chapters on bounded arithmetic

&

on provability logic

domenico zambella

chapters on bounded arithmetic

&

on provability logic

ILLC Dissertation Series 1994-6

institute for logic, language and computation

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
phone: +31-20-5256090
fax: +31-20-5255101
e-mail: illc@fwi.uva.nl

chapters on bounded arithmetic

&

on provability logic

Academisch Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit van Amsterdam,
op gezag van de Rector Magnificus
Prof.dr P.W.M. de Meijer
in het openbaar te verdedigen in de
Aula der Universiteit
(Oude Lutherse Kerk, ingang Singel 411, hoek Spui)
op dinsdag 6 september 1994 te 14.00 uur

door

Domenico Zambella

geboren te Monselice

Promotor: prof. dr. A. Troelstra

Co-promotores: dr. D. H. J. de Jongh

dr. A. Visser

Faculteit der Wiskunde en Informatica
Plantage Muidergracht 24
1018 TV Amsterdam

ISBN: 90–74795–10–2

Contents

Preface

Part I. Bounded arithmetic.

Chapter 1. Notes on polynomially bounded arithmetic.

Chapter 2. End extensions of linearly bounded arithmetic.

Part II. Provability logic.

Introduction.

Chapter 3. On the proofs of arithmetical completeness for interpretability logic.

Chapter 4. Shavrukov's theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + exp$.

Samenvatting

Preface

Monolithic this thesis is not. It is neatly divided in two parts. Each part consists of two articles. The first part deals with theories of weak arithmetic. I.e., fragments of Peano arithmetic that do not prove the totality of exponentiation. We hope to provide sufficient evidence that both technically and heuristically it is useful to interpret these fragments as second-order theories. The first article contains a general introduction and a brief discussion on the 'philosophical' motivations. Both articles are meant to be reasonably self-contained. Some general familiarity with (first-order) models of arithmetic is the only prerequisite to part I. The reader can refresh her memory by consulting the book of R. Kaye, *Models of Peano Arithmetic.* Oxford University Press, Oxford (1991).

Chapter 3 and 4 of the second part are reprints of D. Zambella, On the proofs of arithmetical completeness for interpretability logic, *Notre Dame Journal of Formal Logic*, vol. 35 (1992) pp.542-551 and of D. Zambella, Shavrukov's theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + exp$, *Notre Dame Journal of Formal Logic*, vol. 35 (1994) pp. 147-157. Part II contains a short introduction to these articles. Finally, for connections between the first and the second part of this thesis we would like to refer the reader to R. Verbrugge, *Efficient Metamathematics*, Ph.D. Thesis, Universiteit van Amsterdam, ILLC Dissertation series, 1993-3, (1993).

Each chapter of this thesis may be read separately and contains a separate list of references as well as scientific acknowledgements. My supervisors, Dick de Jongh and Albert Visser have helped me in many different ways during the last four years. A thanks is due also to Professor A. Troelstra who kindly areed to be my official promotor. Finally, I want to express my gratitude to all colleagues who in one way or another have contributed to create a fruitful atmosphere around me in the last few years. In particular I would like to mention Dick, Albert, Michiel, Rineke, Volodya, Marc, Harry, Leen, Peter, Andreja, Maarten, Erik, Martijn and Bas.

Part I. Bounded arithmetic

Chapter 1. Notes on polynomially bounded arithmetic

Abstract

We characterize the collapse of Buss' bounded arithmetic in terms of the provable collapse of the polynomial time hierarchy. We include also some general model-theoretical investigations on fragments of bounded arithmetic.

Contents

0 Introduction and motivation. **2**

1 Preliminaries. **4**

1.1 The polynomially bounded hierarchy. 4
1.2 The axioms of second-order bounded arithmetic. 5
1.3 Rudimentary functions. 6
1.4 Other fragments. 7
1.5 Polynomial time computable functions. 7
1.6 Relations among fragments. 9
1.7 Relations with Buss' bounded arithmetic. 11

2 Witnessing theorems and conservativity results. **12**

2.1 Closures 12
2.2 A model-theoretical version of Buss' witnessing theorem. 14
2.3 A model theoretical characterization of choice. 15
2.4 Ultrapowers 20

3 The collapse of BA versus the collapse of PH **23**

3.1 An interpolation theorem 23
3.2 Sufficient conditions for the collapse of BA 25
3.3 Necessary conditions for the collapse of BA 25
3.4 Krajíček, Pudlák and Takeuti's method 28

0 Introduction and motivation.

In every model of $I\Delta_0$ numbers code finite sets. Sets coded by numbers are Δ_0-definable. In general, the converse is not true. Weak theories, which do not prove the totality of exponentiation, do not prove the existence of a code for every finite Δ_0-definable set. So, a natural way of strengthening $I\Delta_0$ is by adding to the language second-order variables X, Y, Z, etc. ranging over finite sets of numbers and introducing axioms of finite comprehension ensuring the existence of sets of the form $\{x<a : \varphi(x)\}$ for $\varphi(x)$ ranging over some class of second-order formulas. Interesting theories arise when we restrict the schema of finite comprehension to bounded formulas. These are formulas where all quantifiers are of the form $Qx<t$ or $QX<t$ where t is a first-order term (i.e., a polynomial). Note that second-order bounded quantifiers range over sets whose elements are bounded by t, so, by the absence of exponentiation, their nature is radically different from that of first-order quantifiers. We introduce the classes Σ_i^p and Π_i^p counting alternations of (polynomially) bounded second-order quantifiers. Restricting the strength of the schema of finite comprehension to formulas of a certain complexity one obtains the hierarchy of theories that we call Σ_i^p-*comp*. The union of all these theories (i.e., finite comprehension for all bounded second-order formulas) is called second-order bounded arithmetic *BA*. We study the relative strength of various fragments of *BA* and in particular their provably total functions. Interestingly, all provably recursive functions of *BA* are of polynomial growth. In this article we prove some theorems of partial conservativity are proved for some of these theories and the connection with complexity theory is briefly discussed.

In the last decades two subsystems of arithmetic, $I\Delta_0$ and S_2, have been studied especially for their connections with complexity theory (see e.g., [17] and [3] or [8]). In particular, Buss' S_2 is the most extensively studied. The theory S_2 coincides with (an extension by definition of) the equally well-known $I\Delta_0+\Omega_1$. These theories are first-order strengthenings of $I\Delta_0$. In the case of $I\Delta_0+\Omega_1$ or S_2 the motivation for the strengthening is somehow technical; it arises from metamathematical and/or syntactical considerations. In fact, in order to have a reasonable formalization of computation and/or syntax one needs to be able to perform operations on strings such as the substitution of substrings. Such operations increases the code of the string superpolynomially and so, this is not provably total in $I\Delta_0$. Adding to $I\Delta_0$ an axiom (i.e., Ω_1) asserting the totality of this function one obtains a stronger theory in which it is possible to formalize almost all basic notions of metamathematics. Buss introduced a hierarchy of theories S_2^i whose union is S_2. These theories are obtained by some weakening of the axiom of induction (while introducing sufficiently many new primitives to allow smooth bootstrapping).

It is not surprising that *BA* coincides with Buss' S_2, modulo an appropriate translation. Namely, to each (first-order) model $\mathbf{M}'$ of S_2 corresponds a (second-order) model $\mathbf{M}''$ of *BA*. The first-order objects of $\mathbf{M}''$ are the logarithmic numbers of $\mathbf{M}'$ (i.e., numbers belonging to the domain of exponentiation). The smash function guarantees that these numbers are closed under multiplication. The second-order objects of $\mathbf{M}'$ are those finite sets which have a code in $\mathbf{M}'$. In this way, Σ_i^p-formulas get transformed into Σ_i^b-formulas of Buss' language (see e.g., [3] or chapter V of [8]) in a very natural way, so, the constructed second-order model verifies finite comprehension for all bounded formulas. Vice versa, from a model $\mathbf{M}''$ of *BA* one obtains a (first-order) model $\mathbf{M}'$ of S_2 by the inverse procedure. As domain of $\mathbf{M}'$ we take the second-order objects of $\mathbf{M}''$. In $\mathbf{M}''$ we define the primitives of S_2 as set operations.

Intuitively, we think of a finite set X as the numbers $\sum_{x\in X} 2^x$ and define operations lead by this idea. We shall see that BA disposes over enough second-order recursion to formalize these operations and to prove that the axioms of S_2 to hold in $\mathbf{M}'$. Note, parenthetically, that the cartesian product of two sets is mapped to a first-order function with the growth rate of the smash function. This procedure actually maps models of Σ^p_i-*comp* into models of S^i_2 and vice versa (for all $i>0$). A few details on this isomorphism (which was discovered in different ways by many authors) are contained in Section 1.7. Readers who are mainly interested in S^i_2 are advised to read that section first. In fact, afterwards they will be able to translate most of the results reported here into theorems about fragments of S_2. In particular, Lemma 2.2 is a strengthening of the main theorem of [3]. Our proof is model-theoretic and it is formally identical to an unpublished model-theoretic argument for the conservativity of $I\Sigma_1$ over PRA by Albert Visser. In fact, formal similarities between $I\Sigma_1$ and Σ^p_i-*comp* are apparent when primitive recursive functions are replaced by polynomial time computable functions. Other conservativity results are obtainable with the same method. The author's personal motivation for using a second-order framework is that this approach allows economy of primitives, natural definitions and (again in the author's opinion) a clear heuristic.

In the hierarchy of fragments of BA very few inclusions are known to be strict. In general the problem of proving inclusions to be strict seems to be a very difficult one. A more realistic goal is to characterize the collapse of theories in terms of the provable collapse of some complexity classes. A corollary of Lemma 2.2 is that, if $\mathcal{P}$-*def* (i.e., the $\forall\Sigma^p_1$ fragment of $\Sigma^p_1(\equiv S^1_2)$) proves $\Sigma^p_2 = \Pi^p_2$, then all of BA collapses to $\mathcal{P}$-*def*. So, a very satisfactory result would be to prove the converse. One of the best known results in this direction is the celebrated KPT theorem (see Theorem 3.4): in [12] Krajíček, Pudlák and Takeuti, proved that if $\mathcal{P}$-*def* proves Σ^p_1-*comp*, then in the standard model the polynomial time hierarchy collapses to the second level. Unfortunately, it is still unclear whether their proof is formalizable in BA, so, their result cannot be used to answer questions like: if $\mathcal{P}$-*def* proves Σ^p_1-*comp* does BA collapse?

The main achievement of this paper is the following theorem. It gives a satisfactory characterization of the collapse of BA in terms of the provable collapse of PH. (On the right hand side we include the translation into Buss' language. For uniformity, we set $T^0_2 := PV_1$. For the definition of $BB\Sigma^b_{i+1}$ see [8].)

Theorem. The following are equivalent

(i)	$\mathcal{P}_i$-*def* $\vdash \Sigma^p_{i+1}$-*comp*	$T^i_2 \vdash S^{i+1}_2$
(ii)	$\mathcal{P}_i$-*def* $\vdash \Sigma^p_{i+1}\subseteq\Pi^p_{i+1}/poly$	$T^i_2 \vdash \Sigma^b_{i+1}\subseteq\Pi^b_{i+1}/poly$
(iii)	$\mathcal{P}_i$-*def* $\vdash BA$	$T^i_2 \vdash S_2$
(iv)	$\mathcal{P}_i$-*def* $+ \Sigma^p_{i+1}$-*choice* $\vdash \Sigma^p_{i+1}$-*comp*	$T^i_2 + BB\Sigma^b_{i+1} \vdash_0 S^{i+1}_2$

The implication from **(i)** to **(ii)** is Theorem 3.3. The implication from **(i)** to **(iii)** can be reconstructed from the proof of Theorem 3.2. (To read these two proofs the reader needs only to rush through Section 1.) From Theorem 3.2 it actually follows that **(ii)** implies **(iii)** while in Corollary 2.3 is proved that **(iv)** implies **(i)**.

Acknowledgments. The numerous discussions with Rineke Verbrugge, Harry Buhrman and Volodya Shavrukov have been pleasant and stimulating. When the first draft of this manuscript was ready I had interesting discussions with Sam Buss. I owe him various observations and corrections. Buss independently proved [5] that condition **(i)** above implies **(ii)** and **(iii)**. He observed that from **(i)** it follows that PH (provably) collapses to $Boole(\Sigma^p_{i+2})$. His result inspired the interpolation theorem of Section 3.1. The supervision of Dick de Jongh and Albert Visser has assisted me through the numerous stadia of preparation of this work.

1 Preliminaries.

Here we introduce the necessary definitions. Lemma 1.3 provides a smooth bootstrapping. The class of polynomial time computable functions is concisely introduced in section 1.5 in a machine independent way. The (standard) comparison of strength of the various fragments is sketched in Section 1.6. In Section 1.7 the relation with Buss' S^i_2 is sketched.

1.1 The polynomially bounded hierarchy.

We define the analogue of the analytical hierarchy for finite sets. The language L_2 is the language of second-order arithmetic; it consists of two symbols for constants: 0, 1, two symbols for binary functions: $+$, $\cdot$ and two symbols for binary relations: $<$, $\in$. Moreover, there are two sorts of variables: first and second-order. Lower case Latin letters $x, y, z, ..$ denote first-order variables and capital Latin letters X, Y, Z, .. second-order variables. First and second-order variables are meant to range respectively over numbers and finite sets of numbers. Terms are constructed from first-order variables only. The formula $x{<}y$ is to be read "x is less than y". The intended meaning of $X{<}y$ is: "all elements of X are less than y". Let t be a term of L_2 in which x does not occur. We adopt the following abbreviations with the usual meaning

$$(Qx{<}t)\varphi,\ (Qx{\in}Y)\varphi,\ (QX{<}t)\varphi,$$

where Q is either $\forall$ or $\exists$. Quantifiers occurring in either of these contexts are called **(polynomially) bounded quantifiers**. The class of bounded formulas is denoted by PH. Note that first-order quantifiers range over elements of sets while second-order quantifiers range over subsets of sets. Here, first-order bounded quantifiers play the role that **sharply bounded quantifiers** have in first-order bounded arithmetic (see e.g., [3] or chapter V of [8]).

A formula is **(polynomially) bounded** if all of its quantifiers are. Counting alternations of second-order quantifiers we classify bounded formulas in the **(polynomially) bounded hierarchy**. We use either one of the symbols Π^p_0 or Σ^p_0 for formulas containing only bounded first-order quantifiers. We define inductively Σ^p_{i+1} as the minimal class of formulas containing Π^p_i, closed under disjunction, conjunction and bounded existential quantification. The class Π^p_{i+1} is the minimal class of formulas containing Σ^p_i, closed under disjunction, conjunction and bounded universal quantification. So, PH equals $\bigcup_{i\in\omega}\Sigma^p_i$ and

$\bigcup_{i\in\omega} \Pi^p_i$.

The class $\Sigma^p_0(\Sigma^p_i)$ is the smallest set of formulas containing Σ^p_i, closed under Boolean operations and bounded first-order quantification. Sometimes we add to the language L_2 some set $\mathcal{F}$ of new symbols for functions. We define the (relativized) classes of bounded $L_2(\mathcal{F})$-formulas: $\Sigma^p_i(\mathcal{F})$, $\Pi^p_i(\mathcal{F})$, etc. similarly to those of the language L_2. (We allow terms of $L_2(\mathcal{F})$ to occur in the bounds of the quantifiers.)

The domain of an L_2 structure $\mathbf{M}$ is composed of two disjoint parts: the numbers and the sets of $\mathbf{M}$. Truth in $\mathbf{M}$ is defined as usual but first-order variables are restricted to range over numbers while second order variables range over sets. To denote elements of a model, we use the same convention as for variables, so, we write $A\in\mathbf{M}$ for 'A is a set of $\mathbf{M}$' and $a\in\mathbf{M}$ for 'a is a number of $\mathbf{M}$'. For models we use bold face capitals, for the class of first-order objects of a model $\mathbf{M}$ we use the corresponding lower case bold face letter $\mathbf{m}$. The disjoint union of ω and $\mathcal{P}_{<\omega}(\omega)$ constitutes the **standard model**, functions and relations are interpreted in the natural way. We loosely denote the standard model by ω.

Computational complexity theory and second-order arithmetic are our main sources of inspiration, concrete intuition and terminology. For our digressions to computational complexity theory it is convenient to think of finite sets as strings i.e., we identify $\mathcal{P}_{<\omega}(\omega)$ and $2^{<\omega}$. So, sets of finite sets may be identified with languages. The actual form of the isomorphism is immaterial. We stipulate that the length of the string associated to a finite set $X\subseteq\omega$ equals (up to some additive constant) the least upper bound of the set X which we henceforth denote by $|X|$. To begin with, the reader may wish to check that Σ^p_1-formulas define languages in NP, i.e., if $\varphi(X)\in\Sigma^p_1$ then the language $\{X \ : \ \omega \models \varphi(X)\}$ is in NP. Vice versa for every language $L\subseteq 2^{<\omega}$ in NP there is a formula $\varphi(X)$ in Σ^p_1 such that L is $\{X \ : \ \omega \models \varphi(X)\}$. In the same way, Π^p_1-formulas coincide with *coNP* languages and, in general, each level of the bounded hierarchy coincides with one of the Meyer-Stockmeyer polynomial time hierarchy (with the only exception of ground level $i = 0$ which corresponds to uniform-AC^0 languages). When digressing to computational complexity theory, we identify each number $x\in\omega$ with the set of its predecessors and so, with a string of ones of length x. Therefore, a formula $\varphi(x)$ with one free first-order variable defines a tally language i.e., a language which is contained in $\{1\}^{<\omega}$.

1.2 The axioms of second-order bounded arithmetic.

The theory Θ is axiomatized by the following formulas: (The expressions $a \leq b$, $A = \emptyset$ and $A\subseteq B$ stand for the usual abbreviations.)

hskip1cm$0 \neq 1$	$a.(b+1) = (a.b) + a$
$a + 0 = a$	$a \leq b \leftrightarrow a < b + 1$
$a + 1 = b + 1 \rightarrow a = b$	$a \leq b + 1 \leftrightarrow a < b$
$a + (b+1) = (a+b) + 1$	$A < b \leftrightarrow (\forall x\in A)\ x < b$
$a \neq 0 \leftrightarrow (\exists x < a)\ x + 1 = a$	$A = B \leftrightarrow A\subseteq B \wedge B\subseteq A$
$a.0 = 0$	$A \neq \emptyset \rightarrow (\exists x\in A)(A < x + 1)$

These are the axioms of Robinson arithmetic plus the defining axioms for the relation $<$, the axiom of extensionality, the least number principle and the axioms of finiteness (i.e., all sets have an upper bound). The theory Σ_i^p-*comp* is axiomatized by Θ and the schema of **(finite) comprehension** for Σ_i^p-formulas i.e., for all φ in Σ_i^p in which X does not occur free,

$$\Sigma_i^p\text{-}comp : (\exists X{<}a)(\forall x{<}a)\ [x{\in}X \ \leftrightarrow\ \varphi(x)].$$

The theory of **second-order bounded arithmetic**, BA, is the union of Σ_i^p-*comp* for $i{\in}\omega$. The theories Π_i^p-*comp* and $\Sigma_0^p(\Sigma_i^p)$-*comp* are defined in a similar way and are easily seen to be equivalent to Σ_i^p-*comp*.

1.3 Rudimentary functions.

In order to keep formulas to a readable size we need to introduce new function symbols. To begin with, let us give some informal definitions. We write $|A|$ for the least upper bound of A and $|\vec{a},\vec{A}|$ for the least upper bound of $\{1, a_1, .., a_n, |A_1|, .., |A_m|\}$. It should be clear that Σ_0^p-*comp* suffices to prove the existence of $|\vec{a},\vec{A}|$. We call **rudimentary** those functions which are obtained by Σ_0^p comprehension or by Σ_0^p minimalization, i.e., those functions definable in either one of the two following ways:

$$F_{\varphi,p}(\vec{a},\vec{A}) := \{x{<}|\vec{a},\vec{A}|^p \ :\ \varphi(x,\vec{a},\vec{A})\}, \qquad f_{\varphi,p}(\vec{a},\vec{A}) := \mu_{x<|\vec{a},\vec{A}|^p}\ \varphi(x,\vec{a},\vec{A}),$$

for some $\varphi{\in}\Sigma_0^p$ and $p{\in}\omega$ (in the definition of $F_{\varphi,p}$ and $f_{\varphi,p}$, we have stressed that these functions are polynomially bounded).

Let $\mathcal{R}$ be a set of new primitives, one for each (definition of a) rudimentary function. Let $\mathcal{R}$-*def* be the theory axiomatized by Θ plus the (obvious) defining axioms for the functions in $\mathcal{R}$. Clearly, Σ_0^p-*comp* suffices to prove every rudimentary function to be total. So, $\mathcal{R}$-*def* is a conservative expansion of Σ_0^p-*comp*. The following lemma ensures us that there is no danger in considering formulas of the expanded language $L_2(\mathcal{R})$ as abbreviations of L_2-formulas. In fact, the 'translation' does not increase the complexity of the formula. Namely, the following lemma shows that $\Sigma_0^p = \Sigma_0^p(\mathcal{R})$ provably in $\mathcal{R}$-*def*.

Lemma 1.3. *For every $\psi{\in}\Sigma_0^p(\mathcal{R})$, there is $\psi^*{\in}\Sigma_0^p$ such that $\mathcal{R}$-def $\vdash \psi \leftrightarrow \psi^*$.*

Proof. The lemma is proved by a method which we believe to be well-known to the reader, so, we do not need give it in full detail. One has to unfold the definitions of the rudimentary functions inside the $\Sigma_0^p(\mathcal{R})$-formulas ψ. We can assume that ψ has only one occurrence of a single rudimentary function $F_{\varphi,p}(\vec{a},\vec{A})$ (we also assume this function is a set function; the case of a number function is similar). First, one must rewrite ψ to have all occurrences of rudimentary set functions on the right of the symbol $\in$. Then replace each subformula of the form $x{\in}F_{\varphi,p}(\vec{a},\vec{A})$ with

$$x{<}|\vec{a},\vec{A}|^p \wedge \varphi(x,\vec{a},\vec{A}).$$

Finally, replace subformulas of the form $x{<}|\vec{a},\vec{A}|^p$ with an equivalent Σ_0^p-formula. The defining axioms of $F_{\varphi,p}$ ensure that the formula obtained is equivalent to the original ψ. In the resulting formula no rudimentary set functions occur. ∎

A noteworthy corollary of this lemma is that rudimentary functions are closed under composition. From the Lemma it follows also that $\Sigma_i^p(\mathcal{R})$-*comp* + $\mathcal{R}$-*def* is equivalent to Σ_i^p-*comp* + $\mathcal{R}$-*def* and hence an extension by definitions of Σ_i^p-*comp*. Below, we list a few rudimentary functions that we often use.

$\langle a, b\rangle := \mu_z\, 2z = (a+b)(a+b+1)$, the pairing function,

$A \times B := \{\langle x, y\rangle \ : \ x{\in}A \wedge y{\in}B\}$, the cartesian product,

$A^{[b]} := \{y \ : \ \langle b, y\rangle{\in}A\}$, the b-th row of the 'matrix' A,

$A(b) := \mu_z\ z{\in}A^{[b]}$, the value of the 'function' A at b,

$[x] := \{y \ : \ y{<}x\}$, the set of predecessors of x,

$\{x\} := \{y \ : \ x = y\}$, the singleton of x.

1.4 Other fragments.

In this section we present some other interesting axiomatizations of BA. In the next sections we study the relative strength of their fragments. We agree that all theories we introduce in this section contain, by definition, Σ_0^p-*comp*. The theories Σ_i^p-*ind*, Σ_i^p-*dc* and Σ_i^p-*coll* (i.e., of **induction, dependent choice** and **strong collection** for Σ_i^p-formulas) are axiomatized by the following schemas, for $\varphi{\in}\Sigma_i^p$.

Σ_i^p-*ind* : $\varphi(0) \wedge \forall x[\varphi(x) \rightarrow \varphi(x+1)] \ \rightarrow \varphi(a)$,

Σ_i^p-*dc* : $\forall x(\forall X{<}b)(\exists Y{<}b)\varphi(x, X, Y) \ \rightarrow \exists Z(\forall x{<}a)\varphi(x, Z^{[x]}, Z^{[x+1]})$

Σ_i^p-*coll* : $\exists Z\ (\forall x{<}a)[(\exists Y{<}b)\varphi(x, Y) \rightarrow \varphi(x, Z^{[x]})]$

(in the last two schemas Z should not occur free in φ). The schema of dependent choices is inspired by second-order arithmetic. We show (cf. Lemma 1.6) that dependent choice, induction and strong collection are all equivalent to comprehension. A rather intriguing role is played by the following schema of **choice**

Σ_i^p-*choice* : $(\forall x{<}a)(\exists X{<}b)\varphi(x, X) \ \rightarrow \ \exists Z\ (\forall x{<}a)\varphi(x, Z^{[x]})$,

where φ is in Σ_i^p. It asserts that the Σ_i^p-formulas are closed under first-order bounded quantifications.

1.5 Polynomial time computable functions.

In this section we introduce the classes of functions $\mathcal{P}_i$. These correspond to classes which have been intensively studied in computational complexity theory, i.e., the functions which are polynomial time computable with an oracle for Σ_i^p (also denoted in the leterature by $\square_{i+1}^p$). For expository reasons we prefer to introduce them in an axiomatic way avoiding direct reference to any model of computation. Formally, our approach is self-contained. A definition of rudimentary function has already been given in Section 1.3. We include here a different one. It is easy to see that these two definitions define the same class of functions. The reader may consult [7] for some details.

To begin with, let us work in the standard model, i.e., natural numbers and finite sets of natural numbers. The functions we introduce are of two sorts, number functions and set functions, denoted respectively with lower case and capital letters. Functions take as inputs tuples of numbers and sets and they output either a number (number functions) or a set (set functions). Numbers, as input and/or output, are introduced merely as a useful device to express 'logarithmically many iterations'.

The class $\mathcal{R}$ of **rudimentary** functions is the smallest set of functions closed under composition and under the following schemas

1 (a 0-ary function),

$F(a) = [a]$

$f(a_1, .., a_n, A_1, .., A_m) = a_i + a_j$ for $0<i, j \leq n$ and $0 \leq m$,

$f(a_1, .., a_n, A_1, .., A_m) = a_i \cdot a_j$ for $0<i, j \leq n$ and $0 \leq m$,

$F(a_1, .., a_n, A_1, .., A_m) = A_i \cup A_j$ for $0 \leq n$ and $0<i, j \leq m$,

$F(a_1, .., a_n, A_1, .., A_m) = A_i \setminus A_j$ for $0 \leq n$ and $0<i, j \leq m$,

$f(A) = \mu_x(x \in A)$ (for $A = \emptyset$ this is defined to be 0)

$F(\vec{a}, Y, \vec{A}) = \bigcup_{y \in Y} G(\vec{a}, y, \vec{A})$ for G in $\mathcal{R}$.

The functions defined by the first seven schemas are called **basic rudimentary**. We shall refer to the last schema as **rudimentary collection**. The class $\mathcal{P}$ is, by definition, also closed under the following schema of **second-order (polynomially bounded) recursion**

$$F(0, \vec{x}, \vec{X}) = G(\vec{x}, \vec{X}); \qquad F(y+1, \vec{x}, \vec{X}) = [|y, \vec{x}, \vec{X}|^p] \cap H(y, \vec{x}, \vec{X}, F(y, \vec{x}, \vec{X}))$$

for any G, H in $\mathcal{P}$ and $p \in \omega$.

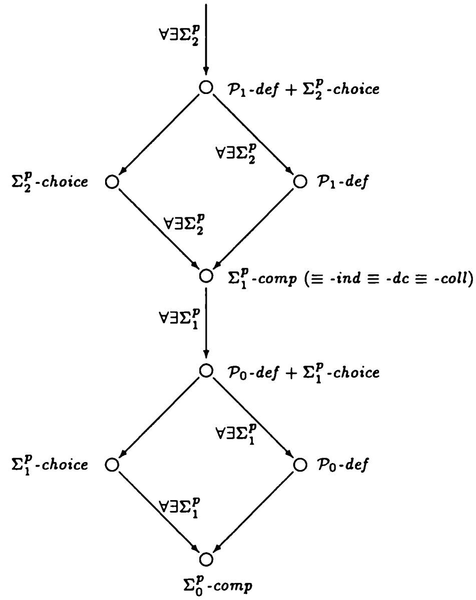
The recursion schema introduced above is polynomially bounded for two reasons. We bound both the size of the output and the depth of the recursion. So, no more than polynomially many nested iterations of functions are possible.

The class $\mathcal{P}$ is also denoted $\mathcal{P}_0$. In general, the classes $\mathcal{P}_i$ are obtained by adding to $\mathcal{P}$ **Turing oracles** for Σ_i^p-formulas and closing under rudimentary collection and second-order recursion. Turing oracles for Σ_i^p-formulas are functions of the form

$$F(\vec{a}, \vec{A}) = \begin{cases} \{0\} & \text{if } \varphi(\vec{a}, \vec{A}) \\ \emptyset & \text{otherwise,} \end{cases}$$

for φ in Σ_i^p.

Now, going back to theories of second-order arithmetic, let us use $\mathcal{P}_i$ to indicate also some sets of symbols for functions, a different symbol for each definition of a function in the corresponding class. Let $L_2(\mathcal{P}_i)$ be the corresponding expansions of L_2. Let $\mathcal{P}_i$-*def* be the theories axiomatized by Θ and the defining axioms of the functions in $\mathcal{P}_i$. We choose the obvious defining axioms of functions obtained by the basic schemas. Namely, for Turing oracles the defining axioms are those given above. If F is (the symbol of the function) obtained by rudimentary collection from $G \in \mathcal{R}$ we take as defining axiom of F



$$x{\in}F(\vec{a}, Y, \vec{A}) \leftrightarrow (\exists y{\in} Y)\ x{\in} G(\vec{a}, y, \vec{A}).$$

If F is obtained by second-order recursion from $G, H{\in}\mathcal{P}$ and $p{\in}\omega$, then the defining axiom is

$$F(0, \vec{a}, \vec{A}) = G(\vec{a}, \vec{A}) \quad\wedge\quad F(y+1, \vec{a}, \vec{A}) = H(y,\ \vec{a},\ \vec{A},\ F(y, \vec{a}, \vec{A}))\ \cap\ [|y, \vec{a}, \vec{A})|^p].$$

1.6 Relations among fragments.

We assume the reader to be familiar with fragments of first-order arithmetic (see e.g., [8]), so, we merely sketch proofs. It is easy to see that the comprehension schemas for Σ_i^p, Π_i^p and $\Sigma_0^p(\Sigma_i^p)$-formulas are equivalent. Also, we may contract quantifiers, so, Σ_{i+1}^p-*dc* and Σ_{i+1}^p-*choice* are respectively equivalent to Π_i^p-*dc* and Π_i^p-*choice* (these last two theories are defined in the obvious way). The theory Σ_{i+1}^p-*choice* proves that Σ_{i+1}^p-formulas are closed under first-order bounded quantification. In the schemas of Σ_i^p-*choice*, Σ_i^p-*dc* and Σ_i^p-*coll* we can in addition require the set Z to be a subset of $[a+1]{\times}[b]$ without strengthening the schema. The easy proofs of these facts are left to the reader.

The content of the Lemma we are going to prove in this section is summarized in the picture below. An arrow means provability. Next to the arrow we write the partial conservativity we shall prove in Sections 2.2 and 2.3.

Lemma 1.6. *For all $i{\in}\omega$,*

(i) Σ_{i+1}^p-*ind* $\Longrightarrow$ Σ_{i+1}^p-*choice* $\Longrightarrow$ Σ_i^p-*comp*

(ii) Σ_{i+1}^p-*comp* $\Longleftrightarrow$ Σ_{i+1}^p-*ind* $\Longleftrightarrow$ Σ_{i+1}^p-*dc* $\Longleftrightarrow$ Σ_{i+1}^p-*coll*

(iii) Σ^p_{i+1}*-comp* $\Longrightarrow$ $\mathcal{P}_i$*-def* $\Longrightarrow$ Σ^p_i*-comp*

We understand the first inclusion of **(iii)** *as: every model of* Σ^p_{i+1}*-comp has a unique expansion to a model of* $\mathcal{P}_i$*-def*

Proof of (i). For the first inclusion, it is sufficient to prove Π^p_i*-choice*. This is proved in a straightforward manner. By the observation above, the quantifier $\exists Z$ in the schema of choice can be bounded. So, assuming the antecedent of the implication one can prove the consequent by induction on the parameter a. The second implication is proved by induction on i. Assume that Σ^p_{i+1}*-choice* proves Σ^p_i*-comp* (this is true by definition if $i = 0$, for $i>0$ it follows from our induction hypothesis), we show that Σ^p_{i+2}*-choice* proves Σ^p_{i+1}*-comp*. Reason in a model of Σ^p_{i+2}*-choice*. Let $\varphi(x)\in\Sigma^p_{i+1}$. For some b and some $\psi\in\Pi^p_i$ the formula φ is equivalent to $(\exists X{<}b)\psi(x,X)$. We have

$$(*)\ (\forall x{<}a)\ (\exists X{<}b)(\forall Y{<}b)[\psi(x,X) \vee \neg\psi(x,Y)].$$

We may apply the axiom of choice to get a set $Z\subseteq[a]\times[b]$ such that for all $x<a$, eiether $\psi(x,Z^{[x]})$ or $(\forall Y{<}b)\neg\psi(x,Y)$. So, $\psi(x,Z^{[x]})$ is equivalent to $\varphi(x)$. Therefore, Σ^p_i*-comp* suffices to prove the existence of the set $\{x{<}a\ :\ \varphi(x)\}$.

Proof of (ii). Since all three theories above prove Σ^p_{i+1}*-choice*, in the following proof we use without explicit mention that Σ^p_{i+1}-formulas are closed under bounded first-order quantification.

It is immediate that Σ^p_{i+1}*-comp* contains Σ^p_{i+1}*-ind*. For the converse inclusion, reason in a model of Σ^p_{i+1}*-ind*; let $\varphi\in\Sigma^p_{i+1}$ and choose a parameter a. We want a set $X{<}a$ such that $x\in X \leftrightarrow \varphi(x)$ for all $x{<}a$. We are done if we can find a set of maximal cardinality among those such that $x\in X \rightarrow \varphi(x)$ for all $x{<}a$. In fact, for such an X, also the converse implication holds. Formally, we write $Y : [c] \hookrightarrow X$ for the Σ^p_0-formula saying that Y is an injection of $[c]$ into X or, in other words, that the cardinality of X is at least c. By Σ^p_1*-ind*, there exists a largest $c{<}a$ such that

$$(\exists X{<}a)(\exists Y{<}\langle c,a\rangle)\ \Big[(Y : [c] \hookrightarrow X) \wedge (\forall x{\in}X)\varphi(x)\Big].$$

The X, witnessing the existential quantifier for c maximal, is the required set satisfying $x\in X \leftrightarrow \varphi(x)$ for all $x{<}a$. This completes the proof of the first equivalence.

To prove that Σ^p_{i+1}*-ind* implies Σ^p_{i+1}*-dc* it is convenient to derive Π^p_i*-dc*. This is done by straightforward induction as for the schema of choice in previous lemma. The converse implication is proved by induction on i. Reason in a model of Σ^p_{i+1}*-dc*. We show that for every $\psi\in\Sigma^p_{i+1}$,

$$(*)\ \psi(0) \wedge (\forall x{<}a)[\psi(x) \rightarrow \psi(x+1)] \rightarrow\ \psi(a).$$

Without loss of generality, we may assume that $\psi(x)$ is equivalent to $(\exists X{<}b)\ \varphi(x,X)$ for some $\varphi(x)$ in Π^p_i and some parameter b. Assume the antecedent of (*), then

$$(\forall x{<}a)(\forall X{<}b)(\exists Y{<}b)[\varphi(x,X) \rightarrow \varphi(x+1,Y)].$$

The formula between square brackets is equivalent to a Σ^p_{i+1}-formula, so, (after few a manipulations) one can apply Σ^p_{i+1}*-dc* to get a set $Z\subseteq[a+1]\times[b]$ such that

$$Z^{[0]} = A \wedge (\forall x{<}a)[\varphi(x,Z^{[x]}) \rightarrow \varphi(x+1,Z^{[x+1]})],$$

where A is any set such that $\varphi(0, A)$. Since Σ_i^p-*ind* holds (by induction hypothesis if $i>0$ or, by definition, if $i = 0$), we can apply induction on x to the formula $\varphi(x, Z^{[x]})$ to prove $\varphi(a, Z^{[a]})$ and hence $\psi(a)$. This completes the proof of the second equivalence.

We leave the proof that Σ_i^p-*comp* is equivalent to Σ_i^p-*coll* to the reader.

Proof of (iii). The second implication is true by definition if $i = 0$. For $i>0$ this holds because $\mathcal{P}_i$ contains Σ_i^p Turing oracles and is closed under rudimentary collection. For the first implication, consider first the case $i = 0$. Given a model $\mathbf{M}$ of Σ_1^p-*comp* we show that there is a unique way of defining new functions on $\mathbf{M}$ which satisfy the axioms of $\mathcal{P}$-*def*. We proceed by induction on the definition of $F \in \mathcal{P}$. The new primitives are added in order to have that for some Σ_1^p-formula φ

$$\mathbf{M} \models F(\vec{x}, \vec{X}) = Y \leftrightarrow \varphi(\vec{x}, \vec{X}, Y)$$

$$\mathbf{M} \models \forall \vec{x}, \vec{X} \exists! Y \varphi(\vec{x}, \vec{X}, Y)$$

The proof is actually standard and need not be reported here in detail. The key step is when F is obtained by recursion. In this case Σ_1^p-*dc* is used. For $i>0$ let $\mathcal{T}_i$ be the set of Turing oracles. Clearly there is a unique way to add to a model $\mathbf{M}$ new primitives for $\mathcal{T}_i$-functions and having them satisfy their defining axioms. Now, it is easily seen that, if $\mathbf{M}$ models Σ_{i+1}^p-*comp*, then it models $\Sigma_1^p(\mathcal{T}_i)$-*comp* too. From this point on the proof proceeds as in the case $i = 0$. ∎

1.7 Relations with Buss' bounded arithmetic.

In the introduction we mentioned that Σ_i^p-*comp* coincides with Buss' S_2^i by a suitable translation of formulas. This translation has been found independently by many authors (see e.g., [15], [16], [11]). It is not necessary to include full details here, but, to give some clue to the reader, we quickly show how to transform a model of S_2^i into a model of Σ_i^p-*comp* and vice versa.

Let $\mathbf{M}_1$ be a model of S_2^i. Let $\mathbf{M}_2$ be the second-order structure having as first-order objects the elements a of $\mathbf{M}_1$ such that 2^a exists and as second-order objects those finite subsets of $\mathbf{M}_1$ which are coded in the usual way by elements of $\mathbf{M}_1$. I.e., for every $a \in \mathbf{M}_1$ we add the set A to $\mathbf{M}_2$ such that

$$a = \textstyle\sum_{x \in A} 2^x$$

Functions and relations of $\mathbf{M}_2$ are defined in the natural way. Note that multiplication of first-order elements is a total operation in $\mathbf{M}_2$. In fact if 2^a and 2^b exist in $\mathbf{M}_1$ then $2^{a \cdot b}$ exists too, since it is equal to $2^a \# 2^b$. It is easy to see that $\mathbf{M}_2$ models Σ_i^p-*comp*. In fact, it is sufficient to note that for every second-order formula $\varphi(x, X) \in \Sigma_i^p$ there is a first-order formula $\varphi^*(x, y) \in \Sigma_i^b$ such that for every $a, A \in \mathbf{M}_2$

$$\mathbf{M}_2 \models \varphi(a, A) \iff \mathbf{M}_1 \models \varphi^*(a, \textstyle\sum_{x \in A} 2^x).$$

To see the other direction, we apply the inverse procedure. Let $\mathbf{M}_2$ be a model of Σ_i^p-*comp*. We think of sets of $\mathbf{M}_2$ as representing numbers, i.e., we think of the set X as the number

$n(X) := \sum_{x \in X} 2^x$

Clearly, in general such a number need not exist in $\mathbf{M}_2$. Still, formalizing the natural algorithm for addition and multiplication of binary numbers, we may define in $\mathbf{M}_2$ some set functions $X \oplus Y$ and $X \otimes Y$ such that

$$n(X \oplus Y) = n(X) + n(Y) \quad \text{and} \quad n(X \otimes Y) = n(X) \cdot n(Y).$$

It is well known that such an algorithm is computable in polynomial time, so, $X \oplus Y$ and $X \otimes Y$ are total functions in every model of $\mathcal{P}$-*def*. Let $X \# Y$ be the set $\{\, |X| \cdot |Y| \,\}$ which exists because $\mathbf{M}_2$ models Σ_0^p-*comp*. Also, all other functions of the language of S_2 can be defined in a similar way. Now, one can construct a model of S_2^i having as its domain the second-order elements of $\mathbf{M}_2$ and as functions and relations the ones just defined. The reader may check that the 32 axioms of BASIC hold in $\mathbf{M}_1$. Because $\mathbf{M}_2$ is a model of Σ_i^p-*ind*, it is not difficult to see that $\mathbf{M}_1$ satisfies logarithmic inductions for Σ_i^b-formulas. Hence $\mathbf{M}_1$ is a model of S_2^i.

This **first-second-order isomorphism** tranforms models of $\mathcal{P}_i$-*def* into models of T_2^i for all positive i and vice versa. The first-order theory corresponding to $\mathcal{P}_0$-*def* is known as PV_1. Second-order models of Σ_{i+1}^p-*choice* correspond to first-order models of $BB\Sigma_{i+1}^b$ (cf. chapter V of [8]), i.e., models of S_2^0 and the schema

$$(\forall x{<}|t|)(\exists y{<}s)\varphi(x,y) \rightarrow \exists w(\forall x{<}|t|)\varphi(x,(w)_x).$$

where φ is in Σ_{i+1}^b.

2 Witnessing theorems and conservativity results.

Buss was the first to give an extensive characterization of complexity classes as classes of functions definable and provably total in some weak fragment of arithmetic. However, the very idea of the proofs we report here goes back to the Mints-Parsons' famous partial conservativity result of $I\Sigma_1$ over PRA [13], [14]. Buss', Parsons' and Mints' proofs are proof-theoretical. Wilkie gave a model-theoretic proof (unpublished) of Buss' theorem (see [8]). Here we adapt a model-theoretical proof of the Mints-Parsons' theorem given by Albert Visser (unpublished).

2.1 Closures

Let $\mathbf{M}$ be a model of Σ_0^p-*comp* and let $\mathbf{W}$ be a subset of $\mathbf{M}$. We say that $\mathbf{W}$ is closed under $\mathcal{R}$-functions if $F(\vec{c},\vec{C}), f(\vec{c},\vec{C}) \in \mathbf{W}$ for every $\vec{c},\vec{C} \in \mathbf{W}$ and $F, f \in \mathcal{R}$. The $\mathcal{R}$-closure of $\mathbf{W}$ in $\mathbf{M}$ is the minimal $\mathcal{R}$-closed subset of $\mathbf{M}$ containing $\mathbf{W}$, i.e.,

$$\langle\!\langle \mathbf{W} \rangle\!\rangle_{\mathcal{R}} := \{F(\vec{c},\vec{C}), f(\vec{c},\vec{C}) \ : \ \vec{c},\vec{C} \in \mathbf{W} \text{ and } F, f \in \mathcal{R}\}.$$

We interpret $\mathcal{R}$-closed subsets of $\mathbf{M}$ as substructures in the canonical way: the functions and relations of $\mathbf{N}$ are the restriction of those of $\mathbf{M}$. In the same way we define $\mathcal{P}_i$-closed sets in models of $\mathcal{P}_i$-*def*.

We say that $\mathbf{N}\subseteq\mathbf{M}$ is a Σ_i^p-elementary substructure of $\mathbf{M}$, if for every Σ_i^p-formula φ and every $\vec{a}, \vec{A}\in\mathbf{N}$

$$\mathbf{N} \models \varphi(\vec{a}, \vec{A}) \implies \mathbf{M} \models \varphi(\vec{a}, \vec{A})$$

We write $\mathbf{N} \prec_{\Sigma_i^p} \mathbf{M}$ if $\mathbf{N}$ is a Σ_i^p-elementary substructure of $\mathbf{M}$. A similar notation is used also for other classes of formulas.

Lemma 2.1. *(Definability of Skolem functions)*

(i) *$\mathcal{R}$-closed substructures of models of Σ_0^p-comp are Σ_0^p-elementary (so, in particular, they are models of Σ_0^p-comp).*

(ii) *$\mathcal{P}_i$-closed substructures of models of $\mathcal{P}_i$-def are $\Sigma_i^p(\mathcal{P}_i)$-elementary (so, in particular, they are models of $\mathcal{P}_i$-def).*

Proof. For **(i)**, observe that first-order Skolem functions for Σ_0^p-formulas are in $\mathcal{R}$. The proof of **(ii)** when $i = 0$ is obvious. For $i>0$ it suffices to show that among the $\mathcal{P}_i$-functions there are Skolem functions for $\Sigma_i^p(\mathcal{P}_i)$-formulas. I.e., for every $\Sigma_i^p(\mathcal{P}_i)$-formula φ there is a function F in $\mathcal{P}_i$ such that

$$\exists Y<|\vec{a}, \vec{A}|^p \varphi(\vec{a}, \vec{A}, Y) \to \varphi(\vec{a}, \vec{A}, F(\vec{a}, \vec{A})).$$

To see this we shall define a function F that, by binary search, produces the minimal (in the lexicographic order) set $Y<|\vec{a}, \vec{A}|^p$ satisfying $\varphi(\vec{a}, \vec{A}, Y)$. Let us define the function G by recursion in the following way (omitting parameters and bounds)

$$G(0, \vec{a}, \vec{A}) = \emptyset$$

$$G(y+1, \vec{a}, \vec{A}) = \begin{cases} G(y, \vec{a}, \vec{A}) \text{ if } & (\exists Y<|\vec{a}, \vec{A}|^p)[(G(y, \vec{a}, \vec{A})\subseteq Y) \wedge \varphi(\vec{a}, \vec{A}, Y) \wedge y \notin Y] \\ G(y, \vec{a}, \vec{A}) \cup \{y\} \text{ otherwise} & \end{cases}$$

(recall that $\mathcal{P}_i$ is closed under definition by $\Sigma_i^p(\mathcal{P}_i)$-cases since it contains the characteristic functions of Σ_i^p-formulas and is closed under composition). Finally, we define

$$F(\vec{a}, \vec{A}) = G(|\vec{a}, \vec{A}|^p + 1).$$

We leave to the reader the verification that F produces a witness of $\exists Y<|\vec{a}, \vec{A}|^p\varphi(\vec{a}, \vec{A}, Y)$, if one exists, and is $\emptyset$ otherwise. ∎

The class of $\mathcal{P}_i$-functions is closed under Σ_i^p-definition by cases, so, an easy compactness argument proves the following witnessing theorem for $\mathcal{P}_i$-*def*.

Corollary 2.1. *(Witnessing theorem for $\mathcal{P}_i$-def.) Each $\forall\exists\Sigma_{i+1}^p$ sentence provable in $\mathcal{P}_i$-def has a witnessing function in $\mathcal{P}_i$.*

Proof. We have to prove that, for all $\varphi\in\Sigma_{i+1}^p$, there is a function F in $\mathcal{P}_i$ such that

$$\mathcal{P}_i\text{-}def \vdash \forall\vec{X}, \vec{x}\ \exists Y \varphi(\vec{x}, \vec{X}, Y) \qquad \implies \qquad \mathcal{P}_i\text{-}def \vdash \forall\vec{X}, \vec{x}\ \varphi(\vec{x}, \vec{X}, F(\vec{x}, \vec{X})).$$

By contraction of quantifiers it suffices to show that the implication above holds for Π_i^p-

formulas. So, let φ be a Π^p_i-formula such that for no $F \in \mathcal{P}_i$

(*) $\mathcal{P}_i\text{-}def \vdash \forall \vec{X}, \vec{x}\ \varphi(\vec{x}, \vec{X}, F(\vec{x}, \vec{X}))$.

Let $\vec{c}, \vec{C}$ be fresh constants and consider the theory

(**) $\mathcal{P}_i\text{-}def + \{\neg\varphi(\vec{c}, \vec{C}, F(\vec{c}, \vec{C}))\ :\ F \in \mathcal{P}_i\}$

This theory is consistent. Otherwise by compactness, for a finite set of functions $\{F_1, .., F_n\}$ in $\mathcal{P}_i$,

$$\mathcal{P}_i\text{-}def \vdash \forall \vec{x}, \vec{X}\ \left[\varphi(\vec{x}, \vec{X}, F_1(\vec{x}, \vec{X})) \vee .... \vee \varphi(\vec{x}, \vec{X}, F_n(\vec{x}, \vec{X}))\right].$$

So, since $\mathcal{P}_i$-functions are closed under definition by Σ^p_i-cases, one can combine $F_1, .., F_n$ together to find a function $F \in \mathcal{P}_i$ satisfying (*). Now, choose a model $\mathbf{M}$ of the theory (**) and let $\mathbf{N}$ be the $\mathcal{P}_i$-closure of $\vec{c}, \vec{C}$. By the previous lemma $\mathbf{N}$ is a model of $\mathcal{P}_i\text{-}def$. The same lemma excludes the possibility of having in $\mathbf{N}$ a set Y such that $\varphi(\vec{c}, \vec{C}, Y)$. Thus $\mathcal{P}_i\text{-}def$ does not prove $\forall \vec{x}, \vec{X}\ \exists Y \varphi(\vec{x}, \vec{X}, Y)$ and the corollary follows. ∎

2.2 A model-theoretical version of Buss' witnessing theorem.

We derive our version of Buss' witnessing theorem from the following lemma.

Lemma 2.2. *Every model $\mathbf{M}$ of $\mathcal{P}_i\text{-}def$ has an $\exists\Sigma^p_{i+1}$-elementary extension to a model $\mathbf{N}$ of Σ^p_{i+1}-comp such that for every Π^p_i-formula φ there is a function $F \in \mathcal{P}_i$ with (undisplayed) parameters from $\mathbf{N}$ such that (*) below holds*

(*) $\mathbf{N} \models \forall X\ \exists Y \varphi(X, Y) \rightarrow \forall X\ \varphi(X, F(X))$.

Proof. We claim that, if we succeed in satisfying condition (*), we obtain also that $\mathbf{N}$ models Σ^p_{i+1}-*comp*. To prove the claim it is sufficient to check that in $\mathbf{N}$ the schema of dependent choices holds for Π^p_i-formulas. Assume in $\mathbf{N}$ holds $\forall x \forall X \exists Y\ \varphi(x, X, Y)$ where a bound b on X and Y is implicit in φ. Let $a \in \mathbf{N}$, we want to find a Z such that $(\forall x < a)$ $\varphi(x, Z^{[x]}, Z^{[x+1]})$. By (*), for some $F \in \mathcal{P}_i$ and for all x and X, $\varphi(x, X, F(x, X))$. Define the following function G by second-order recursion (F can be bounded by b):

$$G(0) = \emptyset, \quad G(x+1) = F(x+1,\ G(x)).$$

Finally, Z is obtained by rudimentary collection: $\bigcup_{x<a+1}\{x\} \times G(x)$. This proves our claim.

Now, let $\mathbf{M}$ be a model of $\mathcal{P}_i\text{-}def$. The required model $\mathbf{N}$ is constructed as the union of an $\exists\Sigma^p_{i+1}$-elementary chain of models of $\mathcal{P}_i\text{-}def$,

$$\mathbf{M} = \mathbf{M}_0 \prec_{\exists\Sigma^p_{i+1}} \mathbf{M}_1 \prec_{\exists\Sigma^p_{i+1}} \mathbf{M}_2 \prec_{\exists\Sigma^p_{i+1}} ....$$

The chain is constructed by stages. Each link of the chain is constructed using a model $\mathbf{W}$ as intermediate step, as in the following diagram

$$\begin{tikzcd}
 & & \mathbf{W} & \\
 {} \arrow[r, "\exists\Sigma^p_{i+1}"] & \mathbf{M}_s \arrow[ur] \arrow[r, "\exists\Sigma^p_{i+1}"] & \mathbf{M}_{s+1} \arrow[u, "\Sigma^p_i"'] \arrow[r, "\exists\Sigma^p_{i+1}"] & {}
\end{tikzcd}$$

Suppose $\mathbf{M}_s$ has already been constructed. Let φ_s be the s-th Π^p_i-formula of an enumeration (to be specified below) of Π^p_i-formulas with parameters in $\mathbf{M}_s$. Let $(*)_s$ be $(*)$ with φ_s for φ. We shall construct a model $\mathbf{M}_{s+1}$ realizing $(*)_s$ for some function $F \in \mathcal{P}_i$. Observe $(*)_s$ is a $\exists\forall\Pi^p_{i+1}$-formula, so, its truth is preserved upwards in the chain and finally inherited by the union $\mathbf{N}$. It is easy to choose the enumeration such that eventually all Π^p_i-formulas with parameters in $\mathbf{N}$ are considered. The details of the enumeration are as follows. At each stage s we fix an arbitrary enumeration $\{\psi^s_t\}_{t\in\omega}$ of all Π^p_i-formulas with parameters in $\mathbf{M}_s$. Finally, let φ_s be ψ^r_t for $s = \langle r, t\rangle$. To define $\mathbf{M}_{s+1}$ proceed as follows. If $(*)_s$ holds for φ_s, we do nothing, i.e., we define $\mathbf{M}_{s+1} := \mathbf{M}_s$. Otherwise, we try to make the antecedent of $(*)_s$ false in $\mathbf{M}_{s+1}$. We construct $\mathbf{M}_{s+1}$ and $C \in \mathbf{M}_{s+1}$ where $\exists Y \varphi_s(C, Y)$ fails. Since $(*)_s$ does not hold in $\mathbf{M}_s$, the following theory has a model $\mathbf{W}$

$$\mathit{Diag}(\mathbf{M}_s) + \{\neg\varphi_s(C, F(C)) \ : \ F \in \mathcal{P}_i \text{ with parameters in } \mathbf{M}_s\},$$

where C is a fresh constant and $\mathit{Diag}(\mathbf{M}_s)$ is the elementary diagram of $\mathbf{M}_s$ (to check the consistency, argue by compactness). $\mathbf{W}$ is elementarily equivalent to $\mathbf{M}_s$, so, in particular, it is a model of $\mathcal{P}_i$-*def*. Define

$$\mathbf{M}_{s+1} := \langle\!\langle \mathbf{M}_s + C \rangle\!\rangle_{\mathcal{P}_i}$$

Closure to be taken in $\mathbf{W}$. Clearly $\mathbf{M}_{s+1}$ is a Σ^p_i-elementary substructure of $\mathbf{W}$ which is elementary equivalent to $\mathbf{M}_s$, so, every $\exists\Sigma^p_{i+1}$-formula true in $\mathbf{M}_{s+1}$ will be true in $\mathbf{W}$ and hence in $\mathbf{M}_s$. In $\mathbf{M}_{s+1}$ there is no witness of $\exists Y \varphi_s(C, Y)$. This completes the proof of the lemma. ∎

Corollary 2.2. *($\forall\exists\Sigma^p_{i+1}$-conservation and witnessing theorem for Σ^p_{i+1}-comp.) Σ^p_{i+1}-comp is $\forall\exists\Sigma^p_{i+1}$-conservative over $\mathcal{P}_i$-def, therefore every $\forall\exists\Sigma^p_{i+1}$ sentence provable in Σ^p_{i+1}-comp has a witnessing function in $\mathcal{P}_i$.*

Proof. Immediate from the previous lemma and from Lemma 2.1. ∎

2.3 A model theoretical characterization of choice.

We now introduce the concept of $\mathcal{R}$-extension. This is an extension where all second-order objects are constructible relative to the extended model. This notion may be viewed also as a second-order generalization of cofinal extension. It will be used to give a model theoretical characterization of Σ^p_{i+1}-*choice* over Σ^p_i-*comp*. An useful application of this notion is given in the proof of the Corollary below. (The conservativity result in Corollary 2.3 **(b)** will find applications in the following sections to characterize the collapse of *BA*.)

Let $\mathbf{M}$ and $\mathbf{N}$ be models of Σ^p_0-*comp*. Recall that their first-order parts are denoted respectively by $\mathbf{m}$ and $\mathbf{n}$. We say that $\mathbf{N}$ is an **$\mathcal{R}$-extension** of $\mathbf{M}$ if

(o) $\mathbf{m}$ is cofinal in $\mathbf{n}$, i.e., for all $a\in\mathbf{N}$ there is $b\in\mathbf{M}$, such that $a<b$.

(i) $\mathbf{M} \prec_{\Sigma_0^p} \mathbf{N}$.

(ii) $\mathbf{N} = \langle\!\langle \mathbf{M}+\mathbf{n}\rangle\!\rangle_{\mathcal{R}}$, i.e., for every $A\in\mathbf{N}$ there are $a\in\mathbf{N}$ such that $\mathbf{N} \models A = F(a)$ for some $F\in\mathcal{R}$ with parameters in $\mathbf{M}$.

We write $\mathbf{M} \prec_{\mathcal{R}} \mathbf{N}$ if $\mathbf{N}$ is an $\mathcal{R}$-extension of $\mathbf{M}$.

Fact 2.3. *Let* $\mathbf{N}$ *be an* $\mathcal{R}$*-extension of* $\mathbf{M}$.

(a) $\mathbf{M} \prec_{\exists\Sigma_1^p} \mathbf{N}$.

(b) $\mathbf{M} \models \Sigma_{i+1}^p\text{-}choice \implies \mathbf{M} \prec_{\exists\Sigma_{i+2}^p} \mathbf{N}$.

(c) $\mathbf{M} \models \Sigma_i^p\text{-}comp \iff \mathbf{N} \models \Sigma_i^p\text{-}comp$.

(d) $\mathbf{M} \models \mathcal{P}_i\text{-}def \iff \mathbf{N} \models \mathcal{P}_i\text{-}def$.

Proof of (a). If $\mathbf{N} \models \exists Y\varphi(Y)$ for some Σ_0^p-formula φ with parameters in $\mathbf{M}$, then for some $b\in\mathbf{M}$, and some $F\in\mathcal{R}$ with parameters in $\mathbf{M}$

$$\mathbf{N} \models (\exists x<b)\varphi(F(x)),$$

so, by Σ_0^p-elementarity, this holds in $\mathbf{M}$ too. This proves **(a)**.

Proof of (b). Let $\mathbf{M}$ be a model of Σ_{i+1}^p-*choice*. Let $a\in\mathbf{M}$ and $\varphi\in\Sigma_i^p$ with parameters in $\mathbf{M}$ and suppose $\mathbf{N} \models \exists Y(\forall X<a)\varphi(X, Y)$. It suffices to show that the same formula holds in $\mathbf{M}$ too. As induction hypothesis we assume $\exists\Sigma_{i+1}^p$-elementarity. Since $\mathbf{N}$ is an $\mathcal{R}$-extension, for some $F\in\mathcal{R}$ with parameters in $\mathbf{M}$,

$$\mathbf{N} \models \exists y\ (\exists x<y)(\forall X<a)\varphi(X, F(x)).$$

Then, clearly,

$$\mathbf{N} \models \exists y\ (\forall Z\subseteq\langle a, y\rangle)(\exists x<y)\varphi(Z^{[x]}, F(x)).$$

So, by $\exists\Sigma_{i+1}^p$-elementarity,

$$\mathbf{M} \models \exists y\ (\forall Z\subseteq\langle a, y\rangle)(\exists x<y)\varphi(Z^{[x]}, F(x)).$$

Finally, by Σ_{i+1}^p-*choice*,

$$\mathbf{M} \models \exists y\ (\exists x<y)(\forall X<a)\varphi(X, F(x)).$$

This proves **(b)**.

Proof of (c). The 'left to right' direction of Fact **(c)** is true by definition when $i = 0$. For $i>0$ it follows from **(b)** and Lemma 1.6. In fact, these imply that $\mathbf{N}$ is an $\exists\Sigma_{i+1}^p$-elementary extension of $\mathbf{M}$. So, let φ be any Σ_i^p-formula with parameters in $\mathbf{N}$. By **(ii)**, we can assume that all second-order parameters $\vec{c}$ of φ belong to $\mathbf{M}$. Let $a\in\mathbf{N}$ be arbitrary. Choose in $\mathbf{M}$ a $b>a, \vec{c}$. Since $\mathbf{M} \models \Sigma_i^p$-*comp* there is a set $A\in\mathbf{M}$ such that

$$\mathbf{M} \models (\forall x, \vec{y}<b)[\langle x, \vec{y}\rangle\in A \leftrightarrow \varphi(x, \vec{y})]$$

where the variables $\vec{y}$ replace all first-order parameters of φ. By $\exists\Sigma_{i+1}^p$-elementarity, A satisfies the same property in $\mathbf{N}$ too. So, in $\mathbf{N}$ the set $B := \{x<a : \langle x, \vec{c}\rangle \in A\}$ verifies,

$$\mathbf{N} \models (\forall x<a)[x\in B \leftrightarrow \varphi(x, \vec{c})]$$

This proves that $\mathbf{N}$ is a model of Σ_i^p-*comp*.

The converse direction ('right to left') is also true by definition when $i = 0$. So, assume it true for i and let us prove it for $i+1$. Let $\mathbf{N}$ be a model of Σ_{i+1}^p-*comp*. By induction hypothesis, $\mathbf{M}$ is a model of Σ_i^p-*comp* and by Fact **(b)**, a $\exists\Sigma_{i+1}^p$-elementary substructure of $\mathbf{N}$. Let $\varphi(x, Y)$ be a Π_i^p-formula with parameters in $\mathbf{M}$ and an implicit bound on Y. It suffices to find in $\mathbf{M}$ a set A such that

$$(\forall x<a)[x\in A \leftrightarrow \exists Y\varphi(x, Y)].$$

By Lemma 1.6, $\mathbf{N}$ models Σ_{i+1}^p-*coll*, so, for some set Z

$$\mathbf{N} \models (\forall x<a)[\exists Y\varphi(x, Y) \leftrightarrow \varphi(x, Z^{[x]})].$$

Then, for some $b\in\mathbf{N}$ and function $F\in\mathbf{N}$ with parameters in $\mathbf{M}$,

$$\mathbf{N} \models (\forall x<a)[\exists Y\varphi(x, Y) \leftrightarrow \varphi(x, F(b)^{[x]})].$$

Consider, in $\mathbf{M}$, the set $A := \{x<a : \exists y\ \varphi(x, F(y)^{[x]})\}$. We claim this is the required one. We only need to show that $(\forall x<a)[\exists Y\varphi(x, Y) \rightarrow x\in A]$, because the converse implication is obvious. If $\exists Y\varphi(x, Y)$ holds in $\mathbf{M}$ for some $x<a$, then this will be also true in the Σ_i^p-elementary extension $\mathbf{N}$. Then $\exists y\ \varphi(x, F(y)^{[x]})$ holds in $\mathbf{N}$ and, again by Σ_i^p-elementarity, is true in $\mathbf{M}$ too. Therefore $x\in A$.

Proof of (d). To prove the 'left to right' direction we use Lemma 2.2. This lemma characterizes models of $\mathcal{P}_i$-*def* as those having an $\exists\Sigma_{i+1}^p$-elementary extension to a model $\mathbf{M}'$ of Σ_{i+1}^p-*comp*. So, it suffices to show that there exists a model $\mathbf{W}$ satisfying the following diagram of (restricted) elementary extensions

$$\begin{tikzcd}
\mathbf{M} \arrow[r, "\exists\Sigma_{i+1}^p"] \arrow[d, "\mathcal{R}"'] & \mathbf{M}' \models \Sigma_{i+1}^p\text{-}comp \arrow[d] \\
\mathbf{N} \arrow[r, "\exists\Sigma_{i+1}^p"] & \mathbf{W}.
\end{tikzcd}$$

Consider the theory $Diag(\mathbf{M}') + Diag_{\Pi_{i+1}^p}(\mathbf{N})$. This theory has a model; otherwise, suppose that for some $\varphi\in\Sigma_i^p$,

$$(\forall X<|\vec{c}|^p)\varphi(X, \vec{c})\in Diag_{\Pi_{i+1}^p}(\mathbf{N}) \quad \text{and} \quad Diag(\mathbf{M}') \vdash \neg\forall X\varphi(X, \vec{c})$$

where we assume that a bound on X is implicit in φ. Also, since $\mathbf{M} \prec_{\mathcal{R}} \mathbf{N}$, we can assume that all other parameters of φ except $\vec{c}$ are in $\mathbf{M}$. Let $a\in\mathbf{M}$ be such that $\vec{c}<a$. Replacing the constants $\vec{c}\notin\mathbf{M}'$ with variables and quantifying we obtain

$$\mathbf{M}' \models (\forall\vec{x}<a)\exists X\neg\varphi(X, \vec{x}).$$

We may apply Σ^p_{i+1}-*choice* to get,

$$\mathbf{M}' \models \exists Z(\forall\vec{x}{<}a)\neg\varphi(Z^{[\vec{x}]},\vec{x})$$

so, by $\exists\Sigma^p_{i+1}$-elementarity,

$$\mathbf{M} \models \exists Z(\forall\vec{x}{<}a)\neg\varphi(Z^{[\vec{x}]},\vec{x}).$$

Recall that $\mathbf{M}$ models Σ^p_i-*comp*, so, by **(b)**, $\mathcal{R}$-extensions of $\mathbf{M}$ are $\exists\Pi^p_i$-elementary. So,

$$\mathbf{N} \models \exists Z(\forall\vec{x}{<}a)\neg\varphi(Z^{[\vec{x}]},\vec{x}).$$

Therefore,

$$\mathbf{N} \models (\forall\vec{x}{<}a)\exists X\neg\varphi(X,\vec{x}).$$

A contradiction since we assumed that $\forall X\varphi(X,\vec{c})\in Diag_{\Pi^p_{i+1}}(\mathbf{N})$.

Let $\mathbf{W}'$ be a model of the theory above and let

$$\mathbf{W} := \{a, A\in\mathbf{W}' : a, A<b \text{ for some } b\in\mathbf{M}\}$$

Clearly $\mathbf{W}'$ is a model of Σ^p_{i+1}-*comp* and consequently also $\mathbf{W}$. To prove $\mathbf{N} \prec_{\exists\Sigma^p_{i+1}} \mathbf{W}$, it suffices to observe that $\mathbf{N} \prec_{\Sigma^p_{i+1}} \mathbf{W}'$, that $\mathbf{W} \prec_{PH} \mathbf{W}'$ and that $\mathbf{N}$ is cofinal in $\mathbf{W}$. This proves the left to right direction of **(d)**.

For the converse, assume $\mathbf{N}$ is a model of $\mathcal{P}_i$-*def*. By Lemma 2.2, there is a model $\mathbf{N}'$ such that

$$\begin{tikzcd}
\mathbf{M} \arrow[d, "\mathcal{R}"'] \arrow[dr, "\exists\Sigma^p_{i+1}"] & \\
\mathbf{N} \arrow[r, "\exists\Sigma^p_{i+1}"] & \mathbf{N}' \models \Sigma^p_{i+1}\text{-}comp
\end{tikzcd}$$

where the diagonal arrow follows from **(b)** since, by **(c)**, $\mathbf{M}$ is a model of Σ^p_i. ∎

Lemma 2.3. *Every model* $\mathbf{M}$ *of* Σ^p_i-*comp has an* $\mathcal{R}$-*extension to a model* N *of* Σ^p_{i+1}-*choice.*

Proof. The proof is similar to that of Theorem 2.2. The model $\mathbf{N}$ is constructed as the union of a chain

$$\mathbf{M} = \mathbf{M}_0 \prec_{\mathcal{R}} \mathbf{M}_1 \prec_{\mathcal{R}} \mathbf{M}_2 \prec_{\mathcal{R}} \ldots$$

By the Fact above we have that we actually construct a $\exists\Sigma^p_{i+1}$-elementary chain of models of Σ^p_i-*comp*. Let $\{\varphi_s\}_s\in\omega$ be an enumeration with infinitely many repetitions of all formulas with parameters in $\mathbf{N}$, such that all parameters of φ_s are in $\mathbf{M}_s$ (see Theorem 2.2 for details on this enumeration). The chain is constructed so that for all $\varphi_s\in\Pi^p_i$, either (1) or (2) below holds.

(1) For every $a\in\mathbf{M}$ there is a $Z\in\mathbf{M}$ such that $\mathbf{M}_s \models (\forall x{<}a)\varphi_s(x, Z^{[x]})$.

(2) There is a $c \in \mathbf{M}_{s+1}$ such that $\mathbf{M}_{s+1} \models \forall Y \neg\varphi_s(c, Y)$.

Each link of the chain is constructed using a model $\mathbf{W}$ as intermediate step, as in the following diagram

$$\begin{tikzcd}
 & & \mathbf{W} & \\
{} \arrow[r, "\mathcal{R}"] & \mathbf{M}_s \arrow[ur] \arrow[r, "\mathcal{R}"] & \mathbf{M}_{s+1} \arrow[u, "\Sigma_0^p"'] \arrow[r, "\mathcal{R}"] & {}
\end{tikzcd}$$

Suppose $\mathbf{M}_s$ has already been constructed. If (1) holds in $\mathbf{M}_s$ then let $\mathbf{M}_{s+1} := \mathbf{M}_s$. Otherwise, let $\mathbf{W}$ be any model of

$$\mathit{Diag}(\mathbf{M}_s) + (c<a) + \{\neg\varphi_s(c, F(c)) \ : \ F \in \mathcal{R} \text{ with parameters in } \mathbf{M}\}.$$

Such a model exists, otherwise, for some n

$$\mathbf{M}_s \models (\forall x<a) \ \bigvee_{m=0}^{n} \varphi_s(x, F_n(x)).$$

Using Σ_i^p-*comp* one can define a set Z in $\mathbf{M}$ such that for all $(x<a)$,

$$Z^{[x]} = F_m(x) \text{ for the minimal } m<n \text{ such that } \varphi_s(x, F_m(x)) \text{ holds.}$$

But we assumed such a set does not exist.

Clearly $\mathbf{W}$ is, up to isomorphism, an elementary superstructure of $\mathbf{M}_s$. Let $\mathbf{M}_{s+1} := \langle\!\langle \mathbf{M} + c \rangle\!\rangle_{\mathcal{R}}$(closures to be taken in $\mathbf{W}$). To check that $\mathbf{M}_s \prec_{\mathcal{R}} \mathbf{M}_{s+1}$ note that $\mathbf{M}_{s+1}$ is a Σ_0^p substructure of $\mathbf{W}$. Also, observe that all elements of $\mathbf{M}_{s+1}$ are generated by elements of $\mathbf{M}_s$ and the first-order element $c<a$, so, conditions **(o)**, **(i)** and **(ii)** in the definition of $\mathcal{R}$-extension are fulfilled.

To check that (2) holds, suppose not, for a contradiction. If $\exists Y \varphi_s(c, Y)$ held in $\mathbf{M}_{s+1}$, then we would have $\varphi_s(c, F(c))$ for some $F \in \mathcal{R}$ with parameters in $\mathbf{M}_s$. We will reach a contradiction by showing that instead $\varphi_s(c, F(c))$ must fail in $\mathbf{M}_s$. By construction, we have $\neg\varphi_s(c, F(c))$ in $\mathbf{M}_{s+1}$. To pull this back to M_s we reason as follows. Since $\mathbf{M}_s$ is a model of Σ_i^p-*comp*, for some $A \in \mathbf{M}_s$,

$$(\forall x<a)[x \in A \leftrightarrow \varphi_s(x, F_n(x))].$$

So, by the elementary equivalences proved above, this holds also in $\mathbf{W}$ and in $\mathbf{M}_{s+1}$. In $\mathbf{W}$ we have $c \notin A$ and, by Σ_0^p equivalence this holds in $\mathbf{M}_{s+1}$. So, $\neg\varphi_s(c, F(c))$, a contradiction.

Finally, $\mathbf{N}$ is a model of Σ_{i+1}^p-*choice*, since the truth of both formulas in (1) and in (2) is preserved along the $\exists\Sigma_{i+1}^p$-chain. ∎

Now we can easily prove the characterization announced above.

Theorem 2.3. *For every* $\mathbf{M} \models \Sigma_i^p$-*comp the following are equivalent*

(i) $\mathbf{M} \models \Sigma_{i+1}^p$-*choice*

(ii) *Every $\mathcal{R}$-elementary extension of* $\mathbf{M}$ *is* $\exists\Sigma^p_{i+2}$*-elementary.*

Proof. That (i) implies (ii) has already been observed in the fact above. For the converse, let $\varphi\in\Pi^p_i$ and suppose that $(\forall x{<}a)(\exists Y{<}b)\varphi(x, Y)$ holds in $\mathbf{M}$. Let $\mathbf{N}$ be the $\mathcal{R}$-elementary extension of $\mathbf{M}$ to a model of Σ^p_{i+1}*-choice* as guaranteed by the Lemma above. Then, by $\exists\Sigma^p_{i+2}$-elementarity, $(\forall x{<}a)(\exists Y{<}b)\varphi(x, Y)$ holds also in $\mathbf{N}$. Let Z in $\mathbf{N}$ be such that $(\forall x{<}a)\varphi(x, Z^{[x]})$. By the definition of $\mathcal{R}$-extension,

$$\mathbf{N} \models (\exists y{<}c)(\forall x{<}a)\varphi(x, F(y)^{[x]})$$

for some $c\in\mathbf{M}$ and some $F\in\mathcal{R}$ with parameters in $\mathbf{M}$. So, by $\exists\Sigma^p_{i+2}$-elementarity this formula holds also in $\mathbf{M}$. ∎

The following conservativity results are consequences of the lemma above.

Corollary 2.3.

(a) Σ^p_{i+1}*-choice is* $\forall\exists\Sigma^p_{i+1}$*-conservative over* Σ^p_i*-comp.*

(b) $\mathcal{P}_i$*-def* $+ \Sigma^p_{i+1}$*-choice* $\vdash \Sigma^p_{i+1}$*-comp* $\implies$ $\mathcal{P}_i$*-def* $\vdash \Sigma^p_{i+1}$*-comp.*

(c) Σ^p_{i+1}*-choice* $\vdash \Sigma^p_{i+1}$*-comp* $\implies$ Σ^p_i*-comp* $\vdash \Sigma^p_{i+1}$*-comp.*

Proof. (a) follows from the Lemma and Fact (b) above. The proof of (b) and (c) are similar. Let us prove (b). Assume $\mathcal{P}_i$*-def* $+ \Sigma^p_{i+1}$*-choice* proves Σ^p_{i+1}*-comp*. Let $\mathbf{M}$ be any model of $\mathcal{P}_i$*-def*. In particular, $\mathbf{M}$ is a model of Σ^p_i*-comp*, so, by the Lemma above, it has an $\mathcal{R}$-extension $\mathbf{N}$ to a model of Σ^p_{i+1}*-choice*. By Fact (d) above, $\mathbf{N}$ is also a model of $\mathcal{P}_i$*-def*. So, by our assumption, $\mathbf{N}$ models Σ^p_{i+1}*-comp*. By Fact (c) above, $\mathbf{M}$ is also a model of Σ^p_{i+1}*-comp*. ∎

2.4 Ultrapowers

In this section we present an ultrapower construction corresponding to the construction in Lemma 2.3. Let $\mathbf{M}$ be a model of Σ^p_0*-comp* and $a\in\mathbf{M}$. Let $\mathcal{U}_a$ be an ultrafilter on the set $\{b\in\mathbf{M} :\ M \models b{<}a\}$. We define on $\mathbf{M}$ the following equivalence relations

$$A \sim_1 B \text{ iff } \{c\in\mathbf{M} :\ M \models c{<}a \wedge A(c) = B(c)\} \in \mathcal{U}_a$$

$$A \sim_2 B \text{ iff } \{c\in\mathbf{M} :\ M \models c{<}a \wedge A^{[c]} = B^{[c]}\} \in \mathcal{U}_a$$

The definitions of $A(c)$ and $A^{[c]}$ are given in the end of Section 1.3. The $\sim_1$ equivalence class of A is denoted by $A_{/1}$, the $\sim_2$ equivalence class of A with $A_{/2}$ (the filter $\mathcal{U}_a$ is usually clear from the context). We shall systematically confuse equivalence classes with their representatives. Let $\mathbf{M}/\mathcal{U}_a$ be the model whose first-order elements are the $\sim_1$ equivalence classes and whose second-order elements the $\sim_2$ equivalence classes. The relations and the functions of $\mathbf{M}/\mathcal{U}_a$ are defined in the canonical way. To every element c, $C\in\mathbf{M}$ we associate the elements $id(c)$ and $id(C)$ of $\mathbf{M}/\mathcal{U}_a$ in the usual way,

$$id(c) = \textstyle\bigcup_{x<a}\{x\} \times \{c\} \quad\text{and}\quad id(C) = \bigcup_{x<a}\{x\} \times C.$$

The map id is an embedding of $\mathbf{M}$ into $\mathbf{M}/\mathcal{U}_a$. So, as usual, we consider $\mathbf{M}$ as a substructure of $\mathbf{M}/\mathcal{U}_a$.

Fact 2.4.

(a) *If* $\mathbf{M}$ *is a model of* Σ^p_i*-comp then for every* $\Sigma^p_0(\Sigma^p_i)$*-formula* φ

$$\mathbf{M}/\mathcal{U}_a \models \varphi(\vec{B}_{/1}, \vec{C}_{/2}) \quad \Longleftrightarrow \quad \{d{\in}\mathbf{M} : \ \mathbf{M} \models (d{<}a) \wedge \varphi(\vec{B}(d), \vec{C}^{[d]})\} \in \mathcal{U}_a.$$

(b) $\mathbf{M}/\mathcal{U}_a$ *is an* $\mathcal{R}$*-extension of* $\mathbf{M}$.

Proof of (a). For atomic φ the lemma holds by definition. The inductive step for the Boolean connectives is immediate. We show that of the existential quantifiers. The $\Longrightarrow$ direction is again immediate. For the converse let us first consider first-order quantifiers. Suppose the lemma holds for $\varphi{\in}\Sigma^p_0(\Sigma^p_i)$ and let us show that it holds also for $\exists y\ \varphi$, where the bound on y is implicit in φ. Let us write φ_x for $\varphi(\vec{B}(x), \vec{C}^{[x]})$ and assume

$$\{x{\in}\mathbf{M} : \ \mathbf{M} \models (x{<}a) \wedge \exists y \varphi_x(y)\} \in \mathcal{U}_a.$$

Then by Σ^p_i comprehension there is a set A in $\mathbf{M}$ such that

$$\mathbf{M} \models \forall y\ (\forall x{<}a)[\varphi_x(y) \leftrightarrow \langle x, y\rangle{\in}A]$$

so,

$$\{x{\in}\mathbf{M} : \ \mathbf{M} \models (x{<}a) \wedge \varphi_x(A(x))\} \in \mathcal{U}_a,$$

and, by induction hypothesis, $\mathbf{M}/\mathcal{U}_a$ models $\varphi(A_{/1})$ and hence $\exists y \varphi(y)$. Now, let us consider second-order quantifiers. The proof is similar; we can assume $i{>}0$ otherwise there is nothing to prove. Suppose the lemma holds for $\varphi{\in}\Sigma^p_i$ and let us show that it is holds also for $\exists Y \varphi$, where the bound on Y is implicit in φ. If

$$\{x{\in}\mathbf{M} : \ \mathbf{M} \models (x{<}a) \wedge \exists Y \varphi_x(Y)\} \in \mathcal{U}_a,$$

by Σ^p_i comprehension, (since $i{>}0$) there is a set A in $\mathbf{M}$ such that

$$\mathbf{M} \models \forall Y\ (\forall x{<}a)[\varphi_x(Y) \leftrightarrow \varphi_x(A^{[x]})]$$

so,

$$\{x{\in}\mathbf{M} : \ \mathbf{M} \models (x{<}a) \wedge \varphi_x(A^{[x]})\} \in \mathcal{U}_a,$$

and, by induction hypothesis, $\mathbf{M}/\mathcal{U}_a$ models $\varphi(A_{/2})$ and hence $\exists Y \varphi(Y)$.

Proof of (b). From (a) we have immediately that $\mathbf{M}/\mathcal{U}_a \prec_{\Sigma^p_0} \mathbf{M}$. To check cofinality, observe that $A_{/1}{<}id(|A|)$. It remains to check condition **(ii)** in the definition of $\mathcal{R}$-extension. Let $D_{/1}$ the diagonal element of $\mathbf{M}/\mathcal{U}_a$, i.e., $D = \{\langle x, x\rangle : \ x{<}a\}$. All elements $A_{/2}$ of $\mathbf{M}/\mathcal{U}_a$ are such that

$$\mathbf{M}/\mathcal{U}_a \models A_{/2} = \{y : \langle D_{/1}, y\rangle{\in}id(A)\}$$

In fact, observe that for all first-order elements $B_{/1}$ of $\mathbf{M}/\mathcal{U}_a$,

$$\langle D_{/1}, B_{/1}\rangle = \{x{<}a : \langle x, \langle x, B(x)\rangle\rangle\}_{/1},$$

so, by the definition of $id(A)$, $\langle D_{/1}, B_{/1}\rangle{\in}id(A)$ is equivalent to

$$\{x<a : \langle x, B(x)\rangle \in A\} \in \mathcal{U}_a.$$

But this is the same as

$$\{x<a : B(x) \in A^{[x]}\} \in \mathcal{U}_a,$$

which turns out to coincide, by definition, with $B_{/1} \in A_{/2}$. So, every second-order element $A_{/2}$ of $\mathbf{M}/\mathcal{U}_a$ is Σ_0^p-definable over the first-order element $D_{/1}$ and the second-order element $id(A)$. The latter is identified with an element of $\mathbf{M}$ since it is the image of A via the canonical embedding of $\mathbf{M}$ into $\mathbf{M}/\mathcal{U}_a$. Thus, also condition **(ii)** is fulfilled. ∎

Theorem 2.4. *Every model* $\mathbf{M}$ *of* Σ_i^p*-comp has an* $\mathcal{R}$*-extension to a model of* Σ_{i+1}^p*-choice.*

Proof. We construct a chain of $\mathcal{R}$-extensions

$$\mathbf{M} = \mathbf{M}_0 \prec_{\mathcal{R}} \mathbf{M}_1 \prec_{\mathcal{R}} \mathbf{M}_2 \prec_{\mathcal{R}} \ldots$$

by means of ultrapowers. By Fact 2.3, the chain is automatically Σ_{i+1}^p and all models in it are models of Σ_{i+1}^p*-comp*. Let $\{\varphi_s\}_{s}\in\omega$ be an enumeration with infinitely many repetitions of all formulas with parameters in $\mathbf{N}$, such that all parameters of φ_s are in $\mathbf{M}_s$ (see Theorem 2.2 for details on this enumeration). The chain is constructed so that for all $\varphi_s \in \Pi_i^p$, either (1) or (2) below holds.

(1) For every $a \in \mathbf{M}$ there is a $Z \in \mathbf{M}$ such that $\mathbf{M}_s \models (\forall x<a)\varphi_s(x, Z^{[x]})$.

(2) There is a $c \in \mathbf{M}_{s+1}$ such that $\mathbf{M}_{s+1} \models \forall Y \neg\varphi_s(c, Y)$.

Suppose $\mathbf{M}_s$ has already been constructed. If (1) holds in $\mathbf{M}_s$ then let $\mathbf{M}_{s+1} := \mathbf{M}_s$. Otherwise let $a \in \mathbf{M}_s$ be any element witnessing the failure of (1). Consider the ultrafilter on $[a]$ generated by the sets

$$\{x<a : \mathbf{M}_s \models \neg\varphi_s(x, Z^{[x]})\}$$

for Z in M. Since $\mathbf{M}_s \models \Sigma_i^p$*-comp*, the sets above enjoy the finite intersection property, so, such an ultrafilter actually exists. Let $\mathbf{M}_{s+1} := M_s/\mathcal{U}_a$.

To check that (2) holds, suppose not for a contradiction. Let D be the diagonal in $M_s/\mathcal{U}_a$. If $\exists Y \varphi_s(D_{/1}, Y)$ held in $\mathbf{M}_{s+1}$, then, by Fact **(a)** above, we would have that for some Y in M,

$$\{x<a : \ M_s \models \varphi_s(x, Y^{[x]})\} \in \mathcal{U}_a,$$

quod non since the complement of this set is, by construction, also in $\mathcal{U}_a$.

Finally, $\mathbf{N}$ is a model of Σ_{i+1}^p*-choice*, since the truth of both formulas in (1) and in (2) is preserved along the the $\exists\Sigma_{i+1}^p$-chain. ∎

3 The collapse of BA versus the collapse of PH

It is not known whether BA collapses, i.e., whether it is equal to some of its fragments. The only collapse that we are able to exclude is Σ_1^p-*choice* = $\mathcal{P}$-*def*. In fact, rudimentary functions $\mathcal{R}$ are the only Σ_1^p provably total functions of Σ_1^p-*choice* and a simple diagonalization argument shows these are strictly included in the polynomial time computable functions $\mathcal{P}$. Actually one can also see that the $\forall\Sigma_0^p$ fragment of $\mathcal{P}$-*def* strictly includes that of Σ_0^p-*comp* (and, by conservativity, that of Σ_1^p-*choice*). In fact, in [2] Ajtai has constructed a model $\mathbf{M}$ of $I\Delta_0(R)$ such that

$$\mathbf{M} \models \exists x\ R : [x] \hookrightarrow [x+1]$$

i.e., R is an injection of $[x]$ into $[x+1]$. We can expand $\mathbf{M}$ to a model $\mathbf{M}'$ of Σ_0^p-*comp* taking as sets of $\mathbf{M}'$ the finite $\Delta_0(R)$-definable sets of $\mathbf{M}$. Then $\mathbf{M}'$ falsifies the pigeonhole principle i.e., the sentence

$$\forall X\ :\ \forall x\ \neg\ X : [x+1] \hookrightarrow [x].$$

while the sentence above is easily seen to be provable in $\mathcal{P}$-*def*.

For stronger fragments we can only produce relativized results. The main result of this section is to prove that the collapse of BA is equivalent to the provable collapse of PH.

3.1 An interpolation theorem

The following is the 'bounded version' of a general interpolation theorem for classical predicate logic.

Theorem 3.1. *Let φ and ψ be $\forall\Pi_{i+2}^p$-formulas and let T be a $\forall\Pi_{i+2}^p$-axiomatized theory. If $T \vdash \varphi \to \neg\psi$ then there is a Boolean combination β of Σ_{i+1}^p-formulas such that, $T \vdash \varphi \to \beta$ and $T \vdash \beta \to \neg\psi$. Moreover all free variables of β occur free in $\varphi \to \neg\psi$.*

Proof. Let φ and ψ be as above and suppose that the required interpolant does not exist. We intend to show that $T + \varphi + \psi$ is consistent. (When the context suggests it, the free variables of $\varphi \to \neg\psi$ need to be replaced by fresh constants.) To show this, it is sufficient to show that there are two $\forall\Pi_{i+2}^p$-theories $U \supseteq T + \varphi$ and $V \supseteq T + \psi$ such that U and V have the same $\forall\Sigma_i^p$-consequences (we say also that they are mutually $\forall\Sigma_i^p$-conservative). In fact, we claim that, for any pair U and V of mutually $\forall\Sigma_i^p$-conservative theories which are $\forall\Pi_{i+2}^p$-axiomatizable, $U + V$ is consistent (and, actually, also has the same $\forall\Sigma_i^p$-consequences). Let us first prove this claim and then proceed to the construction of U and V. We construct a Σ_i^p-elementary chain of models,

$$\mathbf{M}_0 \prec_{\Sigma_i^p} \mathbf{M}_1 \prec_{\Sigma_i^p} ....,$$

such that $\mathbf{M}_{2s}$ is a model of U and $\mathbf{M}_{2s+1}$ is a model of V. It is possible to find $\mathbf{M}_{2s+1}$ and $\mathbf{M}_{2s+2}$ such that

$$\mathbf{M}_{2s+1} \models Diag_{\Pi_i^p}(\mathbf{M}_{2s}) + V \quad \text{and} \quad \mathbf{M}_{2s+2} \models Diag_{\Pi_i^p}(\mathbf{M}_{2s+1}) + U.$$

In fact, if we assume as induction hypothesis that $\mathbf{M}_{2s}$ is a model of U, then $Diag_{\Pi_i^p}(\mathbf{M}_{2s})+V$ is consistent, otherwise, for some $\theta\in\Pi_i^p$

$$V \vdash \forall\vec{x}\vec{X}\neg\theta \quad \text{and} \quad \mathbf{M}_{2s} \models \exists\vec{x}\vec{X}\ \theta$$

which contradicts the $\forall\Sigma_i^p$-conservativity of V over U. The symmetric argument works for odd stages. Finally, recall that both U and V are $\forall\Pi_{i+2}^p$-theories (and hence conserved by unions of Σ_i^p-chains). So, the union of the chain

$$\mathbf{N} := \bigcup_{s\in\omega}\mathbf{M}_s = \bigcup_{s\in\omega}\mathbf{M}_{2s} = \bigcup_{s\in\omega}\mathbf{M}_{2s+1}$$

is a model of both U and V. This proves the claim.

Now we construct U and V. Let $\vec{X}$ be all free variables occurring in $\varphi \to \neg\psi$. Let $\mathcal{B}_{i+1}$ denote the class of formulas with free variables among $\vec{X}$ of the form $\forall Y\beta$ and such that $\forall Y{<}|\vec{X}|^p\beta$ is a Boolean combination of Σ_{i+1}^p-formulas (for $p\in\omega$). Let us say that two theories U and V are $\mathcal{B}_{i+1}$ inseparable (in the following simply inseparable) if $V + Th_{\mathcal{B}_{i+1}}(U)$ is consistent. In other words, if there is no $\forall Y\beta\in\mathcal{B}_{i+1}$ such that $U \vdash \forall Y\beta$ and $V \vdash \neg\forall Y\beta$. Let $U_0 := T + \varphi$ and $V_0 := T + \psi$. If no interpolant exists, U_0 and V_0 are inseparable. In fact, suppose for a contradiction that

$$T + \varphi \vdash \forall Y\beta \quad \text{and} \quad T + \psi \vdash \neg\forall Y\beta$$

where $\forall Y\beta$ is in $\mathcal{B}_{i+1}$. Since T is axiomatized by $\forall PH$ sentences, we can apply a well-known theorem of Parikh's, to find a $p\in\omega$ such that

$$T \vdash \forall\vec{X}\ [\psi \to \neg\forall Y{<}|\vec{X}|^p\beta].$$

Therefore $\forall Y{<}|\vec{X}|^p\beta$ would be an interpolant of φ and ψ of the required complexity. Now, we show, by induction on s that the following theories are inseparable:

$$U_{s+1} = U_s + Th_{\mathcal{B}_{i+1}}(V_s) \quad \text{and} \quad V_{s+1} = V_s + Th_{\mathcal{B}_{i+1}}(U_s).$$

We have already shown the case $s = 0$. Suppose U_s and V_s are inseparable. If, for a contradiction, for some $\forall Y\beta$ in $\mathcal{B}_{i+1}$,

$$U_s + Th_{\mathcal{B}_{i+1}}(V_s) \vdash \forall Y\beta \quad \text{and} \quad V_s + Th_{\mathcal{B}_{i+1}}(U_s) \vdash \neg\forall Y\beta$$

then, for some $\forall Z\beta'\in Th_{\mathcal{B}_{i+1}}(V_s)$,

$$U_s \vdash \forall Z\beta' \to \forall Y\beta.$$

Applying again Parikh's theorem, for some $p\in\omega$,

$$U_s \vdash \forall Y\ [\forall Z{<}|Y|^p\beta' \to \beta].$$

therefore,

$$\forall Y\ [\forall Z{<}|Y|^p\beta' \to \beta] \in Th_{\mathcal{B}_{i+1}}(U_s).$$

But $V_s \vdash \forall Z\beta'$, so, $V_s + Th_{\mathcal{B}_{i+1}}(U_s)$ is inconsistent. This contradicts our induction hypothesis. Finally, let $U := \bigcup_{s\in\omega} U_s$ and $V := \bigcup_{s\in\omega} V_s$. Clearly,

$$Th_{\mathcal{B}_{i+1}}(U) = Th_{\mathcal{B}_{i+1}}(V).$$

So, in particular, U and V have the same $\forall\Sigma_i^p$-consequences. ∎

3.2 Sufficient conditions for the collapse of BA

Let us introduce some terminology. We say that a theory proves $\Pi^p_i = \Sigma^p_i$ if every Π^p_i-formula is provably equivalent to a Σ^p_i-formula (with the same free variables). In this case we also say that PH provably collapses to $\Pi^p_i = \Sigma^p_i$. We say that a theory proves $\Pi^p_{i+1} = \Sigma^p_{i+1}/poly$, if for every $\theta{\in}\Sigma^p_{i+1}$ there is a $\psi{\in}\Pi^p_{i+1}$ and a $p{\in}\omega$ such that, provably

$$(\exists W{<}c^p)(\forall X{<}c)\Big[\theta(X) \leftrightarrow \psi(X,W)\Big].$$

(All variables are shown.) The W is usually called a **(polinomial) advice**. Observe that, if $\Pi^p_{i+1} = \Sigma^p_{i+1}/poly$, then every bounded formula of the form $X{<}c \wedge \varphi(X)$ is equivalent to a Σ^p_{i+1}-formula depending on additional parameters. The following is an interesting consequence of Lemma 2.2 and Lemma 2.3.

Theorem 3.2. *The following are sufficient conditions for $\mathcal{P}_i$-def $\vdash BA$*

(a) $\mathcal{P}_i\text{-}def + \Sigma^p_{i+1}\text{-}choice \;\vdash\; \Pi^p_{i+2} = \Sigma^p_{i+2}$,

(b) $\mathcal{P}_i\text{-}def + \Sigma^p_{i+1}\text{-}choice \;\vdash\; \Pi^p_{i+1} = \Sigma^p_{i+1}/poly$.

Proof. By Corollary 2.3 **(b)**, in both cases it is sufficient to show that $\mathcal{P}_i$-*def* + Σ^p_{i+1}-*choice* proves BA. Let us prove **(a)**. Every model $\mathbf{M}$ of $\mathcal{P}_i$-*def* + Σ^p_{i+1}-*choice* has an Σ^p_{i+1}-elementary extension to a model of Σ^p_{i+1}-*comp*. By the provable collapse of PH every bounded formula is equivalent both to a Π^p_{i+2} and to a Σ^p_{i+2}-formula. Therefore every Σ^p_{i+1}-elementary extension is actually PH-elementary. So $\mathbf{M}$ is a model of Σ^p_{i+1}-*comp* too. By the interpolation lemma above every PH-formula is equivalent to a Boolean combination of Σ^p_{i+1}-formulas. For this class of formulas comprehension is provable in Σ^p_{i+1}-*comp*.

Let us prove **(b)**. We show that the choice schema holds for every bounded formula. We may use Σ^p_{i+1}-*choice*. By $\Pi^p_{i+1} = \Sigma^p_{i+1}/poly$, every bounded formula is equivalent to a Σ^p_{i+1}-formula depending on some additional parameters (i.e., the advices which transform universal in existential quantifiers and vice versa). ∎

3.3 Necessary conditions for the collapse of BA

Here we show that if $\mathcal{P}_i$-*def* proves Σ^p_{i+1}-*comp* then it proves the collapse of PH and BA reduces to $\mathcal{P}_i$-*def*. We need the following lemma of [12] which is known as the KPT witnessing theorem.

Lemma 3.3. *For every $\varphi{\in}\Pi^p_i$ if $\mathcal{P}_i$-def proves $\forall X\exists Y\forall Z\varphi(X,Y,Z)$, then there are $F_0,..,F_{n-1}$ in $\mathcal{P}_i$ such that $\mathcal{P}_i$-def proves*

$$\forall X, Z_0,..,Z_{n-1}\ \bigvee\begin{cases}\varphi(X,F_0(X),Z_0)\\ \varphi(X,F_1(X,Z_0),Z_1)\\ \dots\\ \dots\\ \varphi(X,F_{n-1}(X,Z_0,\dots,Z_{n-2}),Z_{n-1})\end{cases}$$

Proof. Let $\{F_n\}_{n\in\omega}$ be an enumeration of all the functions in $\mathcal{P}_i$ with infinitely many repetitions. Let C, $\{D_n\}_{n\in\omega}$ be fresh constants. Consider the theory

$$\mathcal{P}_i\text{-}def + \{\neg\varphi(C, F_n(C, \vec{D}_n), D_n) \ : \ n\in\omega\}$$

where $\vec{D}_n$ stands for $D_1, .., D_{n-1}$. If this theory is inconsistent, our claim follows by compactness. So, we suppose for a contradiction that this theory has a model. Let $\mathbf{M}$ be the $\mathcal{P}_i$-closure of C, $\{D_n\}_{n\in\omega}$ in the model of the theory above. By Lemma 2.1, $\mathbf{M}$ is a $\Sigma_i^p(\mathcal{P}_i)$-elementary substructure, so,

$$\mathbf{M} \models \neg\varphi(C, F_n(C, \vec{D}_n), D_n)$$

But, in $\mathbf{M}$, every possible witness of $\exists Y \forall Z \varphi(C, Y, Z)$ is of the form $F_n(C, \vec{D}_n)$. A contradiction. ∎

For the next theorem we use ideas of [9] as we learned them from Harry Buhrman.

Theorem 3.3. $\mathcal{P}_i\text{-}def \vdash \Sigma_{i+1}^p\text{-}comp \implies \mathcal{P}_i\text{-}def \vdash \Pi_{i+1}^p = \Sigma_{i+1}^p/poly$.

Proof. Consider an arbitrary formula of the form $\exists Z \varphi(X, Z)$ for $\varphi\in\Pi_i^p$ where a bound on Z is implicit in φ. We shall find a formula $\psi\in\Pi_{i+1}^p$ such that $\mathcal{P}_i\text{-}def$ proves

$$\exists W (\forall X{<}c) \left[\exists Z \varphi(X, Z) \leftrightarrow \psi(X, W)\right].$$

Since, by lemma 1.6, $\Sigma_{i+1}^p\text{-}comp$ is equivalent to $\Sigma_{i+1}^p\text{-}coll$, we can assume that $\mathcal{P}\text{-}def$ proves the following sentence

$$\forall X \exists Y (\forall x{<}a) \left[\exists Z \varphi(X^{[x]}, Z) \rightarrow \varphi(X^{[x]}, Y^{[x]})\right].$$

This sentence says that for every string of sets $X^{[0]}, .., X^{[a-1]}$ there is a string $Y^{[0]}, .., Y^{[a-1]}$ coding witnesses, (if any exists) of $\exists Z\varphi(X^{[0]}, Z), .., \exists Z\varphi(X^{[a-1]}, Z)$. So, assume this is provable in $\mathcal{P}_i\text{-}def$, move the quantifiers $\exists Z$ as far to the left as possible and apply the previous lemma to this formula. Then fix $a = n$ and, for better readability, let us suppose $n = 2$.

$$\forall X, A, B \ \bigvee \begin{cases} (\forall x{<}2) \left[\varphi(X^{[x]}, A) \ \rightarrow \ \varphi(X^{[x]}, F_1^{[x]}(X))\right] \\ (\forall x{<}2) \left[\varphi(X^{[x]}, B) \ \rightarrow \ \varphi(X^{[x]}, F_2^{[x]}(X, A))\right] \end{cases}$$

We can replace universal quantification with conjunction. Also, to streamline notation, let us use two variables X, Y in place of $X^{[0]}$ and $X^{[1]}$ and introduce the functions F, G and H, K in place of the two components of F_1 and F_2. The formula above can be rewritten as $\forall X, Y, A \ \gamma(X, Y, A)$ where

$$\gamma(X, Y, A) :\equiv \bigvee \begin{cases} \bigwedge \begin{cases} \varphi(X, A) \rightarrow \varphi(X, F(X, Y)) \\ \varphi(Y, A) \rightarrow \varphi(Y, G(X, Y)) \end{cases} \\ \\ \bigwedge \begin{cases} \exists B \varphi(X, B) \rightarrow \varphi(X, H(X, Y, A)) \\ \exists B \varphi(Y, B) \rightarrow \varphi(Y, K(X, Y, A)) \end{cases} \end{cases}$$

Let ξ stand for the first disjunct of γ, i.e., for the formula

$$\xi(X, Y, A) :\equiv \wedge \begin{cases} \varphi(X, A) \to \varphi(X, F(X, Y)) \\ \varphi(Y, A) \to \varphi(Y, G(X, Y)) \end{cases}$$

Now, we define the formula $\psi(X, W)$ to be

$$\vee \begin{cases} \varphi(X, F(X, W)) \\ c \in W \ \wedge \ (\forall Y<c)\, \forall A\ [\neg\xi(Y, X, A) \to \varphi(X, K(Y, X, A))] \end{cases}$$

Recall that a polynomial bound for the quantifier $\forall A$ is implicit in φ. So, $\psi(X, W)$ is a Π^p_{i+1}-formula. To complete the proof we have to show that for every c there is an advice W such that $\exists Y\varphi(X, Y) \leftrightarrow \psi(X, W)$ for every $X<c$. Let c be given, we proceed in a nonuniform way. We consider two possibilities.

($\circ$) Suppose there is a $Y<c$ such that $\xi(X, Y, A)$ holds for every $X<c$ and every A. Let $W = Y$. From $\xi(X, W, A)$ it follows that $\exists A\varphi(X, A)$ implies $\varphi(X, F(X, W))$ and so, $\psi(X, W)$. The converse is obvious since we have chosen a $W<c$, so the second disjunct is always false.

($\circ\circ$) Suppose case ($\circ$) does not obtain, i.e., (reversing the roles of X and Y) suppose for all X, $(\exists Y<c)\ \exists A\ \ \neg\xi(Y, X, A)$. We chose a W which informs us of this fact: $W = \{c\}$. If $\exists A\varphi(X, A)$ does not hold then in particular $\varphi(X, F(X, W))$ and $\neg\varphi(X, K(Y, X, A))$ for all W, Y and A. So, for A and Y such that $\neg\xi(Y, X, A)$, $\psi(X, W)$ fails. Vice versa, assume $\exists B\varphi(X, B)$. For all Y and A such that $\neg\xi(Y, X, A)$, the second disjunct in $\gamma(Y, X, A)$ must be true. So, since $\exists B\varphi(X, B)$, we have $\varphi(X, K(Y, X, A))$. Thus the second disjunct of $\psi(X, W)$ holds.

This completes the proof under the condition $n = 2$. The general case is similar. One has to consider n cases in place of 2 and the advice W must inform of which case actually obtains for a given c. Details are left to the reader. ∎

Corollary 3.3.

(a) $\mathcal{P}_i\text{-}def \ \vdash \ \Sigma^p_{i+1}\text{-}comp \quad \Longrightarrow \quad \mathcal{P}_i\text{-}def \ \vdash \ BA$

(b) $\mathcal{P}_i\text{-}def \ \vdash \ \Sigma^p_{i+1}\text{-}comp \quad \Longrightarrow \quad \mathcal{P}_i\text{-}def \ \vdash \ \Pi^p_{i+3} = \Sigma^p_{i+3}$

Proof. **(a)** follows immediately from Theorem 3.2. To prove **(b)**, we can assume that $\theta(A) \in \Pi^p_{i+3}$ has the form $(\forall X<c)(\exists Y<c)\varphi(X, Y, A)$ for some $\varphi \in \Pi^p_{i+1}$. We want a Σ^p_{i+3}-formula equivalent to θ. From $\Pi^p_{i+1} = \Sigma^p_{i+1}/poly$ we have that, provably in $\mathcal{P}_i$-def, for some $\psi \in \Sigma^p_{i+1}$, (omitting the bound on W)

$$\exists W (\forall X, Y, A<c)\ \big[\varphi(X, Y, A) \leftrightarrow \psi(X, Y, W, A)\big]$$

Note that the formula saying that W is a good advice for all $X, Y<c$,

$$(\forall X, Y, A<c)\ \big[\varphi(X, Y, A) \leftrightarrow \psi(X, Y, A, W)\big]$$

is Π^p_{i+2}. So, let $\zeta(W, A)$ stand for this formula. Provably in $\mathcal{P}_i$-def,

$$(\forall X<c)(\exists Y<c)\varphi(X, Y, A) \leftrightarrow \exists W[\zeta(W, A) \wedge (\forall X<c)(\exists Y<c)\psi(X, Y, W, A)]. \quad ∎$$

3.4 Krajíček, Pudlák and Takeuti's method

Krajíček, Pudlák and Takeuti have shown in [12] that if $\mathcal{P}_i$-*def* proves Σ^p_{i+1}-*comp* then $\Sigma^p_{i+1} = \mathcal{P}_i/poly$ in the standard model (and hence $\Sigma^p_{i+2} = \Pi^p_{i+2}$). We show how their result can be obtained by sharpening the reasoning of the previous section. The combinatoral methods used in the following proof are of a more complex nature than those needed in the previous section. It is still unknown whether this proof can be formalized in BA.

We say that $\Sigma^p_{i+1} = \mathcal{P}_i/poly$ if for every Σ^p_{i+1}-formula $\exists Y\varphi(X, Y)$ there is a $\mathcal{P}_i$-function F such that for some $p\in\omega$,

$$(\exists W<c^p)(\forall X<c)\left[\exists Y\varphi(X, Y) \rightarrow \varphi(X, F(X, W))\right].$$

Theorem 3.4. *If $\mathcal{P}_i$-def $+ \Sigma^p_{i+1}$-choice $\vdash$ Σ^p_{i+1}-comp then in the standard model $\Sigma^p_{i+1} = \mathcal{P}/poly$.*

Proof. By Corollary 2.3 we can as well assume that $\mathcal{P}_i$-*def* $\vdash \Sigma^p_{i+1}$-*comp*. Let $\exists Y\varphi(X, Y)$ be in Σ^p_{i+1}. Reasoning as in the proof Theorem 3.3 (so, assuming again that the KPT witnessing theorem holds with $n = 2$ for the formula under consideration) we obtain that the formula $\forall X, Y, A\ \gamma(X, Y, A)$ defined there is provable in $\mathcal{P}_i$-*def*. In particular, it holds in the standard model. For the rest of the argument let us work in ω. We say that X has information about Y if one of the following cases hold

(a) $\varphi(Y, F(Y, X))$,

(b) $\varphi(Y, K(X, Y, A))$, for all A such that $\varphi(X, A)$.

Observe that if X has information about Y, then knowing any witness of $\exists A\varphi(X, A)$ we can compute a witness of $\exists A\varphi(Y, A)$. Now, we claim that for any pair of sets $X, Y<c$ such that $\exists A\varphi(X, A)$ and $\exists A\varphi(Y, A)$ either X has information about Y or vice versa. To prove the claim, suppose X has no information about Y. In particular $\varphi(Y, F(Y, X))$ does not hold. Let A be any witness of $\exists A\varphi(Y, A)$ then, by $\gamma(Y, X, A)$, (the roles of X and Y are interchanged) $\exists B\varphi(X, B) \rightarrow \varphi(X, K(Y, X, A))$ must hold. Therefore, $\varphi(X, K(Y, X, A))$ follows, so, by **(b)**, Y has information about X.

Consider now the class $Q = \{X<c\ :\ \exists A\varphi(X, A)\ \}$ and reason in the standard model. There is a $X\in Q$ such that X has information about at least half of the sets in Q. To see this, let $i(X, Y)$ be 1 if X has information about Y, -1 otherwise. Then, by our claim above, $\sum_{X,Y\in Q} i(X, Y) = 0$, so, for some X in Q, $\sum_{Y\in Q} i(X, Y) \geq 0$. Clearly such an X has information about at least half of the Y in Q. Iterating the argument above, since Q contains at most 2^c-elements, we obtain $W<\langle c, c\rangle$ such that $W^{[0]}, .., W^{[c-1]}$ have information about all elements of Q. Let V be such that $\varphi(W^{[i]}, V^{[i]})$ for $i = 0, .., c-1$. Then, we have that for all $X<c$

$$\exists A\varphi(X, A) \leftrightarrow (\exists x<c)\left[\varphi(X, F(X, W^{[x]})) \vee \varphi(X, K(W^{[x]}, X, V^{[x]}))\right].$$

That is, for some function $F'\in\mathcal{P}_i$ and some W' coding W and V,

$$(\forall X<c)\left[\exists Y\varphi(X, Y) \leftrightarrow \varphi(X, F'(X, W'))\right].$$

Recall a bound on $\exists Y$ is implicit in φ so, the size of V can be bounded by some standard ower of c. Hence $W'<c^p$ for some $p\in\omega$.

The general case (for $n>2$) is similar. ∎

References

[1] M. Ajtai, Σ_1^1-formulas on finite structures. *Annals of Pure and Applied Logic*, **24**, (1983), 1-48.

[2] M. Ajtai, The complexity of the pigeonhole principle. *IEEE*, (1988), 346-355.

[3] S. R. Buss, *Bounded arithmetic.* Bibliopolis, Naples, (1986).

[4] S. R. Buss, Axiomatizations and conservations results for fragments of bounded arithmetic. In *Logic and Computation*, (proceeding of a workshop held at Carnegie-Mellon University, 1987), Contemporary Mathematics, A.M.S., **106**, (1990), 57-84.

[5] S. R. Buss, Relating the bounded Arithmetic and polynomial time hierarchies. To appear (199?).

[6] A. Cobham, The intrinsic computational difficulty of functions. In: A.L.Selman (ed.), Structure in complexity theory, *Lecture Notes in Computational Science*, **221**, (1986), 125-146.

[7] K. J. Devlin, *Constructibility.* Springer Verlag, Berlin, 1984.

[8] P. Hájek, P. Pudlák, *Metamathematics of First-order Arithmetic.* Springer-Verlag, Berlin, 1993

[9] J. Kadin, The polynomial time hierarchy collapses if the Boolean hierarchy collapses. *IEEE*, (1988), 278-292.

[10] J. B. Paris and L. A. S. Kirby, Σ_n-collection schemas in arithmetic. In *Logic Colloquium 77*, A. Macintyre, L. Pacholski, J. Paris (eds.). North-Holland, Amsterdam (1978), 199-209.

[11] J. Krajíček, Exponentiation and second order bounded arithmetic. *Annals of Pure and Applied Logic*, **48**, (1990), 261-276

[12] J. Krajíček, P. Pudlák, G. Takeuti, Bounded arithmetic and the polynomial time hierarchy. *Annals of Pure and Applied Logic*, **52**, (1991), 143-153.

[13] G. E. Mints, Quantifier-free and one-quantifier systems. In: Yu. V. Matijasevich and A. O. Slisenko ed., *Zapiski Nauchnykh Seminarov* **20**, (1971), 115-133. (In Russian. English translation: *Journal of Soviet Mathematics* **1**, (1973), 211-266.)

[14] C. Parsons, On a number theoretic choice schema and its relation to induction. In *Intuitionism and proof theory*, eds. Kino, Myhill and Vesley, North-Holland, Amsterdam, (1970), 459-473.

[15] A. Razborov, An equivalence between second order bounded domain bounded arithmetic and first order bounded arithmetic, in *Proof Theory and Computational Complexity.* In: P. Clote and J. Krajíček ed., Oxford University press, Oxford, 1993, 247-277

[16] G. Takeuti, *RSUV isomorphisms.* In: P. Clote and J. Krajíček (ed.). *Arithmetic, Proof Theory and Computational Complexity.* Oxford University press, Oxford, 1993, 364-386

[17] A. J. Wilkie, Modelès non standard de l'arithmétique et complexité algorithmique. In: A. J. Wilkie and J. Ressayre, *Modelès non standard de l'arithmétique et theorie des ensembles.* Publications Mathématique de l'Université Paris VII, Paris 1-45.

Chapter 2. End extensions of models of linearly bounded arithmetic

Abstract

We prove that every model of $I\Delta_0$ has an end extension to a model of a theory extending Buss' S_2^0 in which all logspace computable function are formalizable. We also show the existence of an isomorphism between models of $I\Delta_0$ and models of LA (i.e., second-order Presburger arithmetic with finite comprehension for bounded formulas).

Contents

0 Introduction **31**

1 Preliminaries **31**

1.1 Linearly and polynomially bounded arithmetic 32

1.2 The first-second-order isomorphism 33

1.3 A digression on fragments and complexity theory 35

2 Bootstrapping **36**

2.1 First-order multiplication 36

2.2 First-order exponentiation 38

2.3 A well-ordering of the sets 40

3 Proof of the main theorem **41**

4 Appendix **45**

0 Introduction

When defining functions in bounded arithmetic, one often uses -more or less explicitly- the following schema,

$$(\forall x{<}a)(\exists y{<}a)\varphi(x,y) \to (\forall x{<}a)\forall b\,\exists z\,[(z)_0 = x \wedge (\forall w{<}|b|)\varphi((z)_w,(z)_{w+1})],$$

where $\varphi(x,y,w)$ is a bounded formula and $(z)_w$ is the w-th value of the string z under some reasonable coding of strings. This schema of *dependent choices* allows one to iterate logarithmically many times functions which are definable by bounded formulas. In fact, if we think of $\varphi(x,y)$ as defining a function $F(x)=y$, then the string z codes the values of the iterations: $F(x), F^2(x), \ldots, F^w(x), \ldots, F^{|b|}(x)$. The bound a in the schema takes care that the final output remains bounded. This schema assumes the existence of $2^{|a|\cdot|b|}$. In fact, this is the typical size of a number coding a string of $|b|$ numbers less than a. So, in general, it is not true in models of $I\Delta_0$ unless Ω_1 (see [5]) holds there. Anyhow, when a is not too large, it is still possible to find a Δ_0-function $Z(x,w)$ which produces the values $F^w(x)$ as above. Actually if $2^{|a|^{1+\epsilon}}$ exists for some positive standard rational ϵ the definition of Z may be found by repeated applications of the divide and conquer method.

This fact is the essential ingredient of the construction that we are going to present. We construct an end extension conteining all numbers z which code functions like Z above.

The end extension we are going to construct is a model of a theory axiomatized by S^0_2 plus an axiom which says that in every (coded) directed graph without terminating nodes there is a (coded) path of arbitrary (but, clearly, logarithmic) length. This theory is one of the weakest fragments for which the problem of whether it coincides or not with S_2 is non-trivial[1]. Recall also that it is not known whether $I\Delta_0{+}\Omega_1$ is a conservative extension of $I\Delta_0$ or whether every model of $I\Delta_0$ has an end extension to a model of $I\Delta_0{+}\Omega_1$.

For independent reasons we are also interested in giving a translation of $I\Delta_0$ into the second-order theory of addition. The technical difficulties involved in constructing the translation and in proving it correct will be circumvented by using the second-order version of the end extension result mentioned above.

Acknowledgments. Discussions with Mark Jumelet and Albert Visser have been fruitful.

1 Preliminaries

Our basic languages are $L_2(+,\cdot)$ and $L_2(+)$, i.e., that of second-order arithmetic with and respectively without the symbol of multiplication. Specifically, $L_2(+,\cdot)$ consists of two symbols for constants: 0, 1, two symbols for binary functions: $+$, $\cdot$ and two symbols for binary relations: $<$, $\in$. Moreover, there are two sorts of variables: first and second-order. Lower

[1]This theory, let us denote it by *Logrec*, is contained in PV_1 (or, in the notation of [6], $\mathcal{P}$-*def*). We know (see [6] and [2]) that if $PV_1 = S^1_2$ then $PV_1 = S_2$ and PV_1 proves the collapse of the polynomial time hierarchy. An intriguing question of Sam Buss is whether and what we can get more from the hypothesis $Logrec = S_2$.

case Latin letters $x, y, z, \ldots$ denote first-order variables and capital Latin letters $X, Y, Z, \ldots$ second-order variables. The language $L_2(+)$ coincides with $L_2(+,\cdot)$ but for the absence of the symbol $\cdot$ of multiplication.

First and second-order variables are meant to range respectively over numbers and finite sets of numbers. Terms are constructed from first-order variables only. The formula $x<y$ is to be read "x is less than y". The intended meaning of $X<y$ is: "all elements of X are less than y". Let t be a term in which x does not occur. We adopt the following abbreviations with the usual meaning

$$(Qx<t)\varphi, \ (Qx\in Y)\varphi, \ (QX<t)\varphi,$$

where Q is either $\forall$ or $\exists$. Quantifiers occurring in one of these contexts are called **bounded quantifiers**. Specifically, we shall speak about **polynomial quantifiers** or **linear quantifiers** according to whether the bounding term t is in $L_2(+,\cdot)$ or in $L_2(+)$. A formula is **polynomial**, respectively **linear**, if all of its quantifiers are. The set of polynomial formulas denoted by PH. The set of linear formulas by LH.

We classify bounded formulas of PH and LH in the **polynomial hierarchy** and **linear hierarchy** by counting alternations of second-order quantifiers. We use one of the symbols Π_0^p or Σ_0^p for the class of polynomial formulas containing atomic formulas and closed under Boolean connectives and polynomial first-order quantifiers. We define inductively Σ_{i+1}^p as the minimal class of formulas containing Π_i^p, closed under disjunction, conjunction and polynomial existential quantification. The class Π_{i+1}^p is the minimal class of formulas containing Σ_i^p, closed under disjunction, conjunction and polynomial universal quantification. So, PH equals $\bigcup_{i\in\omega}\Sigma_i^p$ and $\bigcup_{i\in\omega}\Pi_i^p$. The classes Σ_i^l and Π_i^l are defined analogously but atomic formulas and all quantifiers are required to be linear. Clearly Σ_i^l and Π_i^l coincide with the intersections of Σ_i^p and Π_i^p with $L_2(+)$.

If Γ is a class of formulas we write $\Sigma_i^l(\Gamma)$ for the class defined exactly as Σ_i^l but staring with Γ in place of the open formulas. Similarly for Σ_i^p.

1.1 Linearly and polynomially bounded arithmetic

Second-order **polynomially bounded arithmetic** (BA) is axiomatized by the following set of proper axioms plus the schema below where φ is a polynomial formulas not containing free occurrences of the variable X. This schema is called of **finite comprehension**. (The expressions $a \leq b$, $A = \emptyset$ and $A\subseteq B$ stand for the usual abbreviations.)

$0 \neq 1$ $\qquad$ $a{\cdot}(b{+}1) = (a{\cdot}b){+}a$

$a{+}0 = a$ $\qquad$ $a \leq b \longleftrightarrow a{<}b{+}1$

$a{+}1 = b{+}1 \rightarrow a = b$ $\qquad$ $a \leq b{+}1 \longleftrightarrow a{<}b$

$a{+}(b{+}1) = (a{+}b){+}1$ $\qquad$ $A{<}b \longleftrightarrow (\forall x{\in}A)\ x{<}b$

$a \neq 0 \longleftrightarrow (\exists x{<}a)\ x{+}1 = a$ $\qquad$ $A = B \longleftrightarrow A{\subseteq}B \wedge B{\subseteq}A$

$a{\cdot}0 = 0$ $\qquad$ $A \neq \emptyset \ \rightarrow\cdot\ (\exists x{\in}A)(A{<}x{+}1)$

$$(\exists X{<}a)(\forall x{<}a){\cdot}\ x{\in}X \ \longleftrightarrow\ \varphi(x).$$

The set of proper axioms above is denoted by Θ^p. The set of those axioms of Θ^p which are formulas of the language $L_2(+)$ is denoted by Θ^l. The last axiom deserves some special remark. It is the conjunction of a bounding axiom and a least number principle (it claims the existence of the least upper bound of every set). The least upper bound of the set X will be denoted by $|X|$; the largest element of a non-empty X is then $|X|-1$. The theory of **linearly bounded arithmetic** (LA) is axiomatized by those axioms above which are formulas of $L_2(+)$, i.e., Θ^l plus finite comprehension for linear formulas.

1.2 The first-second-order isomorphism

Second-order models are composed of two disjoint parts: the numbers and the sets. The disjoint union of ω and $\mathcal{P}_{<\omega}(\omega)$ constitutes the **standard model**. There, functions and relations are interpreted in the natural way. Non-standard second-order models are always denoted with boldface capital letters, first-order models (or the first-order parts of second-order models) are denoted by (the respective) boldface lower-case letters. We often identify second-order elements of $\mathbf{M}$ with their extensions, i.e., with actual subsets of $\mathbf{m}$. Given a first-order model $\mathbf{m}$ in which the usual notion of logarithm and of binary string are formalizable, we construct a second-order model with a canonical procedure. Namely, let $\log x$ and $\operatorname{Log} x$ denote respectively the logarithm of $x{\in}\mathbf{m}$ and the set

$$\operatorname{Log} x := \{y{<}\log x \ :\ (x)_y = 1\}$$

(where $(x)_y$ is the y-th digit in the binary expansion of x). The second-order model $\mathbf{Log\,m}$ is defined

$$\mathbf{Log\,m} := \{\log x, \operatorname{Log} x \ :\ x{\in}\mathbf{m}\}.$$

Relations and functions in $\mathbf{Log\,m}$ are defined in the natural way. It is routine to check that if $\mathbf{m}$ is a model of $I\Delta_0{+}\Omega_1$ then $\mathbf{Log\,m}$ is a model of BA. In general, if $\mathbf{m}$ is only a first-order model of $I\Delta_0$, $\mathbf{Log\,m}$ is a model of LA. In fact, the first-order sentence Ω_1 asserts exacly the closure of the logarithmic cut under multiplication.

A natural question is whether every second-order model $\mathbf{M}$ of BA or of LA is the Logarithm of some $\mathbf{n}$ model of $I\Delta_0{+}\Omega_1$ or, respectively, $I\Delta_0$. The answer is affermative. For BA this is relatively simple to check. The domain of the model $\mathbf{n}$ consists of the second-order elements of $\mathbf{M}$. They are interpreted as numbers, namely as the numbers

$$\sum_{x \in X} 2^x.$$

One has to define 0_2, 1_2, $+_2$, $\cdot_2$ and $<_2$ in $\mathbf{n} := \{X : X \in \mathbf{M}\}$ in order that $\mathbf{n}$ satisfies all axioms of $I\Delta_0+\Omega_1$. The definition of 0_2, 1_2 and $<_2$ is immediate

$$0_2 := \emptyset, \quad 1_2 := \{0\}, \quad X <_2 Y \overset{def}{\longleftrightarrow} X \neq Y \wedge |X \triangle Y|-1 \in Y$$

(recall that $|X \triangle Y|-1$ is the largest element which is in Y but not in X or vice versa). For addition and multiplication one has to formalize more or less directly the primary school algorithms for the arithmetical operations on numbers written in a binary base. In fact, these can be easily translated in polynomial formulas and BA will prove both totality and the recursive equations for the new second-order functions. Finally, we have to check (see Section 2.3 below) that the $<_2$-least number principle holds in $\mathbf{n}$ and that $\mathbf{M}$ is actually (isomorphic to) $\mathbf{Log}\,\mathbf{n}$. Note that for $X \in \mathbf{n}$, $\omega_1(X)$ is $\{|X|^2\}$.

In principle, a similar procedure works also for LA. But a direct formalization of the school algorithms is not possible anymore. The absence of multiplication force us to repeated use of divide and conquer techniques to formalize these algorithms. So, the final check of the recursive equations cannot be fairly left to the reader. So, we shall avoid the use of the first-second-order isomorphism for LA but obtain it indirectly from our main theorem.

The first-second-order isomorphism holds also for fragments of BA. Clearly the first-order theory corresponding to these fragments will only be a subtheory of Buss' S_2 (see [1]). One of the weakest fragments of BA that can still be interpreted as first-order theory is Σ_0^p-*rec*. It is axiomatized by Σ_0^p-*comp* (i.e., Θ^p plus finite comprehension for Σ_0^p-formulas) and an axiom which allows recursion on first-order Σ_0^p-definable functions:

$$(\forall x{<}a)(\exists y{<}a)\varphi(x,y) \rightarrow (\forall x{<}a)\exists Z\ [Z(0) = x \wedge (\forall w{<}b)\varphi(Z(w), Z(w{+}1))],$$

where φ is Σ_0^p and $Z(x)$ is the value at x of the function coded by the set Z (in some natural coding of functions as sets). This schema (that by Σ_0^p-*comp* may also be given as a proper axiom) says that given a directed graph without terminating nodes and given an arbitrary node x in the graph, there is a set Z which codes an arbitrary long path with starting node x. It is easy to prove that this schema follows from BA but it is not known whether these two theories coincide. In the appendix we shall extensively comment on the more delicate details connected with the first-second-order isomorphism for models of this theory.

The main result of this paper is that every model of LA has an end extension to a model of Σ_0^p-*rec*. The result announced in the abstract follows from the following easy (but somewhat lengthy to check) considerations: transform any given model of $I\Delta_0$ into a model of LA as explained above. Then end extend this to a model of Σ_0^p-*rec* and, finally, apply the first-second-order isomorphism to it. Check that the first-order model obtained is actually an end extention of the original model of $I\Delta_0$ and that it is a model of S_2^0 plus the following schema (which is the obvious translation in a first-order language of the schema of recursion given above)

$$(\forall x{<}|a|)(\exists y{<}|a|)\varphi(x,y) \rightarrow (\forall x{<}|a|)\exists z\ [(z)_0 = x \wedge (\forall w{<}|b|)\varphi((z)_w, (z)_{w+1},$$

where $(z)_w$ is the w-th element of the string z and φ is Σ_0^b.

The first-second-order isomorphism for LA follows also from the main theorem. Given a model $\mathbf{M}$ of LA, we end extend it to a model of Σ_0^p-*rec*; this is isomorphic to some first-order model $\mathbf{n}$; the restriction of this isomorphism to $\mathbf{M}$ is what we are looking for. In fact, the image of $\mathbf{M}$ is an initial segment of $\mathbf{n}$. The Δ_0 least number principle holds in this segment since it corresponds to the second-order number principle in $\mathbf{M}$ (see Section 2.3).

1.3 A digression on fragments and complexity theory

We shall not consider fragments of the form Γ-*comp* here, i.e., fragments obtained by restricting the schema of comprehension to the formulas in some class Γ of LH, resp., PH, nor shall we study the relations between complexity theory and bounded arithmetic. For fragments of BA we refer the reader to [6] for a brief introduction or to [3] for a more comprehensive review. For fragments of LA such a systematic study is still lacking. We conclude this preliminary section with some observations on the difficulties arising in this field.

It is routine to show that classes definable by Σ_1^p-formulas coincide with NP languages and that, in general, the levels of the Meyer-Stockmeyer polynomial time hierarchy coincide with Σ_{i+1}^p and Π_{i+1}^p. Clearly, one needs to interpret finite sets as binary strings in the natural way. Linear formulas define sets recognisable by alternating linear time machines and the converse is also true. Unfortunately it is not immediate that classes Σ_{i+1}^l and Π_{i+1}^l correspond exactly to those of the linear time hierarchy. Observe that Σ_{i+1}^l and Π_{i+1}^l are not closed under bounded first-order quantification. When defining the polynomial hierarchy, to close or not to close Σ_i^p and Π_i^p under first-order bounded quantification is merely a stylistic question. Up to provable equivalence over the relative weak theory Σ_{i+1}^p-*choice*, this classes turn out to be closed under first-order bounded quantification. In fact, we can code pairs of numbers by first-order objects. So, a single set Z can code a whole sequence of sets $Z^{[0]}, \dots, Z^{[x]}$ where $Z^{[x]} := \{y : \langle x, y\rangle \in Z\}$. Consequently, the alternations of quantifiers $\forall x\, \exists Y\, \varphi(x, Y)$ turns out to be equivalent to $\exists Z\, \forall x\, \varphi(x, Z^{[x]})$ (where we omit bounds on the quantifiers). In the linear hierarchy the situation is much less trivial. Let Σ_{i*}^l be the minimal class containing Σ_i^l and closed under disjunction, conjunction, bounded existential quantification and bounded first-order quantification. Even in the standard model, we do not know whether Σ_{i*}^l coincides, up to equivalence, with Σ_i^l. It would be interesting to know whether for some j, Σ_j^l contains Σ_{i*}^l. A connected problem is to find a j such that Σ_j^l contains $\Sigma_0^l(\Sigma_j^l)$. These problems make the translation into complexity theory less smooth. We believe, one could possibly adapt the definition of Σ_{i+1}^l and Π_{i+1}^l in order to have a good coincidence with the classes of the linear time hierarchy but we consider this beyond the scope of the present work. Anyhow the problems just mentioned seem to us to be interesting. These problems have also an arithmetical version. It consists in the comparison of the fragments of LA where the comprehension schema is restricted to the formulas in Σ_i^l, Σ_{i*}^l and, respectively, $\Sigma_0^l(\Sigma_i^l)$. In BA the corresponding problem does not exist. Again the presence of a definable pairing function makes these three theories coincide. Most of the lemmas in the following sections are provable in $\Sigma_0^l(\Sigma_i^l)$-*comp* (or some natural restriction of them is provable). To formalise the whole construction in Σ_1^l -if possible- should require a more careful work.

2 Bootstrapping

This section is dedicated to some routine work of bootstrapping. We shall show that the graphs of multiplication and exponentiation are definable by linear formulas. They are not provably total but their specific recursive equations are provable in LA. In particular their inverse functions, division, roots and logarithms are total functions. In every model of LA every number codes a set and, for small sets, also the converse holds.

Intervals and co-intervals are used to code functions or strings in sets. An interval is a set of the form $\{x : a \leq x<a+b\}$ for some $b>0$ and is denoted by $[a, a+b)$. b is called the **length** of the interval. When $a = 0$ we write $[b)$ for $[0, b)$. An **interval of** Z is an interval which is contained in Z and which is maximal, i.e., it is not properly contained in an interval which is also a subset of Z. A **co-interval of** Z is an interval having empty intersection with Z and which is maximal, i.e., it is not properly contained in any interval having empty intersection with Z. Sometimes, when a set is clear from the context, we shall simply say **interval** and **co-interval**. The formula expressing "$[a, a+b)$ is an interval of Z" is Σ_0^l.

2.1 First-order multiplication

Definition 1 *When $a \geq b>0$, $M(a, b, c)$ is the formula asserting the existence of a set Z such that three conditions below are satisfied*

(o) $Z<c+b$

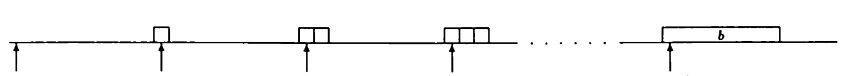
(i) $[0, 2a) \cap Z = \{a\}$.

(ii) *for every $x, y<c$ such that $0<x<y$*
$[x, y)$ *is an interval of Z iff $[x+a, y+a+1)$ is an interval of Z*

(iii) $[c, c+b)$ *is an interval of Z.*

When $b>a>0$, $M(a, b, c)$ assert the same but exchanging the roles of a and b. Finally if either one of a and b is 0 then $M(a, b, c)$ is simply $c = 0$. ∎

The formula $M(a, b, c)$ claims the existence of a set that can be depicted as follows



Let $b \leq a$. It is immediate to check by induction on x that, if Z witnesses $M(a, b, c)$, then for every $x \leq b$ there is a $c'\in Z$ such that $[c', c'+x)$ is an interval of Z and $Z \cap [c'+x)$ witnesses $M(a, x, c')$.

Lemma 2 *(LA) For every a, b there is at most one c such that $M(a, b, c)$.*

Proof. Assume $a \geq b$. Let Z' and Z'' witness respectively $M(a, b, c')$ and $M(a, b, c'')$. Let z the least element of $Z' \triangle Z''$. Say, $z \in Z'$ and $z \notin Z''$. Let $[x, y)$ be the interval of Z' containing z. From **(i)** it follows that there are $x', y' > 0$ such that $x' + a = x$ and $y' + a + 1 = y$. By **(ii)**, $[x', y')$ is an interval of Z' and, by the definition of interval, $y' < x$. So, $[x', y')$ is also an interval of Z'', otherwise z would not be the least of $Z' \triangle Z''$. Applying **(ii)** once again to Z'', we can conclude that $[x' + a, y' + a + 1) = [x, y)$ is an interval of Z'', a contradiction. ∎

Lemma 3 *(LA) $M(a, 0, 0)$ and if $M(a, b, c)$ then $M(a, b+1, c+a)$.*

Proof. The first assertion is true by definition. Assume $a \neq 0 \neq b$, otherwise it is easy. Let Z witness $M(a, b, c)$. There are two cases. If $a \geq b+1$, it is immediate to check that $Z \cup [c+a, c+a+b+1)$ witnesses $M(a, b+1, c+a)$. Otherwise, if $a < b+1$, we construct a set witnessing $M(a, b+1, c+a)$ stretching each co-interval of Z by one. In other words, we move upwards every interval of Z: the y-th interval of Z has to be shifted by y units. This operation is easy to define because the y-th interval of Z has length y. So, it is easy to verify that the set

$$Z' := \bigcup \{[x+y, x+2y) \ : \ [x, x+y) \text{ interval of } Z\},$$

witnesses $M(a, b+1, c+a)$. ∎

Lemma 4 *(LA) If $M(a, b, c)$, then for all $a' \leq a$ and $b' \leq b$ there is a $c' \leq c$ such that $M(a', b', c')$.*

Proof. We suppose $a \geq b$ and we shall prove the lemma for all $a' \geq b'$. By the symmetry of the definition, this is sufficient. For all $b' \leq b$ there is a $c' \leq c$ such that $M(a, b', c')$. In fact, let Z witness $M(a, b, c)$. Let $c' \leq c$ be that element of Z such that $[c', c'+b')$ is an interval of Z. As observed above, $Z' := Z \cap [c'+b')$ witnesses $M(a, b', c')$. Now, consider arbitrary $b' \leq b$ and $a' < a$ where $a' \geq b'$. Suppose, to obtain a contradiction, that for no $c' < c$, $M(a', b', c')$. Choose the least of such a'. So, by minimality, for some c'', there is a Z'' witnessing $M(a'+1, b', c'')$. Consider the set

$$Z' := \bigcup \{[x, x+y) \ : \ [x+y, x+2y) \text{ is an interval of } Z''\}.$$

Z' is obtained by decreasing the length of all co-intervals of Z by one. It is easy to check that Z' is the witness of $M(a', b', c')$. ∎

We write $\left[\frac{b}{a}\right]$ and $b^{\frac{1}{2}}$ for the maximal $q < b$ such that for some $b' < b$, $M(a, q, b')$, respectively, $M(q, q, b')$. Analogously we define for every standard n, $b^{\frac{1}{n}}$. We write $a \cdot b \downarrow$ for $\exists x\, M(a, b, x)$ and $a \cdot b \downarrow \in \mathbf{M}$ for $\mathbf{M} \models \exists x\, M(a, b, x)$. Similarly for $a^n \downarrow$.

Lemma 5 *(LA) If $a^n \downarrow$ then $(a+1)^n \downarrow$ (for n positive standard).*

Proof. Use Lemma 3 to prove, by induction on n standard,

$$(a+1)^n = \sum_{i<n} \binom{n}{i} a^i.$$

∎

Lemma 6 *Every model of $LA + \forall a\, a^2 \downarrow$ has an expansion to a model of BA.*

Proof. By Lemma 4, for all a, b there exists a c such that $M(a,b,c)$. So we expand M to a model $\mathbf{M}'$ of signature $L_2(+,\cdot)$ by defining $a{\cdot}b$ to be the unique c such that $M(a,b,c)$. By Lemma 3, Θ^p is satisfied in the expanded model. To verify that comprehension for all polynomial formulas holds, fix a polynomial formula φ with parameters in $\mathbf{M}'$ and a (large) number b in $\mathbf{M}'$. It suffices to set $b = d^n$ where d is the largest parameter in φ and n is the number of syntactical symbols in φ. Observe that every polynomial formula with parameters in $\mathbf{M}$ is equivalent to one where each atomic subformula contains at most one occurrence of $\cdot$ and where all quantifiers are bounded by b. Now, replacing atomic formulas of the form $r{\cdot}s = t$ with $M(r,s,t)$, we obtain a linear formula equivalent to the original one. So, in $\mathbf{M}'$, polynomial comprehension follows from linear comprehension. ∎

An immediate consequence of this lemma is that for every a model $\mathbf{M}$ of LA, the initial segment of $\mathbf{M}$ with domain

$$\{x, X \ : \ \mathbf{M}_L \models x^n \downarrow \wedge |X|^n \downarrow \quad \text{for all } n{\in}\omega\}$$

is (expandable to) a model of BA. We note that, using with some more care the ideas explained in the proof of Lemma 6 we would obtain the following lemma (cf. Lemma 1.30 of [3])

Lemma 7 *For every polynomial formula $\varphi(\vec{x}, \vec{X})$ there is a linear formula $\varphi^l(q, \vec{x}, \vec{X})$ such that $LA + \Theta^p$ proves*

$$\forall a\, \exists q\, (\forall \vec{x}, \vec{X} {<} a)\Big[\varphi(\vec{x}, \vec{X}) \longleftrightarrow \varphi^l(q, \vec{x}, \vec{X})\Big].$$

Moreover, we can assume that there is some standard n depending on φ such that, for all $q'', q' \geq a^n$ and for all $\vec{x}, \vec{X} {<} a$, $\varphi^l(q', \vec{x}, \vec{X})$ is equivalent over LA to $\varphi^l(q'', \vec{x}, \vec{X})$. ∎

We shall refer to the formula φ^l in the lemma above as the **linear translation** of φ.

2.2 First-order exponentiation

Definition 8 *$E(0,c)$ iff $c = 1$. If $b{>}0$, $E(b,c)$ holds iff there exists a set Z such that the four conditions below are satisfied*

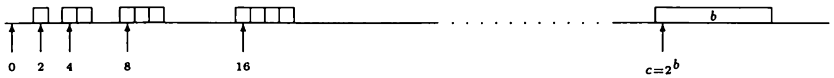
(o) $Z{<}c{+}b$

(i) $[0,4) \cap Z = \{2\}$.

(ii) *for every $x, y{<}c$ such that $0{<}y$*
$[x, x{+}y)$ is an interval of Z iff $[2x, 2x{+}y{+}1)$ is an interval of Z.

(iii) *$[c, c{+}b)$ is an interval of Z.* ∎

The formula $E(b, c)$ claims the existence of the set depicted below



Lemma 9 *(LA)*

(i) *For every b there is at most one c such that $E(b, c)$ and the Z witnessing this formula is unique.*

(ii) *For every b and c, if $E(b, c)$ then $E(b+1, 2c)$.*

(iii) *For every $b'<b$, if $E(b, c)$ there is a $c'<c$ such that $E(b', c')$.*

Proof. The first statement is proved as in Lemma 2; the second assertion is evident. The third statement follows since we can prove by induction on x that, if Z witnesses $E(b, c)$, for every $x \leq b$ there is a $c' \in Z$ such that $[c', c'+x)$ is an interval of Z and $Z \cap [c'+x)$ witnesses $E(b, c')$. ∎

The maximal $b<c$ such that, for some $c'<2c$, $E(b, c')$ is denoted by $\log c$ (if $a = 0$ we agree that $\log a = 0$). We define $(c)_a = 1$ if

$$(\exists d, b<c) \left[E(a, b) \ \wedge \ 2d+1 = \left[\frac{c}{b}\right]\right],$$

$(c)_a = 0$ otherwise. By linear comprehension, for every c there is a set the we denote by $\operatorname{Log} c$ such that

$$\operatorname{Log} c := \{x<\log c \ : \ (c)_x = 1\}.$$

The following lemma proves that small sets are coded by numbers and that this code is unique.

Lemma 10 *(LA) For every a and every $C<\log a$ there is a $c \leq a$ such that $C = \operatorname{Log} c$. Moreover this c is unique.*

Proof. We prove the assertion by induction on a. If $a = 0$ then $C = \emptyset$. It is easy to see that $\emptyset = \operatorname{Log} c$ iff $c = 0$. Suppose the lemma holds for every $C<\log a$ and let prove it for $C<\log(a+1)$. Clerly we may assume that $\log a<\log(a+1)$ otherwise there is nothing to prove. Clearly we have $\log a(a+1) = \log a + 1$ and $E(\log a, a)$. Fix such a $C<\log a + 1$. If $\log a \notin C$ then $C<\log a$ and the claim follows by induction hypothesis. So, assume $\log a \in C$. Apply the induction hypothesis to $C \setminus \{\log a\}$ and let c the unique number such that $C \setminus \{\log a\} = \operatorname{Log} c$. The reader will easily verify that $c+a$ is the code of C. For the unicity suppose that $c' \neq c+a$ is also a code of C'. We can check that

$$c \neq c'-a \quad \text{and} \quad \operatorname{Log} c = \operatorname{Log}(c'-a)\cdot$$

and that this contradicts the induction hypothesis. Details are left to the reader. ∎

Lemma 11 *(LA) For all a, if there exists a c such that $E(a,c)$, then for all standard n, $a^n\downarrow$.*

Proof. The proof, based on Lemma 5, is left to the reader. ∎

Lemma 12 *Let $\mathbf{M}$ be a model of LA and let $q\in\mathbf{M}$ be non-standard. Let $X +_2 Y$ and $X \cdot_2 Y$ be the polynomial functions formalizing second-order addition and multiplication (as explained in Section 1). Let $X +_2^q Y$ and $X \cdot_2^q Y$ be their linear translations as in Lemma 7. Then for all a, b and $c<q$ such that $M(a,b,c)$ we have*

$$\operatorname{Log} a +_2^q \operatorname{Log} b \;=\; \operatorname{Log}(a+b) \quad \textit{and} \quad \operatorname{Log} a \cdot_2^q \operatorname{Log} b \;=\; \operatorname{Log} c.$$

Proof. The proof is lengthy. The idea is the following. We can work in the initial segment of $\mathbf{M}$ where $\{(\log c)^n\}_{n\in\omega}$ or, respectively, $(\log(a+b))^n\}_{n\in\omega}$ is cofinal. This initial segment exists by the previous lemma and is a model of BA by Lemma 6. For a sufficiently large n, $\operatorname{Log} a +_2^q \operatorname{Log} b$ and $\operatorname{Log} a \cdot_2^q \operatorname{Log} b$ are equivalent to $\operatorname{Log} a +_2^{q'} \operatorname{Log} b$ and $\operatorname{Log} a \cdot_2^{q'} \operatorname{Log} b$ where $q' := (\log c)^n$ (resp. $q' := (\log(a+b))^n$. By Lemma 7, $\operatorname{Log} a +_2^{q'} \operatorname{Log} b$ and $\operatorname{Log} a \cdot_2^{q'} \operatorname{Log} b$ are equivalent to $\operatorname{Log} a +_2 \operatorname{Log} b$ and $\operatorname{Log} a \cdot_2 \operatorname{Log} b$. Now the equality can be checked more comfortably in BA. ∎

2.3 A well-ordering of the sets

The interpretation of sets as large numbers suggests the following definition

$$Y <_2 X \;\overset{def}{\longleftrightarrow}\; X \neq Y \;\wedge\; |Y \triangle X|-1 \in X.$$

It is easy to prove that this relation is a discrete linear order. We shall use the abbreviation $(QY <_2 X)$ with the usual meaning. Note that $Y <_2 X$ implies $Y<|X|$, so, the quantifiers $(QY <_2 X)$ are essentially second-order linear quantifier. We are going to prove that for every bounded formula the $<_2$-least number principle is provable in LA. The $<_2$-least number principle is the schema

$$\varphi(A) \;\rightarrow\; (\exists X \leq_2 A)\varphi \wedge (\forall Y <_2 X)\neg\varphi(Y).$$

Lemma 13 *LA proves the $<_2$-least number principle for every linear formula.*

Proof. Let φ be a linear formula. Assume $\varphi(A)$ and let x be the least element of the set

$$\{y \leq |A| \;:\; (\exists X \leq_2 A)\varphi \wedge (\forall Y <_2 X)[\varphi(Y) \rightarrow |X \triangle Y|-1 < y]\}.$$

(This set is non-empty because it contains $|A|$.) If we can show that $x=0$ we are done. So, assume $x=y+1$ and let X be such that

$$(*) \qquad (\forall Y <_2 X)[\varphi(Y) \rightarrow |X \triangle Y|-1 < y+1].$$

By the minimality of $y+1$, there is a $Y <_2 X$ such that

$$(**) \qquad \varphi(Y) \wedge |X \triangle Y|-1 = y.$$

We shall contradict the minimality of $y+1$ by showing that

$$(\forall Z<_2 Y)[\varphi(Z) \rightarrow |Y \triangle Z|-1 < y].$$

Let $Z<_2 Y$ be such that $\varphi(Z)$. By transitivity, $Z<_2 X$ and, by (*), $|X \triangle Z|-1 < y+1$. The latter together with (**) implies $|Y \triangle Z|-1<y+1$. It remains to exclude $|Y \triangle Z|-1 = y$. From $Y<_2 X$ and (**) it follows that $y \notin Y$ but, by definition, $Z <_2 Y$ means $|Y \triangle Z|-1 \in Y$. So, $|Y \triangle Z|-1$ can not be y. The proof is complete. ∎

When sets are small, LA proves that $<_2$ is actually the ordering induced by those of the code of the sets. We ask the reader to prove by induction the following lemma.

Lemma 14 *(LA) If $a<b$ then* $\operatorname{Log} a <_2 \operatorname{Log} b$

3 Proof of the main theorem

Theorem 15 *Every model of LA has an end extension to a model of Σ_0^p-rec.*

Proof. Let $\mathbf{M}_L$ be a model of LA. Assume $\mathbf{M}_L$ is not closed under first-order multiplication, otherwise, by Lemma 6, $\mathbf{M}_L$ is a model of BA and the theorem is trivial. Let $\mathbf{M}_P$ be the maximal cut of $\mathbf{M}_L$ which is closed under multiplication.

$$\mathbf{M}_P := \{x, X \in \mathbf{M}_L \ : \ \mathbf{M}_L \models x^n \downarrow \wedge |X|^n \downarrow \quad \text{for all } n \in \omega\}.$$

Let $\mathbf{m}_0$ be the first-order model of $I\Delta_0+\Omega_1$ obtained from $\mathbf{M}_P$ via the first-second-order isomorphism. I.e., $\mathbf{m}_0 := \{X : X \in \mathbf{M}_P\}$. We expand it to a second-order model $\mathbf{M}_0$. The sets of $\mathbf{M}_0$ are all those bounded subsets of $\mathbf{m}_0$ which are linearly definable over $\mathbf{M}_L$. I.e., the second-order elements of $\mathbf{M}_0$ are those subsets of $\mathbf{m}_0$ of the form

$$C_{\varphi(X)} := \{X : \mathbf{M}_L \models \varphi(X)\};$$

where $\varphi(X)$ is a linear formula depending on parameters in $\mathbf{M}_L$, with exactly one free variable and such that for some $A \in \mathbf{M}_P$, $\varphi(X) \rightarrow X<_2 A$. The relation $\in_2$ is defined in the natural way and we define $C_{\varphi(X)}<_2 A$ iff for all X, $\varphi(X) \rightarrow X<_2 A$.

$\mathbf{M}_0$ is a model of Θ^p. For the first-order part of Θ^p this is clear since $\mathbf{m}_0 \models I\Delta_0$. For the rest, it is sufficient to observe that LA proves the $<_2$-least number principle (see Lemma 13) and that all sets of $\mathbf{M}_0$ are, by definition, bounded. From Lemma 17 it will follow that $\mathbf{M}_0$ is a model of Σ_0^p-*comp*. First we check that, up to isomorphism, $\mathbf{M}_0$ is an end extension of $\mathbf{M}_L$.

Lemma 16 $\mathbf{M}_L$ *is isomorphic to an initial segment of* $\mathbf{M}_0$

Proof. The embedding of $\mathbf{M}_L$ into $\mathbf{M}_0$ sends numbers to sets and sets to bounded classes in the following way

$$a \longmapsto \operatorname{Log} a,$$

$$A \longmapsto C_{\varphi_A(X)} := \{X \ : \ X{<_2}A \wedge (\exists x{\in}A)(X = \operatorname{Log} x)\}.$$

The range of this map is actually in $\mathbf{M}_0$. This is evident for the second-order part while, for the first-order part, it follows from lemma 11. Let us check that the image of $\mathbf{m}_L$ is an initial segment of $\mathbf{m}_0$. If $A{<_2}\operatorname{Log} b$, in particular, $A{<}\log b$ (recall that, by definition, $\operatorname{Log} x{<}\log x$), so, by Lemma 10, there is an $a{<}b$ such that $A = \operatorname{Log} a$. Therefore, A is the image of some $a{<}b$ under the embedding defined above. Now, consider a set of $\mathbf{M}_0$, i.e., a bounded linear class $C_{\psi(X)}$. And suppose that for some $B{\in}\mathbf{M}_0$, $C_{\psi(X)}{<_2}B$, i.e., for all X, $\psi(X) \rightarrow X{<_2}B$. We have just proved that for some $b{\in}\mathbf{M}_L$, $B = \operatorname{Log} b$. Consider the set

$$A := \{x{<}b \ : \ \psi(\operatorname{Log} x) \rightarrow \operatorname{Log} x{<_2}\operatorname{Log} b\}\cdot$$

By Lemma 10, for all X, $\varphi_A(X) \longleftrightarrow \psi(X)$. So, A is mapped to $C_{\psi(X)}$. ∎

From now on let us switch to the usual notations with capital/lower-case letters also for elements of $\mathbf{M}_0$. We define,

$$\mathbf{M}_R := \{x, X{\in}\mathbf{M}_0 \ : \ \mathbf{M}_0 \models x, X{<}p^n \text{ for some } n{\in}\omega,\ p{\in}\mathbf{M}_L\}.$$

We are going to prove that $\mathbf{M}_R$ is a model of Σ_0^p-*rec*. Note that for any $p{\in}\mathbf{M}_L \setminus \mathbf{M}_P$

$$\mathbf{M}_R := \{x, X{\in}\mathbf{M}_0 \ : \ \mathbf{M}_0 \models x, X{<}p^n \text{ for some } n{\in}\omega\}.$$

$$\mathbf{M}_P := \{x, X{\in}\mathbf{M}_L \ : \ \mathbf{M}_L \models x, X{<}p^{\frac{1}{n}} \ \text{ for all } n{\in}\omega\}.$$

The situation is depicted in the figure.

We need a couple of lemmas. A polynomial formula with parameters in $\mathbf{M}_0$ is called quasi-linear if all its second-order quantifiers are bounded by elements of $\mathbf{M}_L$.

Lemma 17 $\mathbf{M}_0$ *satisfies comprehension for all quasi-linear formulas. Also, the* $<_2$*-least number principle holds in* $\mathbf{M}_0$ *for every quasi-linear formula* $\varphi(X)$ *which holds for some* X *in* $\mathbf{M}_L$.

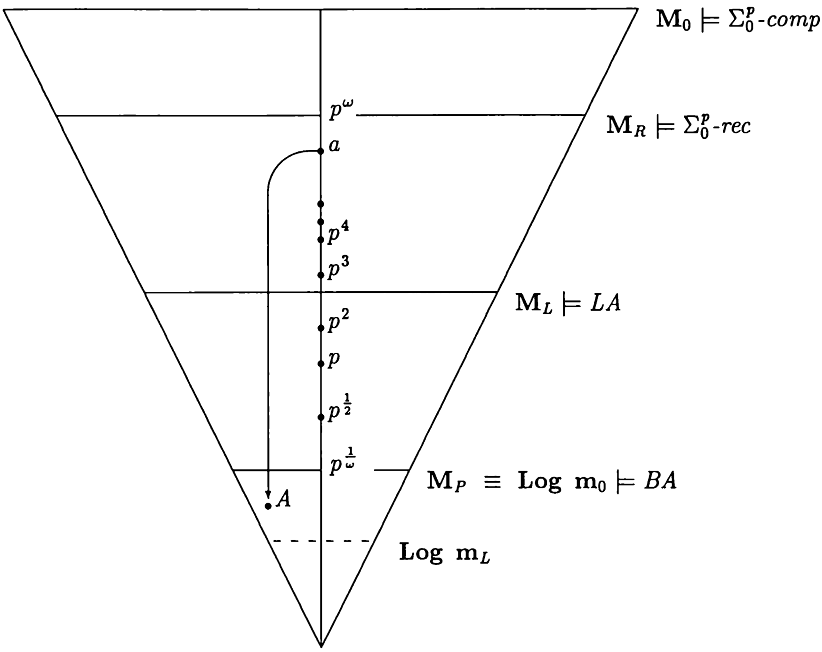
Proof. Consider a formula $\varphi(x, Y)$ as above. We assume without loss of generality that all the second-order quantifiers are bounded by some $c{\in}\mathbf{M}_L \setminus \mathbf{M}_P$. We are going to write a linear formula $\psi(X, Y)$ such that for all $x{\in}\mathbf{M}_0$ and all $Y{<}c$,

$$(*) \qquad \mathbf{M}_0 \models \varphi(x, Y) \ \Leftrightarrow \ \mathbf{M}_L \models \psi(\operatorname{Log} x, Y).$$

By the definition of the sets of $\mathbf{M}_0$, this is sufficient to have comprehension. We can assume without loss of generality that $\varphi(x)$ does not contain nested terms. Also we rename the variables of φ so that the variables v_i and V_i do not both occur in φ. We shall consider the linear translation (as in Lemma 7 with c for q)

$$(X \cdot_2^c Y = Z),\ (X +_2^c Y = Z) \ \text{ and } \ X{<_2^c}Y.$$

such that for all X and Y in M_P these are equivalent to the polynomial formulas



$$(X \cdot_2 Y = Z) \text{ and } (X +_2 Y = Z).$$

This is clearly possible since $c \in \mathbf{M}_L \setminus \mathbf{M}_P$. Note that we can also assume that all second-order quantifiers of $(X \cdot_2^c Y = Z)$ and $(X +_2^c Y = Z)$ are bounded by c.

We proceed by induction on the complexity of the formula φ. If φ is an atomic formula of the form $t \in S$ or $S < t$ (where t and S are either variable or constants) then let ψ be

$$(\exists y \in S)(T = \operatorname{Log} y) \text{ and } (\exists y < c)(S < y \wedge T = \log y).$$

where T is either the capital variable corresponding to t or, if t is a constant, $T = \operatorname{Log} t$. If φ is an atomic formula of the form $(t \cdot s = r)$, $(t + s = r)$ and $t < s$, replace it with, respectively,

$$(X \cdot_2^c Y = Z),\ (X +_2^c Y = Z) \text{ and } X <_2 Y.$$

where, again, T, S and R are either the capital variable corresponding to t, s, and r or the Logarithms of the corresponding constant. It is clear that (*) hold for atomic formulas. The definition for boolean connectives is the natural one. If ψ is the translation of φ than $(Qx < t)\varphi$ is translated with $(QX <_2 T)\psi$. $(\exists x \in T)\varphi$ and $(\forall x \in T)\varphi$ translated with, respectively, $(\exists X < c)[\varphi \wedge (x \in T)']$ and $(\forall X < c)[(x \in T)' \rightarrow \varphi]$ where $(x \in T)'$ is the translation of $(x \in T)$ given above. The reader may check that at each inductive step our translation satisfies (*). ∎

Quasi-linear functions are those defined as

$$F(x_1, .\, , x_n, X_1, .\, , X_n) := \{y < |x_1, .\, , x_n, X_1, .\, , X_n| \ : \ \varphi(y, x_1, .\, , x_n, X_1, .\, , X_n)\},$$

where a is in $\mathbf{M}_L$ and φ is quasi-linear. As usual (see e.g., [6]), when ψ is a quasi-linear formula and F is a quasi-linear function, the formula

$$\psi(y_1,.,y_n,Y_1,.,Y_n,F(x_1,.,x_n,X_1,.,X_n))$$

is considered as the abbreviation of the quasi-linear formula (in the language $L_2(+,\cdot)$) obtained by unfolding the definition of F inside ψ. So, the composition of quasi-linear functions is again a quasi-linear function. We prove below that in $\mathbf{M}_R$ quasi-linear functions are closed under iterations, i.e. under recursion over a first-order variable, provided that their values are not too large. Precisely we prove the following.

Lemma 18 *For every quasi-linear formula φ and every $a\in\mathbf{M}_L$ such that for some standard rational $\epsilon>0$, $a^{1+\epsilon}\downarrow\in\mathbf{M}_L$ the following formula holds true in $\mathbf{M}_R$*

$$(\forall X<a)(\exists Y<a)\varphi(X,Y)\rightarrow(\forall X<a)\forall b\,\exists Z\,[Z[0]=X\wedge(\forall w<b)\varphi(Z[w],Z[w+1])],$$

where we abbreviated $\big(Z\cap[w\cdot a,(w+1)\cdot a)\big)\setminus[wa)$ by $Z[w]$.

Proof. Fix an $a\in\mathbf{M}_L$ such that $a^{1+\epsilon}\downarrow$ and let q be an arbitrary element of $\mathbf{M}_L\setminus\mathbf{M}_P$. If a belongs to $\mathbf{M}_L\setminus\mathbf{M}_P$, let p be a^ϵ, otherwise, let p be the largest element of the set $\{x:\ a\cdot x\downarrow<q\}$. Clearly in both cases p is in $\mathbf{M}_L\setminus\mathbf{M}_P$ and so, $\{p^n\}_{n\in\omega}$ is cofinal in $\mathbf{M}_R$. Also, in both cases $a\cdot p\downarrow\in\mathbf{M}_L$. We shall prove, by induction on n standard, that for all quasi-linear φ,

$$(\forall X<a)(\exists Y<a)\varphi(X,Y)\rightarrow(\forall X<a)\exists Z\,[Z[0]=X\wedge(\forall w<p^n)\varphi(Z[w],Z[w+1])],$$

By the cofinality of $\{p^n\}_{n\in\omega}$ in $\mathbf{M}_R$, this is sufficient. For $n=0$ there is nothing to prove. The case $n=1$ has to be proved separately; it is a direct application of induction on $r<p$ to the quasi-linear formula,

$$(*)\ (\forall X<a)(\exists Y<a)\varphi(X,Y)\rightarrow(\forall X<a)(\exists Z<p\cdot a)\,[Z[0]=X\wedge(\forall w<r)\varphi(Z[w],Z[w+1])],$$

Now we prove by induction on $n>0$ that for every quasi-linear formula φ there is a quasi-linear function $F_n(w,X)$ such that

$$(**)\ (\forall X<a)(\exists Y<a)\varphi(X,Y)\rightarrow(\forall X<a)[F_n(0,X)=X\wedge(\forall w<p^n)\varphi(F_n(w,X),F_n(w+1,X))],$$

For $n=1$ it is true. In fact, the set Z satisfying (*) is definible by a quasi-linear formula (the formula asserting that Z is the $<_2$-least set Z satisfying (*)) and the function $F_1(w,X)$ is trivially definible over this Z.

Now, assume there is a function F_n as in (**). We are going to show that there exists a quasi-linear function $F_{n+1}(w,X)$ such that (**) holds with $n+1$ for n. We apply the induction hypothesis for $n=1$ to the formula $F_n(p^n,X)=Y$. Since clearly

$$(\forall X<a)(\exists Y<a)(F_n(p^n,X)=Y),$$

we conclude that for some F_1'

$$(\forall X<a)[F_1'(0,X)=X\wedge(\forall w<p)[F_n(0,F_1'(w,X))=F_1'(w+1,X)]].$$

Hence we define $F_{n+1}(w,X)$

$$F_{n+1}(w,X) := F_n\left(w - \left[\frac{w}{p^n}\right], F_1'(\left[\frac{w}{p^n}\right], X)\right)$$

(this means: take $\left[\frac{w}{p^n}\right]$ large steps with F_1' and do the fine tuning with F_n). We should check that this definition will do. This is a straightforward induction and is left to the reader. ∎

Finally we show that from this last lemma follows that $\mathbf{M}_R$ is a model of Σ_0^p-*rec*. Fix a quasi-linear formula $\varphi(x, y)$ (so, in particular, a Σ_0^p-formula), fix $a, b \in \mathbf{M}_R$ and $x<a$. Assume that in $\mathbf{M}_R$ holds $(\forall x<a)(\exists y<a)\varphi(x,y)$. Define the formula

$$\varphi'(X, Y) \overset{def}{\longleftrightarrow} (\exists x, y<q)\Big[(X = \operatorname{Log} x) \wedge (Y = \operatorname{Log} y) \wedge \varphi(x,y)\Big],$$

where q is an arbitrary element of $\mathbf{M}_L \setminus \mathbf{M}_R$. Observe that $(\forall X<\log a)(\exists Y<\log a)\varphi'(X, Y)$. We can apply the lemma above because, by Lemma 11, $(\log a)^2 \downarrow \in \mathbf{M}_L$. Lemma 18 yields a set Z such that

$$Z[0] = \operatorname{Log} x \;\wedge\; (\forall w<\log b)\; \varphi'(Z[w], Z[w+1]).$$

We can go back from Logarithms to numbers and obtain a set Z' such that

$$Z'(0) = x \wedge (\forall w<b)\varphi(Z'(w), Z'(w)).$$

This completes the proof of the theorem. ∎

4 Appendix

The first-order theory corresponding to Σ_0^p-*rec* is an extension of Buss' S_2^0. The theory S_2^0 is axiomatized by a set of 32 proper axioms called *BASIC* plus the schema Σ_0^b-*PIND*. This is the schema

$$\varphi(0) \wedge \forall X\ \left[\varphi(\lfloor \tfrac{1}{2} X \rfloor) \to \varphi(X)\right] \to \forall X\ \varphi(X),$$

where φ is a Σ_0^b-formula. The language of S_2^0 is an extension of that of $I\Delta_0+\Omega_1$; the definition of (the translations) of the new primitives $|X|_2$, $X \#_2 Y$ and $\lfloor \frac{1}{2}X \rfloor_2$ is straightforward (e.g., $\lfloor \frac{1}{2}X \rfloor_2$ is $X-1$). Addition $+_2$ and multiplication $\cdot_2$ require more effort.

The following informal discussion should convince the reader that there is a more or less direct way to define $+_2$ and $\cdot_2$ in models of Σ_0^p-*rec*. The first-order part $\mathbf{m}$ of a model $\mathbf{M}$ of Σ_0^p-*rec* satisfies $I\Delta_0$ (this because of Σ_0^p-*comp*). So, we have a Δ_0-definition of (first-order) exponentiation and to every x one can associate a string of length $\log x$ (see [3] Chapter 5 section 3). This makes it possible to formalize computation of a Turing machine whose space resources are bounded by the logarithm of the length of the input. Let us associate each set I with a binary string of length $|I|$ (the least upper bound of I). Our deterministic Turing machine reads the input I in a read-only tape and writes the output O in a write-only tape. The working space is bounded by the logarithm of $|I|$ times some fixed constant n (that we think as a standard number). So, internal states of the Turing machine can be coded by first-order elements $x<a := |I|^n$. Only the working tape is coded in the internal state. We can formalize by a Σ_0^p-formula with I as parameter the next-state relation among states which reads form I. Let $\varphi(x,y)$ be this formula. If we assume that the Turing machine never halts but possibly loops in a state labeled as halting state, the antecedent of the schema above is satisfied by the next-state formula φ. The axiom above claims the existence of (the code of) a computation Z for this Turing machine. The output of the computation can be Σ_0^p-defined reading the write instructions of the interval states coded in the computation Z.

In this way one formalizes logspace computable functions. Natural algorithms for addition and multiplication are in this class (up to some well-known trick: see [3] chapter 5 Section 2) and, when these are formalized, Σ_0^p-*rec* proves their recursive equations.

To check that the model constructed via the first-second-order isomorphism satisfies Σ_0^b-*PIND* is a delicate matter. We would like to proceed in the following way: assume for a contradiction that for some A

$$\varphi(0) \wedge \forall X \left[\varphi(\lfloor \tfrac{1}{2} X \rfloor) \to \varphi(X)\right] \wedge \neg\varphi(A)$$

and show that the minimal x such that $\neg\varphi(A-x)$ cannot exist. Unfortunately Σ_0^b formulas are translated via the isomorphism into $\Sigma_0^p(\Sigma_1^p)$-formulas. So, at first sight it seems that one would need Σ_1^p-*comp* to prove Σ_0^b-*PIND*. We can get around this problem. We observe that $+_1$ and $\cdot_2$, as well as all other primitives of S_2^0, have both a Σ_1^p and a Π_1^p definition. So, Σ_0^b formulas are translated by the isomorphism into formulas which are both Σ_{1*}^p and Π_{1*}^p. (Recall that Σ_{1*}^p is the smallest class of formulas containing Π_0 and closed under second-order polynomial existential quantification, conjunction, disjunction and first-order polynomial quantification. Π_{1*}^p is the dual class.) The following two lemmas show that there is a $\forall\exists\Sigma_{1*}^p$ conservative extension of Σ_0^p-*rec* where $\Sigma_{1*}^p = \Sigma_1^p$ and $\Pi_{1*}^p = \Pi_1^p$ and proving comprehnsion for formulas which are both Σ_1^p and Π_1^p. Recall from [6] that the theory Σ_1^p-*choice* is obtained by adding to Σ_0^p-*comp* the following axioms for φ varying in Σ_1^p,

$$(\forall x{<}a)(\exists Y{<}b)\varphi(x,Y) \to \exists Z\, (\forall x{<}a)\varphi(x, Z^{[x]}).$$

Lemma 19 *Σ_0^p-rec+Σ_1^p-choice is a $\forall\exists\Sigma_{1*}^p$ conservative extension of Σ_0^p-rec.*

Proof. From Corollary 2.3 of [6] follows that every model of Σ_0^p-*comp* has an $\forall\exists\Sigma_1^p$-elementary extension to a model of Σ_1^p-*choice*. So, if we start with a model of Σ_0^p-*rec*, the extension is also a model of Σ_0^p-*rec*. From this we can conclude that Σ_0^p-*rec*+Σ_1^p-*choice* is a $\forall\exists\Sigma_1^p$ conservative extension of Σ_0^p-*rec*. The lemma follows from the following claim. For every Σ_{1*}^p formula φ there is a Σ_1^p formula ψ such that

$$\Sigma_0^p\text{-}comp \vdash \psi \to \varphi,$$

moreover, ψ and φ are equivalent over Σ_1^p-*choice*. This is proved by induction on the syntax of φ. ∎

Lemma 20 *For every Σ_1^p-formula φ and every Π_1^p-formula ψ, Σ_1^p-choice proves*

$$\forall x[\varphi(x) \longleftrightarrow \psi(x)] \to (\exists X{<}a)(\forall x{<}a)x{\in}X \longleftrightarrow \varphi(x).$$

Proof. Reason in a model of Σ_1^p-*choice*. Let $\varphi \in \Pi_1^p$ and $\psi \in \Sigma_1^p$ and suppose that for some parameters a and b

$$(*) \qquad (\forall x{<}a)\cdot\ (\exists X{<}b)\varphi(x,X) \longleftrightarrow (\forall X{<}b)\psi(x,X).$$

It suffices to prove the existence of the set $\{x{<}a : (\exists X{<}b)\varphi(x,X)\}$. From (*) we have, $(\exists X{<}b)[\varphi(x,X) \vee \neg\psi(x,X)]$ for all $x{<}a$. Since the formula between square brackets is Σ_1^p, we may apply the axiom of choice to get a set $Z \subseteq [a) \times [b)$ such that

$$(\forall x{<}a)[\varphi(x, Z^{[x]}) \vee \neg\psi(x, Z^{[x]})].$$

It follows immediately that $\varphi(x, Z^{[x]})$ is equivalent to $\psi(x, Z^{[x]})$ and hence to $(\exists X{<}b)\varphi(x,X)$. So, Σ_0^p-*comp* suffices to guarantee the existence of the set $\{x{<}a : (\exists X{<}b)\varphi(x,X)\}$. ∎

References

[1] S. R. Buss, *Bounded arithmetic.* Bibliopolis, Naples, 1986.

[2] S. R. Buss, to appear.

[3] P. Hájek, P. Pudlák, *Metamathematics of First-order Arithmetic.* Springer-Verlag, Berlin, 1993.

[4] A. Razborov, An equivalence between second order bounded domain bounded arithmetic and first order bounded arithmetic. In *Proof Theory and Computational Complexity.* In: P. Clote and J. Krajícek ed., *Arithmetic, Proofs Theory and Computational Complexity.* Oxford University press, Oxford, 1993

[5] A. Wilkie and J.B. Paris, On the scheme of induction for bounded arithmetic formulas. *Annals of Pure and Applied Logic*, vol. 35 (1987) pp. 261-302.

[6] D. Zambella, Notes on polynomially bounded arithmetic. *ILLC prepublication series*, Univesiteit van Amsterdam (1994).

Part II. Provability logic

Introduction

1 Interpretability logics

Part II of this thesis contributes to the genre known as provability logic. We concentrate on two different problems. Chapter 3 is devoted to the completeness theorems for interpretability logic. Specifically, to Albert Visser's theorem that *ILP* is the interpretability logic of finitely axiomatizable theories which prove cut-elimination and to Alessandro Berarducci and Volodya Shavrukov's theorem that *ILM* is the interpretability logic of Peano Arithmetic (actually, of all full reflexive theories). The proofs given originally were not based on the natural semantics of interpretability logic (i.e. Veltman models). We give more direct completeness proofs of *ILP* and *ILM* based on Veltman models. We also provide a general set up for arithmetical completeness proofs of interpretability logic which, we think, is more in the style of Solovay's arithmetical completeness proof of provability logic. Below, we refresh the reader's memory going briefly through the few prerequisites necessary for a smooth understanding of Chapter 3. We shall omit proofs or give only quick sketches. For a more comprehensive introduction we refer the reader to some introductory text such as [8] and to the introduction of [1] and [11].

Fix a first-order theory T which is axiomatized by a recursively enumerable set of axioms and which allows a reasonable formalization of the sentence 'there is a proof of φ from T'. Let $Prov_T(\ulcorner\varphi\urcorner)$ be such a formalization. We shall mainly consider theories expressed in the language of arithmetic (other possible alternatives are e.g., the languages of ZF, of GB or of second-order arithmetic). An interpretation is map $*$ from sentences of the (propositional) modal language $\mathcal{L}(\Box)$ to sentences of the language of T which commute with the Boolean connectives and which transforms $\Box A$ into $(\Box A)^* := Prov_T(\ulcorner A^*\urcorner)$. A modal formula A is called a principle of the provability logic of T if, for every interpretation $*$, the formula A^* is provable in T. Remarkable examples of principles of provability logic are

1 $\Box(A \to B) \to (\Box A \to \Box B)$

2 $\Box A \to \Box\Box A$

3 $\Box(\Box A \to A) \to \Box A$

The first principle is the formalization of modus ponens. Traditionally, the second principle is derived as a particular case of the following theorem[1]

[1]This principle is known as formalized Σ_1 completeness. We shall only consider theories for which the conclusion of Theorem 1 holds. Principle (2) is also derivable in a more direct manner and thus, it is true also in many weak theories for which it is not known whether Theorem 1 holds or not. Provability logic of these weak theories is not completely understood yet. See [10] and [2] for the best results on this topic.

Theorem 1 *Let T be a theory proving the totality of the function exponentiation. Then for every Σ_1 formula $\varphi(x)$*

$$T \vdash \forall x(\varphi(x) \rightarrow Prov(\ulcorner\varphi(x)\urcorner))$$

where $\ulcorner\varphi(x)\urcorner$ is the formalization in the language of arithmetic of the function that given an x produces the Gödel number of the sentence $\varphi(S^x(0))$. ∎

Principle (2) follows from this theorem. In fact, we always assume that a 'reasonable' formalization of the notion of provability has of complexity Σ_1[2]. The third principle is the formalization of Löbs' generalization of Gödel's incompleteness theorem.

It should be clear that other principles of provability logic can be derived just using modal logic derivations. In fact, what can be derived from (1), (2) and (3) (viewed as axiom schemas) by means of the rule of modus ponens and of necessitation is again a principle of provability logic. The converse is also true: *all* other principles of provability logic can be derived from (1), (2) and (3) (this modal logic is named L or, sometimes, G). This remarkable fact is the starting point of modern provability logic. It is the content of Solovay's arithmetical completeness theorem [9]. In his famous article Solovay proved also a modal completeness theorem for the logic L. That is: if a modal formula is not a consequence of (1), (2) and (3), then there is a finite transitive Kripke model in which the formula does not hold (the converse is easily seen to be true). The finiteness of the counter models is noteworthy. In fact, the finite model property offers us a decision procedure to establish whether a principle of provability logic is valid or not. Also, once we have a Kripke model falsifying a principle of provability logic, the fixed point construction of [9] provides a standard procedure to obtain an actual counter example.

Solovay proved his famous theorem with Peano Arithmetic for T. De Jongh, Montagna and Jumelet in [5] observed that his theorem actually holds for all fragments of *PA* for which the principle of Theorem 1 is valid. This amazing stability is generally understood as a somewhat disappointing fact. It seems to exclude the possibility of classifying theories by means of their formalized metamathematics.

A way out of this impasse is offered by the introduction of new modal operators to formalize other metamathematical concepts. We shall consider the operator of interpretability and see that two different modal logics correspond to theories satisfying the full reflection principle (see Theorem 4) and theories which are finitely axiomatized.

Let us, for simplicity, consider theories in the language arithmetic with, as primitives: 0, 1, +, $\cdot$ and $<$. Fix two theories T and S. We say that a theory T *interprets* the theory S if there are formulas which define - within T - the following objects:

1. a set D,
2. two elements $0'$ and $1'$ in D,
3. two binary functions $+'$ and $\cdot'$,
4. a binary relation $<'$,

[2]The adjective 'reasonable' should be understood here as 'traditional'. In fact, formalizations of higher complexity are possible. Feferman's predicate of provability is a remarkable example. The provability logic of Feferman's predicate of provability is quite different from the traditional (Gödel) one. However, it may be investigated in the same modal framework. See [7] for the best known results on this field

(speaking intuitively, together these objects constitute a model of S). Moreover, we ask that T proves all axioms of S when quantifiers are restricted to D and the functions $+$ and $\cdot$ and the relation $<$ are replaced respectively by $+'$, $\cdot'$ and $<'$.

The sentence '$T+\varphi$ interprets $T+\psi$' is formalized in a natural way in the language of T. Let $Int_T(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ be such a formalization. For sufficiently strong theories, the notion of interpretability is actually a generalization of the notion of provability. In fact, we have the following.

Theorem 2 *Let T be a theory containing $I\Sigma_1$ and let φ be any sentence. Then T proves*

$Prov_T(\ulcorner\varphi\urcorner) \longleftrightarrow Int_T(\ulcorner\neg\varphi\urcorner, \ulcorner 0 \neq 0\urcorner)$.

Principles of interpretability logic are now formulas of the modal language $\mathcal{L}(\Box, \rhd)$ where $\rhd$ is a binary modal operator. The concept of arithmetical interpretation of modal formulas is exactly the same as before. The interpretation of $(A \rhd B)$, $(A \rhd B)^*$ is now $Int_T(\ulcorner A^*\urcorner, \ulcorner B^*\urcorner)$. Again, we call a modal formula A a principle of the interpretability logic of T iff for every interpretation $*$, T proves A^*. Examples of principles of interpretability logic are

4 $\Box(A \to B) \to (A \rhd B)$

5 $(A \rhd B) \wedge (B \rhd C) \to (A \rhd C)$

6 $(A \rhd B) \wedge (C \rhd B) \to (A \vee C \rhd B)$

7 $(A \rhd B) \to (\Diamond A \to \Diamond B)$

8 $\Diamond A \rhd A$

These principles hold for every theory T (see [11]). Particularly remarkable principles are: (7) - which formalizes the fact that relative interpretability implies relative consistency - and (8) - which is the formalized version of Gödel's completeness theorem for first-order logic. I.e., it is the formalization of the following theorem. Let $Cons_T(\ulcorner\varphi\urcorner)$ stand for $\neg Prov_T(\ulcorner\neg\varphi\urcorner)$ and $Cons(\ulcorner T\urcorner)$ for $Cons_T(\ulcorner 0 = 0\urcorner)$.

Theorem 3 (Arithmetized completeness theorem) *Let T be a theory containing $I\Sigma_1$. If $T \vdash Cons_T(\ulcorner\varphi\urcorner)$ then T interprets $T+\varphi$.*

The modal logic axiomatized by the schemas (1) to (8) (rules are again modus ponens and necessitation) is known as *IL*.

Another principle is derivable when T is *PA*. This is known as Montagna's principle

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

We sketch the proofs of the main theorems which lead to the proof of Montagna's principle. Let us agree on some notation. For every k let PA_k be the conjunction of the first k axioms in a fixed primitive recursive enumeration of the axioms of *PA*. Recall also that for every n there is a formula $Sat_n(x, \vec{y})$ such that for every Σ_n formula φ,

$$PA \vdash \forall \vec{y} \left[\varphi(\vec{y}) \longleftrightarrow Sat_n(\ulcorner\varphi(\vec{y}))\right]$$

So, for every standard n, the formula 'φ is a true Σ_n sentence' is formalizable in the language of *PA*. With this in mind, we state the following.

Theorem 4 (Reflection principle) *For every n sentence α and every n and every k, PA proves the formalization of the following statement. For every Σ_n sentence φ if PA_k proves φ then φ is true.*

Proof. Suppose that the complexity of PA_k is Σ_n (otherwise, choose a larger n). The usual proof of the cut-elimination theorem is formalizable in PA (actually, it holds in every model of $I\Delta_0$+$SUPEXP$). So, we can assume that the derivation of φ from PA_k uses only axioms which are of complexity Σ_n. Now it is relatively easy to prove, by induction on the length of the cut-free derivations, that all provable sentences are true. ∎

The following theorem of Orey is the formalized version of the compactness theorem for first-order logic.

Theorem 5 (Orey) *Let S be a theory with a recursively enumerable set of axioms. Let T be a theory extending PA. Let S_n be the conjunction of those axioms of S that have been enumerated up to stage n. If for all n, T proves $Cons(\ulcorner\sigma_n\urcorner)$ then T interprets S.*

Proof. See Theorem 8 below.

From this theorems we obtain the characterization of interpretability over PA which is the main ingredient for the proof of the arithmetical completeness theorem.

Theorem 6 *It is provable in PA that $Int_T(\ulcorner\varphi\urcorner,\ulcorner\psi\urcorner)$ iff for every k, PA+φ proves $Cons_{PA_k}(\ulcorner\psi\urcorner)$.*

Proof. See Theorems 7 and 9 below.

The following two theorems are the model-theoretical analogues of Theorems 5 and 6. They are not expressed in the language of PA but they are almost literally formalizable in the language of second-order arithmetic. We sketch a proof of them. The reader may easily check that the argument can be carried out in ACA_0. This is the fragment of second-order arithmetic with axioms,

1. the axioms of Robinson arithmetic Q (or, alternatively, the first 9 axioms of Θ in Chapter 1 Section 1 of this thesis),
2. the least number principle as a single axiom:
$$A \neq \emptyset \rightarrow \exists y\big(y{\in}A \wedge (\forall z{<}y) z{\notin}A\big),$$
3. arithmetical comprehension, i.e., for every formula φ possibly containing first or second-order parameters different from X, but not containing any second order quantifier,
$$\exists X\, \forall x\, [x{\in}X \longleftrightarrow \varphi(x)].$$

It is easy to see that ACA_0 is a conservative extension of PA. In fact, every model of PA can be expanded to a model of ACA_0 by adding to it all (first-order) definable sets. In this model theoretical setting we shall derive Montagna's principle. In Chapter 3 we shall follow [1] and work in ACA_0.

Theorem 7 *PA+α interprets PA+β iff every model of PA+α has an end extension to a model of PA+β.*

Proofsketch. For the direction ' $\Rightarrow$ ', consider the formula δ which define inside a model $\mathbf{M}$ of PA+α a model $\mathbf{N}$ of PA+β. Let $0'$ and $1'$ be the elements of $\mathbf{M}$ which interprets the constants 0 and 1. Let $+'$ be the interpretation of addition. We can define a function from $\mathbf{M}$

to $\mathbf{N}$ inductively: 0 is mapped to $0'$ and if x is mapped to x' then $x+1$ is mapped to $x'+'1'$. It is not difficult to see that this function preserves addition and multiplication and that the range of this map is an initial segment of $\mathbf{N}$. This initial segment will be isomorphic to $\mathbf{M}$. So, the direction ' $\Rightarrow$ ' of the theorem follows. For the converse we use Orey's theorem. If $PA+\alpha$ does not interpret $PA+\beta$, then for some n, PA does not prove $Cons_{PA_n}(\ulcorner\beta\urcorner)$. So, there is a model $\mathbf{M}$ of $PA+\alpha$ where $Prov_{PA_n}(\ulcorner\neg\beta\urcorner)$ holds. This model cannot have an end extension to a model of $PA+\beta$ because, by the preservation of Σ_1 formulas under end extensions this would be a model of $Prov_{PA_n}(\ulcorner\neg\beta\urcorner)$ contradicting the reflection principle. ∎

Theorem 8 *Let* $\mathbf{M}$ *be a model of PA and let, for every* k, T_k *be a set of* Σ_k*-formulas which is definable in* $\mathbf{M}$ *(possibly non-uniformly in* k*). Let* $T := \bigcup_k T_k$*. Assume that for all* k, $\mathbf{M} \models Cons_Q(\ulcorner T_k\urcorner)$*. Then there is an end extension of* $\mathbf{M}$ *to a model of* T*.*
Proof. Let D be the set of Σ_0-formulas which are true in M. This set is definable in $\mathbf{M}$. Clearly, for every φ in D, $Prov_Q(\varphi)$. Therefore, for every k, $\mathbf{M} \models Cons_Q(\ulcorner D+T_k\urcorner)$. Expand the language with an infinite set C of constants. Working outside $\mathbf{M}$ construct a sequence of theories T'_n in the expanded language such that (writing T' for $\bigcup_k T_k$),

- $\mathbf{M} \models Cons_Q(T'_n)$
- $T+D \subseteq T'$
- T' is complete.
- for every φ in T' there is a constant c such that the formula $\exists x\varphi(x) \to \varphi(c)$ is in T'.

Let $\mathbf{N}$ be the canonical model of T' (as in the usual Henkin construction). It is clear that $\mathbf{N}$ is an end extension of $\mathbf{M}$. In fact, the formula

$$(\forall x{<}S^a 0) \bigvee_{b<a} x = S^b 0$$

is in D, so, it must hold in $\mathbf{N}$. ∎

Note that the theorem above is formalizable in ACA_0 whenever T_k is definable there. For our application, the following immediate corollary is important.

Theorem 9 *Let* $\mathbf{M}$ *be a model of PA and let* a *be an element of* $\mathbf{M}$*. The following are equivalent*

1. *for all* k, $\mathbf{M} \models Cons_{PA_k}(\ulcorner\beta(a)\urcorner)$
2. *there is an end extension of* $\mathbf{M}$ *to a model of* $PA+\beta(a)$*.*

Proof. The direction from (2) to (1) is easy. It is a corollary of the reflection principle 4 and of the fact that Π_1 formulas are preserved in initial segments. Clearly, we can assume that for sufficiently large k, $PA_k+\beta(a)$ has complexity Σ_k and $Q \subseteq PA_k$. So, from the previous theorem follows that (1) implies (2). ∎

Though not used in this thesis, it is worthwhile to observe that the following classical theorem of MacDowell and Specker follows from Theorem 8.

Theorem 10 (MacDowell and Specker 1961) *Every model of PA has an elementary end extension.*

Proof. The set T_k of true Σ_k-formulas is (non-uniformly) definable in $\mathbf{M}$. We also have that, by Theorem 4 for every k, $\mathbf{M} \models Cons_Q(T_k)$. So we may apply Theorem 8. The model $\mathbf{N}$ obtained in this way is clearly an elementary extension of $\mathbf{M}$. ∎

We can now derive Montagna's principle from Theorem 7. In fact, suppose that $Int_T(\ulcorner\alpha\urcorner, \ulcorner\beta\urcorner)$. Then, reasoning in ACA_0, every model of $PA+\alpha$ has an end extension to a model of $PA+\beta$. Let σ be an arbitrary Σ_1 formula (so, σ may be of the form $Prov(\ulcorner\varphi\urcorner)$). In particular, every model of $PA+\alpha+\sigma$ has an end extension to a model of $PA+\beta$. This end extension is also a model of σ simply because Σ_1 formulas are preserved under end extensions. So, $Int_T(\ulcorner\alpha+\sigma\urcorner, \ulcorner\beta+\sigma\urcorner)$. Montagna's principle follows by the conservativity of ACA_0 over PA.

A different principle holds for finitely axiomatized theories:

(P) $\quad A \rhd B \rightarrow \Box(A \rhd B).$

This principle follows immediately from the syntactical complexity that the formula $Int_T(x, y)$ has when T is finitely axiomatized. The formula $Int_T(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ claims the existence of an interpretation and the existence of a (single) proof in $T+\varphi$ of the conjunction of all the translated axioms of $T+\psi$. So, this principle follows from theorem 1.

Let *ILM* and *ILP* be the modal logic axiomatized by the axioms (1) to (8) plus (M) and, respectively (P). The inference rules are again modus ponens and necessitation. The two main theorems of interpretability logic are the following generalizations of Solovay's Theorem.

Theorem 11 (Berarducci-Shavrukov) *A modal formula A of the modal language $\mathcal{L}(\Box, \rhd)$ is a principle of interpretability logic of PA iff it is derivable in ILM.*

Theorem 12 (Visser) *A modal formula A of the modal language $\mathcal{L}(\Box, \rhd)$ is a principle of interpretability logic of a finitely axiomatized theory containing $I\Delta_0+SUPEXP$ iff it is derivable in ILP.*

These theorems are strengthened by the presence of a good semantics for these two modal logics. The modal completeness theorems of De Jongh and Veltman prove that a formula of the modal language $\mathcal{L}(\Box, \rhd)$ is provable in *ILM* (resp. *ILP*) iff it holds in every finite model in a certain class (see below). So, again, it is decidable whether or not a given modal formula is or is not a principle of interpretability logic. So, the combined proofs of the modal and arithmetical completeness theorems provide us with a method that can be used to produce, given a principle of interpretability logic, either a proof of it or an actual counter example.

A Veltman frame consists of a set W of possible worlds, a transitive and conversely well-founded relation R on W, a reflexive and transitive relation S_w for every world $w \in W$ such that the following properties hold for every w, v and v in W

1 if uS_wu, then $wRu \wedge wRv$,

2 if $wRuRv$, then uS_wv,

A Veltman model is a Veltman frame together with a forcing relation $\Vdash$. This is a subset of $P \times W$ where P is the set of propositional letters of the modal language. The forcing relation is then extended in the usual way to all formulas of $\mathcal{L}(\Box, \rhd)$. This extension is the usual one for the propositional connectives and for the modal operator $\Box$. For the modal operator $\rhd$ the recursive definition is as follows

$$w \Vdash A \rhd B \quad \Leftrightarrow \quad \forall v(uRv \wedge v \Vdash A \quad \Rightarrow \quad \exists w(vS_u w \wedge w \Vdash B)).$$

We state precisely the modal completeness theorems mentioned above.

Theorem 13 (D. de Jongh and F. Veltman)

1. *A modal formula of $\mathcal{L}(\Box, \rhd)$ is provable in IL iff it is true in all finite Veltman models.*
2. *A modal formula of $\mathcal{L}(\Box, \rhd)$ is provable in ILM iff it is true in all finite Veltman models which enjoy the following property*

M if $uS_w vRz$, then uRz.

3. *A modal formula of $\mathcal{L}(\Box, \rhd)$ is provable in ILP iff it is true in all finite Veltman models which enjoy the following property*

P if $xS_w yRz$ then xRy.

Veltman's semantics for interpretability logic is very natural but it seemed at first sight not easy to prove an arithmetical completeness based on it. So, the (independent) proofs of Berarducci and of Shavrukov were based on a different semantics: the so-called 'Visser's simplified models'.

Visser simplified models are Veltman models where the relations S_w are all a subset of a global relation S. Precisely, a Veltman model is a Visser simplified model if there is a binary relation S on W such that, for every $w \in W$, $S_w = \{\langle u, v\rangle \in S \ : \ wRu \wedge wRv\}$. The modal completeness theorem holds also for Visser simplifies models. The toll to be paid for having a global relation S is the failure of the finite model property. There are formulas which are not derivable in *ILM* which have no finite Visser counter-model. Visser's completeness proof is based on De Jongh and Veltman's. In fact, he constructs a bisimulation between Veltman and Visser models.

2 Π_1-conservativity logic.

In this thesis we shall not consider the logic of Π_1 conservativity. But this subject is so closely connected with interpretability logic that a few words are probably due here.

The following notion can be naturally formalized in the language of the arithmetic: 'every Π_1 sentence provable in T is provable in S'. Principles of Π_1-conservativity logic are formulas of the modal language $\mathcal{L}(\Box, \rhd)$. The binary modal operator $A \rhd B$ is now interpreted as $T{+}B^*$ is Π_1-conservative over $T{+}A^*$. For all theories proving Σ_1 induction *ILM* is the provability logic of Π_1-conservativity. This was proved by Hájek and Montagna with a proof based on Berarducci's proof of the arithmetical completeness for interpretability logic. Later the proof was simplified by Albert Visser (unpublished). For an elegant proof of this theorem the reader is referred to a forthcoming review paper of Dick de Jongh and Georgi Dzhaparidze [3]. There proof is based on (finite) Veltman models with the same Solovay function that we use in Chapter 3 to reproduce the Berarducci-Shavrukov result.

Albert Visser noted that the soundness of the principles of *ILM* holds also for weaker theories than $I\Sigma_1$ while the completeness proof seems to need Σ_1 induction. The problem seems at first sight somewhat technical but it is worth to take a closer look at it by it because it could lead to the discovery of new principles for Π_1-conservativity logic of weaker theories.

The technical problem consists in the impossibility of proving in, e.g., *PRA* (i.e., $I\Delta_0$ plus the recursive equations for all primitive recursive functions) or $I\Delta_0+exp$ that the Berarducci-Shavrukov (or the Dzhaparidze function) has a limit. It is noteworthy that the weakness of the theories in this context is of a different nature than that studied in the Part I of this thesis. In fact, the same difficulties would arise in theories which are strong in the sense of their provably recursive functions, e.g., the theory axiomatized by the set of the Π_2 consequences of *PA* (or any other sound r.e. set of Π_2 sentences).

The technical problem can be explained as follows. The relations S_w in the Veltman model are, in general, not well-founded, therefore, if no Σ_1 least number principle is available in the theory, the function could loop forever taking S-jumps (see Chapter 3 of this thesis). So, if new principles of Π_1-conservativity logic hold for weaker theories these are likely to tell us something about sentences of the form $A \rhd B \wedge B \rhd A$.

For instance, a sound principle could be inspired by the following lemma.

Lemma 14 *Let T and S be two theories axiomatized by Π_2 axioms. If T and S have the same Π_1 consequences then $T+S$ is consistent and has no more Π_1 consequences than T or S.*

Proof. Assume that T and S have the same Π_1 consequences. Model-theoretically this means that every model of T has a Σ_0-equivalent extension to a model of S and vice versa. It suffices to prove that every model of T has a Σ_0-elementary extension to a model of $T+S$. Choose an arbitrary model $\mathbf{M}_0$ of T. There is a Σ_0-chain

$$\mathbf{M}_0 \prec_{\Sigma_0} \mathbf{M}_1 \prec_{\Sigma_0} \mathbf{M}_2 \prec_{\Sigma_0} \dots$$

such that all the $\mathbf{M}_{2i}$'s are models of T and the $\mathbf{M}_{2i+1}$'s are models of S. Consider the union of the chain $\mathbf{N}$. Clearly, $\mathbf{N}$ coincides with the union of the sub-chain $\{\mathbf{M}_{2i}\}_{i\in\omega}$ and with the union of the sub-chain $\{\mathbf{M}_{2i+1}\}_{i\in\omega}$. Since Π_2 theories are preserved under Σ_0-chains, $\mathbf{N}$ is a model both of S and T. ∎

The formalization of this theorem in *PRA* would lead to new principles for the Π_1-conservativity logic of this theory (recall that $(A \rhd B)^*$ is a Π_2 sentence). Unfortunately, such a formalization seems non-trivial. In general, model-theoretical arguments are formalizable in theories which are at least strong enough to prove Σ_1 induction. To the best of our knowledge not much is known about model theory inside *PRA*. In our opinion, this subject deserves attention for its own sake.

3 Diagonalizable algebras

Recently, Volodya Shavrukov pioneered the study of subalgebras of diagonalizable algebras of theories of arithmetic. He almost completely classified them. His results hold for every theory containing $I\Sigma_1$. Experience shows that results involving only the formalized notion of provability are valid for all theories containing $I\Delta_0+exp$ (more precisely, theories which prove the formalized Σ_1 completeness principle). This seems to be a sort of "physical boundary" for the field. So, it is natural to ask whether Shavrukov's result makes us face a new physical boundary or whether we can surmount the limit using some technical improvement. Actually, the technical improvements needed are provided in Chapter 4. There we show that all the results of Shavrukov hold for theories containing $I\Delta_0+exp$.

We refer the readers to [7] for anything they might want to know about diagonalizable algebras.

References

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

[2] A. Berarducci, R. Verbrugge, On the provability logic of bounded arithmetic. *Annal of Pure and Appied Logic*, vol. 61 (1993) 75-93.

[3] G. Dzhaparidze and D. de Jongh, to appear.

[4] D. de Jongh and F. Veltman, Provability logics for relative interpretability. In *Proceedings of Heyting '88*, ed. P. Petkov, Plenum Press, New York, (1990), pp. 31-42

[5] D. de Jongh, F. Montagna and M. Jumelet, On the proof of Solovay's theorem. *Studia Logica*

[6] V. Shavrukov, Logic of relative interpretability over Peano arithmetic. Preprint No. 5, Steklov Mathematical Institute, Academy of Sciences of the USSR, Moscow, December (1988) (in Russian).

[7] V. Shavrukov, *Adventures in Diagonalizable Algebras*, Ph.D. Thesis, Universiteit van Amsterdam (1994)

[8] C. Smoryński, *Self-reference and modal logic.* Springer-Verlag, Berlin (1985)

[9] R. M. Solovay, Provability interpretations of modal logic logic. *Israel Journal of Mathematics*, vol. 25 (1976), pp. 287-304

[10] R. Verbrugge, *Efficient Metamathematics*, Ph.D. Thesis, Universiteit van Amsterdam (1993)

[11] A. Visser, Interpretability logic. In *Mathematical Logic*, ed. P. Petkov, Plenum Press, New York, (1990), pp. 175-209

Chapter 3. On the proofs of arithmetical completeness for interpretability logic.

Abstract

Visser proved that ILP is the interpretability logic of any finitely axiomatizable theory containing $I\Delta_0$+SUPEXP. Berarducci and Shavrukov that ILM is the interpretability logic of PA. All these proof are not based directly on the natural semantics of interpretability logic (i.e. Veltman models). We give simpler alternative proofs of the arithmetical completeness of ILP and ILM directly based on finite Veltman models. We will provide a general set up for arithmetical completeness proofs of interpretability logic which is in the style of Solovay's arithmetical completeness proof of provability logic.

0. Introduction.

Visser [7] introduced the binary modal logic IL (interpretability logic) and its extensions ILM (interpretability logic with Montagna's axiom) and ILP (interpretability logic with a persistent relation in its models) to describe the interpretability logic of PA and the interpretability logic of any sufficiently strong theory T which is finitely axiomatizable and Σ_1 sound. The modal completeness of IL, ILP and ILM was provided by de Jongh and Veltman [3] using so called Veltman models. These are a very natural generalization of Kripke models. Visser [8] obtained the arithmetical completeness for ILP and more recently, Berarducci [1] and Shavrukov [5] have shown ILM to be complete for arithmetical interpretation over PA . All these proofs of arithmetical completeness do not directly use the Veltman models. Using a bisimulation Visser [8] showed ILP to be modal complete with respect to his so called Friedman models and then used these to prove arithmetical completeness. Berarducci and Shavrukov also used a bisimulation due to Visser [7] showing that ILM is modal complete with respect to the so called simplified models to prove arithmetical completeness. The use of simplified models in proving arithmetical completeness for ILM adds an additional complication due to the fact that in general these cannot be taken to be finite.

Our aim is to provide simpler and more natural proofs of arithmetical completeness for ILP and ILM . For both we shall use the original Veltman models. As all proofs of arithmetical completeness known so far, ours are based on the ideas exposed in the pioneering work of Solovay [6] and made explicit in [4].

Ths paper is organized as follows: in the next section we recall the axioms of ILM and ILP and the corresponding classes of Veltman frames. We shall not give any details. We refer the reader to the literature (see e.g. [7], [3] and [1]) both for details and comments as well as for the proofs of soundness of the axioms. In section 2 we present a general technique inspired by Solovay 's work to obtain arithmetical completeness for theories containing IL, provided that we already have modal completeness w.r.t. a certain class of finite frames. The preparatory work of section 2 is used in the last two sections for the two arithmetical completeness proofs.

I would like to thank Albert Visser for correcting and simplifying some of my arguments, Dick de Jongh and Rineke Verbrugge for their continuous and patient help.

1. Interpretability logics.

The language of the logic of interpretability contains (atomic) propositional letters $p_0, p_1, \ldots$, logical connectives $\rightarrow$, $\neg$ and a binary modal operator $\cdot \rhd \cdot$. All other connectives, as $\wedge, \vee$ and $\leftrightarrow$ are defined in the usual way. We use $\perp$ for falsum and $\top$ for true. The unary modal operator $\Box \cdot$ is defined as $\cdot \rhd \perp$. The axiom of IL are:

(L0)	All tautologies of the propositional calculus.
(L1)	$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$.
(L2)	$\Box A \rightarrow \Box\Box A$.
(L3)	$\Box(\Box A \rightarrow A) \rightarrow \Box A$.
(J1)	$\Box(A \rightarrow B) \rightarrow A \rhd B$.
(J2)	$(A \rhd B \wedge B \rhd C) \rightarrow A \rhd C$.
(J3)	$A \rhd B \rightarrow (\Diamond A \rightarrow \Diamond B)$.
(J4)	$\Diamond A \rhd A$.

The deduction rules of IL are modus ponens and necessitation The following two other axioms are the characteristic axioms of ILP and ILM.

(P)	$A \rhd B \rightarrow \Box(A \rhd B)$.
(M)	$A \rhd B \rightarrow (A \wedge \Box C \rhd B \wedge \Box C)$.

A *Veltman frame* is a triple $<W,S,R>$ where W is a set called *universe* , R and S are respectively a binary and a ternary relation on W. The elements of W are called *nodes* . We shall write xRy for $<x,y> \in R$ and yS_xz for $<x,y,z> \in S$. It is further required that R is transitive and conversely well founded and that for every $x \in W$, S_x is a reflexive and transitive relation on $\{y \mid xRy\} \subseteq W$. Moreover for every $x,y,z \in W$, xRyRz implies yS_xz.

A *Veltman model* is a Veltman frame together with a *forcing* relation $\Vdash$ between elements of W and the formulas of IL commuting with the logical connectives and satisfying the following:

$$x \Vdash \Box A \text{ iff } \forall y\, (xRy \Rightarrow y \Vdash A),$$
$$x \Vdash A \rhd B \text{ iff } \forall y\, [(xRy \;\&\; y \Vdash A) \Rightarrow (\exists z\, yS_xz \;\&\; z \Vdash B)].$$

As usual we shall improperly use the same letter W both for the model, the frame and the underlying universe. If W is a frame we write $W \models A$ iff for all forcing relations on W and all nodes of W, $x \Vdash A$.

We shall consider two other possible properties of Veltman frames:

P: If xS_wy then xS_zy for every z such that wRzRx.
M: If xS_wyRz then xRz.

We call W a *P-Veltman model* (resp. *M-Veltman model*) if the underlying frame satisfies **P** (resp. **M**).

The modal completeness of IL, ILP and ILM has been proved by de Jongh and Veltman. In particular, they proved the following three theorems:

(1) IL$\vdash$A iff for every finite Veltman frame W, W$\models$A.
(2) ILP$\vdash$A iff for every finite P-Veltman frame W, W$\models$A.
(3) ILM$\vdash$A iff for every finite M-Veltman frame W, W$\models$A.

2. A Solovay style strategy.

We want to find a general strategy for proving the arithmetical completeness of the interpretability logic for various arithmetical theories. Let T be a theory in the language of the arithmetic which is Σ_1 sound and Σ_1 complete and enough strong to formalize syntax. Given two arithmetical sentences α and β we shall write $\alpha \rhd \beta$ to mean the arithmetical formalization of the statement: *"T+α interprets T+β"* . It will be always clear from the context to which theory T we refer. We will use Latin letters for modal formulas and Greek letters for arithmetical formulas so that no confusion will arise from the fact that we are using the same symbols $\rhd$ and $\Box$ both for the modal and for the arithmetical operators.

An *interpretation* is a mapping ι from modal formulas to sentences of the language of the arithmetic such that:

(1) $\iota(A \rightarrow B) = \iota(A) \rightarrow \iota(B)$
(2) $\iota(\neg A) = \neg\iota(A)$
(3) $\iota(A \rhd B) = \iota(A) \rhd \iota(B)$

Let us write IL(T) for the set of modal formulas which are provable in T for every interpretation ι, i.e. IL(T)=$\{A \mid \forall\iota\ T\vdash\iota(A)\}$. Let ILX be a modal theory in the language of IL containing IL. We say that ILX is *arithmetically sound* for T if for every modal formula A if ILX$\vdash$A, then for every interpretation ι, T$\vdash\iota(A)$, i.e. if IL(T)$\supseteq$ILX. We say that ILX is *arithmetically complete* for T if the reverse inclusion also holds, i.e. whenever A is not a theorem of ILX then there is an interpretation ι such that $\iota(A)$ is not provable in T.

Claim. Let us suppose there is a class of finite Veltman frames X with respect to which we have modal completeness for the theory ILX. Let us suppose also that IL(T)$\supseteq$IL. If for any frame W$\in$X, there is a set $\{\lambda_x \mid x\in W\}$ of arithmetical sentences such that if (o)-(iv) below are satisfied, then IL(T)$\subseteq$ILX.

(o) for every $x,y \in W$ if $x \neq y$ then $T \vdash \neg(\lambda_x \wedge \lambda_y)$
(i) for every $x \in W$, $T+\lambda_x$ is consistent.
(ii) for every $x \in W$, $T \vdash \lambda_x \rightarrow \Box \bigvee_{xRy} \lambda_y$.
(iii) for every $x,y,z \in W$ such that yS_xz, $T \vdash \lambda_x \rightarrow \lambda_y \rhd \lambda_z$
(iv) for every $x,y \in W$ such that xRy, $T \vdash \lambda_x \rightarrow \neg(\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z)$

Proof of the claim. We assume $ILX \nvdash C$ and define an interpretation ι such that $T \nvdash \iota(C)$. By the modal completeness there is a finite model W with frame in X such that $W \nVdash C$. Let $\{\lambda_x \mid x \in W\}$ be a set of arithmetical sentences satisfying conditions (o)–(iv). Let ι the interpretation which maps the atomic proposition p occurring in C to $\iota(p) := \bigvee \{\lambda_x \mid x \Vdash p\}$. We shall show by induction on the complexity of the modal formula A that for every $x \in W$:

(a) $x \Vdash A \Rightarrow T \vdash \lambda_x \rightarrow \iota(A)$
(b) $x \nVdash A \Rightarrow T \vdash \lambda_x \rightarrow \neg\iota(A)$.

This will suffice to prove the arithmetical completeness, because if $W \nVdash C$ then for some forcing relation on W and some $x \in W$, $x \nVdash C$, from which then by (b), $T \vdash \lambda_x \rightarrow \neg\iota(C)$. By (i), λ_x is consistent with T, as is therefore $\neg\iota(C)$. Hence $T \nvdash \iota(C)$.

It remains only to prove (a) and (b) by induction on the complexity of the formula A. By condition (o) it is clear that (a) and (b) hold for atomic sentences. The inductive step for $\rightarrow$ and $\neg$ are straightforward, so let us consider just the inductive steps for $\rhd$.

Let us prove first (a). Assume $x \Vdash A \rhd B$. Then for every y such that xRy, if $y \Vdash A$, there is a node z such that $yS_xz \Vdash B$. By the induction hypothesis we can write: for every y such that xRy, if $y \Vdash A$, there is a node z such that yS_xz and $T \vdash \lambda_z \rightarrow \iota(B)$. Using (iii) and Σ_1 completeness and the soundness of IL (i.e. making a few deductions in IL) we get $T \vdash \lambda_x \rightarrow \bigwedge_{xRy \Vdash A} (\lambda_y \rhd \iota(B))$ and finally $T \vdash \lambda_x \rightarrow (\bigvee_{xRy \Vdash A} \lambda_y \rhd \iota(B))$. On the other hand, by (ii) and using the induction hypothesis (b) we obtain $T \vdash \iota(A) \rightarrow \neg \bigvee_{y \nVdash A} \lambda_y$, from which, since we assumed $T \vdash \lambda_x \rightarrow \Box \bigvee_{xRy} \lambda_y$, we get $T \vdash \lambda_x \rightarrow \Box(\iota(A) \rightarrow \bigvee_{xRy \Vdash A} \lambda_y)$. Again by the soundness of IL, $T \vdash \lambda_x \rightarrow \iota(A) \rhd \bigvee_{xRy \Vdash A} \lambda_y$. Thus the proof of (a) follows.

We prove now (b). Assume $x \nVdash A \rhd B$. Then there is a y such that xRy and $y \Vdash A$ and for every node z such that yS_xz, $z \nVdash B$. Thus, for some y such that xRy we have: $y \Vdash A \wedge \bigwedge_{yS_xz} z \nVdash B$. By the inductive hypotheses we have $T \vdash \lambda_y \rightarrow \iota(A)$ and $T \vdash \bigvee_{yS_xz} \lambda_z \rightarrow \neg\iota(B)$. By Σ_1 completeness we have $T \vdash \Box[\lambda_y \rightarrow \iota(A)]$ and $T \vdash \Box[\iota(B) \rightarrow \neg \bigvee_{yS_xz} \lambda_z]$, from which by the soundness of IL we get $T \vdash \lambda_y \rhd \iota(A)$ and $T \vdash \iota(B) \rhd \neg \bigvee_{yS_xz} \lambda_z$. Reason in T and assume λ_x. Assume for a contradiction that $\iota(A) \rhd \iota(B)$. By the soundness of IL we would have $\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z$, so from (iv) we obtain the desired contradiction. This completes the proof of the claim.

We conclude this section by remarking that conditions (o)-(iv) are not in general necessary; we believe that with a little additional work one can obtain more general, sufficient and necessary, conditions as is done in [2] for the case of provability logic.

3. The interpretability logic of finitely axiomatizable theories.

In this section T can be any finitely axiomatizable Σ_1 sound theory extending $I\Delta_0$+SUPEXP. The main property which distinguishes interpretability over these theories is that the interpretability predicate in T is Σ_1 from which the soundness of the modal axiom P follows immediately. In T it is possible to characterize interpretability as follows. Let Δ_{EXP} be tableaux provability in $I\Delta_0$+EXP, Δ tableaux provably in T and $\nabla=\neg\Delta\neg$, i.e. the tableaux consistency in T. According to the Friedman-Visser characterization [8], α interprets β iff $\Delta_{EXP}(\nabla\alpha\rightarrow\nabla\beta)$.

We want to prove that IL(T)=ILP. We leave, as usual, the proof of soundness to the reader and we shall prove only IL(T)$\subseteq$ILP. We shall find sentences (o)-(iv) as in the previous section. The method is as in Solovay [6]. We define a function F using the fixed point theorem and let the λ_x be some limit statements concerning F.

Assume for convenience W has been given as a finite set of non zero natural numbers. We shall use the symbols x,y and z only for elements of W. Let λ_x be the sentence $\lim_n F(n)=x$ and $\lambda_0:=\forall nF(n)=0$. Together with the function F we will define also an auxiliary function G which will aid us in book keeping. The function G will always "follow" the function F, i.e. if for some n, F(n)=x then G(n)=F(m) for some m≤n. Speaking informally, G(n)≠F(n) will warn us of the fact that there is no proof of code less then n of $\neg\lambda_{F(n)}$. This has to be considered as a "dangerous signal" since we would like in the end to have $\lambda_x\rightarrow\Box\neg\lambda_x$. When such a situation occurs then only "safe" moves are allowed, i.e. F as well as G will move only to a node y for which there is a proof of $\neg\lambda_y$.

The definition of F and G is the following:

(a) F(0)=G(0)=0. If F(n)=0 and for some $x\in W$, n witnesses $\Delta\neg\lambda_x$, then F(n+1)=G(n+1)=x.
(b) If F(n)=G(n)=$x\in W$ and for some node y such that xRy, n witnesses $\Delta_{EXP}(\nabla\lambda_y\rightarrow\nabla\neg\bigvee_{yS_xz}\lambda_z)$, then F(n+1)=y and G(n+1)=G(n).
(c) If F(n)=y and G(n)=x, for some z, yS_xz and n witnesses $\Delta\neg\lambda_z$, then F(n+1)=G(n+1)=z.
(d) In all other cases F(n+1)=F(n) and G(n+1)=G(n).

Let μ_x be the sentence $\lim_n G(n)=x$. We shall eventually prove that the two functions have the same limit, i.e. $\mu_x\leftrightarrow\lambda_x$, but for proving this we need the cut elimination theorem. The formalization of the cut elimination theorem is provable in T since T contains SUPEXP but is surely not provable in EXP. To carry on with our proof we need to know what $I\Delta_0$+EXP proves about the functions F and G, hence the following:

Lemma 1. $I\Delta_0$+EXP proves the following:

.1 For every $w\in W$, $\mu_w\rightarrow\Delta\bigvee_{wRx}\lambda_x$.
.2 For every $w,x\in W$, if $x\neq w$ then $\mu_w\wedge\lambda_x\rightarrow\Delta\bigvee_{xS_wy}\lambda_y$.
.3 For every $w,y\in W$ if wRy then $\mu_w\wedge\lambda_w\rightarrow\nabla\lambda_y$.
.4 For every $x,y,w\in W$, if xS_wy then $\mu_w\wedge\lambda_x\rightarrow\nabla\lambda_y$.

Proof. Directly from the definition of F, $I\Delta_0$+EXP proves that if, for some n, G(n)=w then after stage n the function F remains either in w or in the upper cone above w. Thus the limit of F is either w or is some node above w. If G(n)=w then by provable Σ_1 completeness, Δ_{EXP}(G(n)=w) and a fortiori Δ(G(n)=w). The proof of (.1) follows by combining all this with the fact that G(n)=w implies $\Delta\neg\lambda_w$. To prove (.2) assume that for some x≠w we have $\mu_w \wedge \lambda_x$. Then for some n $\Delta_{EXP}(G(n)=w \wedge F(n)=x)$. Again, using the definition of the functions F and G, it is easy to argue that whenever $G(n)=w \wedge F(n)=x$ for some w≠x, the function F never leaves the set of nodes which are in S_w relation with x. This gives (.2). To prove (.3) assume wRy, λ_w and μ_w and let n be such that for all m>n, F(m)=G(m)=w. If $\neg\lambda_y$ where cut free provable, then some m>n would witness $\Delta\neg\lambda_y$. (Here and in the following, it is assumed that a cut free provable theorem has infinitely many cut free proofs.) So $\Delta_{EXP}(\nabla\lambda_y \rightarrow \nabla\neg\bigvee_{yS_xz}\lambda_z)$ and then at stage m+1, F would move to y, against our assumption that at stage n F has already reached is limit. To prove (.4) assume λ_x, μ_w and xS_wy. Then wRy, and therefore w≠y. Let n be such that for all m>n, F(m)=x and G(m)=w. Suppose, by contradiction, that $\Delta\neg\lambda_y$. Let m>n a witness of $\Delta\neg\lambda_y$. Then at stage m+1 both F and G move to y, by condition (c). This contradicts our assumption that at stage n G has already reached its limit. (Note that clearly y≠w since xS_wy and then wRy.)

For the following lemma we need that the formula $(\nabla\alpha \wedge \alpha \rhd \beta) \rightarrow \nabla\beta)$ is provable in T. It is easy to check that T (or even $I\Delta_0$+EXP) proves $(\Diamond\alpha \wedge \alpha \rhd \beta) \rightarrow \Diamond\beta)$, and since in T the formalization of the cut elimination theorem is provable, we can substitute tableaux consistency with normal consistency, so also the former formula is derivable in T. We can prove the following:

Lemma 2. For every $x \in W$, $T \vdash \mu_x \leftrightarrow \lambda_x$.
Proof. Reason in T and assume for a contradiction that $\lambda_x \wedge \neg\mu_x$. Then for some wRx we have μ_w. This implies $\nabla\lambda_x$, for otherwise the function G would have jump to x. Since x≠w the last move of the function F has been from w to x using condition (b) and therefore $\lambda_x \rhd \neg\bigvee_{xS_wy}\lambda_y$. By the remark above we get immediately $\neg\Delta\bigvee_{xS_wy}\lambda_y$. From lemma 1.2 we get also $\Delta\bigvee_{xS_wy}$ λ_y. Thus we have the desired contradiction.

Lemma 3. For every x,y,z∈W such that yS_xz, $T \vdash \lambda_x \rightarrow \lambda_y \rhd \lambda_z$.
Proof. Reason in T and assume λ_x. We want to show that for every y,z such that yS_xz, $\lambda_y \rhd \lambda_z$, i.e. $\Delta_{EXP}(\nabla\lambda_y \rightarrow \nabla\lambda_z)$. By lemma 2 we have μ_x and by provable Σ_1 completeness we have that for some k, Δ_{EXP}(G(k)=x). Reason in $I\Delta_0$+EXP. Assume $\nabla\lambda_y$ and let w be the limit of the function G. Since G(k)=x, the limit w is either x or is above x. By lemma 1.1, from $\nabla\lambda_y$ we know that w has to be strictly below y. Thus either x=wRy or xRwRy and, by the characteristic property of the P-Veltman frames, from yS_xz we get yS_wz. Let u be the limit of F. If u=w from wRz and lemma 1.3 the lemma follows immediately. Otherwise by lemma 1.2 and $\nabla\lambda_y$ one has uS_wy. By the transitivity of S_w we obtain uS_wz and thus finally, by lemma 1.4, $\nabla\lambda_z$.

Lemma 4. For every $x \in W$, $T \vdash \lambda_x \rightarrow \Delta\bigvee_{xRy}\lambda_y$
Proof. Immediate by lemmas 1.1 and 2.

We can now easily check that the set of sentences $\{\lambda_x \mid x\in W\}$ satisfies (o)-(iv). In fact (o) is trivial, the proof of (i) is completely standard, (ii) derives from lemma 4 and the provability in T of the cut elimination theorem. Condition (iii) is lemma 3 and (iv) is obvious by the definition of F and lemma 2. This concludes the proof of the completeness theorem.

4. The interpretability logic of PA.

In this section we want to prove that IL(PA)=ILM. The main characteristic of the interpretability in Peano arithmetic is the Orey-Hajek characterization: let $\Box_k\beta$ be the formalization of the sentence *"there is a proof of β which uses only the first k axioms of PA"* , let $\Diamond_k\equiv\neg\Box_k\neg$, then it is provable in PA that α interprets β iff $\forall k\Box(\alpha\rightarrow\Diamond_k\beta)$. Another characteristic property of PA is that it proves full reflection for any of its finite subtheories, moreover this is formalizable in PA, namely: for every α, PA$\vdash\forall k\Box(\Box_k\alpha\rightarrow\alpha)$. These facts would be sufficient to carry out the following proof, but for sake of better readability we shall, following Berarducci, work in ACA_0 rather then in PA. The second order theory ACA_0 is a conservative extension of PA; in ACA_0 we can speak of models of PA and easy theorems of basic model theory are formalizable and provable in ACA_0. In particular in ACA_0 we have the following characterization of the interpretability over PA: *"PA+α interprets PA+β iff every model of PA+α has an end extension to a model of PA+β"*. In ACA_0 the *standard model* is the set $\{x \mid x=x\}$ with the obvious choice of operations, any other *non-standard model* has an initial segment which is isomorphic to it. Numbers belonging to this initial segment are called as usual *standard numbers* . Full reflection translates in ACA_0 in the following manner: *"for every model Y of PA and every standard number k,* $Y\vDash\Box_k\alpha\rightarrow\alpha$".

As in the previous section we shall prove only that IL(PA)$\subseteq$ILM, leaving the converse to the reader. The sentences which are meant to satisfy (o)-(iv) are defined as limits of a recursive function F exactly as in the previous proof . Define, as in [1] for every $x\in W$, rank(x,n):=*"the minimal k such that there is a witness $\leq n$ of* $\Box_k\neg\lambda_x$" . If k is a number, $x,y\in W$, xRy then we define the sentence $\alpha_{x,y}(k)$ as $\forall j\geq k[F(j)=x\vee F(j)=y]$. Our definition of the function F resembles Berarducci's as far as it is concerned with the S–jumps but it differs in the R-jumps. Roughly speaking we allow the function F to make an R–jump if there is a proof that this will not be the last move. We assume for convenience that W has been coded as a finite set of non zero natural numbers, we shall use the symbols w,x,y,...etc. only for elements of W.

(a) Let F(0)=0 and if F(n)=0 and for some $x\in W$, n witnesses $\Box\neg\lambda_x$, then F(n+1)=x.
(b) If F(n)=x and for some $y\in W$ and some k<n such that $\forall j\in[k,n]$ F(j)=x and xRy, n witnesses $\Box\neg\alpha_{x,y}(\dot{k})$ ($\dot{k}$ is the numeral of k), then F(n+1)=y.
(c) If F(n)=x and for some nodes y and z, xS_zy and $\exists i\leq n[\mathrm{rank}(y,n)\leq i<\mathrm{rank}(x,n)\wedge F(i)=z]$, then F(n+1)=y. (If this condition obtains for two different nodes, choose the one with minimal code.)
(d) In all the other cases F(n+1)=F(n).

Note that any two points in the orbit of F are connected by an S-and/or R-arrow. We shall write $Y\vDash$ ...x...y if, according to the model Y the function F goes from x to y (possibly in a non-standard number of steps). We write $Y\models ...xRy...$ (resp. $Y\models ...xS_zy...$) if, in the model Y, F moves in one step from x to y and xRy (resp. xS_zy). If in a model Y the function F moves at stage n from x to y, then we say F moves with an R-step (resp. with S-step) if at stage n condition (b) (resp. condition (c)) has been applied. If, at stage n, F moves from 0 to some node x, we say that F moves with an (a)-step.

Lemma 1. In PA it is provable that the function F has a limit.
Proof. This is not obvious since the S-relations are in general not well founded. It is clear that if h is the height of the frame the function cannot make more than h consecutive R-moves. By the property **M** of the M-frame F cannot make more than h R-moves, whether they are consecutive or not. Thus eventually F is allowed only to make S-moves. If S would not have a limit we could construct a definable infinite decreasing sequence of ranks. This is provably false in PA.

We are eventually going to prove $\lambda_x \rightarrow \Box\neg\lambda_x$, but to achieve this goal we need to prove first a weaker form of it.

Lemma 2. For every $x\in W$ and for every $k\in\omega$, $PA\vdash F(k)=x \rightarrow \Box\exists j>\dot{k}\ F(j)\neq x$.
Proof. Assume F(k)=x. Reasoning in ACA_0 we claim that for every model Y of PA, $Y\models \exists j>k\ F(j)\neq x$. If F moved to x with an (a)-step or with an S-step we would have $\Box\neg\lambda_x$ and then $Y\models\neg\lambda_x$ so our claim would hold trivially. So, assume that the last move of F has been an R-step, and that say at stage h, the function F moves from z to x. Then for some i<h such that $\forall j\in[i,h]\ F(j)=z$, h codes a proof of $\neg\alpha_{z,x}(i)$. So, $Y\models\exists j\geq i\,[F(j)\neq z\wedge F(j)\neq x]$. We have assumed $\forall j\in[i,k]\ [F(j)=z \vee F(x)]$, this is a Σ_1 statement so, by provable Σ_1 completeness, it is true also in Y. Thus $Y\models\exists j>k\ F(j)\neq x$ and our claim is proved.

Lemma 3. For every $x\in W$, $PA\vdash\lambda_x\rightarrow\Box\bigvee_{xRy}\lambda_y$.
Proof. It is sufficient to prove that for every x and y, if $\neg xRy$ then $PA\vdash\lambda_x\rightarrow\Box\neg\lambda_y$. Reason in ACA_0 and assume for a contradiction that λ_x, $\Diamond\lambda_y$ and $\neg xRy$. Choose k such that F(k)=x and let Y be a model of λ_y. By provable Σ_1 completeness we have that $Y\models F(k)=x$. Now, in Y, let z be the last node that the function passes through before arriving to y. The last step must be an S-step otherwise zRy and by the **M** property of the M-Veltman frames we would have xRy. We shall picture the situation as $Y\models ...x...zS_wy$. (We recall that either z or y might be equal to x, the previous lemma guarantees only that after stage k the function has moved at least once.) We assumed $\neg xRy$ thus, since zS_wy implies wRy, we have that $w\neq x$. By the definition of F we have that at some stage n, for some $i\leq n$, $rank(y,n)\leq i<rank(z,n)$ and F(i)=w. By the reflection principle rank(y,n) has to be non-standard in Y, and since we have chosen k standard, $rank(y,n)\geq k$. Thus also $i\geq k$ and so $Y\models ...F(k)....F(i)$ and therefore $Y\models ...x...w...zS_wy$. By the **M** property of the M-Veltman frames from wRy we get xRy. Contradiction.

Lemma 4. For every $x,y,z\in W$ such that yS_xz, $PA\vdash\lambda_x\rightarrow\lambda_y\rhd\lambda_z$.
Proof. Assume λ_x and yS_xz. We shall prove in ACA_0 that, for arbitrary large k, in any model Y of PA, $\lambda_y\rightarrow\Diamond_k\lambda_z$. Let k be such that F(k)=x. Suppose for a contradiction that there exists a model $Y\models\lambda_y\wedge\Box_k\neg\lambda_z$. Then for n large enough we have $Y\models rank(z,n)\leq k<n$. Suppose n is also large enough so that (in Y) F has already reached its limit. By the reflection principle rank(y,n)

must be non-standard in Y. Then $Y \models \mathrm{rank}(z,n) \leq k < \mathrm{rank}(y,n) \wedge F(k)=x$. So, $Y \models F(n+1)=z$ which contradicts the fact that F has already reached its limit.

Lemma 5. for every $x,y \in W$ such that xRy, $PA \vdash \lambda_x \rightarrow \neg(\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z)$.

Proof. Reason in ACA_0 and assume λ_x. To prove $\neg(\lambda_y \rhd \neg \bigvee_{yS_xz} \lambda_z)$ it will suffice to find a model Y of λ_y which has no end extension to a model of $\neg \bigvee_{yS_xz} \lambda_z$. Let fix k such that $\forall j \geq k\ F(j)=x$, since xRy we have: $\Diamond \alpha_{x,y}(\dot{k})$ otherwise the function would jump from x to y contradicting λ_x. Then we can choose our model Y such that $Y \models \forall j>k[F(j)=x \vee F(j)=y]$; since we have assumed λ_x and therefore (by lemma 3) $Y \models \neg\lambda_x$, we can conclude that $Y \models \lambda_y$. Let Z be any end extension of such a model Y and let z such that $Z \models \lambda_z$. The proof is complete if we can show that yS_xz. Let n be the minimal number in Z such that such that $Z \models F(n+1)=z$. By provable Σ_1 completeness and the fact that Σ_1 formulas are conserved by end extensions, we have $Z \models \ldots xRy \ldots z$. Let w be the last node reached with an R-step i.e. for some u, $Z \models \ldots xRy \ldots uRw \ldots z$ and between w and z only S-steps occur. Then the rank of all the steps between w and z is larger than rank(z,n). By the reflection principle rank(z,n) is a non-standard number in Z. If all the steps between w and z are S_x-steps, we are done, otherwise let S_t be the last non S_x-step between w and z i.e. $Z \models \ldots xRy \ldots uRw \ldots S_tv\ S_x \ldots S_xz$. Let $i \geq \mathrm{rank}(z,n)$, be such that $F(i)=t$. Since rank(z,n) is non-standard in Z, t cannot occur in the orbit of F before x, so either t=y or $Z \models \ldots xRy \ldots t \ldots S_tvS_x \ldots S_xz$. In both cases one can conclude that yRv and hence yS_xz.

We can now easily check that the set of sentences $\{\lambda_x \mid x \in W\}$ satisfies (o)-(iv). In Fact (o) is trivial, the proof of (i) is completely standard, (ii) is lemma 3, (iii) is lemma 4 and (iv) is lemma 5. This concludes the proof of the completeness theorem.

References.

[1] A. Berarducci, The interpretability logic of Peano arithmetic. *Jornal of Symbolic Logic* **56**, 1059-1089 (1990).

[2] A. Berarducci and R. Verbrugge, On the metamathematics of weak theories. *ITLI Prepublication Series*, **ML-91-02**, University of Amsterdam (1991).

[3] D. de Jongh and F. Veltman, Provability logic for relative interpretability. In: *Mathematical logic*, 31-42, edited by P. P. Petkov, Plenum Press, New York (1990).

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

[5] V. Y. Shavrukov, The logic of relative interpretability over Peano arithmetic. (Preprint in Russian) Steklov Mathematical Institute Moscow (1988).

[6] R. Solovay, The provability interpretation of modal logic. *Israel Journal of Mathematics* **25**, 287-304, (1976).

[7] A. Visser, Preliminary notes on interpretability logic. *Logic Group Preprint Series*, **29**, University of Utrecht (1988).

[8] A. Visser, Interpretability logic. In: *Mathematical logic*, 175-209, edited by P. P. Petkov, Plenum Press, New York (1990).

Chapter 4. Shavrukov's theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0$+exp.

Abstract.

Recently Volodya Shavrukov [1] pioneered the study of subalgebras of diagonalizable algebras of theories of arithmetic. We show that his results extend to weaker theories (namely to theories containing $I\Delta_0$+exp).

0. Introduction.

A diagonalizable algebra [2,3,4,5,6] is a Boolean algebra $(\mathcal{D},\rightarrow,\bot)$ with an additional operator $\Box$ which satisfies the axioms:

$$\forall x,y \quad \Box(x\rightarrow y)\rightarrow(\Box x\rightarrow\Box y) = \top\,, \qquad \forall x \quad \Box(\Box x\rightarrow x)\rightarrow\Box x = \top, \qquad \Box\top = \top.$$

Let T be a sufficiently strong axiomatized theory in the language of arithmetic. The predicate of provability of T generates in a natural way an operator on the Lindenbaum algebra of T. The resulting diagonalizable algebra $\mathcal{D}_T$ is called the *diagonalizable algebra of T*. The subalgebras of $\mathcal{D}_T$ have been studied in [1], in particular the general problem of when a diagonalizable algebra $\mathcal{D}$ is embeddable in $\mathcal{D}_T$ has been considered there. We intend to present a modification of Shavrukov's construction that allows us to prove these results for a wider class of theories (all those containing $I\Delta_0$+exp).

We will translate questions about subalgebras into problems of provability logic. For this we need some notation. Let $\mathcal{L}$ be the set of modal formulas generated by the language $(\rightarrow,\Box,\bot,\{p_i\}_{i\in\omega})$. We write $B\vDash A$ if A can be derived using modus ponens and necessitation from the formula B and Löb's axioms (hence $\vDash A$ means that A is a theorem of Löb's logic and $B\vDash A$ means $\vDash\boxdot B\rightarrow A$, where $\boxdot B$ is $B\wedge\Box B$), we write $B\Vdash A$ iff $\vDash B\rightarrow A$. When $\mathcal{A}$ is a set of modal formulas in the language $\mathcal{L}$ we write $\mathcal{A}\vDash A$ and $\mathcal{A}\Vdash A$ if for some conjunction B of formulas in $\mathcal{A}$, $B\vDash A$, resp. $B\Vdash A$. Given a set $\mathcal{A}$, consider the equivalence relation on $\mathcal{L}$: $A\approx_{\mathcal{A}}B$ iff $\mathcal{A}\vDash A\leftrightarrow B$, and let $\mathcal{L}/\mathcal{A}$ be the sets of $\approx_{\mathcal{A}}$-equivalence classes. The operator which maps the equivalence class of A to that of $\Box A$ is a well defined operator on $\mathcal{L}/\mathcal{A}$ which turns it into a diagonalizable algebra. For every (denumerable) diagonalizable algebra $\mathcal{D}$ there is a set $\mathcal{A}$ such that $\mathcal{D}$ is isomorphic to $\mathcal{L}/\mathcal{A}$.

Let T be an axiomatized theory in the language of the arithmetic and let Thm(.) be the provability predicate of T. A *T-interpretation* is a map ι which maps formulas of $\mathcal{L}$ to sentences of the language of arithmetic such that T proves:

(i) $\iota(\Box A) \leftrightarrow \mathrm{Thm}[\ulcorner \iota(A) \urcorner]$; (ii) $\neg\,\iota(\bot)$; (iii) $\iota(A \rightarrow B) \leftrightarrow (\iota(A) \rightarrow \iota(B))$.

(In the following we shall simply say an *interpretation* since the theory T will be fixed.) If for every formula A in $\mathcal{L}$, $\mathcal{A} \models A$ iff $T \vdash \iota(A)$ we say that ι *interprets* $\mathcal{A}$ in T. We say that $\mathcal{A}$ is *interpretable* in T if there exists an interpretation which interprets $\mathcal{A}$ in T.

Given an interpretation of $\mathcal{A}$ in T one can construct in a natural way an embedding of $\mathcal{L}/\mathcal{A}$ in $\mathcal{D}_T$ and vice versa: from an embedding one can easily construct an interpretation. So, for any given theory T, the problem of classifying the subalgebras of $\mathcal{D}_T$ reduces to classifying the sets of modal formulas $\mathcal{A}$ which are interpretable in T.

We write as usual $\Box^0\bot$ for $\bot$ and $\Box^{n+1}\bot$ for $\Box\Box^n\bot$; the minimal n such that $\mathcal{A} \models \Box^n\bot$ is called the *height* of $\mathcal{A}$. If such an n does not exist, we say that $\mathcal{A}$ has *infinite height*. We say that $\mathcal{A}$ has the *strong disjunction property (s.d.p.)* or, equivalently, that $\mathcal{A}$ is *strongly disjunctive (s.d.)* iff $\mathcal{A}$ is consistent and for all formulas A and B if $\mathcal{A} \models \Box A \vee \Box B$ then $\mathcal{A} \models A$ or $\mathcal{A} \models B$. The same classification is, mutatis mutandis, applied to diagonalizable algebras. In the following T will be a fixed axiomatized theory (i.e. the theory is given along with a Kalmar elementary axiomatization of it). The language of T contains the language of arithmetic and -only for the sake of convenience- a symbol for exponentiation. Thm(.) is the provability predicate of T. We write $\mathrm{Thm}^0(\bot)$ for the sentence $0 \neq 0$ and $\mathrm{Thm}^{n+1}(\bot)$ for $\mathrm{Thm}(\mathrm{Thm}^n(\bot))$ (in the following we shall always omit the Gödel-number symbols $\ulcorner\ \urcorner$). The minimal n such that $T \vdash \mathrm{Thm}^n(\bot)$ is called the *height* of T. If such an n does not exist we say that T has *infinite height*. The height of T is in fact the height of its diagonalizable algebra $\mathcal{D}_T$. If all Σ_1-sentences provable in T are true in the standard model, then T is *Σ_1-sound*, otherwise T is *Σ_1-ill*. Shavrukov proved that every r.e. set of modal formulas is interpretable in the diagonalizable algebra of every (sufficiently strong) Σ_1-ill theory provided it has the same height as the theory. Moreover an r.e. set of modal formulas is interpretable in the diagonalizable algebra of every (sufficiently strong) Σ_1-sound theory if and only if it is s.d.. Recall that the Gödel numbering of arithmetical sentences gives a natural recursive enumeration of a set $\mathcal{A}$ such that $\mathcal{L}/\mathcal{A}$ is isomorphic to $\mathcal{D}_T$. So, an interesting consequence is that diagonalizable algebras of Σ_1-sound theories are mutually embeddable. The same holds for Σ_1-ill theories of any fixed height.

The results mentioned above have been proved in [1] for theories which contain Σ_1-induction. In fact, the construction makes use of a Solovay function which ranges over a Kripke model. In the case of infinite height theories the models used have nonstandard height so Σ_1-induction is needed to guarantee the existence of the limit. In section 2 we show in Theorem 1 and 2 that the use of Σ_1-induction is inessential and the result is valid for all theories containing $I\Delta_0$+exp. (Actually Theorems 1 and 2 only consider theories of infinite height. In fact, in the case of finite height the proof in [1] goes through for $I\Delta_0$+exp with minor modifications.)

For Σ_1-ill theories a stronger result holds. In [1] it has been proved that a diagonalizable algebra is embeddable in the diagonalizable algebra of a Σ_1-ill theory provided it has the same height as the theory. Also this theorem holds for weaker theories then those considered in [1]. We shall not give a proof of this fact since it is easily derivable from Shavrukov's as follows. To embed $\mathcal{D}$ in the diagonalizable algebra of some "weak" theory T, first apply the result of [1] to embed $\mathcal{D}$ in the diagonalizable algebra of some sufficiently "strong" theory T*. Finally embed $\mathcal{D}_{T^*}$ in $\mathcal{D}_T$. Composing the two embeddings one obtains the desired subalgebra.

I wish to thank Volodya Shavrukov for numerous suggestions and corrections. I owe very much also to the stimulating criticisms and friendly encouragements of Lev Beklemishev. Comments of Dick de Jongh and Alessandro Berarducci have helped to make this paper more readable.

1. A lemma.

In this section we prove a lemma which will be used to characterize the r.e. sets of modal formulas interpretable in a theory $T \supseteq I\Delta_0$+exp. We assume the reader to be familiar with the techniques introduced in [7].

A finite tree-like Kripke model k (in the sequel simply a *model*) is a triple $(W,R,\Vdash)$ where (W,R) is a finite tree with nodes $w \in W$ strictly ordered by the relation R and $\Vdash$ is a finite subset of $W \times \omega$. We call W the *universe of k* and (W,R) the *frame of k*. We write $w \Vdash p_i$ if $(w,i) \in \Vdash$. The relation $w \Vdash A$ (w *forces* A) is then extended to all the formulas of $\mathcal{L}$ in the usual way. We say that $k'=(W',R',\Vdash')$ is a generated submodel (in the sequel simply a *submodel*) of $k=(W,R,\Vdash)$ if the universe of k' is $W'=\{w\} \cup \{u \mid wRu\}$ for some node w of k, R' and $\Vdash'$ are the restrictions of R and $\Vdash$. We write $k \Vdash A$ (*k forces A*) iff the formula A is forced at the root of the model coded by k, we write $k \vDash A$ (k *is a model of A*) if every node of k forces A. Then we have that k is a model of A iff k forces $\Box A$. If $\mathcal{A}$ is a finite set of formulas we write $k \Vdash \mathcal{A}$ (resp. $k \vDash \mathcal{A}$) if for every $A \in \mathcal{A}$, $k \Vdash A$ (resp. $k \vDash A$). Then it easy to check that, if $\mathcal{A}$ is finite, then $\mathcal{A} \vDash A$ iff every model of $\mathcal{A}$ is a model of A, and $\mathcal{A} \Vdash A$ iff every model which forces $\mathcal{A}$ forces A (if $\mathcal{A}$ is infinite this may not be the case since we will deal only with finite models).

In a first-order formula an occurrence of a quantifier is said to be bounded if it is of the form $\forall x<t$ or $\exists x<t$ where t is a term of the language of T. The Δ_0-formulas of T are the formulas provably equivalent to formulas with only bounded quantifiers (having assumed exponentiation as a primitive function of the language we should properly write Δ_0(exp) but in the present paper there will be no risk of confusion). The Σ_1-formulas are those equivalent to a Δ_0-formula preceded by an existential quantifier. The theory whose axioms are those of Robinson arithmetic plus the characteristic axioms for exponentiation and the induction schema for Δ_0-formulas is called $I\Delta_0$+exp; the theory which contains also the schema of Σ_1-induction is called $I\Sigma_1$. We refer the reader to [8] for more details on these theories.

We fix a natural coding of modal formulas and of models in arithmetic; we shall use the same symbol both for a formula (resp. model) and its code. We require that the coding assigns to proper submodels of k a smaller code than to k itself. Having exponentiation as a primitive function, we may require without loss of generality that $k \Vdash A$ and $k \vDash A$ translate into Δ_0-formulas. We also use in the following that the completeness theorem of Löb's logic with respect to (finite) models is formalizable in $I\Delta_0$+exp. Given an r.e. set $\mathcal{A}$ of modal formulas we may find, formalizing in the language of arithmetic the algorithm enumerating $\mathcal{A}$, a Δ_0-formula "$A \in \mathcal{A}_x$" (here A and x are the free variables of the formula) such that for every $A \in \mathcal{L}$, $A \in \mathcal{A}$ iff $\exists n \in \omega$, $T \vdash A \in \mathcal{A}_n$. We also require that (provably in T) if $A \in \mathcal{A}_x$ then $A<x$ i.e., the code of A is less then x. We call such a formula a *description* of $\mathcal{A}$ (in T). We may formalize in T also the notion of Löb's derivability so that we can use the expression $\mathcal{A}_n \vDash A$ both when arguing in the real world and in the theory. Formalizing the proof of the completeness theorem for Löb's logic in

$I\Delta_0$+exp one can find a Δ_0-formula describing the relation $\mathcal{A}_{,n}\models A$. We shall also use the expression "$\mathcal{A}\models A$" when reasoning in T; this stands for $\exists x\,(\mathcal{A}_{,x}\models A)$.

Once we fix a description of $\mathcal{A}$, it makes perfect sense to say *"T proves that $\mathcal{A}$ is s.d."*. This simply means:

$$T\vdash \neg(\mathcal{A}\models \bot)\wedge\forall A,B\,(\mathcal{A}\models \Box A\vee\Box B)\rightarrow(\mathcal{A}\models A \vee \mathcal{A}\models B).$$

Obviously, an r.e. set of formulas $\mathcal{A}$ may have different descriptions and for one description the theory T may prove that $\mathcal{A}$ is s.d. while for another description it may not. Note also that possibly the "opinion" of T about $\mathcal{A}$ may be incorrect. In fact, when T is Σ_1-ill, there are descriptions of $\mathcal{A}$ which do not satisfy: $A\in\mathcal{A}$ iff $T\vdash\exists x(A\in\mathcal{A}_{,x})$. So, it may happen that T proves $\mathcal{A}$ is s.d. while this fails to reflect real life. We essentially use this fact in the next section; for the moment we keep the description fixed and assume T proves that $\mathcal{A}$ is s.d..

Lemma 1. Let T be an axiomatized theory of infinite height containing $I\Delta_0$+exp and $\mathcal{A}$ an r.e. set of modal formulas. If there is a description of $\mathcal{A}$ in T such that T proves that $\mathcal{A}$ is s.d. then $\mathcal{A}$ is interpretable in T.

Proof. Let T be an axiomatized theory and "$A\in\mathcal{A}_{,n}$" be a description of an r.e. set of modal formulas as in the hypothesis of the lemma. We shall define a Solovay function h(n) whose value is either 0 or the code of a model of $\mathcal{A}_{,m}$ for some $m\leq n$. We agree that $0\Vdash A$ is some fixed provably false sentence (e.g. $0\neq0$), so the expression $h(n)\Vdash A$ will always have a meaning. The Solovay function is defined, simultaneously with the formulas λ_0 and λ_A, by an arithmetical fixed point. The definition is the following.

Let λ_0 be the sentence $\forall n\, h(n)=0$. We order the modal formulas by increasing code and let A_i be the i-th formula in this order (this enumeration of formulas is redundant, since here formulas are actually codes, but we introduce it for better readability). For every i and every string $\sigma\in 2^i$ define a formula:

$$A_\sigma:=\bigwedge\{A_n \mid n<i \text{ and } \sigma(n)=1\}\wedge\bigwedge\{\neg A_n \mid n<i \text{ and } \sigma(n)=0\}.$$

The formula λ_A (with free variable A) is:

$$\lambda_A := \exists\sigma\in 2^{i+1}[\sigma(i)=1\wedge\exists^\infty n\; h(n)\Vdash A_\sigma \wedge\forall\tau\in 2^{i+1}\,(\tau<\sigma\rightarrow\forall^\infty n\; h(n)\nVdash A_\tau)],$$

where i is such that $A=A_i$ and $\tau<\sigma$ has to be read as τ precedes σ in the lexicographic order. $\exists^\infty n$ is an abbreviation of $\forall m\exists n>m$ and $\forall^\infty n$ of $\neg\exists^\infty n\neg$.

Let h(0)=0. For n+1, if n codes a proof of $\lambda_0\vee\lambda_A$ for some formula A, then:

(a) if h(n)=0 and $\mathcal{A}_{,n}\not\models A$, then choose the minimal model k of $\mathcal{A}_{,n}$ which forces $\neg A$ and define h(n+1)=k.

(b) if $h(n)=h\neq0$ and the root of some submodel of h forces $\neg A$ then let k be the minimal such submodel and define h(n+1)=k.

(c) in all other cases let h(n+1)=h(n).

Note that (provably in T) the graph of h is Δ_0. A straightforward formalization of the completeness theorem for Löb's modal logic shows that h(n) is (roughly) bounded by 2^{2^n} (h increases only if at stage n case (a) obtains; at that stage the code of $\neg A$ and of all the formulas in $\mathcal{A}_{,n}$ is bounded by n). So, Δ_0 induction shows that h is a total function.

If the theory T is strong enough, one can use for λ_A simply the formula $\exists m \forall n{>}m\ h(n) \Vdash A$. Then $\lambda_0 \vee \lambda_A$ simply means that the limit of h is either 0 or a model which forces the formula A, in particular, if h moved to h(n+1) because n codes a proof of $\lambda_0 \vee \lambda_A$, there will be a proof that h(n+1) is not the limit of the function (in fact h(n+1) is chosen so that $h(n+1) \Vdash \neg A$). But in $I\Delta_0$+exp we do not know how to prove that the limit of the Solovay function exists (one needs Σ_1-induction). In particular it cannot be excluded that for some formula A both $h(n) \Vdash A$ and $h(n) \Vdash \neg A$ occurs for infinitely many n; thus one would not have as desired, $\lambda_{\neg A} \leftrightarrow \neg\lambda_A$. To help the reader's intuition we present the following semi-formal description of λ_A which should clarify the definition above. To each formula A we attach an infinite set C(A) such that either $\forall n \in C(A)\ h(n) \Vdash A$ or $\forall n \in C(A)\ h(n) \Vdash \neg A$. The set C(A) is defined in the following way. Let $C(A_0) = \{n \mid h(n) \Vdash \neg A_0\}$ if this is infinite, $C(A_0)=\{n \mid h(n) \Vdash A_0\}$ otherwise. Let $C(A_{i+1}) = \{n \in C(A_i) \mid h(n) \Vdash \neg A_{i+1}\}$ if this is infinite, $C(A_{i+1}) = \{n \in C(A_i) \mid h(n) \Vdash A_{i+1}\}$ otherwise. Finally, let λ_A be $\forall n \in C(A)\ h(n) \Vdash A$.

Claim 1. T proves $\forall n\ [\ h(n) \neq 0 \rightarrow \mathrm{Thm}[\exists m\ h(m)$ is a proper submodel of $h(\dot{n})]\ ]$.
Proof. In fact if $h(n)\neq 0$ then at some stage s<n for some formula A, s codes a proof $\lambda_0 \vee \lambda_A$ and $h(s+1)=h(n) \Vdash \neg A$. By provable Σ_1 completeness $\mathrm{Thm}[\neg\lambda_0]$. This together with $\mathrm{Thm}[\lambda_0 \vee \lambda_A]$ yields $\mathrm{Thm}[\lambda_A]$ and in particular $\mathrm{Thm}[\exists^\infty n\ h(n) \Vdash A]$. From $h(n) \Vdash \neg A$ we get $\mathrm{Thm}[h(\dot{n}) \Vdash \neg A]$ by provable Σ_1 completeness, thus the claim follows.

Claim 2. $\forall n \in \omega\ \exists m \in \omega$ such that T proves $h(n)\neq 0 \rightarrow \mathrm{Thm}^m(\perp)$. (So, since T has infinite height, for every standard n, h(n)=0.)
Proof. This is an easy corollary of the previous claim.

To define $\iota(A)$ we need to assign "ad hoc" a model to 0. Following Shavrukov we shall construct a formula $\mathcal{T}$ in such a way that for all standard formulas A and B the following properties are provable in T.

(1) $\neg\mathcal{T}(\perp)$

(2) $\mathcal{T}(A \rightarrow B) \leftrightarrow (\mathcal{T}(A) \rightarrow \mathcal{T}(B))$

(3) $\mathcal{A} \vDash A \rightarrow \mathcal{T}(A)$.

(4) $\mathcal{T}(\Box A) \rightarrow \mathcal{A} \vDash A$.

(Roughly speaking the formula $\mathcal{T}(A)$ says that A belongs to some maximal consistent set $\mathcal{T}$ containing $\mathcal{A} \cup \{\neg\Box A \mid \mathcal{A} \nvDash \Box A\}$. Such a set $\mathcal{T}$ exists (within T) since otherwise for some $A_0,\ldots,A_n$ such that $\mathcal{A} \nvDash \Box A_0,\ldots, \mathcal{A} \nvDash \Box A_n$ we would have $\mathcal{A} \vDash \Box A_0 \vee \ldots \vee \Box A_n$. This contradicts the provable s.d.p. of $\mathcal{A}$.) For the proof of the lemma only (1)-(4) are needed, so we prefer to postpone the definition of $\mathcal{T}$ and the proof of (1)-(4) after the proof of the lemma.

We define τ_A as $\lambda_0 \wedge \mathcal{T}(A)$, and finally define: $\iota(A) := \lambda_A \vee \tau_A$, i.e. $\lambda_A \vee [\lambda_0 \wedge \mathcal{T}(A)]$. We shall prove that ι is an interpretation (claim 5) and that ι interprets $\mathcal{A}$ in T (claim 6).

Claim 3. For every $A \in \mathcal{L}$, T proves $(\forall^\infty n\ h(n) \Vdash A) \rightarrow \lambda_A$.
Proof. Since A is standard we can replace in the definition of λ_A the quantifications over strings by finite conjunctions and disjunctions. So the claim is trivial.

Claim 4. For every $A \in \mathcal{L}$, T proves $\forall n\ [h(n)=0 \wedge \mathcal{A}_{,n} \models A \rightarrow \iota(A)]$.
Proof. Assume $h(n)=0$ and $\mathcal{A}_{,n} \models A$. Reasoning in T we want to show $\lambda_A \vee \tau_A$. Since $h(n)=0$ and $\mathcal{A}_{,n} \models A$, the function can leave 0 only to a model of A and eventually move to some submodel of it. So $\neg\lambda_0$ implies $\forall^\infty n\ h(n) \models A$. By the previous claim, this implies λ_A. On the other hand, by (3), we have $\mathcal{T}(A)$, so, λ_0 implies τ_A.

Claim 5. The function ι is an interpretation (i.e. properties (i)-(iii) are provable in T.)
Proof. We have to prove that for every standard formula A properties (i)-(iii) are provable in T, i.e. $\iota(\Box A) \leftrightarrow \mathrm{Thm}[\iota(A)]$, $\neg\iota(\bot)$ and $\iota(A \rightarrow B) \leftrightarrow (\iota(A) \rightarrow \iota(B))$. The proof is more readable if we derive them both from $T+\lambda_0$ and from $T+\neg\lambda_0$. In fact under the hypothesis λ_0 the sentence $\iota(A)$ is equivalent to $\mathcal{T}(A)$ (by our convention that $0 \nVdash A$), while, under the hypothesis $\neg\lambda_0$, $\iota(A)$ is equivalent to λ_A.

$T+\lambda_0 \vdash \iota(\Box A) \rightarrow \mathrm{Thm}[\iota(A)]$. Assume $\iota(\Box A)$ and λ_0 and reason in T. As we just remarked, under the assumption λ_0, $\iota(\Box A)$ reduces to $\mathcal{T}(\Box A)$. By (4) we obtain $\mathcal{A} \models A$, so, for some n, $\mathcal{A}_{,n} \models A$. Since we assumed λ_0, $h(n)=0$. Both $\mathcal{A}_{,n} \models A$ and $h(n)=0$ are Σ_1-formulas, so by provable Σ_1-completeness we have $\mathrm{Thm}[\mathcal{A}_{\dot n} \models A]$ and $\mathrm{Thm}[h(\dot n)=0]$. By claim 4 we have $\mathrm{Thm}[\iota(A)]$.

$T+\lambda_0 \vdash \mathrm{Thm}[\iota(A)] \rightarrow \iota(\Box A)$. Assume $\mathrm{Thm}[\lambda_A \vee \tau_A]$ and λ_0. It suffices to show, reasoning in T, that $\mathcal{T}(\Box A)$. Since $\mathrm{Thm}[\lambda_A \vee \tau_A]$, a fortiori $\mathrm{Thm}[\lambda_0 \vee \lambda_A]$. Let n be the code of a proof of $\lambda_0 \vee \lambda_A$; Since we assumed λ_0, $h(n)=0$. Then $\mathcal{A}_{,n} \models A$, otherwise the function would leave 0 at stage n+1, contradicting λ_0. Then $\mathcal{A} \models \Box A$ and so, by (3), $\mathcal{T}(\Box A)$.

$T+\lambda_0 \vdash \neg\iota(\bot)$. Immediate from (1).

$T+\lambda_0 \vdash \iota(A \rightarrow B) \leftrightarrow (\iota(A) \rightarrow \iota(B))$. Immediate from (2).

$T+\neg\lambda_0 \vdash \iota(\Box A) \rightarrow \mathrm{Thm}[\iota(A)]$. Assume $\iota(\Box A)$ and $\neg\lambda_0$. It suffices to prove $\mathrm{Thm}[\lambda_A]$ in T. By our assumption $\lambda_{\Box A}$ holds, in particular for some n, $h(n) \Vdash \Box A$. The latter is a Σ_1-formula so $\mathrm{Thm}[h(\dot n) \Vdash \Box A]$. Since $h(n) \neq 0$, by claim 1 we have Thm["$\exists m$ h(m) is a submodel of $h(\dot n)$"], thus $\mathrm{Thm}[\forall^\infty n\ h(n) \Vdash A]$. By claim 3, $\mathrm{Thm}[\lambda_A]$ follows.

$T+\neg\lambda_0 \vdash \mathrm{Thm}[\iota(A)] \rightarrow \iota(\Box A)$. Assume $\mathrm{Thm}[\lambda_A \vee \tau_A]$ and $\neg\lambda_0$. It suffices to derive $\lambda_{\Box A}$ reasoning in T. Since $\mathrm{Thm}[\lambda_A \vee \tau_A]$, a fortiori $\mathrm{Thm}[\lambda_0 \vee \lambda_A]$. Let n be a code of a proof of $\lambda_0 \vee \lambda_A$ which is large enough to have $h(n) \neq 0$. (Such an n exists since we assumed $\neg\lambda_0$ and any provable sentence has arbitrary large proofs.) If $h(n) \Vdash \Box A$ then $h(n+1)=h(n)$, otherwise; $h(n+1)$ will be the least submodel of h(n) forcing $\neg A$. In both cases $h(n+1) \Vdash \Box A$ (recall that the code of a model is larger than the code of its proper submodels). Afterwards, h remains confined in a submodel of h(n+1) so, we can conclude that $\forall^{\infty}n\ h(n)\Vdash\Box A$. Thus $\lambda_{\Box A}$ follows by claim 3.

$T+\neg\lambda_0\vdash\neg\iota(\bot)$. Immediate.

$T+\neg\lambda_0\vdash\iota(A\rightarrow B)\leftrightarrow(\iota(A)\rightarrow\iota(B))$. Is left to the reader.

Claim 6. For every $A\in\mathcal{L}$, $\mathcal{A}\vDash A$ iff $T\vdash\iota(A)$.
Proof. ($\Longrightarrow$) Assume $\mathcal{A}\vDash A$. Then for some $\mathcal{A}_{,n}\vDash A$. Since n is standard h(n)=0 and, by Σ_1-completeness, $T\vdash h(n)=0\wedge\mathcal{A}_{,n}\vDash A$. So $\iota(A)$ by claim 4. Vice versa, ($\Longleftarrow$), if $T\vdash\iota(A)$ we have in particular that $T\vdash\lambda_0\vee\lambda_A$. Assume for a contradiction that $\mathcal{A}\nvDash A$ and let n be the code of the proof of $\lambda_0\vee\lambda_A$. In particular we have that $\mathcal{A}_{,n}\nvDash A$ then $h(n+1)\neq 0$. This n is a standard number, so this contradicts the fact that h will spend all of its standard life in 0.

The proof of the lemma is complete but for the definition of the predicate $\mathcal{T}$. We introduce the formula $V(\sigma)$ which roughly says: A_σ is $\Box$-conservative over $\mathcal{A}$, namely

$$V(\sigma):=\forall A\,[\,(\mathcal{A}\vDash A_\sigma\rightarrow\Box A)\rightarrow(\mathcal{A}\vDash\Box A)\,].$$

Assume strings have been coded into numbers in some natural way, (e.g. choose $\Sigma_{\sigma(i)t=1}\,2^i$ as code of σ) so that on strings of equal length the relation "<" coincides with the relation "precedes lexicographically" or, when strings are thought of as nodes of a binary tree, "is to the left of". Let $U(\sigma)$ be the formula which says that σ is the leftmost string satisfying $V(\sigma)$,

$$U(\sigma):=V(\sigma)\wedge\forall\tau\in 2^{i+1}(\tau<\sigma\rightarrow\neg\ V(\tau))].$$

If $A=A_i$ let $\mathcal{T}(A)$ hold if there is $\sigma\in 2^{i+1}$ such that $U(\sigma)$ and $\sigma(i)=1$. We have to show that for every standard formula properties (1) to (4) of $\mathcal{T}$ are provable in T. As a first thing let us remark that for all standard i, T proves $\exists\sigma\in 2^{i+1}\,U(\sigma)$, i.e. i.e. there exists a leftmost string σ satisfying $V(\sigma)$. Reason in T. A string satisfying $V(\sigma)$ must exist, otherwise for every $\sigma\in 2^{i+1}$ there would be a modal formula C_σ such that $\mathcal{A}\vDash A_\sigma\rightarrow\Box C_\sigma$ and $\mathcal{A}\nvDash\Box C_\sigma$. Since $\bigvee_{\sigma\in 2^{i+1}}A_\sigma$ is a tautology, one would have $\mathcal{A}\vDash\bigvee_{\sigma\in 2^{i+1}}\Box C_\sigma$. By the s.d.p. of $\mathcal{A}$ (provable in T) $\mathcal{A}\vDash\Box C_\sigma$ for some σ, a contradiction. Now, once we know that one string σ exists satisfying $V(\sigma)$, the existence of the minimal one is again a consequence of the standardness of i since the quantifiers over strings in 2^{i+1} may be transformed in finite conjunctions and disjunctions. This proves our remark. Now we check in turn that the properties (1) to (4) which we required for $\mathcal{T}$ are provable in T.

(1) $\neg\mathcal{T}(\bot)$

(2) $\mathcal{T}(A\rightarrow B)\leftrightarrow(\mathcal{T}(A)\rightarrow\mathcal{T}(B))$

(3) $\mathcal{A}\vDash A\rightarrow\mathcal{T}(A)$.

(4) $\mathcal{T}(\Box A)\rightarrow\mathcal{A}\vDash A$.

We reason in T. It is obvious that for no string σ such that $V(\sigma)$, $\sigma(\bot)=1$, so (1) holds. (We write $\sigma(A)$ for $\sigma(i)$ where $A=A_i$.) To prove (2) assume first that $\mathcal{T}(A\rightarrow B)$ and $\mathcal{T}(A)$. Let σ be a sufficiently long string such that $U(\sigma)$ and $\sigma(A\rightarrow B)=\sigma(A)=1$. Then $\sigma(B)=1$ otherwise $A_\sigma\leftrightarrow\bot$

and surely could not satisfy $V(\sigma)$. The converse is similar. Property (3) is also a direct consequence of the existence of an arbitrary (standard) long string satisfying $U(\sigma)$. For such a string we must have $\sigma(A)=1$ otherwise $\mathcal{A}\Vdash A_\sigma \rightarrow \perp$ and, by the definition of $V(\sigma)$, $\mathcal{A}\Vdash \perp$. Last, to prove (4) assume that $\mathcal{T}(\Box A)$. Let σ be a sufficiently long string such that $U(\sigma)$ and $\sigma(\Box A)=1$. Then $\mathcal{A}\Vdash A_\sigma \rightarrow \Box A$, so, by the definition of $V(\sigma)$, $\mathcal{A}\Vdash \Box A$. By the s.d.p. of $\mathcal{A}$ we get $\mathcal{A}\Vdash A$.

This completes the proof of lemma 1.□

2. The theorems.

We shall use lemma 1 to prove the two theorems announced in the introduction. They characterize the r.e. sets interpretable in a theory of infinite height.

Theorem 1. If $\mathcal{A}$ is an r.e. set of modal formulas and T is a Σ_1 sound theory containing $I\Delta_0$+exp, then $\mathcal{A}$ is interpretable in T iff $\mathcal{A}$ is s.d..

Theorem 2. If $\mathcal{A}$ is an r.e. set of modal formulas and T is a Σ_1 ill theory of infinite height containing $I\Delta_0$+exp, then $\mathcal{A}$ is interpretable in T iff $\mathcal{A}$ has infinite height .

The "only if" part of both theorems is trivial. To prove the first theorem we show that, if $\mathcal{A}$ is an r.e. set with the s.d.p. and T is a Σ_1-sound theory, then we can find a description of $\mathcal{A}$ in T such that T proves the s.d.p. of $\mathcal{A}$. Analogously for the second theorem. For the sake of readability we shall give these proofs in an informal style, namely we shall merely describe algorithms and take for granted their formalization in the language of T.

Suppose $\mathcal{A}$ is an r.e. set of modal formulas and let $A\in\mathcal{A}_s$ be any description of $\mathcal{A}$. To this description we associate in a natural way the algorithm $\{\mathcal{A}_{,s}\}_{s\in\omega}$ enumerating $\mathcal{A}$, i.e. an increasing recursive sequence of finite sets $\{\mathcal{A}_{,s}\}_{s\in\omega}$ such that $\mathcal{A}=\bigcup_{s\in\omega}\mathcal{A}_{,s}$. We shall construct a new algorithm $\{\mathcal{V}_{,s}\}_{s\in\omega}$ enumerating the same set $\mathcal{A}$ such that the canonical translation of $\{\mathcal{V}_{,s}\}_{s\in\omega}$ in the language of the arithmetic yields a description with the desired properties.

The proofs of theorems 1 and 2 need two modal lemmas, respectively lemma 2 and 3. These are the adaptations of some lemmas of [1]. We shall present them in a form which is easily formalized and proved in $I\Delta_0$+exp. Their proofs are moved to the end of this section.

A finite set $\mathcal{C}$ of formulas is said to be *adequate* if it is closed under subformulas and (up to provable equivalence) closed under Boolean connectives. Namely, if: (i) $\perp\in\mathcal{C}$, (ii) all subformulas of every $B\in\mathcal{C}$ are in $\mathcal{C}$, (iii) for every $B,C\in\mathcal{C}$ there exists $D\in\mathcal{C}$ such that $\Vdash D\leftrightarrow(B\rightarrow C)$.

Lemma 2. Let $\mathcal{C}$ be a finite adequate set and let $\mathcal{A}\subseteq\mathcal{C}$. The following are equivalent:

(a) $\mathcal{A}$ is s.d. (b) $\mathcal{A}\nVdash\perp$ and $\forall B,C\in\mathcal{C}$ $\mathcal{A}\Vdash \Box B\vee\Box C \Rightarrow \mathcal{A}\Vdash B$ or $\mathcal{A}\Vdash C$.□

Proof of theorem 1. We are now ready to present the algorithm required to prove theorem 1. Without loss of generality we may assume that if $\mathcal{A} \models A$ then $A \in \mathcal{A}$. We may code finite sets of formulas with natural numbers. The property "s codes an adequate set" is Δ_0. Consider the following algorithm $\{\mathcal{V}_{,s}\}_{s\in\omega}$.

(Stage 0) $\mathcal{V}_{,0} = \emptyset$.
(Stage s+1) Let A be the minimal formula (if such exists) such that $A \in \mathcal{A}_{,s} - \mathcal{V}_{,s}$. If for some adequate set C of code less than s, $A \in C$, $\mathcal{V}_{,s} \subseteq \mathcal{A}_{,s} \cap C$, and condition (b) of lemma 2 holds for $\mathcal{A}_{,s} \cap C$ then, let $\mathcal{V}_{,s+1} = \mathcal{A}_{,s} \cap C$. Otherwise let $\mathcal{V}_{,s+1} = \mathcal{V}_{,s}$.

We check by induction on the code of the (standard) formula A that $A \in \mathcal{A}$ iff $A \in \bigcup_{s\in\omega} \mathcal{V}_{,s}$. Since $\mathcal{V}_{,s} \subseteq \mathcal{A}_{,s}$, only one implication needs to be proved. Suppose for a contradiction there is a formula such that $A \in \mathcal{A}_{,s} - \mathcal{V}_{,s}$ for all large enough $s \in \omega$. Fix A and s such that for all $r \geq s$, A is the least formula in $\mathcal{A}_{,r} - \mathcal{V}_{,r}$. Fix an adequate set C, such that $\{A\} \cup \mathcal{V}_{,s} \subseteq C$ (such an adequate set exists since A and s are standard). Let $n > s$ be larger than the code of C and such that $\mathcal{A} \cap C \subseteq \mathcal{A}_{,n} \cap C$. Clearly $\mathcal{V}_{,s} \subseteq \mathcal{A}_{,n} \cap C$. Since $\mathcal{A}$ is s.d. and we assumed it closed under $\models$, condition (b) of lemma 2 holds for $\mathcal{A}_{,n} \cap C$. So, $\mathcal{V}_{,n+1} = \mathcal{A}_{,n} \cap C$, a contradiction. It remains to be checked that T proves the s.d.p. of $\bigcup_s \mathcal{V}_{,s}$. For this we need a formalized version of lemma 2 in $I\Delta_0$+exp so we invite the reader to check that all models used in the proof given below are bounded by a few nested exponentiations of the code of the given adequate set C. Consequently, the theorem holds in any model of $I\Delta_0$+exp. From lemma 2 it follows that for all stages s the sets $\mathcal{V}_{,s}$ are s.d., which clearly suffices.□

Lemma 3. Let C be a finite adequate set containing $\mathcal{A}$. The following are equivalent:

(1) $\mathcal{A}$ has infinite height (2) there exists $B \in C$ such that B is s.d. and $B \models \bigwedge \mathcal{A}$.□

Proof of theorem 2. Given a Σ_1-ill theory T choose a Δ_0-formula $\sigma(x)$ such that $T \vdash \exists x \sigma(x)$ and $\omega \models \forall x \neg\sigma(x)$. In every model of T there is a Δ_0 definable nonstandard number n, namely the minimal witness of $\exists x\ \sigma(x)$. The idea of the proof is the following: given any algorithm $\mathcal{A}_{,s}$ enumerating $\mathcal{A}$ we construct a new algorithm which simulates $\mathcal{A}_{,s}$ until the nonstandard stage n, but once this stage is reached we stop the simulation and enumerate some arbitrary s.d. set containing $\mathcal{A}_{,n}$. In the real world this stage n is never reached, so this new algorithm enumerates the same set as the old one. But in any model of T this algorithm enumerates a nonstandard finite s.d. set. Lemma 3 is used to guarantee that some s.d. formula $B \models \mathcal{A}_{,s}$ always exists.

(Stage 0) $\mathcal{V}_{,0} = \emptyset$.
(Stage s+1) Let A be the minimal formula (if such exists) such that $A \in \mathcal{A}_{,s} - \mathcal{V}_{,s}$. If for some adequate set C of code less than s, $A \in C$, $\mathcal{V}_{,s} \subseteq \mathcal{A}_{,s} \cap C$, for some $B \in C$ condition (b) of Lemma 2 holds and $B \models \mathcal{A}_{,s} \cap C$, then
case 1: if $\forall x \leq s\ \neg\sigma(x)$ let $\mathcal{V}_{,s+1} = \mathcal{A}_{,s} \cap C$,
case 2: if $\exists x < s\ \sigma(x)$ let $\mathcal{V}_{,s+1} = \{B\}$ for some s.d. formula $B \in C$ such that $B \models \mathcal{A}_{,s} \cap C$.
Otherwise let $\mathcal{V}_{,s+1} = \mathcal{V}_{,s}$.

We check by induction on the code of the formula A that $A \in \mathcal{A}$ iff $A \in \bigcup_{s\in\omega}\mathcal{V}_{,s}$. Since $\mathcal{V}_{,s} \subseteq \mathcal{A}_{,s}$, only one implication needs to be proved. We need consider only standard stages (recall that a description of $\mathcal{A}$ should verify: $A \in \mathcal{A}$ iff $\exists s \in \omega$, $T \vdash A \in \mathcal{V}_{,s}$), so case 2 never obtain. Suppose for a contradiction there is a formula such that $A \in \mathcal{A}_{,s} - \mathcal{V}_{,s}$ for all $s \in \omega$. Fix A and s such that for all $r \geq s$, A is the least formula in $\mathcal{A}_{,r} - \mathcal{V}_{,r}$. Fix an adequate set C, such that $\{A\} \cup \mathcal{V}_{,s} \subseteq C$ (such an adequate set exists since A is standard). Let $n>s$ be larger than the code of C and such that $\mathcal{A} \cap C \subseteq \mathcal{A}_{,n} \cap C$. Clearly $\mathcal{V}_{,s} \subseteq \mathcal{A}_{,n} \cap C$ and, since $\mathcal{A}$ has infinite height, so does $\mathcal{A}_{,n} \cap C$. Thus, condition (2) of lemma 3 holds for $\mathcal{A}_{,n} \cap C$. We may conclude that $\mathcal{V}_{,n+1} = \mathcal{A}_{,n} \cap C$, a contradiction. To check that T proves the s.d.p. of $\bigcup_s \mathcal{V}_{,s}$ recall that in every model of T, $\bigcup_s \mathcal{V}_{,s} = \bigcup_{s<n+1} \mathcal{V}_{,s}$, where n is the least number such that $\sigma(n)$ and $\bigcup_{s<n+1} \mathcal{V}_{,s}$ is equivalent to a single s.d. formula B.$\square$

Proof of lemma 2. The direction (a)$\Longrightarrow$(b) is trivial. For the converse, assume (b). Fix a set $\mathcal{A}t \subseteq C$ such that:

$$\mathcal{A}t := \{G \in C \mid \forall C \in C \text{ either } G \Vdash C \text{ or } G \Vdash \neg C\}.$$

The elements of $\mathcal{A}t$ are called *atoms* ; roughly, they are conjunctions of maximal consistent subsets of C. By the adequacy of C, for every $C \in C$, if $\nVdash \neg C$ then there is some atom $G \Vdash C$. Also, $\Vdash \bigvee \mathcal{A}t$, otherwise, for some atoms G, $G \Vdash \neg \bigvee \mathcal{A}t$ quod non. Let $\gamma = \{G \in \mathcal{A}t \mid \mathcal{A} \nvDash \neg G\}$. From $\Vdash \bigvee \mathcal{A}t$ and $\mathcal{A} \nvDash \bot$ we can conclude $\gamma \neq \emptyset$. We claim that there is a model of $\mathcal{A} \cup \{\Diamond G \mid G \in \gamma\}$. In fact, if not, then $\mathcal{A} \vDash \bigvee_{G\in\gamma} \Box \neg G$. By (b), there is $G \in \gamma$ such that $\mathcal{A} \vDash \neg G$ quod non. This proves the claim.

Suppose now that for some formulas B_1, B_2 both $\mathcal{A} \nvDash B_1$ and $\mathcal{A} \nvDash B_2$, so we may assume that there are two models k_1 and k_2 of $\mathcal{A}$ forcing respectively $\neg B_1$ and $\neg B_2$. We shall show that $\mathcal{A} \nvDash \Box B_1 \vee \Box B_2$ by constructing a model k' of $\mathcal{A}$ which contains k_1 and k_2 as proper submodels. The s.d.p. of $\mathcal{A}$ will follow.

Let k be a model of $\mathcal{A} \cup \{\Diamond G \mid G \in \gamma\}$. Let r, r_1 and r_2 be the roots of respectively k,k_1 and k_2. Let R,R_1 and R_2 be the respective accessibility relations. Let k' be the model obtained grafting k_1 and k_2 above the root of k. More precisely, the universe of k' is the disjoint union of the universes of k,k_1 and k_2 and the accessibility relation of k' is the transitive closure of the relation $R \cup R_1 \cup R_2 \cup \{(r,r_1),(r,r_2)\}$. The forcing relation of k' is the union of the forcing relations of k,k_1 and k_2.

We claim that k' is a model of $\mathcal{A}$ and $k' \Vdash \neg \Box B_1 \wedge \neg \Box B_2$. Obviously k' forces $\neg \Box B_1 \wedge \neg \Box B_2$ because k_1 and k_2 are submodels of k' forcing respectively $\neg B_1$ and $\neg B_2$. To show that k' is a model of $\mathcal{A}$, we prove by induction on the complexity of subformulas $C \in C$ that $k' \Vdash C$ iff $k \Vdash C$. The basis step is trivial as well as the induction for Boolean connectives. We prove the induction step for $\Box$. Assume $k' \Vdash \neg \Box C$. Then for some proper submodel w' of k', $w' \Vdash \neg C$. The model w' is a submodel of k_1 or of k_2 or is a proper submodel of k. If w' is a proper submodel of k, then $k \Vdash \neg \Box C$ follows. Otherwise, let G be the atom forced in w'; since $C \in C$, by the definition of atom, either $G \Vdash C$ or $G \Vdash \neg C$. But $G \Vdash C$ leads immediately to contradiction so, $G \Vdash \neg C$. Since both k_1 and k_2 are models of $\mathcal{A}$, $G \in \gamma$. By our choice of k, $k \Vdash \bigwedge_{G\in\gamma} \Diamond G$, so there is a proper submodel w of k which forces G. Hence $w \Vdash \neg C$ and k$\Vdash \neg\Box$C. Vice versa if k$\Vdash \neg\Box$C, then for some proper submodel w of k, w$\Vdash\neg$C. Since w is also a proper submodel of k', k'$\Vdash\neg\Box$C follows. This completes the proof of the lemma.$\Box$

Proof of lemma 3. ($1 \Leftarrow 2$) is immediate. ($1 \Rightarrow 2$) List the formulas of $\mathcal{C}=\{C_1,\ldots,C_n\}$. Define $\mathcal{A}_0:=\mathcal{A}$ and for all i$\leq$n let $\mathcal{A}_{i+1}:=\mathcal{A}_i\cup\{C_i\}$ if this has infinite height, $\mathcal{A}_{i+1}:=\mathcal{A}_i$ otherwise. Finally choose in $\mathcal{C}$ a formula B equivalent to $\bigwedge\mathcal{A}_{n+1}$. If B$\vDash \Box C_i \vee \Box C_j$ then $B\wedge C_i$ or $B\wedge C_j$ has infinite height. (For suppose for some n both $B\wedge C_i \vDash \Box^n\bot$ and $B\wedge C_j\vDash\Box^n\bot$ then $B\vDash\Box C_i\rightarrow\Box^{n+1}\bot$ and $B\vDash\Box C_j\rightarrow\Box^{n+1}\bot$. Thus $B\vDash\Box^{n+1}\bot$, quod non.) So, one of C_i and C_j, say C_i, has been enumerated in $\mathcal{A}_{n+1}$, so $B\vDash C_i$. By lemma 2, B is s.d..$\Box$

References.

[1] V. Yu. Shavrukov, Subalgebras of diagonalizable algebras of theories containing arithmetic, *Dissertationes Mathematicae* CCCXXIII, (1993), 82 pp.

[2] R. Magari, Representation and duality theory for diagonalizable algebras (the algebraization of the theories which express Theor.: II), *Bolletino dell' unione Matematica Italiana* (4) 12, 117-125, (1975)

[3] R. Magari, The diagonalizable algebras (the algebraization of the theories which express Theor.: IV), *Studia Logica* 34, 305-313, (1975)

[4] C. Bernardi, On the equational class of diagonalizable algebras (the algebraization of the theories which express Theor.: VI), *Studia Logica* 34, 321-331, (1975)

[5] F. Bellisima, On the modal logic corresponding to diagonalizable algebra theory, *Bolletino dell' Unione Matematica Italiana* B (5) 15, 915-930, (1978)

[6] F. Montagna, On the diagonalizable algebra of Peano arithmetic, *Bolletino dell' Unione Matematica Italiana* B (5) 16, 795-812, (1979)

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

[8] P. Hájek and P. Pudlák, *Methamathematics of first order arithmetic* . Springer Verlag, (1993).

Samenvatting

Dit Proefschrift bestaat uit twee delen. Het eerste deel is aan de begrensde rekenkunde gewijd. Het eerste hoofdstuk daarvan bevat een inleidende paragraaf waarin ook op de motivatie van het onderzoek wordt ingegaan. Ik bestudeer uitbreidingen van zwakke fragmenten van de Peano–rekenkunde tot tweede–orde theorieën. Tweede orde variabelen staan voor eindige verzamelingen van natuurlijke getallen. Ik beperk me tot zwakke fragmenten van de Peano–rekenkunde d.w.z. theorieën die niet kunnen bewijzen dat de exponentiatiefunctie totaal is. Dat houdt in dat er eindige verzamelingen zijn die, hoewel ze te definiëren zijn met begrensde formules, niet gecodeerd kunnen worden door natuurlijke getallen. Dat maakt deze tweede–orde taal echt expressiever. Ik definiëer een hiërarchie van begrensde formules door het tellen van de wisselingen van tweede–orde begrensde kwantoren. Daarna wordt een hiërarchie van theorieën gedefinieerd door het introduceren van comprehensie–axioma's voor formules in deze klassen.

Het is niet bekend of de bovengenoemde hiërarchie van begrensde formules een echte hiërarchie is; ook niet als we ons beperken tot het standaardmodel. Dit blijkt een uiterst moeilijk probleem want het is equivalent met de vraag of de polynomiale hiërarchie instort. Een ermee verbonden vraag is of de hiërarchie van fragmenten van de begrensde rekenkunde ook instort. Hoewel dit tweede probleem sterk op het eerste lijkt is de relatie ertussen nog niet volledig begrepen. Ik laat zien dat, als de begrensde rekenkunde gelijk is aan een van haar fragmenten, dan is het bewijsbaar (in de begrensde rekenkunde) dat de polynomiale–tijd–hiërarchie instort.

In het tweede hoofdstuk behandel ik een fragment van de begrensde rekenkunde van een andere soort. Hier wordt het comprehensie–axioma voor alle begrensde formules aangenomen maar de vermenigvuldigingsfunctie wordt uit de taal weggelaten. Ik noem deze theorie lineaire (begrensde) rekenkunde omdat de termen van de taal lineair zijn. Ik bewijs dat elk model van de lineaire rekenkunde een eindextensie heeft tot een fragment van de begrensde rekenkunde waarin vermenigvuldiging totaal is. Dat gaat echter ten koste van comprehensie.

Het tweede deel van dit proefschrift is gewijd aan de bewijsbaarheidslogica. De grondbegrippen van dit vak zijn in een korte inleiding samengevat. In hoofdstuk drie geven we nieuwe bewijzen van de aritmetische volledigheid van ILP and ILM. Albert Visser bewees dat ILP de modale logica voor de interpreteerbaarheid over eindig geaxiomatiseerde theorieën is. Volodya Shavrukov en Alessandro Berarducci hebben (onafhankelijk van elkaar) laten zien dat ILM de interpreteerbaarheidslogica van essentieel reflexieve theorieën is. Mijn bewijs van deze twee stellingen onthult de gemeenschappelijke aspecten van deze twee stellingen.

Het vierde hoofdstuk gaat over diagonaliseerbare algebra's, met name over subalgebra's van de diagonaliseerbare algebra van aritmetische theorieën. Naar aanleiding van een stelling van Volodya Shavrukov behandel ik de vraag of zijn resultaten ook voor zwakkere theorieën geldig zijn. Ik laat zien dat het bewijs van Volodya Shavrukov kan worden aangepast om de stelling ook voor deze zwakkere theorieën te bewijzen.

Previous titles in the ILLC Dissertation Series:

Transsentential Meditations; Ups and downs in dynamic semantics
Paul Dekker
ILLC Dissertation series, 1993-1

Resource Bounded Reductions
Harry Buhrman
ILLC Dissertation series, 1993-2

Efficient Metamathematics
Rineke Verbrugge
ILLC Dissertation series, 1993-3

Extending Modal Logic
Maarten de Rijke
ILLC Dissertation series, 1993-4

Studied Flexibility
Herman Hendriks
ILLC Dissertation series, 1993-5

Aspects of Algorithms and Complexity
John Tromp
ILLC Dissertation series, 1993-6

The Noble Art of Linear Decorating
Harold Schellinx
ILLC Dissertation series, 1994-1

Generating Uniform User-Interfaces for Interactive Programming Environments
Jan Willem Cornelis Koorn
ILLC Dissertation series, 1994-2

Process Theory and Equation Solving
Nicoline Johanna Drost
ILLC Dissertation series, 1994-3

Calculi for Constructive Communication, a Study of the Dynamics of Partial States
Jan Jaspars
ILLC Dissertation series, 1994-4

Executable Language Definitions, Case Studies and Origin Tracking Techniques
Arie van Deursen
ILLC Dissertation series, 1994-5