

Reflection principles and provability algebras in formal arithmetic

L.D. Beklemishev*

Steklov Mathematical Institute
Gubkina str. 8, 117966 Moscow, Russia
e-mail: `bekl@mi.ras.ru`
Utrecht University, the Netherlands

May 3, 2005

Abstract

We study reflection principles in fragments of Peano arithmetic and their applications to the questions of comparison and classification of arithmetical theories.

Bibliography: 95 items.

Contents

1	Preliminaries	5
1.1	Elementary arithmetic	5
1.2	Provability predicate and Gödel's theorems	7
1.3	Provability logic	12
2	Reflection principles	13
2.1	Forms of reflection	13
2.2	Partial reflection principles	15
2.3	n -provability and n -consistency	16
2.4	Unboundedness theorems	18
2.5	Hierarchies of partial reflection principles	21
3	Provably total computable functions and fragments of PA	27
3.1	Fragments of PA	28
3.2	Provably total computable functions and 1-consistency	30

*Supported by the Russian Foundation for Basic Research and the Council for support of leading scientific schools.

4	Induction and reflection	34
4.1	Tait calculus	34
4.2	Induction schemata with parameters	36
4.3	Induction rules	40
4.4	Parameter-free induction	43
4.5	Reflection schemata and reflection rules	47
5	Provability algebras and ordinal analysis of PA	51
5.1	Lindenbaum algebras	51
5.2	Provability algebras and logics	52
5.3	Graded provability algebras and Japaridze's logic	53
5.4	Reduction property	55
5.5	Closed formulas of GLP and a system of ordinal notation for ε_0	56
5.6	A Gentzen-style consistency proof for PA	58
5.7	A PA -unprovable combinatorial statement	63
5.8	Independence of WP	65
6	Iterated reflection principles	68
6.1	Smooth Turing progressions	69

Introduction

A well-known consequence of Gödel's incompleteness theorems is that there is no single privileged universal axiomatic system: all sufficiently rich recursively enumerable consistent theories are incomplete. Formal systems may differ by the expressivity of their languages as well as by the logical strength of their axioms. Besides, even for the theories with the same body of expressible and provable formulas, the properties of proofs, in particular their complexity, can be substantially different. Thus, one is confronted with the question of choice of a suitable axiomatic system and with those of comparison and classification of different systems.

Proof theory accumulated a good deal of experience in the study of specific formal theories such as Peano arithmetic **PA** or some fragments of the second order arithmetic or set theory. (See, for example, a comprehensive survey [68].) These investigations, in particular, show that the “strength” of a sufficiently expressive theory can be characterized by certain countable ordinals. For many concrete theories these ordinals can be explicitly calculated, which constitutes the subject of the so-called *ordinal analysis* of axiomatic systems [67, 72, 73].

Despite the large amount and depth of the results accumulated in this area, a number of principal problems remain open.

Firstly, there is no clear understanding in what terms proof-theoretic ordinals should be specified. In fact, there is no general answer to the question what it means “to compute the proof-theoretic ordinal of a theory.” This problem is known in several variants since the 50s of the previous century as *the problem of canonicity of ordinal notations* [49, 50, 39]. At present, suitable natural descriptions of ordinals are only known for a number of concrete theories. However,

this does not yet allow one to speak of a general classification of theories by their proof-theoretic ordinals.

Secondly, the ordinals for specific strong theories, such as the second order arithmetic with the full comprehension schema or Zermelo–Fraenkel set theory, have not yet been computed. These questions remain among the most well-known and difficult open problems in mathematical logic.

Technical difficulties arising from the attempts to realize the program of ordinal analysis have recently increased the interest of specialists in some less traditional approaches to these questions and to the attempts to simplify and to better understand the applicable methods. Among such attempts one can mention papers [32, 33] where a relationship between cut-elimination procedures à la Gentzen and à la Novikov–Schütte is studied, papers [16, 17, 15] on a model-theoretic approach to ordinal analysis, and papers [58, 11, 12, 91] on the so-called ε -substitution method.

In author’s papers [27, 26], a new approach to the ordinal analysis of Peano arithmetic based on reflection principles and the so-called *provability algebras* was suggested. In the present paper we systematize currently known results in this direction and survey the results concerning the applications of reflection principles to the study of fragments of arithmetic.

Reflection principles emerged in mathematical logic immediately after Gödel’s fundamental work on the incompleteness theorems [42].¹ For a given theory T , these principles are schemata formalizing variants of the statement “*if a formula φ is provable in T , then φ is true.*” They deliver examples of true but unprovable sentences generalizing the very first example of this kind — Gödel’s consistency assertion for T .

Turing [92] introduced progressions of theories obtained by iterating the process of extension of a theory by a reflection principle and suggested a possible approach to the ordinal classification of theories by means of such progressions. Turing’s ideas were later analyzed and further developed by Feferman in [38]. However, the examples constructed by Turing and Feferman showed that there were substantial difficulties in the way of such a classification related, in particular, to the question of a natural representation of ordinals in arithmetic.

Kreisel and Lévy [51] showed that reflection principles were a convenient tool in the study of logical complexity of axiomatization of formal theories. They proved that the so-called uniform reflection principle for the primitive recursive arithmetic was deductively equivalent to the full induction schema. It follows, in particular, that Peano arithmetic cannot be axiomatized by a set of arithmetical formulas of restricted quantifier complexity. In the same paper, they proved that the uniform reflection principle for Peano arithmetic was equivalent to the schema of transfinite induction up to the ordinal ε_0 . Later, relationships between reflection principles and some other true unprovable statements were established, including a well-known combinatorial principle due to Paris and

¹Reflection principles seem to have been explicitly formulated for the first time in a paper by Rosser of 1937 [76]. Rosser quotes some unpublished results by Kleene who already in 1935 considered a logical rule equivalent, using a modern terminology, to the uniform reflection schema (cf. §2.1 below).

Harrington. A survey of the early (until 1977) results on reflection principles is contained in Smoryński [85].

We demonstrate that systematically using reflection principles allows one to obtain a lot of information about a given theory and its relationships with the other theories. In particular, we obtain by this method the main known results on independence, axiomatizability, conservativity, and the classes of provably total computable functions for the fragments of PA defined by various restrictions of the induction schema over the elementary arithmetic.

Provability algebras. A finite reflection schema R over a theory T can be considered as an operator acting on the Lindenbaum boolean algebra of T . The study of such operators naturally leads to algebraic structures expanding the Lindenbaum algebra that we call *provability algebras* [26]. The simplest example of such a structure is the *Magari algebra of T* where one considers Gödel's consistency formula for T (or its dual provability formula) as R [56]. The variety generated by such algebras was actively studied in the context of the so-called *provability logic* (cf. [87, 31] and also [81, 83]). An extension of provability logic corresponding to some stronger reflection principles was studied in [7, 45].

It is shown in [26] that provability algebras are related to the proof-theoretic ordinals of axiomatic systems. Interesting connections found allow one to view several classical results in proof theory from a novel standpoint. In particular, one can canonically extract the standard ordinal notation system for ε_0 from the provability algebra of the elementary arithmetic and to obtain a new Gentzen-style consistency proof for Peano arithmetic by transfinite induction. The study of provability algebras also yields a characterization of the class of provably total computable functions of PA and new simple combinatorial statements independent from PA [27].

Plan of the paper. In §1 we introduce basic concepts concerning formal arithmetic. We also consider basic properties of the provability formula and give short proofs of theorems due to Gödel, Rosser and Löb.

In §2 we study general properties of local and uniform reflection principles and prove the so-called unboundedness theorems of Kreisel–Lévy.

In §3 basic facts on provably total recursive functions of theories and their relationships with reflection principles are presented.

In §4 the results relating induction and reflection schemata in arithmetic are treated. We show that the fragments of PA defined by various restrictions of the schema of induction precisely correspond (are deductively equivalent) to certain reflection schemata over the elementary arithmetic. We also establish here a conservation result for the uniform reflection schemata over reflection rules, the so-called reduction property, which plays an important role in what follows.

Together with the results of §2 this allows us to obtain by a uniform method answers to a number of questions concerning various properties of fragments of PA. Thus, Corollary 4.20 yields a characterization of provably total computable functions in a theory axiomatized by the induction schema for Π_2 -formulas

without parameters. These functions, in fact, coincide with the primitive recursive functions. An extension of this theory by the schema of induction for Σ_1 -formulas *with* parameters has a wider class of provably total computable functions, which coincides with the class of doubly recursive functions in the sense of Peter (Corollary 6.3). These results, obtained in [24], are perhaps the most interesting applications of the techniques of §§2 and 4.

In §5 the notion of provability algebra is introduced, which forms a basis for our approach towards ordinal analysis of Peano arithmetic. We give a new consistency proof of PA by transfinite induction up to the ordinal ε_0 . We also give a simple example of a true combinatorial statement motivated by provability algebras that is unprovable in Peano arithmetic. Finally, we establish a connection between provability algebras and transfinite progressions of iterated reflection principles à la Turing–Feferman. Using this relationship, we obtain a characterization of Π_n -consequences of PA in terms of such progressions and draw some other interesting corollaries.

1 Preliminaries

1.1 Elementary arithmetic

The language of arithmetic is a first order language containing binary predicate symbols $=$ and \leq ; binary function symbols $+$ and \cdot ; unary function symbols S and \exp ; and a constant 0 . The *standard model* of arithmetic is a model with the universe $\mathbb{N} = \{0, 1, 2, \dots\}$ such that all the symbols have their usual interpretation: $=$ is the equality relation; \leq is the ordering relation; $+$ and \cdot are the addition and multiplication operations; $S(x) = x + 1$; $\exp(x) = 2^x$.

Formulas in the above language are called *arithmetical*. The expressions $\forall x \leq t \varphi(x)$ and $\exists x \leq t \varphi(x)$ abbreviate the formulas $\forall x (x \leq t \rightarrow \varphi(x))$ and $\exists x (x \leq t \wedge \varphi(x))$, respectively, where t is any term (not containing the variable x). Occurrences of quantifiers of this kind are called *bounded*, and *elementary formulas* are those, all of whose quantifiers are bounded. Notice that, by definition, quantifier-free formulas are elementary.

Arithmetical formulas are classified according to their logical complexity into the *arithmetical hierarchy*. For $n \geq 0$, the classes of Σ_n - and Π_n -formulas are inductively defined as follows. Σ_0 - and Π_0 -formulas are the elementary formulas. Σ_{n+1} -formulas are those of the form $\exists x_1 \dots \exists x_m A(x_1, \dots, x_m)$, where A is a Π_n -formula. Π_{n+1} -formulas are those of the form $\forall x_1 \dots \forall x_m A(x_1, \dots, x_m)$, where A is a Σ_n -formula.

From the prenex normal form theorem we know that every arithmetical formula is logically equivalent to a Σ_n -formula, for some n . By extension of terminology, we shall sometimes call Σ_n any formula logically equivalent to a Σ_n -formula in the sense of our official definition.

From the computational point of view, the most interesting class of formulas is Σ_1 . A predicate on \mathbb{N} is definable by a Σ_1 -formula if and only if it is recursively enumerable (r.e.). Elementary predicates are decidable; a rough estimate of the complexity of the decision procedure yields an upper bound of order 2_n^x , where

n is some constant and x is the size of input. Here 2_n^x denotes the function

$$2_0^x = x; \quad 2_{n+1}^x = 2^{2_n^x}.$$

Functions of the form 2_n^x , for a fixed n , are called *multi-exponential*; thus, elementary predicates are decidable in multi-exponential number of steps. The converse is also true (see [35, 74]).

Peano Arithmetic **PA** is a first order theory with equality formulated in the arithmetical language which has, apart from the logical axioms and the axioms of equality, the following mathematical axioms:

$$\text{P1. } \neg S(a) = 0$$

$$\text{P2. } S(a) = S(b) \rightarrow a = b$$

$$\text{P3. } a + 0 = a$$

$$\text{P4. } a + S(b) = S(a + b)$$

$$\text{P5. } a \cdot 0 = 0$$

$$\text{P6. } a \cdot S(b) = a \cdot b + a$$

$$\text{P7. } \exp(0) = S(0)$$

$$\text{P8. } \exp(S(a)) = \exp(a) + \exp(a)$$

$$\text{P9. } a \leq 0 \leftrightarrow a = 0$$

$$\text{P10. } a \leq S(b) \leftrightarrow (a \leq b \vee a = S(b))$$

and the *induction axiom schema* for all arithmetical formulas $\varphi(x)$ (possibly containing parameters, that is, free variables besides x):

$$\text{IA: } \quad \varphi(0) \wedge \forall x (\varphi(x) \rightarrow \varphi(S(x))) \rightarrow \forall x \varphi(x).$$

Elementary Arithmetic **EA** is axiomatized over P1–P10 by the induction schema for bounded formulas $\varphi(x)$.

Elementary arithmetic is a sufficiently weak fragment of **PA**. It is well known that, similarly to **PA**, **EA** has an equivalent formulation in a language without the symbol of exponentiation. A theory $I\Delta_0$ is given in this language by the axioms P1–P6, P9, P10 together with the induction schema for all Δ_0 -formulas (i.e., bounded formulas in the arithmetical language without \exp). The graph of 2^x can be naturally defined by a certain Δ_0 -formula $\text{Exp}(x, y)$ ([43], c. 299). For this formula, the analogs of axioms P7, P8 and the monotonicity and functionality properties are provable in $I\Delta_0$. $\text{Exp}(x, y)$ defines an interpretation of **EA** in the theory $I\Delta_0 + \text{Exp}$, where **Exp** is the axiom $\forall x \exists y \text{Exp}(x, y)$ expressing the totality of \exp ([43], c. 37). In fact, **EA** is deductively equivalent to a conservative definitional extension of $I\Delta_0 + \text{Exp}$ by a symbol for \exp . It is not difficult to convince oneself that this interpretation preserves the classes of Σ_n - and Π_n -formulas, for $n \geq 1$, modulo derivability in $I\Delta_0 + \text{Exp}$. (However, the class

of elementary predicates strictly contains that of Δ_0 -predicates. Elementary predicates coincide with provably Δ_1 -predicates in $I\Delta_0 + \text{Exp}$.)

Unlike PA, the theories $I\Delta_0 + \text{Exp}$ and EA are finitely axiomatizable ([43], p. 366). It is easy to see that EA is axiomatizable by Π_1 -formulas, and $I\Delta_0 + \text{Exp}$ is axiomatizable by Π_2 -formulas of the arithmetical hierarchy (in a language without exp).

(*Kalmar*) *elementary functions* are the ones that can be obtained from $Z(x) \equiv 0$, S , $+$, \cdot , 2^x , cut-off subtraction

$$x \dot{-} y := \begin{cases} 0, & \text{if } x \leq y, \\ x - y, & \text{otherwise,} \end{cases}$$

and projection functions $I_i^n(x_1, \dots, x_n) = x_i$ by the operations of composition and bounded minimization [74]:

$$\mu i \leq z \ R(i, \vec{x}) := \begin{cases} y, & \text{if } y \leq z \text{ and } R(y, \vec{x}) \wedge \forall i < y \neg R(i, \vec{x}) \\ 0, & \text{if } \forall i \leq z \neg R(i, \vec{x}). \end{cases}$$

(Here R is a predicate of the form $g(i, \vec{x}) = 0$, for some previously defined function g .)

The class of all elementary functions is denoted \mathcal{E} . It coincides with the class of functions computable in multi-exponential time on a Turing machine (cf. [35]). \mathcal{E} is closed under the operations of bounded sum, bounded product, and bounded primitive recursion. Modulo composition and a sufficiently large set of initial functions, each of these operations is equivalent to bounded minimization (cf. [74], Chapter 5).

Symbols for all the elementary functions can be naturally introduced in a conservative definitional extension of EA. In this theory, every bounded formula of the extended language is equivalent to a quantifier-free one. Moreover, the extension itself can be axiomatized by a set of purely universal (or quantifier-free) formulas. On the other hand, every bounded formula of the extended language is equivalent to a bounded arithmetical formula, and the same is true for the other classes of the arithmetical hierarchy. In particular, the induction schema for bounded formulas of the extended language is provable in the definitional extension of EA. Thus, the formulations of EA in the standard and in the extended language are in a strong sense equivalent, and we shall often identify them. All the statements mentioned above are folklore; a quantifier-free formulation of EA is thoroughly treated, e.g., in [70].

1.2 Provability predicate and Gödel's theorems

The standard arithmetization of syntax of first order theories — carefully developed by Feferman in [37] for the case of PA or by Smoryński [87] for the case of primitive recursive arithmetic — works in EA without any substantial changes. De facto, all the applications of the operator of primitive recursion necessary to define the main syntactical concepts are multi-exponentially bounded and hence verifiable in EA (cf. also [70]).

Arithmetical formulas are naturally identified with words in a suitable finite alphabet. In turn, words can be naturally one-to-one encoded by numbers,

for example, by transforming them first into binary words and subsequently applying a binary decoding. (Notice that the latter is easily expressible in EA using the exponentiation function and the arithmetical operations.) Thus, we fix a standard global encoding — or a Gödel numbering — of all the expressions of the arithmetical language: variables, terms, formulas, etc. The code, or the *Gödel number*, of an expression τ will be denoted $\ulcorner \tau \urcorner$. It is easy to see that under such an encoding many natural syntactical relations, such as those given below, will be elementarily definable. Moreover, their simplest properties established by elementary induction are provable in EA (cf. [37]).

$x \in St$	“ x is a Gödel number of an arithmetical sentence”
$x \in \Pi_n$	“ x is a Gödel number of a Π_n -formula”
$LogAx(x)$	“ x is a Gödel number of a logical axiom”
$EqAx(x)$	“ x is a Gödel number of an equality axiom”
$MP(x, y, z)$	“formula z follows from formulas x and y by <i>modus ponens</i> rule”
$Gen(x, y)$	“formula y follows from formula x by generalization”

We also fix a one-to-one elementary coding of finite sequences of natural numbers. $\langle x_1, \dots, x_n \rangle$ denotes the code of a sequence x_1, \dots, x_n and is an elementary function of x_1, \dots, x_n , for every fixed n . $(x)_i$ denotes $i + 1$ -st element of the sequence encoded by x , and is an elementary function of x and i . $\langle \rangle$ denotes the code of the empty sequence. The following functions and relations are also elementary and naturally representable in EA:

$Seq(x)$	“ x is a code of a sequence”
$lh(x)$	length of a sequence x
$end(x)$	the last element of a sequence x
$x * y$	concatenation of sequences x and y

We assume that the coding of sequences is monotone in the sense that the code of a sequence increases when the sequence is getting longer or when any of its elements increases.

By a *theory* we shall mean a first order theory with equality. As a rule, when the opposite is not explicitly stated, we shall assume theories to be formulated in the arithmetical language and to contain the axioms of EA. A theory is called *elementary presented* if a bounded formula $Ax_T(x)$ is specified that is true if and only if x codes a (mathematical) axiom of T . All the usual theories such as EA, PA and the others encountered in this paper are elementary presented. Moreover, by a well-known *Craig’s trick* [37], any r.e. theory has an equivalent elementary presentation; therefore, elementary presentability of a theory is not an essentially restrictive condition. A theory T is *sound* if all theorems of T are true in the standard model of arithmetic. T is Σ_n -*sound* if all its theorems of arithmetical complexity Σ_n are true in the standard model. Theories T and U are called *deductively equivalent* if they have the same set of theorems (denoted $T \equiv U$). If T and U are elementary presented, then $T + U$ denotes a theory whose axioms are defined by the formula $Ax_T(x) \vee Ax_U(x)$.

With any elementary presented theory T one associates the predicate “ y codes a proof of a formula x in T ” expressed by an elementary formula $\text{Prf}_T(y, x)$. The formula $\text{Prf}_T(y, x)$ is constructed in the following way, literally formalizing Hilbert’s definition of derivation in first order logic:

$$\begin{aligned} \text{Prf}_T(y, x) \quad &:= \text{Seq}(y) \wedge \text{end}(y) = x \wedge \\ &\forall n < \text{lh}(y) [\text{LogAx}((y)_n) \vee \text{EqAx}((y)_n) \vee \text{Ax}_T((y)_n) \\ &\quad \vee \exists i, j < n \text{MP}((y)_i, (y)_j, (y)_n) \\ &\quad \vee \exists i < n \text{Gen}((y)_i, (y)_n)]. \end{aligned}$$

In other words, y codes a sequence of formulas ending with x such that every element y is either a logical axiom, or an equality axiom, or a mathematical axiom of T , or follows from the previous formulas by the rules of *modus ponens* or generalization.

Gödel’s provability formula $\Box_T(x)$ for T is defined as $\exists y \text{Prf}_T(y, x)$. We use the following standard abbreviations. Closed terms of the form $S(S(\dots S(0)\dots))$ (n times) are called *numerals* and denoted \bar{n} . We shall also omit the bar in those cases when n cannot be confused with a variable. If φ is a formula, we write $\Box_T\varphi$ instead of $\Box_T(\ulcorner\varphi\urcorner)$. The symbol \perp denotes logical falsity, and $\text{Con}(T) := \neg\Box_T\perp$ is the consistency assertion for T . The expression $\ulcorner\sigma(\dot{x}_1, \dots, \dot{x}_k)\urcorner$ denotes a natural definable term for the function mapping n_1, \dots, n_k to the Gödel number $\ulcorner\sigma(\bar{n}_1, \dots, \bar{n}_k)\urcorner$ of the result of substituting the numerals $\bar{n}_1, \dots, \bar{n}_k$ for the variables x_1, \dots, x_k in σ . We also write $\Box_T\sigma(\dot{x}_1, \dots, \dot{x}_n)$ for $\Box_T(\ulcorner\sigma(\dot{x}_1, \dots, \dot{x}_n)\urcorner)$.

Proposition 1.1 (Löb’s conditions). *For any sentences φ, ψ ,*

- L1. $T \vdash \varphi \implies \text{EA} \vdash \Box_T\varphi$
- L2. $\text{EA} \vdash \Box_T(\varphi \rightarrow \psi) \rightarrow (\Box_T\varphi \rightarrow \Box_T\psi)$
- L3. $\text{EA} \vdash \Box_T\varphi \rightarrow \Box_T\Box_T\varphi$

Similar statements also hold for formulas with free variables.

- L1°. $T \vdash \varphi(x) \implies \text{EA} \vdash \Box_T\varphi(\dot{x})$
- L2°. $\text{EA} \vdash \Box_T(\varphi(\dot{x}) \rightarrow \psi(\dot{x})) \rightarrow (\Box_T\varphi(\dot{x}) \rightarrow \Box_T\psi(\dot{x}))$
- L3°. $\text{EA} \vdash \Box_T\varphi(\dot{x}) \rightarrow \Box_T\Box_T\varphi(\dot{x})$
- L4°. $\text{EA} \vdash \Box_T\forall x\varphi(x) \rightarrow \forall x \Box_T\varphi(\dot{x})$

Notice that L4° follows from L1° and L2° by the axiom $\forall x\varphi(x) \rightarrow \varphi(x)$. Properties L3 and L3° are corollaries of a more general fact known as *provable Σ_1 -completeness* [43, 87]:

Proposition 1.2. (i) *For any Σ_1 -sentence σ , $\text{EA} \vdash \sigma \rightarrow \text{Prov}_T(\ulcorner\sigma\urcorner)$.*

- (ii) For any Σ_1 -formula $\sigma(x_1, \dots, x_n)$ with all the free variables explicitly shown,

$$\text{EA} \vdash \sigma(x_1, \dots, x_n) \rightarrow \text{Prov}_T(\ulcorner \sigma(\dot{x}_1, \dots, \dot{x}_n) \urcorner).$$

Σ_1 -completeness of T is formalizable in EA because of a multi-exponential upper bound in n on the size of an EA-proof of one of the formulas $\varphi(\bar{n})$ or $\neg\varphi(\bar{n})$, for any elementary formula $\varphi(x)$ (cf. [87] and [43], p. 175).

The following arithmetical *fixed point lemma* plays a key role in the proofs of Gödel's incompleteness theorems and many other results [37, 87].

Proposition 1.3. (i) For any formula $\varphi(x)$ there is a formula ψ , in which only the variables of φ except for x can occur, such that

$$\text{EA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

- (ii) For any formula $\varphi(x, x_1, \dots, x_n)$ there is a $\psi(x_1, \dots, x_n)$ such that

$$\text{EA} \vdash \psi(x_1, \dots, x_n) \leftrightarrow \varphi(\ulcorner \psi(\dot{x}_1, \dots, \dot{x}_n) \urcorner).$$

The following generalization of Gödel's famous second incompleteness theorem is known as Löb's theorem [55]. Because of its importance for the present paper we present a short proof.

Theorem 1 (Löb). For any elementary presented theory T containing EA and any sentence φ ,

$$T \vdash \Box_T \varphi \rightarrow \varphi \iff T \vdash \varphi.$$

Proof. The implication (\Leftarrow) is obvious. For a proof of the opposite implication, by the fixed point lemma we obtain a sentence ψ such that

$$\text{EA} \vdash \psi \leftrightarrow (\Box_T \psi \rightarrow \varphi).$$

Using Löb's derivability conditions and the fact that T contains EA we successively derive in T :

1. $\Box_T(\psi \rightarrow (\Box_T \psi \rightarrow \varphi))$
2. $\Box_T \psi \rightarrow \Box_T(\Box_T \psi \rightarrow \varphi)$
3. $\Box_T \psi \rightarrow (\Box_T \Box_T \psi \rightarrow \Box_T \varphi)$
4. $\Box_T \psi \rightarrow \Box_T \varphi$
5. $\Box_T \psi \rightarrow \varphi$, because $T \vdash \Box_T \varphi \rightarrow \varphi$
6. ψ
7. $\Box_T \psi$
8. φ . \square

Substituting \perp for φ we obtain Gödel's second incompleteness theorem as a corollary [42].

Theorem 2 (Gödel). *For any elementary presented theory T containing EA,*

- (i) *if T is consistent, then $T \not\vdash \text{Con}(T)$;*
- (ii) *if, in addition, T is Σ_1 -sound, then $T \not\vdash \neg \text{Con}(T)$.*

Remark 1.4. Löb's theorem can in turn be considered as a corollary of Gödel's second incompleteness theorem for $T + \neg\varphi$ provided the axioms of this theory are defined by the formula $\text{Ax}_{T+\neg\varphi}(x) := \text{Ax}_T(x) \vee x = \ulcorner \neg\varphi \urcorner$. For the corresponding provability predicate there holds

$$\text{EA} \vdash \text{Con}(T + \neg\varphi) \leftrightarrow \neg \Box_T \varphi,$$

by a formalization of the deduction theorem in EA (cf. [37]).

The following statement is, in fact, a formalization of Löb's theorem and can be inferred from the non-formalized variant of this theorem (cf. [31]).

Corollary 1.5. *For any sentence φ ,*

$$\text{EA} \vdash \Box_T(\Box_T \varphi \rightarrow \varphi) \rightarrow \Box_T \varphi.$$

Proof. Let ψ denote the formula $\Box_T(\Box_T \varphi \rightarrow \varphi) \rightarrow \Box_T \varphi$. Using Löb's conditions it is easy to prove that

$$\text{EA} \vdash \Box_T \psi \rightarrow \psi. \tag{1}$$

Hence, by Theorem 1, $T \vdash \psi$, and thus $\text{EA} \vdash \Box_T \psi$ and $\text{EA} \vdash \psi$ by (1).

We shall also mention a well-known theorem due to Rosser, generalizing Gödel's first incompleteness theorem [75].

Theorem 3 (Rosser). *For any consistent r.e. theory T containing EA, there is a Σ_1 -sentence φ such that $T \not\vdash \varphi$ and $T \not\vdash \neg\varphi$.*

Proof. Consider any elementary presentation of T . As the required formula φ take the following fixed point:

$$\text{EA} \vdash \varphi \leftrightarrow \exists y (\text{Prf}_T(y, \ulcorner \neg\varphi \urcorner) \wedge \forall z < y \neg \text{Prf}_T(z, \ulcorner \varphi \urcorner)).$$

Assume $T \vdash \neg\varphi$. If m is a code of a derivation of $\neg\varphi$, we obtain

$$\text{EA} \vdash \text{Prf}_T(\overline{m}, \ulcorner \neg\varphi \urcorner) \wedge \forall z < \overline{m} \neg \text{Prf}_T(z, \ulcorner \varphi \urcorner),$$

by Σ_1 -completeness, whence $T \vdash \varphi$ follows by the definition of φ and we have a contradiction in T . If, on the other hand, $T \vdash \varphi$ and n codes a proof of φ , we obtain by the definition of φ

$$T \vdash \exists y \leq \overline{n} \text{Prf}_T(y, \ulcorner \neg\varphi \urcorner).$$

Since T is consistent, $T \not\vdash \neg\varphi$, that is, the Σ_1 -formula

$$\forall y \leq \overline{n} \neg \text{Prf}_T(y, \ulcorner \neg\varphi \urcorner)$$

is true. Then it has to be provable in EA and we obtain a contradiction in T .

1.3 Provability logic

The basic Gödel–Löb provability **GL** is formulated in the language of propositional calculus endowed with a unary modality connective \Box . The expressions $\Diamond\varphi$, $\Box^+\varphi$ and $\Box^n\varphi$ abbreviate $\neg\Box\neg\varphi$, $\varphi \wedge \Box\varphi$ and $\Box\Box\ldots\Box\varphi$ (n times), respectively. **GL** is defined by the following axioms and inference rules:

- Axioms:** A1. Boolean tautologies;
 A2. $\Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$;
 A3. $\Box\varphi \rightarrow \Box\Box\varphi$;
 A4. $\Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi$.

Inference rules: *modus ponens* and $\varphi \vdash \Box\varphi$ (*necessitation*).

By an *arithmetical realization* of the language of **GL** we mean any function $(\cdot)^*$ that maps propositional variables to arithmetical sentences. For a modal formula φ , $(\varphi)_T^*$ denotes the result of substituting for all the variables of φ the corresponding arithmetical sentences and of translating \Box as the provability predicate \Box_T :

$$\begin{aligned} (\varphi \rightarrow \psi)_T^* &:= ((\varphi)_T^* \rightarrow (\psi)_T^*) \\ (\neg\varphi)_T^* &:= \neg(\varphi)_T^* \\ (\Box\varphi)_T^* &:= \Box_T(\varphi)_T^*. \end{aligned}$$

Under this interpretation, axioms A2, A3 and the necessitation rule can be seen to directly correspond to the three Löb’s derivability conditions, and axiom A4 is a formalization of Löb’s theorem (cf. Corollary 1.5). It follows that, for each modal formula φ , **GL** $\vdash \varphi$ implies **EA** $\vdash (\varphi)_T^*$, for every arithmetical realization $(\cdot)^*$ of the variables of φ . The opposite implication, for the case of a Σ_1 -sound theory T , is also valid; this is the content of the important *arithmetical completeness theorem* for **GL** due to Solovay (cf. [88, 31]).

Theorem 4 (Solovay). *Assume an elementary presented theory T is Σ_1 -sound and contains **EA**. Then*

$$\mathbf{GL} \vdash \varphi \iff T \vdash (\varphi)_T^*, \text{ for every realization } (\cdot)^* \text{ of the variables of } \varphi.$$

Kripke models are a convenient tool of study of provability logic. A *Kripke model* is a triple (W, R, \Vdash) , where

1. W is a finite nonempty set;
2. R is an irreflexive partial order on W ;
3. \Vdash is a binary relation of *forcing* between elements of W (*nodes*) and modal formulas such that

$$\begin{aligned} x \Vdash \neg\varphi &\iff x \not\Vdash \varphi, \\ x \Vdash (\varphi \rightarrow \psi) &\iff (x \not\Vdash \varphi \text{ or } x \Vdash \psi), \\ x \Vdash \Box\varphi &\iff \forall y \in W (xRy \Rightarrow y \Vdash \varphi). \end{aligned}$$

We write $\mathcal{W} \models \varphi$ if $\forall x \in W \ x \Vdash \varphi$. A proof of the following theorem, due to Segerberg [79], can also be found, e.g., in [87, 31].

Theorem 5. *For any formula φ ,*

$$\mathbf{GL} \vdash \varphi \iff \mathcal{W} \models \varphi, \text{ for every Kripke model } \mathcal{W}.$$

This theorem provides a useful (and effective) criterium of testing the provability of modal formulas in **GL**. Thus, **GL** is decidable and has the finite model property. It is also well known that **GL** enjoys the Craig interpolation property and admits a natural sequential cut-free formulation.

The study of provability logic and its generalizations is a rather large field, which goes beyond the scope of the present paper. This topic is covered in a number of surveys [36, 14, 54, 93] and monographs [29, 87, 31]. In the present paper, we shall make use of the completeness of **GL** w.r.t. Kripke models. However, we shall only rely on the *soundness* of **GL** w.r.t. the arithmetical interpretation, that is, we do not use the most difficult part of Solovay's theorem.

2 Reflection principles

Reflection principles for a given theory T are formal schemata expressing the soundness of T . We have to deal with the schemata because, by a well-known theorem due to Tarski, the property of *truth* of an arithmetical formula is not expressible in the arithmetical language. Hence, the statement “every provable sentence is true” cannot be formulated by a single arithmetical formula. Since the soundness of T implies its consistency, reflection principles for T are not expected to be provable in T itself. In this section we shall consider the main forms and basic properties of reflection principles.

2.1 Forms of reflection

Consider an elementary presented theory T . The two main forms of reflection principle are:

- *Local* reflection schema $\text{Rfn}(T)$:

$$\Box_T \varphi \rightarrow \varphi,$$

for all arithmetical sentences φ ;

- *Uniform* reflection schema $\text{RFN}(T)$:

$$\forall x_1, \dots, x_n (\Box_T \varphi(\dot{x}_1, \dots, \dot{x}_n) \rightarrow \varphi(x_1, \dots, x_n)),$$

for all arithmetical formulas $\varphi(x_1, \dots, x_n)$.

Obviously, for any theory T ,

$$\text{EA} + \text{RFN}(T) \vdash \text{Rfn}(T) \quad \text{и} \quad \text{EA} + \text{Rfn}(T) \vdash \text{Con}(T).$$

Hence, by Gödel's theorem neither of the two schemata is provable in T provided T is consistent. Löb's theorem actually tells us more: no nontrivial instances of the local, let alone the uniform, reflection schema can be provable in T .

Other natural forms of reflection are usually reducible to one of the main two. For example, the standard elementary coding of sequences allows one to convert several free variables into a single one; hence the schema

$$\forall x (\Box_T \varphi(\dot{x}) \rightarrow \varphi(x)),$$

for formulas $\varphi(x)$ with a single free variable x , is equivalent to $\text{RFN}(T)$ over EA . Feferman [38] proved

Proposition 2.1. *The following schemata and inference rules are equivalent over EA :*

- (a) $\text{RFN}(T)$;
- (b) $\forall x \Box_T \varphi(\dot{x}) \rightarrow \forall x \varphi(x)$;
- (c) $\frac{\forall x \Box_T \varphi(\dot{x})}{\forall x \varphi(x)}$ (Kleene's rule).

Proof. We shall use a simple lemma from [38], which is sometimes called *explicit reflection principle* because it asserts the provability of a reflection schema for the “standard” or “explicit” proofs.

Lemma 2.2. (i) *For any formula $\varphi(x)$ and any $m, n \in \mathbb{N}$,*

$$T \vdash \text{Prf}_T(\overline{m}, \ulcorner \varphi(\overline{n}) \urcorner) \rightarrow \varphi(\overline{n}).$$

(ii) *Statement (i) is formalizable in EA , that is,*

$$\text{EA} \vdash \forall x, y \Box_T (\text{Prf}_T(\dot{y}, \ulcorner \varphi(\dot{x}) \urcorner) \rightarrow \varphi(\dot{x})).$$

Proof. It is sufficient to prove (ii). On the one hand,

$$\begin{aligned} \text{EA} \vdash \text{Prf}_T(y, \ulcorner \varphi(\dot{x}) \urcorner) &\rightarrow \Box_T \varphi(\dot{x}) \\ &\rightarrow \Box_T (\text{Prf}_T(\dot{y}, \ulcorner \varphi(\dot{x}) \urcorner) \rightarrow \varphi(\dot{x})). \end{aligned}$$

On the other hand, by the provable Σ_1 -completeness

$$\begin{aligned} \text{EA} \vdash \neg \text{Prf}_T(y, \ulcorner \varphi(\dot{x}) \urcorner) &\rightarrow \Box_T \neg \text{Prf}_T(\dot{y}, \ulcorner \varphi(\dot{x}) \urcorner) \\ &\rightarrow \Box_T (\text{Prf}_T(\dot{y}, \ulcorner \varphi(\dot{x}) \urcorner) \rightarrow \varphi(\dot{x})). \end{aligned}$$

Combining these two things together we obtain the result.

Applying Kleene's rule (c) to the provable formula from part (ii) of the previous lemma we obtain a proof of the corresponding instance of $\text{RFN}(T)$. Therefore, (c) implies (a). The implications (a) \Rightarrow (b) and (b) \Rightarrow (c) are obvious.

Remark 2.3. Unlike Kleene's rule, the rule of local reflection, that is the rule

$$\frac{\Box_T \varphi}{\varphi}$$

for *sentences* φ , is admissible in T provided T is Σ_1 -sound. The admissibility of this rule is expressed by the formula

$$\forall x (\Box_T \Box_T(\dot{x}) \rightarrow \Box_T(x)),$$

which is an instance of the uniform reflection principle and therefore is not provable in T . Hence, the local reflection rule is admissible but not provably admissible in T . R. Parikh [64] has shown that adding this rule to T yields a substantial speed-up of some derivations in T , which exceeds any provably total computable function in T .

2.2 Partial reflection principles

Partial reflection principles are obtained from the local and the uniform schemata by imposing a restriction that φ may only range over a certain subclass Γ of the class of arithmetical sentences (formulas). Such schemata will be denoted $\text{Rfn}_\Gamma(T)$ and $\text{RFN}_\Gamma(T)$, respectively, and for Γ one usually takes one of the classes Σ_n or Π_n of the arithmetical hierarchy.

Lemma 2.4. *The following schemata are equivalent over EA, for each $n \geq 1$:*

- (i) $\text{RFN}_{\Sigma_n}(T) \equiv \text{RFN}_{\Pi_{n+1}}(T)$;
- (ii) $\text{RFN}_{\Pi_1}(T) \equiv \text{Rfn}_{\Pi_1}(T) \equiv \text{Con}(T)$.

Proof. (i) Consider a Π_{n+1} -formula $\forall y \varphi(y, x)$ where $\varphi(y, x) \in \Sigma_n$. Within $\text{EA} + \text{RFN}_{\Sigma_n}(T)$ one can derive using (L4 $^\circ$)

$$\begin{aligned} \Box_T \forall y \varphi(y, \dot{x}) &\rightarrow \forall y \Box_T \varphi(y, \dot{x}) \\ &\rightarrow \forall y \varphi(y, x). \end{aligned}$$

- (ii) Let $\varphi(x) \in \Pi_1$. We derive using provable Σ_1 -completeness in EA:

$$\begin{aligned} \Box_T \varphi(\dot{x}) \wedge \neg \varphi(x) &\rightarrow \Box_T \neg \varphi(\dot{x}) \\ &\rightarrow \Box_T (\varphi(\dot{x}) \wedge \neg \varphi(\dot{x})) \\ &\rightarrow \Box_T \perp \end{aligned}$$

Hence, $\neg \Box_T \perp$ implies $\Box_T \varphi(\dot{x}) \rightarrow \varphi(x)$.

Remark 2.5. For $n \geq 1$, the arithmetical complexity of all instances of the schema $\text{RFN}_{\Pi_n}(T)$ is Π_n .

Partial uniform reflection principles enjoy an important property of finite axiomatizability. It is well known that the classes of Σ_n - and Π_n -formulas have truth-definitions within EA.

Proposition 2.6. *For each n , there is an arithmetical Π_n -formula $\text{True}_{\Pi_n}(x)$ such that, for all Π_n -formulas $\varphi(x_1, \dots, x_n)$,*

$$\text{EA} \vdash \varphi(x_1, \dots, x_n) \leftrightarrow \text{True}_{\Pi_n}(\ulcorner \varphi(\dot{x}_1, \dots, \dot{x}_n) \urcorner).$$

Similarly for the classes Σ_n .

Formula $\text{True}_{\Pi_n}(x)$ is called a *truth-definition* for Π_n -formulas in EA.

Lemma 2.7. *Over EA, the schema $\text{RFN}_{\Pi_n}(T)$ is equivalent to its universal instance*

$$\forall z (\Box_T \text{True}_{\Pi_n}(\dot{z}) \rightarrow \text{True}_{\Pi_n}(z)). \quad (2.1)$$

A similar statement also holds for the class Σ_n .

Proof. To derive $\text{RFN}_{\Pi_n}(T)$ from formula (2.1), notice that Proposition 2.6 implies

$$\text{EA} \vdash \forall x_1, \dots, x_n \Box_T(\varphi(\dot{x}_1, \dots, \dot{x}_n) \leftrightarrow \text{True}_{\Pi_n}(\ulcorner \varphi(\dot{x}_1, \dots, \dot{x}_n) \urcorner)),$$

and hence one can infer

$$\begin{aligned} \Box_T \varphi(\dot{x}_1, \dots, \dot{x}_n) &\rightarrow \Box_T \text{True}_{\Pi_n}(\ulcorner \varphi(\dot{x}_1, \dots, \dot{x}_n) \urcorner) \\ &\rightarrow \text{True}_{\Pi_n}(\ulcorner \varphi(\dot{x}_1, \dots, \dot{x}_n) \urcorner), \quad (\text{by (2.1) with } z = \ulcorner \varphi(\dot{x}_1, \dots, \dot{x}_n) \urcorner) \\ &\rightarrow \varphi(x_1, \dots, x_n). \end{aligned}$$

Corollary 2.8. *The schemata $\text{RFN}_{\Pi_n}(T)$ and $\text{RFN}_{\Sigma_n}(T)$, for $n \geq 1$, are finitely axiomatizable over EA.*

2.3 n -provability and n -consistency

One can consider uniform partial reflection principles as analogs of Gödel's consistency assertion.

For $n \geq 1$, let $\text{Th}_{\Pi_n}(\mathbb{N})$ denote the set of all true arithmetical Π_n -sentences. A theory T is called *n -consistent*, if $T + \text{Th}_{\Pi_n}(\mathbb{N})$ is consistent. n -consistency of T can be expressed by the following formula, denoted $n\text{-Con}(T)$:

$$\forall z (\text{True}_{\Pi_n}(z) \rightarrow \neg \Box_T \neg \text{True}_{\Pi_n}(\dot{z})).$$

Informally, this sentence asserts the compatibility with T of any true Π_n -sentence. Modulo contraposition $n\text{-Con}(T)$ coincides with the universal instance of $\text{RFN}_{\Sigma_n}(T)$, that is, we obtain the following obvious corollary of Lemma 2.7.

Lemma 2.9. *For $n \geq 1$, $\text{RFN}_{\Sigma_n}(T)$ is equivalent to $n\text{-Con}(T)$ over EA.*

The dual *n -provability formula*

$$[n]_T \varphi := \neg n\text{-Con}(T + \neg \varphi)$$

expressing the provability of φ in $T + \text{Th}_{\Pi_n}(\mathbb{N})$ is naturally defined. Notice that $n\text{-Con}(T)$ has arithmetical complexity Π_{n+1} , and $[n]_T \varphi$ is, therefore, a

Σ_{n+1} -formula. We assume that for $n = 0$ the formulas of n -provability and n -consistency coincide with the usual formulas \Box_T and $\text{Con}(T)$. We shall also use $\langle n \rangle_T \varphi$ as an abbreviation for $n\text{-Con}(T + \varphi)$.

Properties of the n -provability formula are similar to those of the usual one. Firstly, $[n]_T$ satisfies Löb's derivability conditions.

Proposition 2.10.

L1ⁿ. $T \vdash \varphi \Rightarrow \text{EA} \vdash [n]_T \varphi$;

L2ⁿ. $\text{EA} \vdash [n]_T(\varphi \rightarrow \psi) \rightarrow ([n]_T \varphi \rightarrow [n]_T \psi)$;

L3ⁿ. $\text{EA} \vdash [n]_T \varphi \rightarrow [n]_T [n]_T \varphi$.

Similar statements also hold for formulas with free variables.

Proof. It is easy to check the first two conditions taking into account that

$$\text{EA} \vdash [n]_T \varphi \leftrightarrow \exists z (\text{True}_{\Pi_n}(z) \wedge \Box_T(\text{True}_{\Pi_n}(\dot{z}) \rightarrow \varphi).$$

The third condition follows from the more general fact of *provable Σ_{n+1} -completeness*.

Proposition 2.11. *For any Σ_{n+1} -formula $\sigma(x_1, \dots, x_k)$, where all free variables of σ are explicitly shown,*

$$\text{EA} \vdash \sigma(x_1, \dots, x_k) \rightarrow [n]_T \sigma(\dot{x}_1, \dots, \dot{x}_k).$$

Proof. Let $\sigma = \exists x \sigma_0(x)$ where $\sigma_0 \in \Pi_n$. Then we have, ignoring all the other free variables,

$$\text{EA} \vdash \sigma_0(x) \leftrightarrow \text{True}_{\Pi_n}(\ulcorner \sigma_0(\dot{x}) \urcorner)$$

and

$$\text{EA} \vdash \Box_T(\sigma_0(\dot{x}) \leftrightarrow \text{True}_{\Pi_n}(\ulcorner \sigma_0(\dot{x}) \urcorner)).$$

Therefore,

$$\text{EA} \vdash \sigma_0(x) \rightarrow \exists z (\text{True}_{\Pi_n}(z) \wedge \Box_T(\text{True}_{\Pi_n}(\dot{z}) \rightarrow \sigma_0(\dot{x}))),$$

where one can take the value of the term $\ulcorner \sigma_0(\dot{x}) \urcorner$ for z . Thus,

$$\begin{aligned} \text{EA} \vdash \sigma_0(x) &\rightarrow [n]_T \sigma_0(\dot{x}), \\ &\rightarrow [n]_T \exists x \sigma_0(x), \\ &\rightarrow [n]_T \sigma. \end{aligned}$$

Hence, $\text{EA} \vdash \exists x \sigma_0(x) \rightarrow [n]_T \sigma$.

As a standard corollary of Löb's conditions and the fixed point lemma we obtain an analog of the formalized Löb's theorem.

Proposition 2.12. *For any formula φ ,*

$$\text{EA} \vdash [n]_T([n]_T \varphi \rightarrow \varphi) \rightarrow [n]_T \varphi.$$

One can also extend to the n -provability formula the notion of provability logic. Consider the language of propositional logic with a modality $[n]$ and let \mathbf{GL}^n denote a variant of \mathbf{GL} formulated in this language. An arithmetical interpretation $(\varphi)_T^*$ of a modal formula φ is defined as usual, except that now $[n]$ is understood as n -provability in T :

$$([n]\varphi)_T^* = [n]_T(\varphi)_T^*.$$

Propositions 2.10 and 2.12 yield

Corollary 2.13. *If $\mathbf{GL}^n \vdash \varphi$ then $\mathbf{EA} \vdash (\varphi)_T^*$, for any arithmetical realization $(\cdot)^*$ of the variables of φ .*

In fact, C. Smoryński [87] has shown that Solovay's theorem can be generalized to the n -provability logic without any substantial changes.

Proposition 2.14. *Suppose an elementary presented theory T is Σ_{n+1} -sound and contains \mathbf{EA} . Then*

$$\mathbf{GL}^n \vdash \varphi \iff T \vdash (\varphi)_T^*, \text{ for any realization } (\cdot)^* \text{ of the variables of } \varphi.$$

Remark 2.15. The notion of n -consistency plays an important role in this paper. We remark that Gödel has considered a related, but not equivalent, notion of ω -consistency. A theory T is ω -inconsistent if there is a formula $\varphi(x)$ such that simultaneously

- (i) $T \vdash \exists x \varphi(x)$;
- (ii) $T \vdash \neg \varphi(0), \neg \varphi(1), \neg \varphi(2), \dots$

Otherwise, T is called ω -consistent.

In the modern expositions of Gödel's theorems the condition of ω -consistency is often replaced by a weaker condition of Σ_1 -soundness. C. Smoryński [84] showed that the formula naturally expressing ω -consistency of T is equivalent to $\mathbf{RFN}_{\Pi_3}(T + \mathbf{RFN}(T))$. Hence, in this sense the ω -consistency formula is reducible to the more convenient reflection principles.

2.4 Unboundedness theorems

Many applications of reflection principles to the analysis of formal theories are based on the so-called *unboundedness theorems* due to Kreisel and Lévy.

We say that a theory U is a *finite extension* of a theory T if it can be axiomatized over T by finitely many sentences. U is called an *extension of T of complexity Γ* , or a Γ -extension of T if U is axiomatizable over T by a set of sentences from Γ . For Γ one usually takes one of the classes of the arithmetical hierarchy. (Notice that the set of Gödel numbers of the axioms of U is not supposed to be r.e. or even arithmetical. We are talking here about the complexity of formulas axiomatizing U .) A theory U is called Γ -conservative over T if, for any sentence $\varphi \in \Gamma$, $U \vdash \varphi$ implies $T \vdash \varphi$ (denoted $U \subseteq_{\Gamma} T$).

Proposition 2.16. *Let T be an elementary presented theory.*

- (i) $\text{Rfn}_{\Pi_n}(T)$ is not contained in any consistent finite extension of T of complexity Σ_n (and dually for $\text{Rfn}_{\Sigma_n}(T)$).
- (ii) $\text{RFN}_{\Pi_n}(T)$ is not contained in any consistent extension of T of complexity Σ_n .

Proof. (i) Assume $T + \varphi \vdash \text{Rfn}_{\Pi_n}(T)$ where $\varphi \in \Sigma_n$. Then, in particular,

$$T + \varphi \vdash \Box_T \neg \varphi \rightarrow \neg \varphi.$$

Therefore,

$$\begin{aligned} T &\vdash \varphi \rightarrow (\Box_T \neg \varphi \rightarrow \neg \varphi), \\ &\vdash \Box_T \neg \varphi \rightarrow \neg \varphi. \end{aligned}$$

Löb's Theorem then implies $T \vdash \neg \varphi$, i.e., $T + \varphi$ is inconsistent.

(ii) Assume that U is a Σ_n -extension of T and $U \vdash \text{RFN}_{\Pi_n}(T)$. Since $\text{RFN}_{\Pi_n}(T)$ is finitely axiomatizable, there exists a finite subtheory $U_0 \subseteq U$ such that $U_0 \vdash \text{RFN}_{\Pi_n}(T)$. Then U_0 also contains $\text{Rfn}_{\Pi_n}(T)$ and by (i) $T + U_0$ has to be inconsistent; hence so is U .

Corollary 2.17. *Let T be an elementary presented theory.*

- (i) $\text{Rfn}(T)$ is not contained in any consistent finite extension of T .
- (ii) $\text{RFN}(T)$ is not contained in any consistent extension of T of bounded arithmetical complexity.

Example 2.18. By a theorem of Kreisel and Lévy, Peano arithmetic PA proves $\text{RFN}(\text{EA})$. Hence PA is neither finitely axiomatizable, nor is contained in any consistent extension of EA of bounded arithmetical complexity.

Corollary 2.19. *If a theory T is Σ_1 -sound, then*

$$T + \text{Rfn}(T) \not\vdash \text{RFN}(T).$$

Proof. Let U be the extension of T by the set of all true Π_1 -sentences. (U is not r.e. but is a Π_1 -extension of T .) Since T does not prove false Σ_1 sentences, U is consistent. By the previous corollary $U \not\vdash \text{RFN}(T)$. On the other hand, $U \vdash \text{Rfn}(T)$ because

$$\begin{aligned} T \vdash \varphi &\Rightarrow T \vdash \Box_T \varphi \rightarrow \varphi \quad \text{and} \\ T \not\vdash \varphi &\Rightarrow \mathbb{N} \models \neg \Box_T \varphi \Rightarrow \\ U \vdash \neg \Box_T \varphi &\Rightarrow U \vdash \Box_T \varphi \rightarrow \varphi. \end{aligned}$$

Remark 2.20. The Σ_1 -soundness condition in this well-known statement can be weakened, however this requires some deeper considerations (cf. Corollary 2.38 below).

Part (i) of Proposition 2.16 and of Corollary 2.17 has an important strengthening. Namely, the word ‘finite’ there can be replaced by ‘recursively enumerable.’ This follows from the following result.

Lemma 2.21. *Let U be a consistent r.e. extension of T of complexity Σ_n . Then there is a single Σ_n -sentence φ such that $T + \varphi$ is consistent and contains U . (Dually for Π_n -extensions.)*

Proof. For the theories containing the so-called Σ_1 -collection schema, in particular for PA, the result directly follows from Theorem 4 of Lindström [53]. However, for weaker theories a modification of this construction is needed, which is presented below.

By Craig’s trick one can assume the set of Gödel numbers of Σ_n -axioms of U over T to be given by an elementary formula $\text{Ax}_U(x)$. Besides, a truth-definition for Σ_n -formulas has the form

$$\text{EA} \vdash \text{True}_{\Sigma_n}(x) \leftrightarrow \exists z \text{ Wit}_{\Sigma_n}(z, x),$$

for some formula $\text{Wit}_{\Sigma_n}(z, x) \in \Pi_{n-1}$. Consider a fixed point φ such that

$$\text{EA} \vdash \varphi \leftrightarrow \exists x \exists z (\forall y \leq x (\text{Ax}_U(y) \rightarrow \text{Wit}_{\Sigma_n}((z)_y, y)) \wedge \text{Prf}_T(x, \ulcorner \neg \varphi \urcorner)).$$

We claim that φ satisfies the conditions of lemma.

1) The theory $T + \varphi$ is consistent. Indeed, otherwise there is an m such that

$$\text{EA} \vdash \text{Prf}_T(\overline{m}, \ulcorner \neg \varphi \urcorner).$$

On the other hand, for each axiom $\sigma \in \Sigma_n$ of U such that $\ulcorner \sigma \urcorner = y \leq m$ we have $U \vdash \exists z_y \text{ Wit}_{\Sigma_n}(z_y, \overline{y})$. Since for a fixed m there are only finitely many such σ ’s, all the corresponding z_y can be collected into one sequence $z := \langle z_0, \dots, z_m \rangle$ where $z_y := 0$ for all y which are not the codes of axioms of U . Thus,

$$U \vdash \exists z \bigwedge_{y \leq m, y \in \text{Ax}_U} \text{Wit}_{\Sigma_n}((z)_{\overline{y}}, \overline{y}),$$

whence

$$U \vdash \exists z \forall y \leq \overline{m} (\text{Ax}_U(y) \rightarrow \text{Wit}_{\Sigma_n}((z)_y, y)).$$

It follows that $U \vdash \varphi$ and $U \vdash \perp$, since we have assumed $T + \varphi$ inconsistent.

2) $T + \varphi$ contains U . Consider any formula $\sigma \in \text{Ax}_U$, $\ulcorner \sigma \urcorner = m$. First of all, notice that by the explicit reflection principle

$$T + \varphi \vdash \forall x \leq \overline{m} \neg \text{Prf}_T(x, \ulcorner \neg \varphi \urcorner).$$

Reasoning in $T + \varphi$ from this we can derive

$$\begin{aligned} \vdash \forall y \leq x (\text{Ax}_U(x) \rightarrow \text{Wit}_{\Sigma_n}((z)_y, y)) \wedge \text{Prf}_T(x, \ulcorner \neg \varphi \urcorner) &\rightarrow \overline{m} \leq x \wedge \text{Ax}_U(\overline{m}) \\ &\rightarrow \text{Wit}_{\Sigma_n}((z)_{\overline{m}}, \overline{m}) \\ &\rightarrow \text{True}_{\Sigma_n}(\overline{m}) \\ &\rightarrow \sigma. \end{aligned}$$

It follows that $T + \varphi \vdash \sigma$.

Corollary 2.22. (i) $\text{Rfn}_{\Pi_n}(T)$ is not contained in any consistent r.e. extension of T of complexity Σ_n (and dually for $\text{Rfn}_{\Sigma_n}(T)$).

(ii) $\text{Rfn}(T)$ is not contained in any consistent r.e. extension of T of bounded arithmetical complexity.

2.5 Hierarchies of partial reflection principles

In this section we consider in some detail the hierarchies of partial reflection principles. We show that metamathematical properties of local reflection principles sharply contrast with the well-known properties of uniform reflection principles. We recall the following facts from §2.1.

1) Partial uniform reflection schemata, for a given elementary presented theory T , form a linearly ordered hierarchy by inclusion (cf. Fig. 1).

2) Except for the full $\text{RFN}(T)$, all the schemata in this hierarchy are finitely axiomatizable (Corollary 2.8).

The following lemma demonstrates that, in a sense, $\text{RFN}_{\Pi_{n+1}}(T)$ is much stronger than $\text{RFN}_{\Pi_n}(T)$.

Lemma 2.23. For any finite Σ_n -extension U of a theory T , $\text{RFN}_{\Pi_n}(T)$ is deductively equivalent to $\text{RFN}_{\Pi_n}(U)$ over U .

Proof. It is sufficient to derive $\text{RFN}_{\Pi_n}(U)$ in $U + \text{RFN}_{\Pi_n}(T)$. By the formalized deduction theorem, for every Π_n -formula $\varphi(x)$ it is provable in $\text{EA} + \text{RFN}_{\Pi_n}(T)$ that

$$\begin{aligned} \vdash \Box_U \varphi(\dot{x}) &\rightarrow \Box_T(\psi \rightarrow \varphi(\dot{x})) \\ &\rightarrow (\psi \rightarrow \varphi(x)), \end{aligned}$$

where ψ denotes the conjunction of all Σ_n -axioms of U . Since ψ is provable in U , we obtain

$$U + \text{RFN}_{\Pi_n}(T) \vdash \Box_U \varphi(\dot{x}) \rightarrow \varphi(x).$$

Corollary 2.24. $T + \text{RFN}_{\Pi_{n+1}}(T)$ proves $\text{Con}(T + \text{RFN}_{\Pi_n}(T))$.

Thus, the hierarchy of partial uniform reflection principles does not collapse (i.e., all the theories of the hierarchy are different) if and only if $T + \text{RFN}(T)$ is consistent. Furthermore, if the hierarchy is proper, none of the extensions at any level of the hierarchy is conservative even for Π_1 -sentences.

For the hierarchy of local reflection principles the picture essentially changes. First of all, we have the following proposition.

Proposition 2.25. All the schemata $\text{Rfn}_{\Sigma_n}(T)$ and $\text{Rfn}_{\Pi_n}(T)$ are pairwise distinct provided the theory $T + \text{Rfn}(T)$ is consistent. In particular, $\text{Rfn}_{\Sigma_n}(T)$ is strictly contained in $\text{Rfn}_{\Pi_{n+1}}(T)$, and none of $\text{Rfn}_{\Sigma_n}(T)$ and $\text{Rfn}_{\Pi_n}(T)$, for any $n > 1$, implies the other (see Figure 1).

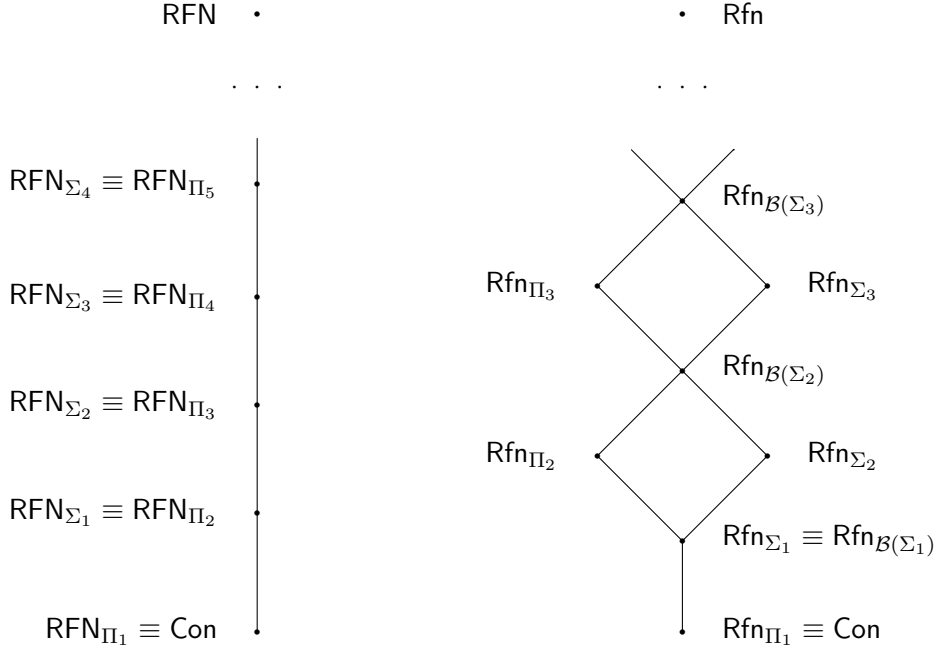


Figure 1: Hierarchies of partial reflection principles

Proof. By the first part of Proposition 2.16, $\text{Rfn}_{\Pi_1}(T)$ is strictly weaker than $\text{Rfn}_{\Sigma_1}(T)$. Hence, it will be sufficient to only prove the second statement of this proposition.

Notice that, for $n > 1$, $\text{Rfn}_{\Sigma_n}(T)$ is an r.e. (in fact, an elementary) set of Σ_n -sentences. So, by Corollary 2.22,

$$T + \text{Rfn}_{\Sigma_n}(T) \not\vdash \text{Rfn}_{\Pi_n}(T).$$

The statement that

$$T + \text{Rfn}_{\Pi_n}(T) \not\vdash \text{Rfn}_{\Sigma_n}(T)$$

is proved symmetrically.

Let $\mathcal{B}(\Sigma_n)$ denote the set of all boolean combinations of Σ_n -sentences. We have the following useful lemma.

Lemma 2.26. *$\text{Rfn}_{\Sigma_1}(T)$ is equivalent to $\text{Rfn}_{\mathcal{B}(\Sigma_1)}(T)$ over EA.*

Proof. We prove that

$$\text{EA} + \text{Rfn}_{\Sigma_1}(T) \vdash \Box_T \varphi \rightarrow \varphi,$$

for any boolean combination of Σ_1 -sentences φ . φ is logically equivalent to a formula of the form $\bigwedge_{i=1}^n (\pi_i \vee \sigma_i)$, for some sentences $\pi_i \in \Pi_1$ and $\sigma_i \in \Sigma_1$. Since \Box_T commutes with conjunction, it is sufficient to derive in $\text{EA} + \text{Rfn}_{\Sigma_1}(T)$ the formulas

$$\text{Prov}_T(\ulcorner (\pi_i \vee \sigma_i) \urcorner) \rightarrow (\pi_i \vee \sigma_i),$$

for any i . Using provable Σ_1 -completeness and Löb's derivability conditions, we obtain

$$\begin{aligned} \vdash \text{Prov}_T(\ulcorner (\pi_i \vee \sigma_i) \urcorner) \wedge \neg \pi_i &\rightarrow \text{Prov}_T(\ulcorner \neg \pi_i \urcorner), \\ &\rightarrow \text{Prov}_T(\ulcorner \sigma_i \urcorner), \\ &\rightarrow \sigma_i, \end{aligned}$$

by $\text{Rfn}_{\Sigma_1}(T)$. Hence,

$$\text{EA} + \text{Rfn}_{\Sigma_1}(T) \vdash \text{Prov}_T(\ulcorner (\pi_i \vee \sigma_i) \urcorner) \rightarrow (\pi_i \vee \sigma_i).$$

We note that in view of Proposition 2.25 a straightforward generalization of this statement to the classes Σ_n , for $n > 1$, does not hold. (A reasonable generalization using relativized reflection principles will be given in Section 4.4.)

Further, for local reflection principles we have the following results.

1) None of the schemata $\text{Rfn}_{\Sigma_n}(T)$ and $\text{Rfn}_{\Pi_{n+1}}(T)$, for $n \geq 1$, is finitely axiomatizable over T . (The only exception is $\text{Rfn}_{\Pi_1}(T)$, which is equivalent to $\text{Con}(T)$.)

2) All the restricted local reflection schemata, with an obvious exception of $\text{Rfn}_{\Pi_1}(T)$, are mutually Π_1 -conservative over each other, i.e., they prove the same Π_1 -sentences. In particular, $T + \text{Rfn}_{\Pi_{n+1}}(T)$ does not prove the consistency of $T + \text{Rfn}_{\Pi_n}(T)$, for $n > 1$ [5].

3) Actually, an even stronger result holds [21]: the full local reflection schema $\text{Rfn}(T)$ is Π_n - (respectively, Σ_n -) conservative over $T + \text{Rfn}_{\Pi_n}(T)$ (respectively, over $T + \text{Rfn}_{\Sigma_n}(T)$). This conservation result is optimal w.r.t. the arithmetical complexity in the sense that Σ_n - (respectively, Π_n -) sentences are already not conserved.

The proofs of these facts are beautiful applications of the techniques of provability logic. The first results of this type were found by Boolos [28] and Artemov [3] who independently proved, using provability logic methods, that n times iterated consistency assertion for a sound theory T cannot be derived from less than n instances of local reflection principle. We shall obtain the results mentioned above using a more general construction found in [21]. These results will then be applied to the study of parameter-free induction schemata in arithmetic.

Lemma 2.27. *Let modal formulas Q_i be defined by*

$$Q_1 := p, \quad Q_{i+1} := Q_i \vee \Box Q_i,$$

where p is a propositional variable. Then, for any variables p_1, \dots, p_m ,

$$\mathbf{GL} \vdash \Box^+(\bigwedge_{i=1}^m (\Box p_i \rightarrow p_i) \rightarrow p) \rightarrow (\bigwedge_{i=1}^m (\Box Q_i \rightarrow Q_i) \rightarrow p). \quad (2.2)$$

Proof. Rather than exhibiting an explicit proof of the formula (2.2) we shall argue semantically using the standard Kripke model characterization of \mathbf{GL} .

Consider any Kripke model (W, R, \Vdash) in which the conclusion $(\bigwedge_{i=1}^m (\Box Q_i \rightarrow Q_i) \rightarrow p)$ is false at some node $x \in W$. This means that $x \not\Vdash p$ and $x \Vdash \Box Q_i \rightarrow Q_i$, for each $i \leq m$. An obvious induction on i then shows that $x \not\Vdash Q_i$, for all $i \leq m+1$, in particular, $x \not\Vdash Q_{m+1}$.

Unwinding the definition of Q_i we observe that in W there is a sequence of nodes

$$x = x_{m+1} R x_m R \dots R x_1$$

such that, for all $i \leq m+1$, $x_i \not\Vdash Q_i$. Since R is irreflexive and transitive, all x_i 's are pairwise distinct. Moreover, it is easy to see by induction on i that, for all i ,

$$\mathbf{GL} \vdash p \rightarrow Q_i.$$

Hence, for each $i \leq m+1$, $x_i \not\Vdash p$.

Now we notice that each formula $\Box p_i \rightarrow p_i$ can be false at no more than one node of the chain x_{m+1}, \dots, x_1 . Therefore, by the pigeon-hole principle, there must exist a node z among the $m+1$ nodes x_i such that

$$z \Vdash \bigwedge_{i=1}^m (\Box p_i \rightarrow p_i) \wedge \neg p.$$

In case z coincides with $x = x_{m+1}$ we have

$$x \not\Vdash \bigwedge_{i=1}^m (\Box p_i \rightarrow p_i) \rightarrow p.$$

In case $z = x_i$, for some $i \leq m$, we have $x R z$ by transitivity of R , and thus

$$x \not\Vdash \Box \left(\bigwedge_{i=1}^m (\Box p_i \rightarrow p_i) \rightarrow p \right).$$

This shows that the formula (2.2) is forced at every node of any Kripke model; hence it is provable in \mathbf{GL} .

Corollary 2.28. $\mathbf{GL} \vdash \Box^+ \neg \bigwedge_{i=1}^m (\Box p_i \rightarrow p_i) \rightarrow \Box^m \perp$.

Proof. Substituting \perp for p in (2.2) observe that

$$\mathbf{GL} \vdash Q_i(p/\perp) \leftrightarrow \Box^{i-1} \perp,$$

and therefore

$$\mathbf{GL} \vdash \bigwedge_{i=1}^m (\Box Q_i(p/\perp) \rightarrow Q_i(p/\perp)) \leftrightarrow \neg \Box^m \perp.$$

A progression of theories based on *iteration of consistency assertion* is defined as follows.

$$T_0 := T, \quad T_{n+1} := T_n + \mathbf{Con}(T_n), \quad T_\omega := \bigcup_{n \geq 0} T_n.$$

Lemma 2.29. *Over EA, $\text{Con}(T_n)$ is provably equivalent to $\neg \Box_T^{n+1} \perp$.*

Proof. By induction on n using Löb's derivability conditions.

The formula $\text{Con}(T_n)$ is called *n times iterated consistency assertion for T* . By Gödel's theorem, the theories T_n form a strictly increasing sequence of Σ_1 -sound extensions of T , whenever the initial theory T is Σ_1 -sound.

The following statement is obtained in [28, 4].

Proposition 2.30. *$\text{Con}(T_n)$ is not derivable from any n instances of local reflection schema for a theory T provided T_n is consistent.*

Proof. Assume

$$T \vdash \bigwedge_{i=1}^n (\Box_T \varphi_i \rightarrow \varphi_i) \rightarrow \text{Con}(T_n).$$

Then, by Lemma 2.29 and contraposition,

$$T \vdash \Box_T^{n+1} \perp \rightarrow \neg \bigwedge_{i=1}^n (\Box_T \varphi_i \rightarrow \varphi_i),$$

and by Löb's derivability conditions,

$$\text{EA} \vdash \Box_T^{n+2} \perp \rightarrow \Box_T \neg \bigwedge_{i=1}^n (\Box_T \varphi_i \rightarrow \varphi_i).$$

Considering an arithmetical realization mapping each variable p_i to φ_i , from Corollary 2.28 we obtain

$$\text{EA} \vdash \Box_T^{n+2} \perp \rightarrow \Box_T^{n+1} \perp,$$

whence, by Löb's theorem, $\text{EA} \vdash \Box_T^{n+1} \perp$. It follows that T_n is inconsistent.

Remark 2.31. Formulas $\Box_T^{i+1} \perp \rightarrow \Box_T^i \perp$, for $i \leq n$, provide an example of $n+1$ instances of local reflection schema (in fact, instances of $\text{Rfn}_{\Sigma_1}(T)$) which, taken together, imply $\text{Con}(T_n)$ over EA. Notice that this also proves the inclusion $T_\omega \subseteq T + \text{Rfn}_{\Sigma_1}(T)$.

Proposition 2.32. *For $n \geq 1$, none of the schemata $\text{Rfn}_{\Sigma_n}(T)$ and $\text{Rfn}_{\Pi_{n+1}}(T)$ is finitely axiomatizable over T provided T_ω is consistent.*

Proof. By Proposition 2.30, any m instances of these schemata are insufficient to prove $\text{Con}(T_m)$. On the other hand, $\text{Con}(T_m)$ is provable in T_ω ; hence, it follows from any of the schemata $\text{Rfn}_{\Sigma_n}(T)$ and $\text{Rfn}_{\Pi_{n+1}}(T)$.

The following result due to Goryachev [5] is another expression of the same idea.

Proposition 2.33. *Any n instances of the local reflection principle for T are Π_1 -conservative over T_n .*

Proof. Assume $\pi \in \Pi_1$ and

$$T \vdash \bigwedge_{i=1}^n (\Box_T \varphi_i \rightarrow \varphi_i) \rightarrow \pi.$$

Then, by Löb's conditions,

$$T \vdash \Box_T \neg \pi \rightarrow \Box_T \neg \bigwedge_{i=1}^n (\Box_T \varphi_i \rightarrow \varphi_i),$$

and

$$T \vdash \Box_T^+ \neg \pi \rightarrow \Box_T^+ \neg \bigwedge_{i=1}^n (\Box_T \varphi_i \rightarrow \varphi_i).$$

By arithmetical interpretation, from Corollary 2.28 we obtain

$$T \vdash \Box_T^+ \neg \pi \rightarrow \Box_T^n \perp,$$

whence

$$T \vdash \neg \pi \rightarrow \Box_T^n \perp,$$

by provable Σ_1 -completeness. Thus, $T \vdash \neg \Box_T^n \perp \rightarrow \pi$ and $T_n \vdash \pi$.

Corollary 2.34. *For any elementary presented theory T , $T + \text{Rfn}(T)$ and T_ω prove the same Π_1 -sentences.*

It also yields the following corollary.

Corollary 2.35. *$T + \text{Rfn}(T)$ is consistent if and only if so is T_ω .*

Obviously, $T + \text{Rfn}_{\Sigma_n}(T)$ contains T_ω , for each n . Hence, in contrast with the uniform reflection principles, all such schemata prove the same Π_1 -sentences. Actually, for local reflection principles an even stronger result holds [21].

Theorem 6. *For $n > 1$, the following conservation results hold over T :*

- (i) $\text{Rfn}(T) \subseteq_{\Pi_n} \text{Rfn}_{\Pi_n}(T)$;
- (ii) $\text{Rfn}(T) \subseteq_{\Sigma_n} \text{Rfn}_{\Sigma_n}(T)$;
- (iii) $\text{Rfn}(T) \subseteq_{\mathcal{B}(\Sigma_1)} \text{Rfn}_{\Sigma_1}(T)$.

Proof. Assume

$$T \vdash \bigwedge_{i=1}^m (\Box_T \varphi_i \rightarrow \varphi_i) \rightarrow \pi.$$

By Löb's conditions, we also have

$$T \vdash \Box_T \left(\bigwedge_{i=1}^m (\Box_T \varphi_i \rightarrow \varphi_i) \rightarrow \pi \right).$$

Considering an arithmetical realization $(\cdot)^*$ mapping p to the sentence π and p_i to φ_i , for all i , by Lemma 2.27 we conclude that

$$T \vdash \Box_T \left(\bigwedge_{i=1}^m (\Box_T \psi_i \rightarrow \psi_i) \rightarrow \pi \right),$$

where $\psi_i := (Q_i)_T^*$. Now for $n > 1$ we observe that, if $\pi \in \Pi_n$, then $\psi_i \in \Pi_n$, for all i . Hence,

$$T + \text{Rfn}_{\Pi_n}(T) \vdash \pi.$$

Similarly, Σ_n -consequences of $\text{Rfn}(T)$ follow from $\text{Rfn}_{\Sigma_n}(T)$, and $\mathcal{B}(\Sigma_1)$ -consequences of $\text{Rfn}(T)$ follow from $\text{Rfn}_{\mathcal{B}(\Sigma_1)}(T)$, which is equivalent to $\text{Rfn}_{\Sigma_1}(T)$, by Lemma 2.26.

The following lemma shows that the above theorem is optimal w.r.t. the arithmetical complexity.

Lemma 2.36. *For $n > 1$, there is a Σ_n - (respectively, Π_n -) sentence provable in $T + \text{Rfn}(T)$ but not in $T + \text{Rfn}_{\Pi_n}(T)$ (respectively, $T + \text{Rfn}_{\Sigma_n}(T)$) provided T_ω is consistent.*

Proof. By Proposition 2.25, for such a formula one can take an appropriate instance of $\text{Rfn}_{\Sigma_n}(T)$ (respectively, of $\text{Rfn}_{\Pi_n}(T)$).

Remark 2.37. At present, the question of optimality of Statement (iii) of the previous theorem seems to be open. In particular, it is unknown if $\text{Rfn}(T)$ has to be Π_2 -conservative over $T + \text{Rfn}_{\Sigma_1}(T)$.

Finally, we notice an interesting strengthening of Corollary 2.19.

Corollary 2.38. *$T + \text{Rfn}(T) \not\vdash \text{RFN}(T)$ if and only if T_ω is consistent.*

Proof. By Corollary 2.35, if T_ω is inconsistent, then so is $T + \text{Rfn}(T)$, that is, $T + \text{Rfn}(T) \vdash \text{RFN}(T)$. On the other hand, if $T + \text{Rfn}(T) \vdash \text{RFN}(T)$, then using finite axiomatizability of $\text{RFN}_{\Sigma_1}(T)$ one finds n instances of $\text{Rfn}(T)$ which imply $\text{RFN}_{\Sigma_1}(T)$. However, $T + \text{RFN}_{\Sigma_1}(T)$ contains T_ω , and therefore $\text{Con}(T_n)$. Thus, by Proposition 2.30, T_n and T_ω are inconsistent.

3 Provably total computable functions and fragments of PA

The concept of provably total computable function² was introduced by Georg Kreisel in the 50's. This notion currently plays a very important role in proof theory in connection with the questions of ordinal analysis and constructive bounds, as well as with the more applied questions such as the extraction of programs from formal proofs. It is also tightly related with the construction of independent combinatorial statements for Peano arithmetic and its extensions.

²Sometimes an earlier but somewhat misleading term for this notion is used — *provably recursive function*.

Definition 3.1. A function $f : \mathbb{N}^k \rightarrow \mathbb{N}$ is called *provably total computable* in a theory T if there is a Σ_1 -formula $\varphi(\vec{x}, y)$ such that

- (i) $f(\vec{x}) = y \iff \mathbb{N} \models \varphi(\vec{x}, y)$;
- (ii) $T \vdash \forall \vec{x} \exists! y \varphi(\vec{x}, y)$.

The class of all provably total computable functions of a theory T is denoted $\mathcal{F}(T)$.

Notice that Condition (i) is equivalent to the computability of f . For theories T containing **EA**, $\mathcal{F}(T)$ contains all elementary functions \mathcal{E} and is closed under composition. In general, this class is not closed under bounded minimization (cf. [23]). We also note that $\mathcal{F}(T)$ only depends on Π_2 -consequences of T ; hence, if $T \subseteq_{\Pi_2} U$, then $\mathcal{F}(T) \subseteq \mathcal{F}(U)$.

Descriptions of the classes $\mathcal{F}(T)$ in recursion-theoretic terms were obtained for many natural theories T . For example, by a well-known theorem due to Parsons [65] and independently to Mints [59], $\mathcal{F}(I\Sigma_1)$ coincides with the class of primitive recursive functions. Similarly, it follows from the results of Ackermann [10] and Kreisel [48] that $\mathcal{F}(\text{PA})$ coincides with the class of the so-called $<\varepsilon_0$ -recursive functions. Characterizing the classes $\mathcal{F}(T)$ for strong theories T in terms of transfinite hierarchies of recursive functions constitutes one of the main tasks of *proof-theoretic ordinal analysis*. See [68] for a survey of the modern developments in this important area of proof theory.

3.1 Fragments of PA

Main fragments of arithmetic are obtained by restricting, in one way or another, the schema of induction axiomatizing **PA**. Firstly, one can impose a restriction on the arithmetical complexity of the induction formulas. Secondly, one can disallow the use of parameters in the induction schema. Thirdly, induction is sometimes applied in the form of an inference rule rather than a schema. Induction rule yields an equivalent axiomatization of **PA**. However, the schema and the rule for formulas of restricted quantifier complexity define different hierarchies of fragments of arithmetic. A systematic study of these hierarchies, as well as of the other natural fragments of **PA**, was initiated by Parsons at the beginning of the 70's. By now it has grown into a lively area (cf. [43] for an extensive bibliography).

Fig. 2 shows the relationships between the fragments of **PA** defined by various restrictions of the induction schema over **EA**. We use the following abbreviations.

The theory $I\Sigma_n$ is axiomatized over **EA** by the induction schema for Σ_n -formulas with parameters. $I\Sigma_n$ is equivalent to III_n (cf. Lemma 4.1 below). Theories $I\Sigma_n^-$ and III_n^- are axiomatized by the corresponding schemata without parameters. $I\Sigma_n^R$ is the closure of **EA** under the induction rule

$$\text{IR} : \frac{\varphi(0), \quad \forall x (\varphi(x) \rightarrow \varphi(x+1))}{\forall x \varphi(x)},$$

where $\varphi \in \Sigma_n$. Parsons [66] has shown that $I\Sigma_n^R$ is equivalent to III_{n+1}^R (see Corollary 4.14 below). It is shown in [24] that parameter-free induction rules

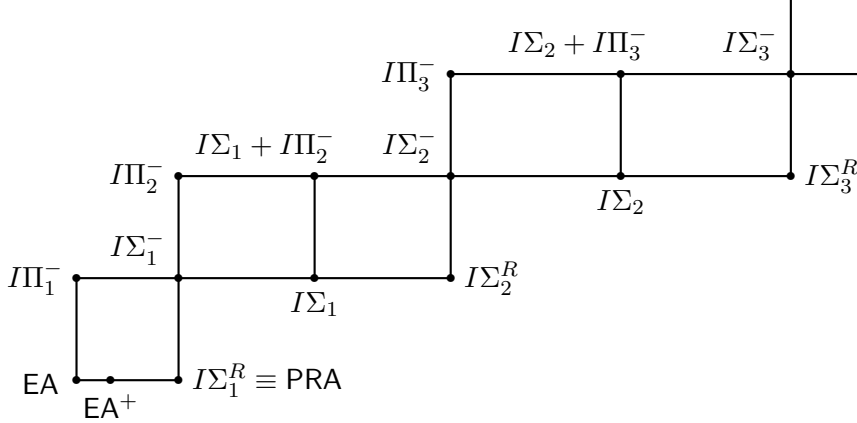


Figure 2: Fragments of PA axiomatized by restricted induction

are equivalent to the corresponding induction rules with parameters (cf. Lemma 4.6). $I\Sigma_1^R$ is equivalent to the system of *primitive recursive arithmetic* PRA discussed below.

Let \mathbf{EA}^+ denote the extension of \mathbf{EA} by an axiom asserting that the value of the function 2_n^x is defined for all values of the arguments n, x (an equivalent theory is denoted $I\Delta_0 + \text{Supexp}$ in [43]). The graph of 2_n^x can be easily defined by an elementary formula because

$$2_n^x = y \leftrightarrow \exists s \in \text{Seq}((s)_0 = x \wedge \forall i < n (s)_{i+1} = \exp((s)_i) \wedge (s)_n = y),$$

and the existential quantifier on s can be bounded by an elementary function of y , depending on the choice of the coding of sequences. We note that one can introduce within \mathbf{EA}^+ a new symbol for the function 2_n^x , and the induction schema for bounded formulas in the extended language will be provable in the corresponding definitional extension (cf. [43], p. 271).

\mathbf{EA}^+ is the most interesting fragment of PA among those between \mathbf{EA} and $I\Sigma_1^R$, for it is sufficiently strong to prove the cut-elimination theorem for the predicate logic and some of its prominent consequences such as Herbrand's theorem (cf. §4.1 below and [43, 94]). In fact, modulo \mathbf{EA} the axiom of \mathbf{EA}^+ is equivalent to the cut-elimination theorem. This follows from the well-known upper and lower bounds on the length of proofs after cut-elimination due to Statman [90] and Orevkov [2].

Proof-theoretic properties of fragments of PA often considerably differ from those of PA itself. The standard questions one asks about a given fragment are, for example:

- Finite axiomatizability of the theory;
- The optimal arithmetical complexity of its axiomatization;
- The class of provably total computable functions of the theory;
- How much reflection is provable in the given theory over a weaker fragment;

- Whether the fragment is conservative over a weaker fragment for sentences of a particular arithmetical complexity.

In §§4,5 we shall prove the main relationships of this kind between the fragments defined by restricted induction. Here we shall discuss in more detail general properties of provably total computable functions and their relationship with reflection principles.

3.2 Provably total computable functions and 1-consistency

Proposition 3.2 below more or less immediately follows from Herbrand's theorem. Recall that Herbrand's theorem states, in the simplest case, that for any first order theory T axiomatized by a set of purely universal (or quantifier-free) formulas, if $T \vdash \exists x \varphi(a, x)$, where $\varphi(a, x)$ is quantifier-free, then there is a sequence of terms $t_1(a), \dots, t_k(a)$ such that

$$T \vdash \varphi(a, t_1(a)) \vee \dots \vee \varphi(a, t_k(a)).$$

This theorem is proved by bounded induction on a cut-free derivation and hence is formalizable in \mathbf{EA}^+ (cf. [78]). There also is a simple (but less constructive) alternative model-theoretic proof based on compactness theorem.

For a given function f , let $\mathbf{C}(f)$ denote the closure of $\mathcal{E} \cup \{f\}$ under composition.

Proposition 3.2. *Let $\pi = \forall x \exists y \varphi(x, y)$ be a true Π_2 -sentence, with φ elementary. Then $\mathcal{F}(\mathbf{EA} + \pi) = \mathbf{C}(f)$, where $f(x) = \mu y. \varphi(x, y)$.*

Proof. Let T denote $\mathbf{EA} + \pi$. The graph of f can be defined by $\varphi(x, y) \wedge \forall z < y \neg \varphi(x, z)$, whence $\mathbf{C}(f) \subseteq \mathcal{F}(T)$.

For a proof of the opposite inclusion consider a purely universal axiomatization of \mathbf{EA} in a language with the symbols for all elementary functions. Add a new function symbol f together with the axiom $\forall x \varphi(x, f(x))$. Using suitable elementary terms one can get rid of all the bounded quantifiers in φ . Hence, the resulting theory is a conservative definitional extension of T and has a purely universal axiomatization.

Now assume that $T \vdash \forall x \exists! y \exists z \psi_0(x, y, z)$, where ψ_0 is elementary (and in our formulation equivalent to a quantifier-free formula). By Herbrand's theorem we obtain some terms $t_1, \dots, t_k, u_1, \dots, u_k$ of the extended language such that

$$T \vdash \psi_0(a, t_1(a), u_1(a)) \vee \dots \vee \psi_0(a, t_k(a), u_k(a)).$$

Obviously, the terms t_i and u_i represent functions in $\mathbf{C}(f)$. Then we set

$$t(x) := \begin{cases} t_1(x), & \text{if } \psi_0(x, t_1(x), u_1(x)), \\ t_2(x), & \text{if } \psi_0(x, t_2(x), u_2(x)) \text{ and } \neg \psi_0(x, t_1(x), u_1(x)), \\ \dots & \dots \\ t_k(x), & \text{if } \psi_0(x, t_k(x), u_k(x)) \text{ and } \neg \psi_0(x, t_i(x), u_i(x)) \text{ for all } i < k, \\ 0, & \text{otherwise.} \end{cases}$$

A function $u(x)$ is defined similarly, with the terms u_i in place of t_i . Since the function

$$\text{Cond}(x, y, z) := \begin{cases} x, & \text{if } z = 0 \\ y, & \text{if } z \neq 0 \end{cases}$$

is elementary, $\mathbf{C}(f)$ is closed under definitions by cases, hence $t(x)$ and $u(x)$ are definable by some terms from $\mathbf{C}(f)$. For these terms we obviously have $T \vdash \psi_0(a, t(a), u(a))$. It follows that

$$T \vdash \forall x \exists z \psi_0(x, t(x), z)$$

and, by the functionality of $\psi(x, y) = \exists z \psi_0(x, y, z)$,

$$T \vdash \forall x, y (t(x) = y \leftrightarrow \psi(x, y)).$$

Since T is sound, ψ defines the graph of $t(x)$ in the standard model.

Corollary 3.3. (i) $\mathcal{F}(\text{EA}) = \mathcal{E}$;

(ii) $\mathcal{F}(\text{EA}^+) = \mathbf{C}(\text{supexp})$, where $\text{supexp}(n, x) = 2_n^x$.

Fix some natural numbering of partial computable functions (in one variable) and an associated Σ_1 -formula $\varphi_e(x) = y$ expressing the statement that the computable function with index e on input x halts and outputs y . One usually takes the codes of Turing machines computing f as the indices of computable functions f . It will be more convenient for us here to consider an alternative indexing by the Gödel numbers of Σ_1 -formulas defining the graph of f in the standard model of arithmetic. Both numberings are equivalent in the sense that there is a provable one-to-one elementary correspondence between the indices of the two kinds. In our case, $\varphi_e(x) = y$ will be expressed using a truth-definition for Σ_1 -formulas by $\text{True}_{\Sigma_1}(\text{sub}(e, x, y))$, where $\text{sub}(e, x, y)$ is an elementary term computing $\ulcorner \sigma(\bar{x}, \bar{y}) \urcorner$ given x, y, e if $\ulcorner \sigma \urcorner = e$.

Just as for the partial computable functions, we can introduce a natural numbering of provably total computable functions in T .

Definition 3.4. Let T be an elementary presented theory. A number e is a T -index of a function f if $e = \langle e_1, e_2 \rangle$, where

- e_1 is an index of f ;
- e_2 codes a T -proof of $\forall x \exists y \varphi_{\bar{e}_1}(x) = y$.

With this indexing of $\mathcal{F}(T)$ a universal function ψ^T is naturally associated:

$$\psi_e^T(x) := \begin{cases} \varphi_{e_1}(x), & \text{if } e = \langle e_1, e_2 \rangle \text{ is a } T\text{-index;} \\ 0, & \text{otherwise.} \end{cases}$$

The usual diagonalization argument shows that ψ^T , as a function of arguments e and x , does not belong to $\mathcal{F}(T)$. Therefore, the statement of its totality is true but unprovable in T . The following lemma establishes the equivalence of this statement and the uniform Π_2 -reflection principle.

Lemma 3.5. $\text{EA} \vdash \forall e, x \exists y \psi_e(x) = y \leftrightarrow \text{RFN}_{\Pi_2}(T)$.

Proof. The totality of ψ is expressed by the formula

$$\forall e_1, e_2, x (\text{Prf}_T(e_2, \ulcorner \forall x \exists y \varphi_{e_1}(x) = y \urcorner) \rightarrow \exists y \varphi_{e_1}(x) = y). \quad (2)$$

Modulo logical equivalence, this is an instance of $\text{RFN}_{\Pi_2}(T)$. Any Π_2 -sentence is equivalent to the one of the form $\forall x \exists y \varphi_{e_1}(x) = y$ for a suitable index e_1 , hence (2) implies $\text{RFN}_{\Pi_2}(T)$.

If $f(\vec{x})$ is a function whose graph is definable, let $f \downarrow$ denote the formula $\forall \vec{x} \exists y f(\vec{x}) = y$.

The statement of totality of a function f whose graph is definable will be denoted $f \downarrow$. We presuppose that a formula defining the graph of f is fixed. As we have seen, any Π_2 -sentence can be considered as a statement of totality of a suitable function with an elementary graph.

Consider now a theory T of the form $\text{EA} + f \downarrow$, where f is such a function. Proposition 3.2 delivers a numbering of the functions from $\mathcal{F}(T) = \mathbf{C}(f)$ which does not depend on the notion of proof. The terms built-up from those for elementary functions and the symbol f have a natural Gödel numbering. We can consider these Gödel numbers as indices of provably total computable functions in T . Another universal function $\theta_e^f(x)$ is naturally associated with this numeration, which computes the value of a term with the Gödel number e on argument x . We notice that both numberings are provably equivalent in EA^+ because Herbrand's theorem is verifiable in EA^+ and allows to extract an explicit term from a proof of totality of a computable function.

Suppose a function f has an elementary graph, is non-decreasing and $f(x) \geq 2^x$. The *iteration* $f^{(k)}(x)$ of f is defined by

$$f^{(0)}(x) = x, \quad f^{(k+1)}(x) = f(f^{(k)}(x)).$$

Lemma 3.6. $\mathbf{C}(\theta^f) = \mathbf{C}(f^*)$, where $f^*(x) = f^{(x)}(x)$.

Proof. For every fixed k , $f^{(k)} \in \mathbf{C}(f)$ and the index of $f^{(k)}$ is an elementary function of k , say $i(k)$. Then

$$f^*(x) = \theta_{i(x)}^f(x). \quad (3)$$

For a proof of the opposite implication we use the monotonicity of f . Under the given assumptions every $g \in \mathbf{C}(f)$ is majorized by an iteration of f :

$$\forall x (g(x) \leq f^{(k)}(x)).$$

Here, the number k is an elementary function of the index $e = \ulcorner g \urcorner$, say $j(e)$. We may assume j to be monotone in e .

Computation of a value $y = \theta_e^f(x)$ reduces to computations of the values of subterms of g . We call a protocol of a computation of $g(x)$ a tree whose leaves are labeled by the argument x , the root is labeled by the pair $\langle e, y \rangle$, and internal nodes are labeled by pairs $\langle i, y_i \rangle$, where i is the index of a subterm of g and y_i is

the value of the subterm i on x . It is required that y_i is obtained by elementary functions or f from the corresponding values of the successors of the given node.

The total number of subterms and their indices are bounded by the index of g . Hence, all the intermediate values of subterms are bounded by $f^{(j(e))}(x) \leq f^*(\max(j(e), x))$. It follows that the code of the protocol of the computation of $\theta_e^f(x)$ is bounded by some function $b(e, x) \in \mathbf{C}(f^*)$. From this we obtain a representation of θ^f in the form:

$$\theta_e^f(x) = U(\mu z \leq b(e, x). T(e, x, z)), \quad (4)$$

where $T(e, x, z)$ is an elementary predicate expressing “ z is a protocol of a computation of a function with index e on input x ”, and $U(z)$ is an elementary function recovering the value of a function from its computation protocol.

The following lemma is obtained by formalizing the previous statement in EA.

Lemma 3.7. $\text{EA} \vdash f^* \downarrow \leftrightarrow \langle 1 \rangle_{\text{EA}} f \downarrow$.

Proof. Let $T := \text{EA} + f \downarrow$. By Lemma 3.5, $\langle 1 \rangle_{\text{EA}} f \downarrow$ is equivalent to $\psi^T \downarrow$ in EA. The formula $\psi^T \downarrow$, in particular, implies $\text{supexp} \downarrow$ and hence $\theta^f \downarrow$. The converse also holds, therefore it is sufficient to prove that $f^* \downarrow$ is equivalent to the totality of θ^f .

Obviously, if θ^f is total, then f^* is also total by (3).

For a proof of the opposite implication, by (4) it is sufficient to show $\forall e, x \exists z T(e, x, z)$ assuming the totality of f^* . Reasoning in EA consider any e, x . Since f^* is total, there is a number $a = b(e, x)$ which bounds the code of the protocol from (4). Then by induction on e_1 we can prove

$$\forall e_1 \leq_s e \exists z_1 \leq a T(e_1, x, z_1),$$

where $e_1 \leq_s e$ means that e_1 is a subterm of e . It is easy to see that the induction formula is bounded (a, e and x play the role of parameters), therefore this reasoning is formalizable in EA.

For $f(x) = \exp(x)$ we obtain the following

Corollary 3.8. $\text{EA}^+ \equiv \text{EA} + \text{RFN}_{\Pi_2}(\text{EA})$.

As a generalization of this example consider the so-called Grzegorzczuk hierarchy given by the sequence of functions

$$F_0(x) := 2^x; \quad F_{n+1}(x) := F_n^{(x)}(x).$$

Functions F_n are primitive recursive and have elementary graphs. Denote $\mathcal{E}^n := \mathbf{C}(F_n)$. It is well known that the sequence of classes $\mathcal{E}^0 \subset \mathcal{E}^1 \subset \mathcal{E}^2 \subset \dots$ exhausts all primitive recursive functions (cf. [74]). An extension of EA by the axioms $F_n \downarrow$, for all $n \geq 1$, is an alternative axiomatization of *primitive recursive arithmetic* PRA.

Proposition 3.9. (i) $\text{EA}_\omega^1 \equiv I\Sigma_1^R \equiv \text{PRA}$.

(ii) $\mathcal{F}(\text{EA}_\omega^1)$ coincides with the class of primitive recursive functions.

Proof. (i) The totality of all F_n can easily be established by the rule of Σ_1 -induction. The inclusion $\text{EA}_\omega^1 \subseteq \text{PRA}$ is obtained by iterating Lemma 3.7. We prove that EA_ω^1 is closed under the Σ_1 -induction rule.

For a given $\varphi \in \Sigma_1$ let ψ denote the universal closure of

$$\varphi(0) \wedge \forall x (\varphi(x) \rightarrow \varphi(x+1)).$$

If $\text{EA}_k^1 \vdash \psi$, then

$$\text{EA}_{k+1}^1 \vdash \text{RFN}_{\Pi_2}(\text{EA} + \psi).$$

On the other hand, it is easy to see that, for every fixed $n \in \mathbb{N}$, $\text{EA} + \psi \vdash \varphi(\bar{n})$. This fact is formalizable in EA by a bounded induction on n , therefore

$$\text{EA} \vdash \forall x \Box_{\text{EA}+\psi} \varphi(\dot{x}).$$

Hence, using Π_2 -reflection we obtain $\text{EA}_{k+1}^1 \vdash \forall x \varphi(x)$, as required.

(ii) Any function from $\mathcal{F}(\text{PRA})$ is primitive recursive because so are all F_n . A natural proof of totality of any primitive recursive function is formalizable in $I\Sigma_1^R$.

4 Induction and reflection

In this and the following paragraphs we present the main results concerning interrelations between reflection schemata and various forms of induction in arithmetic. The first and the most well-known result of this type was obtained by Kreisel and Lévy [51] who showed that PA was equivalent to the full uniform reflection principle over the primitive recursive arithmetic. Later this theorem was strengthened by Leivant [52] who has established a sharp correspondence between the fragments $I\Sigma_n$ and the hierarchy of partial uniform reflection principles. Below, we also give exact characterizations of induction rules and of parameter-free induction schemata in terms of reflection principles obtained in [20, 22]. On the basis of the results of §2, these relationships allow to obtain some interesting corollaries on the structure of these theories and their classes of provably total computable functions.

4.1 Tait calculus

We rely on the standard cut-elimination techniques for a variant of sequent calculus due to Tait that is presented, for example, in Schwichtenberg [78]. *Formulas* in Tait's calculus are constructed as in the first order logic from atomic formulas and their negations by the connectives \wedge , \vee and the quantifiers \forall , \exists . It is also convenient to assume that the alphabets of free and bound variables are disjoint. Negation is recursively defined as an operation on formulas by de Morgan's rules. *Sequents* are finite sets of formulas, denoted Γ , Δ , \dots and understood as disjunctions. $\Gamma \cup \{\varphi\}$ is written as Γ, φ .

Tait's calculus is given by the axioms $\Gamma, \varphi, \neg\varphi$, where φ is an atomic formula, and the following inference rules:

$$\begin{array}{c} \frac{\Gamma, \varphi \quad \Gamma, \psi}{\Gamma, \varphi \wedge \psi} (\wedge) \quad \frac{\Gamma, \varphi}{\Gamma, \varphi \vee \psi} (\vee_0) \quad \frac{\Gamma, \psi}{\Gamma, \varphi \vee \psi} (\vee_1) \\[10pt] \frac{\Gamma, \varphi(a)}{\Gamma, \forall x \varphi(x)} (\forall), \quad \text{where } a \text{ does not occur (free) } \Gamma \\[10pt] \frac{\Gamma, \varphi(t)}{\Gamma, \exists x \varphi(x)} (\exists) \quad \frac{\Gamma, \varphi \quad \Gamma, \neg\varphi}{\Gamma} (Cut) \end{array}$$

It is not difficult to show that a sequent Γ is provable in Tait's calculus if and only if the formula $\bigvee \Gamma$ is provable in pure predicate calculus PC. The cut-elimination theorem (cf. [78]) states that there is an effective procedure transforming any derivation of Γ in Tait's calculus to a derivation of Γ in which there are no applications of the rule (Cut) . Cut-free proofs have the property that any formula occurring in such a proof is a subformula of a formula occurring in the last sequent. (The property of being a subformula is understood here modulo substituting arbitrary terms for the free variables.) In particular, a cut-free proof of a Π_n -formula consists entirely of Π_n -formulas.

From [78] one can extract a rough upper bound of order 2_{cn}^x on the size of the cut-free proof where x is the size of the original proof, n is its cut-rank, and c is some constant. This allows to directly formalize in \mathbf{EA}^+ the usual inductive proof of the cut-elimination theorem (cf. also [43], p. 376).

For the language of Tait's calculus it is technically convenient to extend the syntactical notion of Π_n -formula to all formulas obtained from Σ_{n-1} - and Π_{n-1} -formulas using \wedge, \vee , and \forall . In this and the following sections, we shall work in Tait's calculus for the language of arithmetic without the symbol of exponentiation. A truth-definition for Δ_0 -formulas expressed by an elementary formula in \mathbf{EA} is constructed in [43], p. 364. As in [43], p. 58–59, one can construct from it the truth-definitions for the classes Σ_n and Π_n of the arithmetical hierarchy, for all $n \geq 1$.

Apart from Proposition 2.6 we shall need some additional properties of the truth-definition for Π_n -formulas, called *Tarski's conditions*. Tarski's conditions formally express the fact that $\mathbf{True}_{\Pi_n}(x)$ is correctly defined on atomic formulas and provably in \mathbf{EA} commutes with the boolean connectives and quantifiers on the class of Π_n -formulas. In other words, \mathbf{EA} proves: *for any formulas $\varphi, \psi, \theta, \alpha, \gamma$ such that $\varphi, \neg\varphi, \psi, \theta, \forall x\gamma(x), \exists x\alpha(x) \in St \cap \Pi_n$,*

$$\begin{array}{l} \mathbf{True}_{\Pi_n}(\ulcorner \neg\varphi \urcorner) \leftrightarrow \neg \mathbf{True}_{\Pi_n}(\ulcorner \varphi \urcorner), \\ \mathbf{True}_{\Pi_n}(\ulcorner \theta \wedge \psi \urcorner) \leftrightarrow \mathbf{True}_{\Pi_n}(\ulcorner \theta \urcorner) \wedge \mathbf{True}_{\Pi_n}(\ulcorner \psi \urcorner), \\ \mathbf{True}_{\Pi_n}(\ulcorner \theta \vee \psi \urcorner) \leftrightarrow \mathbf{True}_{\Pi_n}(\ulcorner \theta \urcorner) \vee \mathbf{True}_{\Pi_n}(\ulcorner \psi \urcorner), \\ \mathbf{True}_{\Pi_n}(\ulcorner \exists x\alpha(x) \urcorner) \leftrightarrow \exists x \mathbf{True}_{\Pi_n}(\ulcorner \alpha(\dot{x}) \urcorner), \\ \mathbf{True}_{\Pi_n}(\ulcorner \forall x\gamma(x) \urcorner) \leftrightarrow \forall x \mathbf{True}_{\Pi_n}(\ulcorner \gamma(\dot{x}) \urcorner). \end{array}$$

Let us stress that the above conditions are provable in **EA** not only for individual formulas φ, ψ, \dots , but also with the universal quantifiers over their Gödel numbers. (In the latter case $\ulcorner \alpha(\dot{x}) \urcorner$ should be understood as a definable term for the elementary function mapping the Gödel number of a formula α and a number x to $\ulcorner \alpha(\bar{x}) \urcorner$. Below we shall only use Tarski's conditions in an informal context, therefore we do not introduce any special notations for such functions.)

We also assume that the formalization of the operation $\bigvee \Gamma$ on sequents Γ consisting of Π_n -formulas **EA**-provably agrees with $\text{True}_{\Pi_n}(x)$, in the sense that the truth of $\bigvee \Gamma$ does not depend on the ordering of Γ in a disjunction. In particular, **EA** proves: *for all* $\varphi \in \Gamma \subseteq St \cap \Pi_n$

$$\text{True}_{\Pi_n}(\ulcorner \varphi \urcorner) \rightarrow \text{True}_{\Pi_n}(\ulcorner \bigvee \Gamma \urcorner). \quad (5)$$

We assume that for an empty sequent $\Gamma \quad \bigvee \Gamma := \perp$ and therefore $\neg \text{True}_{\Pi_n}(\ulcorner \bigvee \Gamma \urcorner)$ is provable.

In **EA**, we also have a reasonable universal function for terms in the language without exponentiation, that is, a definable elementary function $\text{eval}(u, x)$ satisfying³

$$\text{EA} \vdash \text{eval}(\ulcorner t \urcorner, \langle x_0, \dots, x_n \rangle) = t(x_0, \dots, x_n),$$

for any term $t(x_0, \dots, x_n)$. The function $\text{eval}(u, x)$ usually appears in the construction of a truth-definition for the evaluation of atomic and bounded formulas ([43], p. 363–364). Together with the provable totality of $\text{eval}(u, x)$ this implies that the truth-definition and the universal function for terms agree in the sense that **EA** proves: *for any* Π_n -formula $\varphi(z_0, \dots, z_m)$ *and any terms* $t_0(\vec{x}), \dots, t_m(\vec{x})$

$$\bigwedge_{i=0}^m \text{eval}(\ulcorner t_i \urcorner, \langle \vec{x} \rangle) = y_i \rightarrow (\text{True}_{\Pi_n}(\ulcorner \varphi(t_0(\vec{x}), \dots, t_m(\vec{x})) \urcorner) \leftrightarrow \text{True}_{\Pi_n}(\ulcorner \varphi(\dot{y}_0, \dots, \dot{y}_m) \urcorner)). \quad (6)$$

Here and below, \vec{x} is an abbreviation for x_0, \dots, x_k ; $\dot{\vec{x}}$ stands for $\dot{x}_0, \dots, \dot{x}_k$; $\bar{\vec{x}}$ denotes $\bar{x}_0, \dots, \bar{x}_k$; and $\langle \vec{x} \rangle$ is $\langle x_0, \dots, x_k \rangle$.

4.2 Induction schemata with parameters

We begin with a basic lemma on restricted induction schemata established in [66].

Lemma 4.1. $I\Sigma_n \equiv I\Pi_n$, for any $n \geq 1$.

Proof. Let $\varphi(x)$ be a Π_n -formula. To prove the instance of **IA** for $\varphi(x)$ by Σ_n -induction, we consider the Σ_n -formula $\psi(a, x) := \neg \varphi(a \div x)$. Then $\forall x (\varphi(x) \rightarrow \varphi(x+1))$ implies $\forall x (\psi(a, x) \rightarrow \psi(a, x+1))$, hence $\psi(a, 0) \rightarrow \psi(a, a)$, by $I\Sigma_n$. But this means $\varphi(0) \rightarrow \varphi(a)$.

Notice the role of parameter a in this argument: the proof does not work for $I\Sigma_n^-$. The following theorem was proved by Leivant [52] for the case of **PRA** and $n \geq 2$. The case $n = 1$ requires a weakening of **PRA** to **EA**. The inclusion of theories from left to right was first noticed, probably, by Ono [63].

³We assume that all variables of the arithmetical language are ordered in a sequence x_0, \dots, x_n, \dots

Theorem 7. $I\Sigma_n \equiv \text{EA} + \text{RFN}_{\Sigma_{n+1}}(\text{EA})$, for any $n \geq 1$.

Proof. (\subseteq) Let $\sigma(x, y) \in \Sigma_n$ where without loss of generality we consider only one parameter y . Let $\psi(y)$ denote the formula

$$\sigma(0, y) \wedge \forall x (\sigma(x, y) \rightarrow \sigma(x+1, y)).$$

By an obvious external induction on k we obtain:

$$\forall k, m \in \mathbb{N} \quad \text{EA} \vdash \psi(\overline{m}) \rightarrow \sigma(\overline{k}, \overline{m}).$$

Since the size of the EA-proof above is estimated by an elementary function in k and m , the external induction on k is formalizable in EA:

$$\text{EA} \vdash \forall x, y \quad \Box_{\text{EA}}(\psi(y) \rightarrow \sigma(x, y)).$$

Notice that $\psi(y) \rightarrow \sigma(x, y)$ is logically equivalent to a Σ_{n+1} -formula, hence by the uniform Σ_{n+1} -reflection we infer

$$\text{EA} + \text{RFN}_{\Sigma_{n+1}}(\text{EA}) \vdash \forall x, y (\psi(y) \rightarrow \sigma(x, y)),$$

which is equivalent to the corresponding instance of the induction schema.

For a proof of $\text{RFN}_{\Sigma_{n+1}}(\text{EA})$ in $I\Sigma_n$, we first notice that it is sufficient to prove in $I\Sigma_n$ the uniform Σ_{n+1} -reflection principle $\text{RFN}_{\Sigma_{n+1}}(\text{PC})$ for the *pure predicate calculus* in the arithmetical language without exponentiation. This follows from the provable equivalence of EA and $I\Delta_0 + \text{Exp}$ and from the fact that $I\Delta_0 + \text{Exp}$ has a finite Π_2 -axiomatization (over PC), whence, by an analog of Lemma 2.23,

$$\begin{aligned} \text{EA} \vdash \text{RFN}_{\Sigma_{n+1}}(\text{EA}) &\leftrightarrow \text{RFN}_{\Sigma_{n+1}}(I\Delta_0 + \text{Exp}) \\ &\leftrightarrow \text{RFN}_{\Sigma_{n+1}}(\text{PC}). \end{aligned}$$

Now, given a Σ_{n+1} -formula $\sigma(z) := \exists y_1, \dots, y_m \alpha(y_1, \dots, y_m, z)$ with $\alpha \in \Pi_n$ and a number $k \in \mathbb{N}$, consider a cut-free derivation of a sequent $\sigma(\overline{k}), \Pi$, for a set of Π_n -formulas Π . By the subformula property, any formula occurring in this derivation either (a) has the form

$$\exists y_i \dots \exists y_m \alpha(t_0, \dots, t_{i-1}, y_i, \dots, y_m, \overline{k}),$$

for some $0 \leq i \leq m$ and terms t_0, \dots, t_{i-1} ; or (b) is a Π_n -formula.

Let $I_\sigma(h, k)$ be a Π_n -formula naturally expressing the following:

“For all p , if p is a cut-free derivation of a sequent of the form $\Gamma, \Pi(\vec{a})$ where Γ is a set formulas of type (a) above, $\Pi(\vec{a})$ is a set of Π_n -formulas, and if the height of p is smaller than h , then $\forall \vec{x} \text{True}_{\Pi_n}(\ulcorner \bigvee \Pi(\vec{x}) \urcorner)$.”

Lemma 4.2. $\text{EA} \vdash \neg \sigma(z) \rightarrow (I_\sigma(0, z) \wedge \forall h (I_\sigma(h, z) \rightarrow I_\sigma(h+1, z)))$.

Proof. We shall prove the lemma reasoning informally within EA. $I_\sigma(0, z)$ trivially holds. We show that $I_\sigma(h, z)$ implies $I_\sigma(h+1, z)$. Thus, we are given a cut-free derivation, of height h , of a sequent of the form $\Gamma, \Pi(\vec{a})$ where Γ and Π are as above, and we must show that the disjunction of Π is true, in the sense of True_{Π_n} , under every substitution of numerals for the free variables in Π . We distinguish several cases, according to the form of the last rule applied in the given derivation. The statement $I_\sigma(h, z)$ will be called the induction hypothesis.

CASE 1. The sequent $\Gamma, \Pi(\vec{a})$ is a logical axiom, i.e., it has the form $\Delta, \varphi(\vec{a}), \neg\varphi(\vec{a})$, for some atomic φ . Since all the formulas of type (a) contain at least one existential quantifier and therefore are neither atomic nor negated atomic, both φ and $\neg\varphi$ must belong to Π . Tarski's conditions then imply that, for all \vec{x} ,

$$\text{True}_{\Pi_n}(\ulcorner \varphi(\vec{x}) \urcorner) \vee \text{True}_{\Pi_n}(\ulcorner \neg\varphi(\vec{x}) \urcorner),$$

whence we infer $\text{True}_{\Pi_n}(\ulcorner \bigvee \Pi(\vec{x}) \urcorner)$ by (5).

CASE 2. The sequent $\Gamma, \Pi(\vec{a})$ is obtained by a rule introducing a boolean connective or a quantifier into a formula from $\Pi(\vec{a})$. All these rules are treated similarly using the subformula property of cut-free derivations and Tarski's conditions for True_{Π_n} . For example, the rule for the universal quantifier has the form

$$\frac{\Gamma, \Pi'(\vec{a}), \varphi(\vec{a}, b)}{\Gamma, \Pi'(\vec{a}), \forall x \varphi(\vec{a}, x)},$$

where b does not occur free in $\Gamma, \Pi'(\vec{a})$. We must show that the formula $\bigvee \Pi'(\vec{x}) \vee \forall y \varphi(\vec{x}, y)$ is true. By the induction hypothesis, since b does not occur free in Π' , we know that, for each y , $\bigvee \Pi'(\vec{x}) \vee \varphi(\vec{x}, \bar{y})$ is true. Commuting True_{Π_n} with \vee we conclude that, for each y , either $\bigvee \Pi'(\vec{x})$ or $\varphi(\vec{x}, \bar{y})$ is true. Since $\text{True}_{\Pi_n}(\ulcorner \bigvee \Pi'(\vec{x}) \urcorner)$ does not depend on y , it follows that either $\Pi'(\vec{x})$ is true or, for every y , $\varphi(\vec{x}, \bar{y})$ is true. Commuting True_{Π_n} with the universal quantifier and then backwards with the disjunction we conclude that $\Pi'(\vec{x}) \vee \forall y \varphi(\vec{x}, y)$ is true, q.e.d.

CASE 3. The last inference introduces the existential quantifier in front of α , i.e., our derivation has the form

$$\frac{\Gamma', \alpha(t_0(\vec{a}), \dots, t_{m-1}(\vec{a}), t_m(\vec{a}), \bar{z}), \Pi(\vec{a})}{\Gamma', \exists y_m \alpha(t_0(\vec{a}), \dots, t_{m-1}(\vec{a}), y_m, \bar{z}), \Pi(\vec{a})}.$$

By the induction hypothesis, either the disjunction of $\Pi(\vec{x})$ or $\alpha(t_0(\vec{x}), \dots, t_m(\vec{x}), \bar{z})$ is true. We must, reasoning inside EA, refute the second alternative.

Notice that, although, in general, t_i are “nonstandard” terms, α is a fixed “standard” Π_{n+1} -formula. Therefore, Proposition 2.6 can be applied to α after evaluating the terms t_i . Thus, by (6) and Proposition 2.6, we obtain:

$$\begin{aligned} \text{True}_{\Pi_n}(\ulcorner \alpha(t_0(\vec{x}), \dots, t_m(\vec{x}), \bar{z}) \urcorner) \wedge \bigwedge_{i=0}^m \text{eval}(\ulcorner t_i \urcorner, \langle \vec{x} \rangle) = y_i &\rightarrow \\ &\rightarrow \text{True}_{\Pi_n}(\ulcorner \alpha(\dot{y}_0, \dots, \dot{y}_m, \bar{z}) \urcorner) \\ &\rightarrow \alpha(y_0, \dots, y_m, \bar{z}). \end{aligned}$$

Since the universal function is provably total in EA, it follows that

$$\text{True}_{\Pi_n}(\ulcorner \alpha(t_0(\vec{x}), \dots, t_m(\vec{x})), \dot{z} \urcorner)$$

implies $\exists y_0 \dots y_m \alpha(y_0, \dots, y_m, z)$, that is, it contradicts the hypothesis $\neg\sigma(z)$. Thus, for any t_0, \dots, t_m and \vec{x} , the formula $\alpha(t_0(\vec{x}), \dots, t_m(\vec{x}), \bar{z})$ cannot be true, hence the disjunction of $\Pi(\vec{x})$ is true.

CASE 4. $\Gamma, \Pi(\vec{a})$ is obtained by a rule introducing any other existential quantifier into a formula from Γ . Then our claim immediately follows from the induction hypothesis because the Π -part of the premise is in this case the same as that of the conclusion.

Since, for $n \geq 1$, $I\Sigma_n$ is equivalent to III_n and contains EA^+ , an immediate corollary of the above lemma is that

$$I\Sigma_n \vdash \neg\sigma(z) \rightarrow \forall h I_\sigma(h, z). \quad (7)$$

In order to derive $\text{RFN}_{\Sigma_{n+1}}(\text{PC})$ we now reason within $I\Sigma_n$, for every particular Σ_{n+1} -formula $\sigma(z)$, as follows.

Suppose $\neg\sigma(z)$ and $\Box_{\text{PC}} \sigma(\dot{z})$, then the sequent consisting of the single formula $\sigma(\bar{z})$ is provable in Tait's calculus. By the (formalized) cut-elimination theorem we obtain a cut-free proof of this sequent, and by (7) we conclude that $\text{True}_{\Pi_n}(\ulcorner \perp \urcorner)$ holds. But this implies \perp , that is, a contradiction.

The following statement was obtained by Kreisel and Lévy [51].

Corollary 4.3. $\text{PA} \equiv \text{EA} + \text{RFN}(\text{EA})$.

As a corollary of the unboundedness theorems for the uniform reflection principles we obtain the following results, of which the first one is due to Rabin [71] and the second one to Leivant [52].

Corollary 4.4. (i) *PA is not contained in any consistent extension of EA of bounded arithmetical complexity.*

(ii) *$I\Sigma_n$ is not contained in any consistent extension of EA of complexity Σ_{n+2} .*

We note that the complexity of the obvious axiomatization of $I\Sigma_n$ is Π_{n+2} .

Corollary 4.5. *Neither III_{n+1}^R , nor III_{n+1}^- contains $I\Sigma_n$.*

Proof. The arithmetical complexity of their axiomatizations is, respectively, Π_{n+1} and Σ_{n+2} (all conclusions of instances of the rule Π_{n+1} -IR have complexity Π_{n+1}).

4.3 Induction rules

First, we establish some general terminology and notation concerning inference rules in arithmetic. We say that a *rule* is a set of *instances*, that is, expressions of the form

$$\frac{\varphi_1, \dots, \varphi_n}{\psi}, \quad (8)$$

where $\varphi_1, \dots, \varphi_n$ and ψ are arithmetical formulas. Derivations using rules are defined in the standard way; $T + R$ denotes the closure of a theory T under a rule R and first order logic. $[T, R]$ denotes the closure of T under *non-nested applications* of R , that is, the theory axiomatized over T by all formulas ψ such that, for some $\varphi_1, \dots, \varphi_n$ derivable in T , (8) is an instance of R .

A rule R_1 is *derivable* from R_2 if, for every theory T containing EA, $T + R_1 \subseteq T + R_2$. This means, in other words, that every derivation from premises using R_1 can be replaced by an equivalent derivation using EA and R_2 . A rule R_1 is *reducible* to R_2 if, for every theory T containing EA, $[T, R_1] \subseteq [T, R_2]$. R_1 and R_2 are *congruent* if they are mutually reducible (denoted $R_1 \cong R_2$). For any theory U containing EA, we say that R_1 and R_2 are *congruent modulo U* if, for every extension T of U , $[T, R_1] \equiv [T, R_2]$.

Next we shall consider in more detail the induction rules for Σ_n - and Π_n -formulas, denoted Σ_n -IR and Π_n -IR, respectively. In general, we allow parameters to occur in the induction formulas of these rules, however the following useful lemma holds.

Lemma 4.6. (i) Π_n -IR is reducible to parameter-free Π_n -IR.

(ii) Σ_n -IR is reducible to parameter-free Σ_n -IR.

Proof. An application of IR to a formula $\varphi(x, a)$ with the induction variable x and a parameter a is obviously reducible to an application of IR for $\forall z \varphi(x, z)$, which proves the first statement of the lemma.

Now assume that $\varphi(x, y, a) \in \Pi_{n-1}$ and

$$T \vdash \exists y \varphi(0, y, a) \text{ and} \quad (9)$$

$$T \vdash \forall x (\exists y \varphi(x, y, a) \rightarrow \exists y \varphi(x+1, y, a)). \quad (10)$$

Let

$$\varphi'(x, y) := \forall i \leq x \varphi((i)_0, (y)_i, (i)_1).$$

Then, by (9) and by the monotonicity of the coding of sequences, $T \vdash \exists y \varphi'(0, y)$. For a proof of

$$T \vdash \forall x (\exists y \varphi'(x, y) \rightarrow \exists y' \varphi'(x+1, y'))$$

assume that $\forall i \leq x \varphi((i)_0, (y)_i, (i)_1)$. If $(x+1)_0 = 0$, then by (9) there is an element z such that $\varphi(0, z, (x+1)_1)$, and one can take the sequence $y * \langle z \rangle$ for y' . If, on the other hand, $(x+1)_0 > 0$, then the code of the pair $p := \langle (x+1)_0 - 1, (x+1)_1 \rangle$ is strictly smaller than $x+1$ and hence, by the induction hypothesis, there is a $z = (y)_p$ such that $\varphi((x+1)_0 - 1, z, (x+1)_1)$. It follows from (10) that for some z' there holds $\varphi((x+1)_0, z', (x+1)_1)$. Hence, one can take $y * \langle z' \rangle$ for y' . Thus, an application of IR to $\exists y \varphi(x, y, a)$ reduces to an application of IR for the Σ_n -formula $\exists y \varphi'(x, y)$ without parameters.

Corollary 4.7. $I\Sigma_n^R \subseteq I\Sigma_n^-, \quad I\Pi_n^R \subseteq I\Pi_n^-.$

Now we shall introduce another inference rule related to the principle of reflection. Given an elementary presented theory T , the *metareflection rule over T* is the rule

$$\text{RR}^n(T) : \quad \frac{\varphi(\vec{x})}{\langle n \rangle_T \varphi(\vec{x})}.$$

We let $\Pi_m\text{-RR}^n$ denote the above rule with a restriction that φ is a Π_m -sentence. If $T = \text{EA}$ we write $\Pi_m\text{-RR}^n$.

Notice that $\Pi_m\text{-RR}^n(T)$ is reducible to its parameter-free variant because one can derive by this rule from $\varphi(\vec{x})$ the formula $\langle n \rangle_T \forall \vec{x} \varphi(\vec{x})$, whence $\forall \vec{x} \langle n \rangle_T \varphi(\vec{x})$ follows by Löb's conditions.

The following theorem is obtained in [8, 20].

Theorem 8. (i) $\Pi_{n+1}\text{-IR} \cong \Pi_{n+2}\text{-RR}^n$, for $n > 0$;
(ii) $\Pi_1\text{-IR} \cong \Pi_2\text{-RR}^0$ (modulo EA^+).

In order to reduce $\Pi_{n+1}\text{-IR}$ to $\Pi_{n+2}\text{-RR}^n$, assume a theory T proves

$$\varphi(0) \wedge \forall x (\varphi(x) \rightarrow \varphi(x+1)), \quad (11)$$

where we can assume that φ does not contain additional parameters. Notice that (11) is logically equivalent to a Π_{n+2} -sentence, say ψ . Obviously, for every particular $n \in \mathbb{N}$ we have $\text{EA} + \psi \vdash \varphi(\bar{n})$. Formalizing this fact in EA we obtain:

$$\text{EA} \vdash \forall x \Box_{\text{EA}+\psi} \varphi(\dot{x}).$$

On the other hand, by metareflection from ψ we infer

$$[T, \Pi_{n+2}\text{-RR}^n] \vdash \text{RFN}_{\Pi_{n+1}}(\text{EA} + \psi),$$

whence

$$[T, \Pi_{n+2}\text{-RR}^n] \vdash \forall x \varphi(x),$$

as required.

Now we have to show that, for T containing EA (or EA^+ for $n = 0$),

$$[T, \Pi_{n+1}\text{-IR}] \vdash \text{RFN}_{\Pi_{n+1}}(\text{EA} + \psi),$$

for any $\psi \in \Pi_{n+2}$ such that $T \vdash \psi$. Without loss of generality we can assume T to be formulated in the arithmetical language without exponentiation and $T \equiv I\Delta_0 + \text{Exp} + \psi$. Using finite axiomatizability of $I\Delta_0 + \text{Exp}$ we also assume that the only non-logical axiom of T has the form $\forall x_0 \dots \forall x_m \neg \alpha(x_0, \dots, x_m)$ where $\alpha \in \Pi_n$. In particular, this formula accumulates all the equality axioms and a finite Π_2 -axiomatization of $I\Delta_0 + \text{Exp}$.

For a proof of $\text{RFN}_{\Pi_{n+1}}(T)$ we consider arbitrary cut-free derivations of

$$\exists x_0 \dots \exists x_m \alpha(x_0, \dots, x_m), \Pi$$

in Tait's calculus, where Π is a set of Π_{n+1} -formulas. In this situation Lemma 4.2 is directly applicable: $\exists x_0 \dots \exists x_m \alpha(x_0, \dots, x_m)$, that is, the negation of the axiom of T , plays the role of σ . In this case, parameter z is absent.

Since T contains **EA**, by Lemma 4.2,

$$T \vdash I_\sigma(0) \wedge \forall h (I_\sigma(h) \rightarrow I_\sigma(h+1)).$$

Hence, by Π_{n+1} -IR we can infer

$$[T, \Pi_{n+1}\text{-IR}] \vdash \forall h I_\sigma(h),$$

which implies $\text{RFN}_{\Pi_{n+1}}(T)$ over EA^+ . It remains us to notice that, for $n > 0$, the theory $[\text{EA}, \Pi_{n+1}\text{-IR}]$ contains EA^+ .

Let, for a fixed $n \geq 0$, T_k^n denote a progression of theories based on iteration of the n -consistency assertion over T :

$$T_0^n := T, \quad T_{k+1}^n := T_k^n + n\text{-Con}(T_k^n), \quad T_\omega^n := \bigcup_{k \geq 0} T_k^n.$$

Since $n\text{-Con}(T)$ has complexity Π_{n+1} , Theorem 8 can be applied several times and we obtain the following

Corollary 4.8. *Let T be a finite Π_{n+2} -extension of **EA** (or EA^+ , for $n = 0$). Then*

$$T + \Pi_{n+1}\text{-IR} \equiv T_\omega^n.$$

One can also remark that the k -fold application of the induction rule precisely corresponds to the k -times iterated reflection principle over T .

Corollary 4.9. $\text{I}\Pi_{n+1}^R \equiv \text{EA}_\omega^n$, for any $n > 0$.

An interesting particular case of Corollary 4.8 concerns the induction rule for Π_1 -formulas.

Corollary 4.10. *For finite Π_2 -axiomatized theories T containing EA^+ ,*

$$T + \Pi_1\text{-IR} \equiv T_\omega \equiv T + \text{Con}(T) + \text{Con}(T + \text{Con}(T)) + \dots \quad (12)$$

Corollary 4.11. *For Π_{n+1} -extensions T of **EA** (or of EA^+ for $n = 1$), the theory $T + \Pi_n\text{-IR}$ is not finitely axiomatizable provided it is consistent.*

Corollary 4.12. *The theories $\text{I}\Pi_n^R$, for $n > 1$, and also $\text{EA}^+ + \Pi_1\text{-IR}$ are not finitely axiomatizable.*

Remark 4.13. A characterization of $\Pi_1\text{-IR}$ modulo congruence in **EA** can be obtained on the basis of the ideas of Wilkie and Paris [94]. In this situation, a formula $\text{Con}^{\text{cf}}(T)$ expressing the consistency of T in the sense of the cut-free provability plays the role of $\text{Con}(T)$. Notice that both formulas are equivalent within EA^+ , by the formalizability of the cut-elimination theorem.

De facto, all applications of the cut-rule in our proof of Theorem 8 have a bounded (standard) rank, hence they can already be eliminated provably within EA. With this in mind, all the reasoning goes through in EA, which shows that $\Pi_1\text{-IR}$ is congruent to

$$\frac{\varphi}{\text{Con}^{\text{cf}}(\text{EA} + \varphi)}, \quad \varphi \in \Pi_2.$$

By the results of [94], this also implies that $I\Pi_1^R$ is not finitely axiomatizable.

A proof of the following theorem is more involved and we omit it (cf. [20]).

Theorem 9. (i) $\Sigma_1\text{-IR} \cong \Pi_2\text{-RR}^1$;

(ii) $\Sigma_n\text{-IR} \cong \Pi_{n+1}\text{-RR}^n$ modulo $I\Sigma_{n-1}$, for $n > 1$.

Since it is easy to see that $[\text{EA}, \Sigma_n\text{-IR}]$ contains $I\Sigma_{n-1}$, Statement (ii) implies that the rules $\Pi_{n+1}\text{-RR}^{n+1}$ and $\Sigma_n\text{-IR}$ are interderivable, for any $n \geq 1$. In fact, k nested applications of $\Pi_{n+1}\text{-RR}^n$ are reducible to $k+1$ applications of $\Sigma_n\text{-IR}$, the first one is used to infer $I\Sigma_{n-1}$. The question whether k applications of $\Sigma_n\text{-IR}$ would be sufficient here is still open.

Corollary 4.14. $I\Sigma_n^R \equiv \text{EA}_\omega^n \equiv I\Pi_{n+1}^R$, for all $n > 0$.

Notice that we have already proved this statement for $n = 1$ (Proposition 3.9).

4.4 Parameter-free induction

In this section we establish a relationship between reflection principles and parameter-free induction schemata in arithmetic. In view of an exact correspondence between induction schemata and partial uniform reflection principles over EA (Proposition 7), it seems natural to assume that parameter-free induction schemata should correspond to local reflection principles. However, it is not difficult to see that local reflection principles are far too weak for this goal: $\text{Rfn}(\text{EA})$ is contained in the extension of EA by the set of all true Π_1 -sentences, whereas by Corollaries 4.7, 4.9 and Proposition 2.16 none of the theories $I\Pi_n^-$, for $n > 1$, is. It turns out that in order to obtain a sharp characterization one needs to relativize the provability predicate (cf. §2.3).

Relativized local reflection principles are defined by analogy as follows:

$$\begin{aligned} \text{Rfn}^n(T) &:= \{[n]_T \varphi \rightarrow \varphi \mid \varphi \in St\}, \\ \text{Rfn}_{\Sigma_m}^n(T) &:= \{[n]_T \sigma \rightarrow \sigma \mid \sigma \in St \cap \Sigma_m\}. \end{aligned}$$

For $n = 0$ all these schemata coincide, by definition, with their non-relativized analogs. Notice that, by contraposition, $\text{Rfn}_{\Sigma_m}^n(T)$ is equivalent over EA to

$$\{\pi \rightarrow \langle n \rangle_T \pi \mid \pi \in St \cap \Pi_m\}.$$

The following characterization was obtained in [22, 24].

Theorem 10. *For $n \geq 1$ the following theories are equivalent*

- (i) $I\Sigma_n^- \equiv \mathbf{EA} + \text{Rfn}_{\Sigma_{n+1}}^n(\mathbf{EA})$;
- (ii) $I\Pi_{n+1}^- \equiv \mathbf{EA} + \text{Rfn}_{\Sigma_{n+2}}^n(\mathbf{EA})$;
- (iii) $\mathbf{EA}^+ + I\Pi_1^- \equiv \mathbf{EA}^+ + \text{Rfn}_{\Sigma_2}(\mathbf{EA}) \equiv \mathbf{EA}^+ + \text{Rfn}_{\Sigma_2}(\mathbf{EA}^+)$.

Proof. All the statements are proved similarly using Theorems 8 and 9. We only give a proof of (ii).

For the inclusion (\subseteq) we have to prove

$$\varphi(0) \wedge \forall x (\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x),$$

for each Π_{n+1} -formula $\varphi(x)$ with a single free variable x . Let ψ denote a Π_{n+2} -sentence logically equivalent to

$$\varphi(0) \wedge \forall x (\varphi(x) \rightarrow \varphi(x+1)).$$

As usually, an external induction on m allows one to easily establish, for each m , $\mathbf{EA} + \psi \vdash \varphi(\overline{m})$. This fact is formalizable in \mathbf{EA} , hence

$$\mathbf{EA} \vdash \forall x \Box_{\mathbf{EA}+\psi} \varphi(\dot{x}). \quad (13)$$

From this we conclude

$$\begin{aligned} \mathbf{EA} + \text{Rfn}_{\Sigma_{n+2}}^n(\mathbf{EA}) + \psi &\vdash \text{RFN}_{\Pi_{n+1}}(\mathbf{EA} + \psi) \\ &\vdash \forall x (\Box_{\mathbf{EA}+\psi} \varphi(\dot{x}) \rightarrow \varphi(x)) \\ &\vdash \forall x \varphi(x), \quad \text{by (13).} \end{aligned}$$

Consequently,

$$\mathbf{EA} + \text{Rfn}_{\Sigma_{n+2}}^n(\mathbf{EA}) \vdash \psi \rightarrow \forall x \varphi(x),$$

as required.

For a proof of (\supseteq) notice that, for any Π_{n+2} -sentence φ , the theory $I\Pi_{n+1}^- + \varphi$ contains $\mathbf{EA} + \varphi + \Pi_{n+1}\text{-IR}$, by Lemma 4.6. Hence, by Theorem 8,

$$I\Pi_{n+1}^- + \varphi \vdash \langle n \rangle_{\mathbf{EA}} \varphi.$$

Thus,

$$I\Pi_{n+1}^- \vdash \varphi \rightarrow \langle n \rangle_{\mathbf{EA}} \varphi,$$

which yields the statement of the theorem.

Theorem 10 allows to apply the machinery developed in §2.5 to the study of parameter-free induction. Since both the usual and the relativized provability formulas satisfy Löb's conditions and, in fact, share the same provability logic, all the results on local reflection principles proved so far generalize to the relativized schemata.

The analog of Proposition 2.33 is as follows.

Lemma 4.15. *Any m instances of $\text{Rfn}^n(T)$ are Π_{n+1} -conservative over T_m^n .*

Proof. Everywhere in the proof of Proposition 2.33 one should read $[n]_T$ in place of \Box_T . The applicability of Lemma 2.27 is guaranteed by Corollary 2.13. By Lemma 2.9, iterated n -consistency assertions correspond to iterated uniform Π_{n+1} -reflection schemata over T .

The following corollary comes from [22, 24].

Proposition 4.16. *For each $n \geq 1$,*

- (i) *Any m instances of III_{n+1}^- are Π_{n+1} -conservative over EA_m^n .*
- (ii) *III_{n+1}^- is Π_{n+1} -conservative over $EA_\omega^n \equiv I\Sigma_n^R$.*

Proof. Statement (ii) follows from (i). A proof of Statement (i) is based on the previous lemma and a remark that our characterization of parameter-free induction schemata in terms of reflection principles respects the number of instances of the corresponding schemata.

Lemma 4.17. *For any instance B of III_{n+1}^- , there is a Σ_{n+2} -sentence σ such that $[n]_{EA}\sigma \rightarrow \sigma$ implies B over EA . Conversely, for any such σ , there is an instance B of III_{n+1}^- such that $EA + B$ proves $[n]_{EA}\sigma \rightarrow \sigma$.*

Proof. This follows by verification of our proof of Theorem 10. For a proof of the second part of this lemma we use Theorem 8 (i) stating that

$$[EA + \neg\sigma, \Pi_{n+1}\text{-IR}] \vdash \text{RFN}_{\Pi_{n+1}}(EA + \neg\sigma).$$

We also notice that any finite number of *non-nested* applications of $\Pi_{n+1}\text{-IR}$ is reducible to a single application which, in turn, is reducible to a single instance of III_{n+1}^- .

Remark 4.18. A similar statement also holds for $I\Sigma_n^-$, but the second part seems to be only valid over $I\Sigma_{n-1}$. In general, $m + 1$ instances of $I\Sigma_n^-$ are needed in order to derive m instances of $\text{Rfn}_{\Sigma_{n+1}}^n(EA)$ (the first one is used to derive $I\Sigma_{n-1}$).

By Lemma 4.17, any m instances of III_{n+1}^- are contained in an extension of EA by m instances of $\text{Rfn}^n(EA)$, which is Π_{n+1} -conservative over EA_m^n , by Lemma 4.15. This concludes the proof of Proposition 4.16.

Remark 4.19. A statement similar to 4.16 also holds for the schema III_1^- over EA^+ , see Theorem 10 (iii).

We obtain the following interesting corollary [22, 24].

Corollary 4.20. *The classes of provably total computable functions of III_{n+1}^- and of $I\Sigma_n^R$ coincide. In particular, $\mathcal{F}(III_2^-)$ coincides with the set of all primitive recursive functions.*

Proof. By Part (ii) of Proposition 4.16, III_{n+1}^- is a Π_{n+1} -conservative extension of $I\Sigma_n^R$. By Proposition 3.9, $\mathcal{F}(I\Sigma_1^R)$ coincides with the set of primitive recursive functions.

A relativization of Theorem 6 (iii) is as follows.

Lemma 4.21. *For every $n \geq 0$, the theory $T + \text{Rfn}^n(T)$ is conservative over $T + \text{Rfn}_{\Sigma_{n+1}}^n(T)$ for $\mathcal{B}(\Sigma_{n+1})$ -sentences.*

Proof repeats the given proof of Theorem 6.

Corollary 4.22. *For any $n \geq 1$, III_{n+1}^- is conservative over $I\Sigma_n^-$ for $\mathcal{B}(\Sigma_{n+1})$ -sentences.*

One can also infer from Proposition 4.16 the impossibility of finite axiomatization of III_n^- that was earlier established by Kaye, Paris and Dimitracopoulos [46] by model-theoretic methods. We shall obtain a generalization of this result below. To this end let us first state a formally more general version of Proposition 4.16.

Proposition 4.23. *Suppose T is an extension of EA by a finite number of Π_{n+2} -sentences, $n \geq 1$. Then,*

- (i) *the extension of T by any m instances of III_{n+1}^- is Π_{n+1} -conservative over T_m^n ;*
- (ii) *$T + III_{n+1}^-$ is a Π_{n+1} -conservative extension of T_ω^n .*

Proof. For a given theory T , by Theorem 10, we have

$$\begin{aligned} T + III_{n+1}^- &\equiv T + \text{Rfn}_{\Sigma_{n+2}}^n(\text{EA}) \\ &\equiv T + \text{Rfn}_{\Sigma_{n+2}}^n(T), \end{aligned}$$

by Lemma 2.23. By Lemma 4.15, the second statement follows. (The fact that $T + III_{n+1}^-$ contains T_ω^n follows from Corollaries 4.8 and 4.7.) The first statement is similarly obtained from Proposition 4.16 (i).

Corollary 4.24. *No consistent theory extending III_{n+1}^- by a set of Π_{n+2} -sentences is finitely axiomatizable.*

Proof. Assume, towards a contradiction, that such an extension exists. Without loss of generality we may assume that it has the form $T + U$, for some m instances U of the schema III_{n+1}^- , where T is a finite Π_{n+2} -extension of EA . Then, by Proposition 4.23 (i), the Π_{n+1} -consequences of $T + U$ coincide with the Π_{n+1} -consequences of T_m^n . However, by 4.23 (ii),

$$T + III_{n+1}^- \vdash \text{RFN}_{\Pi_{n+1}}(T_m^n).$$

The latter reflection schema is equivalent to a Π_{n+1} -formula and is unprovable in T_m^n .

We also notice the following corollary of the unboundedness theorems for local reflection principles.

Proposition 4.25. III_{n+1}^- is not contained in any consistent r.e. extension of EA by a set of Π_{n+2} -sentences.

Proof. By Theorem 10, III_{n+1}^- contains $\text{Rfn}_{\Sigma_{n+2}}^n(\text{EA})$ and, thus, also the weaker schema $\text{Rfn}_{\Sigma_{n+2}}(\text{EA})$. The result follows by Corollary 2.22.

Corollary 4.26. $III_{n+1}^- \not\subseteq I\Sigma_{n+1}^R$.

Notice that the complexity of a natural axiomatization of III_{n+1}^- is Σ_{n+2} and the complexity of $I\Sigma_n^-$ is $\mathcal{B}(\Sigma_{n+1})$. We have the following version of the unboundedness theorems for $\text{Rfn}_{\Sigma_{n+1}}^n(T)$.

Lemma 4.27. $\text{Rfn}_{\Sigma_{n+1}}^n(T)$ is not contained in any consistent extension of T by $\mathcal{B}(\Sigma_{n+1})$ -sentences.

Proof. An obvious relativization of Lemma 2.26 shows that the schemata $\text{Rfn}_{\Sigma_{n+1}}^n(T)$ and $\text{Rfn}_{\mathcal{B}(\Sigma_{n+1})}^n(T)$ are deductively equivalent over EA. If the latter is contained in $T + \varphi$, where $\varphi \in \mathcal{B}(\Sigma_{n+1})$, then $T + \varphi \vdash [n]_T \neg\varphi \rightarrow \neg\varphi$, and hence $T \vdash \Box_T \neg\varphi \rightarrow \neg\varphi$. Löb's theorem yields $T \vdash \neg\varphi$.

As a corollary we obtain the following proposition.

Proposition 4.28. $I\Sigma_n^-$ is not contained in any consistent extension of EA by $\mathcal{B}(\Sigma_{n+1})$ -sentences.

Corollary 4.29. Any consistent $\mathcal{B}(\Sigma_{n+1})$ -extension of $I\Sigma_n^-$ is not finitely axiomatizable.

Proof. This follows from Proposition 4.28 and the fact that $I\Sigma_n^-$ has a $\mathcal{B}(\Sigma_{n+1})$ -axiomatization.

The following corollary was obtained by a different method in [46].

Corollary 4.30. For any $n \geq 1$, the theories $I\Sigma_n^-$ and III_n^- are not finitely axiomatizable.

Proof. The claim follows from Corollaries 4.29 and 4.24. For the case of III_1^- it is sufficient to note that if this theory is finitely axiomatizable, so is $\text{EA}^+ + III_1^-$, which is impossible in view of Remark 4.19.

4.5 Reflection schemata and reflection rules

In this section we establish a connection between the schemata and the rules of reflection. We show that, under certain conditions, $\text{RFN}_{\Pi_{n+2}}(T)$ is a Π_{n+1} -conservative extension of the closure of T under the rule $\Pi_{n+1}\text{-RR}^n(T)$, that is of the theory T_ω^n .

Schmerl [77] proved this result for $T = \text{PRA}$. However, it will be important for the following that this statement holds under a substantially more general assumptions on T . We shall show that such a generalization yields Parsons' theorem on partial conservativity of induction schemata over induction rules

and a theorem of Kaye, Paris and Dimtracopoulos on partial conservativity of parametric induction schemata over the parameter-free ones. We also obtain a corollary on Σ_2 -conservativity of $\text{RFN}_{\Sigma_1}(T)$ over $\text{Rfn}_{\Sigma_1}(T)$.

The theorem formulated below [25] also plays an important role in the next paragraph, where it provides — in combination with the techniques of provability algebras — a basis for a new consistency proof for PA in the style of Gentzen and for a characterization of provably total computable functions in PA .

Theorem 11. *Let T be an elementary presented theory containing EA , and let U be any Π_{n+2} -extension EA . Then $U + \text{RFN}_{\Sigma_{n+1}}(T)$ is Π_{n+1} -conservative over $U + \Pi_{n+1}\text{-RR}^n(T)$.*

Proof. For any term s and a formula $\varphi(a)$ let $\ulcorner \psi(\dot{s}) \urcorner$ denote the result of substituting s for a in $\ulcorner \psi(\dot{a}) \urcorner$.

Lemma 4.31. *For any term $s(\vec{x})$, where \vec{x} exhausts all the variables of s , and any $\varphi(a)$ (where s is substitutable for a in φ)*

$$\text{EA} \vdash \forall \vec{x} \left(\Box_T \varphi(s(\vec{x})) \leftrightarrow \Box_T \varphi(\dot{s}) \right).$$

Proof. Obviously,

$$\text{EA} \vdash s(\vec{x}) = y \rightarrow (\varphi(s(\vec{x})) \leftrightarrow \varphi(y)).$$

Hence, by provable Σ_1 -completeness and Löb's derivability conditions,

$$\begin{aligned} \text{EA} \vdash s(\vec{x}) = y &\rightarrow \Box_T(s(\vec{x}) = \dot{y}) \\ &\rightarrow \Box_T(\varphi(s(\vec{x})) \leftrightarrow \varphi(\dot{y})) \\ &\rightarrow (\Box_T \varphi(s(\vec{x})) \leftrightarrow \Box_T \varphi(\dot{y})) \end{aligned} \tag{14}$$

On the other hand, by the definition of $\ulcorner \varphi(\dot{s}) \urcorner$,

$$\text{EA} \vdash s(\vec{x}) = y \rightarrow (\Box_T \varphi(\dot{s}) \leftrightarrow \Box_T \varphi(\dot{y})),$$

which together with (14) yields the result.

We shall show that the standard cut-elimination procedure can be viewed as a reduction of $\text{RFN}_{\Sigma_{n+1}}(T)$ to $\Pi_{n+1}\text{-RR}^n(T)$. Consider a cut-free proof of a sequent of the form

$$\neg U, \neg \text{RFN}_{\Sigma_{n+1}}(T), \Pi, \tag{15}$$

where Π is a set of Π_{n+1} -formulas, $\neg U$ is a finite set of negated axioms of U , and $\neg \text{RFN}_{\Sigma_{n+1}}(T)$ is a finite set of negated instances of $\text{RFN}_{\Sigma_{n+1}}(T)$ of the form

$$\exists y \exists x [\text{Prf}_T(y, \ulcorner \neg \varphi(\dot{x}) \urcorner) \wedge \varphi(x)],$$

for some $\varphi(x) \in \Pi_{n+1}$. Let $R_\varphi(x, y)$ denote the formula in square brackets. We can also assume the axioms of U to have the form $\forall x_1 \dots \forall x_m \neg A(x_1, \dots, x_m)$, for some Π_{n+1} -formulas $A(\vec{x})$.

By the subformula property, any formula occurring in a derivation of a sequent Γ of the form (15) either (a) is a Π_{n+1} -formula, or (b) has the form $\neg\text{RFN}_{\Sigma_{n+1}}(T)$, $\exists x R_\varphi(t, x)$ or $R_\varphi(t, s)$, for some terms s, t , or (c) has the form

$$\exists x_{i+1} \dots \exists x_n A(t_1, \dots, t_i, x_{i+1}, \dots, x_n),$$

for some $i < m$ and terms t_1, \dots, t_i . Let Γ^- denote the result of deleting all formulas of types (b) and (c) from Γ .

Lemma 4.32. *If a sequent Γ of the form (15) is cut-free provable, then $\bigvee \Gamma^-$ is provable in $U + \Pi_{n+1}\text{-RR}^n(T)$.*

Proof goes by induction on the height d of a derivation of Γ . It is sufficient to consider the case that a formula of type (b) or (c) is introduced by the last application of a rule in d ; besides, it is sufficient to only consider the formulas $R_\varphi(t, s)$ and $\exists x_m A(t_1, \dots, t_{m-1}, x_m)$ because in all the other cases the premise and the conclusion of the rule coincide after applying the operation $(\cdot)^-$.

Thus, let us assume that a derivation d has the form

$$\frac{\text{Prf}_T(t, \ulcorner \neg \varphi(\dot{s}) \urcorner), \Delta \quad \varphi(s), \Delta}{R_\varphi(t, s), \Delta} (\wedge)$$

where $\varphi \in \Pi_{n+1}$. Then, by the induction hypothesis, we obtain some derivations in $U + \Pi_{n+1}\text{-RR}^n(T)$ of the formulas

$$\text{Prf}_T(t, \ulcorner \neg \varphi(\dot{s}) \urcorner) \vee \bigvee \Delta^- \quad (16)$$

and

$$\varphi(s) \vee \bigvee \Delta^-. \quad (17)$$

Since Δ^- consists of Π_{n+1} -formulas, $\Pi_{n+1}\text{-RR}^n(T)$ is applicable to (17) and we obtain a derivation of

$$\langle n \rangle_T (\varphi(s(\vec{x})) \vee \bigvee \Delta^-(\vec{x})).$$

Using Löb's conditions we successively derive

1. $\langle n \rangle_T \bigvee \Delta^-(\vec{x}) \vee \langle n \rangle_T \varphi(s(\vec{x}))$
2. $\langle n \rangle_T (\bigvee \Delta^-(\vec{x}) \rightarrow \bigvee \Delta^-(\vec{x}))$ (by provable Σ_{n+1} -completeness)
3. $\bigvee \Delta^-(\vec{x}) \vee \langle n \rangle_T \varphi(s(\vec{x}))$ (by 1, 2 and Löb's conditions)
4. $\bigvee \Delta^-(\vec{x}) \vee \langle n \rangle_T \varphi(\dot{s})$ (by Lemma 4.31).

On the other hand, replacing t by an existential quantifier and weakening \Box_T to $[n]_T$ from (16) we obtain

$$\bigvee \Delta^-(\vec{x}) \vee [n]_T \neg \varphi(\dot{s}).$$

Together with 4, by the rule of cut, this yields a derivation of $\bigvee \Delta^-$ in $U + \Pi_{n+1}\text{-RR}^n(T)$.

If the last application of a rule in d has the form

$$\frac{A(t_1, \dots, t_{m-1}, t_m), \Delta}{\exists x_n A(t_1, \dots, t_{m-1}, x_m), \Delta} (\exists)$$

then by the induction hypothesis we obtain a derivation of

$$A(t_1, \dots, t_{m-1}, t_m) \vee \bigvee \Delta^-$$

in $U + \Pi_{n+1}\text{-RR}^n(T)$. From this formula we infer

$$\exists x_1 \dots \exists x_m A(x_1, \dots, x_m) \vee \bigvee \Delta^-,$$

whence $\bigvee \Delta^-$ follows by an application of the cut-rule with the axiom $\forall x_1 \dots \forall x_n \neg A(x_1, \dots, x_m)$ of U .

Corollary 4.33. *If T is a Π_{n+2} -extension of \mathbf{EA} , then $T + \mathbf{RFN}_{\Sigma_{n+1}}(T)$ is a Π_{n+1} -conservative extension of T_ω^n .*

We note that this statement was obtained by a different method for $T = \mathbf{PRA}$ in [77] and for $n = 1$ and $T = \mathbf{EA}$ in [19]. If $T = \mathbf{EA}$, it yields the following result due to Parsons [66].

Proposition 4.34. *For $n \geq 1$, $I\Sigma_n$ is Π_{n+1} -conservative over $I\Sigma_n^R$. In particular, $I\Sigma_1$ is Π_2 -conservative over $I\Sigma_1^R \equiv \mathbf{PRA}$, and $\mathcal{F}(I\Sigma_1)$ is the class of all primitive recursive functions.*

We also have a stronger conservation result for parameter-free schemata.

Lemma 4.35. *If U is a Π_{n+2} -extension of \mathbf{EA} , then $U + \mathbf{RFN}_{\Sigma_{n+1}}(T)$ is a Σ_{n+2} -conservative extension of $U + \mathbf{Rfn}_{\Sigma_{n+1}}^n(T)$.*

In particular, for $n = 0$ and $U = \mathbf{EA}$ we obtain that $\mathbf{RFN}_{\Sigma_1}(T)$ is Σ_2 -conservative over $\mathbf{Rfn}_{\Sigma_1}(T)$.

Proof. Assume $U + \mathbf{RFN}_{\Sigma_{n+1}}(T) \vdash \sigma$, with a sentence $\sigma \in \Sigma_{n+2}$. Then $U + \neg\sigma + \mathbf{RFN}_{\Sigma_{n+1}}(T) \vdash \perp$ and, by Theorem 11, $U + \neg\sigma + \Pi_{n+1}\text{-RR}^n(T) \vdash \perp$. From this we conclude that $U + \neg\sigma + \mathbf{Rfn}_{\Sigma_{n+1}}^n(T) \vdash \perp$ and, by the deduction theorem, $U + \mathbf{Rfn}_{\Sigma_{n+1}}^n(T) \vdash \sigma$.

In view of Theorem 10, this immediately implies the following corollary, first proved by Kaye, Paris and Dimitracopoulos [46] using model-theoretic methods.

Proposition 4.36. *For $n \geq 1$, $I\Sigma_n$ is a Σ_{n+2} -conservative extension of $I\Sigma_n^-$.*

To close this section we mention a few facts concerning the diagram of fragments of \mathbf{PA} on Fig. 3.1.

Thus, $I\Pi_{n+1}^- \not\subseteq I\Sigma_{n+1}^R$ follows from Corollary 4.26. $I\Sigma_n \not\subseteq I\Pi_{n+1}^-$ follows from the fact that $I\Pi_{n+1}^-$ has a Σ_{n+2} -axiomatization, whereas $I\Sigma_n$ proves $\mathbf{RFN}_{\Pi_{n+2}}(\mathbf{EA})$, by Theorem 7. $I\Sigma_{n+1}^R \not\subseteq I\Sigma_n + I\Pi_{n+1}^-$ follows from the fact that $I\Sigma_n + I\Pi_{n+1}^-$ is an extension of $I\Sigma_n$ by a set of Σ_{n+2} -sentences, whereas $I\Sigma_{n+1}^R$ contains $\mathbf{RFN}_{\Pi_{n+2}}(I\Sigma_n)$. Thus, all inclusions corresponding to the edges of the diagram are strict.

5 Provability algebras and ordinal analysis of PA

In the previous paragraph we have shown that various fragments of PA are expressible, in one way or another, over the elementary arithmetic. General properties of reflection principles allowed us then to easily establish various characteristics of the theories in question. Thus, the structure of reflection schemata over a theory T plays the role of a “map” and a useful tool to study the extensions of T .

The structure of reflection principles over T , as well as a more comprehensive structure of arbitrary extensions of T , can be understood from a more general algebraic point of view. This point of view happens to be fruitful for further applications of reflection principles to ordinal analysis and to general questions of classification of theories. We shall begin with a well-known concept of the Lindenbaum algebra of a theory.

5.1 Lindenbaum algebras

Given a theory T , consider the set \mathcal{L}_T of all arithmetical sentences factorized by the relation of provable equivalence in T :

$$\varphi \sim \psi \iff T \vdash \varphi \leftrightarrow \psi.$$

The equivalence class of φ is denoted $\{\varphi\}$. Logical operations \wedge, \vee, \neg , together with the order relation

$$\{\varphi\} \leq \{\psi\} \iff T \vdash \varphi \rightarrow \psi,$$

provide this set with the structure of a boolean algebra, which is called the Lindenbaum algebra of T . The maximal \top and the minimal \perp elements of the algebra are, respectively, the classes of all provable and all refutable sentences of T . The remaining elements of the Lindenbaum algebra are the equivalence classes of T -independent sentences. Deductively closed extensions of T correspond to *filters* of the boolean algebra \mathcal{L}_T .

Recall that a boolean algebra \mathcal{B} is called *atomless* if

$$\forall x, y \in \mathcal{B} (x < y \Rightarrow \exists z x < z < y).$$

Rosser’s theorem implies the following statement.

Proposition 5.1. *Let T be a consistent r.e. theory. Then \mathcal{L}_T is a countable atomless boolean algebra.*

Proof. Consider φ and ψ such that $\{\varphi\} < \{\psi\}$, that is, $T \vdash \varphi \rightarrow \psi$ and $T \not\vdash \psi \rightarrow \varphi$. Then the theory $U = T + \psi + \neg\varphi$ is consistent. By Theorem 3, there is a sentence θ independent from U . Then one can take the class $\{\varphi \vee (\theta \wedge \psi)\}$ for z .

A basic result in the theory of boolean algebras (cf. [9]) states that all countable atomless boolean algebras are pairwise isomorphic.

Corollary 5.2. *If T and U are consistent r.e. theories, then $\mathcal{L}_T \simeq \mathcal{L}_U$.*

In particular, the algebras \mathcal{L}_{EA} , \mathcal{L}_{PA} , \mathcal{L}_{ZF} are all pairwise isomorphic. Actually, a stronger statement holds.

\mathcal{L}_T can be considered as a (positively) numerated structure, where the Gödel numbers of formulas play the role of the codes of elements of the algebra (cf. [1]). Kripke and Pour-El [69] proved that all Lindenbaum algebras for consistent r.e. arithmetical theories, considered as numerated algebras, are recursively isomorphic. De facto, this statement follows from the effectiveness of Rosser's theorem.

5.2 Provability algebras and logics

The result on the isomorphism stated above shows that the Lindenbaum algebra of a theory does not really bear any proof-theoretically interesting information on its “strength.” A question arises if the structure of Lindenbaum algebra can be enriched in such a way that such properties would become expressible.

Tarski (cf. [44]) has introduced the notion of cylindric algebra, which allows one to express the quantifiers on the algebraic language. However, since the language of cylindric algebras is expressively very strong, the resulting structures are complex and, seemingly, do not provide significant advantages for proof theory compared to the more traditional language of first order logic.

Another expansion of Lindenbaum algebras was suggested by Magari [56]. Given an elementary presented theory T , Gödel's provability formula induces an operator

$$\Box_T : \{\varphi\} \longmapsto \{\Box_T \varphi\},$$

acting on \mathcal{L}_T . By Löb's conditions, this operator is correctly defined on the equivalence classes. The structure $\mathcal{M}_T = (\mathcal{L}_T, \Box_T)$ is called the *provability algebra*, or the *Magari algebra*, for T .⁴

The language of Magari algebras allows us to express many interesting facts about provability, for example, Gödel's second incompleteness theorem:

$$\mathcal{M}_T \models \forall x (x \neq \perp \Rightarrow x \not\leq \neg \Box_T \neg x).$$

Notice that the terms in the language of Magari algebras can be identified with propositional modal formulas. Accordingly, Solovay's theorem describes all terms $\varphi(\vec{x})$ such that the identity $\varphi(\vec{x}) = \top$ holds in \mathcal{M}_T :

$$\mathbf{GL} \vdash \varphi(\vec{x}) \iff \mathcal{M}_T \models \forall \vec{x} \varphi(\vec{x}) = \top.$$

An arbitrary identity $\psi(\vec{x}) = \theta(\vec{x})$ of a Magari algebra reduces to the one of the form $\varphi(\vec{x}) = \top$, where $\varphi = (\psi \leftrightarrow \theta)$. Thus, from an algebraic point of view, Solovay's theorem provides a characterization of the set of identities of provability algebras for Σ_1 -sound theories T .

A boolean algebra with a unary operator \Box satisfying the identities

$$\text{M1. } \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi) = \top,$$

⁴Magari himself called such algebras *diagonalizable*.

M2. $\Box\varphi \rightarrow \Box\Box\varphi = \top$,

M3. $\Box(\Box\varphi \rightarrow \varphi) \rightarrow \Box\varphi = \top$,

M4. $\Box\top = \top$,

and thus all the theorems of **GL**, is called a *Magari algebra*. By Solovay's theorem, the set of Magari algebras is a variety generated by an algebra \mathcal{M}_T , for any Σ_1 -sound theory T , for example, **PA**.

General properties of Magari algebras were actively investigated in the context of provability logic (cf. [56, 57, 60, 61, 62, 86, 13]). From the point of view of the study of provability algebras of arithmetical theories the most important results were obtained by Shavrukov (cf. [81, 80, 83, 82], and also [95]). In particular, Shavrukov gave a description of all possible subalgebras of \mathcal{M}_T . He also obtained an important result on the undecidability of the elementary (first order) theory of \mathcal{M}_T . We note that the decidability of the purely universal theory of this structure follows from the decidability of the truth provability logic **S** (cf. [13]), whereas the question whether the $\forall^*\exists^*$ -theory is decidable remains a difficult open problem. Shavrukov also established a number of interesting results on the isomorphism of provability algebras. In particular, he showed that the algebras \mathcal{M}_T and \mathcal{M}_U are not isomorphic if $U \vdash \text{RFN}_{\Sigma_1}(T)$.⁵ The question of isomorphism of the provability algebras of theories such as **PA** and **PA** + **Con**(**PA**) is still open.

5.3 Graded provability algebras and Japaridze's logic

Gödel's provability formula has arithmetical complexity Σ_1 . At the same time, the theories we are interested in (such as $I\Sigma_1$ or **PA**) often have a higher, or even unbounded, level of arithmetical complexity of their axioms. This suggests that the language of Magari algebras is still not expressive enough for a representation and a fruitful study of such theories.

The n -provability formulas $[n]_T$ and the corresponding operators on \mathcal{M}_T provide a needed generalization. This leads to a notion of *graded provability algebra* introduced in [26]. In the previous paragraph we have seen that already the simple use of $[n]_T$ as a modality of **GL** allows to obtain a number of nontrivial results on parameter-free induction schemata. Jointly using several provability operators allows for a deeper analysis of arithmetical theories and yields such results as a new consistency proof for Peano arithmetic by transfinite induction up to the ordinal ε_0 in the style of Gentzen and a characterization of provably total computable functions of **PA**. It also leads to simple combinatorial statements independent from Peano arithmetic.

Each formula $[n]_T$ induces an operator

$$[n]_T : \{\varphi\} \longmapsto \{[n]_T\varphi\}$$

⁵Shavrukov considered a somewhat more restrictive condition, which is equivalent to this one in view of certain results in [20].

acting on the Lindenbaum algebra of T . The structure $\mathcal{M}_T^\infty = (\mathcal{L}_T, [0]_T, [1]_T, \dots)$ will be called the *graded provability algebra* of T . Sometimes we shall also consider \mathcal{M}_T^∞ as a many-sorted algebra, that is, a structure with a distinguished family of subsets

$$\Pi_1 \subset \Pi_2 \subset \Pi_3 \subset \dots \subset \mathcal{M}_T^\infty,$$

defined by the corresponding classes of the arithmetical hierarchy.

We consider $\mathcal{M}_{\text{EA}}^\infty$ as the most important example. The familiar fragments of PA correspond to the filters of this algebra generated by some specific elements. Thus, by the results of the previous paragraph, $I\Sigma_n$ corresponds to the element $\langle n+1 \rangle_{\text{EA}} \top$ and the associated principal filter.

The rules of Σ_n - and Π_{n+1} -induction correspond to an operator $p \mapsto \langle n \rangle_{\text{EA}} p$, for $p \in \Pi_{n+1}$ and $p \in \Pi_{n+2}$, respectively. Therefore, $I\Pi_{n+1}^R$ is a filter generated from \top by this operator, that is, the upwards closure of $\{\langle n \rangle_{\text{EA}}^k \top : k < \omega\}$.

Parameter-free induction schemata are sets of elements of $\mathcal{M}_{\text{EA}}^\infty$ of the form $p \rightarrow \langle n \rangle_{\text{EA}} p$, where $p \in \Pi_{n+1}$ and $p \in \Pi_{n+2}$, respectively.

Terms of \mathcal{M}_T^∞ are identified with propositional *polymodal formulas* in a language with the modalities $[0], [1], \dots$. The identities of this algebra are exactly characterized by the following system **GLP** introduced by Japaridze [7, 6].

Axioms: J1. Axioms of **GL** for each modality $[n]$;

J2. $[m]\varphi \rightarrow [n]\varphi$, for $m \leq n$;

J3. $\langle m \rangle \varphi \rightarrow [n]\langle m \rangle \varphi$, for $m < n$.

Inference rules: modus ponens, $\varphi \vdash [n]\varphi$.

Arithmetical interpretation $(\varphi)_T^*$ of a polymodal formula φ is defined as usual, that is, for all n , $[n]$ is translated as the n -provability formula:

$$([n]\varphi)_T^* = [n]_T(\varphi)_T^*.$$

Lemma 5.3. *If $\text{GLP} \vdash \varphi$, then $\text{EA} \vdash (\varphi)_T^*$, for any realization $(\cdot)^*$ of the variables of φ .*

Proof. The soundness of all axioms of **GL** follows from Löb's conditions and Löb's theorem for the n -provability formulas. The validity of (J3) follows from the provable Σ_{n+1} -completeness. Axiom (J2) follows by formalizing in EA a triviality that any (true) Π_n -formula is a (true) Π_{n+1} -formula.

By analogy with the Magari algebras we call *Japaridze algebras* any boolean algebras with operators $[0], [1], \dots$ satisfying all the identities of **GLP**. Japaridze [7, 6] obtained the following deep generalization of Solovay's arithmetical completeness theorem.

Theorem 12. *If T is sound, then for every polymodal formula φ ,*

$$\text{GLP} \vdash \varphi(\vec{x}) \iff \mathcal{M}_T^\infty \models \forall \vec{x} (\varphi(\vec{x}) = \top).$$

Therefore, the variety of Japaridze algebras is generated by \mathcal{M}_T^∞ , for any sound theory T .

Originally, Japaridze formulated his result for a somewhat different interpretation of modalities $[n]$. The history is as follows.

Boolos [30] investigated the modal logic of ω -provability (the notion dual to Gödel's notion of ω -consistency) and observed that Solovay's theorems can be transferred to this case without any significant changes. ω -provability can be described as provability in arithmetic by non-nested applications of ω -rule, that is, provability in the theory

$$T' := T + \{\forall x \varphi(x) : \forall n \ T \vdash \varphi(\bar{n})\}.$$

Therefore, the logic of ω -provability coincides with **GL**.

Japaridze [6] obtained a more interesting characterization of the bimodal logic of provability and ω -provability for **PA**. In fact, he formulated the polymodal logic **GLP** with the interpretation of $[1]$, $[2]$, etc. as the provability in **PA'**, **PA''**, etc.

Later, Ignatiev [45] simplified Japaridze's work and thoroughly investigated modal-logical properties of **GLP**. He observed that Japaridze's theorem also held under more general assumptions than originally stated. In particular, **GLP** was proved to be arithmetically complete for the n -provability interpretation and, more generally, for the interpretation of $[n]$ as arithmetical predicates satisfying some sufficiently broad assumptions. Graded provability algebras were introduced in [26], where their relationship with the ordinal analysis of Peano arithmetic was established.

Japaridze's logic is decidable and can be studied using Kripke semantics. Strictly speaking, **GLP** is not Kripke complete, however it has a simple translation into a weaker logic **GLP**[−], obtained from **GLP** by replacing axioms (J2) by

$$[m]\varphi \rightarrow [n][m]\varphi, \quad \text{for } m \leq n.$$

GLP[−] is already sound and complete with respect to a nice class of finite Kripke frames. Ignatiev has also obtained a normal form theorem for the closed formulas of **GLP** which play a significant role in what follows. A detailed treatment of Japaridze's logic is given in [31, 45]. In this paper we shall only use the arithmetical soundness of **GLP** expressed by Lemma 5.3.

5.4 Reduction property

Theorem 11 proved in §4.5 expresses an important property of graded provability algebras. Let U and V be subsets of \mathcal{M}_T^∞ . We write $U \equiv_n V$ if the filters generated by U and V (i.e., the theories containing T) are mutually Π_{n+1} -conservative.

Proposition 5.4. *Assume T is a Π_{n+2} -extension of **EA**. Then, for all $\varphi \in \mathcal{M}_T^\infty$,*

$$\{\langle n+1 \rangle_T \varphi\} \equiv_n \{Q_k^n(\varphi) : k < \omega\},$$

where

$$\begin{aligned} Q_0^n(\varphi) &= \langle n \rangle_T \varphi, \\ Q_{k+1}^n(\varphi) &= \langle n \rangle_T (Q_k^n(\varphi) \wedge \varphi). \end{aligned}$$

Proof. The statement follows if one takes $U = T$ and $T = T + \varphi$ in the formulation of Theorem 11. Notice that the rule $\Pi_{n+1}\text{-RR}^n(T + \varphi)$ is equivalent to

$$\frac{\psi}{\langle n \rangle_T (\varphi \wedge \psi)}, \quad \psi \in \Pi_{n+1}.$$

The theory $T + \{Q_k^n(\varphi) : k < \omega\}$ is the closure of \top under this rule. Indeed, if $T + Q_k^n(\varphi) \vdash \psi$, then

$$\begin{aligned} T + Q_{k+1}^n(\varphi) &\vdash \langle n \rangle_T (\varphi \wedge Q_k^n(\varphi)) \\ &\vdash \langle n \rangle_T (\varphi \wedge \psi). \end{aligned}$$

Therefore, all Π_{n+1} -consequences of an element $\langle n+1 \rangle_T \varphi \in \mathcal{M}_T^\infty$ of complexity Π_{n+2} are generated by Π_{n+1} -elements $Q_k^n(\varphi)$. We call this property of \mathcal{M}_T^∞ *the reduction property*. It will be important for us that the elements $Q_k^n(\varphi)$ are expressible in terms of the operations of \mathcal{M}_T^∞ .

As an example, consider $\varphi = \top$ in the formulation of Proposition 5.4. Then we obtain

$$\{\langle n+1 \rangle_T \top\} \equiv_n \{\langle n \rangle_T^{k+1} \top : k < \omega\}.$$

It is easy to see that this statement expresses Corollary 4.33 in the language of graded provability algebras.

Corollary 5.5. *If $m \leq n$, then*

$$\text{EA}^+ \vdash \langle m \rangle \langle n+1 \rangle \varphi \leftrightarrow \forall k \langle m \rangle Q_k^n(\varphi).$$

Proof. In $\mathcal{M}_{\text{EA}}^\infty$, by the reduction property,

$$\{\langle n+1 \rangle \varphi\} \equiv_n \{Q_k^n(\varphi) : k < \omega\},$$

and this equivalence is established within EA^+ . Hence, a Π_{m+1} -sentence follows from $\langle n+1 \rangle \varphi$ if and only if it follows from $Q_k^n(\varphi)$, for some k . Therefore, the statements of Π_{m+1} -soundness for the theories on the left and on the right hand side are equivalent.

5.5 Closed formulas of GLP and a system of ordinal notation for ε_0

The collection of closed fomrulas of **GLP** (modulo provable equivalence) can also be described as the free 0-generated Japaridze algebra \mathcal{M}_0^∞ , that is, the free algebra generated from the constants \top and \perp by the operators \wedge , \neg and $\langle n \rangle$, for all n , and satisfying the identities of **GLP**. Notice that, by Japaridze's theorem, \mathcal{M}_0^∞ is naturally embedded in \mathcal{M}_T^∞ , if T is sound, and is its minimal subalgebra:

$$\mathcal{M}_0^\infty \models \varphi = \psi \iff \mathcal{M}_T^\infty \models (\varphi)_T^* = (\psi)_T^*.$$

The structure of \mathcal{M}_0^∞ was studied by Ignatiev [45], who has obtained a normal form theorem for the closed fragment of **GLP** and an elementary decision procedure for it. We state a simplified but a somewhat weaker version of that theorem.

Closed formulas of **GLP** play a central role in the ordinal analysis of Peano arithmetic because they can be considered as finite representations of ordinals up to ε_0 .

Recall that ε_0 is defined as $\sup\{\omega, \omega^\omega, \omega^{\omega^\omega}, \dots\}$, where ω is the order type of the set of natural numbers. By the Cantor normal form theorem any ordinal $\alpha < \varepsilon_0$, $\alpha \neq 0$ can be uniquely represented as $\alpha = \omega^{\alpha_1} + \dots + \omega^{\alpha_n}$, where $\alpha > \alpha_1 \geq \dots \geq \alpha_n$. Similarly, decomposing $\alpha_1, \dots, \alpha_n$ and all the smaller ordinals that appear in the process we obtain that any $\alpha < \varepsilon_0$ can be represented as a term built up from the constant 0 and the functions $+$ and ω^x . We call such a term a *representation* of α . We say that a representation $\alpha = \omega^{\alpha_1} + \dots + \omega^{\alpha_n}$ is in *normal form* if $\alpha_1 \geq \dots \geq \alpha_n$ and all subterms α_i are also represented in a normal form. A representation of $\alpha < \varepsilon_0$ in a normal form is unique.

It is well known that the relation *to be a representation in a normal form* and the ordering $\alpha < \beta$, as a relation on the Gödel numbers of terms, are elementary and naturally definable in **EA**. The structure

$$(\varepsilon_0; 0, <, +, \omega^x),$$

together with the numbering described above, is called *the standard system of ordinal notation for ε_0* .

Let S denote the set of formulas generated from \top by $\langle 0 \rangle, \langle 1 \rangle, \dots$. Elements of S have the form

$$\alpha = \langle n_1 \rangle \langle n_2 \rangle \dots \langle n_k \rangle \top.$$

We identify such formulas with words in the alphabet \mathbb{N} :

$$\alpha = n_1 n_2 \dots n_k.$$

The empty word Λ is identified with \top . Let S_n denote the restriction of S to the alphabet $\{n, n+1, \dots\}$.

For each n define an ordering $<_n$ on S_n by

$$\beta <_n \alpha \iff \mathbf{GLP} \vdash \alpha \rightarrow \langle n \rangle \beta.$$

It is easy to see that $<_n$ is a transitive relation. This relation is also irreflexive because $\mathbf{GLP} \vdash \alpha \rightarrow \langle n \rangle \alpha$ implies $\mathbf{GLP} \vdash \neg \alpha$, by Löb's axiom. But then, the arithmetical interpretation $(\alpha)_{\mathbf{EA}}^*$ is refutable in **EA**, hence it is false contradicting the soundness of **EA**.

This argument relies on a strong assumption of soundness of **EA**, which is provable in **PA** but not in **EA**. A proof of irreflexivity of $<_n$ formalizable in **EA** can be given on the basis of Ignatiev's techniques of Kripke models for **GLP**, however we shall not need it here. The same remark concerns the following proposition established in [26].

Proposition 5.6. $(S_n, <_n)$ is well-founded of height ε_0 . Modulo provable equivalence in **GLP** the ordering is linear.

Since the closed fragment of **GLP** has an elementary decision procedure, the ordering $(S, <_0)$ is also elementary. Therefore, one can consider the elements of S as alternative ordinal notations up to ε_0 . We formulate a simple correspondence between S and the ordinals below ε_0 .

Definition 5.7. Let $o(0^k) = k$. If $\alpha = \alpha_1 0 \alpha_2 0 \cdots 0 \alpha_n$, where all $\alpha_i \in S_1$ and not all of them are empty, then

$$o(\alpha) = \omega^{o(\alpha_n^-)} + \cdots + \omega^{o(\alpha_1^-)}.$$

Here, β^- is obtained from $\beta \in S_1$ replacing every letter $m + 1$ by m .

Similarly, γ^+ will denote the result replacing in γ every letter m by $m + 1$. Notice that $o(\gamma^+) = \omega^{o(\gamma)}$ and the mapping $o : S \rightarrow \varepsilon_0$ is obviously surjective. We also notice that $o(\alpha)$ is a limit ordinal if and only if α begins with an $m > 0$.

Proposition 5.6 can be more explicitly formulated (for the case $n = 0$) as follows: for all $\alpha, \beta \in S$,

$$\begin{aligned} \mathbf{GLP} \vdash \alpha \leftrightarrow \beta &\iff o(\alpha) = o(\beta); \\ \mathbf{GLP} \vdash \beta \rightarrow \Diamond \alpha &\iff o(\alpha) < o(\beta). \end{aligned}$$

For example, consider a word 2101. Then

$$o(2101) = \omega^{o(0)} + \omega^{o(10)} = \omega + \omega^{\omega^0 + \omega^1} = \omega^\omega = o(2).$$

Accordingly, $\mathbf{GLP} \vdash 2101 \leftrightarrow (21 \wedge 01) \leftrightarrow 21 \leftrightarrow 2$. The first of these equivalences follows from Lemma 5.9(iv) proved below.

We mention without proof the following version of Ignatiev's normal form theorem (cf. [45, 26]).

Proposition 5.8. Every closed formula is equivalent in **GLP** to a boolean combination of formulas from S .

5.6 A Gentzen-style consistency proof for PA

Here we give a proof of consistency of PA which only relies, apart from the narrow finitistic means of \mathbf{EA}^+ , on the transfinite induction along ε_0 . The first proof of this kind was obtained by G. Gentzen in his famous papers [40, 41].

The reduction property of graded provability algebras will play a central role in our proof (Proposition 5.4). First, we show that modulo provability in **GLP** the set S is closed under the operations $\langle n + 1 \rangle \alpha \mapsto Q_k^n(\alpha)$.

Lemma 5.9. The following formulas are derivable in **GLP**:

- (i) If $m \leq n$, then $\vdash nm\alpha \rightarrow m\alpha$;
- (ii) If $m < n$, then $\vdash \langle n \rangle \varphi \wedge \langle m \rangle \psi \leftrightarrow \langle n \rangle (\varphi \wedge \langle m \rangle \psi)$;

(iii) If $\alpha \in S_{n+1}$, then $\vdash \alpha\gamma \wedge n\beta \leftrightarrow \alpha(\gamma \wedge n\beta)$;

(iv) If $\alpha \in S_{n+1}$, then $\vdash \alpha \wedge n\beta \leftrightarrow \alpha n\beta$.

Proof. Statement (i) follows from (J2) and axiom $[m]\varphi \rightarrow [m][m]\varphi$ of \mathbf{GL}^m .
Statement (ii):

$$\begin{aligned} \mathbf{GLP} \vdash \langle n \rangle \varphi \wedge \langle m \rangle \psi &\rightarrow [n]\langle m \rangle \psi, \quad \text{by axiom (J3)} \\ &\rightarrow \langle n \rangle (\varphi \wedge \langle m \rangle \psi). \end{aligned}$$

The converse implication follows from (i). Statement (iii) is obtained by applying (ii) several times, and (iv) follows from (iii) for $\gamma = \Lambda$.

Lemma 5.10. (i) If $\alpha = \langle n+1 \rangle \beta$ and $\beta \in S_{n+1}$ then, for all k ,
 $\mathbf{GLP} \vdash Q_k^n(\beta) \leftrightarrow (n\beta)^{k+1}$.

(ii) If $\alpha = \langle n+1 \rangle \beta$, where $\beta = \gamma m\delta$, $\gamma \in S_{n+1}$ and $m \leq n$, then, for all k ,
 $\mathbf{GLP} \vdash Q_k^n(\beta) \leftrightarrow (n\gamma)^{k+1} m\delta$.

Proof. (i) We reason by induction on k . For $k = 0$ we have $Q_0^n(\beta) = \langle n \rangle \beta \in S$.
Induction step:

$$\begin{aligned} \mathbf{GLP} \vdash Q_{k+1}^n(\beta) &\leftrightarrow \langle n \rangle (\beta \wedge (n\beta)^{k+1}) \\ &\leftrightarrow \langle n \rangle (\beta (n\beta)^{k+1}), \quad \text{by Lemma 5.9(iv)} \\ &\leftrightarrow (n\beta)^{k+2}. \end{aligned}$$

(ii) Similarly, the induction step follows from

$$\begin{aligned} \mathbf{GLP} \vdash Q_{k+1}^n(\beta) &\leftrightarrow \langle n \rangle (\gamma m\delta \wedge (n\gamma)^{k+1} m\delta) \\ &\leftrightarrow \langle n \rangle (\gamma (m\delta \wedge (n\gamma)^{k+1} m\delta)) \quad \text{by Lemma 5.9(iii)} \\ &\leftrightarrow (n\gamma)^{k+2} m\delta \quad \text{by Lemma 5.9(i).} \quad \square \end{aligned}$$

Definition 5.11. For a nonempty $\alpha \in S$ and $k \in \mathbb{N}$ define $\alpha[k]$ as follows:

- If $\alpha = 0\beta$ then $\alpha[k] = \beta$.
- If $\alpha = \langle n+1 \rangle \beta$ then $\alpha[k]$ is defined as the expression for $Q_k^n(\beta)$ from the previous lemma.

Corollary 5.12. For any k and a nonempty $\alpha \in S$, $\alpha[k] <_0 \alpha$.

Proof. By induction on k we easily obtain

$$\mathbf{GLP} \vdash \langle n+1 \rangle \beta \rightarrow Q_{k+1}^n(\beta).$$

Hence, by Lemma 5.10 for $\alpha = \langle n+1 \rangle \beta$,

$$\begin{aligned} \mathbf{GLP} \vdash \alpha &\rightarrow \langle n \rangle \alpha[k] \\ &\rightarrow \langle 0 \rangle \alpha[k]. \quad \square \end{aligned}$$

Let $\text{TI}^R(\Pi_1, <_0)$ denote the following inference rule expressing transfinite induction along the ordering $(S, <_0)$ for Π_1 -formulas φ :

$$\frac{\forall \alpha \in S (\forall \beta <_0 \alpha \varphi(\beta) \rightarrow \varphi(\alpha))}{\forall \alpha \in S \varphi(\alpha)}.$$

We shall prove the consistency of PA by one application of this rule over EA^+ .

Theorem 13. $[\text{EA}^+, \text{TI}^R(\Pi_1, <_0)] \vdash \text{Con}(\text{PA})$.

Proof. We shall work in $\mathcal{M}_{\text{EA}}^\infty$. Let α^* denote the arithmetical interpretation $(\alpha)_{\text{EA}}^*$ of $\alpha \in S$. The function $(\cdot)^*$, as a mapping between Gödel numbers of formulas, is elementary and hence definable by a term in EA. We shall write \Diamond instead of \Diamond_{EA} .

Recall that, by Corollary 4.3,

$$\text{PA} \subseteq \text{EA} + \{\langle n \rangle \top : n < \omega\}.$$

A proof of this fact, that is, a derivation of an induction axiom from the correspondent instance of the uniform reflection principle, is easily formalizable in EA, hence we have

$$\text{EA} \vdash \forall n \Diamond(\langle n \rangle \top)^* \rightarrow \text{Con}(\text{PA}). \quad (18)$$

We prove $\forall \alpha \in S \Diamond \alpha^*$ by transfinite induction over EA^+ , where the Π_1 -formula $\varphi(\alpha) = \Diamond \alpha^*$ plays the role of the induction formula (with α the induction variable). We claim that

$$\text{EA}^+ \vdash \forall \alpha \in S (\forall \beta <_0 \alpha \Diamond \beta^* \rightarrow \Diamond \alpha^*).$$

Reasoning in EA^+ assume that $\forall \beta <_0 \alpha \Diamond \beta^*$. Consider two cases:

- If $\alpha = 0\beta$ then $\Diamond \beta^*$, whence $\Diamond \Diamond \beta^*$ using $\text{RFN}_{\Sigma_1}(\text{EA})$ in EA^+ .
- If $\alpha = \langle n+1 \rangle \beta$ then $\forall k \Diamond \alpha[k]^*$ because $\alpha[k] <_0 \alpha$. Notice that, by Corollary 5.5,

$$\text{EA}^+ \vdash \Diamond \alpha^* \leftrightarrow \forall k \Diamond \alpha[k]^*.$$

Hence, $\forall k \Diamond \alpha[k]^*$ implies $\Diamond \alpha^*$.

Thus,

$$\begin{aligned} [\text{EA}^+, \text{TI}^R(\Pi_1, <_0)] &\vdash \forall \alpha \in S \Diamond \alpha^* \\ &\vdash \text{Con}(\text{PA}), \quad \text{by (18),} \end{aligned}$$

as required.

Now we describe how to derive from Theorem 13 the same result for the standard ordinal notations for ε_0 based on the Cantor normal form.

Let a relation $<'_0$ on S be defined by

$$\alpha <'_0 \beta \iff o(\alpha) < o(\beta),$$

where o is the function given in Definition 5.7. As we already know from Proposition 5.6, $<'_0$ coincides with $<_0$ in the real world. However, a proof of this fact within **EA**, which can be given using Kripke models for the closed fragment of **GLP**, is sufficiently complicated. In order to avoid this problem, we use a formalization in **EA** of the relation $<'_0$, whose well-foundedness follows from that of ε_0 .

Lemma 5.13. $\text{TI}^R(\Pi_1, <'_0)$ is reducible to $\text{TI}^R(\Pi_1, \varepsilon_0)$.

Proof. We notice that o , as a function from the Gödel numbers of S to the standard ordinal notations for ε_0 , is elementary. Hence, the induction for $\varphi(\alpha)$ along $(S, <'_0)$ can be reduced to the induction for

$$\psi(x) := \forall \alpha \in S (o(\alpha) = x \rightarrow \varphi(\alpha)),$$

where x is an ordinal below ε_0 .

Lemma 5.14. Let $\alpha = \alpha_0^+ 0 \dots 0 \alpha_k^+$ and $o(\alpha)$ is a limit ordinal. Then

$$o(\alpha \llbracket n \rrbracket) = \begin{cases} \omega^{o(\alpha_k)} + \dots + \omega^{o(\alpha_1)} + \omega^{o(\beta)} \cdot (n+1) + 1, & \text{if } o(\alpha_0) = o(\beta) + 1; \\ \omega^{o(\alpha_k)} + \dots + \omega^{o(\alpha_1)} + \omega^{o(\alpha_0 \llbracket n \rrbracket)}, & \text{if } o(\alpha_0) \text{ is a limit ordinal.} \end{cases}$$

Proof. Since $o(\alpha)$ is a limit ordinal, α_0 is nonempty, that is $\alpha_0^+ = \langle m+1 \rangle \beta$, for some $\beta \in S_1$. $\alpha \llbracket n \rrbracket$ is defined according to one of the following two cases:

- (a) $m = 0$. Then $\alpha \llbracket n \rrbracket = (0\beta)^{n+1} 0 \alpha_1^+ 0 \dots 0 \alpha_k^+$ because $\beta \in S_1$ and β is followed by 0.
- (b) $m > 0$. Then $\alpha \llbracket n \rrbracket = \alpha'_0 0 \alpha_1^+ 0 \dots 0 \alpha_k^+$, where $\alpha'_0 = \alpha_0^+ \llbracket n \rrbracket = (\alpha_0 \llbracket n \rrbracket)^+$.

This implies the statement of the lemma.

Corollary 5.15. For any $\alpha \neq \Lambda$ and any k , $\alpha \llbracket k \rrbracket <'_0 \alpha$.

Now we can repeat the proof of Proposition 13 and derive

$$\text{EA}^+ \vdash \forall \alpha (\forall \beta <'_0 \alpha \diamond \beta^* \rightarrow \diamond \alpha^*).$$

Hence, **Con(PA)** follows by an application of $\text{TI}^R(\Pi_1, <'_0)$ and we obtain the following theorem.

Theorem 14. $[\text{EA}^+, \text{TI}^R(\Pi_1, \varepsilon_0)] \vdash \text{Con(PA)}$.

Remark 5.16. Say that $\alpha \in S$ is in a normal form if so is the representation of the ordinal $o(\alpha)$, that is, if $\alpha = \alpha_0^+ 0 \dots 0 \alpha_k^+$, where all α_i are in a normal form and $o(\alpha_k) \geq \dots \geq o(\alpha_0)$. Words in a normal form constitute a one-to-one notation system for ε_0 . A proof of Theorem 14 also works for words in a normal form. For such words the sequence $o(\alpha \llbracket n \rrbracket)$ differs from the “standard” fundamental sequence for $o(\alpha)$ only by a term “+1” in the expression from Lemma 5.14.

Theorems 13 and 14 have some interesting strengthenings. First of all, we notice that $\text{TI}^R(\Pi_1, \varepsilon_0)$ is reducible to a formally weaker rule of transfinite induction for elementary formulas. Indeed, if a given Π_1 -formula has the form $\varphi(\alpha) = \forall x \varphi_0(\alpha, x)$, then it is sufficient to define an elementary formula $\psi(\beta)$ in such a way that

$$\psi(\omega\alpha + n) \leftrightarrow \varphi_0(\alpha, n),$$

for all $\alpha < \varepsilon_0$ and $n < \omega$. Then an application of the induction rule for φ is reducible to the one for ψ .

Secondly, the formula $\text{Con}(\text{PA})$ is the strongest sentence that can be derived from EA^+ by one application of $\text{TI}^R(\Pi_1, \varepsilon_0)$.

Proposition 5.17. $[\text{EA}^+, \text{TI}^R(\Pi_1, \varepsilon_0)] \equiv \text{EA}^+ + \text{Con}(\text{PA})$.

A short proof can be found in [26].

Consider a theory $T = \text{EA} + \text{Con}(\text{PA})$ and the corresponding algebra. Notice that T has a Π_1 -axiomatization, hence \mathcal{M}_T^∞ enjoys the reduction property. It is not difficult to see that

$$\begin{aligned} \text{EA} \vdash \langle n \rangle_T \top &\leftrightarrow \langle n \rangle_{\text{EA}} \text{Con}(\text{PA}) \\ &\leftrightarrow \text{Con}(\text{PA}) \wedge \langle n \rangle_{\text{EA}} \top. \end{aligned}$$

Hence,

$$T + \text{PA} \equiv T + \{\langle n \rangle_T \top : n < \omega\}.$$

Therefore, the proof of Theorem 13 for the algebra \mathcal{M}_T^∞ yields the following result:

$$[\text{EA}^+ + T, \text{TI}^R(\Pi_1, \varepsilon_0)] \equiv \text{EA}^+ + \text{Con}(T + \text{PA}).$$

Iterating this argument, we obtain a characterization of the closure of EA^+ under the rule $\text{TI}^R(\Pi_1, \varepsilon_0)$.

Proposition 5.18. $\text{EA}^+ + \text{TI}^R(\Pi_1, \varepsilon_0)$ is equivalent to the theory

$$\text{EA}^+ + \text{Con}(\text{PA}) + \text{Con}(\text{PA} + \text{Con}(\text{PA})) + \dots$$

One can similarly obtain a characterization of the transfinite induction rule for Π_{n+1} -formulas. In this situation iterated n -consistency assertions for PA are emerging instead of the ordinary consistency assertions. Transfinite induction rule for *arbitrary* arithmetical formulas happens to be equivalent to the full transfinite induction schema:

$$\text{TI}(\varepsilon_0) : \quad \forall \alpha (\forall \beta < \alpha \varphi(\beta) \rightarrow \varphi(\alpha)) \rightarrow \forall \alpha \varphi(\alpha).$$

This implies the following well-known result obtained by Kreisel and Lévy [51] using the techniques of Gentzen.

Theorem 15. $\text{EA} + \text{TI}(\varepsilon_0) \equiv \text{EA} + \text{RFN}(\text{PA}) \equiv \text{PA} + \text{RFN}(\text{PA})$.

Proof. By the previous result, it is sufficient to notice that the n -consistency formulas over PA (for any n) follow from $\text{RFN}(\text{PA})$.

5.7 A PA-unprovable combinatorial statement

Here we present a simple statement of combinatorial nature that is independent of Peano Arithmetic and is motivated by the provability algebraic point of view. The principle asserts the termination of a certain sequence similar to the well-known *Hercules and Hydra* game of Kirby and Paris [47].

The game deals with finite words in the alphabet \mathbb{N} , which will be called *worms* in analogy with the hydras. The last element of a worm $w = x_0x_1 \dots x_n$ will be called its *head*.

Informally speaking, the game starts with an arbitrary worm and can be described as follows. At each step Hercules hits the worm, so that its head decreases by 1. In response, the worm regenerates according to the two simple rules below. Unlike in the original Hercules and Hydra battle, the emerging sequence of words is fully deterministic.

More formally, we specify a function $\text{next}(w, m)$, where $w = x_0x_1 \dots x_n$ is a worm and $m \in \mathbb{N}$ is a step of the game:

1. If $x_n = 0$, then $\text{next}(w, m) := x_0 \dots x_{n-1}$. In this case the head of the worm is being cut away.
2. If $x_n > 0$ set $k := \max\{i < n : x_i < x_n\}$.

The worm w (with the head decreased by 1) consists of two parts: the *good*⁶ one $r := x_0 \dots x_k$, and the *bad* one $s := x_{k+1} \dots x_{n-1}(x_n - 1)$. Set

$$\text{next}(w, m) := r * \underbrace{s * s * \dots * s}_{m+1 \text{ times}}.$$

The sequence of worms starting from an arbitrary worm w is defined as follows:

$$w_0 := w \quad \text{and} \quad w_{n+1} := \text{next}(w_n, n + 1).$$

As a typical example, consider a worm $w = 2031$. At the first step we obtain $k = 1$; $r = 20$; $s = 30$; $\text{next}(w, 1) = 203030$. Then the game proceeds as follows:

$$\begin{aligned} w_0 &= 2031 \\ w_1 &= 203030 \\ w_2 &= 20303 \\ w_3 &= 20302222 \\ w_4 &= 203022212221222122212221 \\ w_5 &= 2030(22212221222122212220)^6 \\ &\dots \end{aligned}$$

Notice that w_n is defined by primitive recursion. In fact, w_n is an elementary function of n and (the code of) w . This can be seen from the simple estimate

$$|w_n| \leq (n + 2)! |w_0|$$

⁶This part can also be empty.

showing that the length of a worm grows only elementarily in the course of the game. Also notice that the maximal element of a worm can only decrease. This allows one to write out an elementary formula in three variables expressing that $w_n = u$.

The intended true PA-unprovable principle WP asserts that Hercules eventually defeats any initial worm: $\forall w \exists n w_n = \Lambda$.

Theorem 16. *WP is true but unprovable in PA. In fact, WP is equivalent to 1-Con(PA) in EA.*

Proof. Worms have a natural interpretation in terms of the graded provability algebra $\mathcal{M}_{\text{EA}}^\infty$. Let a word $\alpha \in S$ be converse to w , that is, α coincides with w written in the opposite order. Then it is easy to see that the worm $\text{next}(w, n)$ will be converse to $\alpha \llbracket n \rrbracket$ (this is not accidental, since we have written off the rules of the game from the definition of $\alpha \llbracket n \rrbracket$). This implies the truth of WP because, by Corollary 5.15, the sequence of the corresponding ordinals decreases.

In fact, this reasoning shows that WP is provable by one application of the rule $\text{TI}^R(\Pi_2, \varepsilon_0)$. Therefore, WP also follows from 1-Con(PA), by an analog of Proposition 5.17. However, here we would like to prove that $\text{EA} + 1\text{-Con(PA)} \vdash \text{WP}$ in a more direct (and more interesting) way, without using ordinals.

Define an arithmetical interpretation $w^* := (\alpha^+)_{\text{EA}}^*$, where α is converse to w . As before, we write $\langle m \rangle$ instead of $\langle m \rangle_{\text{EA}}$, for all m .

Lemma 5.19. *For any w , $\text{PA} \vdash w^*$.*

Proof. We argue by induction on $|w|$. If $w = vn$ and m exceeds any letter in w , then

$$\begin{aligned} \text{EA} \vdash v^* \wedge \langle m+1 \rangle \top &\rightarrow \langle m+1 \rangle v^*, & \text{by Lemma 5.9} \\ &\rightarrow \langle n+1 \rangle v^*. \end{aligned}$$

By Corollary 4.3 and the induction hypothesis,

$$\text{PA} \vdash v^* \wedge \langle m+1 \rangle \top,$$

therefore $\text{PA} \vdash \langle n+1 \rangle v^*$, which yields the induction step.

Lemma 5.20. *For any w ,*

$$\text{EA} \vdash \forall n (w_n \neq \Lambda \rightarrow \Box(w_n^* \rightarrow \langle 1 \rangle w_{n+1}^*)).$$

Proof. It is sufficient to prove in EA

$$\forall w \neq \Lambda \forall n \text{ EA} \vdash w^* \rightarrow \langle 1 \rangle \text{next}(w, n)^*.$$

Let α be converse to w . If α begins with 0, the claim is obvious. If α begins with $k+1$, then, by a remark above, $\alpha \llbracket n \rrbracket$ is converse to $\text{next}(w, n)$. By Corollary 5.12 (formalizable in EA),

$$\text{GLP} \vdash \alpha \rightarrow \Diamond \alpha \llbracket n \rrbracket.$$

Theorems of **GLP** are stable under $(\cdot)^+$, hence

$$\mathbf{GLP} \vdash \alpha^+ \rightarrow \langle 1 \rangle \alpha \llbracket n \rrbracket^+.$$

This proves the claim by the arithmetical soundness of **GLP**.

Lemma 5.21. *For any w ,*

$$\mathbf{EA} \vdash \langle 1 \rangle w_0^* \rightarrow \exists n w_n = \Lambda.$$

Proof. We shall prove $\forall n w_n \neq \Lambda \rightarrow \forall n [1] \neg w_n^*$ using Löb's theorem.

$$\begin{aligned} \mathbf{EA} \vdash \forall n w_n \neq \Lambda \wedge [1] \forall n [1] \neg w_n^* &\rightarrow [1] \forall n [1] \neg w_{n+1}^* \\ &\rightarrow \forall n [1] [1] \neg w_{n+1}^* \\ &\rightarrow \forall n [1] \neg w_n^*, \quad \text{by Lemma 5.20;} \end{aligned}$$

$$\begin{aligned} \mathbf{EA} \vdash [1] \forall n w_n \neq \Lambda &\rightarrow [1] ([1] \forall n [1] \neg w_n^* \rightarrow \forall n [1] \neg w_n^*) \\ &\rightarrow [1] \forall n [1] \neg w_n^*, \quad \text{by Löb's theorem;} \end{aligned}$$

$$\begin{aligned} \mathbf{EA} \vdash \forall n w_n \neq \Lambda &\rightarrow [1] \forall n w_n \neq \Lambda, \quad \text{by } \Sigma_2\text{-soundness} \\ &\rightarrow [1] \forall n [1] \neg w_n^* \\ &\rightarrow \forall n [1] \neg w_n^* \\ &\rightarrow [1] \neg w_0^*, \end{aligned}$$

as required.

We complete the proof of Theorem 16. From Lemmas 5.19 and 5.21 we obtain that, for each w ,

$$\begin{aligned} \mathbf{PA} &\vdash \langle 1 \rangle w^*, \\ \mathbf{EA} &\vdash \langle 1 \rangle w^* \rightarrow \exists n w_n = \Lambda. \end{aligned}$$

Hence, $\forall w \mathbf{PA} \vdash \exists n w_n = \Lambda$. This fact is formalizable in **EA**, therefore $1\text{-Con}(\mathbf{PA})$ implies $\forall w \exists n w_n = \Lambda$.

5.8 Independence of WP

Let $w \llbracket n \rrbracket := \text{next}(w, n)$ and

$$w \llbracket n \dots n + k \rrbracket := w \llbracket n \rrbracket \llbracket n + 1 \rrbracket \dots \llbracket n + k \rrbracket.$$

We introduce an analogue of the so-called Hardy functions as follows. Let $h_w(n)$ be the smallest k such that

$$w \llbracket n \dots n + k \rrbracket = \Lambda.$$

We need some nice properties of h established by elementary reasoning formalizable in **EA**. Notice that the function

$$W(w, n, k) := w \llbracket n \dots n + k \rrbracket$$

is elementary because it can be defined by bounded recursion on k in a way similar to the sequence w_n . This yields a natural elementary presentation of the relation $h_w(n) = k$ in **EA**.

The following notion will be useful to establish the monotonicity of the functions h_w . Let $v \trianglelefteq u$ if and only if $v = u[0][0] \dots [0]$. This means that v is an initial segment of u except possibly for the last letter, which should not exceed the corresponding letter in u . Notice that the totality of the functions h_w cannot, in general, be established within **EA**. For a formalization of the following lemmas in **EA**, we shall always mention this as an additional condition.

Lemma 5.22. *If $h_w(m)$ is defined and $u \trianglelefteq w$, then*

$$\exists k \ w[m \dots m + k] = u.$$

Proof. The rules of the game are such that the n -th letter in w can only change if all letters to the right of it are deleted. Hence, if $w[m \dots m + s] = \Lambda$ then, for some $k < s$, $w[m \dots m + k] = u$.

Corollary 5.23. *If $h_w(n)$ is defined, then $\forall m \leq n \ \exists k \ w[n \dots n + k] = w[m]$.*

Proof. $w[m]$ is an initial segment of $w[n]$.

Lemma 5.24. *If $v \trianglelefteq u$ and $x \leq y$, then $h_v(x) \leq h_u(y)$.*

Proof. Applying Corollary 5.23 several times we obtain s_0, s_1, \dots such that

$$\begin{aligned} u[y \dots y + s_0] &= v[x], \quad \text{where } y + s_0 \geq x; \\ u[y \dots y + s_0 + s_1] &= v[x][x + 1], \quad \text{where } y + s_0 + s_1 \geq x + 1; \\ &\dots \end{aligned}$$

Hence, all elements of the sequence starting with v occur in the sequence for u .

Lemma 5.25. $h_{u0v}(n) = h_u(n + h_v(n) + 2) + h_v(n) + 1 > h_u(h_v(n))$.

Proof. Nothing can happen to the 0 between u and v until the v part is eliminated. So, the worm $u0v$ first rewrites to $u0$ in $h_v(n)$ steps, and then to Λ .

Corollary 5.26. *If $w \in S_1$ then $h_{w1}(n) > h_w^{(n)}(n)$.*

Proof. It is sufficient to note that $w1[n] = w0w0 \dots w0$.

Now we formulate our main lemma. As usual, $h_w \downarrow$ denotes the formula $\forall x \exists y \ h_w(x) = y$. We also define $w^* := \alpha^*$, where $\alpha \in S$ is converse to w .

Lemma 5.27. $\text{EA} \vdash \forall w \in S_1 \ (h_{1111w} \downarrow \rightarrow \langle 1 \rangle w^*)$.

From this result one easily infers Theorem 16.

Proof of the independence of WP:

$$\begin{aligned} \text{EA} \vdash \forall w \exists n \, w_n = \Lambda &\rightarrow \forall w \in S_1 \, h_w \downarrow \\ &\rightarrow \forall n \, \langle 1 \rangle \langle n \rangle \top \\ &\rightarrow \text{1-Con}(\text{PA}). \end{aligned}$$

Here, the first implication holds because for every worm w and every number x one can find another worm $w' := w0^x$ such that $w' \llbracket 0 \dots x-1 \rrbracket = w$. Hence, Hercules defeats w' iff $h_w(x)$ is defined.

In the proof of this lemma we essentially use Lemma 3.7. Notice that, by Corollary 5.26, we have $h_{111}(x) > 2^x$, therefore the same inequality holds for the function $h_{111w}(x)$, where w is any word.

Proof of Lemma 5.27. Reason in EA. By Löb's theorem, we can use as an additional assumption

$$\forall w \in S_1 \, [1](h_{1111w} \downarrow \rightarrow \langle 1 \rangle w^*). \quad (5.2)$$

If $1111w = v1$ then $h_{v1} \downarrow \rightarrow \lambda x. h_v^{(x)}(x) \downarrow$.

The function h_v is increasing, has an elementary graph and grows at least exponentially. So, for $w = \Lambda$ the claim is obvious: $h_{1111} \downarrow$ implies the totality of 2_n^x , and hence $\langle 1 \rangle \top$. If w is nonempty, we reason as follows:

$$\begin{aligned} \lambda x. h_v^{(x)}(x) \downarrow &\rightarrow \langle 1 \rangle h_v \downarrow, \quad \text{by Lemma 3.7} \\ &\rightarrow \langle 1 \rangle \langle 1 \rangle v^*, \quad \text{by assumption (5.2)} \\ &\rightarrow \langle 1 \rangle w^*. \end{aligned}$$

If $1111w = v$ ends with $m > 1$, then

$$\begin{aligned} h_v \downarrow &\rightarrow \lambda x. h_{v \llbracket x \rrbracket}(x+1) \downarrow \\ &\rightarrow \forall n \, h_{v \llbracket n \rrbracket} \downarrow. \end{aligned}$$

For a proof of the last implication fix an arbitrary n . If $x \leq n$ then $h_{v \llbracket n \rrbracket}(x) \leq h_{v \llbracket n \rrbracket}(n+1)$. If $x \geq n$ then $h_{v \llbracket n \rrbracket}(x) \leq h_{v \llbracket x \rrbracket}(x+1)$. In both cases the larger value is defined, hence so is the smaller one.

Let us continue the argument:

$$\begin{aligned} \forall n \, h_{v \llbracket n+1 \rrbracket} \downarrow &\rightarrow \forall n \, h_{v \llbracket n \rrbracket} \downarrow \quad (\text{as } v \llbracket n \rrbracket 1 \trianglelefteq v \llbracket n+1 \rrbracket) \\ &\rightarrow \forall n \, \langle 1 \rangle h_{v \llbracket n \rrbracket} \downarrow \quad (\text{as before}) \\ &\rightarrow \forall n \, \langle 1 \rangle \langle 1 \rangle w \llbracket n \rrbracket^* \\ &\rightarrow \langle 1 \rangle w^* \quad (\text{by the reduction property}). \end{aligned}$$

This ends the proofs of Lemma 5.27 and of Theorem 16.

6 Iterated reflection principles

Here we shall briefly consider connections between provability algebras and the hierarchies of iterated reflection schemata in the sense of Turing–Feferman [92, 38]. Historically the first and, probably, the most important example of such a hierarchy is the so-called *Turing transfinite recursive progression* defined according to the following clauses:

T1. $T_0 = T$, T being a given “initial” theory;

T2. $T_{\alpha+1} = T_\alpha + \text{Con}(T_\alpha)$;

T3. $T_\alpha = \bigcup_{\beta < \alpha} T_\beta$, for α a limit ordinal.

By Gödel’s theorem, whenever the initial theory T is Σ_1 -sound, the theories T_α form a strictly increasing transfinite sequence of Σ_1 -sound extensions of T . This sequence can be used to associate an ordinal $\text{ord}_T(U)$ to any extension U of T as follows:

$$\text{ord}_T(U) := \text{the least } \alpha \text{ such that } U \not\vdash \text{Con}(T_\alpha).$$

For the most natural extensions of T one can usually show that, if $\alpha = \text{ord}_T(U)$, then T_α exhausts all arithmetical Π_1 -consequences of U , that is,

$$U \equiv_{\Pi_1} T_\alpha.$$

Thus, $\text{ord}_T(U)$ can be thought of as an ordinal measuring the “strength” of U with respect to T in the sense of the provability of Π_1 -sentences. It is natural to take for T some fixed weak Π_1 -axiomatizable theory, for example, **EA**.

This approach has an advantage over the other methods of associating ordinals with theories because it yields the finest classification, which allows to distinguish the theories already by their arithmetical Π_1 -consequences. For example, **PA** and $\text{PA} + \text{Con}(\text{PA})$ have the same class of provably total computable functions and the same bounds on the provability of transfinite induction, equal to ε_0 . However (cf. Corollaries 6.1 and 6.4 below), $\text{ord}_{\text{EA}}(\text{PA}) = \varepsilon_0$, whereas $\text{ord}_{\text{EA}}(\text{PA} + \text{Con}(\text{PA})) = \varepsilon_0 \cdot 2$.

Similarly, replacing in the definition of the Turing progression $\text{Con}(T)$ by the n -consistency formula $n\text{-Con}(T)$ we obtain a definition of Π_n^0 -ordinal of a theory.

A well-known difficulty in the way of this program roots in the fact that the clauses T1–T3 do not uniquely define the sequence of theories T_α , that is, T_α depends on a formal representation of the ordinal α within arithmetic rather than on the ordinal itself.

For the analysis of this problem Feferman [38] considered families of theories of the form $(T_c)_{c \in \mathcal{O}}$ satisfying T1–T3 along every path within \mathcal{O} , where \mathcal{O} is Kleene’s universal system of ordinal notation. Using an idea of Turing, he showed that every true Π_1 -sentence is provable in T_c , for a suitable ordinal notation $c \in \mathcal{O}$ with $|c| = \omega + 1$.⁷ It follows that there are two ordinal notations

⁷Here and below $|c|$ denotes the ordinal coded by c .

$a, b \in \mathcal{O}$ with $|a| = |b| = \omega + 1$ such that T_a proves $\text{Con}(T_b)$. This observation seems to break down the program of associating ordinals with formal theories as described above, at least in the general case.

Nonetheless, a possibility remains that for *natural* (mathematically meaningful) theories U one can exhaust all Π_1 -consequences of U using only specific *natural* ordinal notations, and a careful choice of such notations should yield proper ordinal bounds. This idea has been developed in the work of Schmerl [77] who showed among other things that, for the standard ordinal notation system up to ε_0 and for a specific arithmetization of conditions T1–T3, the theories PA and $\text{PRA}_{\varepsilon_0}$ prove the same Π_1 -sentences. Thus, $\text{ord}_{\text{PRA}}(\text{PA}) = \varepsilon_0$, which coincides with the ordinal associated with PA by some other proof-theoretic methods.

A simplified approach to the treatment of iterated reflection principles based on a notion of *smooth Turing's progression* was suggested in [18]. In the same paper, existence and uniqueness theorems for such progressions were proved. Thus, the notion of smoothness gives a satisfactory answer to the question of arithmetization of conditions T1–T3 and their analogs for the other reflection principles. This provided correct definitions of Π_n^0 -ordinals of theories according to the above approach, for any fixed system of ordinal notation. See [25] for a detailed discussion of possible definitions of proof-theoretic ordinals of theories and a survey of main results on iterated reflection principles.

Here we shall show that the system of ordinal notation emerging from the graded provability algebra for EA allows one to compute the Π_n^0 -ordinals of the standard fragments of PA, for all n . We also easily obtain a variant of the so-called *Schmerl's formula* relating different progressions of iterated reflection principles. The treatment in this paper is a simplification of the one in [25].

6.1 Smooth Turing progressions

We call an *elementary partial ordering* (D, \prec) a pair of elementary formulas $x \in D$ and $x \prec y$ such that EA proves that \prec strictly partially orders $D \subseteq \mathbb{N}$. An elementary *well-ordering* is an elementary ordering which is well-founded in the standard model of arithmetic. (Recall that the notion of well-foundedness is not expressible by a first order arithmetical formula.)

Suppose we are given an elementary presented “initial” theory T and an elementary well-ordering (D, \prec) . It is shown in [18, 25] that, for each n , there is a unique modulo provable equivalence in EA family of elementary presented theories $(T_\alpha^n)_{\alpha \in D}$ such that EA proves

$$\forall \alpha \in D \quad T_\alpha^n \equiv T + \{n\text{-Con}(T_\beta^n) : \beta \prec \alpha, \beta \in D\}.$$

Such a sequence will be called a *smooth progression based on the iteration of the n -consistency assertion*. T_α will denote the smooth Turing progression T_α^0 . Π_{n+1}^0 -ordinal of a theory U is defined as the first $\alpha \in D$ such that $\text{EA}_{\alpha+1}^n \not\subseteq U$.

Take $(S_n, <_n)$ as an elementary well-ordering (D, \prec) and consider the corresponding progression T_α^n .

Theorem 17. *If T is an elementary presented Π_{n+1} -extension of EA^+ , then, provably in EA^+ ,*

$$\forall \alpha \in S_n, \quad T + \alpha_T^* \equiv_n T_\alpha^n.$$

Proof. By Löb's theorem, it is sufficient to prove

$$\text{EA}^+ \vdash \forall \beta <_n \alpha \quad \Box_{\text{EA}^+}(T + \beta_T^* \equiv_n T_\beta^n) \rightarrow T + \alpha_T^* \equiv_n T_\alpha^n.$$

We prove the equivalence $T + \alpha_T^* \equiv_n T_\alpha^n$ by presenting an informal argument in EA^+ ; the premise of the above implication will be called the reflexive induction hypothesis. Notice that this hypothesis implies

$$\forall \beta <_n \alpha \quad \text{EA}^+ \vdash \langle n \rangle_T \beta^* \leftrightarrow n\text{-Con}(T_\beta^n).$$

(\supseteq) By the definition of $<_n$, if $\beta <_n \alpha$ then $T \vdash \alpha_T^* \rightarrow \langle n \rangle_T \beta_T^*$. By the reflexive induction hypothesis,

$$\text{EA}^+ \vdash \langle n \rangle_T \beta_T^* \rightarrow n\text{-Con}(T_\beta^n).$$

It follows that, for all $\beta <_n \alpha$,

$$T \vdash \alpha_T^* \rightarrow n\text{-Con}(T_\beta^n),$$

hence $T + \alpha_T^* \vdash T_\alpha^n$.

(\subseteq_n) Assume $T + \alpha_T^* \vdash \pi$ with $\pi \in \Pi_n$. If $\alpha = n\beta$ then we have $T + \langle n \rangle_T \beta_T^* \vdash \pi$, which implies $T + n\text{-Con}(T_\beta^n) \vdash \pi$, by the reflexive induction hypothesis. Since obviously $\beta <_n \alpha$, we conclude that $T_\alpha^n \vdash \pi$.

If $\alpha = \langle m+1 \rangle \beta$ with $m \geq n$, then

$$\{\alpha_T^*\} \equiv_m \{(\alpha[k])_T^* : k < \omega\},$$

by the reduction property. Hence, $T + \alpha[k]_T^* \vdash \pi$, for some k . Since $\alpha[k] <_n \alpha$, we obtain $T_{\alpha[k]}^n \vdash \pi$, by the reflexive induction hypothesis.

Notice that the previous theorem does not depend on the property of the ordinal notation system being actually well-founded. The fact that T contains EA^+ is only essential for $n = 0$; for $n > 0$ it is sufficient to assume $T \supseteq \text{EA}$. This also applies to the corollaries formulated below.

Similarly to Gentzen's theorem, we have a variant of this result for the standard system of ordinal notation for ε_0 . Let $o_n(\alpha)$ denote the Cantor normal form of a representation of the ordinal $o(\alpha^{-n})$, where α^{-n} is obtained from $\alpha \in S_n$ by replacing all letters m by $m - n$.

Theorem 18. *If T is an elementary presented Π_{n+1} -extension of EA^+ , then, provably in EA^+ ,*

$$\forall \alpha \in S_n, \quad T + \alpha_T^* \equiv_n T_{o_n(\alpha)}^n.$$

This theorem allows us to compute Π_n^0 -ordinals and to obtain characterizations of Π_n -consequences of all previously considered fragments of PA. For the induction schemata with parameters, an answer is given by the following theorem. Define $\omega_0(\alpha) = \alpha$, $\omega_{n+1}(\alpha) = \omega^{\omega_n(\alpha)}$, $\omega_n = \omega_n(1)$. Notice that $o(n) = \omega_n$.

Corollary 6.1. (i) $\text{PA} \equiv_m \bigcup_{\alpha < \varepsilon_0} (\text{EA}^+)_\alpha^m$;

(ii) $I\Sigma_n \equiv_m (\text{EA}^+)_{\omega_{n+1-m}}^m$, if $m \leq n$.

Proof. Statement (ii) follows from the fact that $I\Sigma_n$ is $\langle n+1 \rangle_{\text{EA}} \top$ and $o_m(n+1) = o(n+1-m) = \omega_{n+1-m}$. The first statement is obtained from it by taking a limit in n for a fixed m .

A characterization of the Π_2 -ordinal of a theory in terms of the iterated 1-consistency assertions over EA implies a characterization of the class of provably total computable functions of the theory in terms of the so-called extended Grzegorzczuk hierarchy \mathcal{E}^α : $\mathcal{F}(\text{EA}_\alpha^2) = \mathcal{E}^\alpha$, for all α . Hence, for any theory U such that $U \equiv_1 \text{EA}_\alpha^2$, one also has $\mathcal{F}(U) = \mathcal{E}^\alpha$.

This result holds for sufficiently general systems of ordinal notation and is proved by a transfinite iteration of Lemma 3.7 ([25], see also [89]). Hence, we obtain a new proof of a well-known theorem due to Schwichtenberg and Wainer (cf. [34, 25]).

Corollary 6.2. $\mathcal{F}(\text{PA}) = \bigcup_{\alpha < \varepsilon_0} \mathcal{E}^\alpha$.

Consider now a more exotic theory $I\Sigma_1 + \text{III}_2^-$. This system is interesting because it formalizes the most natural proof of the totality of the Ackermann function. The following theorem, in particular, yields a characterization of the class of provably total computable functions of that theory.

Theorem 19. $I\Sigma_n + \text{III}_{n+1}^- \equiv_m (\text{EA}^+)_\alpha^m$, where $\alpha = \omega_{n+1-m}(2)$, $m \leq n$.

Proof. By Proposition 4.23, $I\Sigma_n + \text{III}_{n+1}^- \equiv_n (I\Sigma_n)_\omega^n$. By the reduction property for $\mathcal{M}_{\text{EA}^+}^\infty$, we also have

$$(I\Sigma_n)_\omega^n \equiv_n \langle n+1 \rangle (I\Sigma_n) \equiv \langle n+1 \rangle \langle n+1 \rangle \top.$$

By Theorem 18, the Π_n -ordinal of this theory equals

$$o_m(\langle n+1 \rangle \langle n+1 \rangle \top) = o(\langle n+1-m \rangle \langle n+1-m \rangle \top) = \omega_{n+1-m}(2).$$

Corollary 6.3. $\mathcal{F}(I\Sigma_n + \text{III}_{n+1}^-) = \mathcal{E}^{\omega_n(2)}$. In particular, $\mathcal{F}(I\Sigma_1 + \text{III}_2^-) = \mathcal{E}^{\omega^2}$, that is, it coincides with the class of doubly recursive functions in the sense of Peter.

Finally, let us compute the Π_1^0 -ordinal of the theory $\text{PA} + \text{Con}(\text{PA})$. We shall work with the standard system of ordinal notation for ε_1 , which is the structure $(\varepsilon_1; \leq, 0, \varepsilon_0, +, \omega^x)$. Ordinals are coded by normal form terms built up from the constants 0 and ε_0 by the functions $+$ and ω^x .

Corollary 6.4. $\text{PA} + \text{Con}(\text{PA}) \equiv_0 \text{EA}_{\varepsilon_0 + \varepsilon_0}^+$.

Proof. Consider the theory $T = \text{EA}^+ + \text{Con}(\text{PA})$ and the corresponding graded provability algebra \mathcal{M}_T^∞ . Notice that T is a Π_2 -extension of EA , hence \mathcal{M}_T^∞ enjoys the reduction property. As we have seen,

$$\text{PA} + \text{Con}(\text{PA}) \equiv T + \{\langle n \rangle_T \top : n < \omega\}.$$

It follows by Theorem 18 that

$$\text{PA} + \text{Con}(\text{PA}) \equiv_0 T_{\varepsilon_0}.$$

On the other hand, provably in EA^+ we have $\text{PA} \equiv_0 \text{EA}_{\varepsilon_0}^+$, whence

$$T \equiv \text{EA}_{\varepsilon_0+1}^+.$$

By the uniqueness theorem for smooth progressions, if $U \equiv V$, then provably $U_\alpha \equiv V_\alpha$, for all α . Therefore,

$$\text{PA} + \text{Con}(\text{PA}) \equiv_0 T_{\varepsilon_0} \equiv (\text{EA}_{\varepsilon_0+1}^+)_{\varepsilon_0}.$$

The properties of $+$ for the standard system of ordinal notation allow us to prove in EA the obvious property $(U_\alpha)_\beta \equiv U_{\alpha+\beta}$, for all α, β . Hence, $(\text{EA}_{\varepsilon_0+1}^+)_{\varepsilon_0} \equiv \text{EA}_{\varepsilon_0+\varepsilon_0}^+$, as required.

References

- [1] Ю.Л. Ершов. *Теория нумераций*. Наука, Москва, 1977.
- [2] В.П. Оревков. Нижние оценки удлинения выводов после устранения сечения. *Записки науч. сем. ЛОМИ*, 88:137–162, 1979. English translation in *Journal of Soviet Mathematics*, 20 (1982), 2337–2350.
- [3] С.Н. Артёмов. *Расширения арифметики и модальные логики*. Дисс. канд. физ.-мат. наук, Математический институт им. В.А. Стеклова, Москва, 1979.
- [4] С.Н. Артёмов. Приложения модальной логики в теории доказательств. In *Вопросы кибернетики: Неклассические логики и их применения*, pages 3–20. Наука, Москва, 1982.
- [5] С.В. Горячев. Об интерпретируемости некоторых расширений арифметики. *Математические заметки*, 40:561–572, 1986. English translation in *Math. Notes*, 40.
- [6] Г.К. Джапаридзе. Модально-логические средства исследования доказуемости. Дисс. канд. филос. наук, Москва, МГУ, 1986.
- [7] Г.К. Джапаридзе. Полимодальная логика доказуемости. In *Интенциональные логики и логическая структура теорий: материалы 4-го советско-финского симпозиума по логике, Телави, 20–24 мая 1985 г.*, pages 16–48. Мецниереба, Тбилиси, 1988.

- [8] Л.Д. Беклемишев. Об ограниченном правиле индукции и итерированных схемах рефлексии над Кальмаровской элементарной арифметикой. In О.Б. Лупанов, editor, *Теоретические и прикладные аспекты математических исследований*, pages 36–39. МГУ, Москва, 1994.
- [9] С.С. Гончаров. *Счётные булевы алгебры и разрешимость*. Сибирская школа алгебры и логики. Научная Книга, Новосибирск, 1996. English translation: *Countable boolean algebras and decidability*. Plenum Press, New York, 1997.
- [10] W. Ackermann. Zur Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 117:162–194, 1940.
- [11] T. Arai. Epsilon substitution method for theories of jump hierarchies. *Archive for Mathematical Logic*, 41(2):123–153, 2002.
- [12] T. Arai. Epsilon substitution method for $ID_1(\Pi_1^0 \vee \Sigma_1^0)$. *Annals of Pure and Applied Logic*, 121(2–3):163–208, 2003.
- [13] S.N. Artëmov and L.D. Beklemishev. On propositional quantifiers in provability logic. *Notre Dame Journal of Formal Logic*, 34:401–419, 1993.
- [14] S.N. Artemov and L.D. Beklemishev. Provability logic. Logic Group Preprint Series 234, University of Utrecht, November 2004. <http://preprints.phil.uu.nl/lgps/>. To appear in *Handbook of Philosophical Logic*, vol. 10, D. Gabbay and F. Guenther, eds., Kluwer, Dordrecht.
- [15] J. Avigad. An ordinal analysis of admissible set theory using recursion on ordinal notations. *Journal of Mathematical Logic*, 2(1):91–112, 2002.
- [16] J. Avigad and R. Sommer. A model-theoretic approach to ordinal analysis. *Bulletin of Symbolic Logic*, 3(1):17–52, 1997.
- [17] J. Avigad and R. Sommer. The model-theoretic ordinal analysis of theories of predicative strength. *The Journal of Symbolic Logic*, 64(1):327–349, 1999.
- [18] L.D. Beklemishev. Iterated local reflection versus iterated consistency. *Annals of Pure and Applied Logic*, 75:25–48, 1995.
- [19] L.D. Beklemishev. Remarks on Magari algebras of PA and $ID_0 + EXP$. In P. Agliano and A. Ursini, editors, *Logic and Algebra*, pages 317–325. Marcel Dekker, New York, 1996.
- [20] L.D. Beklemishev. Induction rules, reflection principles, and provably recursive functions. *Annals of Pure and Applied Logic*, 85:193–242, 1997.
- [21] L.D. Beklemishev. Notes on local reflection principles. *Theoria*, 63(3):139–146, 1997.

- [22] L.D. Beklemishev. Parameter free induction and reflection. In G. Gottlob, A. Leitsch, and D. Mundici, editors, *Lecture Notes in Computer Science 1289. Computational Logic and Proof Theory, 5-th K.Gödel Colloquium KGC'97, Proceedings*, pages 103–113. Springer-Verlag, Berlin, 1997.
- [23] L.D. Beklemishev. A proof-theoretic analysis of collection. *Archive for Mathematical Logic*, 37:275–296, 1998.
- [24] L.D. Beklemishev. Parameter free induction and provably total computable functions. *Theoretical Computer Science*, 224(1–2):13–33, 1999.
- [25] L.D. Beklemishev. Proof-theoretic analysis by iterated reflection. *Archive for Mathematical Logic*, 42:515–552, 2003. DOI: 10.1007/s00153-002-0158-7.
- [26] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128:103–123, 2004.
- [27] L.D. Beklemishev. The Worm principle. Logic Group Preprint Series 219, University of Utrecht, March 2003. <http://preprints.phil.uu.nl/lgps/>.
- [28] G. Boolos. Reflection principles and iterated consistency assertions. *The Journal of Symbolic Logic*, 44:33–35, 1979.
- [29] G. Boolos. *The Unprovability of Consistency: An Essay in Modal Logic*. Cambridge University Press, Cambridge, 1979.
- [30] G. Boolos. Omega-consistency and the diamond. *Studia Logica*, 39:237–243, 1980.
- [31] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.
- [32] W. Buchholz. Explaining Gentzen’s consistency proof within infinitary proof theory. In G. Gottlob, A. Leitsch, and D. Mundici, editors, *Lecture Notes in Computer Science 1289. Computational Logic and Proof Theory, 5-th K.Gödel Colloquium KGC'97, Proceedings*, pages 4–17. Springer-Verlag, Berlin, 1997.
- [33] W. Buchholz. Explaining the Gentzen–Takeuti reduction steps: A second-order system. *Archive for Mathematical Logic*, 40(4):255–272, 2001.
- [34] W. Buchholz and S. Wainer. Provably computable functions and the fast growing hierarchy. In *Contemporary Math.* 65, pages 179–198, 1987.
- [35] N. Cutland. *Computability. An introduction to recursive function theory*. Cambridge University Press, Cambridge, 1980.
- [36] D. de Jongh and G. Japaridze. The Logic of Provability. In S.R. Buss, editor, *Handbook of Proof Theory*. Studies in Logic and the Foundations of Mathematics, Vol.137, pages 475–546. Elsevier, Amsterdam, 1998.

- [37] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.
- [38] S. Feferman. Transfinite recursive progressions of axiomatic theories. *The Journal of Symbolic Logic*, 27:259–316, 1962.
- [39] S. Feferman. Three conceptual problems that bug me. Lecture text for 7-th Scandinavian Logic Symposium, <ftp://math.stanford.edu/pub/papers/feferman/conceptualprobs.ps.gz>, 1996.
- [40] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Math. Ann.*, 112(4):493–565, 1936.
- [41] G. Gentzen. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und Grundlegung der exakten Wissenschaften*, 4:19–44, 1938.
- [42] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931.
- [43] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic*. Springer-Verlag, Berlin, Heidelberg, New York, 1993.
- [44] L. Henkin, J.D. Monk, and A. Tarski. *Cylindric algebras. Part I, II. 2nd printing*. Studies in Logic and the Foundations of Mathematics, Vols. 64, 115. North-Holland, Amsterdam, 1985.
- [45] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *The Journal of Symbolic Logic*, 58:249–290, 1993.
- [46] R. Kaye, J. Paris, and C. Dimitracopoulos. On parameter free induction schemas. *The Journal of Symbolic Logic*, 53(4):1082–1097, 1988.
- [47] L.A.S. Kirby and J.B. Paris. Accessible independence results for Peano arithmetic. *Bull. London Math. Soc.*, 14:285–293, 1982.
- [48] G. Kreisel. On the interpretation of non-finitist proofs, II. *The Journal of Symbolic Logic*, 48:182–184, 1952.
- [49] G. Kreisel. A survey of proof theory. *The Journal of Symbolic Logic*, 33:321–388, 1968.
- [50] G. Kreisel. Wie die Beweistheorie zu ihren Ordinalzahlen kam und kommt. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 78(4):177–223, 1977.
- [51] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 14:97–142, 1968.

- [52] D. Leivant. The optimality of induction as an axiomatization of arithmetic. *The Journal of Symbolic Logic*, 48:182–184, 1983.
- [53] P. Lindström. On partially conservative sentences and interpretability. *Proceedings of the AMS*, 91(3):436–443, 1984.
- [54] P. Lindström. Provability logic – a short introduction. *Theoria*, 62(1-2):19–61, 1996.
- [55] M.H. Löb. Solution of a problem of Leon Henkin. *The Journal of Symbolic Logic*, 20:115–118, 1955.
- [56] R. Magari. The diagonalizable algebras (the algebraization of the theories which express Theor.:II). *Bollettino della Unione Matematica Italiana*, Serie 4, 12, 1975. Suppl. fasc. 3, 117–125.
- [57] R. Magari. Representation and duality theory for diagonalizable algebras (the algebraization of theories which express Theor.:IV). *Studia Logica*, 34:305–313, 1975.
- [58] G. Mints and S. Tupailo. Epsilon-substitution method for the ramified language and Δ_1^1 -comprehension rule. In A. et al. Cantini, editor, *Logic and foundations of mathematics. Selected contributed papers of the 10th international congress of logic, methodology and philosophy of science, Florence, August 19–25, 1995*. Synth. Libr. 280, pages 107–130, Dordrecht, 1999. Kluwer.
- [59] G.E. Mints. Quantifier free and one quantifier systems. *Zapiski nauchnykh seminarov LOMI*, 20:115–133, 1971. In Russian.
- [60] F. Montagna. On the algebraization of a Feferman’s predicate (the algebraization of theories which express Theor; X). *Studia Logica*, 37:221–236, 1978.
- [61] F. Montagna. On the diagonalizable algebra of Peano arithmetic. *Bollettino della Unione Matematica Italiana*, B (5), 16:795–812, 1979.
- [62] F. Montagna. Undecidability of the first order theory of diagonalizable algebras. *Studia Logica*, 39:347–354, 1980.
- [63] H. Ono. Reflection principles in fragments of Peano Arithmetic. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 33(4):317–333, 1987.
- [64] R. Parikh. Existence and feasibility in arithmetic. *The Journal of Symbolic Logic*, 36:494–508, 1971.
- [65] C. Parsons. On a number-theoretic choice schema and its relation to induction. In A. Kino, J. Myhill, and R.E. Vessley, editors, *Intuitionism and Proof Theory*, pages 459–473. North Holland, Amsterdam, 1970.
- [66] C. Parsons. On n -quantifier induction. *The Journal of Symbolic Logic*, 37(3):466–482, 1972.

- [67] W. Pohlers. A short course in ordinal analysis. In A. Axcel and S. Wainer, editors, *Proof Theory, Complexity, Logic*, pages 867–896. Oxford University Press, Oxford, 1993.
- [68] W. Pohlers. Subsystems of set theory and second order number theory. In S.R. Buss, editor, *Handbook of Proof Theory*, pages 210–335. Elsevier, North-Holland, Amsterdam, 1998.
- [69] M.B. Pour-El and S. Kripke. Deduction-preserving “recursive isomorphisms” between theories. *Fundamenta Mathematicae*, 61:141–163, 1967.
- [70] L.J. Pozsgay. Gödel’s Second Theorem for Elementary Arithmetic. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 14:67–80, 1968.
- [71] M.O. Rabin. Non-standard models and independence of the induction axiom. In *Essays on the Foundations of Mathematics: Dedicated to A. Fraenkel on his 70th anniversary*, pages 287–299. North-Holland, Amsterdam, 1961.
- [72] M. Rathjen. Recent advances in ordinal analysis: Π_2^1 -CA and related systems. *Bulletin of Symbolic Logic*, 1(4):468–485, 1995.
- [73] M. Rathjen. The realm of ordinal analysis. In S.B. Cooper and J.K. Truss, editors, *Sets and proofs. London Math. Soc. Lect. Note Series 258*, pages 219–279. Cambridge University Press, Cambridge, 1999.
- [74] H.E. Rose. *Subrecursion: Functions and Hierarchies*. Clarendon Press, Oxford, 1984.
- [75] J.B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, 1:87–91, 1936.
- [76] J.B. Rosser. Gödel Theorems for non-constructive logics. *The Journal of Symbolic Logic*, 2:129–137, 1937.
- [77] U.R. Schmerl. A fine structure generated by reflection formulas over Primitive Recursive Arithmetic. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium’78*, pages 335–350. North Holland, Amsterdam, 1979.
- [78] H. Schwichtenberg. Some applications of cut-elimination. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 867–896. North Holland, Amsterdam, 1977.
- [79] K. Segerberg. *An essay in classical modal logic*. Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet, Uppsala, 1971.
- [80] V.Yu. Shavrukov. A note on the diagonalizable algebras of PA and ZF. *Annals of Pure and Applied Logic*, 61:161–173, 1993.
- [81] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes Mathematicae*, 323, 1993.

- [82] V.Yu. Shavrukov. Isomorphisms of diagonalizable algebras. *Theoria*, 63(3):210–221, 1997.
- [83] V.Yu. Shavrukov. Undecidability in diagonalizable algebras. *The Journal of Symbolic Logic*, 62(1):79–116, 1997.
- [84] C. Smoryński. ω -consistency and reflection. In *Colloque International de Logique (Colloq. Int. CNRS)*, pages 167–181. CNRS Inst. B. Pascal, Paris, 1977.
- [85] C. Smoryński. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North Holland, Amsterdam, 1977.
- [86] C. Smoryński. The finite inseparability of the first order theory of diagonalizable algebras. *Studia Logica*, 41:347–349, 1982.
- [87] C. Smoryński. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin, 1985.
- [88] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.
- [89] R. Sommer. Transfinite induction within Peano arithmetic. *Annals of Pure and Applied Logic*, 76(3):231–289, 1995.
- [90] R. Statman. Bounds for proof-search and speed-up in the predicate calculus. *Annals of Mathematical Logic*, 15:225–287, 1978.
- [91] S. Tupailo. Epsilon substitution method for Δ_1^1 -CR: A constructive termination proof. *Logic Journal of the IGPL*, 11(3):367–377, 2003.
- [92] A.M. Turing. System of logics based on ordinals. *Proc. London Math. Soc.*, ser. 2, 45:161–228, 1939.
- [93] A. Visser. An overview of interpretability logic. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakhariashev, editors, *Advances in Modal Logic, v.1, CSLI Lecture Notes, No. 87*, pages 307–359. CSLI Publications, Stanford, 1998.
- [94] A. Wilkie and J. Paris. On the scheme of induction for bounded arithmetic formulas. *Annals of Pure and Applied Logic*, 35:261–302, 1987.
- [95] D. Zambella. Shavrukov’s theorem on the subalgebras of diagonalizable algebras for theories containing $I\Delta_0 + \exp$. *Notre Dame Journal of Formal Logic*, 35:147–157, 1994.