INTERPRETABILITY SUPREMA IN PEANO ARITHMETIC

PAULA HENK AND ALBERT VISSER

ABSTRACT. This paper develops the philosophy and technology needed for adding a supremum operator to the interpretability logic ILM of Peano Arithmetic (PA). It is well-known that any theories extending PA have a supremum in the interpretability ordering. While provable in PA, this fact is not reflected in the theorems of the modal system ILM, due to limited expressive power. Our goal is to enrich the language of ILM by adding to it a new modality for the interpretability supremum. We explore different options for specifying the exact meaning of the new modality. Our final proposal involves a unary operator, the dual of which can be seen as a (nonstandard) provability predicate satisfying the axioms of the provability logic GL.

1. INTRODUCTION

The aim of this paper is to bring closer together the two main approaches to the study of interpretability. As we will explain below, one of them currently lacks expressive power for properly talking about the other.

Given first-order theories T and S, possibly in different languages, we say that T interprets S, and write $T \rhd S$, if there is some structure-preserving translation from the language of S to the language of T, such that the translations of all theorems of S are provable in T.

Interpretability can be seen as *generalized provability*: T is required to prove everything that S proves modulo some well-behaved translation. As such, it allows us to *compare* theories talking about different objects, such as Peano Arithmetic (PA) and Zermelo-Fraenkel set theory (ZF). In fact, the notion of interpretability is a natural candidate for giving a precise meaning to the intuitive idea of one theory being stronger than another one. Seen from a semantic perspective, an interpretation of S in T gives rise to a uniform way of constructing a model of S inside a given model of T. Interpretations therefore give rise to relative consistency proofs.

The study of interpretability may roughly be divided into two traditions, briefly outlined below. For a more comprehensive overview, the reader is referred to [32].

The lattice of degrees. Seeing interpretability as a measuring stick for theories naturally leads one to study the space of all theories ordered by this relation. A *degree* is a collection of all theories that are equally strong as a given theory, i.e. that all mutually interpret each other. We write $[S]$ for the degree of S.

Among the first results concerning interpretability degrees is a strengthening of Gödel's Second Incompleteness Theorem by Feferman. Denote by $\mathrm{Con_{PA}}$ the usual consistency statement of PA. In [4, Theorem 6.5], it is shown that not only is $\mathrm{Con_{PA}}$ unprovable, it is

also "uninterpretable": $[\mathsf{PA}] \not\rhd [\mathsf{PA} + \mathrm{Con}_{\mathsf{PA}}]$. Jeroslow ([12, Theorem 3.1, 3.2]) showed that the degrees intermediate between $[\mathsf{PA}]$ and $[\mathsf{PA} + \mathrm{Con}_{\mathsf{PA}}]$ form a dense partial order. It follows from his work that the interpretability ordering is dense (see [28, p.798]). Montague ([20, Theorem 1]) proved the existstence of an infinte set of finitely axiomatized subtheories of PA, all of which are mutually incomparable with respect to the interpretability ordering.

A systematic study of interpretability degrees was undertaken independently by Švejdar and Lindström ([7, p.402]). Švejdar studied the structure $(\mathscr{V}, \rhd)$ of the degrees of finite extensions of PA. He proved, among other things, that this structure is a distributive lattice ([28, Theorems 4.4, 4.7]). Lindström was concerned with the structure of the degrees of *all* r.e. extensions of PA ([15], [17]). However he also showed that this structure is in fact isomorphic to $(\mathscr{V}, \rhd)$ ([15, p.348, Theorem 3]).

Our interest in provability logic makes it natural to focus on $(\mathscr{V}, \rhd)$. Note first that $[\mathsf{PA} + \top]$ is the minimum and $[\mathsf{PA} + \bot]$ the maximum element of this structure[1]. It is easy to see that the infimum of $[\mathsf{PA}+A]$ and $[\mathsf{PA}+B]$ in $(\mathscr{V}, \rhd)$ is $[\mathsf{PA} + (A \vee B)]$. However the supremum of $[\mathsf{PA}+A]$ and $[\mathsf{PA}+B]$ is in general *not* $[\mathsf{PA} + (A \wedge B)]$. Švejdar shows that the supremum of $[\mathsf{PA}+A]$ and $[\mathsf{PA}+B]$ may taken to be $[\mathsf{PA} + \vartheta]$, where ϑ is obtained by the Diagonal Lemma as a sentence such that[2]:

$$(1) \qquad\qquad \mathsf{PA} \vdash \vartheta \leftrightarrow \forall x \left(\mathrm{Con}_x(\vartheta) \to \mathrm{Con}_x(A) \wedge \mathrm{Con}_x(B) \right).$$

Interpretability logic. Interpretability, like provability, is a syntactical notion, and can therefore be formalized in the language of arithmetic. We can thus ask: which statements concerning provability and interpretability (between its finite extensions) are provable in PA?

It turns out that using *modal logic* allows one to answer this question in a neat way. The system GL of propositional provability logic contains a unary modality $\Box$, interpreted as formalized provability in PA. It follows from the work of Hilbert and Bernays ([11]), Löb ([18]), and Solovay ([27]), that the theorems of GL are exactly the propositional schemata involving formalized provability that are provable in PA.

The system ILM of interpretability logic is obtained by adding to GL a binary modality $\rhd$ for interpretability. It was proven independently by Berarducci ([1]) and Shavrukov ([24]) that the theorems of ILM are exactly the propositional schemata involving formalized provability and interpretability that are provable in PA.

Plan of this paper. Given the two traditions in the study of interpretability, it is natural to ask how they relate to each other. The starting point for this paper is the following question: does the modal logic ILM "know" that $(\mathscr{V}, \rhd)$ is a lattice?

The fact that the infimum in $(\mathscr{V}, \rhd)$ is given by disjunction is indeed reflected in the axioms of ILM. On the other hand, Švejdar's construction of a supremum employs a language much more complex than that available in propositional modal logic. Indeed, the supremum turns out not to be definable in ILM.

We want to boost the expressive power of ILM by adding to it a *new modality* whose intended interpretation is a supremum operator in $(\mathscr{V}, \rhd)$. As we will see, Švejdar's construction is only one out of many ways of obtaining the supremum of given elements in

---

[1]Someone used to Boolean algebras has to think of the lattice $(\mathscr{V}, \rhd)$ as being *upside down*: the weakest theory $\top$ is the bottom element, and the strongest theory $\bot$ is the top element of $(\mathscr{V}, \rhd)$.

[2]$\mathrm{Con}_x(C)$ denotes the consistency statement for $\mathrm{I}\Sigma_x + C$, where $\mathrm{I}\Sigma_x$ is as usual PA with induction restricted to Σ_x–formulas. In Švejdar's original construction, the sentence $\mathrm{Con}_x(C)$ is taken to be the consistency statement for C together with the first x axioms of PA. For our purposes, it is more convenient to use the stratification sequence based on $\mathrm{I}\Sigma_x$ instead.

this structure. Since each of these can, in principle, be used to specify the meaning of the new modality, we are faced with a design choice.

Section 3 discusses various considerations to be taken into account in the midst of this embarrassment of riches. Švejdar's construction, along with its dual discovered by Visser, are studied in Section 4. Section 5 introduces our favourite way of interpreting the new modality: a combination of conjunction with a certain unary operator that we call a *supremum adapter*. The idea and first examples of supremum adapters are due to Shavrukov. Supremum adapters may be seen as certain nonstandard provability predicates. This perspective is explored in Section 6, where we study the bimodal logic of such a nonstandard notion of provability, together with ordinary provability.

2. PRELIMINARIES

This section introduces the basic notions used in the paper. The reader is advised to go through it lightly in order to return when some fact or definition is used.

2.1. **Arithmetical Theories.** We consider first-order theories in the language $\mathscr{L}$ of arithmetic containing 0, S (successor), $+$, $\times$, and $\leq$. As usual, a formula is said to be Δ_0 (equivalently: Σ_0 or Π_0) if all its quantifiers are bounded. A formula is Σ_{n+1} (Π_{n+1}) if it is of the form $\exists x_0 \ldots x_n A$, with A a Π_n (Σ_n)-formula. For each natural number n, we define a term $\overline{n}$ of $\mathscr{L}$ by letting $\overline{0} = 0$, and $\overline{n+1} = S\overline{n}$. Given that, we will often write n instead of $\overline{n}$.

The basic facts concerning 0, S, $+$, $\times$, and $\leq$ are given by the axioms of the theory Q of Robinson Arithmetic ([7, Definition I.1.1]). The theory Q is Σ_1-complete, meaning that it proves every true Σ_1-sentence. The theory PA of Peano Arithmetic results from adding to Q the induction schema for all $\mathscr{L}$-formulas.

As usual, $I\Sigma_n$ is the fragment of PA obtained by restricting the induction schema to Σ_n-formulas. Note that $I\Sigma_n \subseteq I\Sigma_{n+1}$ for all n. We recall that for $n > 0$, $I\Sigma_n$ is finitely axiomatizable ([7, Theorem I.2.52]). The theory $I\Delta_0 + \exp$ results from adding to $I\Delta_0$ an axiom stating the totality of exponentiation. We note that $I\Delta_0 + \exp$ is finitely axiomatizable ([7, Theorem V.5.6]) and contained in $I\Sigma_1$ ([7, Theorem I.1.50]).

Metamathematics. It is well-known that arithmetization of syntax can be carried out in $I\Delta_0 + \exp$. We assume as given some standard gödelnumbering of $\mathscr{L}$-formulas, and write $\ulcorner A \urcorner$ for the gödelnumber of A. We shall often identify a formula with its gödelnumber, writing for example $B(A)$ instead of $B(\ulcorner A \urcorner)$.

Let S be a recursively enumerable (r.e.) extension of $I\Delta_0 + \exp$. Since any r.e. relation can be represented in a natural way in $I\Delta_0 + \exp$ by a Σ_1-formula ([3]), there is a Σ_1-formula σ representing the axioms of S in $I\Delta_0 + \exp$, i.e. we have

$$A \text{ is an axiom of } S \quad \text{iff} \quad I\Delta_0 + \exp \vdash \sigma(A).$$

Using σ, one can define in a natural way (see [4, Definition 4.1]) a Σ_1-formula $\Pr_\sigma$ representing (inside $I\Delta_0 + \exp$) provability in S. In this paper, we will write $\Pr_S$ instead of $\Pr_\sigma$, having in mind some standard Σ_1-representation σ of the axioms of S in $I\Delta_0 + \exp$. To further simplify notation, we shall write $\Box_S$ for the formula $\Pr_S$. $\Box$ and $\Box_0$ will be used as shorthand for $\Box_{PA}$ and $\Box_{I\Delta_0+\exp}$ respectively. By $\Box_x$ we denote the formula containing x as a free variable, and such that for $n > 0$, $\Box_n$ (the result of substituting $\overline{n}$ for x in $\Box_x$) is $\Box_{I\Sigma_n}$. The symbol $\Diamond_S$ is used for the dual of $\Box_S$, i.e. as an abbreviation for $\neg\Box_S\neg A$.

We use the dot notation as usual, thus $\Box_S A(\dot{x})$ means that the numeral for the value of x has been substituted for the free variable of the formula A inside $\Box_S$. If the intended meaning is clear from the context, we will often simply write $\Box_S A(x)$ instead of $\Box_S A(\dot{x})$.

We recall that any theory S extending $I\Delta_0+\exp$ is provably Σ_1-complete, meaning that for any Σ_1-formula A,

$$I\Delta_0+\exp \vdash A(x) \to \Box_S A(\dot{x}).$$

It is well-known that if S is as above, then the *Hilbert-Bernays-Löb derivability conditions* hold for $\Box_S$ verifiably in $I\Delta_0+\exp$:

(1) if $S \vdash A$, then $I\Delta_0+\exp \vdash \Box_S A$
(2) $I\Delta_0+\exp \vdash \Box_S(A \to B) \to (\Box_S A \to \Box_S B)$
(3) $I\Delta_0+\exp \vdash \Box_S A \to \Box_S \Box_S A$

In fact, 2 and 3 also hold with internal variables ranging over the sentences A and B.

The following theorem is known as the Diagonal Lemma, or Gödel-Carnap Fixed Point Lemma.

Theorem 1. *Let A be a Σ_n (Π_n)-formula whose free variables are exactly $x_0, \ldots, x_n$. There is a Σ_n (Π_n)-formula β with exactly the same free variables, and such that*

$$I\Delta_0+\exp \vdash \beta(x_1, \ldots, x_n) \leftrightarrow A(\ulcorner\beta(v_1, \ldots, v_n)\urcorner, x_1, \ldots, x_n).$$

We say that $\beta(x_1, \ldots, x_n)$ is a fixed point of $A(x_0 \ldots, x_n)$.

Theorem 1, together with the Hilbert-Bernays-Löb derivability conditions for $\Box_S$, implies that Löb's principle for $\Box_S$ is verifiable in $I\Delta_0+\exp$ ([2, Theorem 3.2]):

$$I\Delta_0+\exp \vdash \Box_S(\Box_S A \to A) \to \Box_S A.$$

Principles valid in the Gödel-Löb provability logic GL (see Section 2.3) can thus be used when reasoning about S in $I\Delta_0+\exp$.

Consider the sequence of theories $\{T_n\}_{n\in\omega}$, where $T_0 = I\Delta_0+\exp$, and for $n > 0$, $T_n = I\Sigma_n$. The basic facts concerning $\{T_n\}_{n\in\omega}$ are verifiable in $I\Delta_0+\exp$:

(2) $I\Delta_0+\exp \vdash \Box A \leftrightarrow \exists x \Box_x A$

(3) $I\Delta_0+\exp \vdash \forall x, y \, (\Box_x A \wedge x < y \to \Box_y A)$

Furthermore, $I\Delta_0+\exp$ verifies that each T_{n+1} proves *uniform Π_{n+2}–reflection* for T_n:

(4) $I\Delta_0+\exp \vdash \forall x \, (\Box_{x+1}(\forall A \in \Pi_{x+2} \to \forall y \, (\Box_x A(\dot{y}) \to A(y))))$.

That (4) holds was first stated in [21]; an inspection of the proof shows that it can be verified in $I\Delta_0+\exp$. Throughout this paper, we shall refer to properties (3) and (4) as *monotonicity* and *reflection* respectively.

Oracles. We recall that for $n > 1$, there is a partial truth definition tr_{Π_n} for Π_n-sentences in $I\Delta_0+\exp$ ([7, Section V.5(b)]). The formula tr_{Π_n} is Π_n and satisfies Tarski's conditions ([7, Definition I.1.74].

Let S be an r.e. extension of $I\Delta_0+\exp$. Using tr_{Π_n}, one can define in a natural way the provability predicate $\Box_S^{\Pi_n}$ of S together with a Π_n-oracle. Thus $\Box_x^{\Pi_1}$ is an intensionally correct provability predicate for $I\Sigma_x$ together with a Π_1-oracle. Note that $\Box_x^{\Pi_1}$ is a Σ_2-formula. It follows from (4) that $I\Delta_0+\exp \vdash \forall x \Box_{x+1}\neg\Box_x\bot$. Inspection of the proof shows that it goes through when we add Π_1-sentences that are known to be true to formal proofs, thus we have: $I\Delta_0+\exp \vdash \forall x \Box_{x+1}\neg\Box_x^{\Pi_1}\bot$. Also this property will be referred to as *reflection*.

2.2. **Interpretability.** The notion of interpretability that we are interested in is that of *relative interpretability*, first introduced and carefully studied by Tarski, Mostowski and Robinson ([29]). Since all theories considered here have pairing, it is safe to focus our attention on one-dimensional interpretations.

Definition 2. Let S and T be first-order theories whose languages are $\mathscr{L}_S$ and $\mathscr{L}_T$. An *interpretation* j of S in T is a tuple $\langle \delta, \tau \rangle$, where δ is an $\mathscr{L}_T$-formula with one free variable, and τ a mapping from relation symbols[3] R of $\mathscr{L}_S$ to formulas R^τ of $\mathscr{L}_T$, where the number of free variables of R^τ is equal to the arity of R. We extend τ to a translation from all formulas of $\mathscr{L}_S$ to formulas of $\mathscr{L}_T$ by requiring:

   i. $(R(x_1, \ldots x_n))^\tau = R^\tau(x_1, \ldots x_n)$
   ii. $(A \to B)^\tau = A^\tau \to B^\tau$
   iii. $\perp^\tau = \perp$
   iv. $(\forall x A)^\tau = \forall x (\delta(x) \to A^\tau)$

Finally, we require that $T \vdash \exists x\, \delta(x)$, and $T \vdash A^\tau$ for all axioms A of S.

We write $j : T \triangleright S$ if j is an interpretation of S in T, and $T \triangleright S$ if $j : T \triangleright S$ for some j. We say that T and S are mutually interpretable, and write $T \equiv S$, if $T \triangleright S$ and $S \triangleright T$. In this paper we are concerned with interpretability between finite extensions of PA, i.e. theories of the form $\mathsf{PA} + A$, where A is an $\mathscr{L}$-sentence. We write $A \triangleright B$ as an abbreviation for $\mathsf{PA} + A \triangleright \mathsf{PA} + B$.

Interpretability, like provability, is a syntactical notion, and can therefore be formalized in PA. We write $A \triangleright B$ for the arithmetical sentence expressing that $\mathsf{PA} + A$ interprets $\mathsf{PA} + B$.

The following theorem concerning interpretability over PA is referred to as the Orey-Hájek Characterization.

Theorem 3 ($\mathrm{I}\Delta_0 + \exp$). *The following are equivalent:*

   i. $A \triangleright B$
   ii. *For all n,* $\mathsf{PA} \vdash A \to \Diamond_n B$
   iii. *For any* $C \in \Pi_1$, $\mathsf{PA} \vdash B \to C$ *implies* $\mathsf{PA} \vdash A \to C$

Theorem 3 is implicit in [22], and was first explicitly stated in [6] and in [8]. Item iii was added in [5]. Inspection of the proof shows that it can be verified in $\mathrm{I}\Delta_0 + \exp$.

2.3. **Modal logic.** We denote by $\mathscr{L}_\square$ the language of propositional modal logic containing a unary modality $\square$, and by $\mathscr{L}_{\square\triangleright}$ the language $\mathscr{L}_\square$ together with a binary modality $\triangleright$. Provability logic. The axiom schemata of the Gödel-Löb provability logic GL include all propositional tautologies in the language $\mathscr{L}_\square$, and furthermore:

(K) $\square(A \to B) \to (\square A \to \square B)$

(L) $\square(\square A \to A) \to \square A$

The inference rules of GL are modus ponens and necessitation:

$$\text{if } \mathsf{GL} \vdash A, \text{ then } \mathsf{GL} \vdash \square A.$$

We note that $\mathsf{GL} \vdash \square A \to \square\square A$ (see e.g. [2, p.11]).

The axiom schemata of the interpretability logic ILM include all propositional tautologies in the language $\mathscr{L}_{\square\triangleright}$, the axiom schemata of GL, and:

(J1) $\square(A \to B) \to A \triangleright B$
(J2) $(A \triangleright B) \wedge (B \triangleright C) \to (A \triangleright C)$
(J3) $(A \triangleright C) \wedge (B \triangleright C) \to (A \vee B) \triangleright C$
(J4) $A \triangleright B \to (\Diamond A \to \Diamond B)$
(J5) $\Diamond A \triangleright A$

---

[3]We assume here that S is formulated in a purely relational way. This restriction is not essential – function symbols can be replaced by relation symbols by a well-known algorithm (for details, see [31, Section 7.3]).

(M) $A \rhd B \to (A \wedge \Box C) \rhd (B \wedge \Box C)$

The rules of ILM are modus ponens, and necessitation for $\Box$.

Definition 4. Let φ be an $\mathscr{L}$-formula with one free variable. A φ-*realization* is a function $*$ from the propositional letters of $\mathscr{L}_\Box$ to $\mathscr{L}$-sentences. The domain of $*$ is extended to all $\mathscr{L}_\Box$-formulas by requiring that it commutes with the propositional connectives, and furthermore

$$(\Box A)^* := \varphi(\ulcorner A^* \urcorner).$$

Given an $\mathscr{L}$-formulas φ and ψ with one and two free variables respectively, the notion of a φ-ψ-*realization* (a function from $\mathscr{L}_{\Box\rhd}$-formulas to $\mathscr{L}$-sentences) is defined in a similar way.

Theorem 5. *Let T be a r.e. and Σ_1-sound extension of $I\Delta_0 + \mathsf{exp}$, and let Pr_T be an intensionally correct provability predicate of T. Write $\mathscr{R}$ for the set of all Pr_T-realizations. Then for all $A \in \mathscr{L}_\Box$,*

$$\mathsf{GL} \vdash A \quad \textit{if and only if} \quad \textit{for all } * \in \mathscr{R}, \ I\Delta_0 + \mathsf{exp} \vdash A^*.$$

The direction from left to right follows by the fact that the Hilbert-Bernays-Löb derivability conditions for Pr_T are verifiable in $I\Delta_0 + \mathsf{exp}$ ([11], [18]). The proof of the other direction with $T = \mathsf{PA}$ is due to Solovay ([27]). De Jongh, Jumelet and Montagna ([13]) extended the result to Σ_1-sound theories containing $I\Delta_0 + \mathsf{exp}$.

We recall that a theory is said to be *essentially reflexive* if it proves the consistency of each of its finite subtheories, and the same holds for every consistent extension in the same language. It is well-known that PA is essentially reflexive ([7, Theorem III.2.35]).

Theorem 6. *Let T be r.e., essentially reflexive, and Σ_1–sound. Let Pr_T be an intensionally correct provability predicate of T, and Int an $\mathscr{L}$-formula (with two free variables) representing interpretability in T. Write $\mathscr{R}$ for the set of all Pr_T–Int–realizations. For all $A \in \mathscr{L}_{\Box\rhd}$,*

$$\mathsf{ILM} \vdash A \quad \textit{if and only if} \quad \textit{for all } * \in \mathscr{R}, \ T \vdash A^*.$$

Theorem 6 was proven independently by Shavrukov ([24]) and Berarducci ([1]).

3. Methodological considerations

Write $(\mathscr{V}, \rhd)$ for the degrees of finite extensions of PA, ordered by the relation of interpretability. As mentioned in Section 1, it was proven by Švejdar that the structure $(\mathscr{V}, \rhd)$ is a lattice. Our question is whether this fact is seen from the perspective of the interpretability logic ILM.

Indeed, the fact that $(\mathscr{V}, \rhd)$ is a lower semilattice is reflected in the axioms of ILM. Principles J1 and J2 imply that the ordering given by $\rhd$ is reflexive and transitive, and furthermore that $A \rhd A \vee B$ and $B \rhd A \vee B$, i.e. that $[A \vee B]$ is a lower bound of $[A]$ and $[B]$ in $(\mathscr{V}, \rhd)$. Principle J3 states that it is in fact the *greatest* lower bound, i.e. the infimum of $[A]$ and $[B]$.

To see that, in general, the supremum of $[A]$ and $[B]$ is not $[A \wedge B]$, it suffices to consider the so-called *Orey sentences* discovered by Orey in [22, Theorem 2.4]. An Orey sentence is any sentence O with $\top \rhd O$ and $\top \rhd \neg O$. An Orey sentence and its negation are thus both in $[\top]$, and hence also their supremum is an element of $[\top]$, which is clearly not the case for the sentence $O \wedge \neg O$.

A quick way to convince oneself of the existence of suprema in $(\mathscr{V}, \rhd)$ is via a detour through the structure $(\mathscr{D}, \rhd)$ of interpretability degrees of all r.e. extensions of PA. It

follows from Theorem 3 that the supremum of $[\mathsf{PA}+A]$ and $[\mathsf{PA}+B]$ in $(\mathscr{D},\rhd)$ is $[S]$, where S is the following infinite theory:

$$S := \mathsf{PA} + \{\Diamond_n A \wedge \Diamond_n B \mid n \in \omega\}.$$

By Lindström's result ([15, Theorem 3]), we know that $(\mathscr{D},\rhd)$ and $(\mathscr{V},\rhd)$ are isomorphic; thus the supremum of $[\mathsf{PA}+A]$ and $[\mathsf{PA}+B]$ in $(\mathscr{V},\rhd)$ must also exist.

In Section 1, we saw Švejdar's direct construction of suprema in $(\mathscr{V},\rhd)$ (see (1), or Theorem 10 in Section 4). The latter, however, employs a language much more complicated than that available in the propositional system ILM. Indeed, as shown in appendix A.1, the existence of interpretability suprema lies beyond the expressive power of ILM.

We thus want to enrich the language of ILM by adding to it a new binary modality $\oslash$, together with the following axiom S, stating that $\oslash$ is an interpretability supremum:

(S) $\quad (C \rhd A) \wedge (C \rhd B) \leftrightarrow (C \rhd A \oslash B)$.

We want to interpret $\oslash$ as *an arithmetical* formula. The following definition states what we are looking for.

Definition 7. An $\mathscr{L}$-formula σ with two free variables is a *supremum implementation* if for all $\mathscr{L}$-sentences A,B,C, we have that $\mathsf{PA} \vdash (C \rhd A) \wedge (C \rhd B) \leftrightarrow C \rhd \sigma(A,B)$.

Given a supremum implementation σ, we can extend the notion of an arithmetical realization to include formulas of the extended modal language:

(5) $$(A \oslash B)^* := \sigma(\ulcorner A^* \urcorner, \ulcorner B^* \urcorner).$$

Let $\mathscr{R}$ be the set of all realizations where the ILM part is as in Theorem 6, and $\oslash$ is interpreted as in (5). Denote by ILMS the system ILM together with axiom S. Then it is clear that for all modal formulas A,

$$\text{If ILMS} \vdash A \ , \text{ then for all } * \in \mathscr{R}, \ \mathsf{PA} \vdash A^*.$$

In other words ILMS is arithmetically sound. For arithmetical completeness, we need to make sure that all propositional schemata involving σ that are theorems of PA are also provable in ILMS. However, it turns out that what exactly is provable in PA about a supremum implementation is far from determined by the fact that the latter satisfies Definition 7. For example, while for any such σ it is clear that $\sigma(A,B)$ and $\sigma(B,A)$ are, verifiably in PA, mutually interpretable, there is no a priori reason why

(6) $$\sigma(A,B) \leftrightarrow \sigma(B,A)$$

should be provable in PA, or even true (see Appendix A.2). On the other hand, the supremum implementations we encounter in the following sections all satisfy (6).

Given the above, should we add $A \oslash B \leftrightarrow B \oslash A$ to the axioms of ILMS? The answer depends on which supremum implementation(s) we have in mind. In contrast to formalized provability and interpretability, there is no strong intuition as to what constitutes a *natural* supremum implementation. Thus our choice will depend on practical and esthetical criteria. For example, we prefer implementations that allow for a nice Kripke semantics. The following definition states some nice features that a supremum implementation could have.

Definition 8. Let A,A',B,B' range over all $\mathscr{L}$-sentences, and let σ be an $\mathscr{L}$-formula with two free variables. We say that σ is *extensional* if

$$\mathsf{PA} \vdash A \leftrightarrow A' \text{ and } \mathsf{PA} \vdash B \leftrightarrow B' \quad \Rightarrow \quad \mathsf{PA} \vdash \sigma(A,B) \leftrightarrow \sigma(A',B'),$$

and σ is *provably extensional* if the above is verifiable in PA, i.e.

$$\mathsf{PA} \vdash \Box(A \leftrightarrow A') \wedge \Box(B \leftrightarrow B') \rightarrow \Box(\sigma(A,B) \leftrightarrow \sigma(A',B')).$$

We say that σ is *monotone* if

$$\text{PA} \vdash A \to A' \text{ and } \text{PA} \vdash B \to B' \quad \Rightarrow \quad \text{PA} \vdash \sigma(A,B) \to \sigma(A',B'),$$

and *provably monotone* if the above is verifiable in PA, i.e.

$$\text{PA} \vdash \Box(A \to A') \wedge \Box(B \to B') \to \Box(\sigma(A,B) \to \sigma(A',B')).$$

We note that the above properties are ordered by strength, with extensionality being the weakest.

We conclude this section by considering a supremum implementation property that is so nice that we did not even think of it before Shavrukov gave us an example.

Definition 9. A supremum implementation σ is *adapted* if there is an $\mathscr{L}$-formula τ containing one free variable, and such that for all $\mathscr{L}$-sentences A and B,

$$\sigma(A,B) = \tau(A) \wedge \tau(B).$$

We call a formula τ as in Definition 9 a *supremum adapter*. A supremum adapter is thus a *unary* operator that, in combination with conjunction, provides us with a supremum implementation[4].

4. Varieties of supremum implementations

This section studies the supremum implementation based on Švejdar's argument showing that $(\mathscr{V}, \rhd)$ is a lattice, as well as a dual construction discovered by Visser. These implementations are given as fixed points of certain arithmetical formulas. We prove the existence of unique explicit fixed points for these formulas. The purpose of this section is mainly methodological; it can be safely skipped without affecting insight into the rest of the article.

4.1. **Definition and verification of the target property.** As mentioned above, it follows from Theorem 3 that the *infinite* theory

$$(7) \qquad\qquad\qquad \text{PA} + \{\Diamond_n A \wedge \Diamond_n B \mid n \in \omega\}$$

is an interpretability supremum of $\text{PA} + A$ and $\text{PA} + B$. The idea of Švejdar's construction is to compress the information contained in (7) into a single sentence.

Theorem 10 ([28, Theorem 4.4]). *By Theorem 1, let ϑ be such that*

$$\text{PA} \vdash \vartheta \leftrightarrow \forall x (\Diamond_x \vartheta \to \Diamond_x A \wedge \Diamond_x B).$$

Then ϑ is (verifiably in $\text{I}\Delta_0 + \text{exp}$) a supremum of A and B in $(\mathscr{V}, \rhd)$.

We note that the complexity of ϑ is Π_2. A dual construction, yielding a Σ_2-supremum, was discovered by Visser.

Theorem 11. *By Theorem 1, let ϑ be such that*

$$\text{PA} \vdash \vartheta \leftrightarrow \exists x (\Box_x \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B)).$$

Then ϑ is (verifiably in $\text{I}\Delta_0 + \text{exp}$) a supremum of A and B in $(\mathscr{V}, \rhd)$.

Proof. Argue in $\text{I}\Delta_0 + \text{exp}$. We show that

  i. $\vartheta \rhd A$ and $\vartheta \rhd B$
  ii. Whenever $C \rhd A$ and $C \rhd B$, also $C \rhd \vartheta$

---

[4]We think of τ as *adapting A and B*, so that their conjunction can be used to get the supremum.

We first show i. By Theorem 3, it suffices to show that ϑ implies $\diamondsuit_n A$ and $\diamondsuit_n B$ for any standard n. Assume ϑ. Thus there is some a with $\square_a \neg \vartheta$, $\diamondsuit_a A$, and $\diamondsuit_a B$. By essential reflexivity, a must be greater than any standard n. By monotonicity, we thus have $\diamondsuit_n A$ and $\diamondsuit_n B$ for any such n.

For ii, let C be such that $C \triangleright A$ and $C \triangleright B$. We claim that $C \wedge \neg \vartheta \triangleright \vartheta$. Note that $\neg \vartheta$ is the sentence $\forall x ((\diamondsuit_x A \wedge \diamondsuit_x B) \to \diamondsuit_x \vartheta)$. Since $C \triangleright A$ and $C \triangleright B$, by Theorem 3 we have that C implies $\diamondsuit_n A$ and $\diamondsuit_n B$ for any standard n. But then $C \wedge \neg \vartheta$ implies $\diamondsuit_n \vartheta$ for any such n, whence $C \wedge \neg \vartheta \triangleright \vartheta$ again by Theorem 3. Finally, note that clearly $C \wedge \vartheta \triangleright \vartheta$. By axiom J3 of ILM, we therefore have $(C \wedge \vartheta) \vee (C \wedge \neg \vartheta) \triangleright \vartheta$, i.e. $C \triangleright \vartheta$. □

Strictly speaking, theorems 10 and 11 do not yet provide us with supremum implementations in the sense of Definition 7. Corresponding to Theorem 10, we would want a formula σ with two free variables, such that for any sentences A and B,

$$(8) \qquad \mathsf{PA} \vdash \sigma(A, B) \leftrightarrow \forall x (\diamondsuit_x \sigma(A, B) \to \diamondsuit_x A \wedge \diamondsuit_x B).$$

Such a formula can be found by using Theorem 1 (see Appendix B.1). Having this in mind, we shall from now on say *Švejdar's supremum (implementation)* to refer to a formula σ as in (8); similarly for Visser's supremum.

4.2. **Existence of unique explicit fixed points.** The suprema introduced above are given as fixed points of certain formulas. Before proving the existence of unique and explicit fixed points for these formulas, we examine a surprising feature of Švejdar's construction. Švejdar's supremum of A and B is obtained as a fixed point of the following formula[5]:

$$\forall x (\diamondsuit_x Y \to (\diamondsuit_x A \wedge \diamondsuit_x B)).$$

In particular the supremum of $\bot$ and $\bot$ is given as a fixed point of

$$(9) \qquad \forall x (\diamondsuit_x Y \to \diamondsuit_x \bot).$$

Example 12. Write $\Psi_\bot(Y)$ for the formula in (9). We have the following:

  i. If $\mathsf{PA} \vdash \vartheta \leftrightarrow \Psi_\bot(\vartheta)$, then $\mathsf{PA} \vdash \vartheta \leftrightarrow \bot$
  ii. $\mathsf{PA} \nvdash \bot \leftrightarrow \Psi_\bot(\bot)$

In other words every fixed point of $\Psi_\bot(Y)$ is equivalent to $\bot$, however $\bot$ itself is not a fixed point of $\Psi_\bot(Y)$ To see that i holds, suppose that $\mathsf{PA} \vdash \vartheta \leftrightarrow \Psi_\bot(\vartheta)$, i.e.

$$\mathsf{PA} \vdash \vartheta \leftrightarrow \forall x (\square_x \top \to \square_x \neg \vartheta).$$

Since $\mathsf{PA} \vdash \forall x \square_x \top$, it follows that $\mathsf{PA} \vdash \vartheta \leftrightarrow \square_0 \neg \vartheta$. Since $I\Sigma_1 \vdash \square_0 \neg \vartheta \to \neg \vartheta$ by reflection, we thus have $\mathsf{PA} \vdash \vartheta \leftrightarrow \bot$. For ii note that $\Psi_\bot(\bot)$ is $\forall x (\diamondsuit_x \bot \to \diamondsuit_x \bot)$, i.e. it is a tautology and thus clearly not provably equivalent to $\bot$. □

Example 12 demonstrates that the formula $\Psi_\bot(Y)$ is not extensional, i.e. it is *not* the case that for all sentences A and B,

$$(10) \qquad \mathsf{PA} \vdash A \leftrightarrow B \quad \Rightarrow \quad \mathsf{PA} \vdash \Psi_\bot(A) \leftrightarrow \Psi_\bot(B).$$

Indeed, if ϑ is a fixed point of $\Psi_\bot(Y)$, then as shown above $\mathsf{PA} \vdash \vartheta \leftrightarrow \bot$. On the other hand $\Psi_\bot(\bot)$ is a tautology and therefore not provably equivalent to $\Psi_\bot(\vartheta)$ – since the latter is equivalent to ϑ and hence to $\bot$.

---

[5] The capital variable Y indicates that we are interested in fixed points with respect to this variable.

We thus cannot apply to $\forall x\,(\Diamond_x Y \to (\Diamond_x A \wedge \Diamond_x B))$ Smoryński's method ([26]) for establishing the uniqueness of fixed points of arithmetical formulas[6]. The following theorem, due to Shavrukov, shows that unique explicit fixed points nevertheless exist.

Theorem 13. *Let $\varepsilon(A,B)$ be the formula*

$$(\Box_0 \neg A \vee \Box_0 \neg B \to \Box_0 \bot) \wedge \forall x\,(\Box_{x+1} \neg A \vee \Box_{x+1} \neg B \to \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B)),$$

where A and B are regarded as internal variables ranging over $\mathscr{L}$-sentences. Then

  i. $\mathsf{PA} \vdash \varepsilon(A,B) \leftrightarrow \forall x\,(\Diamond_x \varepsilon(A,B) \to \Diamond_x A \wedge \Diamond_x B)$
  ii. *For any ϑ, if $\mathsf{PA} \vdash \vartheta \leftrightarrow \forall x\,(\Diamond_x \vartheta \to \Diamond_x A \wedge \Diamond_x B)$, then $\mathsf{PA} \vdash \vartheta \leftrightarrow \varepsilon(A,B)$*

Proof. See Appendix B.2. $\qquad\square$

A similar result holds for Visser's supremum implementation.

Theorem 14. *Let $\varepsilon(A,B)$ be the formula*

$$\exists x\,(\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B))$$

where A and B are regarded as internal variables ranging over $\mathscr{L}$-sentences. Then

  i. $\mathsf{PA} \vdash \varepsilon(A,B) \leftrightarrow \exists x\,(\Box_x \neg \varepsilon(A,B) \wedge (\Diamond_x A \wedge \Diamond_x B))$
  ii. *For any ϑ, if $\mathsf{PA} \vdash \vartheta \leftrightarrow \exists x\,(\Box \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B))$, then $\mathsf{PA} \vdash \vartheta \leftrightarrow \varepsilon(A,B)$*

Proof. See Appendix B.2. $\qquad\square$

It follows that any formula ε as in Theorem 13 or in Theorem 14 is a supremum implementation (using items i. and the proofs of theorems 10 and 11). With the explicit versions at hand, we can learn a bit more about Švejdar's and Visser's suprema.

Theorem 15. *Švejdar's and Visser's supremum are both extensional.*

Proof. See Appendix B.3. $\qquad\square$

Theorem 16. *Visser's supremum is not monotone.*

Proof. See Appendix B.4.

$\qquad\square$

Open Question 17. Is Švejdar's supremum monotone? $\qquad\square$

5. SUPREMUM ADAPTERS

The idea and the first examples of supremum adapters are due to Shavrukov. The versions introduced here were obtained by analyzing and simplifying the latter. Define

$$\sigma(A) := \forall x\,(\Diamond_x^{\Pi_1} \top \to \Diamond_{x+1} A)$$
$$\tau(A) := \exists x\,(\Box_x^{\Pi_1} \bot \wedge \Diamond_x A)$$

where A is regarded as an internal variable ranging over sentences. Write $\triangledown A$ for either $\sigma(A)$ or $\tau(A)$. In this section, we show that $\triangledown$ is a supremum adapter, i.e. that for all $\mathscr{L}$-sentences A, B, and C, it is provable in PA that

$$(C \triangleright A) \wedge (C \triangleright B) \;\leftrightarrow\; C \triangleright (\triangledown A \wedge \triangledown B).$$

In fact, we show that the above is already verifiable in $\mathsf{I}\Delta_0 + \mathsf{exp}$.

---

[6]Smoryński's method for showing the uniqueness of fixed points of an arithmetical formula $\Psi(Y)$ relies on $\Psi(Y)$ being PA–*substitutable*, meaning that $\mathsf{PA} \vdash \Box(A \leftrightarrow B) \to (\Psi(A) \leftrightarrow \Psi(B))$ for all A and B. Note that if $\Psi(Y)$ is PA–substitutable, then it is also extensional in the sense of Definition 8.

Lemma 18 ($I\Delta_0 + \exp$). *For all n, $\mathsf{PA} \vdash \triangledown A \to \lozenge_n A$.*

Proof. Straightforward from the definition of σ, τ, and the fact that $I\Sigma_{n+1} \vdash \lozenge_n^{\Pi_1} \top$ for all n by reflection. $\square$

Lemma 19. $I\Delta_0 + \exp \vdash \forall z \left(\Box_z(\neg\triangledown A \vee \neg\triangledown B) \to \Box_z(\Box_z \neg A \vee \Box_z \neg B) \right).$

Proof. Argue $I\Delta_0 + \exp$, assuming

$$\text{(11)} \qquad \Box_z(\neg\triangledown A \vee \neg\triangledown B).$$

We first show that (11) implies

$$\text{(12)} \qquad \Box_z(\Box_z^{\Pi_1} \bot \to (\Box_z \neg A \vee \Box_z \neg B)).$$

Suppose first that (11) is $\Box_z(\neg\sigma(A) \vee \neg\sigma(B))$, i.e.

$$\text{(13)} \qquad \Box_z\left(\exists x\,(\lozenge_x^{\Pi_1} \top \wedge \Box_{x+1} \neg A) \vee \exists x\,(\lozenge_x^{\Pi_1} \top \wedge \Box_{x+1} \neg B)\right).$$

It follows that

$$\text{(14)} \qquad \Box_z \exists x\,(\lozenge_{x-1}^{\Pi_1} \top \wedge (\Box_x \neg A \vee \Box_x \neg B)).$$

We reason in $\Box_z$. By (14), let y be such that $\neg\Box_y^{\Pi_1} \bot \wedge (\Box_{y+1} \neg A \vee \Box_{y+1} \neg B)$. Assuming $\Box_z^{\Pi_1} \bot$, we have $y < z$, whence by monotonicity $\Box_z \neg A \vee \Box_z \neg B$. We exit $\Box_z$. Note that we have shown (12). The other possibility is that (11) is $\Box_z(\neg\tau(A) \vee \neg\tau(B))$, i.e.

$$\text{(15)} \qquad \Box_z\left(\forall x\,(\Box_x^{\Pi_1} \bot \to \Box_x \neg A) \vee \forall x\,(\Box_x^{\Pi_1} \bot \to \Box_x \neg B)\right).$$

From this it follows that $\Box_z \forall x\,(\Box_x^{\Pi_1} \bot \to (\Box_x \neg A \vee \Box_x \neg B))$, whence clearly also (12). Thus we have (12) whenever (11) holds. We now show that (12) implies

$$\text{(16)} \qquad \Box_z(\Box_z^{\Pi_1}(\Box_z \neg A \vee \Box_z \neg B) \to \Box_z \neg A \vee \Box_z \neg B).$$

To see that (16) holds, reason inside $\Box_z$, assuming

$$\text{(17)} \qquad \Box_z^{\Pi_1}(\Box_z \neg A \vee \Box_z \neg B).$$

If $\Box_z \neg A \vee \Box_z \neg B$, we are done. So assume that not: $\Box_z \neg A \vee \Box_z \neg B$, i.e.

$$\text{(18)} \qquad \lozenge_z A \wedge \lozenge_z B.$$

Since (18) is a Π_1-sentence, we have $\Box_z^{\Pi_1}(\lozenge_z A \wedge \lozenge_z B)$ and thus $\Box_z^{\Pi_1} \bot$ in combination with (17). The desired conclusion $\Box_z \neg A \vee \Box_z \neg B$ now follows by (12). Exiting the world inside $\Box_z$, we have established (16). From the latter it clearly follows that

$$\text{(19)} \qquad \Box_z(\Box_z(\Box_z \neg A \vee \Box_z \neg B) \to \Box_z \neg A \vee \Box_z \neg B),$$

and thus $\Box_z(\Box_z \neg A \vee \Box_z \neg B)$ by Löb's principle for $I\Sigma_z$. $\square$

Lemma 20 ($I\Delta_0 + \exp$). *For all n, $\mathsf{PA} \vdash \lozenge_n A \wedge \lozenge_n B \to \lozenge_n(\triangledown A \wedge \triangledown B)$.*

Proof. Argue in $I\Delta_0 + \exp$. By contraposition, it suffices to show that for all n,

$$I\Sigma_{n+1} \vdash \Box_n(\neg\triangledown A \vee \neg\triangledown B) \to \Box_n \neg A \vee \Box_n \neg B.$$

Fix an n. From Lemma 19 it follows that

$$I\Sigma_{n+1} \vdash \Box_n(\neg\triangledown A \vee \neg\triangledown B) \to \Box_n(\Box_n \neg A \vee \Box_n \neg B).$$

Since $I\Sigma_{n+1} \vdash \Box_n(\Box_n \neg A \vee \Box_n \neg B) \to (\Box_n \neg A \vee \Box_n \neg B)$ by reflection, we are done. $\square$

Theorem 21 ($I\Delta_0 + \exp$). *For all $\mathscr{L}$-sentences A, B, and C,*

$$(C \triangleright A) \wedge (C \triangleright B) \;\leftrightarrow\; C \triangleright (\triangledown A \wedge \triangledown B).$$

Proof. Straightforward consequence of lemmas 18, 20, and Theorem 3. $\square$

6. Nonstandard provability predicates

The supremum adapters introduced in the previous section can be seen as consistency statements corresponding to certain (nonstandard) provability predicates. We shall now investigate the consequences of this perspective, showing in particular that these provability predicates satisfy the principles of the provability logic GL. In Section 6.3, we establish some principles for the bimodal provability logic of such a nonstandard notion of provability, together with ordinary provability.

We use modal notation, writing $\blacktriangledown A$ and $\triangledown A$ for the sentences $\sigma(A)$ and $\tau(A)$ as in the previous section. The provability predicates $\blacktriangle$ and $\triangle$ are defined as their duals, i.e. $\blacktriangle A := \neg \blacktriangledown \neg A$ and $\triangle A := \neg \triangledown \neg A$. Spelling this out:

$$\blacktriangle A = \exists x \,(\Box_{x+1} A \wedge \Diamond_x^{\Pi_1} \top)$$

$$\triangle A = \forall x \,(\Box_x^{\Pi_1} \bot \to \Box_x A)$$

Thus an $\mathscr{L}$-formula A is $\blacktriangle$–provable just in case there is some n such that A is provable (in the usual sense) in $I\Sigma_{n+1}$, and the theory $I\Sigma_n$ together with a Π_1-oracle is consistent. Note that since for all n it is true (in the standard model) that $I\Sigma_n$ together with a Π_1-oracle is consistent, the extension of $\blacktriangle$ in the standard model coincides with ordinary PA-provability defined by $\Box$. However this might not be the case in a model where PA together with a Π_1-oracle is inconsistent. This is why we say that $\blacktriangle$ is a *nonstandard* provability predicate for PA. Note also that $\blacktriangle$ is a Σ_3-formula.

As for the other triangle, we note that a formula A is $\triangle$-provable if for all n, a proof of inconsistency of $I\Sigma_n$ together with a Π_1-oracle implies the existence of a usual $I\Sigma_n$-proof of A. We note that $\triangle$ is a Π_2-formula.

6.1. **Relating the two triangles.** Arguing in PA, let $\mu + 1$ be the smallest number such that $I\Sigma_{\mu+1}$ together with a Π_1-oracle is inconsistent. It is easy to see that if $\mu + 1$ exists, then

$$\blacktriangle A \leftrightarrow \triangle A \leftrightarrow \Box_{\mu+1} A$$

If $\mu + 1$ does not exist, i.e. if PA together with a Π_1-oracle is consistent, then $\blacktriangle A$ is equivalent to $\Box A$, while $\triangle A$ is equivalent to $\top$. Both $\blacktriangle$ and $\triangle$ are therefore equivalent to $\Box_{\mu+1}$, the only difference being in how the latter is interpreted in case $\mu + 1$ does not exist. This leads to the following alternative definition of $\blacktriangle$ and $\triangle$:

$$(20) \qquad\qquad \blacktriangle A := \begin{cases} \Box_{\mu+1} A & \text{if } \Box^{\Pi_1} \bot \\ \Box A & \text{otherwise} \end{cases}$$

$$(21) \qquad\qquad \triangle A := \begin{cases} \Box_{\mu+1} A & \text{if } \Box^{\Pi_1} \bot \\ \top & \text{otherwise} \end{cases}$$

Using the above observations, we see that the two triangles are closely related.

Lemma 22.   *i.* $\mathsf{PA} \vdash \blacktriangle A \leftrightarrow (\Diamond^{\Pi_1} \top \wedge \Box A) \vee (\Box^{\Pi_1} \bot \wedge \triangle A)$
ii. $\mathsf{PA} \vdash \triangle A \leftrightarrow (\blacktriangle A \vee \Diamond^{\Pi_1} \top)$

Proof. For i, reason in PA. First suppose $\blacktriangle A$. From the definition of $\blacktriangle$, it follows that also $\Box A$. If $\Diamond^{\Pi_1} \top$, we are done. If $\Box^{\Pi_1} \bot$, then $\blacktriangle A$ and $\triangle A$ are equivalent, whence $\triangle A$. For the other direction, suppose that $\Diamond^{\Pi_1} \top$ and $\Box A$. Then $\blacktriangle A$ follows straight from the definition. And if $\Box^{\Pi} \bot$, then $\triangle A$ if and only of $\blacktriangle A$. The proof of ii is similar. $\square$

Given existential sentences $A = \exists x A'$ and $B = \exists y B'$, we write $A \leq B$ for the sentence $\exists x (A' \wedge \forall y < x \neg B')$. Thus $A \leq B$ says that the formula A has a witness, and this witness is smaller than or equal to the smallest witness of B. Similarly, we write $A < B$ for the sentence $\exists x (A' \wedge \forall y \leq x \neg B')$. If C is of the form $A \leq B$ we write $C^\perp$ for $B < A$, and if C is of the form $A < B$, we write $C^\perp$ for $B \leq A$. Note that $C^\perp$ implies $\neg C$, but not the other way around. Thus $C^\perp$ can be seen as a strong negation of C. We write $\square_* A$ for $\exists x \square_x A$, and similarly $\square_*^{\Pi_1} A$ for $\exists x \square_x^{\Pi_1} A$.

It is easy to see that

$$\mathsf{PA} \vdash \blacktriangle A \leftrightarrow \square_* A \leq \square_*^{\Pi_1} \bot.$$

Now, $\blacktriangle^\perp A$ is the sentence $\square_*^{\Pi_1} \bot < \square_* A$. Rewriting the definition of $\triangle$ according to the witness comparison notation, we see that $\triangle$ is equivalent the negation of the latter:

$$\mathsf{PA} \vdash \triangle A \leftrightarrow \neg(\square_*^{\Pi_1} \bot < \square_* A).$$

Thus $\mathsf{PA} \vdash \triangle A \leftrightarrow \neg \blacktriangle^\perp A$, whence $\triangle$ is, in a sense, the double negation of $\blacktriangle$.

6.2. **Provability logic of $\triangle$.** We use the symbol $\triangle$ to refer to either $\blacktriangle$ or $\triangle$. We show that $\mathrm{I}\Delta_0 + \exp$ verifies the principles of GL for $\triangle$.

Theorem 23.   *(1) If $\mathsf{PA} \vdash A$, then $\mathsf{PA} \vdash \triangle A$*
    (2) $\mathrm{I}\Delta_0 + \exp \vdash \triangle(A \to B) \to (\triangle A \to \triangle B)$
    (3) $\mathrm{I}\Delta_0 + \exp \vdash \triangle A \to \triangle\triangle A$

Proof. Item 1 follows by reflection, and item 2 by principle K for $\mathrm{I}\Sigma_x$. For 3, we first consider $\blacktriangle A$. Ague in $\mathrm{I}\Delta_0 + \exp$, assuming $\exists x (\square_{x+1} A \wedge \neg \square_x^{\Pi_1} \bot)$. We want to show

$$(22) \qquad \exists x (\square_{x+1} \exists y (\square_{y+1} A \wedge \neg \square_y^{\Pi_1} \bot) \wedge \neg \square_x^{\Pi_1} \bot).$$

Let x be such that $\square_{x+1} A$ and $\neg \square_x^{\Pi_1} \bot$. We have $\square_{x+1} \square_{x+1} A$ by GL for $\mathrm{I}\Sigma_x$, and $\square_{x+1} \neg \square_x^{\Pi_1} \bot$ by reflection. Therefore $\square_{x+1}(\square_{x+1} A \wedge \neg \square_x^{\Pi_1} \bot)$, and thus clearly also (22). It remains to consider $\triangle A$. Assuming $\forall x (\square_x^{\Pi_1} \bot \to \square_x A)$, we want to show:

$$\forall x (\square_x^{\Pi_1} \bot \to \square_x \forall y (\square_y^{\Pi_1} \bot \to \square_y A)).$$

Suppose $\square_x^{\Pi_1} \bot$. We get $\square_x A$ by assumption, and thus also $\square_x \square_x A$ and $\square_x \forall y \geq x \square_y A$. On the other hand, we have by reflection $\square_x \forall y < x \neg \square_y^{\Pi_1} \bot$. Combining the above yields $\square_x \forall y (\square_y^{\Pi_1} \bot \to \square_y A)$ as required. $\square$

Theorem 24. By examining the proof, it is clear that Theorem 23 also holds with internal quantifiers in items 2 and 3. $\square$

As explained in Section 2.1, it follows from theorems 23 and 1 that

$$\mathsf{PA} \vdash \triangle(\triangle A \to A) \to \triangle A,$$

and thus that GL is arithmetically sound with respect to $\triangle$. We note that since $\triangle$ is not a Σ_1-formula, arithmetical completeness does not follow by the usual method. However, as has been shown by Shavrukov, GL is nevertheless arithmetically complete with respect to $\blacktriangle$ (see the forthcoming preprint [10]).

Open Question 25. Is GL arithmetically complete with respect to $\triangle$? $\square$

6.3. **Joint provability logic of $\triangle$ and $\square$.** We establish some principles for the joint provability logic of a supremum adapter provability, together with ordinary provability. As before, we use the symbol $\triangle$ to refer to either $\blacktriangle$ or $\vartriangle$.

Lemma 26. $I\Delta_0+\exp \vdash \square A \rightarrow \triangle\square A$.

Proof. It is easy to see that $S \rightarrow \triangle S$ holds (verifiably in $I\Delta_0+\exp$) for any Σ_1-formula S. The desired result follows by noting that $\square A$ is a Σ_1-formula. $\square$

Lemma 27. $I\Delta_0+\exp \vdash \square(\square B \rightarrow A) \leftrightarrow \square(\square B \rightarrow \triangle A)$

Proof. Argue in $I\Delta_0+\exp$. By Theorem 21, $\neg A \equiv \triangledown\neg A$, whence by Theorem 3, $\neg A$ is (provably in PA) Π_1-conservative over $\triangledown\neg A$ and vice versa. Since $\diamondsuit B$ is Π_1, this means that $\square(\neg A \rightarrow \diamondsuit\neg B) \leftrightarrow \square(\triangledown\neg A \rightarrow \diamondsuit\neg B)$. The desired result follows by contraposition. $\square$

Taking $\top$ for B in Lemma 27, it follows that $I\Delta_0+\exp \vdash \square A \leftrightarrow \square\triangle A$.
Monotonicity. Returning to the business of supremum implementations, we use the above lemmas to show that supremum implementations based on our supremum adapters are monotone.

Lemma 28. $\vdash \square(A \rightarrow B) \rightarrow \square(\triangledown A \rightarrow \triangledown B)$.

Proof. Argue in PA. If $\square(A \rightarrow B)$, then $\square(\neg B \rightarrow \neg A)$, and thus $\square\triangle(\neg B \rightarrow \neg A)$ by the remark under Lemma 27. By the principles of GL for $\triangle$ and for $\square$, this implies $\square(\triangle\neg B \rightarrow \triangle\neg A)$, which in turn implies $\square(\triangledown A \rightarrow \triangledown B)$. $\square$

6.4. **Joint provability logic of $\blacktriangle$ and $\square$.** We consider the joint provability logic of $\blacktriangle$ and $\square$. There are several reasons that make $\blacktriangle$ a preferred object of study compared to $\vartriangle$. As explained above, in contrast to $\vartriangle$ the formula $\blacktriangle$ is an *extensionally correct* provability predicate for PA (albeit a nonstandard one). Second, we know that GL is arithmetically complete with respect to $\blacktriangle$, while we do not yet know whether the same holds for $\vartriangle$. Finally, as is easily seen from the definition, we have that $\blacktriangle A \rightarrow \square A$. On the other hand, there seems to be no modal principle relating $\vartriangle$ and $\square$ in such a simple way.

Theorem 29. We note the similarlity of $\blacktriangle$ to the Feferman provability predicate:

$$\square_f A := \exists x (\square_x A \wedge \diamondsuit_x \top)$$

As shown in [4, Theorem 5.9], $\square_f$ is an extensionally correct provability predicate for PA, however at the same time $PA \vdash \diamondsuit_f \top$. The existence of $\square_f$ illustrates the need for a more careful formulation of Gödel's Second Incompleteness Theorem than what is commonly stated. The Feferman provability predicate $\square_f$ has been further studied in [19], [30], and [25].

Recall that $\blacktriangle A = \exists x (\square_{x+1} A \wedge \diamondsuit_x^{\Pi_1} \top)$. Thus $\blacktriangle$ differs from $\square_f$ by a Π_1-oracle, and by a "+1". Removing the Π_1-oracle, we obtain the formula $\exists x (\square_{x+1} A \wedge \diamondsuit_x \top)$ that is easily seen to be equivalent to $\square A$. Remove the "+1" we obtain:

$$(23) \qquad\qquad\qquad \vartriangle_f A := \exists x (\square_x A \wedge \diamondsuit_x^{\Pi_1} \top).$$

Inspecting the proof of Theorem 21, it is clear that everything works when taking the dual $\triangledown_f$ of $\vartriangle_f$ for $\triangledown$. Thus $\triangledown_f$ is also a supremum adapter.

On the other hand, the modal principles valid for $\vartriangle_f$ are rather different from those valid for $\blacktriangle$. It is easy to see that $PA \vdash \triangledown_f \top$, from which it follows that $\vartriangle_f$ cannot satisfy the principles of GL. Indeed, examining the proof of Theorem 23, the "+1" in the definition of $\blacktriangle$ seems to be essential for establishing transitivity ($\triangle A \rightarrow \triangle\triangle A$).

There is no a priori reason for preferring a supremum implementation based on ▲ to one based on $\triangle_f$. The advantage of the GL-satisfying version is that it satisfies a very well studied modal logic. As we will now see, also the joint provability logic of ▲ and □ has a relative in the existing literature. □

Denote by $\mathscr{L}_{\square\triangle}$ the language of propositional modal logic containing two unary modalities □ and △.

Definition 30. The axiom schemata of GLS include all propositional tautologies in the language $\mathscr{L}_{\square\triangle}$, the axiom schemata of GL for both □ and △, and furthermore

(T1) $\triangle A \rightarrow \square A$
(T2) $\square A \rightarrow \triangle \square A$
(T3) $\square A \rightarrow \square \triangle A$
(T4) $\square(\square B \rightarrow \triangle A) \rightarrow \square(\square B \rightarrow A)$

We note that (T3) and (T2) imply $\square(\square B \rightarrow A) \rightarrow \square(\square B \rightarrow \triangle A)$. In contrast, (T4) does not follow from the weaker $\square \triangle A \rightarrow \square A$. We denote by GLT the system GLS, with axiom (T4) replaced by $\square\triangle A \rightarrow \square A$.

Open Question 31. Is GLS arithmetically complete with respect to □ and ▲? □

The system GLT was studied by Lindström ([16]) due to its relation to Parikh's rule. The latter allows us to infer A, given that we have established the provability of A. It is thus a sort of reflection rule. Since Parikh's rule is admissible in PA, adding it to PA does not yield new theorems. As shown in [23], it does yield speed-up, meaning that some theorems have much shorter proofs when Parikh's rule is allowed. The equivalence of Parikh provability and ordinary provability is however not verifiable in PA. Writing $\square_p$ for the formula representing Parikh provability, we have

$$\mathsf{PA} \vdash \square_p A \leftrightarrow \exists x \square^x A,$$

where $\square^x$ denotes the x-times iterated version of □. Using the above, it is not difficult to see that the principles of GLT are valid when interpreting the modal symbol △ as ordinary provability, and □ as Parikh provability. Indeed, it was proven by Lindström in [16] that GLT is arithmetically complete with respect to this interpretation. In [9], it is shown that GLT is also the arithmetically complete joint provability logic of slow and ordinary provability.

We note that both of the above interpretations of GLT are with respect to a pair of provability predicates where one is the ordinary provability predicate, and the other a *nonstandard Σ_1 provability predicate* for PA. Our conjecture for the system GLS on the other hand involves the ordinary provability predicate, together with a *nonstandard Σ_3-predicate*. Dealing with the complexity of ▲ while trying to answer Question 31 remains a challenge for future work.

REFERENCES

[1] Berarducci, A.: The interpretability logic of Peano arithmetic. The Journal of Symbolic Logic **55**, 1059–1089 (1990)

[2] Boolos, G.: The logic of provability. Cambridge University Press, Cambridge (1993)

[3] Ehrenfeucht, A., Feferman, S.: Representability op recursively enumerable sets in formal theories. Archiv für mathematische Logik und Grundlagenforschung **5**(1), 37–41 (1961). DOI 10.1007/BF01977641. URL http://dx.doi.org/10.1007/BF01977641

[4] Feferman, S.: Arithmetization of metamathematics in a general setting. Fundamenta Mathematicae **49**, 35–92 (1960)

[5] Guaspari, D.: Partially conservative extensions of arithmetic. Transactions of the American Mathematical Society **254**, 47–68 (1979)

[6] Hájek, P.: On interpretability in set theories I. Comm. Math. Univ. Carolinae **12**, 73–79 (1971)

[7] Hájek, P., Pudlák, P.: Metamathematics of First-Order Arithmetic. Springer-Verlag (1998)

[8] Hájková, M., Hájek, P.: On interpretability in theories containing arithmetic. Fundamenta Mathematicae **76**(2), 131–137 (1972). URL http://eudml.org/doc/214455

[9] Henk, P., Pakhomov, F.: Slow and ordinary provability for Peano Arithmetic. ArXiv e-prints (2016)

[10] Henk, P., Shavrukov, V.: A Solovay function for the least 1-inconsistent subtheory of PA. ILLC technical notes series, ILLC, University of Amsterdam (2016)

[11] Hilbert, D., Bernays, P.: Grundlagen der Mathematik II. Springer, Berlin (1939). Second edition: 1970

[12] Jeroslow, R.G.: Consistency statements in formal theories. Fundamenta Mathematicae **72**, 17–40 (1971)

[13] de Jongh, D., Jumelet, M., Montagna, F.: On the proof of Solovay's theorem. Studia Logica **50**, 51–70 (1991)

[14] de Jongh, D., Veltman, F.: Provability logics for relative interpretability. In: Mathematical logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria, pp. 31–42. Plenum Press, Boston (1990)

[15] Lindström, P.: Some results on interpretability. In: F. Jensen, M. B.H., M. K.K. (eds.) Proceedings of the 5th Scandinavian Logic Symposium 1979, pp. 329–361. Aalborg University Press, Aalborg (1979)

[16] Lindström, P.: 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, Uppsala Philosophical Studies 53, pp. 279–288 (2006)

[17] Lindstrm, P.: On certain lattices of degrees of interpretability. Notre Dame J. Formal Logic **25**(2), 127–140 (1984). DOI 10.1305/ndjfl/1093870573. URL http://dx.doi.org/10.1305/ndjfl/1093870573

[18] Löb, M.: Solution of a Problem of Leon Henkin. Journal of Symbolic Logic **20**, 115–118 (1955)

[19] Montagna, F.: On the algebraization of a Feferman's predicate. Studia Logica **37**, 221–236 (1978)

[20] Montague, R.: Theories incomparable with respect to relative interpretability. J. Symbolic Logic **27**(2), 195–211 (1962). URL http://projecteuclid.org/euclid.jsl/1183734432

[21] Ono, H.: Reflection principles in fragments of Peano arithmetic. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik **33**, 317–333 (1987)

[22] Orey, S.: Relative interpretations. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik **7**, 146–153 (1961)

[23] Parikh, R.: Existence and feasibility in arithmetic. The Journal of Symbolic Logic **36**(3), 494–508 (1971). URL http://www.jstor.org/stable/2269958

[24] Shavrukov, V.: The logic of relative interpretability over Peano arithmetic (in Russian). Tech. Rep. No.5, Steklov Mathematical Institute, Moscow (1988)

[25] Shavrukov, V.: A smart child of Peano's. Notre Dame Journal of Formal Logic **35**, 161–185 (1994)

[26] Smoryński, C.: Arithmetic analogues of McAloon's unique Rosser sentences. Archive for Mathematical logic **28**, 1–21 (1989)

[27] Solovay, R.: Provability interpretations of modal logic. Israel Journal of Mathematics **25**, 287–304 (1976)

[28] Švejdar, V.: Degrees of interpretability. Commentationes Mathematicae Universitatis Carolinae **19**, 789–813 (1978)

[29] Tarski, A., Mostowski, A., Robinson, R.: Undecidable theories. North–Holland, Amsterdam (1953)

[30] Visser, A.: Peano's smart children: A provability logical study of systems with built-in consistency. Notre Dame Journal of Formal Logic **30**, 161–196 (1989)

[31] Visser, A.: An inside view of EXP. Journal of Symbolic Logic **57**, 131–165 (1992)

[32] Visser, A.: An Overview of Interpretability Logic. In: M. Kracht, M. de Rijke, H. Wansing, M. Zakharyaschev (eds.) Advances in Modal Logic, vol 1, CSLI Lecture Notes, no. 87, pp. 307–359. Center for the Study of Language and Information, Stanford (1998)

APPENDIX A. METHODOLOGICAL CONSIDERATIONS

A.1. **Undefinability of suprema in** ILM**.** We show here that the existence of suprema in $(\mathscr{V}, \rhd)$ is not expressible in the modal system ILM.

Lemma 32. *Let A be an $\mathscr{L}_{\Box\rhd}$-formula. Then there exists a formula A^p containing no occurrences of $\Box$ and $\rhd$, and such that*

$$\mathsf{ILM} \vdash \Box\bot \to (A \leftrightarrow A^p)$$

Proof. By induction on the complexity of A. The base case and the propositional cases are trivial. If $A = \Box B$ or $A = B \rhd C$ for some B, C, then we can take $\top$ for A^p (this is easily seen by using soundness and completeness of ILM with respect to Veltman frames, see [14]). $\qquad\square$

Theorem 33. *There is no $\mathscr{L}_{\Box\rhd}$-formula S, such that for all $\mathscr{L}_{\Box\rhd}$-formulas A, B, and C,*

$$\text{ILM} \vdash (C \rhd A) \wedge (C \rhd B) \ \leftrightarrow \ C \rhd S(A,B).$$

Proof. Suppose that such a formula S exists. Then in particular

(24) $$\text{ILM} \vdash (\top \rhd A) \wedge (\top \rhd B) \ \leftrightarrow \ \top \rhd S(A,B).$$

Let S^p be purely propositional as in Lemma 32. Using completeness of ILM with respect to Veltman frames, it is easy to see that

(25) $$\text{ILM} \vdash \Box\Box\bot \to (\top \rhd S(A,B) \leftrightarrow \top \rhd S^p(A,B)).$$

By (25), and arithmetical soundness of ILM, we have

$$\text{PA} \vdash \Box\Box\bot \to (\top \rhd S^*(A^*,B^*) \leftrightarrow \top \rhd (S^p)^*(A^*,B^*)),$$

where $*$ is any arithmetical realisation for ILM. Combining with (24), we get for any arithmetical realization $*$,

(26) $$\text{PA} \vdash \Box\Box\bot \to (\top \rhd (S^p)^*(A^*,B^*) \ \leftrightarrow \ (\top \rhd A^*) \wedge (\top \rhd B^*)).$$

Since S^p is purely propositional, there are 16 possibilities. We will show that in each case, there are $\mathscr{L}$-formulas A and B such that

$$\text{PA} \vdash (\top \rhd S^p(A,B) \ \leftrightarrow \ (\top \rhd A) \wedge (\top \rhd B)) \ \to \Box\bot.$$

Combining the above with (26), we get $\text{PA} \vdash \Box\Box\bot \to \Box\bot$ in all cases, a contradiction.

Argue in PA, assuming $\top \rhd S^p(A,B) \ \leftrightarrow \ (\top \rhd A) \wedge (\top \rhd B)$. We use reasoning warranted by ILM to show $\top \rhd \bot$, which is equivalent to $\Box\bot$.

1. $S^p(A,B) = \top$. We have $\top \rhd \top$, and therefore by assumption $(\to)$ $\top \rhd \bot$ (taking $\bot$ for A).
2. $S^p(A,B) = \bot$. Take $\top$ for A and B. We have $\top \rhd \top$, and therefore by assumption $(\leftarrow)$ $\top \rhd \bot$.
3. $S^p(A,B) = A$. Take $\top$ for A and $\bot$ for B. We have $\top \rhd A$, and thus by assumption $\top \rhd B$, i.e. $\top \rhd \bot$.
4. $S^p(A,B) = B$. Like the previous case, but take $\top$ for B and $\bot$ for A.
5. $S^p(A,B) = \neg A$. Take $\bot$ for A. Then $\top \rhd \neg A$, whence by assumption $\top \rhd A$, i.e. $\top \rhd \bot$.
6. $S^p(A,B) = \neg B$. Like the previous case, but take $\bot$ for B.
7. $S^p(A,B) = A \wedge B$. Take as A an Orey sentence O, and for B the sentence $\neg O$. Then $\top \rhd O$ and $\top \rhd \neg O$, by the fact that O is an Orey sentence, so by assumption $\top \rhd O \wedge \neg O$, i.e. $\top \rhd \bot$.
8. $S^p(A,B) = A \vee B$. Take $\bot$ for A and $\top$ for B. Then $\top \rhd A \vee B$, so by assumption $\top \rhd A$, i.e. $\top \rhd \bot$.
9. $S^p(A,B) = \neg A \wedge \neg B$. Take $\bot$ for both A and B. Then $\top \rhd \neg A \wedge \neg B$, so by assumption $\top \rhd A$, i.e. $\top \rhd \bot$.
10. $S^p(A,B) = \neg A \vee \neg B$. Take $\bot$ for A. Then $\top \rhd \neg A \vee \neg B$, so by assumption $\top \rhd A$, i.e. $\top \rhd \bot$.
11. $S^p(A,B) = A \wedge \neg B$. Take $\top$ for A, and $\bot$ for B. Then $\top \rhd A \wedge \neg B$, so by assumption $\top \rhd B$, i.e. $\top \rhd \bot$.
12. $S^p(A,B) = \neg A \wedge B$. Like the previous case, but take $\bot$ for A and $\top$ for B.

13. $S^p(A,B) = A \to B$. Take $\top$ for B and $\bot$ for A. Then $\top \rhd A \to B$, so by assumption $\top \rhd A$, i.e. $\top \rhd \bot$.
14. $S^p(A,B) = B \to A$. Like the previous case, but take $\bot$ for B and $\top$ for A.
15. $S^p(A,B) = A \leftrightarrow B$. Take $\bot$ for both A and B. Then $\top \rhd A \leftrightarrow B$, so also $\top \rhd A$, i.e. $\top \rhd \bot$ by assumption.
16. $S^p(A,B) = \neg(A \leftrightarrow B)$. Take $\top$ for A and $\bot$ for B. Then $\top \rhd \neg(A \leftrightarrow B)$, so by assumption also $\top \rhd B$, i.e. $\top \rhd \bot$.

$\square$

Theorem 34. The above proof proceeds by showing that if suprema were definable in ILM, then $\Box\Box\bot \to \Box\bot$ would be provable in PA. Note that since $\Box\Box\bot \to \Box\bot$ is *true*, the same argument cannot be used to show that the existence of the supremum is not expressible in ILM$^\omega$, the modal logic of interpretability statements that are true (in the standard model).

$\square$

Open Question 35. Are interpretability suprema definable in ILM$^\omega$? $\square$

A.2. **A Non-commutative supremum.** We construct a supremum implementation τ, such that $\tau(A,B) \leftrightarrow \tau(B,A)$ is not always provable in PA. Let σ be any supremum implementation. Let $\tau(A,B)$ be the formula

$$(A \leq B \to \sigma(A,B) \wedge \Box\neg\sigma(A,B)) \wedge (B < A \to \sigma(A,B) \vee \Diamond\sigma(A,B)),$$

where A and B are regarded as internal variables ranging over sentences. By $A \leq B$ we mean that the gödelnumber of A is smaller than or equal to the gödelnumber of B.

We show first that τ is a supremum implementation. Argue in PA. We first show $\tau(A,B) \rhd A$. In case $A \leq B$, we have $\Box(\tau(A,B) \to \sigma(A,B))$, and thus also $\tau(A,B) \rhd \sigma(A,B)$. Since $\sigma(A,B) \rhd A$, also $\tau(A,B) \rhd A$ by transitivity of $\rhd$. In case $B < A$, we have $\Box(\tau(A,B) \to \sigma(A,B) \vee \Diamond\sigma(A,B))$, and so also $\tau(A,B) \rhd \sigma(A,B) \vee \Diamond\sigma(A,B)$. Since by axiom J5 of ILM, $\Diamond\sigma(A,B) \rhd \sigma(A,B)$ we get

$$\sigma(A,B) \vee \Diamond\sigma(A,B) \rhd \sigma(A,B)$$

by axiom J3 of ILM, and so $\sigma(A,B) \vee \Diamond\sigma(A,B) \rhd A$ by the fact that $\sigma(A,B) \rhd A$, and transitivity of $\rhd$. Thus $\tau(A,B) \rhd A$ follows again by transitivity of $\rhd$. The argument for $\tau(A,B) \rhd B$ is similar. It remains to show that if C is such that $C \rhd A$ and $C \rhd B$, then $C \rhd \tau(A,B)$. We clearly have $C \rhd \sigma(A,B)$ for any such C. Consider first the case that $A \leq B$. Then

$$\Box(\tau(A,B) \leftrightarrow \sigma(A,B) \wedge \Box\neg\sigma(A,B)).$$

Since $D \rhd D \wedge \Box\neg D$ is a theorem of ILM, we have that $\sigma(A,B) \rhd \sigma(A,B) \wedge \Box\neg\sigma(A,B)$, and therefore also $\sigma(A,B) \rhd \tau(A,B)$ by transitivity of $\rhd$. In case $B < A$, we have that $\Box(\tau(A,B) \leftrightarrow \sigma(A,B) \vee \Diamond\sigma(A,B))$, and so $\Box(\sigma(A,B) \to \tau(A,B))$, whence again $\sigma(A,B) \rhd \tau(A,B)$. Thus if $C \rhd \sigma(A,B)$, then also $C \rhd \tau(A,B)$.

Finally, we show τ is not commutative with respect to provability. For that, let A and B be distinct $\mathscr{L}$-sentences (thus under a reasonable gödelnumbering, their gödelnumbers are also different). We assume without loss of generality that $A < B$, and suppose for a contradiction that $\text{PA} \vdash \tau(A,B) \leftrightarrow \tau(B,A)$. By definition of τ, this means that

(27) $\qquad \text{PA} \vdash \sigma(A,B) \wedge \Box\neg\sigma(A,B) \leftrightarrow \sigma(A,B) \vee \Diamond\sigma(A,B).$

It follows from (27) by propositional reasoning that $\text{PA} \vdash \Box\neg\sigma(A,B)$, and thus $\text{PA} \vdash \neg\sigma(A,B)$. But this means that $\sigma(A,B)$, i.e. the supremum of A and B, is in the degree of $[\bot]$, contradicting our assumption.

APPENDIX B. VARIETIES OF SUPREMUM IMPLEMENTATIONS

B.1. **Obtaining supremum implementations.** We show how to turn Švejdar's construction into a supremum implementation as in Definition 7. The case of Visser's construction is similar. Let $\text{Sub}(x,y,z)$ be an intensionally correct formula representing in $I\Delta_0+\text{exp}$ substitution of numerals for free variables in formulas. Thus for all m, n, and all $\mathscr{L}$-formulas A,

$$I\Delta_0+\text{exp} \vdash \text{Sub}(\ulcorner A(u,v)\urcorner, \overline{m}, \overline{n}) = \ulcorner A(\overline{m}, \overline{n})\urcorner,$$

By Theorem 1, there is a formula σ with two free variables, and such that

$$I\Delta_0+\text{exp} \vdash \sigma(A,B) \leftrightarrow \forall x (\Diamond_x \text{Sub}(\ulcorner \sigma(u,v)\urcorner, A, B) \to \Diamond_x A \wedge \Diamond_x B),$$

where A and B are seen as internal variables ranging over $\mathscr{L}$-sentences. By properties of Sub,

$$I\Delta_0+\text{exp} \vdash \sigma(A,B) \leftrightarrow \forall x (\Diamond_x \sigma(A,B) \to \Diamond_x A \wedge \Diamond_x B).$$

Thus the formula σ is exactly what we were looking for.

B.2. **Unique explicit fixed points.** We prove theorems 13 and 14, repeated here as theorems 37 and 41. We start with the following lemma.

Lemma 36. *Let $\varepsilon(A,B)$ be the formula*

$$(\Box_0 \neg A \vee \Box_0 \neg B \to \Box_0 \bot) \wedge \forall x (\Box_{x+1} \neg A \vee \Box_{x+1} \neg B \to \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B)),$$

where A and B are regarded as internal variables ranging over $\mathscr{L}$-sentences. Then

  i. $I\Delta_0+\text{exp} \vdash \Box_0 \neg \varepsilon(A,B) \leftrightarrow \Box_0 \bot$
  ii. $I\Delta_0+\text{exp} \vdash \forall x (\Box_{x+1} \neg \varepsilon(A,B) \leftrightarrow \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B))$

Proof. For the nontrivial direction of i, reason in $I\Delta_0+\text{exp}$, assuming $\Box_0 \neg \varepsilon(A,B)$, i.e.

(28)
$$\Box_0 \left(((\Box_0 \neg A \vee \Box_0 \neg B) \wedge \Diamond_0 \top) \vee \exists x ((\Box_{x+1} \neg A \vee \Box_{x+1} \neg B) \wedge \neg \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B)) \right).$$

(28) implies $\Box_0(\Diamond_0 \top \vee \exists x \neg \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B))$. Since for any C, $\neg \Box_{x+1} C$ implies $\neg \Box_0 \bot$ by monotonicity, it follows from (28) that $\Box_0 \Diamond_0 \top$, and therefore $\Box_0 \bot$ by Löb's principle.

For ii, reason in $I\Delta_0+\text{exp}$, and let x be arbitrary. Assume first

(29)
$$\Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B).$$

To show that $\Box_{x+1} \neg \varepsilon(A,B)$, reason in $\Box_{x+1}$. Suppose for a contradiction that $\varepsilon(A,B)$. By (29), let $\mu \leq x$ be the minimal number s.t. $\Box_\mu \neg A \vee \Box_\mu \neg B$. If $\mu = 0$, then $\Box_0 \bot$ follows by $\varepsilon(A,B)$, which is a contradiction since we have $\neg \Box_0 \bot$ by reflection. Thus $\mu = y+1$ for some y, whence $\varepsilon(A,B)$ now gives us

(30)
$$\Box_{y+1}(\Box_y \neg A \vee \Box_y \neg B).$$

Since $y+1 < x+1$, from (30) we get $\Box_y \neg A \vee \Box_y \neg B$ by reflection, contradicting the minimality of $y+1$. We conclude that $\neg \varepsilon(A,B)$.

For the other direction, assume $\Box_{x+1} \neg \varepsilon(A,B)$. Reason in $\Box_{x+1}$. Note that $\neg \varepsilon(A,B)$ is the sentence

(31) $\quad ((\Box_0 \neg A \vee \Box_0 \neg B) \wedge \Diamond_0 \top) \vee \exists y ((\Box_{y+1} \neg A \vee \Box_{y+1} \neg B) \wedge \neg \Box_{y+1}(\Box_y \neg A \vee \Box_y \neg B)).$

If the first disjunct holds or some $y < x$ witnesses the second disjunct, then by monotonicity we have that $\Box_x \neg A \vee \Box_x \neg B$. If the smallest witness of the second disjunct is some $y \geq x$, then (by monotonicity) $\neg \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B)$. Thus $(\Box_x \neg A \vee \Box_x \neg B) \vee \neg \Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B)$, or in other words

$$\Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B) \to \Box_x \neg A \vee \Box_x \neg B.$$

Leaving the inner $\Box_{x+1}$–world, we have shown: $\Box_{x+1}(\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B) \to \Box_x\neg A \vee \Box_x\neg B)$. By Löb's Theorem we can conclude $\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B)$ as required. $\Box$

Theorem 37. *Let $\varepsilon(A,B)$ be the formula*

$$(\Box_0\neg A \vee \Box_0\neg B \to \Box_0\bot) \wedge \forall x(\Box_{x+1}\neg A \vee \Box_{x+1}\neg B \to \Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B)),$$

where A and B are regarded as internal variables ranging over $\mathscr{L}$-sentences. Then

  i. $\mathrm{I}\Delta_0+\mathsf{exp} \vdash \varepsilon(A,B) \leftrightarrow \forall x(\Diamond_x\varepsilon(A,B) \to \Diamond_xA \wedge \Diamond_xB)$
  ii. For any ϑ, if $\mathsf{PA} \vdash \vartheta \leftrightarrow \forall x(\Diamond_x\vartheta \to \Diamond_xA \wedge \Diamond_xB)$, then $\mathsf{PA} \vdash \vartheta \leftrightarrow \varepsilon(A,B)$

Proof. For i, argue in $\mathrm{I}\Delta_0+\mathsf{exp}$, and assume $\varepsilon(A,B)$. Using contraposition, it suffices to show that for all x, $\Box_x\neg A \vee \Box_x\neg B$ implies $\Box_x\neg\varepsilon(A,B)$. Suppose $\Box_x\neg A \vee \Box_x\neg B$. If $x = 0$, then $\varepsilon(A,B)$ implies $\Box_0\bot$, and hence trivially $\Box_0\neg\varepsilon(A,B)$. If $x = y+1$ for some y, then by $\varepsilon(A,B)$, we have $\Box_{y+1}(\Box_y\neg A \vee \Box_y\neg B)$ and hence $\Box_{y+1}\neg\varepsilon(A,B)$, i.e. $\Box_x\neg\varepsilon(A,B)$ by Lemma 36. For the other direction assume $\neg\varepsilon(A,B)$, i.e. the sentence in (31). By Lemma 36 i, the first disjunct implies $(\Box_0\neg A \wedge \Box_0\neg B) \wedge \Diamond_0\varepsilon(A,B)$, and by Lemma 36 ii the second disjunct implies $\exists x((\Box_{x+1}\neg A \vee \Box_{x+1}\neg B) \wedge \Diamond_{x+1}\varepsilon(A,B))$ In either case $\exists x(\Box_x\neg A \vee \Box_x\neg B \wedge \Diamond_x\varepsilon(A,B))$, i.e. $\neg\forall x(\Diamond_x\varepsilon(A,B) \to \Diamond_xA \wedge \Diamond_xB)$, which is what we wanted to show.

For ii, it now suffices to show that any two fixed points of the formula $\forall x(\Diamond_xY \to (\Diamond_xA \wedge \Diamond_xB))$ are provably equivalent in PA. Let n be s.t.

$$\mathrm{I}\Sigma_n \vdash \vartheta \leftrightarrow \forall x(\Box_x\neg A \vee \Box_x\neg B \to \Box_x\neg\vartheta) \quad \text{and} \quad \mathrm{I}\Sigma_n \vdash \sigma \leftrightarrow \forall x(\Box_x\neg A \vee \Box_x\neg B \to \Box_x\neg\sigma).$$

We show that $\mathrm{I}\Sigma_n \vdash \sigma \leftrightarrow \vartheta$. Using Löb's Theorem, it suffices to show $\mathrm{I}\Sigma_n \vdash \Box_n(\vartheta \leftrightarrow \sigma) \to (\sigma \leftrightarrow \vartheta)$. We argue in $\mathrm{I}\Sigma_n$ and assume $\Box_n(\vartheta \leftrightarrow \sigma)$. By monotonicity and validity of K for $\Box_n$, we have for all $x \geq n$:

$$(32) \qquad\qquad\qquad \Box_x\neg\vartheta \leftrightarrow \Box_x\neg\sigma.$$

Assume σ. In order to show ϑ, it suffices to show $\forall x(\Box_x\neg A \vee \Box_x\neg B \to \Box_x\neg\vartheta)$. Assume $\Box_x\neg A \vee \Box_x\neg B$. By σ, we get $\Box_x\neg\sigma$. By essential reflexivity, it must be that $x \geq n$, and thus $\Box_x\neg\vartheta$ follows by (32). $\Box$

We now turn our attention to Visser's supremum. We first prove a few lemmas.

Lemma 38. *Suppose that $\mathrm{I}\Sigma_n \vdash \vartheta \leftrightarrow \forall x(\Box_x\vartheta \to \Box_x\neg A \vee \Box_x\neg B)$. Then*

$$\mathrm{I}\Sigma_n \vdash \forall x \geq n \Box_x(\forall y < x(\Box_y\vartheta \to \Box_y\neg A \vee \Box_y\neg B)).$$

Proof. By necessitation,

$$\mathrm{I}\Sigma_n \vdash \Box_n(\vartheta \leftrightarrow \forall x(\Box_x\vartheta \to \Box_x\neg A \vee \Box_x\neg B)),$$

whence we can use ϑ and $\forall x(\Box_x\vartheta \to \Box_x\neg A \vee \Box_x\neg B)$ interchangeably in $\Box_x$ if $x \geq n$.

$$\mathrm{I}\Sigma_n \vdash x \geq n \to \Box_x\forall y < x(\Box_y\vartheta \to \vartheta) \qquad\qquad\qquad\qquad \text{(reflection)}$$
$$\to \Box_x\forall y < x(\Box_y\vartheta \to \forall z(\Box_z\vartheta \to \Box_z\neg A \vee \Box_z\neg B)) \quad \text{(fixed point version of } \vartheta\text{)}$$
$$\to \Box_x\forall y < x(\Box_y\vartheta \to \Box_y\neg A \vee \Box_y\neg B)$$

                                                                                                    $\Box$

Lemma 39. *Suppose that $\mathrm{I}\Sigma_n \vdash \vartheta \leftrightarrow \forall x(\Box_x\vartheta \to \Box_x\neg A \vee \Box_x\neg B)$. Then*

$$\mathrm{I}\Sigma_n \vdash \forall x \geq n(\Box_x\vartheta \leftrightarrow \Box_x(\Box_x\neg A \vee \Box_x\neg B)).$$

Proof.

$$
\begin{aligned}
\mathrm{I}\Sigma_n \vdash x \geq n \to (\Box_x \vartheta &\to \Box_x \forall y \, (\Box_y \vartheta \to \Box_y \neg A \vee \Box_y \neg B)) && \text{(fixed point version of } \vartheta) \\
&\to \Box_x (\Box_x \vartheta \to \Box_x \neg A \vee \Box_x \neg B) && \text{(instantiating } \forall) \\
&\to (\Box_x \Box_x \vartheta \to \Box_x (\Box_x \neg A \vee \Box_x \neg B)) && \text{(K-axiom)} \\
&\to \Box_x (\Box_x \neg A \vee \Box_x \neg B) && \text{(since } \Box_x \vartheta \to \Box_x \Box_x \vartheta) \\
&\to \Box_x \forall y \geq x \, (\Box_y \neg A \vee \Box_y \neg B) && \text{(monotonicity)} \\
&\to \Box_x \forall y \geq x \, (\Box_y \vartheta \to \Box_y \neg A \vee \Box_y \neg B) && \\
&\to \Box_x \forall y \, (\Box_y \vartheta \to \Box_y \neg A \vee \Box_y \neg B) && \text{(Lemma 38)} \\
&\to \Box_x \vartheta && \text{(fixed point version of } \vartheta)
\end{aligned}
$$

$\square$

Lemma 40. *Let $\varepsilon(A,B)$ be the formula $\exists x \, (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge \Diamond_x A \wedge \Diamond_x B)$, where A and B are regarded as internal variables ranging over sentences. Then*

$$
\mathrm{I}\Delta_0 + \exp \vdash \forall x \, (\Box_x \neg \varepsilon(A,B) \leftrightarrow \Box_x(\Box_x \neg A \vee \Box_x \neg B)).
$$

Proof.

$$
\begin{aligned}
(33) \quad \mathrm{I}\Delta_0 + \exp \vdash \Box_x \neg \varepsilon(A,B) &\to \Box_x \forall y \, (\Box_y(\Box_y \neg A \vee \Box_y \neg B) \to \Box_y \neg A \vee \Box_y \neg B) \\
(34) \quad &\to \Box_x(\Box_x(\Box_x \neg A \vee \Box_x \neg B) \to \Box_x \neg A \vee \Box_x \neg B) \\
(35) \quad &\to \Box_x(\Box_x \neg A \vee \Box_x \neg B) \\
(36) \quad &\to \Box_x \forall y \geq x \, (\Box_y \neg A \vee \Box_y \neg B) \\
(37) \quad &\to \Box_x \forall y \geq x \, (\Box_y(\Box_y \neg A \vee \Box_y \neg B) \to \Box_y \neg A \vee \Box_y \neg B) \wedge \\
(38) \quad &\qquad \Box_x \forall y < x \, (\Box_y(\Box_y \neg A \vee \Box_y \neg B) \to \Box_y \neg A \vee \Box_y \neg B) \\
(39) \quad &\to \Box_x \forall y \, (\Box_y(\Box_y \neg A \vee \Box_y \neg B) \to \Box_y \neg A \vee \Box_y \neg B) \\
(40) \quad &\to \Box_x \neg \varepsilon(A,B)
\end{aligned}
$$

Step (35) is Löb's Theorem, step (36) monotonicity, step (37) the previous step together with propositional logic, and step (38) reflection. $\square$

Theorem 41. *Let $\varepsilon(A,B)$ be the formula*

$$
(41) \qquad \exists x \, (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B).
$$

where A and B are regarded as internal variables ranging over $\mathscr{L}$-sentences. Then

  i. $\mathsf{PA} \vdash \varepsilon(A,B) \leftrightarrow \exists x \, (\Box_x \neg \varepsilon(A,B) \wedge (\Diamond_x A \wedge \Diamond_x B))$
  ii. *For any ϑ, if $\mathsf{PA} \vdash \vartheta \leftrightarrow \exists x \, (\Box \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B))$, then $\mathsf{PA} \vdash \vartheta \leftrightarrow \varepsilon(A,B)$*

Proof. Item i is an immediate consequence of Lemma 40, together with the definition of $\varepsilon(A,B)$. For ii, assume $\mathrm{I}\Sigma_n \vdash \vartheta \leftrightarrow \exists x \, (\Box_x \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B))$. By contraposition,

$$
\mathrm{I}\Sigma_n \vdash \neg \vartheta \leftrightarrow \forall x \, (\Box_x \neg \vartheta \to (\Box_x \neg A \vee \Box_x \neg B)).
$$

By Lemma 39 and reflection,

$$
(42) \qquad \mathrm{I}\Sigma_n \vdash \neg \vartheta \leftrightarrow \forall x \geq n \, (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \to \Box_x \neg A \vee \Box_x \neg B).
$$

On the other hand, we have by reflection

$$
(43) \qquad \mathrm{I}\Sigma_n \vdash \forall x < n \, (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \to \Box_x \neg A \vee \Box_x \neg B).
$$

Combining (42) and (43), we have $I\Sigma_n \vdash \neg\vartheta \leftrightarrow \forall x\,(\Box_x(\Box_x\neg A \vee \Box_x\neg B) \to \Box_x\neg A \vee \Box_x\neg B)$, whence by contraposition, $I\Sigma_n \vdash \vartheta \leftrightarrow \exists x\,(\Box_x(\Box_x\neg A \vee \Box_x\neg B) \wedge (\Diamond_x A \wedge \Diamond_x B))$, i.e. $I\Sigma_n \vdash \vartheta \leftrightarrow \varepsilon(A,B)$ as required. $\qquad\square$

B.3. **Extensionality.** We prove Theorem 15 from Section 4. We start with a lemma.

Lemma 42 ($I\Delta_0+\mathsf{exp}$)**.** *Let $\varepsilon(A,B)$ be as in Theorem 37. Then*

$$(44) \qquad I\Sigma_{n+1} \vdash \varepsilon(A,B) \leftrightarrow$$

$$(45) \qquad\qquad (\Diamond_n A \wedge \Diamond_n B) \wedge \forall x \geq n(\Diamond_{x+1}(\Diamond_x A \wedge \Diamond_x B) \to \Diamond_{x+1}A \wedge \Diamond_{x+1}B)$$

Proof. By propositional reasoning, we see that $\varepsilon(A,B)$ is equivalent to

$$(46) \qquad (\Diamond_0\top \to \Diamond_0 A \wedge \Diamond_0 B) \wedge \forall x\,(\Diamond_{x+1}(\Diamond_x A \wedge \Diamond_x B) \to \Diamond_{x+1}A \wedge \Diamond_{x+1}B).$$

Argue in $I\Sigma_{n+1}$, assuming (46). By reflection, $\Diamond_0\top$, and therefore $\Diamond_0 A \wedge \Diamond_0 B$ from the first conjunct of (46). Now let $x < n$. By reflection for $I\Sigma_{x+1}$ (remember that we are reasoning inside $I\Sigma_{n+1}$), we have that

$$\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B) \to (\Box_x\neg A \vee \Box_x\neg B),$$

i.e. by contraposition,

$$(47) \qquad\qquad \Diamond_x A \wedge \Diamond_x B \to \Diamond_{x+1}(\Diamond_x A \wedge \Diamond_x B).$$

Since we have $\Diamond_0 A \wedge \Diamond_0 B$, we can apply (47) and the second conjunct of (46) to get $\Diamond_{x+1}A \wedge \Diamond_{x+1}B$ for all $x < n$, and thus $\Diamond_n A \wedge \Diamond_n B$. The other conjunct of (45) clearly follows from (46). This finishes the proof from (44) to (45). For the other direction, we note that $\Diamond_n A \wedge \Diamond_n B$ implies $\Diamond_0 A \wedge \Diamond_0 B$, taking care of the first conjunct of (46). For the other conjunct, it suffices to show that for $1 \leq x < n$,

$$(48) \qquad\qquad \Diamond_{x+1}(\Diamond_x A \wedge \Diamond_x B) \to \Diamond_{x+1}A \wedge \Diamond_{x+1}B.$$

This follows because we have $\Diamond_{x+1}A \wedge \Diamond_{x+1}B$ for all such x by assumption. $\qquad\square$

Theorem 43. *Let $\varepsilon(A,B)$ be as in Theorem 37. Then*

$$I\Delta_0+\mathsf{exp} \vdash \Box(A \leftrightarrow A') \wedge \Box(B \leftrightarrow B') \to \Box(\varepsilon(A,B) \leftrightarrow \varepsilon(A',B'))$$

Proof. Clearly it suffices to show that

$$I\Delta_0+\mathsf{exp} \vdash \Box_x(A \leftrightarrow A') \wedge \Box_x(B \leftrightarrow B') \to \Box_{x+1}(\varepsilon(A,B) \leftrightarrow \varepsilon(A',B')).$$

Argue in $I\Delta_0+\mathsf{exp}$, assuming

$$(49) \qquad\qquad \Box_n(A \leftrightarrow A') \wedge \Box_n(B \leftrightarrow B').$$

Now argue inside $I\Sigma_{n+1}$. Using Lemma 42 it suffices to show that

$$(50) \qquad (\Diamond_n A \wedge \Diamond_n B) \wedge \forall x \geq n(\Diamond_{x+1}(\Diamond_x A \wedge \Diamond_x B) \to \Diamond_{x+1}A \wedge \Diamond_{x+1}B)$$

if and only if

$$(51) \qquad (\Diamond_n A' \wedge \Diamond_n B') \wedge \forall x \geq n(\Diamond_{x+1}(\Diamond_x A' \wedge \Diamond_x B') \to \Diamond_{x+1}A' \wedge \Diamond_{x+1}B').$$

Since (49) is a Σ_1-sentence, its truth carries over to our current surroundings inside $I\Sigma_{n+1}$. By monotonicity and modal reasoning, it follows that for all $x \geq n$, $\Box_x A \leftrightarrow \Box_x A'$ and $\Diamond_x A \leftrightarrow \Diamond_x A'$; similarly for B and B'. Given that, the equivalence of (50) and (51) is clear. $\qquad\square$

Theorem 44. *Let $\varepsilon(A,B)$ be as in Theorem 41. Then*

$$I\Delta_0+\mathsf{exp} \vdash \Box(A \leftrightarrow A') \wedge \Box(B \leftrightarrow B') \to \Box(\varepsilon(A,B) \leftrightarrow \varepsilon(A',B'))$$

Proof. We shall show that

$$\mathrm{I}\Delta_0+\mathsf{exp} \vdash \Box_x(A \leftrightarrow A') \wedge \Box_x(B \leftrightarrow B') \rightarrow \Box_x(\varepsilon(A,B) \leftrightarrow \varepsilon(A',B')).$$

Argue in $\mathrm{I}\Delta_0+\mathsf{exp}$, assuming

$$\text{(52)} \qquad\qquad\qquad \Box_n(A \leftrightarrow A') \wedge \Box_n(B \leftrightarrow B')$$

Argue in $\mathrm{I}\Sigma_n$. We want to show that

$$\text{(53)} \qquad\qquad\qquad \exists x \, (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B).$$

if and only if

$$\text{(54)} \qquad\qquad\qquad \exists x \, (\Box_x(\Box_x \neg A' \vee \Box_x \neg B') \wedge (\Diamond_x A' \wedge \Diamond_x B').$$

Since (52) is a Σ_1-sentence, its truth carries over to the world inside $\mathrm{I}\Sigma_n$. By monotonicity and modal reasoning, it follows that for all $x \geq n$, $\Box_x A \leftrightarrow \Box_x A'$ and $\Diamond_x A \leftrightarrow \Diamond_x A'$, and also $\Box_x(\Box_x A \leftrightarrow \Box_x A')$ (again using that $\Box_x A \leftrightarrow \Box_x A'$ is Σ_1). Similarly for B and B'. Given that, the equivalence of (53) and (54) is clear.

$\square$

B.4. **Failure of monotonicity for Visser's supremum.** We prove Theorem 16.

Theorem 45. *Let $\varepsilon(A,B)$ be $\exists x \, (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B)$, where A and B are seen as internal variables over $\mathscr{L}$-sentences. The formula ε is not monotone.*

Proof. It suffices to show that there is a sentence A for which

$$\text{(55)} \qquad\qquad\qquad \mathsf{PA} \nvdash \varepsilon(A,A) \rightarrow \varepsilon(\top,\top).$$

In fact we can take for A the sentence $\varepsilon(\top,\top)$. We note that:

$$\text{(56)} \qquad\qquad \varepsilon(\top,\top) = \exists x \, (\Box_x \Box_x \bot \wedge \Diamond_x \top)$$

$$\text{(57)} \qquad \varepsilon(\varepsilon(\top,\top),\varepsilon(\top,\top)) = \exists x \, (\Box_x \Box_x \neg \varepsilon(\top,\top) \wedge \Diamond_x \varepsilon(\top,\top))$$

Recall that by Lemma 40,

$$\text{(58)} \qquad\qquad \mathrm{I}\Delta_0+\mathsf{exp} \vdash \forall x \, (\Box_x \neg \varepsilon(\top,\top) \leftrightarrow \Box_x \Box_x \bot)$$

Using (58), we see that, verifiably in $\mathrm{I}\Delta_0+\mathsf{exp}$, (57) is equivalent to

$$\text{(59)} \qquad\qquad\qquad \exists x \, (\Box_x \Box_x \Box_x \bot \wedge \Diamond_x \Diamond_x \top)$$

Thus in order to show (55) it suffices to show

$$\mathsf{PA} \nvdash \exists x \, (\Box_x \Box_x \Box_x \bot \wedge \Diamond_x \Diamond_x \top) \rightarrow \exists x \, (\Box_x \Box_x \bot \wedge \Diamond_x \top).$$

Suppose for a contradiction that $\exists x \, (\Box_x \Box_x \Box_x \bot \wedge \Diamond_x \Diamond_x \top) \rightarrow \exists x \, (\Box_x \Box_x \bot \wedge \Diamond_x \top)$ is provable in some $\mathrm{I}\Sigma_n$. Argue in $\mathrm{I}\Sigma_n$. From our assumption, it follows that in particular

$$\text{(60)} \qquad\qquad (\Box_n \Box_n \Box_n \bot \wedge \Diamond_n \Diamond_n \top) \rightarrow \exists x \, (\Box_x \Box_x \bot \wedge \Diamond_x \top).$$

Assume $\Box_n \Box_n \Box_n \bot \wedge \Diamond_n \Diamond_n \top$, and let x be such that $\Box_x \Box_x \bot \wedge \Diamond_x \top$. Since we have reflection for all $m < n$, it must be that $x \geq n$. But if $x > n$, then $\Box_n \Box_n \Box_n \bot$ implies $\Box_x \bot$ by monotonicity and reflection. Thus it must be that $x = n$. We now exit the world inside $\mathrm{I}\Sigma_n$. We have shown:

$$\mathrm{I}\Sigma_n \vdash (\Box_n \Box_n \Box_n \bot \wedge \Diamond_n \Diamond_n \top) \rightarrow (\Box_n \Box_n \bot \wedge \Diamond_n \top);$$

It follows by propositional logic that

$$\mathrm{I}\Sigma_n \vdash \Box_n \Box_n \Box_n \bot \rightarrow \Box_n \Box_n \bot.$$

By Löb's Theorem, we now have $\mathrm{I}\Sigma_n \vdash \Box_n \Box_n \bot$, a contradiction.

$\square$