow B)$ . Let  $\mathcal{S}_2 := (\mathcal{S}_0 \cup \mathcal{S}_1) \cap \mathcal{S}^*(B)$ . By elementary propositional logic we find that  $\text{taut}^*(\bigwedge(Y_x \cup \mathcal{S}_2^N) \rightarrow B)$  (since the atoms corresponding to elements of  $\mathcal{S}_0 \cup \mathcal{S}_1$  that are not in  $\theta(B)$  are irrelevant for the truth of  $\theta(B)$  for a given assignment). The elements of  $\mathcal{S}_2$  are witnessed below  $z$ . So  $\blacksquare_x B$ .

We prove item (vi). We reason in EA. Consider any  $A$  in the language of  $U$  and any  $x$ . We have, using Lemma 4.8:

$$\begin{aligned}
 & \Box(\blacksquare_x^N A \rightarrow [\exists \mathcal{S} \subseteq \mathcal{S}^*(Y_x \cup \{A\}) \exists z \\
 & \quad (\forall S \in \mathcal{S} \text{ true}_{\Sigma_1}^z(S) \wedge \text{taut}^*(\bigwedge(Y_x \cup \mathcal{S}^N) \rightarrow A))]^N \\
 & \rightarrow \bigvee_{\mathcal{S} \subseteq \mathcal{S}^*(Y_x \cup \{A\})} (\bigwedge_{S \in \mathcal{S}} \text{true}_{\Sigma_1}^N(S) \wedge \text{taut}^{*N}(\bigwedge(Y_x \cup \mathcal{S}^N) \rightarrow A)) \\
 & \rightarrow \bigvee_{\mathcal{S} \subseteq \mathcal{S}^*(Y_x \cup \{A\})} (\bigwedge \mathcal{S}^N \wedge (\bigwedge(Y_x \cup \mathcal{S}^N) \rightarrow A)) \\
 & \rightarrow \bigvee_{\mathcal{S} \subseteq \mathcal{S}^*(Y_x \cup \{A\})} (\bigwedge \mathcal{S}^N \wedge (\bigwedge Y_x \rightarrow A)) \\
 & \rightarrow (\bigwedge Y_x \rightarrow A) \\
 & \rightarrow A)
 \end{aligned}$$

Finally (vii) follows by combining (iii) and (vi) and (viii) is simply the combination of (ii) and (vii).  $\square$

With our new notion of ‘restricted provability’ in hand, we can now proceed to give an ‘Orey Hájek characterization’ for  $\Pi_1$ -conservativity. We have  $\Pi_1$  here rather than  $\forall \Pi_1^b$  because we need the totality of exponentiation to get everything going.

We write  $\blacksquare_{V,M,n}$  for  $\blacksquare$  defined w.r.t.  $V, M$ . We have:

**Theorem 4.10** (Orey-Hájek for  $\Pi_1$ -conservativity). Consider  $U, N$  and  $V, M$ , where  $N$  is an interpretation of EA in  $U$  and  $M$  is an interpretation of EA in  $V$ . Then,  $(U, N) \triangleright_{\Pi_1} (V, M)$  iff, for all  $n$ , we have  $U \vdash \blacklozenge_{V,M,n}^N \top$ .

**Proof.** From left to right: Suppose  $(U, N) \triangleright_{\Pi_1} (V, M)$ . By Lemma 4.9(vi) we have, for any  $n$ , that  $V \vdash \blacklozenge_{V,M,n}^M \top$ . Hence we also have  $U \vdash \blacklozenge_{V,M,n}^N \top$ .

From right to left: Suppose, for all  $n$ ,  $U \vdash \blacklozenge_{V,M,n}^N \top$ . Suppose  $V \vdash P^M$ , for  $P$  in  $\Pi_1$ . It follows that  $U \vdash \blacksquare_{V,M,n^*}^N P^M$ , for sufficiently large  $n^*$ . We can write  $P$  as  $\neg S$ , where  $S$  is in  $\Sigma_1$ . Reason in  $U$ . Suppose  $S^N$ .

Then, we have  $\blacksquare_{V,M,n}^N S^M$ . So,  $\blacksquare_{V,M,n}^N \perp$ . Quod non. Hence, we may conclude  $P^N$ . Leaving  $U$ , we see that  $U \vdash P^N$ .  $\square$

**Open Question 4.11.** Consider  $U, N$  and  $V, M$ , where  $N$  is an interpretation of  $S_2^1$  in  $U$  and  $M$  is an interpretation of  $\text{EA}$  in  $V$ . Can we prove the following?  $(U, N) \triangleright_{\Pi_1^1} (V, M)$  iff, for all  $n$ , we have  $U \vdash \blacklozenge_{V,M,n}^N \top$ .  $\square$

Finally we give our main theorem.

**Theorem 4.12.** Suppose  $U$  is a theory and  $N : \text{EA} \triangleleft U$ . Then,  $A \blacktriangleright_{\text{hbl}_0} B$  iff  $A \triangleright_{\Pi_1} B$ .

**Proof.** From left to right. This is Theorem 4.5.

From right to left. Suppose  $A \triangleright_{\Pi_1} B$ . By the ‘unformalized’ version of Lemma 4.9(vi), we have: for all  $n$ ,  $\vdash B \rightarrow \blacklozenge_n^N B$  and, hence  $(\dagger)$  for all  $n$ ,  $\vdash A \rightarrow \blacklozenge_n^N B$ . We define the following predicate:

$$\blacktriangle C :\leftrightarrow (B \vee (A \wedge \exists x (\blacksquare_x(B \rightarrow C) \wedge \blacklozenge_x B))).$$

We claim that  $\blacktriangle : A \blacktriangleright_{\text{hbl}_0} B$  (w.r.t. for  $U, N$ ). It is easy to see that  $\vdash \blacktriangle \perp \leftrightarrow B$ ,  $\vdash \blacktriangle \top \leftrightarrow (A \vee B)$  and  $\vdash \blacktriangle \blacktriangle \perp \leftrightarrow (A \vee B)$ .

Suppose  $C \vdash D$ . It follows that  $(\ddagger)$  for some  $m$ ,  $\vdash \blacksquare_m(C \rightarrow D)$ . Reason in  $U$ . Suppose  $\blacktriangle C$ . In case we have  $B$ , we are immediately done. Suppose  $\neg B$ . In that case, we have  $A$  and  $\exists x (\blacksquare_x(B \rightarrow C) \wedge \blacklozenge_x B)$ . Suppose  $\blacksquare_{x_0}(B \rightarrow C)$  and  $\blacklozenge_{x_0} B$ . We may assume, by  $(\dagger)$  that  $x_0 \geq m$ . By  $(\ddagger)$  it follows that  $A$  and  $\blacksquare_{x_0}(B \rightarrow D)$  and  $\blacklozenge_{x_0} B$ . So  $\blacktriangle D$ .

Reason in  $U$ . Suppose  $\blacktriangle C$  and  $\blacktriangle(C \rightarrow D)$ . In case  $B$ , we immediately have  $\blacktriangle D$ . Suppose  $\neg B$ . It follows that  $A$  and for some  $x_0, x_1$ , we have  $\blacksquare_{x_0}(B \rightarrow C)$  and  $\blacksquare_{x_1}(B \rightarrow (C \rightarrow D))$  and  $\blacklozenge_{x_0} B$  and  $\blacklozenge_{x_1} B$ . Let  $x := \max(x_0, x_1)$ . We find:  $\blacksquare_x(B \rightarrow C)$  and  $\blacksquare_x(B \rightarrow (C \rightarrow D))$  and  $\blacklozenge_x B$ . It follows that  $A$  and  $\blacksquare_x(B \rightarrow D)$  and  $\blacklozenge_x B$ . I.o.w.,  $\blacktriangle D$ .  $\square$

We have seen, in the previous subsection, that for sententially essentially reflexive theories, interpretability and HBL-transducibility coincide. In the present subsection, we have seen that for extensions of  $\text{EA}$ ,  $\Pi_1$ -conservativity and the  $\text{HBL}_0$ -transducibility coincide.

In the next section we will provide an example that illustrates that L-transducibility does *not* coincide with interpretability for a wide range of theories.

## 5. The Kreisel condition and a separating example

Suppose we have a theory  $U$  and an interpretation  $N$  of  $\text{EA}$  in  $U$ . We assume that the theory  $U$  is  $\Delta_1^b$ -axiomatized. As before, the interpretation  $N$  provides us the Gödel numbers we use. In this section we want to achieve two things at once. In the first place, we want to produce a  $\Sigma_1$ -predicate  $\Box$  for  $U$  such that  $\Box^N$  is an L-predicate that satisfies the Kreisel condition:  $U \vdash \Box^N A$  iff  $U \vdash A$ , for all  $U$ -sentences  $A$ . In the second place, we want  $\Box^N \perp$  to be a separating example between  $\blacktriangleright$  and  $\triangleright$ . Thus, we want:  $U \blacktriangleright (U + \Box^N \perp)$ , but  $U \not\triangleright (U + \Box^N \perp)$ .

Let  $P$  be any formula defining a set of  $N$ -numbers in  $U$ . We assume that  $P$  starts with a quantifier. Note that we can always add a vacuous quantifier to obtain the desired effect. We treat  $P$  as a modal operator, writing  $\Delta A$  for  $P(\ulcorner A \urcorner)$ . Note that, for the moment, we do not demand any further properties from  $P$ .

Consider any set of sentences  $Z$  in the language of  $U$ . The set  $Z$  generates a propositional language as follows. First we define  $\text{sub}(Z)$  as the smallest set  $X$  such that:



- i.  $Z \subseteq X$ ,
- ii. if  $A \wedge B$  is in  $X$  then  $A$  and  $B$  are in  $X$ , and similarly for the other propositional connectives.
- iii. if  $\Delta A$  is in  $X$ , then so is  $A$ .

In our set-up, we treat the formulas starting with quantifiers as atoms. Consider any set of sentences  $Z$ . We define  $Z \vdash_0 C$ , if  $C$  follows from  $Z$  using modus ponens and  $\Delta$ -necessitation, i.e., the rule that if we have derived  $A$ , we may infer  $\Delta A$ . A  $\vdash_0$ -proof from  $Z$  is simply a sequence of formulas  $D_0, \dots, D_{k-1}$ , where that  $D_i$  are either in  $Z$  or follow from earlier elements of the sequence by our two rules.

Suppose  $Z$  is finite. Consider any  $\vdash_0$ -proof  $\pi$  from  $Z$ . Let  $\gamma$  be an occurrence-as-subconclusion of a formula  $C$  in  $\pi$ . We note that if  $C$  is in  $\text{sub}(Z)$ , then all formulas occurring above  $\gamma$  as subconclusions are subformulas of formulas in  $Z$ . If  $C$  is not in  $\text{sub}(Z)$ , then  $C$  is of the form  $\Delta D$  and the last rule applied is  $\Delta$ -necessitation.

Thus, any proof witnessing  $Z \vdash_0 A$  has the following form:  $A = \Delta^n B$  ( $n$  may be 0), where  $B$  is subformula of formula in  $Z$ . From  $B$  to  $A$  we have necessitation inferences, and the proof of  $B$  contains only elements of  $\text{sub}(Z)$ .

If a  $\vdash_0$ -proof containing only elements of  $\text{sub}(Z)$  is longer than the number of subformulas of formulas of  $Z$ , then a certain subconclusion will occur twice sequentially. Thus we can shorten the proof by omitting all but the first occurrence of the subconclusion. Hence, proofs containing only subformulas of formulas in  $Z$  can be reduced to proofs with as length at most the number of subformulas of formulas in  $Z$ .

We may conclude that  $Z \vdash_0 A$  is decidable. We can easily see that our decidability proof can be formalized in EA.

Let  $Y_n$  be the set of  $A$  such that, for some  $p < n$ ,  $\text{proof}_U(p, A)$ . Let  $\boxplus_x^P A$  stand for (the arithmetization of)  $Y_x \vdash_0 A$ , and let  $\boxplus^P A$  stand for  $\exists x \boxplus_x^P A$ . We note that  $P$  only occurs coded in the definition of  $\boxplus_x^P$  and  $\boxplus^P$ .

Consider any  $\Sigma_1$ -sentence  $S$  of the form  $\exists x S_0(x)$ , where  $S_0$  is  $\Delta_0(\text{exp})$ . Using the Gödel Fixed Point Lemma, we find a formula  $\Box$  (or, more explicitly,  $\Box^{[S]}$ ) with:

$$\text{EA} \vdash \Box A \leftrightarrow \boxplus^{\Box^N} A < S.$$

Note that we take  $P := \Box^N$ . We define  $\Box^\perp A$  by  $S \leq \boxplus^{\Box^N} A$ .

**Theorem 5.1.** *We have:*

- i.  $\text{EA} \vdash \forall A, B ((\Box A \wedge \Box(A \rightarrow B)) \rightarrow \Box B)$ ,
- ii.  $\text{EA} \vdash \forall A (\Box A \rightarrow \Box \Box^N A)$ ,
- iii.  $\text{EA} \vdash \forall A (\Box A \rightarrow \Box \Box^N A)$ ,
- iv.  $\text{EA} \vdash \forall A (\Box A \rightarrow \Box A)$ .
- v.  $\text{EA} \vdash \neg S \vdash \forall A (\Box A \leftrightarrow \Box A)$ .
- vi. EA verifies that, if  $S$  is false, then,  $\Box^N$  is an  $L$ -predicate for  $U$ .
- vii.  $\text{EA} \vdash \Box A \rightarrow (\Box A \vee \Box^\perp A)$ .

**Proof.** We reason in EA. We write  $s$  for the minimal witness of  $S$ . In case  $\neg S$ , we treat  $s$  as  $\infty$  in the obvious way.

Ad (i): Suppose  $\Box A$  and  $\Box(A \rightarrow B)$ . It follows that, for some  $x < s$ , we have  $\boxplus_x^{\Box^N} A$  and  $\boxplus_x^{\Box^N} (A \rightarrow B)$ . Hence, since  $\boxplus_x^{\Box^N}$  is closed under modus ponens by construction, we find  $\boxplus_x^{\Box^N} B$ . Ergo,  $\Box B$ .

Ad (ii): Suppose  $\Box A$ . It follows that, for some  $x < s$ , we have  $\boxplus_x^{\Box^N} A$ . Since  $\boxplus_x^{\Box^N}$  is closed under  $\Box^N$ -necessitation by construction, we find  $\boxplus_x^{\Box^N} \Box^N A$ . Ergo,  $\Box \Box^N A$ .

Ad (iii): This is just  $\Sigma_1$ -completeness.

Ad (iv): Suppose  $\Box A$ . Then, for some  $x < s$ , we have  $\boxplus_x^N A$ . We prove by induction on proof-length that, for every  $\vdash_0$ -proof  $p$  from  $Y_x$  of a  $B$ , there is a matching ordinary proof  $q$  of  $B$ . To make the induction possible, we need a multi-exponential bound on the  $q$ . We will discuss this bound after describing the transformations. We note that we can consider the  $\vdash_0$ -proof  $p$  as the witness for  $\boxplus_x^N B$ , since the  $U$ -proofs needed for verifying that an element of the proof is in  $Y_x$  are all bounded by  $x$ .

In case  $B$  is in  $Y_x$ , we are guaranteed a proof  $q < x$  of  $B$ .

Suppose we have concluded  $B$  from  $C$  and  $C \rightarrow B$ . Say, we have  $\vdash_0$ -proofs  $p_0$  of  $C$  and  $p_1$  of  $C \rightarrow B$ , then by the induction hypothesis we have proofs  $q_0$  of  $C$  and  $q_1$  of  $C \rightarrow B$ . Clearly, we can find a proof  $q$  of  $B$  with length linear in the lengths of  $q_0$  and  $q_1$ .

Suppose we have concluded  $B = \Box^N C$  from  $C$ . Suppose our  $\vdash_0$ -proof of  $C$  is  $p$ . Clearly  $p$  witnesses  $\Box C$ . So we can construct an ordinary proof of order  $2^{2^p}$  to show  $\Box \Box^N C$  — following the usual proof of  $\Sigma_1$ -completeness. (Note that we do not need the Induction Hypotheses here.)

On the basis of the two transformations, we can easily see that we can estimate the ordinary proofs  $q$  by  $2^{2^p}$ , where  $p$  is the  $\vdash_0$ -proof from which they are derived.

Ad (v): This is immediate using (iv).

Ad (vi): We reason in EA. By (i) and (iii), we have that:

$$\Box((\Box^N A \wedge \Box^N(A \rightarrow B)) \rightarrow \Box^N B) \text{ and } \Box(\Box^N A \rightarrow \Box^N \Box^N A).$$

Suppose  $\neg S$  and  $\Box A$ . Then, by (v), we find that  $\Box A$ . So, by (iii),  $\Box \Box^N A$ .

Ad (vii): This is immediate since  $\text{EA} \vdash \Box A \rightarrow \boxplus A$ .  $\square$

Next, we find using the Gödel Fixed Point Lemma, a sentence  $R$  such that:

$$\text{EA} \vdash R \leftrightarrow \exists C (\Box \Box^{[R], N} C \leq \boxplus^{\Box^{[R], N}} C).$$

Inspecting the fixed point construction we may arrange it so that  $R$  is of the form:

$$\exists p \exists C < p (\text{proof}(p, \Box^{t, N} C) \wedge \forall x < p \neg \boxplus_x^{t, N} C),$$

where  $t$  is an elementary term that evaluates to (the Gödel number of)  $R$ . We note that  $R$  is of the form  $\exists p R_0(p)$ , where  $R_0$  is  $\Delta_0(\text{exp})$ .

Finally we define  $\Box A :\leftrightarrow \Box^{[R]} A$ . So,

$$\text{EA} \vdash R \leftrightarrow \exists C (\Box \Box^N C \leq \boxplus^{\Box^N} C).$$

**Theorem 5.2.** *We have:*

- $\text{EA} \vdash R \rightarrow \Box \perp$ ,
- $\text{EA} \vdash \Diamond \top \rightarrow (\Box A \leftrightarrow \Box A)$ ,
- $\text{EA} \vdash \Box \Box^N A \leftrightarrow \Box A$ ,
- Suppose that  $U$  is EA-verifiably sequential and essentially reflexive w.r.t.  $N$ . Then,  $\text{EA} \vdash \top \triangleright \Box^N \perp \rightarrow \Box \perp$ .

**Proof.** Ad (a): We reason in EA. Suppose  $R$ . It follows that, for some  $C$ , we have  $\Box \Box^N C \leq \boxplus^{\Box^N} C$ . It follows that  $\Box \Box^N C$ , i.e. (a)  $\Box(\boxplus^{\Box^N} C < R)^N$ . On the other hand,  $R$  implies  $R \leq \boxplus^{\Box^N} C$ . So, by  $\Sigma_1$ -completeness, (b)  $\Box(R \leq \boxplus^{\Box^N} C)^N$ . By (a) and (b), we may conclude that  $\Box \perp$ .

Ad (b): The desired result is immediate by (a) and [Theorem 5.1\(v\)](#).

Ad (c): The right-to-left direction is immediate from [Theorem 5.1\(iii\)](#). We prove left-to-right. We reason in EA. Suppose  $\Box \Box^N A$ . We want to show  $\Box A$ . In case we have  $R$ , we are immediately done by (a). If we have  $\neg R$ , it follows that we cannot have  $\Box \Box^N A < \Box \Box^N A$ . So, we must have  $\Box \Box^N A$ , and hence  $\Box A$ . By [Theorem 5.1\(v\)](#), we find  $\Box A$ .

Ad (d): Suppose that  $U$  is EA-verifiably sequential and essentially reflexive w.r.t.  $N$ . We reason in EA. Suppose  $\top \triangleright \Box^N \perp$ . Then also  $\top \triangleright \neg \Box^{\perp, N} \perp$ . Since  $\Box^{\perp, N} \perp$  is  $\Sigma_1$ , it follows that  $\Box \neg \Box^{\perp, N} \perp$ . By [Theorem 5.1\(vii\)](#),  $\Box(\Box^N \perp \rightarrow (\Box^N \perp \vee \Box^{\perp, N} \perp))$ . Hence,  $(\dagger) \Box(\Box^N \perp \rightarrow \Box^N \perp)$ . It follows that  $\Box(\Box^N \Box^N \perp \rightarrow \Box^N \Box^N \perp)$ . Hence, by (c), we find  $\Box(\Box^N \Box^N \perp \rightarrow \Box^N \perp)$ , and so  $(\ddagger) \Box \Box^N \perp$ . Combining  $(\dagger)$  and  $(\ddagger)$ , we obtain  $\Box \Box^N \perp$ . Hence, again by [Theorem 5.1\(vii\)](#), we may conclude  $\Box \perp$ .  $\square$

Here (c) gives us the promised result that  $\Box^N$  has the Kreisel property. Moreover (d) shows that  $\blacktriangleright$  strictly extends  $\triangleright$ , since we do have  $\top \blacktriangleright \Box^N \perp$ . In fact  $\Box^N \perp$  is a  $\Sigma_1$  Rosser sentence for  $U$ . So, we have an example of a  $\Sigma_1$  Rosser sentence that can be  $\blacktriangleright$ -reached from  $\top$ .

## Appendix A. Basic facts and definitions

In this appendix we explain some basic notions.

### A.1. Theories

Theories are, in this paper, theories of first-order predicate logic, that have a finite signature and that are axiomatized by an axiom set that is represented by a  $\Delta_1^b$ -formula.<sup>5</sup>

The formula specifying the axiom set is part of the data for the theory. Thus, we treat theories *intentionally* and not as mere sets of theorems.

We say that a theory is *finitely axiomatized* if its axiomatization has the form  $\bigvee_{i < n} x = \ulcorner A_i \urcorner$ . Note that  $S_2^1$  may prove that a theory has an axiom-set of, say, less than two axioms, without being able to prove the equivalence of the formula defining the axiom set with any formula of the prescribed form.

Our official signatures are relational, however, via the term-unwinding algorithm, we can also accommodate signatures with functions.

### A.2. Translations and interpretations

We present the notion of *m-dimensional interpretation without parameters*. There are two extensions of this notion: we can consider piecewise interpretations and we can add parameters. We will not treat these extensions in this paper.

Consider two signatures  $\Sigma$  and  $\Theta$ . An  $m$ -dimensional translation  $\tau : \Sigma \rightarrow \Theta$  is a quadruple  $\langle \Sigma, \delta, \mathcal{F}, \Theta \rangle$ , where  $\delta(v_0, \dots, v_{m-1})$  is a  $\Theta$ -formula and where for any  $n$ -ary predicate  $P$  of  $\Sigma$ ,  $\mathcal{F}(P)$  is a formula  $A(\vec{v}_0, \dots, \vec{v}_{n-1})$  in the language of signature  $\Theta$ , where  $\vec{v}_i = v_{i0}, \dots, v_{i(m-1)}$ . Both in the case of  $\delta$  and  $A$  all free variables are among the variables shown. Moreover, if  $i \neq j$  and  $k \neq \ell$ , then  $v_{ik}$  is syntactically different from  $v_{j\ell}$ .

We demand that we have  $\vdash \mathcal{F}(P)(\vec{v}_0, \dots, \vec{v}_{n-1}) \rightarrow \bigwedge_{i < n} \delta(\vec{v}_i)$ . Here  $\vdash$  is provability in predicate logic. This demand is inessential, but it is convenient to have.

We define  $B^\tau$  as follows:

<sup>5</sup> See [\[10\]](#) or [\[21\]](#) for an explanation of the relevant formula classes.

- $(P(x_0, \dots, x_{n-1}))^\tau := \mathcal{F}(P)(\vec{x}_0, \dots, \vec{x}_{n-1})$ .
- $(\cdot)^\tau$  commutes with the propositional connectives.
- $(\forall x A)^\tau := \forall \vec{x} (\delta(\vec{x}) \rightarrow A^\tau)$ .
- $(\exists x A)^\tau := \exists \vec{x} (\delta(\vec{x}) \wedge A^\tau)$ .

There are two worries about this definition. First, what variables  $\vec{x}_i$  on the side of the translation  $A^\tau$  correspond with  $x_i$  in the original formula  $A$ ? The second worry is that substitution of variables in  $\delta$  and  $\mathcal{F}(P)$  may cause variable clashes. These worries are never important in practice: we choose ‘suitable’ sequences  $\vec{x}$  to correspond to variables  $x$ , and we avoid clashes by  $\alpha$ -conversions. However, if we want to give precise definitions of translations and, for example, of composition of translations these problems come into play. These problems are clearly solvable, but they are beyond the scope of this paper.

We allow identity to be translated to a formula that is not identity. There are several important operations on translations.

- $\text{id}_\Sigma$  is the identity translation. We take  $\delta_{\text{id}_\Sigma}(v) := v = v$  and  $\mathcal{F}(P) := P(\vec{v})$ .
- We can compose translations. Suppose  $\tau : \Sigma \rightarrow \Theta$  and  $\nu : \Theta \rightarrow \Lambda$ . Then  $\nu \circ \tau$  or  $\tau \nu$  is a translation from  $\Sigma$  to  $\Lambda$ . We define:
  - $\delta_{\tau\nu}(\vec{v}_0, \dots, \vec{v}_{m_\tau-1}) := \bigwedge_{i < m_\tau} \delta_\nu(\vec{v}_i) \wedge (\delta_\tau(v_0, \dots, v_{m_\tau-1}))^\nu$ .
  - $P_{\tau\nu}(\vec{v}_{0,0}, \dots, \vec{v}_{0,m_\tau-1}, \dots, \vec{v}_{n-1,0}, \dots, \vec{v}_{n-1,m_\tau-1}) := \bigwedge_{i < n, j < m_\tau} \delta_\nu(\vec{v}_{i,j}) \wedge (P(v_0, \dots, v_{n-1})^\tau)^\nu$ .
- Let  $\tau, \nu : \Sigma \rightarrow \Theta$  and let  $A$  be a sentence of signature  $\Theta$ . We define the disjunctive translation  $\sigma := \tau \langle A \rangle \nu : \Sigma \rightarrow \Theta$  as follows. We take  $m_\sigma := \max(m_\tau, m_\nu)$ . We write  $\vec{v} \upharpoonright n$ , for the restriction of  $\vec{v}$  to the first  $n$  variables, where  $n \leq \text{length}(\vec{v})$ .
  - $\delta_\sigma(\vec{v}) := (A \wedge \delta_\tau(\vec{v} \upharpoonright m_\tau)) \vee (\neg A \wedge \delta_\nu(\vec{v} \upharpoonright m_\nu))$ .
  - $P_\sigma(\vec{v}_0, \dots, \vec{v}_{n-1}) := (A \wedge P_\tau(\vec{v}_0 \upharpoonright m_\tau, \dots, \vec{v}_{n-1} \upharpoonright m_\tau)) \vee (\neg A \wedge P_\nu(\vec{v}_0 \upharpoonright m_\nu, \dots, \vec{v}_{n-1} \upharpoonright m_\nu))$ .

Note that in the definition of  $\tau \langle A \rangle \nu$  we used a padding mechanism. In case, for example,  $m_\tau < m_\nu$ , the variables  $v_{m_\tau}, \dots, v_{m_\nu-1}$  are used ‘vacuously’ when we have  $A$ . If we had piecewise interpretations, where domains are built up from pieces with possibly different dimensions, we could avoid padding by building the domain of disjoint pieces with different dimensions.

A translation relates signatures; an interpretation relates theories. An interpretation  $K : U \rightarrow V$  is a triple  $\langle U, \tau, V \rangle$ , where  $U$  and  $V$  are theories and  $\tau : \Sigma_U \rightarrow \Sigma_V$ . We demand: for all axioms  $A$  of  $U$ , we have  $V \vdash A^\tau$ . Here are some further definitions.

- $\text{ID}_U : U \rightarrow U$  is the interpretation  $\langle U, \text{id}_{\Sigma_U}, U \rangle$ .
- Suppose  $K : U \rightarrow V$  and  $M : V \rightarrow W$ . Then,  $KM := M \circ K : U \rightarrow W$  is  $\langle U, \tau_M \circ \tau_K, W \rangle$ .
- Suppose  $K : U \rightarrow (V + A)$  and  $M : U \rightarrow (V + \neg A)$ . Then  $K \langle A \rangle M : U \rightarrow V$  is the interpretation  $\langle U, \tau_K \langle A \rangle \tau_M, V \rangle$ . In an appropriate category  $K \langle A \rangle M$  is a special case of a product.

The notation  $K : U \rightarrow V$  is inspired by the idea of interpretations as arrows in a category. There is also an intuition of interpretability as a generalization of provability. The traditional notations and notions associated to this intuition are:

- $K : U \triangleleft V$  stands for  $K : U \rightarrow V$ .
- $K : V \triangleright U$  stands for  $K : U \rightarrow V$ .
- $U \triangleleft V$  stands for  $\exists K K : U \triangleleft V$ . We say:  $U$  is *interpretable* in  $V$ .

- $V \triangleright U$  stands for  $\exists K K : V \triangleright U$ . We say:  $V$  *interprets*  $U$ .
- $U \equiv V$  stands for  $U \triangleright V$  and  $V \triangleright U$ . We say:  $V$  and  $U$  are *mutually interpretable*.

A basic insight in concerning interpretability is the Gödel–Hilbert–Bernays–Wang–Henkin–Feferman Theorem.

**Theorem A.1.** *Consider  $N : S_2^1 \triangleleft U$ . We assume that  $U$  is  $\Delta_1^b$ -axiomatized. Then, we can construct an interpretation  $H : (U + \Diamond_U^N \top) \triangleright U$ . We call  $H$ : the Henkin interpretation. This interpretation has the additional feature that we can construct inside  $U$  a truth-predicate  $T$  such that for some definable cut  $I$  of  $N$  the commutation conditions for the language coded in  $I$  are  $U$ -verifiable.*

The proof uses the formalized Henkin construction to produce an interpretation  $H : (U + \Diamond_U^N \top) \triangleright U$ . The basic intuition here is, of course, that an interpretation is a uniform internal model construction. The lack of induction in our setting has to be systematically compensated by going to shorter and shorter definable cuts of  $N$ .

### A.3. Sequential theories

A sequential theory provides an interpretation  $N$  of a weak number theory, say  $S_2^1$ , and sequences of all objects of the domain of the theories with projections in  $N$ . We can use these sequences to develop partial satisfaction predicates. Using these we can prove restricted consistency statements of  $U$  in  $U$ .

The notion of sequential theory has an very simple definition discovered by Pavel Pudlák. We first need the definition of a very weak set theory. The theory Adjunctive Set Theory or AS has a binary relation  $\in$ .

AS1  $\vdash \exists x \forall y y \notin x$ ,

AS2  $\vdash \forall x, y \exists z \forall u (u \in z \leftrightarrow (u \in x \vee u = y))$ .

We note that we do not demand extensionality. For example, in AS we could have lots of ‘empty sets’.

An interpretation is *direct* iff it is one-dimensional, unrelativized (that is, it has the trivial domain) and identity preserving (that is, it translates identity to identity).

A theory  $U$  is sequential iff it directly interprets AS. By a substantial bootstrap, we can define, in a sequential theory  $U$ , an interpretation  $N$  of a weak number theory, sequences of all objects, etc.

For details see, for example, [41,42,40,21,54,56].

We can generalize the notion of sequentiality a bit to *poly-sequentiality* by replacing *direct interpretation* in the definition by its obvious generalization to the  $m$ -dimensional case.

### A.4. Complexity measures

In sequential theories we can define partial satisfaction predicates for formulas with complexity below  $n$ , for any  $n$ . The presence of these predicates has as a consequence that for any sequential theory  $U$  and for any  $n$ , we can find an interpretation  $N$  of a weak arithmetic like Buss’  $S_2^1$  in  $U$  such that  $U \vdash \text{con}_n^N(U)$ . See, for example, [52] for more details. We give the relevant definitions of complexity notions.

*Restricted provability* plays an important role in this paper. An  $n$ -proof is a proof from axioms with Gödel number smaller or equal than  $n$  only involving formulas of complexity smaller or equal than  $n$ . To work conveniently with this notion, a good complexity measure is needed. This should satisfy three conditions. (i) Eliminating terms in favor of a relational formulation should raise the complexity only by a fixed standard number. (ii) Translation of a formula via the translation corresponding to an interpretation  $K$  should raise the complexity of the formula by a fixed standard number depending only on  $K$ . (iii) The

tower of exponents involved in cut-elimination should be of height linear in the complexity of the formulas involved in the proof.

A good measure of complexity together with a verification of desideratum (iii) — a form of nesting degree of quantifier alternations — is supplied in the work of Philipp Gerhardy. See [15] and [16]. It is also provided by Samuel Buss in his preliminary draft [11]. Buss also proves that (iii) is fulfilled.

Buss' definition is somewhat more suitable for our purposes, so we will follow his presentation. Buss gives the following formula classes.

- $\Sigma_0^* = \Pi_0^*$  = the class of quantifier-free formulas.
- $\Sigma_n^* ::= \Sigma_{n-1}^* \mid \Pi_{n-1}^* \mid \neg \Pi_n^* \mid (\Sigma_n^* \wedge \Sigma_n^*) \mid (\Sigma_n^* \vee \Sigma_n^*) \mid (\Pi_n^* \rightarrow \Sigma_n^*) \mid \exists \Sigma_n^*$ .
- $\Pi_n^* ::= \Sigma_{n-1}^* \mid \Pi_{n-1}^* \mid \neg \Sigma_n^* \mid (\Pi_n^* \wedge \Pi_n^*) \mid (\Pi_n^* \vee \Pi_n^*) \mid (\Sigma_n^* \rightarrow \Pi_n^*) \mid \forall \Pi_n^*$ .

We may define  $\rho(A)$  as the smallest  $n$  such that  $A$  is in  $\Sigma_n^*$ . This is the same measure, as was employed in [52].

We use  $\text{proof}_{U,n}$  for the proof predicate where only  $U$ -axioms with Gödel numbers  $\leq n$  are allowed and where the formulas occurring in the proof are in the complexity class  $\Gamma_n$  of all formulas of complexity  $\leq n$ . Similarly, we use  $U \vdash_n A$ ,  $\text{con}_n(U)$ ,  $\Box_{U,n} A$ , etc.

We end with some basic facts concerning sequential theories and restricted provability. A finitely axiomatized sequential theory is mutually interpretable with its own restricted consistency over  $S_2^1$ .

**Theorem A.2.** *Suppose  $A$  is finitely axiomatized and sequential. We have:*

$$A \equiv (S_2^1 + \Diamond_{A,\rho(A)} \top).$$

For a proof, see, [42] or [21]. We note that the right-to-left direction of the result is a variant of the Gödel–Hilbert–Bernays–Wang–Henkin–Feferman Theorem. An important point here is that the existence of a truth-predicate for the witnessing Henkin interpretation is lost when we switch from ordinary consistency to restricted consistency. (If this were not the case, we would obtain a contradiction with the Second Incompleteness Theorem.)

We provide an partial analogue of Theorem A.2 for infinitely axiomatized theories. The  $\mathcal{U}$ -functor is given as follows.<sup>6</sup>

- $\mathcal{U}(U) := S_2^1 + \{\Diamond_{U,n} \top \mid n \in \omega\}$ .

The central fact about the  $\mathcal{U}$ -functor is as follows:

**Theorem A.3.** *Suppose  $U$  is sequential. We have:  $U \triangleright_{\text{loc}} V \Leftrightarrow \mathcal{U}(U) \triangleright V$ .*

If we restrict ourselves to sequential theories, the theorem tells us that  $\mathcal{U}$  is the right adjoint of the embedding functor of  $\triangleleft$  considered as a preorder category into  $\triangleleft_{\text{loc}}$  considered as a preorder category. For a proof, see [55]. We note that it follows that  $U \equiv \mathcal{U}(U)$ .

## References

- [1] S. Artemov, L. Beklemishev, Provability logic, in: D. Gabbay, F. Guenther (Eds.), *Handbook of Philosophical Logic*, 2nd ed., vol. 13, Springer, Dordrecht, 2004, pp. 229–403.
- [2] L. Beklemishev, Provability algebras and proof-theoretic ordinals, *Ann. Pure Appl. Logic* 128 (2004) 103–124.

<sup>6</sup> We pronounce  $\mathcal{U}$  as ‘mho’ is such a way that it rhymes with ‘joe’.

- [3] L. Beklemishev, Reflection principles and provability algebras in formal arithmetic, *Russian Math. Surveys* 60 (2) (2005) 197–268.
- [4] L. Beklemishev, The worm principle, in: Z. Chatzidakis, P. Koepke, W. Pohlers (Eds.), *Logic Colloquium'02*, in: *Lecture Notes in Logic*, vol. 27, A.K. Peters and CRC Press, Natick, Massachusetts, 2006, pp. 75–95.
- [5] L. Beklemishev, J. Joosten, M. Vervoort, A finitary treatment of the closed fragment of Japaridze's provability logic, *J. Logic Comput.* 15 (4) (2005) 447–463.
- [6] L. Beklemishev, A. Visser, On the limit existence principles in elementary arithmetic and  $\Sigma_n^0$ -consequences of theories, *Ann. Pure Appl. Logic* 136 (1–2) (2005) 56–74.
- [7] A. Berarducci, The interpretability logic of Peano arithmetic, *J. Symbolic Logic* 55 (1990) 1059–1089.
- [8] G. Boolos, *The Logic of Provability*, Cambridge University Press, Cambridge, 1993.
- [9] G. Boolos, G. Sambin, Provability: the emergence of a mathematical modality, *Studia Logica* 50 (1991) 1–23.
- [10] S. Buss, *Bounded Arithmetic*, Bibliopolis, Napoli, 1986.
- [11] S. Buss, Cut elimination in situ, in: R. Kahle, M. Rathjen (Eds.), *Gentzen's Centenary*, Springer International Publishing, ISBN 978-3-319-10102-6, 2015, pp. 245–277.
- [12] S. Feferman, Arithmetization of metamathematics in a general setting, *Fund. Math.* 49 (1960) 35–92.
- [13] S. Feferman, My route to arithmetization, *Theoria* 63 (3) (1997) 168–181.
- [14] S. Friedman, M. Rathjen, A. Weiermann, Slow consistency, *Ann. Pure Appl. Logic* 164 (3) (2013) 382–393.
- [15] P. Gerhardy, Refined complexity analysis of cut elimination, in: M. Baaz, J. Makowsky (Eds.), *Proceedings of the 17th International Workshop, CSL 2003*, in: *LNCS*, vol. 2803, Springer-Verlag, Berlin, 2003, pp. 212–225.
- [16] P. Gerhardy, The role of quantifier alternations in cut elimination, *Notre Dame J. Form. Log.* 46 (2) (2005) 165–171.
- [17] K. Gödel, Ein Interpretation des intuitionistischen Aussagenkalküls, in: *Ergebnisse eines mathematischen Kolloquiums*, vol. 4, Oxford, 1933, pp. 39–40; reprinted as: An interpretation of the intuitionistic propositional calculus, in: S. Feferman (Ed.), *Gödel Collected Works I*, Publications 1929–1936, Oxford, 1986, pp. 300–303.
- [18] E. Goris, J. Joosten, Modal matters in interpretability logic, *Log. J. IGPL* 16 (4) (2008) 371–412.
- [19] P. Hájek, F. Montagna, The logic of  $\Pi_1$ -conservativity, *Arch. Math. Log. Grundle.forsch.* 30 (1990) 113–123.
- [20] P. Hájek, F. Montagna, The logic of  $\Pi_1$ -conservativity continued, *Arch. Math. Log. Grundle.forsch.* 32 (1992) 57–63.
- [21] P. Hájek, P. Pudlák, *Metamathematics of First-Order Arithmetic*, Perspectives in Mathematical Logic, Springer, Berlin, 1993.
- [22] V. Halbach, H. Leitgeb, P. Welch, Possible-worlds semantics for modal notions conceived as predicates, *J. Philos. Logic* 32 (2) (2003) 179–223.
- [23] V. Halbach, A. Visser, The Henkin sentence, in: M. Manzano, I. Sain, E. Alonso (Eds.), *The Life and Work of Leon Henkin*, Springer, 2014, pp. 249–263.
- [24] D. Hilbert, P. Bernays, *Grundlagen der Mathematik II*, Springer, Berlin, 1939, second edition: 1970.
- [25] L. Horsten, H. Leitgeb, No future, *J. Philos. Logic* 30 (3) (2001) 259–265.
- [26] G. Hughes, M. Cresswell, *A New Introduction to Modal Logic*, Burns & Oates, 1968.
- [27] K. Ignatiev, Partial conservativity and modal logics, Tech. Rep. X-91-04, ILLC, University of Amsterdam, 1991.
- [28] G. Japaridze, The polymodal logic of provability, in: *Intensional Logics and Logical Structure of Theories: Material from the Fourth Soviet–Finnish Symposium on Logic*, Telavi, 1985, pp. 16–48.
- [29] G. Japaridze, D. de Jongh, The logic of provability, in: S. Buss (Ed.), *Handbook of Proof Theory*, North-Holland Publishing Co., Amsterdam, 1998, pp. 475–546.
- [30] G. Japaridze Dzjaparidze, A simple proof of arithmetical completeness for  $\Pi_1$ -conservativity logic, *Notre Dame J. Form. Log.* 35 (1994) 346–354.
- [31] J. Joosten, A. Visser, The interpretability logic of all reasonable arithmetical theories, *Erkenntnis* 53 (1–2) (2000) 3–26.
- [32] M. Kalsbeek, Towards the Interpretability Logic of  $\text{ID}_0 + \text{EXP}$ , *Logic Group Preprint Series*, vol. 61, Faculty of Humanities, Philosophy, Utrecht University, Janskerhof 13, 3512 BL Utrecht, 1991, <http://www.phil.uu.nl/preprints/lgps/>.
- [33] D. Kaplan, R. Montague, A paradox regained, *Notre Dame J. Form. Log.* 1 (1960).
- [34] G. Kreisel, On a problem of Henkin's, *Indag. Math.* 15 (1953) 405–406.
- [35] P. Lindström, Provability logic – a short introduction, *Theoria* 62 (1–2) (1996) 19–61.
- [36] P. Lindström, On Parikh provability: an exercise in modal logic, in: H. Lagerlund, S. Lindström, R. Sliwinski (Eds.), *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, in: *Uppsala Philosophical Studies*, vol. 53, 2006, pp. 53–287.
- [37] M. Löb, Solution of a problem of Leon Henkin, *J. Symbolic Logic* 20 (1955) 115–118.
- [38] F. Montagna, On the algebraization of a Feferman's predicate (the algebraization of theories which express  $\text{Theor}; X$ ), *Studia Logica* 37 (1978) 221–236.
- [39] R. Montague, Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability, *Acta Philos. Fenn.* 16 (1963) 153–167.
- [40] J. Mycielski, P. Pudlák, A. Stern, A Lattice of Chapters of Mathematics (Interpretations Between Theorems), *Memoirs of the American Mathematical Society*, vol. 426, AMS, Providence, Rhode Island, 1990.
- [41] P. Pudlák, Some prime elements in the lattice of interpretability types, *Trans. Amer. Math. Soc.* 280 (1983) 255–275.
- [42] P. Pudlák, Cuts, consistency statements and interpretations, *J. Symbolic Logic* 50 (2) (1985) 423–441.
- [43] J. des Rivières, H. Levesque, The consistency of syntactical treatments of knowledge, in: *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning About Knowledge*, Morgan Kaufmann Publishers Inc., 1986, pp. 115–130.
- [44] V. Shavrukov, The logic of relative interpretability over Peano arithmetic (in Russian), Tech. Rep. Report No. 5, Steklov Mathematical Institute, Moscow, 1988.
- [45] V. Shavrukov, A smart child of Peano's, *Notre Dame J. Form. Log.* 35 (1994) 161–185.
- [46] R. Solovay, Provability interpretations of modal logic, *Israel J. Math.* 25 (1976) 287–304.

- [47] J. Stern, M. Fischer, Paradoxes of interaction?, *J. Philos. Logic* (2014) 1–22.
- [48] V. Švejdar, On provability logic, *Nord. J. Philos. Log.* 4 (2) (2000) 95–116.
- [49] R.H. Thomason, A note on syntactical treatments of modality, *Synthese* 44 (3) (1980) 391–395.
- [50] A. Visser, Peano’s smart children: a provability logical study of systems with built-in consistency, *Notre Dame J. Form. Log.* 30 (2) (1989) 161–196.
- [51] A. Visser, Interpretability logic, in: P. Petkov (Ed.), *Mathematical Logic, Proceedings of the Heyting 1988 Summer School in Varna, Bulgaria*, Plenum Press, Boston, 1990, pp. 175–209.
- [52] A. Visser, The unprovability of small inconsistency, *Arch. Math. Logic* 32 (4) (1993) 275–298.
- [53] A. Visser, An overview of interpretability logic, in: M. Kracht, M. de Rijke, H. Wansing, M. Zakharyashev (Eds.), *Advances in Modal Logic*, in: *CSLI Lecture Notes*, vol. 87, Center for the Study of Language and Information, Stanford, 1998, pp. 307–359.
- [54] A. Visser, Cardinal arithmetic in the style of Baron von Münchhausen, *Rev. Symb. Log.* 2 (3) (2009) 570–589, <http://dx.doi.org/10.1017/S1755020309090261>.
- [55] A. Visser, Can we make the second incompleteness theorem coordinate free, *J. Logic Comput.* 21 (4) (2011) 543–560, <http://dx.doi.org/10.1093/logcom/exp048>, first published online August 12, 2009.
- [56] A. Visser, What is sequentiality?, in: P. Cégielski, C. Cornaros, C. Dimitracopoulos (Eds.), *New Studies in Weak Arithmetics*, in: *CSLI Lecture Notes*, vol. 211, CSLI Publications and Presses Universitaires du Pôle de Recherche et d’Enseignement Supérieur Paris-est, Stanford, 2013, pp. 229–269.
- [57] A. Visser, The Interpretability of Inconsistency, Feferman’s Theorem and Related Results, In *Logic Group Preprint Series*, vol. 318, Faculty of Humanities, Philosophy, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, 2014, <http://www.phil.uu.nl/preprints/lgps/>.
- [58] A. Visser, Peano Corto and Peano Basso: a study of local induction in the context of weak theories, *Math. Log. Q.* 60 (1–2) (2014) 92–117, <http://dx.doi.org/10.1002/malq.201200102>.
- [59] A. Visser, The second incompleteness theorem, reflections and ruminations, in: *Proceedings of “The Scope and Limits of Mathematics II”*, Bristol, 2014, submitted for publication.
- [60] A. Visser, Oracle Bites Theory, *Logic Group Preprint Series*, vol. 324, Faculty of Humanities, Philosophy, Utrecht University, Janskerkhof 13, 3512 BL Utrecht, 2015, <http://www.phil.uu.nl/preprints/lgps/>.