Substitutions of Σ_1^0-Sentences

explorations between intuitionistic

propositional logic and intuitionistic arithmetic

Albert Visser

Department of Philosophy, Utrecht University

Heidelberglaan 8, 3584 CS Utrecht, The Netherlands

email: `Albert.Visser@phil.uu.nl`

December 14, 2002

Dedicated to Anne Troelstra on the occasion of his 60th birthday.

Abstract

This paper is concerned with notions of consequence. On the one hand we study admissible consequence, specifically for substitutions of Σ_1^0-sentences over Heyting Arithmetic (HA). On the other hand we study preservativity relations. The notion of preservativity of sentences over a given theory is a dual of the notion of conservativity of formulas over a given theory.

We show that admissible consequence for Σ_1^0-substitutions over HA coincides with NNIL-preservativity over intuitionistic propositional logic (IPC). Here NNIL is the class of propositional formulas with *no nestings of implications to the left*.

The identical embedding of IPC-derivability (considered as a preorder and, thus, as a category) into a consequence relation (considered as a preorder) has in many cases a left adjoint. The main tool of the present paper will be an algorithm to compute this left adjoint in the case of NNIL-preservativity.

In the last section we employ the methods developed in the paper to give a characterization the closed fragment of the provability logic of HA.

Key words: Constructive Logic, Propositional Logic, Heyting's Arithmetic, Schema, Admissible Rule, Consequence Relation, Provability Logic

MSC2000 codes: 03B20, 03F40, 03F45, 03F55

Contents

1 Introduction

This paper is a study both of constructive propositional logic and of constructive arithmetic. In the case of arithmetic the focus is not on proof theoretical strength but on the 'logical' properties of arithmetical theories.

I will first explain the contents of the part of the paper that is concerned purely with propositional logic. This part is independent of the arithmetical part and could be studied on its own. After that, I will explain the arithmetical part.

1.1 Propositional Logic

Intuitionistic propositional logic, IPC, differs from classical propositional logic in that it admits natural hierarchies of formulas. The first level of one of these hierarchies is formed by the NNIL-formulas. 'NNIL' stands for *no nestings of implications to the left*. Thus $(p \to (q \to (r \vee (s \to t))))$ is a NNIL-formula and $((p \to q) \to r)$ isn't. NNIL formulas are the analogues of purely universal sentences in the prenex normal form hierarchy for classical predicate logic, in view of the folowing characterization: modulo IPC-provable equivalence this class coincides with ROB the class of formulas that are preserved under taking sub-Kripke-models.[1]

[1] In fact, there is a way of strengthening the characterization, that makes NNIL a natural subclass of the purely universal formulas of a specific theory in predicate logic. We can reformulate Kripke semantics in a familiar way as a translation of IPC into a suitable theory in predicate logic, say T, of Kripke structures. One can characterize the formulas of IPC, modulo T-provable equivalence, as precisely the predicate logical formulas in one free variable that are upwards persistent and preserved under bisimulation. The NNIL-formulas are, modulo provable equivalence, precisely the predicate logical formulas in one free variable that are upwards persistent, preserved under bisiuation and preserved under taking submodels. Thus, the NNIL-formulas are, modulo provable equivalence, precisely the purely universal formulas of T that correspond via the translation to IPC-formulas. See [VvBdJdL95].

Here ROB stands for *robust*. The characterization can be proved in at least two ways. One proof is presented in [VvBdJdL95]. The idea of this proof is due to Johan van Benthem. Van Benthem's proof is analogous to the classical proof of Łoś-Tarski theorem. See also [vB95]. Another proof is contained in the preprints [Vis85] and [Vis94]. This second proof is the one presented in this paper (see section 7). The basic idea of the proof is that, given an arbitrary propositional formula A, one computes its best robust approximation from below, say A^*. The computation will terminate in a NNIL-formula. Since NNIL-formulas are trivially robust, it follows that robust formulas can always be rewritten to IPC-equivalent NNIL-formulas.

Reflecting on the proof, one sees that it is best understood in terms of the following notions. We define IPC,ROB-*preservativity* as follows:

- $A \rhd_{\mathsf{IPC,ROB}} B :\Leftrightarrow \forall C{\in}\mathsf{ROB}\ \ (C \vdash_{\mathsf{IPC}} A \Rightarrow C \vdash_{\mathsf{IPC}} B\,).$

The claim that A^* is the best robust approximation of A, can be analysed to mean the following: $A \rhd_{\mathsf{IPC,ROB}} B \Leftrightarrow A^* \vdash_{\mathsf{IPC}} B$. Thus $(\cdot)^*$ is the left adjoint of the identical embedding of $\vdash_{\mathsf{IPC}}$ considered as a preorder (and thus as a category) into $\rhd_{\mathsf{IPC,ROB}}$ considered as a preorder. As a spin-off of our proof we obtain an 'axiomatization' of $\rhd_{\mathsf{IPC,ROB}}$.

Preservativity can be viewed as dual to conservativity of sentences over a given theory. Note that in the constructive context preservativity and conservativity cannot be transformed into each other (modulo provable equivalence).

Another important notion of consequence is *admissible consequence*, discussed in the next section. In the present paper we will study the two consequence notions, preservativity & admissibility, in tandem.

1.2 Arithmetic

A Σ^0_1-formula is a formula in the language of arithmetic consisting of a block of *existential quantifiers* followed by a formula containing only bounded quantifiers. A Π^0_1-formula is a formula in the language of arithmetic consisting of a block of *universal quantifiers* followed by a formula containing only bounded quantifiers. We will often write 'Σ-formula' and 'Π-formula' for Σ^0_1-formula, respectively Π^0_1-formula. In the classical context, Σ-formulas and Π-formulas are interdefinable using only Boolean conectives: a Π-formula is the negation of Σ-formula and a Σ-formula is the negation of a Π-formula. Constructively, under minimal arithmetical assumptions, a Π-formula is still provably equivalent to the negation of a Σ-formula, but e.g. over Heyting's Arithmetic, we have that if a Σ-sentence is provably equivalent to any Boolean combination of Π-sentences, then it is either provable or refutable.[2]

It is a common observation, that if we have proved some Boolean (or, perhaps, 'Brouwerian') combination of Σ-sentences in Heyting Arithmetic, (HA), then we often know that a *better* Boolean combination of the same Σ-sentences

[2] One way to prove this is by using corollary 9.2 in combination with the disjunction property of Heyting's Arithmetic.

is also provable. Moreover this better Boolean combination can be found independently of the specific Σ-sentences under consideration.

Suppose, for example, that we have found that $\mathsf{HA} \vdash \neg\neg S \to S$. Then, using the Friedman translation, we may show that one also has: $\mathsf{HA} \vdash S \vee \neg S$.

In which sense is $(S \vee \neg S)$ better than $(\neg\neg S \to S)$? Well, we have

$$p \vee \neg p \vdash_{\mathsf{IPC}} \neg\neg p \to p \text{ but not } \neg\neg p \to p \vdash_{\mathsf{IPC}} p \vee \neg p.$$

So *the form* of $(S \vee \neg S)$ is more informative than *the form* of $(\neg\neg S \to S)$. Can we do still better than $(S \vee \neg S)$? Yes and no. *Yes*, since HA has the disjunction property we do have either $\mathsf{HA} \vdash S$ or $\mathsf{HA} \vdash \neg S$. Thus, for a specific S, we can find a further improvement. *No*, if we demand that the improvement depend only on *the form* of the Boolean combination of Σ-sentences, in other words, if we want the improvement to be uniform for arbitrary substitutions of Σ-sentences.

Analyzing these ideas, one arrives at the following explication of what one is looking for. Let $\mathcal{P}$ be a set of propositional variables. $\Sigma^{\mathcal{P}}$ is the set of assignments of Σ-sentences to the propositional variables. Let f be such an assignment of Σ-sentences. We write fA for the result of substituting the fp's for p in A. If we want to stress the fact that we lifted f on $\mathcal{P}$ to a function on the full language of propositional logic over $\mathcal{P}$, we write $[f]$ for the result of the lifting. Thus $[f]$ is a *substitution of* Σ*-sentences*. We define the relation of HA, Σ*-admissible consequence* as follows. Let A, B be IPC-formulas.

- $A \mathrel{\vert\!\sim}_{\mathsf{HA},\Sigma} B \Leftrightarrow \forall f {\in} \Sigma^{\mathcal{P}} \ (\mathsf{HA} \vdash fA \Rightarrow \mathsf{HA} \vdash fB\,)$.

Consider a propositional formula A. Let's call A's desired best improvement A^*. We will show that A^* always exists and satisfies the following claim:

$$A \mathrel{\vert\!\sim}_{\mathsf{HA},\Sigma} B \Leftrightarrow A^* \vdash_{\mathsf{IPC}} B.$$

As the attentive reader will have guessed, $(\cdot)^*$ of subsection 1.1 is identical to $(\cdot)^*$ of subsection 1.2. Thus, we also have: $\mathrel{\vert\!\sim}_{\mathsf{HA},\Sigma} = \rhd_{\mathsf{IPC},\mathsf{ROB}}$.

1.3 Provability logic

Pure provability logic has, in my opinion, two great open problems. The first one is to characterize the provability logics of Wilkie and Paris' theory $I\Delta_0 + \Omega_1$ and of Buss' theory S^1_2. This problem has been extensively studied in [Ver93] and [BV93]. The second problem is to characterize the provability logic of HA. This problem is studied in [Vis81], [Vis82], [Vis85] and [IemXX].

In the present paper we will present full characterizations for two *fragments* of the provabilitity logic of HA. The first is the characterization of all formulas of the form $\Box A \to \Box B$, valid in the provability logic of HA, where A and B are purely propositional. This characterization is due to Rosalie Iemhoff. See her paper [Iem99]. The second one is the characterization of the closed fragment of the provability logic of HA. This is my old result, reported first in [Vis85].[3]

In the appendix of the present paper, I formulate a conjecture about the provability logic of HA. Part of this conjecture is that this logic is best formulated in the richer language containing a binary connective for Σ-preservativity.

1.4 History and Context

The results reported in this paper took a long time from conception to publication. The main part of this research here was done between 1981 and 1984. The primary objects of this research were (i) the characterization of the closed fragment of HA and (ii) generalizing the results of Dick de Jongh on propositional formulas of one variable (see [dJ82]).[4] The results of this research were reported in the preprint [Vis85]. For various reasons, however, the preprint was never published. In 1994 I completely rewrote the paper, resulting in the preprint [Vis94]. However, again, the paper was not published. The present version, written in 2000, is an extensive update of the 1994 preprint.

To put the present paper in a broader context of research, let me briefly sum up some related work.

First there are many results on *logics* of theories (see subsubsection 4.2.1). For simple propositional logic there is De Jongh's theorem saying that the propositonal logic of HA and related theories is precisely IPC. See e.g. [Smo73], [Vis85], [dJV96] for some proofs. In the case of e.g. $\mathsf{HA} + \mathsf{MP} + \mathsf{ECT}_0$, we get a stronger logic. See e.g. [Gav81]. A precise characterization of this logic is still lacking. For de Jongh's theorem for Predicate Logic, see [dJ70], [Lei75], [vO91]. Negative results for the case of intuitionistic predicate logic are contained in [Pli77], [Pli78] and [Pli83]. For the study of schemes in the classical predicate calculus, the reader is referred to [Yav97].

Secondly, we have provability logic. Here the schematic languages are languages of modal logic. The reader is referred to the excellent textbooks [Smo85] and [Boo93]. For the work in this paper, the development of interpretability logic is of particular interest. See the survey papers [JdJ98] and [Vis98a]. Our central notion Σ-preservativity is closely related to the notion of Π^0_1-conservativity, studied in interpretability logic. This connection turns out to be really useful: the expertise generated by research into Kripke models for interpretability logic, is now used by Rosalie Iemhoff in the study of the provability logic of HA and of admissible rules. See [IemXX] and [Iem99].

[3] The problem to characterize the closed fragment of the provability logic of PA was Friedman's 35th problem. See [Fri75]. It was solved indepently by van Benthem, Magari and Boolos.

[4] The full generalization of de Jongh's results had to wait till 1998 (see [Vis99]). The crucial lemma, was proved by Silvio Ghilardi, following a rather different line of enquiry. See [Ghi99]. Ghilardi's lemma characterizes two classes of propositional formulas, the projective formulas and the exact formulas as precisely those with the extension property —a property formulated in terms of Kripke models. Guram Bezhanishvili showed me that Ghilardi's lemma can also be obtained by applying the work of R. Grigolia. See [Gri87].

Thirdly, the work on axiom schemes can be generalized to the study of admissible rules, exact formulas, unification and the like. We refer the reader to [Lei80], [dJ82], [Ryb92], [dJC95], [dJV96], [Ryb97], [Ghi99], [Iem99]. A closely related subject is uniform interpolation. See [Pit92], [GZ95a], [GZ95b] and [Vis96].

Fourthly and finally, there is the issue of formula classes, like NNIL, and their characterization. These matters are taken up in e.g. [Lei81], [vB95], [VvBdJdL95][5], [Bur98] and [Vis98b].

1.5 Acknowledgements

In various stages of research I benefited from the work, the wisdom and/or the advice of: Johan van Benthem, Guram Bezhanishvili, Dirk van Dalen, Rosalie Iemhoff, Dick de Jongh, Karst Koymans, Jaap van Oosten, Piet Rodenburg, Volodya Shavrukov, Rick Statman, Anne Troelstra and Domenico Zambella. Lev Beklemishev spotted a silly mistake in the penultimate version of the paper. I am grateful to the anonymous referee for his/her helpful comments. I thank Sander Hermsen for making the pictures for this paper.

1.6 Prerequisites

Some knowledge of [TvD88a], [TvD88b] is certainly beneficial. At some places I will make use of results from [Vis82] and [dJV96].

2 Formulas of Propositional Logic

In this section we fix some notations. We also introduce the central formula class NNIL.

$\mathcal{P}$, $\mathcal{Q}$, ... will be *sets of propositional variables*. $\vec{p}$, $\vec{q}$, $\vec{r}$, ... will be *finite sets of propositional variables*. We define $\mathcal{L}(\mathcal{P})$ as the smallest set $\mathcal{S}$ such that:

- $\mathcal{P} \subseteq \mathcal{S}$, $\top, \bot \in \mathcal{S}$,
- If $A, B \in \mathcal{S}$, then $(A \wedge B)$, $(A \vee B)$, $(A \to B) \in \mathcal{S}$.

$\mathsf{sub}(A)$ is the set of subformulas of A. By convention we will count $\bot$ as a subformula of any A. $\mathsf{pv}(A)$ is the set of propositional variables occurring in A.

We define a measure of complexity ρ, which counts the left-nesting of $\to$, as follows:

- $\rho p := \rho\bot := \rho\top := 0$
- $\rho(A \wedge B) := \rho(A \vee B) := \max(\rho A, \rho B)$
- $\rho(A \to B) := \max(\rho A + 1, \rho B)$

[5]This paper may be considered as a companion paper of the present paper.

$\mathsf{NNIL}(\mathcal{P}) := \{A \in \mathcal{L}(\mathcal{P}) \mid \rho A \leq 1\}$. In other words NNIL is the class of formulas without nestings of implications to the left. An example of a NNIL-formula is: $(p \to (q \vee (s \to t))) \wedge ((q \vee r) \to s))$. It is easy to see that modulo IPC-provable equivalence each NNIL-formula can be rewritten to a NNIL_0-formula, i.e. a formula in which as antecedents of implications only single atoms occur. For more information about NNIL, see [dJV96] and [VvBdJdL95]. For a generalization of the result of [VvBdJdL95] to the case of predicate logic, see [Vis98b].

3 Semi-consequence

In this section we introduce the basic notion of semi-consequence relation. '$\rhd$' will range over semi-consequence relations. Let $\vdash$ stand for derivability in IPC.

3.1 Basic Definitons

Let $\mathcal{B}$ be a language (for propositional or predicate logic) and let T be a theory in $\mathcal{B}$. A *semi-consequence relation on $\mathcal{B}$ over T* is a binary relation on the set of $\mathcal{B}$-formulas satisfying:

A1 $\quad A \vdash_T B \Rightarrow A \rhd B$

A2 $\quad A \rhd B$ and $B \rhd C \Rightarrow A \rhd C$

A3 $\quad C \rhd A$ and $C \rhd B \Rightarrow C \rhd (A \wedge B)$

The name 'semi-consequence relation' is ad hoc in this paper. We take:

- $A \equiv B :\Leftrightarrow A \rhd B$ and $B \rhd A$.

If we don't specify the theory corresponding to the semi-consequence relation, it is always supposed to be over IPC. A further salient principle is:

- B1 $\;A \rhd C$ and $B \rhd C \Rightarrow (A \vee B) \rhd C$.

A relation satisfying A1-A3 and B1 is called a *nearly-consequence* relation. Note that $\vdash_T$ is a nearly-consequence relation over T.

We will use $X \vdash Y$, $X \rhd Y$ for respectively $\bigwedge X \vdash \bigvee Y$ and $\bigwedge X \rhd \bigvee Y$, where X and Y are finite sets of formulas. Here $\bigwedge \emptyset := \top$ and $\bigvee \emptyset := \bot$. We treat implications similarly, writing $(X \to Y)$ for $(\bigwedge X \to \bigvee Y)$. We write X, Y for $X \cup Y$; X, A for $X \cup \{A\}$, etcetera.

Nearly-consequence relations over T can be alternatively described in Genzen style as follows. Nearly-consequence relations are relations between (finiete sets of) formulas satifying the following conditions.

A1′ $\quad X \vdash_T Y \Rightarrow X \rhd Y$

Thin $\quad X \rhd Y \Rightarrow X, Z \rhd Y, U$

Cut $\quad X \rhd Y, A$ and $Z, A \rhd U \Rightarrow X, Z \rhd Y, U$

We take the permutation rules to be implicit in the set notation. We leave it to the reader to check the equivalence of the Genzen style principles with A1-A3 plus B1.

We will be interested in adjoints involving semi-consequence relations. Some basic facts concerning adjoints and preorders are given in appendix A

3.2 Principles Involving Implication

Principles involving implication play an important role in the paper. In fact, the main reason for the complexity of e.g. the admissible rules of IPC, the admissible rules of HA and the admissible rules of HA for Σ-substitutions is the presence of nested implications and the interplay of implications and disjunctions (in the antecedent).

Regrettably, the principles involving implication needed to characterize admissible consequence and Σ-preservativity are never going to win a beauty contest.

The principes introduced below, allow us to reduce the complexity of formulas. E.g. if a formula of the form $(C \to D) \to (E \vee F)$ is provable in IPC (HA), then also $((C \to D) \to C) \vee ((C \to D) \to E) \vee ((C \to D) \to F)$ is provable in IPC (HA). The removal of the disjunction from the consequent makes further simplification possible.

At this point, we will just introduce the principles in full generality. They will become more understandable in the light of their verifications (for various notions of consequence) in e.g. theorem 6.2 and theorem 9.1.

To introduce the principles we need some syntactical operations. We define the operations $[\cdot](\cdot)$ and $\{\cdot\}(\cdot)$ on propositional formulas as follows.

- – $[B]p := p$, $[B]\top := \top$, $[B]\bot := \bot$,
  – $[\cdot](\cdot)$ commutes with $\wedge$ and $\vee$ in the second argument,
  – $[B](C \to D) := (B \to (C \to D))$.
- – $\{B\}p := (B \to p)$, $\{B\}\top := \top$, $\{B\}\bot := \bot$,
  – $\{\cdot\}(\cdot)$ commutes with $\wedge$ and $\vee$ in the second argument,
  – $\{B\}(C \to D) := (B \to (C \to D))$.

Note that $[\cdot](\cdot)$ and $\{\cdot\}(\cdot)$ do not preserve provable equivalence in the second argument. Note also that $\vdash B \to ([B]C \leftrightarrow C)$ and $\vdash B \to (\{B\}C \leftrightarrow C)$. Define $[B]X := \{[B]C \mid C \in X\}$.

We will study the following principles for implication for semi-consequence relations on $\mathcal{L}(\mathcal{P})$.

B2 Let X be a finite set of implications and let

$$Y := \{C \mid (C \to D) \in X\} \cup \{B\}.$$

Take $A := \bigwedge X$. Then $(A \to B) \rhd [A]Y$

B2′ Let X be a finite set of implications and let

$$Y := \{C \mid (C \to D) \in X\} \cup \{B\}.$$

Take $A := \bigwedge X$. Then $(A \to B) \rhd \{A\}Y$.

B3 $A \rhd B \Rightarrow (p \to A) \rhd (p \to B)$.

It is easy to see that B2′ can be derived from B2 over A1-A3. Note that both B2, B2′ and B3 are non-ordinary schemes. B2 and B2′ involve finite conjunctions and disjunctions and B2 and B3 contain variables ranging over proposition letters. (It is easy to see that one can replace B2′ by an equivalent scheme that makes no special mention of proposition letters. See also [Iem99].) The fact that proposition letters are not 'generic' in B2 and B3, reflects our interest in certain *special substitutions*, like substitutions of Σ-sentences.

If we spell out the conclusion of e.g. B2, we get:

$$(A \to B) \ \rhd \ \bigvee\{[A]C \mid (C \to D) \in X \text{ or } C = B\}.$$

We give an example of a use of B2. Suppose $\rhd$ satisfies B2. Let

$$E := ((p \to q) \wedge (r \to (s \vee t))),\ F := (E \to (u \vee (p \to r))).$$

Then we have $F \rhd (p \vee r \vee u \vee (E \to (p \to r)))$.

Remark 3.1 Rosalie Iemhoff has shown that B2′ cannot be replaced by finitely many conventional schemes. This is an immediate consequence of the methods developed in [Iem99]. I conjecture that a similar result holds for B2. ❑

We say that a relation satisfying A1, A2, A3, B1, B2, B3 is a *σ-relation*. We say that a relation satisfying A1, A2, A3, B1, B2′ is an α-relation. We take $\blacktriangleright_\sigma$ to be the smallest σ-relation, i.o.w $\blacktriangleright_\sigma$ is the semi-consequence relation axiomatized by A1, A2, A3, B1, B2, B3. Similarly, $\blacktriangleright_\alpha$ is the smallest α-relation.

4 Preservativity

In this section we introduce preservativity relations which are special semi-consequence relations. In fact, we will treat some semi-consequence relations that are not preservativity relations: we will also consider relations like *provably deductive consequence*. Treatment of these further relations is postponed till section 12.

4.1 Basic Definitions

Consider again any language $\mathcal{B}$ of propositional or predicate logic. If $A \in \mathcal{L}(\mathcal{P})$ and $f : \mathcal{P} \to \mathcal{B}$, then fA is the result of substituting fp for p in A for each $p \in \mathcal{P}$. If we want to stress that we are speaking of f as $\mathcal{L}(\mathcal{P}) \to \mathcal{B}$ we will use $[f]$. Let $X \subseteq \mathcal{B}$, let T be any theory in $\mathcal{B}$ and let $\mathcal{F}$ be a set of functions from $\mathcal{P}$ to $\mathcal{B}$. We write:

- for $A, B \in \mathcal{B}$, $A \rhd_{T,X} B :\Leftrightarrow \forall C{\in}X\ (C \vdash_T A \Rightarrow C \vdash_T B)$
- for $A, B \in \mathcal{L}(\mathcal{P})$, $A \rhd_{T,X,\mathcal{F}} B :\Leftrightarrow \forall f{\in}\ \mathcal{F}\ fA \rhd_{T,X} fB$.

If $Y \subseteq \mathcal{B}$ and $\mathcal{F} = Y^{\mathcal{P}}$, we will write $\rhd_{T,X,Y}$ for $\rhd_{T,X,\mathcal{F}}$. If X or $\mathcal{F}$ are singletons we will omit the singleton brackets. If $\mathcal{B} = \mathcal{L}(\mathcal{P})$ and $T = \mathsf{IPC}$, we will often omit 'T' in the index. We will call $\rhd_{T,X,\mathcal{F}}$, $T, X, \mathcal{F}$*-preservativity*, etcetera. We will call the $\rhd_{T,X,\mathcal{F}}$ and the $\rhd_{T,X}$ *preservativity relations.* We will call the $\rhd_{T,X}$ *pure preservativity relations.*

Clearly $\rhd_{T,X}$ is a semi-consequence relation over T and $\rhd_{T,X,\mathcal{F}}$ is a semi-consequence relation over IPC. Below we will provide a number of motivating examples for our definitions.

Remark 4.1 It is instructive to compare preservativity with conservativity (of sentences over a theory). Define:

- For $A, B \in \mathcal{B}$, $A \rhd^*_{T,X} B :\Leftrightarrow \forall C{\in}X(B \vdash_T C \Rightarrow A \vdash_T C)$

So $A \rhd^*_{T,X} B$ means that $T + B$ is conservative over $T + A$ w.r.t. X. For *classical* theories T we have:

✠ $A \rhd_{T,X} B \Leftrightarrow \neg B \rhd^*_{T,\neg X} \neg A$, where $\neg X := \{\neg C \mid C \in X\}$.

Thus, in the classical case preservativity is a superfluous notion. In the constructive case, however, the reduction given in ✠ does not work. Note that conservativity as a relation between sentences over a theory is a natural extension of the notion of conservativity between theories. There is no analogue of this for preservativity.

It should be noted that Π^0_2-conservativity between theories is a central conceptual tool for theory comparison. The only other notions of theory comparison of comparable importance are relative interpretability and verifiable relative consistency.

Π^0_1-conservativity for RE extentensions of Peano Arithmetic, PA, in the language of PA is equivalent to relative interpretability for these theories. In the constructive case, relative interpretability plays a much more modest role. Many translations that do not commute with the logical connectives, play a significant role. Moreover, the metamathematical properties of relative interpretability are much different. We certainly do not have anything resembling the connection between relative interpretability and Π_1-conservativity.

Π^0_1-conservativity of sentences over Heyting's Arithmetic, HA, is reducible to Σ-preservativity over HA by:

- $A \rhd^*_{\mathsf{HA},\Pi^0_1} B \Leftrightarrow \neg B \rhd_{\mathsf{HA},\Sigma^0_1} \neg A$.

As an auxiliary notion to prove metamathematical results, Σ-conservativity over HA is the more useful notion, as will be illustrated in te rest of the paper. ❑

4.2 Examples of Preservativity Relations

In this subsection we provide examples of preservativity relations and we review some results from the literature concerning these examples.

4.2.1 The Logic of a Theory

If we take $X := \mathcal{B}$, then $\rhd_{T,X}$ is simply equal to $\vdash_T$. More interestingly we consider $\rhd_{T,\mathcal{B},\mathcal{F}}$. It is easy to see that:

$$A \rhd_{T,\mathcal{B},\mathcal{F}} B \Leftrightarrow \forall f{\in}\mathcal{F}\ fA \vdash_T fB.$$

We define:

- $\Lambda_{T,\mathcal{F}} := \{A \mid \forall f{\in}\mathcal{F}\ T \vdash fA\}$.
- $\Lambda_T := \Lambda_{T,\mathcal{B}^{\mathcal{P}}}$. The theory Λ_T is *the* propositional logic of T.

It is easy to see that:

$$\begin{array}{rcl} A \rhd_{T,\mathcal{B},\mathcal{F}} B & \Leftrightarrow & \Lambda_{T,\mathcal{F}} \vdash (A \to B) \\ \Lambda_{T,\mathcal{F}} \vdash A & \Leftrightarrow & \top \rhd_{T,\mathcal{B},\mathcal{F}} A \end{array}$$

De Jongh's Completeness Theorem for Σ-substitutions tells us that $\Lambda_{\mathsf{HA}} = \Lambda_{\mathsf{HA},\Sigma} = \mathsf{IPC}$. There are many different proofs of De Jongh's theorem, see e.g. [Smo73] or [Vis85] or [dJV96].

4.2.2 Admissible Consequence

If we take $X := \{\top\}$, then, for $A, B \in \mathcal{B}$, we find

$$A \rhd_{T,\top} B \ \Leftrightarrow\ (T \vdash A \Rightarrow T \vdash B).$$

Thus $\rhd_{T,\top}$ is the relation of *deductive consequence* for T. Moreover, for $A, B \in \mathcal{L}(\mathcal{P})$,

$$A \rhd_{T,\top,\mathcal{F}} B \Leftrightarrow \forall f{\in}\mathcal{F}\ (T \vdash fA \Rightarrow T \vdash fB).$$

We find that $\rhd_{T,\top,\mathcal{F}}$ is admissible consequence for T w.r.t. $\mathcal{F}$. We define:

- $A \ {\mid\!\sim}_{T,\mathcal{F}} B :\Leftrightarrow A \rhd_{T,\top,\mathcal{F}} B$,
- $A \ {\mid\!\sim}_{T} B :\Leftrightarrow A \ {\mid\!\sim}_{T,\mathcal{B}} B$.

We will provide more information about admissible consequence in subsection 4.4.

4.2.3 A Result of Pitts

Suppose $\mathcal{Q} \subseteq \mathcal{P}$. Let $\mathcal{R} := \mathcal{P} \setminus \mathcal{Q}$. Consider $\rhd_{\mathsf{IPC},\mathcal{L}(\mathcal{Q})}$. By theorem 5.3 below, in combination with lemma 4.2(2) below, $\rhd_{\mathsf{IPC},\mathcal{L}(\mathcal{Q})}$ is a nearly-consequence relation. Andrew Pitts, in his [Pit92], shows that for every $A \in \mathcal{L}(\mathcal{P})$, there is a formula $(\forall_{\mathcal{R}} A) \in \mathcal{L}(\mathcal{Q})$, such that for all $B \in \mathcal{L}(\mathcal{P})$: $A \rhd_{\mathsf{IPC},\mathcal{L}(\mathcal{Q})} B \Leftrightarrow \forall_{\mathcal{R}} A \vdash B$.

For a semantical treatment of Pitt's result see [GZ95a] or [Vis96]. Of related interest is the paper [GZ95b].

4.3 Basic Facts

We collect some technical facts about preservativity. A formula A of $\mathcal{B}$ is T-*prime* if, for any finite set of $\mathcal{B}$-formulas Z, $A \vdash_T Z \Rightarrow \exists B \in Z\ A \vdash_T B$. Note that if A is T-prime, then $A \nvdash_T \bot$. A class of formulas X is *weakly T-disjunctive* if every $A \in X$ is equivalent to the disjunction of a finite set of T-prime formulas Y, with $Y \subseteq X$.

For any class of formulas X, let $\mathsf{disj}(X)$ be the closure of X under arbitrary disjunctions. In the next lemma we collect a number of noteworthy small facts on preservativity.

Lemma 4.2 1. $\rhd_{T,X}$ is a semi-consequence relation over T. $\rhd_{T,X,\mathcal{F}}$ is a semi-consequence relation over IPC.

2. Suppose X is weakly T-disjunctive, then $\rhd_{T,X}$ and $\rhd_{T,X,\mathcal{F}}$ are nearly-consequence relations.
3. Suppose X is closed under conjunction, then:

$$C \in X \text{ and } A \rhd_{T,X} B \Rightarrow (C \to A) \rhd_{T,X} (C \to B).$$

4. Let $\mathsf{range}(\mathcal{F})$ be the union of the ranges of the elements of $\mathcal{F}$. Suppose that $\mathsf{range}(\mathcal{F}) \subseteq X$ and that X is closed under conjunction, then $\rhd_{T,X,\mathcal{F}}$ satisfies B3.
5. Suppose $A \in X$. Then $A \rhd_{T,X} B \Leftrightarrow A \vdash_T B$.
6. Suppose $X \subseteq Y$ and $\mathcal{F} \subseteq \mathcal{G}$, then $\rhd_{T,Y} \subseteq \rhd_{T,X}$ and $\rhd_{T,Y,\mathcal{G}} \subseteq \rhd_{T,X,\mathcal{F}}$.
7. $\rhd_{T,\mathsf{disj}(X)} = \rhd_{T,X}$ and, hence, $\rhd_{T,\mathsf{disj}(X),\mathcal{F}} = \rhd_{T,X,\mathcal{F}}$.
8. Let $\mathsf{id} : \mathcal{P} \to \mathcal{L}(\mathcal{P})$ be the identical embedding. Then $\rhd_{\mathsf{IPC},X,\mathsf{id}} = \rhd_{\mathsf{IPC},X}$.
9. If T has the disjunction property, then $\vdash_T$ and $\vdash_{T,\mathcal{F}}$ satisfy B1.

❑

Proof

We treat (2) and (3). (2) Suppose X is weakly T-disjunctive and $A \rhd_{T,X} C$ and $B \rhd_{T,X} C$. Let E be any element of X. Suppose Y is a finite set of T-prime formulas in X such that E is equivalent to the disjunction of Y. We have:

$$\begin{array}{rcl} E \vdash_T A \vee B & \Rightarrow & \bigvee Y \vdash_T A \vee B \\ & \Rightarrow & \forall F{\in}Y\ F \vdash_T A \vee B \\ & \Rightarrow & \forall F{\in}Y\ F \vdash_T A \text{ or } F \vdash_T B \\ & \Rightarrow & \forall F{\in}Y\ F \vdash_T C \\ & \Rightarrow & \bigvee Y \vdash_T C \\ & \Rightarrow & E \vdash_T C \end{array}$$

B1 for $\rhd_{T,X,\mathcal{F}}$ is immediate from B1 for $\rhd_{T,X}$.

(3) Suppose that X is closed under conjunction and that $A \rhd B$. Let $C, E \in X$ and suppose that $E \vdash (C \to A)$. Then $(E \wedge C) \vdash A$ and, hence, $(E \wedge C) \vdash B$. Ergo $E \vdash (C \to B)$. ❑

The next lemma is quite useful both to verify left adjointness and to show that certain classes of formulas are equal modulo IPC-provable equivalence.

Lemma 4.3 Let $\rhd$ be a semi-consequence relation over IPC and let $\Psi : \mathcal{L}(\mathcal{P}) \to \mathcal{L}(\mathcal{P})$, let $X \subseteq \mathcal{L}(\mathcal{P})$. Suppose that

1. $A \rhd \Psi A$,
2. $\Psi A \vdash_{\mathsf{IPC}} A$,
3. $\mathsf{range}(\Psi) \subseteq X$,
4. $\rhd \subseteq \rhd_{\mathsf{IPC},X}$.

Then, we have:

- $A \rhd B \Leftrightarrow A \rhd_{\mathsf{IPC},X} B \Leftrightarrow \Psi A \vdash_{\mathsf{IPC}} B$,
- $X = \mathsf{range}(\Psi)$ modulo IPC-provable equivalence.

❑

Proof

Under the assumptions of the lemma. We have:

$$\begin{array}{rcll} A \rhd B & \Rightarrow & A \rhd_{\mathsf{IPC},X} B & \text{assumption (4)} \\ & \Rightarrow & \Psi A \rhd_{\mathsf{IPC},X} B & \text{assumption (2), (4)} \\ & \Rightarrow & \Psi A \vdash_{\mathsf{IPC}} B & \text{lemma 4.2(5)} \\ & \Rightarrow & \Psi A \rhd B & \\ & \Rightarrow & A \rhd B & \text{assumption 1} \end{array}$$

Consider any $B \in X$. We have $B \rhd \Psi B$. Hence, by assumption (4) and lemma 4.2(5), $B \vdash_{\mathsf{IPC}} \Psi B$. Combining this with assumption (2), we find $\mathsf{IPC} \vdash B \leftrightarrow \Psi B$. ❑

We prove a basic theorem about admissible consequence. Define $f \star g := [f] \circ g$. It is easy to see that $\star$ is associative and the identical embedding of $\mathcal{P}$ into $\mathcal{L}(\mathcal{P})$ is the identity for $\star$.

Theorem 4.4 *Suppose, $\Lambda_{T,\mathcal{F}} = \mathsf{IPC}$. Moreover, suppose that $\mathcal{G} \subseteq \mathcal{L}(\mathcal{P})^{\mathcal{P}}$ and that, for any $g \in \mathcal{G}$ and $f \in \mathcal{F}$, $(f \star g) \in \mathcal{F}$. Then, $\mathrel{|\!\sim}_{T,\mathcal{F}} \subseteq \mathrel{|\!\sim}_{\mathsf{IPC},\mathcal{G}}$.*

Proof

Suppose $A \mathrel{|\!\!\sim}_{T,\mathcal{F}} B$ and $\vdash gA$. Consider any $f \in \mathcal{F}$. Clearly $\vdash_T fgA$, i.e. $\vdash_T (f \star g)A$. Since $(f \star g) \in \mathcal{F}$, we find: $\vdash_T (f \star g)B$. Hence, for all $f \in \mathcal{F}$, $\vdash_T fgB$ and, hence, $\vdash gB$. ❑

Note that it follows that $\mathrel{|\!\!\sim}_{\mathsf{IPC}}$ is maximal among the $\mathrel{|\!\!\sim}_T$ with $\Lambda_T = \mathsf{IPC}$.

4.4 A Survey of Results on Preservativity

In this section we give an overview of results concerning preservativity.

4.4.1 Derivability for IPC

The minimal preservativity relation over IPC is $\vdash_{\mathsf{IPC}}$. There is an extension HA^* of HA (see in subsection 8.2), such that $\vdash_{\mathsf{IPC}} = \mathrel{|\!\!\sim}_{\mathsf{HA}^*} = \mathrel{|\!\!\sim}_{\mathsf{HA}^*,\Sigma}$. This result was proved by de Jongh and Visser in their paper [dJV96], adapting a result of Shavrukov (see [Sha93] or [Zam94]). Note that by de Jongh's theorem is also follows that: $\rhd_{\mathsf{HA},\mathcal{A},\mathcal{A}} = \rhd_{\mathsf{HA},\mathcal{A},\Sigma} = \vdash_{\mathsf{IPC}}$.

4.4.2 Admissible Consequence for IPC

$\mathrel{|\!\!\sim}_{\mathsf{IPC}}$ is the maximal relation $\mathrel{|\!\!\sim}_T$ for theories T with $\Lambda_T = \mathsf{IPC}$. This was shown in subsection 4.3. $\mathrel{|\!\!\sim}_{\mathsf{IPC}}$ strictly extends $\vdash_{\mathsf{IPC}}$. E.g. the independence of premiss rule

$$\mathsf{IP} \quad \neg p \to (q \vee r) \mathrel{|\!\!\sim}_{\mathsf{IPC}} (\neg p \to q) \vee (\neg p \to r)$$

is admissible but not derivable. We have the following facts.

1. $\mathrel{|\!\!\sim}_{\mathsf{IPC}}$ is decidable. This result is due to Rybakov. See the paper [Ryb92] or the book [Ryb97]. The result follows also, along a different route, from the results in [Ghi99] or, alternatively, from the results in [Iem99].

2. The embedding of $\vdash_{\mathsf{IPC}}$ into $\mathrel{|\!\!\sim}_{\mathsf{IPC}}$ has a left adjoint, say $(\cdot)^{\triangledown}$. So we have:

$$A \mathrel{|\!\!\sim}_{\mathsf{IPC}} B \Leftrightarrow A^{\triangledown} \vdash_{\mathsf{IPC}} B.$$

   This result is due to Ghilardi. See his paper [Ghi99]. $(\cdot)^{\triangledown}$ does not commute with conjunction.

3. The following semi-consequence relations are identical:

$$\mathrel{|\!\!\sim}_{\mathsf{IPC}}, \quad \mathrel{|\!\!\sim}_{\mathsf{HA}}, \quad \mathrel{|\!\!\sim}_{\mathsf{HA},B\Sigma}, \quad \rhd_{\mathsf{IPC,EX}}, \quad \blacktriangleright_{\alpha}.$$

   We briefly discuss the individual identities.

   - $\mathrel{|\!\!\sim}_{\mathsf{IPC}} = \mathrel{|\!\!\sim}_{\mathsf{HA}}$. This result is due to Visser. See his paper [Vis99]. The result can be obtained in a different way via the results of [Iem99] and the results of this paper. We will give the argument in section 11.

- $\mathrel{|\!\sim}_{\mathsf{IPC}} = \mathrel{|\!\sim}_{\mathsf{HA},B\Sigma}$. This result is due to de Jongh and Visser. See their paper [dJV96], corollary 6.6.
- $\mathrel{|\!\sim}_{\mathsf{IPC}} = \rhd_{\mathsf{IPC,EX}}$. This result is immediate from the results in de Jongh and Visser's paper [dJV96]. Here EX is the set of exact formulas. See e.g. [dJV96] for a definition.

  Ghilardi's result mentioned above in (2) strengthens this theorem, since, as it turns out, the range of $(\cdot)^{\nabla}$ is precisely $\mathsf{disj}(\mathsf{EX})$.
- $\mathrel{|\!\sim}_{\mathsf{IPC}} = \blacktriangleright_{\alpha}$. This result was obtained by Iemhoff. See her paper [Iem99].

4.4.3 Admissible Rules for Σ-Substitutions over HA

Obviously $\mathrel{|\!\sim}_{\mathsf{HA},\Sigma}$ extends $\mathrel{|\!\sim}_{\mathsf{HA}}$, and thus, ipso facto, $\mathrel{|\!\sim}_{\mathsf{IPC}}$. It follows from results to be proved below that $\neg\neg p \to p \mathrel{|\!\sim}_{\mathsf{HA},\Sigma} p \vee \neg p$. However, we do *not* have $\neg\neg p \to p \mathrel{|\!\sim}_{\mathsf{IPC}} p \vee \neg p$, as witnessed by the substitution $p \mapsto \neg q$.

We will connect $\mathrel{|\!\sim}_{\mathsf{HA},\Sigma}$, i.e. HA-admissible consequence for Σ-substitutions, with the formula classes NNIL, ROB and $\mathsf{f\text{-}ROB}$. NNIL is introduced in subsection 2. ROB and $\mathsf{f\text{-}ROB}$ are defined in section 6 below. We show that $\mathsf{NNIL} = \mathsf{ROB} = \mathsf{f\text{-}ROB}$ (mod IPC). The result was first proved in my unpublished paper [Vis85]. We will present a version of the proof in section 7 of this paper. A different proof, due to Johan van Benthem, is contained in [vB95]. A version of van Benthem's proof is presented in [VvBdJdL95].

We will further prove the following results:

1. $\mathrel{|\!\sim}_{\mathsf{HA},\Sigma} = \rhd_{\mathsf{HA},\Sigma,\Sigma} = \rhd_{\mathsf{IPC,NNIL}} = \blacktriangleright_{\sigma}$.
2. The identical embedding of $\vdash_{\mathsf{IPC}}$ into $\mathrel{|\!\sim}_{\mathsf{HA},\Sigma}$ has a left adjoint $(\cdot)^*$. I.o.w. we have $A \mathrel{|\!\sim}_{\mathsf{HA},\Sigma} B \Leftrightarrow A^* \vdash_{\mathsf{IPC}} B$.

These results were essentially already contained in [Vis85].

5 Applications of Kripke Semantics

To prepare ourselves for section 6 we briefly state some basic facts about Kripke models.

5.1 Some Basic Definitions

We suppose that the reader is familiar with Kripke models for IPC. Two good introductions are [TvD88a] and [Smo73]. Our treatment here is mainly intended to fix notations. A model is a structure $\mathbb{K} = \langle K, \leq, \Vdash, \mathcal{P}\rangle$. Here K is a non-empty set of nodes, $\leq$ is a partial ordering and $\Vdash$ is a relation between nodes and propositional atoms in $\mathcal{P}$, satisfying persistence: if $k \leq k'$ and $k \Vdash p$, then $k' \Vdash p$. We call $\Vdash$ the *forcing relation* of $\mathbb{K}$. We use e.g. $\mathcal{P}_{\mathbb{K}}$ for *the set of propositional variables of* $\mathbb{K}$, $\Vdash_{\mathbb{K}}$ for *the forcing relation of* $\mathbb{K}$, etcetera. A model $\mathbb{K}$ is a $\mathcal{P}$-model if $\mathcal{P}_{\mathbb{K}} = \mathcal{P}$.

Consider $\mathbb{K} = \langle K, \leq, \Vdash, \mathcal{P}\rangle$. We define $k \Vdash A$ for $A \in \mathcal{L}(\mathcal{P})$ in the standard way. $\mathbb{K} \Vdash A$ if, for all $k \in K$, we have $k \Vdash A$.

A model $\mathbb{K}$ is *finite* if both $K_{\mathbb{K}}$ and $\mathcal{P}_{\mathbb{K}}$ are finite. A *rooted model* $\mathbb{K}$ is a structure $\langle K, b, \leq, \Vdash, \mathcal{P}\rangle$, where $\langle K, \leq, \Vdash, \mathcal{P}\rangle$ is a model and where $b \in K$ is the bottom element w.r.t. $\leq$.

For any $k \in K$, we define $\mathbb{K}[k]$ as the model $\langle K', k, \leq', \Vdash', \mathcal{P}\rangle$, where $K' := \uparrow k := \{k' \mid k \leq k'\}$ and where $\leq'$ and $\Vdash'$ are the restrictions of $\leq$ respectively $\Vdash$ to K'. (We will often simply write $\leq$ and $\Vdash$ for $\leq'$ and $\Vdash'$.)

5.2 The Henkin Construction

A set $X \subseteq \mathcal{L}(\mathcal{P})$ is *adequate* if it is non-empty and closed under subformulas (and, hence, contains $\bot$). A set Γ is *X-saturated* if:

1. $\Gamma \subseteq X$,
2. $\Gamma \nvdash \bot$,
3. If $\Gamma \vdash A$ and $A \in X$, then $A \in \Gamma$,
4. If $\Gamma \vdash (B \vee C)$ and $(B \vee C) \in X$, then $B \in \Gamma$ or $C \in \Gamma$.

Let X be adequate. The Henkin model $\mathbb{H}_X(\mathcal{P})$ is the $\mathcal{P}$-model with as nodes the X-saturated sets Δ and as accessibily relation $\subseteq$. The atomic forcing in the nodes is given by: $\Gamma \Vdash p \Leftrightarrow p \in \Gamma$. We have, by a standard argument, that $\Gamma \Vdash A \Leftrightarrow A \in \Gamma$, for any $A \in X$. Note that if X and $\mathcal{P}$ are finite, then $\mathbb{H}_X(\mathcal{P})$ is finite. A direct consequence of the Henkin construction is the Kripke Completeness Theorem. Let $\vec{p} \supseteq \mathsf{pv}(A)$, then:

$$\mathsf{IPC} \vdash A \quad \Leftrightarrow \quad \mathbb{K} \Vdash A, \text{ for all (finite) } \vec{p}\text{-models } \mathbb{K}.$$

5.3 More Definitions

Let $\mathcal{K}$ be a set of $\mathcal{P}$-models. $M\mathcal{K}$ is the $\mathcal{P}$-model with nodes $\langle k, \mathbb{K}\rangle$ for $k \in K_{\mathbb{K}}$, $\mathbb{K} \in \mathcal{K}$ and with ordering $\langle k, \mathbb{K}\rangle \leq \langle m, \mathbb{M}\rangle :\Leftrightarrow \mathbb{K} = \mathbb{M}$ and $k \leq_{\mathbb{K}} m$. As atomic forcing we define: $\langle k, \mathbb{K}\rangle \Vdash p :\Leftrightarrow k \Vdash_{\mathbb{K}} p$. In practice we will forget the second components of the new nodes, pretending the domains to be already disjoint.

Let $\mathbb{K}$ be a $\mathcal{P}$-model. $B\mathbb{K}$ is the rooted $\mathcal{P}$-model obtained by adding a new bottom b to $\mathbb{K}$ and by taking $b \Vdash p :\Leftrightarrow \mathbb{K} \Vdash p$. We put $\mathsf{glue}(\mathcal{K}) := BM\mathcal{K}$.

5.4 The Push Down Lemma

We will assume below that $\mathcal{P}$ is fixed. We will often notationally suppress it.

Theorem 5.1 (Push Down Lemma) *Let X be adequate. Suppose Δ is X-saturated and $\mathbb{K} \Vdash \Delta$. Then $\mathsf{glue}(\mathbb{H}_X[\Delta], \mathbb{K}) \Vdash \Delta$.*

Proof

We show by induction on $A \in X$ that $b \Vdash A \Leftrightarrow A \in \Delta$. The cases of atoms, conjunction and disjunction are trivial. If $(B \to C) \in X$ and $b \Vdash (B \to C)$, then $\Delta \Vdash (B \to C)$ and, hence, $(B \to C) \in \Delta$. Conversely, suppose $(B \to C) \in \Delta$. If $b \nVdash B$, we easily done. If $b \Vdash B$, then, by the Induction Hypothesis, $B \in \Delta$, hence $C \in \Delta$ and, again by the Induction Hypothesis: $b \Vdash C$. ❑

We say that Δ is *($\mathcal{P}$-)prime* if it is consistent and if, for every $(C \vee D) \in \mathcal{L}(\mathcal{P})$, $\Delta \vdash (C \vee D) \Rightarrow \Delta \vdash C$ or $\Delta \vdash D$. A formula A is prime if $\{A\}$ is prime.

Theorem 5.2 *Suppose X is adequate and Δ is X-saturated. Then Δ is prime.*

Proof

Δ is consistent by definition. Suppose $\Delta \vdash C \vee D$ and $\Delta \nvdash C$ and $\Delta \nvdash D$. Suppose further $\mathbb{K} \Vdash \Delta$, $\mathbb{K} \nVdash C$, $\mathbb{M} \Vdash \Delta$ and $\mathbb{M} \nVdash D$. Consider $\mathsf{glue}(\mathbb{H}_X(\Delta), \mathbb{K}, \mathbb{M})$. By theorem 5.1 we have: $b \Vdash \Delta$. On the other hand, by persistence, $b \nVdash C$ and $b \nVdash D$. Contradiction. ❑

$\sim C$ exhibited next to a node means that C *is not forced*; this is not to be confused with $\neg C$ exhibited next to a node, which means that $\neg C$ *is forced.*

Theorem 5.3 *Consider any formula A. The formula A can be written (modulo* IPC*-provable equivalence) as a disjunction of prime formulas C. Moreover these C are conjunctions of implications and propositional variables in* $\mathsf{sub}(A)$*.*

Proof

Consider a $\mathsf{sub}(A)$-saturated Δ. Let $\mathsf{ip}(\Delta)$ be the set of implications and atoms of Δ. It is easily seen that $\mathsf{IPC} \vdash \bigwedge \mathsf{ip}(\Delta) \leftrightarrow \bigwedge \Delta$. Take:

$$D := \bigvee \{ \bigwedge \mathsf{ip}(\Delta) \mid \Delta \text{ is } \mathsf{sub}(A)\text{-saturated and } A \in \Delta \}.$$

Trivially: $\mathsf{IPC} \vdash D \to A$. On the other hand if $\mathsf{IPC} \nvdash A \to D$, then by a standard construction there is a $\mathsf{sub}(A)$-saturated set Γ such that $A \in \Gamma$ and $\Gamma \nvdash D$. Quod non. ❑

Remark 5.4 Note that in the definition of D in 5.3, we can restrict ourselves to the $\subseteq$-minimal $\mathsf{sub}(A)$-saturated Δ with $A \in \Delta$. ❑

6 Robust Formulas

In this section we study robust formulas. We will verify that $\rhd_{\mathsf{IPC,ROB}}$ is a σ-relation, and thus that $\blacktriangleright_\sigma \subseteq \rhd_{\mathsf{IPC,ROB}}$.

We aim at application of lemma 4.3 to $\blacktriangleright_\sigma$ in the role of $\rhd$, and ROB in the role of X. The mapping Ψ needed for the application of lemma 4.3, will be the mapping $(\cdot)^*$ provided in section 7.

Consider $\mathcal{P}$-models $\mathbb{K}$ and $\mathbb{M}$. We say that:

- $\mathbb{K} \subseteq \mathbb{M} :\Leftrightarrow \exists f : K_\mathbb{K} \to K_\mathbb{M}$ (f is injective and $f \circ \leq_\mathbb{K} \subseteq \leq_\mathbb{M} \circ f$ and $\forall k{\in}K_\mathbb{K}, p{\in}\mathcal{P}$ $(k \Vdash_\mathbb{K} p \Leftrightarrow f(k) \Vdash_\mathbb{M} p)$).

Modulo the identification of the elements of K with their f-images in M, clearly '$\mathbb{K} \subseteq \mathbb{M}$' means that $\mathbb{K}$ is a submodel of $\mathbb{M}$. ROB is the class of formulas that is preserved under submodels.

A submodel is *full* if the new ordering relation is the restriction of the old ordering relation to the new worlds. We do not demand of our submodels that they are *full*. However, all or results work for full submodels too.

- $A \in \mathsf{ROB} :\Leftrightarrow A$ is robust $:\Leftrightarrow \forall \mathbb{M}\ (\mathbb{M} \Vdash A \Rightarrow \forall \mathbb{K}{\subseteq}\mathbb{M}\ \mathbb{K} \Vdash A)$.

We will let σ, σ', τ, ... range over ROB. It is easy to see that $\mathsf{NNIL} \subseteq \mathsf{ROB}$. In section 7 we will see that modulo IPC-provable equivalence each robust formula is in NNIL.

Let's take as a local convention of this section that $\rhd := \rhd_{\mathsf{IPC,ROB}}$. Clearly we have that $\mathcal{P} \subseteq \mathsf{ROB}$ and that ROB is closed under conjunction. So it follows that $\rhd$ satisfies A1, A2, A3, B3. To verify B1, by lemma 4.2(2), the following theorem suffices.

Theorem 6.1 ROB *is weakly disjunctive.*

Proof

Consider any $A \in \mathsf{ROB}$. We write A in disjunctive normal form D_A as in theorem 5.3 and remark 5.4. Consider any disjunct $C(\Delta)$ of D_A. Here, as in 5.4, Δ is a $\subseteq$-minimal $\mathsf{sub}(A)$-saturated set with $A \in \Delta$ and $C(\Delta)$ is the conjunction of the atoms and implications in Δ. We claim that $C(\Delta)$ is robust. Consider any models $\mathbb{K} \subseteq \mathbb{M} \Vdash C(\Delta)$. Trivially, $\mathbb{M} \Vdash \Delta$. By the Push Down Lemma 5.1, $\mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}[\Delta], \mathbb{M}) \Vdash \Delta$. Hence, $\mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}[\Delta], \mathbb{M}) \Vdash A$. Now clearly,

$$\mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}(\Delta), \mathbb{K}) \subseteq \mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}(\Delta), \mathbb{M}).$$

By the stability of A we get $\mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}(\Delta), \mathbb{K}) \Vdash A$. Consider

$$\Gamma := \{G \in \mathsf{sub}(A) \mid \mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}(\Delta), \mathbb{K}) \Vdash G\}.$$

Clearly $\Gamma \subseteq \Delta$. Moreover, Γ is $\mathsf{sub}(A)$-saturated and $A \in \Gamma$. By the $\subseteq$-minimality of Δ we find $\Gamma = \Delta$. Hence, $\mathsf{glue}(\mathbb{H}_{\mathsf{sub}(A)}(\Delta), \mathbb{K}) \Vdash C(\Delta)$ and so $\mathbb{K} \Vdash C(\Delta)$. Ergo $C(\Delta)$ is robust. ❑

Theorem 6.2 $\rhd$ *is closed under B2.*

Proof

Let X be a finite set of implications and let $Y := \{C \mid (C \to D) \in X\} \cup \{B\}$. Let $A := \bigwedge X$. We have to show: $(A \to B) \rhd [A]Y$. The proof is by contraposition. Consider any $H \in \mathsf{ROB}$ and suppose: $H \nvdash [A]Y$. Let $\mathbb{K} = \langle K, b, \leq, \Vdash, \mathcal{P}\rangle$ be a rooted model such that $\mathbb{K} \Vdash H$ and $\mathbb{K} \nVdash \bigvee [A]Y$, i.e. for all $E \in Y$, we have $\mathbb{K} \nVdash [A]E$.

Let $\mathbb{K}' := \mathbb{K}\{A\} := \langle K', b, \leq', \Vdash', \mathcal{P}\rangle$, be the full submodel of $\mathbb{K}$ on $K' := \{b\} \cup \{k' \in K \mid k' \Vdash A\}$. Note that $\{k' \in K \mid k' \Vdash A\}$ is upwards closed and that on $\{k' \in K \mid k' \Vdash A\}$ the old and the new forcing coincide. Moreover, on this class, for any G, we have that $[A]G$ is equivalent with G.

We claim that, for all F, $b \Vdash' F \Rightarrow b \Vdash [A]F$. The proof of the claim is by a simple induction on F. The cases of atoms, disjunction and conjunction are trivial. Suppose F is an implication and $b \Vdash' F$. Then certainly, for all $k \in K$, $(k \Vdash A \Rightarrow k \Vdash' F)$. Since on the k with $k \Vdash A$, $\Vdash$ and $\Vdash'$ coincide, we find $b \Vdash (A \to F)$, i.e. $b \Vdash [A]F$.

We return to the main proof. Remember that: $b \Vdash H$ and $b \nVdash [A]C$, for all C with $(C \to D) \in X$ and $b \nVdash [A]B$. We show that $b \Vdash' H$ and $b \Vdash' A$ and $b \nVdash' B$.

It is immediate that $b \Vdash' H$, since $\mathbb{K}'$ is a submodel of $\mathbb{K}$ and H is robust.

Remember that A is the conjunction of the $(C \to D)$ in X. So it is sufficient to show that, for each $(C \to D)$ in X, $b \Vdash' (C \to D)$. Consider any $(C \to D) \in X$ and any $k' \geq' b$ with $k' \Vdash' C$. Since $b \nVdash [A]C$, we have, by the claim, $b \nVdash' C$. So $k' \neq b$. But then $k' \Vdash A$, hence $k' \Vdash' A$ and, thus, $k' \Vdash' (C \to D)$. We may conclude that $k' \Vdash' D$ and, hence, $b \Vdash' (C \to D)$.

From $b \nVdash [A]B$ and the claim we have immediately that: $b \nVdash' B$. ❑

All the proofs in this section also work when we replace ROB by $\mathsf{f\text{-}ROB}$, the set of formulas preserved by full submodels. Note that $\mathsf{ROB} \subseteq \mathsf{f\text{-}ROB}$. Our result in section 7 will imply: $\mathsf{ROB} = \mathsf{f\text{-}ROB} = \mathsf{NNIL}$ (modulo IPC-provable equivalence).

7 The NNIL-Algorithm

In this section we produce the algorithm that is the main tool of this paper. The existence of the algorithm establishes the following theorem.

Theorem 7.1 *For all A there is an $A^* \in \mathsf{NNIL}(\mathsf{pv}(A))$ such that $A \blacktriangleright_\sigma A^*$ and $A^* \vdash A$.*

Before proceeding with the proof of 7.1 we interpolate a corollary.

Corollary 7.2

1. Let A and A^* be as promised by 7.1. Then we have

$$A \blacktriangleright_\sigma B \Leftrightarrow A \rhd_{\mathsf{IPC},\mathsf{ROB}} B \Leftrightarrow A^* \vdash_{\mathsf{IPC}} B.$$

2. $\mathsf{NNIL} = \mathsf{ROB} = \mathsf{f\text{-}ROB}$ (mod IPC)

❑

Proof

It is immediate from the fact that both $\rhd_{\mathsf{IPC,ROB}}$ and $\rhd_{\mathsf{IPC,f\text{-}ROB}}$ are σ-relations that $\blacktriangleright_\sigma \subseteq \rhd_{\mathsf{IPC,ROB}}$ and $\blacktriangleright_\sigma \subseteq \rhd_{\mathsf{IPC,f\text{-}ROB}}$. Moreover we have $\mathsf{range}((\cdot)^*) \subseteq \mathsf{NNIL} \subseteq \mathsf{ROB} \subseteq \mathsf{f\text{-}ROB}$. Combining this with the result of theorem 7.1, we can apply lemma 4.3 to obtain the desired results. ❑

7.2(2) was proved by purely model theoretical means by Johan van Benthem. See his [vB95] or, alternatively, [VvBdJdL95]. The advantage of van Benthem's proof is its relative simplicity and the fact that the method employed easily generalizes. The advantage of the present method is the extra information it produces, like the axiomatization of $\rhd_{\mathsf{IPC,ROB}}$ and its usefulness in the arithmetical case, see sections 9, 10, 11 and 12. It is an open question whether van Benthem's proof can be adapted to the arithmetical case.

Remark 7.3 Combining corollary 7.2 with the results of [VvBdJdL95], we obtain the following semantical characterization of $\blacktriangleright_\sigma$.

$$A \blacktriangleright_\sigma B \Leftrightarrow \forall \mathbb{M}\ (\forall \mathbb{K} \subseteq \mathbb{M}\ \ \mathbb{K} \Vdash A \Rightarrow \mathbb{M} \Vdash B).$$

❑

We proceed with the proof of theorem 7.1. For the rest of this section we write $\rhd := \blacktriangleright_\sigma$. For convenience we reproduce the axioms and rules of $\rhd$ here.

A1 $\quad A \vdash B \Rightarrow A \rhd B$

A2 $\quad A \rhd B$ and $B \rhd C \Rightarrow A \rhd C$

A3 $\quad C \rhd A$ and $C \rhd B \Rightarrow C \rhd (A \wedge B)$

B1 $\quad A \rhd C$ and $B \rhd C \Rightarrow (A \vee B) \rhd C$

B2 $\quad$ Let X be a finite set of implications and let

$$Y := \{C \mid (C \to D) \in X\} \cup \{B\}.$$

Take $A := \bigwedge X$. Then, $(A \to B) \rhd [A]Y$

B3 $\quad A \rhd B \Rightarrow (p \to A) \rhd (p \to B)$

Proof of theorem 7.1

We introduce an ordinal measure $\mathfrak{o}$ of complexity on formulas as follows:

- $I(D) := \{E \in \mathsf{sub}(D) \mid E$ is an implication$\}$
- $\mathfrak{i}D := \max\{|I(E)| \mid E \in I(D)\}$
  (Here $|Z|$ is the cardinality of Z)
- $\mathfrak{c}D :=$ the number of occurrences of logical connectives in D

- $\mathfrak{o}D := \omega \cdot \mathfrak{i}D + \mathfrak{c}D$

Note that we count *occurrences* of connectives for $\mathfrak{c}$ and *types* of implications, not occurrences, for $\mathfrak{i}$. We say that an occurence of E in D is an *outer occurrence* if this occurrence is not in the scope of an implication.

The proof of theorem 7.1 proceeds as follows. At every stage we have a formula and associated with it certain designated subformulas, that still have to be simplified. A step operates on one of these subformulas, say A. Either A will lose its designation and a number of its subformulas will get designated or A will be modified, say to A', and a number of subformulas of A' will replace A among the designated formulas. The subformulas B replacing A among the designated formulas will satisfy: $\mathfrak{o}B < \mathfrak{o}A$. Thus, to every stage we can associate a multiset of ordinals below ω^2, viz. the multiset of $\mathfrak{o}$-complexities of the designated formulas. A step will reduce one of the ordinals to an number of strictly smaller ones. This means that we are doing a recursion over ω^{ω^2}.

We will exhibit the properties of $\rhd$ that are used in each step between square brackets.

α **Atoms:** [A1] Suppose A is an atom. Take $A^* := A$.

β **Conjunction:** [A1,A2,A3]: Suppose $A = (B \wedge C)$. Clearly $\mathfrak{o}B < \mathfrak{o}A$ and $\mathfrak{o}C < \mathfrak{o}A$. Take $A^* := B^* \wedge C^*$. Clearly $A^* \in \mathsf{NNIL}(\mathsf{pv}(A))$. The verification of the desired properties is trivial.

γ **Disjunction:** [A1,A2,B1] Suppose $A = (B \vee C)$. Clearly $\mathfrak{o}B < \mathfrak{o}A$ and $\mathfrak{o}C < \mathfrak{o}A$. Take $A^* := B^* \vee C^*$. Clearly $A^* \in \mathsf{NNIL}$. The verification of the desired properties is trivial.

δ **Implication:** [A1,A2,A3,B2,B3] Suppose $A = (B \to C)$. We split the step into several cases.

$\delta 1$ **Outer conjunction in the consequent:** [A1,A2,A3] Suppose C has an outer ocurrence of a formula $D \wedge E$. Pick any $J(q)$ be such that:

- q is a fresh variable,
- q occurs precisely once in J,
- q is not in the scope of an implication in J,
- $C = J[q := (D \wedge E)]$.

Let $C_1 := J[q := D]$, $C_2 := J[q := E]$. As is easily seen C is IPC-provably equivalent to $C_1 \wedge C_2$. Let $A_1 := (B \to C_1)$ and $A_2 := (B \to C_2)$. Clearly A is IPC-provably equivalent to $A_1 \wedge A_2$. We prove that $\mathfrak{o}A_i < \mathfrak{o}A$ for $i = 1, 2$. Since it is clear that $\mathfrak{c}A_i < \mathfrak{c}A$, it is sufficient to show that $\mathfrak{i}A_i \leq \mathfrak{i}A$. Since A and the A_i are implications we have to show that $|I(A_i)| \leq |I(A)|$. We treat the case that $i = 1$. It is sufficient to construct an injective mapping from $I(A_1)$ to $I(A)$.

Consider any implication F in $I(A_1)$. If $F = A_1$, we map F to A. Otherwise $F \in I(B)$ or $F \in I(J)$ or $F \in I(D)$ (since q does not occur in the scope of an implication). In all three cases we can map F to itself. Since A_1 cannot be in $I(B)$ or $I(J)$ or $I(D)$, our mapping is injective. The case that $i = 2$ is similar. Set $A^* := (A_1^* \wedge A_2^*)$.

$\delta 2$ **Outer disjunction in the antecedent:** [A1,A2,A3] This case is completely analogous to the previous one.

If A has no outer disjunction in the antecedent and no outer conjunction in the consequent, then B is a conjunction of atoms and implications and C is a disjunction of atoms and implications. It is easy to see, that applications of associativity, commutativity and idempotency to the conjunction in the antecedent or to the disjunction in the consequent do not raise $\mathfrak{o}$. So we can safely write: $B = \bigwedge X$ and $C = \bigvee Y$, where X and Y are finite sets of atoms or implications. This leads us to the following case.

$\delta 3$ B **is a conjunction of atoms and implications,** C **is a disjunction of atoms and implications** [A1,A2,A3,B2,B3]

$\delta 3.1$ X **contains an atom:** [A1,A2,B3]

$\delta 3.1.1$ X **contains a propositional variable, say** p**:** [A1,A2,B3] Consider $D := \bigwedge(X \setminus \{p\})$ and $E := (D \to C)$. Clearly $\vdash A \leftrightarrow (p \to E)$ and $\mathfrak{o}E < \mathfrak{o}A$. Put $A^* := (p \to E^*)$. Evidently $(p \to E^*)$ is in $\mathsf{NNIL}(\mathsf{pv}(A))$. We have $E^* \vdash E$ by the Induction Hypothesis and, hence, $(p \to E^*) \vdash (p \to E) \vdash A$. We have $E \rhd E^*$ by the Induction Hypothesis. It follows by B3 that $(p \to E) \rhd (p \to E^*)$. From $A \vdash (p \to E)$, we have, by A1, $A \rhd (p \to E)$. Hence, by A2, $A \rhd (p \to E^*)$.

$\delta 3.1.2$ X **contains** $\top$**:** [A1,A2] Left to the reader.

$\delta 3.1.3$ X **contains** $\bot$**:** [A1] Left to the reader.

$\delta 3.2$ X **contains no atoms:** [A1,A2,A3,B2] This case —the last one— is the truly difficult one. To motivate its treatment, let's solve the difficulties one by one. We first look at an example.

Example 7.4 Consider $(p \to q) \to r$. B2 gives us $((p \to q) \to r) \rhd (p \vee r)$. However, we do not have: $(p \vee r) \vdash ((p \to q) \to r)$. We can repair this by noting that $((p \to q) \to r) \vdash (q \to r)$ and $(q \to r) \wedge (p \vee r) \vdash ((p \to q) \to r)$. So the full solution of our example is as follows.

We have $((p \to q) \to r) \rhd ((q \to r) \wedge (p \vee r))$, by:

$a)$	$((p \to q) \to r)$	$\rhd$	$(p \vee r)$	$B2$
$b)$	$((p \to q) \to r)$	$\vdash$	$(q \to r)$	IPC
$c)$	$((p \to q) \to r)$	$\rhd$	$(q \to r)$	$A1$
$d)$	$((p \to q) \to r)$	$\rhd$	$((q \to r) \wedge (p \vee r))$	$a, c, A3$

Moreover, $((q \to r) \wedge (p \vee r)) \vdash ((p \to q) \to r)$ and

$$((q \to r) \wedge (p \vee r)) \in \mathsf{NNIL}(\mathsf{pv}((p \to q) \to r)).$$

❑

We implement this idea for the general case. There will be a problem with $\mathfrak{o}$, but we will postpone its discussion until we run into it. A is of the form $B \to C$, where B is a conjunction of a finite set of implications X and C is a disjunction of a finite set of atoms or implications Y. For any $D := (E \to F) \in X$, let:

$$B{\downarrow}D := \bigwedge((X \setminus \{D\}) \cup \{F\}).$$

Clearly $\mathfrak{o}((B{\downarrow}D) \to C) < \mathfrak{o}A$.

Let $Z := \{E \mid (E \to F) \in X\} \cup \{C\}$. Put $A_0 := \bigvee[B]Z$. We will show that our problem reduces to the question whether A_0^* exists. So for the moment we pretend it does. We have:

1)	$A \rhd A_0$	$B2$
2)	$A_0 \rhd A_0^*$	assumption
3)	$A \rhd A_0^*$	$1, 2, A2$
4)	$\forall D{\in}X\ A \vdash ((B{\downarrow}D) \to C)$	IPC
5)	$\forall D{\in}X\ A \rhd ((B{\downarrow}D) \to C)$	$4, A1$
6)	$\forall D{\in}X\ ((B{\downarrow}D) \to C) \rhd ((B{\downarrow}D) \to C)^*$	IH
7)	$\forall D{\in}X\ A \rhd ((B{\downarrow}D) \to C)^*$	$5, 6, A2$
8)	$A \rhd (\bigwedge\{((B{\downarrow}D) \to C)^* \mid D \in X\} \wedge A_0^*)$	$3, 7, A3$

It is clear that $(\bigwedge\{((B{\downarrow}D) \to C)^* \mid D \in X\} \wedge A_0^*) \in \mathsf{NNIL}(\mathsf{pv}(A))$. We show that:

$$\bigwedge\{((B{\downarrow}D) \to C)^* \mid D \in X\} \wedge A_0^* \vdash A.$$

It is sufficient to show: $\bigwedge\{((B{\downarrow}D) \to C) \mid D \in X\} \wedge [B]E \vdash A$, for each $E \in Z$. In case $E = C$, we are immediately done by: $[B]C \vdash B \to C$. Suppose $(E \to F) \in X$ for some F. Reason in IPC.

Suppose $\bigwedge\{((B{\downarrow}D) \to C) \mid D \in X\}$, $[B]E$ and B. We want to derive C. We have $((\bigwedge(X \setminus \{E \to F\}) \wedge F) \to C)$, $[B]E$ and B. From B we find $\bigwedge(X \setminus \{E \to F\})$ and $(E \to F)$. From B and $[B]E$, we derive E. From E and $(E \to F)$ we get F. Finally we infer from $((\bigwedge(X \setminus \{E \to F\}) \wedge F) \to C)$, $\bigwedge(X \setminus \{E \to F\})$ and F the desired conclusion C. Here ends our discussion inside IPC.

So the only thing left to do is to show that A_0^* exists. If $\mathfrak{i}A_0 = 0$, we are easily done. Suppose $\mathfrak{i}A_0 > 0$. If for each E in Z with $\mathfrak{i}E > 0$, we would have $\mathfrak{i}([B]E) < \mathfrak{i}A$, it would follow that $\mathfrak{i}A_0 < \mathfrak{i}A$. Hence we would be done by the Induction Hypothesis.

We study some examples. These examples show that we cannot generally hope to get $\mathfrak{i}([B]E) < \mathfrak{i}A$. The examples will, however, suggest a way around the problem: we produce a logical equivalent, say Q, of $[B]E$, such that $\mathfrak{i}Q < \mathfrak{i}A$. So we replace A_0 by the disjunction R of the Q's, which is logically equivalent and has $\mathfrak{i}R < \mathfrak{i}A_0$. Put $A_0^* := R^*$.

Example 7.5 Consider $(p \to q) \to (r \to s)$. Take $E := C = (r \to s)$. We have $[B]E = [p \to q](r \to s) = (p \to q) \to (r \to s)$. So no simplification is obtained. However $[B]E$ is IPC-provably equivalent to $((r \wedge (p \to q)) \to s)$, which has lower $\mathfrak{i}$. By A1,A2 we can put $([B]E)^* := ((r \wedge (p \to q)) \to s)^*$. We could have applied the reduction before going to [B]E. We did not choose to do so for reasons of uniformity of treatment. ❑

Example 7.6 Consider $(((p \to q) \to r) \to s)$. Take $E := (p \to q)$. We find $[B]E = [(p \to q) \to r](p \to q) = (((p \to q) \to r) \to (p \to q))$. $[B]E$ is IPC-provably equivalent to $((p \wedge (q \to r)) \to q)$, which has lower $\mathfrak{i}$. ❑

Consider $[B]E$ for $E \in Z = \{E \mid (E \to F) \in X\} \cup \{C\}$. $[B]E$ is the result of replacing outer implications $(G \to H)$ of E by $(B \to (G \to H))$. If there is no outer implication in E, we find that $[B]E = E$ and $\mathfrak{i}([B]E) = 0$. In this case $[B]E$ is implication free and hence a fortiori in NNIL. We can put $([B]E)^* := E$. Suppose E has an outer implication.

Consider any outer implication $(G \to H)$ of E. We first show: $\mathfrak{i}(B \to (G \to H)) \leq \mathfrak{i}(B \to C)$. We define an injection from $I(B \to (G \to H))$ to $I(B \to C)$. Any implication occurring in $I(B)$ or $I(G \to H)$ can be mapped to itself in $I(B \to C)$. None of these implications has as image $(B \to C)$, since $(G \to H) \in I(B) \cup I(C)$. So we can send $(B \to (G \to H))$ to $(B \to C)$.

The next step is to replace $(B \to (G \to H))$ by an IPC-provably equivalent formula K with lower value of $\mathfrak{i}$. Let B' be the result of replacing all occurrences of $(G \to H)$ in B by H and let $K := ((G \wedge B') \to H)$. Clearly K is provably equivalent to $(B \to (G \to H))$. We show that $\mathfrak{i}(K) < \mathfrak{i}(B \to (G \to H))$.

We define a non-surjective injection from $I(K)$ to $I(B \to (G \to H))$. Consider an implication M in $I(K)$. If $M = K$ we send it to $(B \to (G \to H))$. Suppose $M \neq K$, i.e. $M \in I(B') \cup I(G) \cup I(H)$. A subformula N of $I(B) \cup I(G) \cup I(H)$ is a *predecessor* of M if M is the result of replacing all occurrences of $(G \to H)$ in N by H. Clearly M has at least one predecessor and any predecessor of M must be an implication. Send M to one of its shortest predecessors. Clearly our function is injective, since two different implications cannot share a predecessor. Finally $(G \to H)$ cannot be in the range of our injection. If it were, we would have $M = H$, but then H would be a shorter predecessor. A contradiction.

Replace every outer implication in $[B]E$ by an equivalent with lower value of $\mathfrak{i}$. The result is the desired Q.

Remark 7.7 The algorithm given with our proof is non-deterministically specified. However by corollary 7.2 the result is unique modulo provable equivalence. I didn't try to make the algorithm optimally efficient.

In [Vis85] a simple adaptation of the NNIL algorithm is given for $\rhd_{\mathsf{IPC},\top}$. Here the algorithm computes not a value in NNIL, but a value in $\{\top, \bot\}$. Thus we obtain an algorithm that checks for provability. It is easy to use the present algorithm for this purpose too, since there is a p-time algorithm to decide IPC-provability of NNIL_0 formulas, i.e. formulas with only propositional variables in front of arrows. The outputs of our algorithm are in NNIL_0. It has been shown by Richard Statman (see [Sta79]) that checking whether a formula is IPC-provable is p-space complete. This puts an absolute bound on what an algorithm like ours can do. ❑

Example 7.8 [Sample computations] We use some obvious short-cuts.

$$
\begin{array}{ll}
(\neg\neg p \to p) \to (p \vee \neg p) & \equiv \\
(p \to (p \vee \neg p)) \wedge (\, [\neg\neg p \to p]\neg\neg p \vee [\neg\neg p \to p]p \vee [\neg\neg p \to p]\neg p\,) & \equiv \\
\neg\neg p \vee p \vee \neg p & \equiv \\
(\,(\bot \to p) \wedge (\,[\neg p]p \vee [\neg p]\bot\,)\,) \vee p \vee \neg p & \equiv \\
p \vee \neg p &
\end{array}
$$

$$
\begin{array}{ll}
((p \to q) \to r) \to s & \equiv \\
(r \to s) \wedge (\,[(p \to q) \to r](p \to q) \vee [(p \to q) \to r]s\,) & \equiv \\
(r \to s) \wedge (\,((p \wedge (q \to r)) \to q) \vee s\,) & \equiv \\
(r \to s) \wedge (\,(p \to ((q \to r) \to q)) \vee s\,) & \equiv \\
(r \to s) \wedge (\,(p \to ((r \to q) \wedge (\,[q \to r]q \vee [q \to r]q))) \vee s\,) & \equiv \\
(r \to s) \wedge ((p \to q) \vee s). &
\end{array}
$$

❑

8 Basic Facts and Notations in Arithmetic

8.1 Arithmetical Theories

The arithmetical theories T considered in this paper are RE theories in $\mathcal{A}$, the language of HA. These theories all extend $i\text{-}\mathsf{S}^1_2$, the intuitionistic version of Buss' theory S^1_2. Another salient theory is intuitionistic elementary arithmetic, $i\text{-}\mathsf{EA} := i\text{-}I\Delta_0 + \mathsf{Exp}$, the constructive theory of Δ_0-induction with the axiom expressing that the exponentiation function is total. i-EA is finitely axiomatizable; we stipulate that a fixed finite axiomatization is employed. We will use $\Box_T$ for the formalization of provability in T. In the present paper, we are mostly interested in extensions of HA, so we wil not worry too much about details concerning weak theories.

Suppose A is a formula. $\Box_T A$ means $\mathsf{Prov}_T(t(\vec{x}))$, where Prov_T is the arithmetization of the provability predicate of T and where $\vec{x}$ is the sequence of variables occuring in A and $t(\vec{x})$ is the term 'the Gödelnumber of the result of substituting the Gödelnumbers of the numerals of the $\vec{x}$'s for the variables in A'. Suppose e.g. that 11, 12 and 15 code, respectively, '(', ')' and '=', that $*$ is the arithmetization of the syntactical operation of concatenation and that num is the arithmetization of the numeral function. Under these stipulations e.g. $\Box_T(x = x)$ means: $\mathsf{Prov}_T(\underline{11} * \mathsf{num}(x) * \underline{15} * \mathsf{num}(x) * \underline{12})$. [6]

[6] Our notational convention introduces a scope ambiguity. Fortunately by standard metamathematical results, we know that as long as the terms we employ stand for T-provably total recursive functions the different readings are provably equivalent. In this paper we will only employ terms for p-time computable functions, so the ambiguity is harmless.

Suppose T extends i-EA. We write T_n for the theory axiomatized by the finitely many axioms of i-EA, plus the axioms of T which are smaller than n in the standard Gödelnumbering. We write $\mathsf{Prov}_{T,n}$ for the formalization of provability in T_n. We consider $\mathsf{Prov}_{T,n}$ as a form of *restricted provability* in T. The following well known fact is quite important:

Theorem 8.1 *Suppose T is an RE extension of HA in the language of HA. Then T is essentially reflexive (verifiably in HA). I.e. for all n and, for all $\mathcal{A}$-formulas A with free variables $\vec{x}$, $T \vdash \forall \vec{x}\, (\Box_{T,n} A \to A)$. And, using UC for 'the universal closure of', even $\mathsf{HA} \vdash \forall x\, \forall A \in \mathsf{FOR}_{\mathcal{A}}\ \Box_T\, \mathsf{UC}\,(\Box_{T,x} A \to A)$.*

Proof

(Sketch) The proof is roughly as follows. Ordinary cut-elimination for constructive predicate logic (or normalization in case we have a natural deduction system) can be formalized in HA. Reason in HA. Let a number x and a formula A be given. Introduce a measure of complexity on arithmetical formulas that counts both the depth of quantifiers and of implications. Find y such that both the axioms of T_x and A have complexity $< y$. We can construct a truth predicate True_y for formulas of complexity $< y$. We have $\Box_T\, \mathsf{UC}\,(\mathsf{True}_y A \to A)$. Reason inside $\Box_T$. Suppose we have $\Box_{T,x} A$. By cut elimination we can find a T_x-proof p of A in which only formulas of complexity $< y$ occur. We now prove by induction on the subproofs of p, that all subconclusions of p are True_y. So A is True_y. Hence we obtain A. ❑

8.2 A Brief Introduction to HA^*

In this subsection we describe the theory HA^*. This theory will be employed in the proof of theorem 10.2. In section 12 we will study the closed fragment of HA^*.

The theory HA^* was introduced in [Vis82]. HA^* is to Beeson's fp-realizability, introduced in [Bee75], as Troelstra's $\mathsf{HA} + ECT_0$ is to Kleene's r-realizability. This means that for a suitable variant of fp-realizability HA^* is the set of sentences such that their fp-translations are provable in HA. The theory HA^* is HA plus *the Completeness Principle for* HA^*. The Completeness Principle can be viewed as an arithmetically interpreted modal principle. The Completeness Principle viewed modally is:

$\mathsf{C} \quad \vdash A \to \Box A$

The Completeness Principle for a specific theory T is:

$\mathsf{C}[T] \quad \vdash A \to \Box_T A.$

We have $\mathsf{HA}^* = \mathsf{HA} + \mathsf{C}[\mathsf{HA}^*]$.[7] We briefly review some of the results of [Vis82] and [dJV96].

[7] The natural way to define HA^* is by a fixed point construction as: HA plus the *Completeness Principle* for HA^*. Here it is essential that the construction is verifiable in HA.

- Let $\mathfrak{A}$ be the smallest class closed under atoms and all connectives except implication, satisfying: $A \in \Sigma_1$ and $B \in \mathfrak{A} \Rightarrow (A \to B) \in \mathfrak{A}$. Note that modulo provable equivalence in HA all prenex formulas (of the classical arithmetical hierarchy) are in $\mathfrak{A}$. HA^* is conservative w.r.t. $\mathfrak{A}$ over HA. Note that $\mathsf{NNIL}_0(\Sigma) \subseteq \mathfrak{A}$.

- There are infinitely many incomparable T with $T = \mathsf{HA} + \mathsf{C}[T]$. However if $T = \mathsf{HA} + \mathsf{C}[T]$ *verifiably in* HA, then $T = \mathsf{HA}^*$.

- Let KLS:=Kreisel-Lacombe-Shoenfield's Theorem on the continuity of the effective operations. We have $\mathsf{HA}^* \vdash \mathsf{KLS} \to \Box_{\mathsf{HA}^*}\bot$. This immediately gives Beeson's result that $\mathsf{HA} \nvdash \mathsf{KLS}$ [Bee75].

- Every prime RE Heyting algebra $\mathcal{H}$ can be embedded into the Heyting algebra of HA^*. This mapping is primitive recursive in the enumeration of the generators of $\mathcal{H}$ w.r.t which $\mathcal{H}$ is RE. The mapping sends the generators to Σ-sentences [dJV96].

We insert some basic facts on the provability logic of HA^*. These materials will be only needed in section 12. Clearly, HA^* satisfies the Löb conditions.

L1 $\quad \vdash A \Rightarrow \vdash \Box A$

L2 $\quad \vdash \Box(A \to B) \to (\Box A \to \Box B)$

L3 $\quad \vdash \Box A \to \Box\Box A$

L4 $\quad \vdash \Box(\Box A \to A) \to \Box A$

$\mathsf{i\text{-}K}$ is given by IPC+L1,L2. $\mathsf{i\text{-}L}$ is $\mathsf{i\text{-}K}$+L3,L4. We write $\mathsf{i\text{-}K}\{P\}$ for the extension of $\mathsf{i\text{-}K}$ with some principle P. Note that $\mathsf{i\text{-}L}\{\mathsf{C}\}$ is valid for provability interpretations in HA^*.

A principle closely connected to C is the Strong Löb Principle:

SL $\quad \vdash (\Box A \to A) \to A$

As a special case of SL we have $\vdash \neg\neg\Box\bot$.

Theorem 8.2 $\mathsf{i\text{-}L}\{\mathsf{C}\}$ *is interderivable with* $\mathsf{i\text{-}K}\{\mathsf{SL}\}$.

Proof

L4 is immediate from SL. We show "$\mathsf{i\text{-}K}\{\mathsf{SL}\} \vdash \mathsf{C}$". Reason in $\mathsf{i\text{-}K}\{\mathsf{SL}\}$. Suppose A. It easily follows that $\Box(A \wedge \Box A) \to (A \wedge \Box A)$, and, hence, by SL, that $A \wedge \Box A$. We may conclude that $\Box A$.

We show "$\mathsf{i\text{-}L}\{\mathsf{C}\} \vdash \mathsf{SL}$". Reason in $\mathsf{i\text{-}L}\{\mathsf{C}\}$. Suppose $(\Box A \to A)$. It follows, by C, that: $(\Box A \to A) \wedge \Box(\Box A \to A)$. Hence, by Löb's Principle, $(\Box A \to A) \wedge \Box A$ and thus A. ❑

9 Closure Properties of Σ-Preservativity

In this section we verify two closure properties of Σ-preservativity over HA in HA.

9.1 Closure under B1

We will show that HA verifies that Σ-preservativity is closed under B1 for any substitutions. We produce both proofs known to us. The first one employs q-realizability. This form of realizability is a translation from $\mathcal{A}$ to $\mathcal{A}$, due to Kleene. It is defined as follows:

- $x \mathrel{\mathsf{q}} P := P$, for P atomic
- $x \mathrel{\mathsf{q}} (A \wedge B) := ((x)_0 \mathrel{\mathsf{q}} A \wedge (x)_1 \mathrel{\mathsf{q}} B)$
- $x \mathrel{\mathsf{q}} (A \vee B) := (((x)_0 = 0 \to (x)_1 \mathrel{\mathsf{q}} A) \ \wedge \ ((x)_0 \neq 0 \to (x)_1 \mathrel{\mathsf{q}} B))$
- $x \mathrel{\mathsf{q}} (A \to B) := \forall y\ (y \mathrel{\mathsf{q}} A \to \exists z\ (\{x\}y \simeq z \wedge z \mathrel{\mathsf{q}} B)) \wedge (A \to B)$
- $x \mathrel{\mathsf{q}} \exists y\ A(y) := (x)_0 \mathrel{\mathsf{q}} A((x)_1)$
- $x \mathrel{\mathsf{q}} \forall y\ A(y) := \forall y \exists z\ (\{x\}y \simeq z \wedge z \mathrel{\mathsf{q}} A(y)))$

The following facts can be verified in HA.

1. $\mathsf{HA} \vdash x \mathrel{\mathsf{q}} A \to A$
2. for any $A \in \Sigma$ and any set $\{y_1, \ldots, y_n\}$ with $\mathsf{FV}(A) \subseteq y_1, \ldots, y_n$, we can find an index e such that $\mathsf{HA} \vdash A \to \exists z(\{\underline{e}\}(y_1, \ldots, y_n) \simeq z \ \wedge \ z \mathrel{\mathsf{q}} A)$
3. Suppose $B_1, \ldots, B_n, C$ have free variables among $\{y_1, \ldots, y_m\}$ and that $\{x_1, \ldots, x_n\}$ is disjoint from $\{y_1, \ldots, y_m\}$. Suppose $B_1, \ldots, B_n \vdash_{\mathsf{HA}} C$. Then, for some e, we have:

$$x_1 \mathrel{\mathsf{q}} B_1, \cdots, x_n \mathrel{\mathsf{q}} B_n \vdash_{\mathsf{HA}} \exists z\ (\{\underline{e}\}(x_1, \ldots, x_n, y_1, \ldots, y_m) \simeq z \wedge z \mathrel{\mathsf{q}} C).$$

The proofs are all simple inductions. For details the reader is referred to [Tro73], 188–202. We will now show that Σ-preservativity satisfies B1. It is easily seen that our proof is verifiable in HA. We reason as follows:

$\alpha)$	$A \rhd_{\mathsf{HA},\Sigma} C$	assumption
$\beta)$	$B \rhd_{\mathsf{HA},\Sigma} C$	assumption
$\gamma)$	$S \in \Sigma$ and $\mathsf{HA} \vdash S \to (A \vee B)$	assumption
$\delta)$	$x \mathrel{\mathsf{q}} S \vdash_{\mathsf{HA}} \exists z\ (\{\underline{e}\}(x) \simeq z \ \wedge \ z \mathrel{\mathsf{q}} (A \vee B))$	$\gamma,\ e$ provided by (3)
$\epsilon)$	$S \vdash_{\mathsf{HA}} \exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \wedge z \mathrel{\mathsf{q}} (A \vee B))$	$\delta,\ f$ provided by (2)
$\eta)$	$S \wedge \exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \wedge (z)_0 = 0) \vdash_{\mathsf{HA}} A$	ϵ
$\zeta)$	$S \wedge \exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \wedge (z)_0 \neq 0) \vdash_{\mathsf{HA}} B$	ϵ
$\theta)$	$S \wedge \exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \wedge (z)_0 = 0) \vdash_{\mathsf{HA}} C$	η, α
$\iota)$	$S \wedge \exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \wedge (z)_0 \neq 0) \vdash_{\mathsf{HA}} C$	ζ, β
$\kappa)$	$S \vdash_{\mathsf{HA}} \exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \ \wedge \ (z)_0 = 0) \ \vee$ $\exists z\ (\{\underline{e}\}(\underline{f}) \simeq z \ \wedge \ (z)_0 \neq 0)$	ϵ
$\lambda)$	$S \vdash_{\mathsf{HA}} C$	θ, ι, κ

The second proof employs a translation due to Dick de Jongh. For later use our definition is slightly more general than is really needed for the problem at hand. Let C be an $\mathcal{A}$-sentence and let n be a natural number. We write, locally, $\Box_n$ for $\Box_{\mathsf{HA},n}$. Define a translation $[C]_n(\cdot)$ as follows.

- $[C]_n P := P$ for P atomic
- $[C]_n(\cdot)$ commutes with $\wedge, \vee, \exists$
- $[C]_n(A \to B) := ([C]_n A \to [C]_n B) \ \wedge \ \Box_n(C \to (A \to B))$
- $[C]_n \forall y\, A(y) := \forall y\, [C]_n A(y) \wedge \Box_n(C \to \forall y\, A(y))$

Let's first make a few quick observations, that make life easier. We write $[C]_n\Gamma := \{[C]_n D \mid D \in \Gamma\}$. We have:

‡i) $\mathsf{HA} \vdash [C]_n A \to \Box_n(C \to A)$

‡ii) $\mathsf{HA} \vdash [C]_n((A \to B) \wedge (A' \to B')) \leftrightarrow$
$$([C]_n A \to [C]_n B) \wedge ([C]_n A' \to [C]_n B') \ \wedge$$
$$\Box_n(C \to ((A \to B) \ \wedge \ (A' \to B'))).$$
Similarly for conjunctions of more than two implications.

‡iii) $\mathsf{HA} \vdash [C]_n \forall y \forall z\, A(y,z) \leftrightarrow (\forall y \forall z\, [C]_n A(y,z) \ \wedge \ \Box_n(C \to \forall y \forall z\, A(y,z)))$.
Similary for larger blocks of universal quantifiers.

‡iv) $\mathsf{HA} \vdash [C]_n \forall y\, (A(y) \to B(y)) \leftrightarrow$
$$\forall y\, ([C]_n A(y) \to [C]_n B(y)) \wedge \Box_n(C \to \forall y\, (A(y) \to B(y))).$$

‡v) $\mathsf{HA} \vdash [C]_n \forall y{<}z\, A(y) \leftrightarrow \forall y{<}z\, [C]_n A(y)$.

‡vi) For $S \in \Sigma$, $\mathsf{HA} \vdash S \leftrightarrow [C]_n S$.

‡vii) $\Gamma \vdash_{\mathsf{HA},n} A \Rightarrow [C]_n\Gamma \vdash_{\mathsf{HA}} [C]_n A$ (verifiably in HA).

(‡v) is immediate from the well known fact that

$$\mathsf{HA} \vdash \forall y{<}z\, \Box_n A(y) \to \Box_n \forall y{<}z\, A(y).$$

(‡vi) is immediate from (‡v).

Proof of (‡vii) The proof is by induction on the proof witnessing $\Gamma \vdash_{\mathsf{HA},n} A$. We treat two cases.

First case $\Gamma = \emptyset$ and A is an induction axiom, say for $B(x)$, of HA_n. Clearly $[C]_n A$ is HA-provably equivalent to:

$$\begin{array}{llrll} (& (& [C]_n B0 & \wedge & \\ & & \forall x\, ([C]_n Bx \to [C]_n B(x+1)) & \wedge & \\ & & \Box_n(C \to \forall x\, (Bx \to B(x+1))) &) & \to \\ & & (\forall x\, [C]_n Bx \ \wedge \ \Box_n(C \to \forall x\, Bx)) &) & \wedge \\ & & \Box_n(C \to A) & & . \end{array}$$

We have: $\mathsf{HA} \vdash \Box_n A$, and, hence, $\mathsf{HA} \vdash \Box_n(C \to A)$. So it follows that:

$$\mathsf{HA} \vdash \Box_n(C \to \forall x\ (Bx \to B(x+1))) \to \Box_n(C \to \forall x\ Bx).$$

Moreover (as an instance of induction for $[C]_n Bx$):

$$\mathsf{HA} \vdash ([C]_n B0 \wedge \forall x\ ([C]_n Bx \to [C]_n B(x+1))) \to \forall x\ [C]_n Bx.$$

Combining these we find the promised: $\mathsf{HA} \vdash [C]_n A$.

Second case Suppose that $A = (D \to E)$ and that the last step in the proof was by:

$$\Gamma, D \vdash_{\mathsf{HA},n} E \Rightarrow \Gamma \vdash_{\mathsf{HA},n} D \to E.$$

From $\Gamma, D \vdash_{\mathsf{HA},n} E$, we have, by the Induction Hypothesis, $[C]_n\Gamma$, $[C]_n D \vdash_{\mathsf{HA},n} [C]_n E$ and, hence, $[C]_n\Gamma \vdash_{\mathsf{HA},n} [C]_n D \to [C]_n E$. Moreover, for some finite $\Gamma_0 \subseteq \Gamma$, we have $\Gamma_0, D \vdash_{\mathsf{HA},n} E$. Let B be the conjunction of the elements of Γ_0. We find: $[C]_n\Gamma \vdash_{\mathsf{HA},n} \Box_n(C \to B)$ and $\vdash_{\mathsf{HA},n} \Box_n(B \to (D \to E))$. Hence, $[C]_n\Gamma \vdash_{\mathsf{HA},n} \Box_n(C \to (D \to E))$. We may conclude:

$$[C]_n\Gamma \vdash_{\mathsf{HA},n} ([C]_n D \to [C]_n E) \ \wedge\ \Box_n(C \to (D \to E)).$$

Here ends our proof of (‡vii).
We now prove our principle. As is easily seen the argument can be verified in HA.

α)	$A \rhd_{\mathsf{HA},\Sigma} C$	assumption
β)	$B \rhd_{\mathsf{HA},\Sigma} C$	assumption
γ)	$S \in \Sigma$ and $\mathsf{HA} \vdash S \to (A \vee B)$	assumption
δ)	for some n $S \vdash_{\mathsf{HA},n} (A \vee B)$	γ
ϵ)	$[\top]_n S \vdash_{\mathsf{HA}} ([\top]_n A \vee [\top]_n B)$	$\delta, (\ddagger vii)$
η)	$S \vdash_{\mathsf{HA}} (\Box_n A \ \vee\ \Box_n B)$	$\epsilon, (\ddagger vi), (\ddagger i)$
ζ)	$\mathsf{HA} \vdash \Box_n A \to A$	theorem 8.1
θ)	$\mathsf{HA} \vdash \Box_n A \to C$	α, ζ
ι)	$\mathsf{HA} \vdash \Box_n B \to B$	theorem 8.1
κ)	$\mathsf{HA} \vdash \Box_n B \to C$	β, ι
λ)	$\mathsf{HA} \vdash (S \to C)$	η, θ, κ

9.2 A Closure Rule for Implication

To formulate our next closure rule it is convenient to work in a conservative extension of HA. Let $\mathcal{A}^+$ be $\mathcal{A}$ extended with new predicate symbols (including the 0-ary case) for Σ-formulas. Let f be some assignment of Σ-formulas of $\mathcal{A}$ of the appropriate arities to the new predicate symbols. We extend f to $\mathcal{A}^+$ in the obvious way. Define $[\cdot]_n^\circ(\cdot)$ and $[\![\cdot]\!]\,(\cdot) : \mathcal{A}^+ \to \mathcal{A}$ as follows:

- $[\![A]\!]\,P := [A]_n^\circ P := fP$ for P atomic
- $[\![A]\!]\,(\cdot)$ and $[A]_n^\circ(\cdot)$ commute with $\wedge$, $\vee$ and $\exists$

- $[\![A]\!]\,(B \to C) := f(A \to (B \to C))$,
  $[A]^\circ_n(B \to C) := \Box_n f(A \to (B \to C))$

- $[\![A]\!]\,\forall x\, B(x) := f(A \to \forall x\, B(x))$,
  $[A]^\circ_n \forall x\, B(x) := \Box_n f(A \to \forall x\, B(x))$

We have, for B in $\mathcal{A}^+$, extending the numbering of ‡-principles of subsection 9.1:

‡viii) The following implications are derivable in HA: $[fA]_n(fB) \to [A]^\circ_n B$, $[A]^\circ_n B \to [\![A]\!]\,B$ and $[\![A]\!]\,B \to (A \to B)$.

‡ix) $[A]^\circ_n B$ is provably equivalent to a Σ-formula.

The proof of (‡viii) is an easy induction on B. (‡ix) is trivial.

Theorem 9.1 *Suppose X is a finite set of implications in $\mathcal{A}^+$ and let B be in $\mathcal{A}^+$. Say $A := \bigwedge X$. Let $Y := \{C \mid (C \to D) \in X\} \cup \{B\}$. We have, verifiably in* HA, $f(A \to B) \rhd_{\mathsf{HA},\Sigma} \bigvee [\![A]\!]\,Y$.

Before giving the proof of theorem 9.1, we first draw a corollary.

Corollary 9.2 The following facts are verifiable in HA.

1. $\rhd_{\mathsf{HA},\Sigma,\Sigma}$ is closed under B2.
2. $\rhd_{\mathsf{HA},\Sigma,\mathcal{A}}$ is closed under B2′.
3. $\Sigma 8$ of appendix C is HA-valid for Σ-preservativity.

❑

We leave the easy verification of the corollary to the reader. We proceed with the proof of theorem 9.1.

Proof

To avoid heavy notations we suppress 'f'. In the context of HA we assume that an $\mathcal{A}^+$-formula is automatically translated via f to the corresponding $\mathcal{A}$-formula. Let S be a Σ-sentence (of $\mathcal{A}$). Suppose $S \vdash_{\mathsf{HA}} (A \to B)$. It follows that, for some n, $S \vdash_{\mathsf{HA},n} (A \to B)$ and, hence, by (‡vii) $[A]_n S \vdash_{\mathsf{HA}} [A]_n(A \to B)$. By (‡ii) and (‡vi) we find:

$$S \vdash_{\mathsf{HA}} (\bigwedge\{[A]_n C \to [A]_n D \mid (C \to D) \in X\} \wedge \Box_n(A \to A)) \to [A]_n B,$$

and so:

$$S \vdash_{\mathsf{HA}} \bigwedge\{[A]_n C \to [A]_n D \mid (C \to D) \in X\} \to [A]_n B.$$

It follows by (‡viii) that:

$$S \vdash_{\mathsf{HA}} \bigwedge\{[A]^\circ_n C \to [A]_n D \mid (C \to D) \in X\} \to [A]^\circ_n B.$$

Since HA is a subtheory of PA, we get:

$$S \vdash_{\mathsf{PA}} \bigwedge\{[A]_n^\circ C \to [A]_n D \mid (C \to D) \in X\} \to [A]_n^\circ B.$$

By classical logic, we get $S \vdash_{\mathsf{PA}} \bigvee [A]_n^\circ Y$. Remember that PA is, verifiably in HA, conservative over HA w.r.t Π_2-sentences (see [Fri78]). Since $(S \to \bigvee [A]_n^\circ Y)$ is Π_2, we get $S \vdash_{\mathsf{HA}} \bigvee [A]_n^\circ Y$. Ergo, by (‡viii), $S \vdash_{\mathsf{HA}} \bigvee [\![A]\!] Y$. ❑

Before closing this section we insert some remarks on the proof.

Remark 9.3 1. The above proof was obtained after analyzing an argument in [dJ82].

2. The step involving conservativity of PA over HA uses the Gödel-Friedman translation. Closer inspection reveals that the argument at hand just requires the Friedman translation. We give a sketch of the proof below. It follows that our results generalize to all essentially reflexive RE extensions T of HA that are closed both under the Friedman and the de Jongh translation.

3. It seems to me that, from a sufficiently abstract perspective, it should become clear that our present proof is just a variant of the proof of 6.2. Thus, $\{G \mid \Box_n(A \to G)\}$ in the present proof would correspond to the grey part of the Kripke model in the picture below 6.2. $[A]_n(\cdot)$ would correspond to the operation of adding the bottom-node in the picture (via Smoryński's operation) to the grey part. The detour via PA would correspond to the fact that our Kripke model argument is essentially classical. The special behaviour of Σ-sentences under translation would correspond with the fact that whether a Σ-sentence is forced or not (in a model of HA) is dependent just on the node under consideration and not on other nodes. Let me pose it as a problem to further develop such a perspective.

We prove the claim of (2): we show that the double negation translation can be eliminated from the proof of 9.1. We pick up the proof from the point where we have proved:

a) $S \vdash_{\mathsf{HA}} \bigwedge\{[A]_n^\circ C \to [A]_n D \mid (C \to D) \in X\} \to [A]_n^\circ B$.

Let $H := \bigvee [A]_n^\circ Y$. The Friedman translation $(E)^H$ of an arithmetical formula E is (modulo some details to avoid variable-clashes) the result of replacing all atomic formulas P in E by $(P \vee H)$. One easily shows:

b) $\mathsf{HA} \vdash E \Rightarrow \mathsf{HA} \vdash (E)^H$

c) $\mathsf{HA} \vdash H \to (E)^H$

d) for $S \in \Sigma$, $\mathsf{HA} \vdash (S)^H \leftrightarrow (S \vee H)$.

By (a), (b) and (d) we have:

e) $S \vee H \vdash_{\mathsf{HA}}$
$$\bigwedge\{(([A]_n^{\circ}C) \vee H) \to ([A]_n D)^H \mid (C \to D) \in X\} \to (([A]_n^{\circ}B) \vee H).$$

By (e) and propositional logic we find:

f) $S \vdash_{\mathsf{HA}} \bigwedge\{H \to ([A]_n D)^H \mid (C \to D) \in X\} \to H$.

So by (f) and (c):

g) $S \vdash_{\mathsf{HA}} H$.

Hence, by (g) and (‡viii), $S \vdash_{\mathsf{HA}} \bigvee [\![A]\!] Y$. ❑

10 On Σ-Substitutions

In this section we prove our main results on Σ-substitutions. Given the results we already have, this is rather easy.

Theorem 10.1 $\rhd_{\mathsf{HA},\Sigma,\Sigma}$ *is, verifiably in* HA*, a* σ*-relation.*

Proof

Every preservativity relation satisfies A1-3, and, hence, so does $\rhd_{\mathsf{HA},\Sigma,\Sigma}$. Moreover, $\rhd_{\mathsf{HA},\Sigma,\Sigma}$ satisfies B1 in virtue of our result of subsection 9.1. $\rhd_{\mathsf{HA},\Sigma,\Sigma}$ satisfies B2 by corollary 9.2. $\rhd_{\mathsf{HA},\Sigma,\Sigma}$ satisfies B3 by lemma 4.2(4). It is easy to see that all the steps are verifiable in HA. ❑

Let R be any arithmetically definable relation. We write R^{HA} for the relation given by: $mR^{\mathsf{HA}}n :\Leftrightarrow \mathsf{HA} \vdash \underline{m}R\underline{n}$.

Theorem 10.2 *The following relations are coextensive:*

$$(a)\ \blacktriangleright_{\sigma},\ \ (b)\ \rhd^{\mathsf{HA}}_{\mathsf{HA},\Sigma,\Sigma},\ \ (c)\ \rhd_{\mathsf{HA},\Sigma,\Sigma},\ \ (d)\ \mathrel{|\!\!\!\sim}^{\mathsf{HA}}_{\mathsf{HA},\Sigma},\ \ (e)\ \mathrel{|\!\!\!\sim}_{\mathsf{HA},\Sigma}.$$

Proof

By theorem 10.1, $\rhd^{\mathsf{HA}}_{\mathsf{HA},\Sigma,\Sigma}$ is a σ-relation. So, $(a) \subseteq (b)$.

Since HA is sound, we find $(b) \subseteq (c)$.

By lemma 4.2(6), $(c) \subseteq \rhd_{\mathsf{HA},\top,\Sigma} = (e)$.

Since the argument for lemma 4.2(6) is just universal instantiation, it is verifiable in HA. Hence $(b) \subseteq (d)$.

By the soundness of HA, we have $(d) \subseteq (e)$.

We show that $(e) \subseteq (a)$, i.e. $\mathrel{|\!\!\!\sim}_{\mathsf{HA},\Sigma} \subseteq \blacktriangleright_{\sigma}$. Suppose *not* $A \blacktriangleright_{\sigma} B$. Then, by corollary 7.2(1), $A^* \nvdash_{\mathsf{IPC}} B$. A^* is IPC-provably equivalent to a disjunction of prime NNIL-formulas. It follows that for some such disjunct, say C, $C \nvdash_{\mathsf{IPC}} B$. The Heyting algebra $\mathcal{H}$ axiomatized by C is prime and RE. By the embedding theorem proved in [dJV96] (see section 8.2 of this paper) there is an $f \in \Sigma^{\mathcal{P}}$, such that $\mathsf{HA}^* \vdash fC$ and $\mathsf{HA}^* \nvdash fB$. Since $fC \in \mathsf{NNIL}(\Sigma)$, we have by the $\mathsf{NNIL}(\Sigma)$-conservativity of HA^* over HA (proved in [Vis82]; see also section 8.2 for the statement of the full result), that $\mathsf{HA} \vdash fC$. Since HA is a subtheory of HA^*, we have $\mathsf{HA} \nvdash fB$. Since $\mathsf{IPC} \vdash C \to A$, it follows that $\mathsf{HA} \vdash fA$ and $\mathsf{HA} \nvdash fB$. We may conclude: *not* $A \;\vdash\!\sim_{\mathsf{HA},\Sigma} B$. ❑

Remark 10.3 (1) A propositional formula A is HA,Σ-exact iff there is a substitution $f \in \Sigma^{\mathcal{P}}$, such that $\mathsf{HA} \vdash fB \Leftrightarrow A \vdash_{\mathsf{IPC}} B$, i.o.w. if A axiomatizes the 'theory of f', i.e. $\Lambda_{\mathsf{HA},\{f\}}$. See also [dJV96]. The proof of theorem 10.2 establishes also that the prime NNIL-formulas are *precisely* the HA,Σ-exact formulas.

(2) By theorem 10.2, $\vdash\!\sim^{\mathsf{HA}}_{\mathsf{HA},\Sigma}$ is a σ-relation. Note, however, that it is *not* HA-provably a σ-relation. The reason is that closure under B1 is not verifiable. To prove this, suppose HA verifies B1. Let R be the standard Rosser sentence. So R satisfies $\mathsf{HA} \vdash R \leftrightarrow \Box_{\mathsf{HA}}\neg R \leq \Box_{\mathsf{HA}}R$. Let $R^{\perp} := \Box_{\mathsf{HA}}R < \Box_{\mathsf{HA}}\neg R$. We have (a) $\mathsf{HA} \vdash \Box_{\mathsf{HA}}R \to \Box_{\mathsf{HA}}\bot$ and (b) $\mathsf{HA} \vdash \Box_{\mathsf{HA}}R^{\perp} \to \Box_{\mathsf{HA}}\bot$. This is simply the Formalized Rosser Property. Moreover, we have: (c) $\mathsf{HA} \vdash \Box_{\mathsf{HA}}\bot \to (R \vee R^{\perp})$. This is the formalization in HA of the insight that if $\bot$ is witnessed then both R and $\neg R$ will be witnessed; one must be witnessed first, which means that either R or $R^{\perp}$. From (c), it follows that: (d) $\mathsf{HA} \vdash \Box_{\mathsf{HA}}\Box_{\mathsf{HA}}\bot \to \Box_{\mathsf{HA}}(R \vee R^{\perp})$. On the other hand, from (a), (b) and the assumed verifiable closure under $\vdash\!\sim^{\mathsf{HA}}_{\mathsf{HA},\Sigma}$, we find that (e) $\mathsf{HA} \vdash \Box_{\mathsf{HA}}(R \vee R^{\perp}) \to \Box_{\mathsf{HA}}\bot$. Combining (d) and (e), we get $\mathsf{HA} \vdash \Box_{\mathsf{HA}}\Box_{\mathsf{HA}}\bot \to \Box_{\mathsf{HA}}\bot$. Quod non, by Löb's Theorem. ❑

11 The Admissible Rules of HA

The results of Rosalie Iemhoff [Iem99] imply the following theorem.

Theorem 11.1 (Iemhoff) $\vdash\!\sim_{\mathsf{IPC}} = \blacktriangleright_{\alpha}$. *In other words: the admissible rules of* IPC *are axiomatized by A1, A2, A3, B1, B2'.*

From Iemhoff's result, we derive the following theorem.

Theorem 11.2 *The following relations are coextensive:*

$$(a)\ \blacktriangleright_{\alpha},\ \ (b)\ \rhd^{\mathsf{HA}}_{\mathsf{HA},\Sigma,\mathcal{A}},\ \ (c)\ \rhd_{\mathsf{HA},\Sigma,\mathcal{A}},\ \ (d)\ \vdash\!\sim^{\mathsf{HA}}_{\mathsf{HA}},\ \ (e)\ \vdash\!\sim_{\mathsf{HA}},\ \ (f)\ \vdash\!\sim_{\mathsf{IPC}}.$$

Proof

We first show that $(a) \subseteq (b)$. It is clearly sufficient to show that $\rhd_{\mathsf{HA},\Sigma,\mathcal{A}}$ is HA-verifiably an α-relation. Closure under A1-3 is trivial. Moreover, $\rhd_{\mathsf{HA},\Sigma,\mathcal{A}}$

satisfies B1 in virtue of our result of subsection 9.1. $\rhd_{\mathsf{HA},\Sigma,\mathcal{A}}$ satisfies B2 by corollary 9.2.

By the soundness of HA, we have $(b) \subseteq (c)$ and $(d) \subseteq (e)$.

By lemma 4.2(6), $(c) \subseteq \rhd_{\mathsf{HA},\top,\mathcal{A}} = (e)$. Since lemma 4.2(6) is verifiable in HA, we also have $(b) \subseteq (d)$.

Since, by de Jongh's theorem, $\Lambda_{\mathsf{HA}} = \mathsf{IPC}$, we find, by theorem 4.4, that $(e) \subseteq (f)$.

Finally, by theorem 11.1, $(f) \subseteq (a)$. ❏

12 Closed Fragments

In this section we study the closed fragments of the provability logics of three theories. The first is the theory HA^*. The second is Peano Arithmetic, PA. The third is HA. We will first introduce some definitions an prove some basic facts.

12.1 Preliminaries

Consider the language of modal propositional logic without propositional variables. Let's call this language $\mathcal{L}_{\Box,0}$. So $\mathcal{L}_{\Box,0}$ is given as follows.

- $A ::= \bot \mid \top \mid (A \wedge A) \mid (A \vee A) \mid (A \to A) \mid \Box A$.

Let T be any theory extending, say, i-S^1_2, the intuitionistic version of Buss' S^1_2. What is relevant here is just that arithmetization of metamathematics should be possible in T and that T verifies the Löb conditions. We can map the formulas of $\mathcal{L}_{\Box,0}$, say via e_T, to the language of T, by stipulating that e_T commutes with all the propositional connectives including $\top$ and $\bot$ and that

$$\mathsf{e}_T \Box A := \Box_T\ \mathsf{e}_T A := \mathsf{Prov}_T\ (\underline{\mathsf{gn}(\mathsf{e}_T A)}).$$

The closed fragment $\mathcal{C}_T$ of the provability logic of T is the set of A in $\mathcal{L}_{\Box,0}$ such that $T \vdash \mathsf{e}_T A$.

An important class, DF, of closed formulas is formed by the *degrees of falsity* $\Box^\alpha \bot$, where α ranges over $\omega^+ := \omega \cup \{\infty\}$. Here:

- $\Box^0 \bot := \bot$,
- $\Box^{n+1} \bot := \Box\Box^n \bot$,
- $\Box^\infty \bot := \top$,

We will often use the evident mappings $\mathsf{df} : \alpha \mapsto \Box^\alpha\bot$ and $\varpi : \Box^\alpha\bot \mapsto \alpha$.

To link the closed fragments to the framework of this paper, we will study purely propositional theories associated to these fragments. The language of these propositional theories is the usual language of propositional logic with as 'propositional atoms' the degrees of falsity. These atoms function more as constants than as propositional variables. We call this language BDF. So BDF is given as follows.

- $A \ ::= \ \Box^\alpha\bot \mid (A \wedge B) \mid (A \vee B) \mid (A \to B)$.

We will consider theories Θ in BDF, which satisfy (at least) intuitionistic propositional logic and the $\mathtt{up}$ axiom:

$\mathtt{up} \vdash \Box^n\bot \to \Box^{n+1}\bot$.

The minimal such theory will be called UP. So UP is axiomatized by IPC and $\mathtt{up}$. We will say that Θ is DF-*irreflexive* if $\Box^\alpha\bot \vdash_\Theta \Box^\beta\bot \ \Rightarrow \ \alpha \leq \beta$. It is easy to show that UP is DF-irreflexive. Note that subtheories of DF-irreflexive theories are DF-irreflexive.

An arithmetical theory T will have associated to it the propositional theory $\Lambda_{T,\mathsf{e}_T} := \{A{\in}\mathsf{BDF} \mid T \vdash \mathsf{e}_T A\}$. (Strictly speaking e_T here is the restriction of e_T as defined above to BDF.)

We will write $\Box_T^\alpha\bot$ for $\mathsf{e}_T(\Box^\alpha\bot)$. We will say that T is DF-*sound* if, for no $n \in \omega$, $T \vdash \Box_T^n\bot$.

Lemma 12.1 Suppose T is DF-sound. Then, Λ_{T,e_T} is DF-irreflexive. ❑

Proof

Suppose we have $\Box^\alpha\bot \vdash_{\Lambda_{T,\mathsf{e}_T}} \Box^\beta\bot$. We assume, to obtain a contradiction, that $\beta < \alpha$. Then, by definition, $\Box_T^\alpha\bot \vdash_T \Box_T^\beta\bot$. It follows that: $\Box_T^{\beta+1}\bot \vdash_T \Box_T^\beta\bot$ and $\beta \in \omega$. By Löb's Rule, we obtain $T \vdash \Box_T^\beta\bot$. But this is impossible, by DF-soundness. Ergo $\alpha \leq \beta$. ❑

The fundamental structure in the study of closed fragments is $\omega^+ := \omega \cup \{\infty\} = \{0, 1, 2, \cdots \infty\}$ equipped with the obvious partial order. ω^+ with partial order can be extended to a Heyting algebra. The ordering fixes the operations of the Heyting algebra uniquely. We have:

- $\top := \infty$, $\bot := 0$,
- $(\alpha \wedge \beta) := \min(\alpha, \beta)$.
- $(\alpha \vee \beta) := \max(\alpha, \beta)$,
- $(\alpha \to \beta) \ := \ \begin{cases} \infty & \text{if } \alpha \leq \beta \\ \beta & \text{otherwise} \end{cases}$

Par abus de langage we will use ω^+ for both set, partial order and Heyting algebra. We will put $\alpha + \infty = \infty + \alpha = \infty$. We will repeatly use the following convenient property: $\alpha + n \leq \beta + n \Rightarrow \alpha \leq \beta$.

The map $\mathsf{df} : \alpha \mapsto \Box^\alpha \bot$ can be considered as a functor from ω^+, considered as a category, to a theory Θ considered as the category of the preorder $\vdash_\Theta$.

We will be interested in theories where df has a right adjoint ∂_Θ, i.e. where we have:

- $\mathsf{df}(\alpha) \vdash_\Theta A \Leftrightarrow \alpha \leq \partial_\Theta A$.

Lemma 12.2 Suppose Θ is DF-irreflexive and ∂_Θ exists. Then, $\partial_\Theta \Box^\alpha \bot = \alpha$. ❑

Proof

We have, for DF-irreflexive Θ,
$\gamma \leq \partial_\Theta \Box^\alpha \bot \Leftrightarrow \Box^\gamma \bot \vdash_\Theta \Box^\alpha \bot \Leftrightarrow \gamma \leq \alpha$. ❑

We will write ∂_T for $\partial_{\Lambda_{T,\mathsf{e}_T}}$. We can 'lift' ∂_Θ from BDF to $\mathcal{L}_{\Box,0}$ using the auxiliary function $(\cdot)^{\mathsf{aux}} : \mathcal{L}_{\Box,0} \to \mathsf{BDF}$ as follows.

- $(\cdot)^{\mathsf{aux}}$ commutes with the propositional connectives including $\top$ and $\bot$,
- $(\Box A)^{\mathsf{aux}} := \Box \mathsf{df} \partial_\Theta A^{\mathsf{aux}}$

We can now define $\Delta_\Theta : \mathcal{L}_{\Box,0} \to \omega^+$, by $\Delta_\Theta A := \partial_\Theta A^{\mathsf{aux}}$. We show that, for DF-irreflexive Θ, ∂_Θ and Δ_Θ coincide on BDF-formulas.

Lemma 12.3 Suppose Θ is DF-irreflexive. Let $A \in \mathsf{BDF}$. Then $\Delta_\Theta A = \partial_\Theta A$. ❑

Proof

Suppose Θ is DF-irreflexive. First we show, by induction on n, that $\Delta_\Theta \Box^n \bot = n$. For $n = 0$, this is immediate. We have:

$$\begin{array}{rcl}
\Delta_\Theta \Box^{k+1} \bot & = & \partial_\Theta (\Box^{k+1} \bot)^{\mathsf{aux}} \\
& = & \partial_\Theta \Box \mathsf{df} \partial_\Theta (\Box^k \bot)^{\mathsf{aux}} \\
& = & \partial_\Theta \Box \mathsf{df} \Delta_\Theta \Box^k \bot \\
& = & \partial_\Theta \Box \mathsf{df}(k) \\
& = & \partial_\Theta \Box^{k+1} \bot \\
& = & k + 1.
\end{array}$$

It now follows that, for $A \in \mathsf{BDF}$, we have $A^{\mathsf{aux}} = A$. The theorem follows directly from this. ❑

Finally, define, for A, B in BDF:

- $A \ \mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}} \ B :\Leftrightarrow T \vdash \Box_T \ \mathsf{e}_T A \to \Box_T \ \mathsf{e}_T B$.

So $\mathrel{\vdash\!\!\sim}^{\mathsf{prov}}$ is a form of *provable deductive consequence*. Note that if T is Σ-sound, then $\Lambda_{T,\mathsf{e}_T} = \{A \in \mathsf{BDF} \mid \top \ \mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}} \ A\}$. We will write the induced equivalence relation of $\mathrel{\vdash\!\!\sim}^{\mathsf{prov}}$ as $\approx^{\mathsf{prov}}$.

The following theorem is the main lemma for the subsequent development.

Theorem 12.4 *Let ϕ be a function from BDF to ω^+. We put $\Phi A := \Box^{\phi A}\bot$. Note that ϕ and Φ are interdefinable. Let T be a theory containing i-S^1_2. Let Θ be a BDF-theory containing IPC and up. We assume the following.*

Z1 T *is* DF*-sound.*

Z2 Θ *is a T-sound, i.e.* $\Theta \vdash A \Rightarrow T \vdash \mathsf{e}_T A$*, for* $A \in \mathsf{BDF}$.

Z3 $\Phi A \vdash_\Theta A$.

Z4 $A \ \mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}} \ \Phi A$.

Then we have:

1. $\Theta = \Lambda_{T,\mathsf{e}_T}$.
2. *Φ is the left adjoint of the embedding of $\vdash_\Theta$ into $\mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}}$. In other words, we have: $\Phi A \vdash_\Theta B \Leftrightarrow A \ \mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}} \ B$.*
3. *The full subcategory of $\vdash_\Theta$ obtained by restriction to DF is equivalent to $\mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}}$. This subcategory is isomorphic to ω^+.*
4. *The full subcategory of $\mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}}$ obtained by restriction to DF is a skeleton of $\mathrel{\vdash\!\!\sim}_T^{\mathsf{prov}}$. This subcategory is isomorphic to ω^+.*
5. $\phi(\Box^\alpha\bot) = \alpha$ *and* $\phi = \delta_\Theta$.
6. *We have, for any $A \in \mathcal{L}_{\Box,0}$:*
   (a) $\mathcal{C}_T \vdash A \leftrightarrow A^{\mathsf{aux}}$,
   (b) $\mathcal{C}_T \vdash A \Leftrightarrow \Delta_\Theta A = \infty$.

Proof

(under the assumptions of the theorem)

We prove (1). By Z2, Θ is T-sound. We show that it is also T-complete. Suppose $T \vdash \mathsf{e}_T A$. Then $T \vdash \Box_T \mathsf{e}_T A$. By Z4, $T \vdash \Box_T \mathsf{e}_T \Phi A$. By Z1, it follows that $\Phi A = \top$. By Z3, we may conclude that $\Theta \vdash A$.

We prove (2). Note that from Z3,4 it follows that $\Phi C \approx_T^{\mathsf{prov}} C$. We have:

$$\begin{array}{rcl} \Phi A \vdash_\Theta B & \Rightarrow & \mathsf{e}_T \Phi A \vdash_T \mathsf{e}_T B \\ & \Rightarrow & \Box_T \mathsf{e}_T \Phi A \vdash_T \Box_T \mathsf{e}_T B \\ & \Rightarrow & \Box_T \mathsf{e}_T A \vdash_T \Box_T \mathsf{e}_T B \\ & \Rightarrow & \Box_T \mathsf{e}_T \Phi A \vdash_T \Box_T \mathsf{e}_T \Phi B \\ & \Rightarrow & \phi A + 1 \leq \phi B + 1 \\ & \Rightarrow & \phi A \leq \phi B \\ & \Rightarrow & \Phi A \vdash_\Theta \Phi B \\ & \Rightarrow & \Phi A \vdash_\Theta B \end{array}$$

We prove (3). The equivalence of the restriction of $\vdash_\Theta$ to DF and $\mathrel{\vdash\!\sim}_T^{\mathsf{prov}}$, is immediate from lemma A.1(5). Since, by Z1, T is DF-sound, clearly Θ is DF-irreflexive. Hence the restriction of $\vdash_\Theta$ to DF is isomorphic to ω^+.

We prove (4). The restriction of $\mathrel{\vdash\!\sim}_T^{\mathsf{prov}}$ to DF is, by lemma A.1(4) and Z1, equivalent to $\mathrel{\vdash\!\sim}_T^{\mathsf{prov}}$. Trivially the restriction is isomorphic to ω^+.

We prove (5). The first part is immediate from the assumptions of the theorem. The second part follows by:

$$\begin{array}{rcl} \mathsf{df}\alpha \vdash_\Theta A & \Leftrightarrow & \Phi \mathsf{df}\alpha \vdash_\Theta A \\ & \Leftrightarrow & \mathsf{df}\alpha \mathrel{\vdash\!\sim}_T^{\mathsf{prov}} A \\ & \Leftrightarrow & \mathsf{df}\alpha \mathrel{\vdash\!\sim}_T^{\mathsf{prov}} \Phi A \\ & \Leftrightarrow & \alpha \leq \phi A \end{array}$$

Since right adjoints are unique modulo natural isomorphism, it follows that $\phi = \delta_\Theta$.

We prove (6). The proof of (6a) is by induction on A, using that, for $B \in \mathsf{BDF}$, $\mathcal{C}_T \vdash \Box B \leftrightarrow \Box \Phi B$. (6b) is immediate from (6a). ❑

We close this section with a lemma.

Lemma 12.5 Let $B \in \mathsf{BDF}$. Then we have:

$\mathsf{HA} \vdash \mathsf{e}_{\mathsf{HA}}\, B \leftrightarrow \mathsf{e}_{\mathsf{HA}^*}\, B \leftrightarrow \mathsf{e}_{\mathsf{PA}}\, B.$ ❑

Proof

It is well known that both HA^* (see subsection 8.2) and PA are HA-provably Π^0_2-conservative over HA. It follows that $\mathsf{HA} \vdash \Box_{\mathsf{HA}}\bot \leftrightarrow \Box_{\mathsf{HA}^*}\bot \leftrightarrow \Box_{\mathsf{PA}}\bot$. The lemma follows by induction on B. ❑

12.2 The Closed Fragment for HA^*

In subsection 8.2 we introduced HA^* and it's provability logic. We saw that in this logic at least the principles of the constructive version of Löb's Logic plus the Strong Löb Principle SLP, $\vdash (\Box A \to A) \to A$, hold.

The propositional theory UP^* is the BDF-theory, axiomatized by the rules of IPC plus

- $\mathsf{up} \vdash \Box^n \bot \to \Box^{n+1} \bot$,
- $\mathsf{slp} \vdash (\Box^{n+1} \bot \to \Box^n \bot) \to \Box^n \bot$.

We define a mapping $\phi_{\mathsf{HA}^*} := [\![\cdot]\!]$ from BDF to $\omega \cup \{\infty\}$ as follows:

- $[\![\bot]\!] := 0$, $[\![\top]\!] := \infty$,
- $[\![A \wedge B]\!] := \min([\![A]\!], [\![B]\!])$,
- $[\![A \vee B]\!] := \max([\![A]\!], [\![B]\!])$,
- $[\![A \to B]\!] := ([\![A]\!] \to [\![B]\!])$,

Define further $\Phi_{\mathsf{HA}^*} A := A^{\diamondsuit} := \Box^{[\![A]\!]} \bot$. We will verify Z1-4 of theorem 12.4 for HA^*, UP^* and $[\![\cdot]\!]$.

Clearly HA^* is DF-sound. So we have Z1. Moreover, UP^* is HA^*-sound. thus we have Z2. It is easy to see, by an induction on the subformulas of A, that for A in BDF:

£ $\mathsf{UP}^* \vdash A \leftrightarrow A^{\diamondsuit}$.

The only non-trivial step is the case of implication which uses slp. It follows that $[\![\cdot]\!]$ defines an equivalence between the category of UP^* and ω^+.

It is now easily seen that £ immediately implies both Z3 and Z4. It follows that $\mathsf{UP}^* = \Lambda_{\mathsf{HA}^*, \mathsf{e}_{\mathsf{HA}^*}}$. Moreover, since, evidently $(\cdot)^{\diamondsuit}$ is naturally isomorphic to the identity functor on $\vdash_{\mathsf{UP}^*}$, we find that provably deductive consequence and derivable consequence coincide for HA^*, i.e. $\vdash^{\mathsf{prov}}_{\mathsf{HA}^*} = \vdash_{\mathsf{UP}^*}$.

12.3 The Closed Fragment of PA

Friedman's 35th problem was to give a decision procedure for the closed fragment of the provability logic of PA. See [Fri75]. It was indepently solved by van Benthem, Boolos and Magari. To present the solution in our style, we first introduce the theory UP^c. This is classical propositional logic in the language BDF with the principle up. By a simple model theoretic argument one can show that $\mathsf{UP}^c = \Lambda_{\mathsf{PA}, \mathsf{e}_{\mathsf{PA}}}$. This fact will also follow from the application of theorem 12.4 below. Note that, trivially, UP^c is PA-sound.

Consider any A in BDF. In UP^c we can rewrite A to a normal form $\mathsf{nf}(A)$ as follows. First we rewrite A to conjunctive normal form, obtaining a conjunction of disjunctions of degrees of falsity and negations of degrees of falsity. The disjunctions of degrees of falsity and negations of degrees of falsity can be rewritten to implications from conjunctions of degrees of falsity to disjunctions of degrees of falsity. Finally, using up, we can contract conjunctions of degrees of falsity and disjunctions of degrees of falsity to single degrees of falsity. Thus we obtain a formula $\mathsf{nf}(A)$ of the form $\bigwedge_i(\Box^{\alpha_i}\bot \to \Box^{\beta_i}\bot)$. We define: $\Phi_{\mathsf{PA}} := (\cdot)^{\heartsuit} := (\cdot)^{\diamondsuit} \circ \mathsf{nf}$. Here $(\cdot)^{\diamondsuit}$ is the functor from subsection 12.2.

We verify Z1-4 of theorem 12.4 for PA, UP^c and $(\cdot)^{\diamondsuit}$. Z1,2 are trivial. We have:

- $\mathsf{UP}^c \vdash A \leftrightarrow \mathsf{nf}(A)$,
- $(\mathsf{nf}(A))^{\diamondsuit} \vdash_{\mathsf{UP}^c} \mathsf{nf}(A) \vdash_{\mathsf{UP}^c} A$.

The second item is simply by inspecting the computation of $(\cdot)^{\diamondsuit}$ on a formula of the special form $\mathsf{nf}(A)$. From the above, we have Z3. Moreover we have:

$$\begin{aligned} \Box_{\mathsf{PA}}\, \mathsf{e}_{\mathsf{PA}}\, A \;&\vdash_{\mathsf{PA}}\; \Box_{\mathsf{PA}}\, \mathsf{e}_{\mathsf{PA}}\, \mathsf{nf}(A) && (1)\\ &\vdash_{\mathsf{PA}}\; \Box_{\mathsf{PA}}\, \mathsf{e}_{\mathsf{PA}}\, (\mathsf{nf}(A))^{\diamondsuit}. && (2) \end{aligned}$$

The step labeled (2) is by a simple direct computation: first bring the conjunctions outside the box and then apply Löb's theorem to the conjuncts (if appropriate). Thus we have proved Z4.

Remark 12.6 There is an alternative way to prove (2) of the above argument.

$$\begin{aligned} \mathsf{PA} \vdash \Box_{\mathsf{PA}}\, \mathsf{e}_{\mathsf{PA}}\, A \;&\to\; \Box_{\mathsf{PA}}\, \mathsf{e}_{\mathsf{PA}}\, \mathsf{nf}(A) && (3)\\ &\to\; \Box_{\mathsf{HA}}\, \mathsf{e}_{\mathsf{HA}}\, \mathsf{nf}(A) && (4)\\ &\to\; \Box_{\mathsf{HA}^*}\, \mathsf{e}_{\mathsf{HA}^*}\, \mathsf{nf}(A) && (5)\\ &\to\; \Box_{\mathsf{HA}^*}\, \mathsf{e}_{\mathsf{HA}^*}\, (\mathsf{nf}(A))^{\diamondsuit} && (6)\\ &\to\; \Box_{\mathsf{HA}}\, \mathsf{e}_{\mathsf{HA}}\, (\mathsf{nf}(A))^{\diamondsuit} && (7)\\ &\to\; \Box_{\mathsf{PA}}\, \mathsf{e}_{\mathsf{PA}}\, (\mathsf{nf}(A))^{\diamondsuit} && (8) \end{aligned}$$

Step (3) is shown simply by classical logic. (4) uses first the HA-verifiable Π^0_2-conservativity of PA over HA and, then lemma 12.5. (5) is by lemma 12.5. (6) uses the results of subsection 12.2. (7) employs the HA-verifiable Π^0_2-conservativity of HA^* over HA. (8) is by lemma 12.5.

Of course, this second argument is a silly way of proving (2). It is more difficult, it uses more theory and it is less general (if we replace PA by other classical theories). However, it uncovers an analogy with the argument we are going to give for the case of HA. ❑

We have verified the assumptions Z1-4 of theorem 12.4. The theorem gives us the desired characterization of the closed fragment of PA.

Our argument, which is just a form of presenting the classical argument, for the van Benthem-Magari-Boolos result works for all DF-sound RE extensions of Buss's S^1_2.[8] In contrast the proof of Solovay's Arithmetical Completeness Theorem only works *as far as we know* if Exp is present.

[8] However, as we pointed out, the argument of remark 12.6 does not work in general.

12.4 The Closed Fragment of HA

Define $\Phi_{\mathsf{HA}} := (\cdot)^{\clubsuit} := (\cdot)^{\diamondsuit} \circ (\cdot)^{*}$. Here $(\cdot)^{*}$ is the functor described in section 7. We will verify Z1-4 of theorem 12.4 with UP in the role of Θ. Z1 and Z2 are immediate. We have:

$$
\begin{aligned}
(A^*)^{\diamondsuit} &\ \vdash_{\mathsf{UP}}\ A^* && (9)\\
&\ \vdash_{\mathsf{UP}}\ A && (10)
\end{aligned}
$$

(9) uses the fact that $A^* \in \mathsf{NNIL}$ and a simple induction of the complexity of NNIL-formulas. (10) is immediate from the results of section 7. Thus we have proved Z3. Finally we have:

$$
\begin{aligned}
\mathsf{HA} \vdash \Box_{\mathsf{HA}}\, \mathsf{e}_{\mathsf{HA}}\, A &\ \to\ \Box_{\mathsf{HA}}\, \mathsf{e}_{\mathsf{HA}}\, A^* && (11)\\
&\ \to\ \Box_{\mathsf{HA}^*}\, \mathsf{e}_{\mathsf{HA}^*}\, A^* && (12)\\
&\ \to\ \Box_{\mathsf{HA}^*}\, \mathsf{e}_{\mathsf{HA}^*}\, (A^*)^{\diamondsuit} && (13)\\
&\ \to\ \Box_{\mathsf{HA}}\, \mathsf{e}_{\mathsf{HA}}\, (A^*)^{\diamondsuit} && (14)
\end{aligned}
$$

Step (11) follows from theorem 10.2 in combination with the results of section 7. Step (12) uses lemma 12.5. (13) employs the results of subsection 12.2. (14) is by the HA-verifiable Π^0_1-conservativity of HA^* ovr HA and by lemma 12.5. Thus we have verified Z4.

Example 12.7 [Sample computations] Often the characterization of $(\cdot)^{\clubsuit}$ by $A^{\clubsuit} = \Box^{\partial_{\mathsf{UP}} A}\bot$, provides a quick way to compute $A^{\clubsuit}$. First one guesses the best approximation from below in DF of A, then one proves, e.g. using a Kripke model, that the conjectured approximation is indeed best. In the examples below we use our algorithm. We write $\approx$ for $\approx^{\mathsf{prov}}_{\mathsf{HA}}$.

1. Consider $A := \neg\Box\bot \to \Box\Box\bot$. We find that $A^* = \Box\bot \vee \Box\Box\bot$. Hence, $A \approx \Box^2\bot$.

2. Consider $A = (\neg\neg\Box\Box\bot \to \Box\Box\bot) \to (\Box\bot \vee \neg\Box\bot)$. Let $B := \neg\neg\Box\Box\bot \to \Box\Box\bot$. We compute A^*.

$$
\begin{aligned}
A &\approx (\,\Box\Box\bot \to (\Box\bot \vee \neg\Box\bot)\,)\ \wedge\\
&\qquad (\,[B]\neg\neg\Box\Box\bot \vee [B]\Box\bot \vee [B]\neg\Box\bot\,)\\
&\approx (\,\Box\Box\bot \to (\Box\bot \vee \neg\Box\bot)\,) \wedge (\neg\neg\Box\Box\bot \vee \Box\bot\ \vee\ \neg\Box\bot)\\
&\approx (\,\Box\Box\bot \to (\Box\bot \vee \neg\Box\bot)\,) \wedge (\,\Box\Box\bot \vee \Box\bot \vee \neg\Box\bot\,)
\end{aligned}
$$

   Applying $(\cdot)^{\diamondsuit}$, we find that $A \approx \Box\bot$.

3. Consider $A = (\neg\neg\Box\bot \to \Box\bot) \to (\Box\Box\bot \vee \neg\Box\Box\bot)$. Let $B := \neg\neg\Box\bot \to \Box\bot$. We compute A^*, using some shortcuts involving UP-principles.

$$
\begin{aligned}
A &\approx (\,\Box\bot \to (\Box\Box\bot \vee \neg\Box\Box\bot)\,)\ \wedge\\
&\qquad (\,[B]\neg\neg\Box\bot \vee [B]\Box\Box\bot \vee [B]\neg\Box\Box\bot\,)\\
&\approx \neg\neg\Box\bot \vee \Box\Box\bot \vee \neg\Box\Box\bot\\
&\approx \Box\bot \vee \Box\Box\bot \vee \neg\Box\Box\bot
\end{aligned}
$$

Applying $(\cdot)^{\Diamond}$, we find $A \approx \Box^2\bot$.

❑

12.5 Comparing Three Functors

In this last subsection of closed fragments, we compare the functors Φ_{HA}, Φ_{HA^*} and Φ_{PA}. By our previous results, it sufices to compare ∂_{HA}, ∂_{HA^*} and ∂_{PA}. We have:

$$\begin{array}{rcl} \alpha \leq \partial_{\mathsf{HA}} A & \Leftrightarrow & \mathsf{df}(\alpha) \vdash_{\mathsf{UP}} A \\ & \Rightarrow & \mathsf{df}(\alpha) \vdash_{\mathsf{UP}^c} A \\ & \Leftrightarrow & \alpha \leq \partial_{\mathsf{PA}} A \end{array}$$

Ergo, $\partial_{\mathsf{HA}} A \leq \partial_{\mathsf{PA}} A$. By a similar argument, we find $\partial_{\mathsf{HA}} A \leq \partial_{\mathsf{HA}^*} A$. Now consider the following formulas:

- $E_{\alpha,\beta,\gamma} := (\neg\neg\Box^{\alpha}\bot \wedge \Box^{\beta}\bot) \vee ((\neg\neg\Box^{\alpha}\bot \to \Box^{\alpha}\bot) \wedge \Box^{\gamma}\bot)$

Then we have, for $\alpha \leq \beta$ and $\alpha \leq \gamma$, that $\partial_{\mathsf{HA}} E_{\alpha,\beta,\gamma} = \alpha$, $\partial_{\mathsf{HA}^*} E_{\alpha,\beta,\gamma} = \beta$, and $\partial_{\mathsf{PA}} E_{\alpha,\beta,\gamma} = \gamma$. This shows that:

$$\begin{array}{c} \{\langle\alpha,\beta,\gamma\rangle\in\omega^{+3} \mid \exists A\in\mathsf{BDF}\ \partial_{\mathsf{HA}} A = \alpha,\ \partial_{\mathsf{HA}^*} A = \beta,\ \partial_{\mathsf{PA}} A = \gamma\} \\ = \\ \{\langle\alpha,\beta,\gamma\rangle\in\omega^{+3} \mid \alpha \leq \beta,\ \alpha \leq \gamma\}. \end{array}$$

12.6 Questions

1. What are the possible Λ_{T,e_T} for DF-sound RE extensions T of i-S^1_2? It is clear that they must be RE, DF-irreflexive BDF-theories extending UP.

2. Characterize the closed fragments of $\mathsf{HA} + \mathsf{ECT}_0$, $\mathsf{HA} + \mathsf{MP}$ and $\mathsf{HA} + \mathsf{ECT}_0 + \mathsf{MP}$. It seems to me that closer inspection of the results of the present paper should reveal that the closed fragment of $\mathsf{HA} + \mathsf{ECT}_0$ is the same as the closed fragment of HA.

3. Is there a T of the appropriate kind where $\vdash^{\mathsf{prov}}_{T}$ does not have a left adjoint?

References

[Bee75] M. Beeson. The nonderivability in intuitionistic formal systems of theorems on the continuity of effective operations. *The Journal of Symbolic Logic*, 40:321–346, 1975.

[Boo93] G. Boolos. *The logic of provability.* Cambridge University Press, 1993.

[Bur98] W. Burr. Fragments of Heyting-Arithmetic. To appear in the JSL, 1998.

[BV93] A. Berarducci and R. Verbrugge. On the provability logic of bounded arithmetic. *Annals of Pure and Applied Logic*, 61:75–93, 1993.

[dJ70] D.H.J. de Jongh. The maximality of the intuitionistic predicate calculus with respect to Heyting's Arithmetic. *The Journal of Symbolic Logic*, 36:606, 1970.

[dJ82] D.H.J. de Jongh. Formulas of one propositional variable in intuitionistic arithmetic. In A.S. Troelstra and D. van Dalen, editors, *The L.E.J. Brouwer Centenary Symposium*, Studies in Logic and the Foundations of Mathematics, vol. 110, pages 51–64. North Holland, Amsterdam, 1982.

[dJC95] D.H.J. de Jongh and L.A. Chagrova. The decidability of dependency in intuitionistic propositional logic. *The Journal of Symbolic Logic*, 60:495–504, 1995.

[dJV96] D.H.J. de Jongh and A. Visser. Embeddings of Heyting algebras. In *[HHST96]*, pages 187–213, 1996.

[Fri75] H. Friedman. One hundred and two problems in mathematical logic. *The Journal of Symbolic Logic*, 40:113–129, 1975.

[Fri78] H. Friedman. Classically and intuitionistically provably recursive functions. In G.H. Müller and D.S. Scott, editors, *Higher Set Theory*, pages 21–27. Springer Verlag, Berlin, Heidelberg, New York, 1978.

[Gav81] Yu. V. Gavrilenko. Recursive realizability from the inuitionistic point of view. *Soviet Mathematical Dokl.*, 23:9–14, 1981.

[Ghi99] S. Ghilardi. Unification in intuitionistic logic. *The Journal of Symbolic Logic*, 64:859–880, 1999.

[Gri87] R. Grigolia. *Free Algebras of Non-classical Logics*. Metsniereba Press, Tbilisi, 1987. In Russian.

[GZ95a] S. Ghilardi and M. Zawadowski. A sheaf representation and duality for finitely presented Heyting algebras. *Journal of Symbolic Logic*, 60:911–939, 1995.

[GZ95b] S. Ghilardi and M. Zawadowski. Undefinability of Propositional Quantifiers in the Modal System S4. *Studia Logica*, 55:259–271, 1995.

[Háj96] P. Hájek, editor. *Gödel '96, Logical Foundations of Mathematics, Computer Science and Physics —Kurt Gödel's Legacy*. Springer, Berlin, 1996.

[HHST96] W. Hodges, M. Hyland, C. Steinhorn, and J. Truss, editors. *Logic: from foundations to applications*. Clarendon Press, Oxford, 1996.

[Iem99] R. Iemhoff. On the admissible rules of Intuitionistic Propositional Logic. Preprint PP-1990-08, ILLC, Plantage Muidergracht 24, NL-1018TV Amsterdam, 1999. To appear in JSL.

[IemXX] R. Iemhoff. A modal analysis of some principles of the provability logic of Heyting Arithmetic. In *Proceedings of AiML'98*, volume 2, Uppsala, 20XX.

[JdJ98] G. Japaridze and D. de Jongh. The logic of provability. In S. Buss, editor, *Handbook of proof theory*, pages 475–546. North-Holland Publishing Co., amsterdam edition, 1998.

[Lei75] D. Leivant. *Absoluteness in intuitionistic logic*, volume 73. Mathematical Centre Tract, Amsterdam, 1975.

[Lei80] D. Leivant. Innocuous substitutions. *Journal of Symbolic Logic*, 45:363–368, 1980.

[Lei81] D. Leivant. Implicational complexity in intuitionistic arithmetic. *Journal of Symbolic Logic*, 46:240–248, 1981.

[Mac71] S. MacLane. *Categories for the Working Mathematician*. Number 5 in Graduate Texts in Mathematics. Springer, New York, 1971.

[Pit92] A. Pitts. On an interpretation of second order quantification in first order intuitionistic propositional logic. *Journal of Symbolic Logic*, 57:33–52, 1992.

[Pli77] V. E. Plisko. The nonarithmeticity of the class of realizable formulas. *Math. of USSR Izv.*, 11:453–471, 1977.

[Pli78] V. E. Plisko. Some variants of the notion of realizability for predicate formulas. *Math. of USSR Izv.*, 12:588–604, 1978.

[Pli83] V. E. Plisko. Absolute realizability of predicate formulas. *Math. of USSR Izv.*, 22:291–308, 1983.

[Ryb92] V.V. Rybakov. Rules of inference with parameters for intuitionistic logic. *Journal of Symbolic Logic*, 57:912–923, 1992.

[Ryb97] V. V. Rybakov. *Admissibility of logical inference rules*. Studies in Logic. Elsevier, Amsterdam, 1997.

[Sha93] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes mathematicae (Rozprawy matematycne)*, CCCXXIII, 1993.

[Smo73] C. Smoryński. Applications of Kripke Models. In A.S. Troelstra, editor, *Metamathematical Investigations of Intuitionistic Arithmetic and Analysis*, Springer Lecture Notes 344, pages 324–391. Springer, Berlin, 1973.

[Smo85] C. Smoryński. *Self-Reference and Modal Logic*. Universitext. Springer, New York, 1985.

[Sta79] R. Statman. Intuitionistic propositional logic is polynomial-space complete. *Theoretical Computer Science*, 9:67–72, 1979.

[Tro73] A.S. Troelstra. *Metamathematical investigations of intuitionistic arithmetic and analysis.* Springer Lecture Notes 344. Springer Verlag, Berlin, 1973.

[TvD88a] A.S. Troelstra and D. van Dalen. *Constructivism in Mathematics, vol 1.* Studies in Logic and the Foundations of Mathematics, vol. 121. North Holland, Amsterdam, 1988.

[TvD88b] A.S. Troelstra and D. van Dalen. *Constructivism in Mathematics, vol 2.* Studies in Logic and the Foundations of Mathematics, vol. 123. North Holland, Amsterdam, 1988.

[vB95] J.F.A.K. van Benthem. Temporal Logic. In Dov Gabbay et al., editor, *Handbook of Logic in Artificial Intelligence and Logic Programming.* Oxford University Press, 1995.

[Ver93] L.C. Verbrugge. *Efficient metamathematics.* ILLC-disseration series 1993-3, Amsterdam, 1993.

[Vis81] A. Visser. *Aspects of diagonalization and provability.* Ph.D. Thesis, Department of Philosophy, Utrecht University, 1981.

[Vis82] A. Visser. On the completeness principle. *Annals of Mathematical Logic*, 22:263–295, 1982.

[Vis85] A. Visser. Evaluation, provably deductive equivalence in Heyting's Arithmetic of substitution instances of propositional formulas. Logic Group Preprint Series 4, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, 1985.

[Vis94] A. Visser. *Propositional combinations of* Σ*-sentences in Heyting's Arithmetic.* Logic Group Preprint Series 117. Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, 1994.

[Vis96] A. Visser. Uniform interpolation and layered bisimulation. In *[Háj96]*, pages 139–164, 1996.

[Vis98a] A. Visser. An Overview of Interpretability Logic. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakharyaschev, editors, *Advances in Modal Logic, vol 1*, CSLI Lecture Notes, no. 87, pages 307–359. Center for the Study of Language and Information, Stanford, 1998.

[Vis98b] A. Visser. Submodels of Kripke models. Logic Group Preprint Series 189, Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, `http://www.phil.uu.nl/preprints.html`, 1998. To appear in Archive for Mathematical Logic.

[Vis99] A. Visser. Rules and Arithmetics. *The Notre Dame Journal of Formal Logic*, 40(1):116–140, 1999.

[vO91] J. van Oosten. A semantical proof of de Jongh's theorem. *Archive for Mathematical Logic*, 31:105–114, 1991.

[VvBdJdL95] A. Visser, J. van Benthem, D. de Jongh, and G. Renardel de Lavalette. NNIL, a Study in Intuitionistic Propositional Logic. In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra, a Bisimulation Perspective*, CSLI Lecture Notes, no. 53, pages 289–326. Center for the Study of Language and Information, Stanford, 1995.

[Yav97] R.E. Yavorsky. Logical schemes for first order theories. In *Springer LNCS (Yaroslavl'97 volume)*, volume 1234, pages 410–418, 1997.

[Zam94] D. Zambella. Shavrukov's theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + \mathsf{EXP}$. *The Notre Dame Journal of Formal Logic*, 35:147–157, 1994.

A Adjoints in Preorders

The consequence relations we consider are preorders. A preorder is a non-emty set wth a reflexive transitive relation. It is a special kind of category, viz. a category where, between any two objects, we have at most one arrow. In the present paper, we will use some facts from category theory and some facts specific for preorders. We collect most of these facts in the following lemma. The easy proofs are left to the reader.

Lemma A.1 Suppose $\leq$ is a preorder on C and that $\sqsubseteq$ is a preorder on D. Say, $\simeq$ is the induced equivalence relation of $\leq$ and $\equiv$ is the induced equivalence relation of $\sqsubseteq$. Let $L : D \to C$ be the left adjoint function to $R : C \to D$, that is $Ld \leq c \Leftrightarrow d \sqsubseteq Rc$, for every $d \in D$ and every $c \in C$. Then the following holds.

1. $LRc \leq c$ and $d \sqsubseteq RLd$.
2. L and R are order preserving, in other words, L and R are mophisms of preorders, and thus functors of the corresponding categories.
3. $Rc \equiv RLRc$ and $Ld \simeq LRLd$.
4. Suppose that R is a surjection on objects. Then, for any $d \in D$, $d \equiv RLd$.
5. Suppose R is a surjection on objects. Let X be the range of L. Let $\leq_0$ be the full sub-preorder of $\leq$ obtained by restricting $\leq$ to X. Then, $\leq_0$ is equivalent to $\sqsubseteq$. This means that, for $x \in X$, $x \simeq LRx$ and, for $d \in D$, $d \equiv RLd$.
6. $L \circ R$ is uniquely determined modulo $\simeq$ by $X := \mathsf{range}(L)$.
7. Suppose we have sums on the preorders $\leq$ and $\sqsubseteq$, i.e. there are binary functions $\vee$ and $+$ such that
   - $c_1 \leq c'$ and $c_2 \leq c'$ iff $(c_1 \vee c_2) \leq c'$,
   - $d_1 \sqsubseteq d'$ and $d_2 \sqsubseteq d'$ iff $(d_1 + d_2) \sqsubseteq d'$.

   Then $L(d + e) \simeq Ld \vee Le$.

❑

Let $\rhd$ be a semi-consequence relation on $\mathcal{L}(\mathcal{P})$. Both $\langle \mathcal{L}(\mathcal{P}), \vdash \rangle$ and $\langle \mathcal{L}(\mathcal{P}), \rhd \rangle$ can be considered as preorders. A1 tells us that the identical mapping $\Im : A \mapsto A$ is an embedding from $\langle \mathcal{L}(\mathcal{P}), \vdash \rangle$ to $\langle \mathcal{L}(\mathcal{P}), \rhd \rangle$. In this paper we are interested in cases where $\Im$ has a left adjoint L. If such a left adjoint of the embedding functor exists, we call $\langle \mathcal{L}(\mathcal{P}), \vdash \rangle$ a *reflective subcategory* of $\langle \mathcal{L}(\mathcal{P}), \rhd \rangle$. The functor L is called *the reflector*. (See [Mac71] for more on these notions.) Note that $\Im$ is also surjective on objects.

Suppose $\rhd$ is a nearly-consequence relation. B1 tells us that $\vee$ is a sum for the category $\rhd$. Since left adjoints commute with sums, by lemma A.1(7), we have: $\vdash L(A \vee B) \leftrightarrow (LA \vee LB)$

B Characterizations and Dependencies

Consider any consistent RE theory T, extending HA, in the language of HA. We introduce some notions, closely related to Σ-preservativity for T. Define, suppressing the index for T, the following consequence relations:

Provable Deductive Cons.	$A \rhd_{\mathsf{pdc}} B$	$:\leftrightarrow$	$\Box(\Box A \to \Box B)$
Uniform Deductive Cons.	$A \rhd_{\mathsf{udc}} B$	$:\leftrightarrow$	$\forall x \exists y\ (\Box_x A \to \Box_y B)$
Strong Uniform Deductive Cons.	$A \rhd_{\mathsf{sudc}} B$	$:\leftrightarrow$	$\forall x\ (\Box_x A \to \Box_x B)$
Uniform Provably Deductive Cons.	$A \rhd_{\mathsf{updc}} B$	$:\leftrightarrow$	$\forall x \exists y\ \Box(\Box_x A \to \Box_y B)$

We prove the Orey-Hájek characterization for Σ-preservativity and Uniform Provable Deductive Consequence.

Theorem B.1 (Orey-Hájek) *T proves that the following are equivalent:* *(i) $A \rhd_\Sigma B$, (ii) $\forall x\ \Box(\Box_x A \to B)$, (iii) $A \rhd_{\mathsf{updc}} B$.*

Proof

Reason in T. "(i)→(ii)" Suppose $A \rhd_\Sigma B$ and consider any x. We have, by theorem 8.1, $\Box(\Box_x A \to A)$ and $(\Box_x A) \in \Sigma$, hence $\Box(\Box_x A \to B)$.

"(ii)→(iii)" From $\Box(\Box_x A \to B)$, we have that, for some y, $\Box_y(\Box_x A \to B)$ and hence $\Box\Box_y(\Box_x A \to B)$. Ergo: $\Box(\Box_y \Box_x A \to \Box_y B)$. Clearly, for any u, $\Box(\Box_u A \to \Box_y \Box_u A)$. We may conclude $\Box(\Box_x A \to \Box_y B)$. Hence, $A \rhd_{\mathsf{updc}} B$.

"(i)←(iii)" Suppose $A \rhd_{\mathsf{updc}} B$ and $\Box(S \to A)$. It follows that, for some x, $\Box\Box_x(S \to A)$. Moreover, $\Box(S \to \Box_x S)$. So $\Box(S \to \Box_x A)$. Hence, for some y, $\Box(S \to \Box_y B)$. Ergo, by reflection, $\Box(S \to B)$. ❑

Theorem B.2 *We have: $T \vdash A \rhd_\Sigma B \to A \rhd_{\mathsf{udc}} B$.*

Proof

Reason in T. Suppose $A \rhd_\Sigma B$. Consider any x. We can find a y such that we have $\Box_y(\Box_x A \to B)$. Suppose $\Box_x A$. Clearly, for any u, $\Box_u A \to \Box_y \Box_u A$. It follows that $\Box_y B$. ❑

Theorem B.3 (Orey-Hájek for HA^*) *We have:*

$$\mathsf{HA}^* \vdash A \rhd_{\mathsf{HA}^*,\Sigma} B \leftrightarrow A \rhd_{\mathsf{udc},\mathsf{HA}^*} B.$$

Proof

Immediate by the principle C, theorem B.1 and theorem B.2. ❑

Theorem B.4 $\mathsf{HA}^* \vdash (\Box_{\mathsf{HA}^*} A \to B) \to A \rhd_{\mathsf{HA}^*,\Sigma} B$.

Proof

Reason in HA^*. Suppose $\Box A \to B$. It follows that $\forall x\ (\Box_x A \to B)$. Hence, $\forall x\ \Box(\Box_x A \to B)$. Thus we may conclude, $A \rhd_\Sigma B$. ❑

C Modal Logic for Σ-Preservativity

Consider the language of modal propositional logic with a unary modal operator $\Box$ and a binary operator $\rhd$. We define arithmetical interpretations in the language of HA in the usual manner, interpreting $\Box$ as provability in HA and $\rhd$ as Σ-preservativity over HA. We state the principles valid in HA for this logic, known at present. With the exception of Σ4 and Σ8 these principles are the duals of the principles of the interpretability logic ILM. For a discussion of this duality, see [Vis98a].

L1 $\vdash A \Rightarrow \vdash \Box A$

L2 $\vdash \Box(A \to B) \to (\Box A \to \Box B)$

L3 $\vdash \Box A \to \Box\Box A$

L4 $\vdash \Box(\Box A \to A) \to \Box A$

Σ1 $\vdash \Box(A \to B) \to A \rhd B$

Σ2 $\vdash A \rhd B \wedge B \rhd C \to A \rhd C$

Σ3 $\vdash C \rhd A \wedge C \rhd B \to C \rhd (A \wedge B)$

Σ4 $\vdash A \rhd C \wedge B \rhd C \to (A \vee B) \rhd C$

Σ5 $\vdash A \rhd B \to (\Box A \to \Box B)$

Σ6 $\vdash A \rhd \Box A$

Σ7 $\vdash A \rhd B \to (\Box C \to A) \rhd (\Box C \to B)$

Σ8 Let X be a finite set of implications and let

$$Y := \{C \mid (C \to D) \in X\} \cup \{B\}.$$

Take $A := \bigwedge X$. Then, $\vdash (A \to B) \rhd \{A\}Y$. Here $(A \to B) \rhd \{A\}Y$ is short for $(A \to B) \rhd \bigvee\{A\}Y$ and $\{C\}D$ is defined as follows:

- $\{C\}p := (C \to p)$, $\{C\}\bot := \bot$, $\{C\}\top := \top$,
- $\{C\}\Box E := \Box E$, $\{C\}(E \rhd F) = (C \to (E \rhd F))$, $\{C\}(E \to F) := (C \to (E \to F))$,
- $\{C\}(\cdot)$ commutes with $\wedge$ and $\vee$.

The verification of L1-L4, Σ1-Σ3, Σ5 – Σ7 is routine. For Σ4 see 9.1 and for Σ8 see 9.2.

From our principles Leivant's principle can be derived. (This is one of the Stellingen of [Lei75].)

Le $\vdash \Box(A \vee B) \to \Box(A \vee \Box B)$

We leave this as an exercise to the reader. It is open whether our axioms are arithmetically complete.