Title

Reflection ranks and ordinal analysis

Authors

Walsh, James
Pakhomov, Fedor

Publication Date

2020-01-14

REFLECTION RANKS AND ORDINAL ANALYSIS

FEDOR PAKHOMOV AND JAMES WALSH

Abstract. It is well-known that natural axiomatic theories are well-ordered by consistency strength. However, it is possible to construct descending chains of artificial theories with respect to consistency strength. We provide an explanation of this well-orderedness phenomenon by studying a coarsening of the consistency strength order, namely, the Π^1_1 reflection strength order. We prove that there are no descending sequences of Π^1_1 sound extensions of ACA_0 in this ordering. Accordingly, we can attach a rank in this order, which we call reflection rank, to any Π^1_1 sound extension of ACA_0. We prove that for any Π^1_1 sound theory T extending ACA_0^+, the reflection rank of T equals the proof-theoretic ordinal of T. We also prove that the proof-theoretic ordinal of α iterated Π^1_1 reflection is ε_α. Finally, we use our results to provide straightforward well-foundedness proofs of ordinal notation systems based on reflection principles.

1. Introduction

It is a well-known empirical phenomenon that *natural* axiomatic theories are well-ordered[1] according to many popular metrics of proof-theoretic strength, such as consistency strength. This phenomenon is manifest in *ordinal analysis*, a research program wherein recursive ordinals are assigned to theories to measure their proof-theoretic strength. However, these metrics of proof-theoretic strength do *not* well-order axiomatic theories *in general*. For instance, there are descending chains of sound theories, each of which proves the consistency of the next. However, all such examples of ill-foundedness make use of unnatural, artificial theories. Without a mathematical definition of "natural," it is unclear how to provide a general mathematical explanation of the apparent well-orderedness of the hierarchy of natural theories.

In this paper we introduce a metric of proof-theoretic strength and prove that it is immune to these pathological instances of ill-foundedness. Recall that a theory T is Π^1_1 sound just in case every Π^1_1 theorem of T is true. The Π^1_1 soundness of T is expressible in the language of second-order arithmetic by a formula $\mathsf{RFN}_{\Pi^1_1}(T)$. The formula $\mathsf{RFN}_{\Pi^1_1}(T)$ is also known as the *uniform* Π^1_1 *reflection principle for* T.

Definition 1.1. For theories T and U in the language of second-order arithmetic we say that $T \prec_{\Pi^1_1} U$ if U proves the Π^1_1 soundness of T.

This metric of proof-theoretic strength is coarser than consistency strength, but, as we noted, it is also more robust. In practice, when one shows that U proves the consistency of T, one often also establishes the stronger fact that U proves the Π^1_1 soundness of T. Our first main theorem is the following.

Thanks to Lev Beklemishev and Antonio Montalbán for helpful discussion and to an anonymous referee for many useful comments and suggestions.

The first author is supported in part by Young Russian Mathematics award.

[1]Of course, by *well-ordered* here we mean *pre-well-ordered*.

Theorem 1.2. *The restriction of $\prec_{\Pi^1_1}$ to the Π^1_1-sound extensions of ACA_0 is well-founded.*

Accordingly, we can attach a well-founded rank—*reflection rank*—to Π^1_1 sound extensions of ACA_0 in the $\prec_{\Pi^1_1}$ ordering.

Definition 1.3. The *reflection rank* of T is the rank of T in the ordering $\prec_{\Pi^1_1}$ restricted to Π^1_1 sound extensions of ACA_0. We write $|T|_{\mathsf{ACA}_0}$ to denote the reflection rank of T.

What is the connection between the reflection rank of T and the Π^1_1 proof-theoretic ordinal of T? Recall that the Π^1_1 *proof-theoretic ordinal* $|T|_{\mathsf{WO}}$ of a theory T is the supremum of the order-types of T-provably well-founded primitive recursive linear orders. We will show that the reflection ranks and proof-theoretic ordinals of theories are closely connected. Recall that ACA_0^+ is axiomatized over ACA_0 by the statement "for every X, the ω^{th} jump of X exists."

Theorem 1.4. *For any Π^1_1-sound extension T of ACA_0^+, $|T|_{\mathsf{ACA}_0} = |T|_{\mathsf{WO}}$.*

In general, if $|T|_{\mathsf{ACA}_0} = \boldsymbol{\alpha}$ then $|T|_{\mathsf{WO}} \geqslant \boldsymbol{\varepsilon_\alpha}$. We provide examples of theories such that $|T|_{\mathsf{ACA}_0} = \boldsymbol{\alpha}$ and $|T|_{\mathsf{WO}} > \boldsymbol{\varepsilon_\alpha}$. Nevertheless for many theories T with $|T|_{\mathsf{ACA}_0} = \boldsymbol{\alpha}$ we have $|T|_{\mathsf{WO}} = \boldsymbol{\varepsilon_\alpha}$.

To prove these results, we extend techniques from the proof theory of iterated reflection principles to the second-order context. In particular, we focus on iterated Π^1_1 reflection. Roughly speaking, the theories $\mathbf{R}^\alpha_{\Pi^1_1}(T)$ of α-iterated Π^1_1-reflection over T are defined as follows

$$\mathbf{R}^0_{\Pi^1_1}(T) := T$$
$$\mathbf{R}^\alpha_{\Pi^1_1}(T) := T + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\Pi^1_1}\big(\mathbf{R}^\beta_{\Pi^1_1}(T)\big) \text{ for } \alpha \succ 0.$$

The formalization of this definition in arithmetic requires some additional efforts; see §2 for details.

Iterated reflection principles have been used previously to calculate proof-theoretic ordinals. For instance, Schmerl [22] used iterated reflection principles to establish bounds on provable arithmetical transfinite induction principles for fragments of PA. Beklemishev [2] has also calculated proof-theoretic ordinals of subsystems of PA via iterated reflection. These results differ from ours in two important ways. First, these results concern only theories in the language of first-order arithmetic, and hence do not engender calculations of Π^1_1 proof-theoretic ordinals. Second, these results are notation-dependent, i.e., they involve the calculation of proof-theoretic ordinals *modulo* the choice of a particular (natural) ordinal notation system. We are concerned with Π^1_1 reflection. Hence, in light of Theorem 1.2, we are able to calculate proof-theoretic ordinals in a manner that is not sensitive to the choice of a particular ordinal notation system.

Theorem 1.5. *Let α be an ordinal notation system with the order type $|\alpha| = \boldsymbol{\alpha}$. Then $|\mathbf{R}^\alpha_{\Pi^1_1}(\mathsf{ACA}_0)|_{\mathsf{ACA}_0} = \boldsymbol{\alpha}$ and $|\mathbf{R}^\alpha_{\Pi^1_1}(\mathsf{ACA}_0)|_{\mathsf{WO}} = \boldsymbol{\varepsilon_\alpha}$.*

It is possible to prove Theorem 1.4 and Theorem 1.5 by formalizing infinitary derivations in ACA_0 and appealing to cut-elimination, and in an early draft of this paper we did just that. Lev Beklemishev suggested that it might be possible to prove these results with methods from the proof theory of iterated reflection principles, namely conservation theorems in the style of Schmerl [22]. Though these methods have become quite polished for studying subsystems of first-order arithmetic, they have not yet been extended to Π^1_1 ordinal analysis. Thus, we devote a section of the paper to developing these techniques in the context of second-order arithmetic. We thank Lev for encouraging us to pursue this approach. Our main result in this respect is the following conservation theorem, where $\Pi^1_1(\Pi^0_3)$ denotes the complexity class consisting of formulas of the form $\forall X\ \mathsf{F}$ where $\mathsf{F} \in \Pi^0_3$.

Theorem 1.6. $\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)$ *is* $\Pi^1_1(\Pi^0_3)$ *conservative over* $\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)$.

To prove this result, we establish connections between Π^1_1 reflection over second-order theories and reflection over arithmetical theories with free set variables.

Finally, we demonstrate that Theorem 1.2 could be used for straightforward well-foundedness proofs for certain ordinal notation systems. A recent development in ordinal analysis is the use of ordinal notation systems that are based on reflection principles. Roughly, the elements of such notation systems are reflection principles and they are ordered by proof-theoretic strength. Such notation systems have been extensively studied since Beklemishev [3] endorsed their use as an approach to the canonicity problem for ordinal notations. See [12] for a survey of such notation systems. We prove the well-foundedness of Beklemishev's reflection notation system for ε_0 using the well-foundedness of the $\prec_{\Pi^1_1}$-order. Previously, Beklemishev proved the well-foundedness of this system by constructing the isomorphism with Cantor's ordinal notation system for ε_0. We expect that our techniques—or extensions thereof—could be used to prove the well-foundedness of ordinal notation systems for stronger axiomatic theories.

Here is our plan for the rest of the paper. In §2 we fix our notation and introduce some key definitions. In §3 we present our technique for showing that certain classes of theories are well-founded (or nearly well-founded) according to various notions of proof-theoretic strength. Our first application of this technique establishes Theorem 1.2. In §3 we prove various conservation results that connect iterated reflection principles with transfinite induction. The theorems in §3 extend results of Schmerl from first-order theories to pseudo Π^1_1 theories, i.e., to theories axiomatized by formulas with at most free set variables, and to second-order theories. We conclude with a proof of Theorem 1.6. In §4 we establish connections between the reflection ranks and proof-theoretic ordinals of theories, including proofs of Theorem 1.4 and Theorem 1.5. In §5 we show how to use our results to prove the well-foundedness of ordinal notation systems based on reflection principles. In §6 we present an explicit example by proving the well-foundedness of Beklemishev's notation system for ε_0.

2. Definitions and notation

In this section we describe and justify our choice of meta-theory. We then fix some notation and present some key definitions. Finally, we describe a proof-technique that we will use repeatedly throughout the paper, namely, Schmerl's technique of reflexive induction.

2.1. Treatment of theories.

Recall that EA is a finitely axiomatizable theory in the language of arithmetic with the exponential function, i.e., in the signature $(0, 1, +, \times, 2^x, \leqslant)$. EA is characterized by the standard recursive axioms for addition, multiplication, and exponentiation as well as the induction schema for Δ_0 formulas. Note that by Δ_0 formulas we mean Δ_0 formulas in the language with exponentiation. EA is strong enough to facilitate typical approaches to arithmetization of syntax. Moreover, EA proves its own Σ_1 completeness.

We will also be interested in EA^+. EA^+ is a theory in the language of EA. EA^+ extends EA by the additional axiom "superexponentiation is total." By superexponentiation, we mean the function 2^x_x where $2^x_0 = x$ and $2^x_{y+1} = 2^{2^x_y}$. EA^+ is the weakest extension of EA in which the cut-elimination theorem is provable. Indeed, the cut-elimination theorem is equivalent to the totality of superexponentiation over EA. See [4] for details on EA and EA^+; see also [15] for details on EA and EA^+ in a slightly different formalism without an explicit symbol for exponentiation. We will use EA^+ as a meta-theory for proving many of our results.

In this paper we will examine theories in three different languages. First the language of first-order arithmetic, i.e., the language of EA. Second the language of first-order arithmetic extended with one additional free set variable X; we also call this the *pseudo-*Π^1_1 *language*. And finally the language of second-order arithmetic. The language of first-order arithmetic of course is a sublanguage of the other two languages. And we consider the pseudo-Π^1_1 language to be a sublanguage of the language of second-order arithmetic by identifying each pseudo-Π^1_1 sentence F with the second-order sentence $\forall X\ \mathsf{F}$.

In the first-order context we are interested in the standard arithmetical complexity classes Π_n and Σ_n. We write Π_∞ to denote the class of all arithmetical formulas. We write $\mathbf{\Pi}^0_n$ to denote the class of formulas that are just like Π_n formulas except that their formulas (potentially) contain a free set variable X. Formulas in the complexity class $\mathbf{\Pi}^0_n$ *cannot* have set quantifiers, and so contain *only* free set variables. Of course, the class $\mathbf{\Sigma}^0_n$ is defined dually to the class $\mathbf{\Pi}^0_n$. We write $\mathbf{\Pi}^0_\infty$ to denote the class of boldface arithmetical formulas, i.e., the class of arithmetical formulas (potentially) with a free set variable.

In the second-order context we are mostly interested in the standard analytical complexity classes Π^1_1 and Σ^1_1. However, we will also use other complexity classes. Suppose $\mathcal{C} \subset \mathcal{L}_2$ is one of the following classes of formulas: Π^0_m or Σ^0_m, for $m \geqslant 1$. Then we denote by $\Pi^1_n(\mathcal{C})$ the class of all the formulas of the form $\forall X_1 \exists X_2 \ldots QX_n\ \mathsf{F}$, where $\mathsf{F} \in \mathcal{C}$. We define $\Sigma^1_n(\mathcal{C})$ dually.

For a first-order theory T, we use $T(X)$ to denote the pseudo Π^1_1 pendant of T. For example, the theory $\mathsf{PA}(X)$ contains (i) the axioms of PA and (ii) induction axioms for all formulas in the language, including those with free set variables. The theories $\mathsf{EA}(X)$, $\mathsf{EA}^+(X)$, and $\mathsf{I\Sigma}_1(X)$ are defined analogously, i.e., their induction axioms are extended to include formulas with the free set variable X.

Formulas in any of the three languages we are working with can naturally be identified with words in a suitable finite alphabet, which, in turn, are naturally one-to-one encoded by numbers. Accordingly, we can fix a Gödel numbering of these languages. We denote the Gödel number of an expression τ by $\ulcorner \tau \urcorner$. Many natural syntactic relations (x is a logical axiom, z the result of applying Modus Ponens to x and y, x encodes a Π_n formula, etc.) are elementary definable and their simplest properties can be verified within EA. We also fix a one-to-one elementary coding of finite sequences of natural numbers. $\langle x_1, ..., x_n\rangle$ denotes the code of a sequence $x_1, ..., x_n$ and, for any fixed n, is an elementary function of $x_1, ..., x_n$.

We are concerned with recursively enumerable theories. Officially, a theory T is a Σ_1 formula $\mathsf{Ax}_T(x)$ that is understood as a formula defining the (Gödel numbers of) axioms of T in the standard model of arithmetic, i.e., the set of axioms of T

is $\{\varphi : \mathbb{N} \vDash \mathsf{Ax}_T(\varphi)\}$. Thus, we are considering theories *intensionally*, via their axioms, rather than as deductively closed sets of formulas.

Since our base theory EA is fairly weak, we have to be careful with our choice of formalizations of proof predicates. Namely, we want our provability predicate to be Σ_1. And due to this we can't use the straightforwardly defined predicates $\mathsf{PrfNat}_T(x, y)$: x is a Hilbert-style proof of y, where all axioms are either axioms of first-order logic or axioms of T. The predicates $\mathsf{PrfNat}_T(x, y)$ are equivalent to $\forall^b\Sigma_1$-formulas over EA ($\forall^b\Sigma_1$-formulas are the formulas starting with a bounded universal quantifier followed by Σ_1-formula). However, EA is too weak to equivalently transform $\forall^b\Sigma_1$-formulas to Σ_1-formulas; for this one needs the collection scheme $\mathsf{B}\Sigma_1$, which isn't provable in EA. We note that this doesn't affect most natural theories T, in particular, for any T with Δ_0 formula Ax_T, the predicate $\mathsf{PrfNat}_T(x, y)$ is equivalent to a Σ_1 formula over EA.

Nevertheless, to avoid this issue, we work with proof predicates that are forced to be Σ_1 in EA, which are sometimes called *smooth proof* predicates. In the definition of the smooth proof predicate, a "proof" is a pair consisting of an actual Hilbert style proof and a uniform bound for witnesses to the facts that axioms in the proof indeed are axioms. We simply write $\mathsf{Prf}_T(x, y)$ to formalize that x is a "smooth proof" of y in theory T. The predicates $\mathsf{Prf}_T(x, y)$ are Δ_0-formulas. The predicate $\mathsf{Pr}_T(y)$ is shorthand for $\exists x \mathsf{Prf}_T(x, y)$. We use the predicate $\mathsf{Con}(T)$ as shorthand for $\neg\mathsf{Pr}_T(\bot)$, where we fix $\bot$ to be some contradictory sentence.

The closed term $1 + 1 + ... + 1$ (n times) is the numeral of n and is denoted $\underline{n}$. We often omit the bar when no confusion can occur. We also often omit the corner quotes from Gödel numbers when no confusion can occur. For instance, we can encode the notion of a formula φ being provable in a theory T, by saying that there is a T-proof (a sequence subject to certain constraints) the last element of which is the numeral of the Gödel number of φ. However, instead of writing $\mathsf{Pr}_T(\underline{\ulcorner\varphi\urcorner})$ to say that φ is provable we simply write $\mathsf{Pr}_T(\varphi)$.

Suppose T and U are recursively enumerable theories in the same language. We write $T \sqsubseteq U$ if T is a subtheory of U; we can formalize the claim that $T \sqsubseteq U$ in arithmetic with the formula $\forall\varphi\big(\mathsf{Pr}_T(\varphi) \to \mathsf{Pr}_U(\varphi)\big)$. We write $T \equiv U$ if $T \sqsupseteq U$ and $U \sqsupseteq T$. For a class $\mathcal{C}$ of sentences of the language of T we write $T \sqsubseteq_{\mathcal{C}} U$ if the set of $\mathcal{C}$-theorems of T is a subset of $\mathcal{C}$-theorems of U; this could be naturally formalized in arithmetic with the formula $\forall\varphi \in \mathcal{C}\big(\mathsf{Pr}_T(\varphi) \to \mathsf{Pr}_U(\varphi)\big)$. We write $T \equiv_{\mathcal{C}} U$ if $T \sqsubseteq_{\mathcal{C}} U$ and $U \sqsubseteq_{\mathcal{C}} T$.

We will be interested in partial truth-definitions for various classes of formulas for which we could prove Tarski's bi-conditionals. For a class $\mathcal{C}$ of formulas we call a formula $\mathsf{Tr}_{\mathcal{C}}(x)$ a partial truth definition for $\mathcal{C}$ over a theory T, if $\mathsf{Tr}_{\mathcal{C}}(x)$ is from the class $\mathcal{C}$ and

$$T \vdash \varphi(\vec{x}) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\varphi(\vec{x})), \text{ for all } \varphi(\vec{x}) \text{ from } \mathcal{C}.$$

Moreover, we will work only with truth definitions such that the above property is provable in EA.

In the book by Hájek and Pudlák [15, §I.1(d)] there is a construction of partial truth definitions for classes Π_n and Σ_n, $n \geqslant 1$, over $\mathsf{I}\Sigma_1$. However, we will use a sharper construction of partial truth definitions for classes Π_n and Σ_n, $n \geqslant 1$, over EA which could be found in [9, Appendix A]. And we will use truth definitions for classes $\boldsymbol{\Pi}_n$ and $\boldsymbol{\Sigma}_n$, $n \geqslant 1$, over $\mathsf{EA}(X)$ that as well were constructed in [9, Appendix A].

In the case of second-order arithmetic there are partial truth definitions for classes $\Pi^1_n(\Pi^0_m)$, $\Sigma^1_n(\Sigma^0_m)$, $\Sigma^1_n(\Pi^0_m)$, and $\Pi^1_n(\Sigma^0_m)$, where $m \geqslant 1$, over RCA_0. One could easily construct this partial truth definitions from the partial truth definitions for classes $\mathbf{\Pi}_n$ and $\mathbf{\Sigma}_n$ over $\mathsf{EA}(X)$. However, over ACA_0 it is possible to construct partial truth definitions for the classes Π^1_n and Σ^1_n, $n \geqslant 1$. Let $\mathbf{\Sigma}^1_1$ be the class of Σ^1_1-formulas with a set parameter X. It is easy to construct partial truth definitions for classes Π^1_n and Σ^1_n, $n \geqslant 1$, from a partial truth definition for $\mathbf{\Sigma}^1_1$. Simpson [24, Lemma V.1.4] proves that for each Σ^1_1 formula $\varphi(X)$ there exists a Δ^0_0 formula $\theta(x, y)$ such that

$$\mathsf{ACA}_0 \vdash \forall X \big(\varphi(X) \leftrightarrow (\exists f \colon \mathbb{N} \to \mathbb{N})\; \forall m\; \theta_\varphi(X \restriction m, f \restriction m)\big).$$

Here $X \restriction m$ is the natural number encoding the finite set $X \cap \{0, \ldots, m-1\}$ and $f \restriction m$ is the code of the finite sequence $\langle f(0), \ldots, f(m-1)\rangle$. From Simpson's proof it is easy to extract a Kalmar elementary algorithm for constructing the formula θ_φ from a formula φ. And by the same argument as Simpson we show that the Σ^1_1-formula $(\exists f \colon \mathbb{N} \to \mathbb{N})\; \forall m\; \mathsf{Tr}_{\Pi^0_1}(\theta_x(X \restriction m, f \restriction m))$ is a partial truth definition $\mathsf{Tr}_{\mathbf{\Sigma}^1_1}(X, x)$ for the class $\mathbf{\Sigma}^1_1$ over ACA_0.

2.2. **Ordinal notations.** There are many ways of treating ordinal notations in arithmetic. We choose one specific method that will be suitable when we work in the theory EA^+ (and its extensions). Our results will be valid for other natural choices of treatment of ordinal notations, but some of the proofs would have to be tweaked slightly.

Often we will use ordinal notation systems within formal theories that couldn't prove (or even express) the well-foundedness of the notation systems. Also, most of our results are intensional in nature and don't require the notation system to be well-founded from an external point of view. Due to this, our definition of an ordinal notation system does not require it to be well-founded.

Officially, an ordinal notation α is a tuple $\langle \varphi(x), \psi(x, y), p, n\rangle$ where $\varphi, \psi \in \Delta_0$, $\varphi(n)$ is true according to Tr_{Σ_1}, and p is an EA proof of the fact that on the set $\{x \mid \varphi(x)\}$ the order

$$x \prec_\alpha y \stackrel{\text{def}}{\iff} \psi(x, y)$$

is a strict linear order. More formally p is an EA proof of the conjunction of the following sentences:

(1) $\forall x, y, z \big(\varphi(x) \land \varphi(y) \land \varphi(z) \land \psi(x, y) \land \psi(y, z) \to \psi(x, z)\big)$ (Transitivity);
(2) $\forall x \big(\varphi(x) \to \neg\psi(x, x)\big)$ (Irreflexivity);
(3) $\forall x, y \Big(\varphi(x) \land \varphi(y) \land x \neq y \to \big(\psi(x, y) \land \neg\psi(y, x)\big) \lor \big(\psi(y, x) \land \neg\psi(x, y)\big)\Big)$ (Antisymmetry).

We now define a partial order $\prec$ on the set of all notation systems. Any tuples $\alpha = \langle \varphi, \psi, p, n\rangle$ and $\alpha' = \langle \varphi', \psi', p', n\rangle$ are $\prec$-incomparable if either $\varphi \neq \varphi'$, or $\psi \neq \psi'$, or $p \neq p'$. If α, β are of the form $\alpha = \langle \varphi, \psi, p, n\rangle$ and $\beta = \langle \varphi, \psi, p, m\rangle$, we put $\alpha \prec \beta$ if $\mathsf{Tr}_{\Sigma_1}\big(\psi(n, m)\big)$ but $\mathsf{Tr}_{\Sigma_1}\big(\neg\psi(m, n)\big)$.

Clearly the relation $\prec$ and the property of being an ordinal notation system are expressible by Σ_1-formulas. In EA^+ we could expand the language by a definable superexponentiation function 2^x_y. Since the superexponentiation function is EA^+ provably monotone, by a standard technique one could show that EA^+ proves induction for the class $\Delta_0(2^x_y)$ of formulas with bounded quantifiers in the expanded language. It is easy to show that over EA^+ the truth of Δ_0-formulas according to the Σ_1-truth predicate could be expressed by a $\Delta_0(2^x_y)$ formula. Thus, the order $\prec$ and the property of being an ordinal notation system are expressible by $\Delta_0(2^x_y)$ formulas, which allows us to reason about them in EA^+ is a straightforward manner.

Let us show that EA^+ proves that $\prec$ is a disjoint union of linear orders. First we note that the theory EA^+ proves the Π_2 soundness of EA (i.e. $\mathsf{RFN}_{\Pi_2}(\mathsf{EA})$, see section below). And we note that for any $\alpha = \langle \varphi(x), \psi(x,y), p, n\rangle$ the conclusion of p (conjunction of sentences (1)–(3)) is EA-provably equivalent to a Π_1 sentence. Hence for any notation system $\alpha = \langle \varphi(x), \psi(x,y), p, n\rangle$ the theory EA^+ proves that the corresponding conjunction of sentences (1)–(3) is true. Using this we easily prove in EA^+ that $\prec$ is a linear ordering, when restricted to the tuples that share the same first three components.

For an ordinal notation α the value of $|\alpha|$ is either an ordinal or ∞. If the lower cone $(\{\beta \mid \beta \prec \alpha\}, \prec)$ is well-founded, then $|\alpha|$ is the ordinal isomorphic to the well-ordering $(\{\beta \mid \beta \prec \alpha\}, \prec)$. Otherwise, $|\alpha| = \infty$. In other words, $|\alpha|$ is the well-founded rank of α in the $\prec$-order.

An alternative (more standard) approach to treating ordinal notations in arithmetic is to fix an elementary ordinal notation up to some ordinal $\boldsymbol{\alpha}$. This is a fixed linear order $\mathbf{L} = (\mathcal{D}_{\mathbf{L}}, \prec_{\mathbf{L}})$, where both $\mathcal{D}_{\mathbf{L}} \subseteq \mathbb{N}$ and $\prec_{\mathbf{L}} \subseteq \mathbb{N}\times\mathbb{N}$ are given by Δ_0 formulas such that (i) $\mathbf{L}$ is provably linear in EA, (ii) $\mathbf{L}$ is well-founded, and (iii) the order type of $\mathbf{L}$ is $\boldsymbol{\alpha}$. It has been empirically observed that the ordinal notation systems that arise in ordinal analysis results in proof theory are of this kind; see, e.g., [21]. Note that from any $\mathbf{L}$ of this sort we could easily form an ordinal notation (in our sense) α such that there is a Kalmar elementary isomorphism f between $\mathbf{L}$ and $(\{\beta \mid \beta \prec \alpha\}, \prec)$; moreover, the latter is provable in EA^+.

Further we will work with ordinal notation systems that are given by some combinatorially defined system of terms and order on them. The standard example of such a system is the Cantor ordinal notation system up to ε_0. For the notations that we will consider it will be always possible to formalize in EA the definition and proof that the order is linear. Thus, as described above, we will be able to form an ordinal notation α such that there will be a natural isomorphism between $(\{\beta \mid \beta \prec \alpha\}, \prec)$ and the initial combinatorially defined ordinal notation system. We will make transitions from combinatorial definitions of notation systems to ordinal notation systems in our sense without any further comments.

Moreover, we will use expressions like ω^α and ε_α, where α is some ordinal notation system. Let us consider a notation system $\alpha = \langle \varphi(x), \psi(x,y), p, n\rangle$ and define the notation system $\omega^\alpha = \langle \varphi'(x), \psi'(x,y), p', n'\rangle$. We want the order $\prec_{\omega^\alpha}$ to be the order on the terms $\omega^{a_1} + \ldots + \omega^{a_k}$, where $a_1 \succeq_\alpha \ldots \succeq_\alpha a_k$. And the order $\prec_{\omega^\alpha}$ is defined as the usual order on Cantor normal forms, where we compare a_i by the order $\prec_\alpha$. By arithmetizing this definition of $\prec_{\omega^\alpha}$ we get φ', ψ', and p'. We put n' to be the number encoding the term ω^n. Note that, according to this definition, α and ω^α are $\prec$-incomparable. However, if $\alpha \prec \beta$, then $\omega^\alpha \prec \omega^\beta$.

The definition of the notation system ε_α is similar to that of ω^α. The system of terms for ε_α consists of nested Cantor normal forms built up from 0 and elements ε_a, for $a \in \mathsf{dom}(\prec_\alpha)$. The comparission of nested Cantor normal forms is defined in the standard fashion, where we compare elements ε_a and ε_b as $a \prec_\alpha b$.

2.3. **Reflection principles.** Suppose $\mathcal{C}$ is some class of formulas in one of the languages that we consider and T is a theory in the same language. The uniform

$\mathcal{C}$ reflection principle $\mathsf{RFN}_{\mathcal{C}}(T)$ over T is the schema

$$\forall \vec{x}\big(\mathsf{Pr}_T(\varphi(\vec{x})) \to \varphi(\vec{x})\big)$$

for all $\varphi \in \mathcal{C}$, where $\vec{x}$ are free number variables and $\varphi(\vec{x})$ contains no other variables.

In those cases for which we have a truth-definition for $\mathcal{C}$ in T the scheme $\mathsf{RFN}_{\mathcal{C}}(T)$ can be axiomatized by the single sentence

$$\forall \varphi \in \mathcal{C}\big(\mathsf{Pr}_T(\varphi) \to \mathsf{Tr}_{\mathcal{C}}(\varphi)\big).$$

Given an ordinal notation system $\prec$, we informally define the operation $\mathbf{R}^{\cdot}_{\mathcal{C}}(\cdot)$ of iterated $\mathcal{C}$ reflection along $\prec$ as follows.

$$\mathbf{R}^0_{\mathcal{C}}(T) := T$$
$$\mathbf{R}^{\alpha}_{\mathcal{C}}(T) := T + \bigcup_{\beta \prec \alpha} \mathsf{RFN}_{\mathcal{C}}\big(\mathbf{R}^{\beta}_{\mathcal{C}}(T)\big) \text{ for } \alpha \succ 0.$$

More formally, we appeal to Gödel's fixed point lemma in EA. We fix a formula $\mathsf{RFN\text{-}Inst}_{\mathcal{C}}(U, x)$, where U and x are first-order variables, that formalizes the fact that x is an instance of the scheme $\mathsf{RFN}_{\mathcal{C}}(U)$. We now want to define a Σ_1 formula $\mathsf{Ax}_{\mathbf{R}^{\alpha}_{\mathcal{C}}(T)}(x)$ (note that α, T, and x are arguments of the formula) that defines the set of axioms of the theories $\mathbf{R}^{\alpha}_{\mathcal{C}}(T)$. We define the formula as a fixed point:

$$\mathsf{EA} \vdash \mathsf{Ax}_{\mathbf{R}^{\alpha}_{\mathcal{C}}(T)}(x) \leftrightarrow \big(\mathsf{Ax}_T(x) \vee \exists \beta \prec \alpha\ \mathsf{RFN\text{-}Inst}_{\mathcal{C}}(\mathbf{R}^{\beta}_{\mathcal{C}}(T), x))\big),$$

note that when we substitute $\mathbf{R}^{\beta}_{\mathcal{C}}(T)$ in $\mathsf{RFN\text{-}Inst}_{\mathcal{C}}$ we actually substitute (the Gödel number of) $\mathsf{Ax}_{\mathbf{R}^{\beta}_{\mathcal{C}}(T)}$.

Beklemishev introduced this approach to defining progressions of iterated reflection in [1]; the reader can find a more modern version of this approach in [8]. It is easy to prove that this definition of progressions of iterated reflection provides a unique (up to EA provable deductive equivalence) definition of the theories $\mathsf{RFN}^{\alpha}_{\mathcal{C}}(T)$.

2.4. **Reflexive induction.** We often employ Schmerl's technique of *reflexive induction*. Reflexive induction is a way of simulating large amounts of transfinite induction in weak theories. The technique is facilitated by the following theorem; we include the proof of the theorem, which is very short.

Theorem 2.1 (Schmerl)**.** *Let T be a recursively axiomatized theory (in one of the languages that we consider) that contains* EA. *Suppose*

$$T \vdash \forall \alpha\Big(\mathsf{Pr}_T\big(\forall \beta \prec \alpha\ \varphi(\beta)\big) \to \varphi(\alpha)\Big).$$

Then $T \vdash \forall \alpha\ \varphi(\alpha)$.[2]

Proof. Suppose that $T \vdash \forall \alpha\Big(\mathsf{Pr}_T\big(\forall \beta \prec \alpha\ \varphi(\beta)\big) \to \varphi(\alpha)\Big)$. We infer that

$$T \vdash \forall \alpha \mathsf{Pr}_T\big(\forall \beta \prec \alpha\ \varphi(\beta)\big) \to \forall \alpha\ \varphi(\alpha),$$

whence it follows that

$$T \vdash \mathsf{Pr}_T\big(\forall \alpha\ \varphi(\alpha)\big) \to \forall \alpha\ \varphi(\alpha).$$

Löb's theorem then yields $T \vdash \forall \alpha\ \varphi(\alpha)$. ❑

[2]Schmerl proved this result over the base theory PRA. Beklemishev [2] weakened the base theory to EA.

Accordingly, to prove claims of the form $T \vdash \forall\alpha\ \varphi(\alpha)$, we often prove that $T \vdash \forall\alpha\Big(\mathsf{Pr}_T\big(\forall\beta \prec \alpha\ \varphi(\beta)\big) \to \varphi(\alpha)\Big)$ and infer the desired claim by Schmerl's Theorem. While working inside T, we refer to the assumption $\mathsf{Pr}_T\big(\forall\beta \prec \alpha\ \varphi(\beta)\big)$ as the *reflexive induction hypothesis.*

3. Well-foundedness and reflection principles

In this section we develop a technique for showing that certain orders on axiomatic theories exhibit a well-foundeness like properties. The coarsest order that we will consider is Π^1_1 reflection order for which we will prove that its restriction to Π^1_1 sound theories is well-founded. For weaker reflection and consistency orders we will prove only some well-foundedness like properties. Also we note that the same technique is used in [11, Theorem 3.2] to prove certain facts about axiomatic theories of truth and in [18, Theorem 1.1] to prove a recursion-theoretic result concerning the hyper-degrees.

Our technique is inspired by H. Friedman's [13] proof of the following result originally due to Steel [27]; recall that $\leqslant_T$ denotes Turing reducibility.

Theorem 3.1. *Let $P \subset \mathbb{R}^2$ be arithmetic. Then there is no sequence $(x_n)_{n<\omega}$ of reals such that for every n, both $x_n \geqslant_T x'_{n+1}$ and also x_{n+1} is the unique real y such that $P(x_n, y)$.*

Friedman and Steel were not directly investigating the well-foundedness of axiomatic systems, but rather an analogous phenomenon from recursion theory, namely, the well-foundedness of natural Turing degrees under Turing reducibility. The adaptability of Friedman's proof arguably strengthens the analogy between these phenomena.

In this section we study both first and second order theories. The first theory that we treat with our technique is ACA_0, a subsystem of second-order arithmetic that has been widely studied in reverse mathematics. ACA_0 is arithmetically conservative over PA. We then turn to other applications of our technique. We consider RCA_0, another subsystem of second-order arithmetic and familiar base theory from reverse mathematics. RCA_0 is conservative over $\mathsf{I\Sigma}_1$. We then turn to first-order theories, and we study elementary arithmetic EA as our object theory.

3.1. Π^1_1-Reflection.

In this subsection we examine the ordering $\prec_{\Pi^1_1}$ on r.e. extensions of ACA_0, where

$$T \prec_{\Pi^1_1} U \overset{\text{def}}{\iff} U \vdash \mathsf{RFN}_{\Pi^1_1}(T).$$

We will show that there are no infinite $\prec_{\Pi^1_1}$ descending sequences of Π^1_1 sound extensions of ACA_0. We recall that, provably in ACA_0, a theory T is Π^1_1 sound if and only if T is consistent with any true Σ^1_1 statement.

Theorem 3.2. (ACA_0) *The restriction of the order $\prec_{\Pi^1_1}$ to Π^1_1-sound r.e. extensions of ACA_0 is well-founded.*

Proof. In order to prove the result in ACA_0 we show the inconsistency of the theory ACA_0 plus the following statement DS, which says that there *is* a descending sequence of Π^1_1 sound extensions of ACA_0 in the $\prec_{\Pi^1_1}$ ordering:

$$\mathsf{DS} := \exists E\colon \langle T_i \mid i \in \mathbb{N}\rangle(\mathsf{RFN}_{\Pi^1_1}(T_0) \wedge \forall x\ \mathsf{Pr}_{T_x}(\mathsf{RFN}_{\Pi^1_1}(T_{x+1})) \wedge \forall x(T_x \sqsupseteq \mathsf{ACA}_0))$$

Note that $E\colon \langle T_i \mid i \in \mathbb{N}\rangle$ is understood to mean that E is a set encoding a sequence $\langle T_0, T_1, T_2, \ldots\rangle$ of r.e. theories.

If we prove that $\mathsf{ACA}_0 + \mathsf{DS}$ proves its own consistency, then the inconsistency of $\mathsf{ACA}_0 + \mathsf{DS}$ follows from Gödel's second incompleteness theorem. We reason in $\mathsf{ACA}_0 + \mathsf{DS}$ to to prove consistency of $\mathsf{ACA}_0 + \mathsf{DS}$.

Let $E\colon \langle T_i \mid i \in \mathbb{N}\rangle$ be a sequence of theories witnessing the truth of DS. Let us consider the sentence F

$$\exists U\colon \langle S_i \mid i \in \mathbb{N}\rangle (S_0 = T_1 \wedge \forall x\, \mathsf{Pr}_{S_x}(\mathsf{RFN}_{\Pi^1_1}(S_{x+1})) \wedge \forall x (S_x \sqsupseteq \mathsf{ACA}_0)).$$

The sentence F is true since we could take $\langle T_{i+1} : i \in \mathbb{N}\rangle$ as U. It is easy to observe that F is Σ^1_1.

From $\mathsf{RFN}_{\Pi^1_1}(T_0)$ we get that T_0 is consistent with any true Σ^1_1 statement. Thus, we infer that

$$\mathsf{Con}(T_0 + \mathsf{F}).$$

Now using the fact that $\mathsf{Pr}_{T_0}\big(\mathsf{RFN}_{\Pi^1_1}(T_1)\big)$ and that $T_0 \sqsupseteq \mathsf{ACA}_0$ we conclude,

$$\mathsf{Con}(\mathsf{ACA}_0 + \mathsf{RFN}_{\Pi^1_1}(T_1) + \mathsf{F}).$$

But it is easy to see that $\mathsf{RFN}_{\Pi^1_1}(T_1) + \mathsf{F}$ implies DS in ACA_0. In particular, we may take $\langle T_1, T_2, ...\rangle$ as our new witness to DS. Thus, we conclude that $\mathsf{Con}(\mathsf{ACA}_0 + \mathsf{DS})$. ❑

We now observe that a similar result holds over RCA_0. To do so, we consider formulas from the complexity class $\Pi^1_1(\Pi^0_3)$ (see §2.4). It is easy to see that the proof of Theorem 3.3 remains valid if we replace the theory ACA_0 with RCA_0, the complexity class Π^1_1 with $\Pi^1_1(\Pi^0_3)$, and the complexity class Σ^1_1 with $\Sigma^1_1(\Pi^0_2)$. Thus, we also infer the following.

Theorem 3.3. (RCA_0) *The restriction of the order* $\prec_{\Pi^1_1(\Pi^0_3)}$ *to* $\Pi^1_1(\Pi^0_3)$*-sound r.e. extensions of* RCA_0 *theories is well-founded.*

3.2. Π_3 **soundness.** In this subsection we study the complexity of descending sequences of r.e. theories with respect to Π_3 soundness. We recall that (provably in EA) a theory T is Π_3 sound just in case T is 2-consistent, i.e., just in case T is consistent with any true Π_2 sentence.

Theorem 3.4. *There is no recursively enumerable sequence* $(T_n)_{n<\omega}$ *of r.e. extensions of* EA *such that* T_0 *is* Π_3 *sound and such that for every* n, $T_n \vdash \mathsf{RFN}_{\Pi_3}(T_{n+1})$.

Proof. If the theorem fails, then the following sentence is true,

$$\mathsf{DS} := \exists e\colon \langle T_i \mid i \in \mathbb{N}\rangle \Big(\mathsf{RFN}_{\Pi_3}(T_0) \wedge \forall x\, \mathsf{Pr}_{T_x}\big(\mathsf{RFN}_{\Pi_3}(T_{x+1})\big)\Big)$$

where $\exists e : \langle T_i : i \in \mathbb{N}\rangle$ is understood to mean that e is an index for a Turing machine enumerating the sequence $\langle T_0, T_1, ...\rangle$.

We show that $\mathsf{EA} + \mathsf{DS}$ proves its own consistency, whence, by Gödel's second incompleteness theorem, $\mathsf{EA} + \mathsf{DS}$ is inconsistent and hence DS is false.

Work in $\mathsf{EA} + \mathsf{DS}$. Since DS is true, it has some witness $e\colon \langle T_i \mid i \in \mathbb{N}\rangle$. We now consider the sequence e' that results from omitting T_0 from e. More formally, we consider the sequence $e'\colon \langle T'_i \mid i \in \mathbb{N}\rangle$, which is numerated by the Turing functional $\{e'\} : x \mapsto \{e\}(x+1)$. That is, for each i, $T'_i = T_{i+1}$.

From DS we infer that for all x, $T_{x+1} \vdash \mathsf{RFN}_{\Pi_3}(T_{x+2})$. Thus, for every x, $T'_x \vdash \mathsf{RFN}_{\Pi_3}(T'_{x+1})$ by the definition of e'.

From the first conjunct of DS we infer that $\mathsf{RFN}_{\Pi_3}(T_0)$. That is, T_0 is consistent with any Π_2 truth. Thus, we infer that

$$T_0 + \forall x\ \mathsf{Pr}_{T'_x}\big(\mathsf{RFN}_{\Pi_3}(T'_{x+1})\big)$$

is consistent.

On the other hand, from DS we infer that T_0 *proves* the Π_3 soundness of T'_0. So it is consistent that e' witnesses DS. ❑

3.3. **Consistency.** In this subsection we provide a new proof of a theorem independently due to H. Friedman, Smorynski, and Solovay (see [17, 26]). Before stating the theorem we recall that, EA proves the equivalence of, the consistency sentences $\mathsf{Con}(T)$ and the Π_1-reflection principle $\mathsf{RFN}_{\Pi_1}(T)$.

Theorem 3.5. *There is no recursively enumerable sequence $(T_n)_{n<\omega}$ of r.e. extensions of* EA *such that T_0 is consistent and such that* $\mathsf{EA} \vdash \forall x\ \mathsf{Pr}_{T_x}\big(\mathsf{Con}(T_{x+1})\big)$.

Proof. Suppose, toward a contradiction, that there is a recursively enumerable sequence $(T_n)_{n<\omega}$ of r.e. extensions of EA such that T_0 is consistent and such that

$$\mathsf{EA} \vdash \forall x\ \mathsf{Pr}_{T_x}\big(\mathsf{Con}(T_{x+1})\big).$$

Since EA is sound, we also infer that for every n, $T_n \vdash \mathsf{Con}(T_{n+1})$. Thus the following sentence is true.

$$\mathsf{DS} := \exists e \colon \langle T_i \mid i \in \mathbb{N}\rangle \Big(\mathsf{Con}(T_0) \wedge \mathsf{Pr}_{\mathsf{EA}}\big(\forall x\ \mathsf{Pr}_{T_x}(\mathsf{Con}(T_{x+1}))\big) \wedge \forall x\ \mathsf{Pr}_{T_x}\big(\mathsf{Con}(T_{x+1})\big)\Big)$$

where $\exists e : \langle T_i : i \in \mathbb{N}\rangle$ is understood to mean that e is an index for a Turing machine enumerating the sequence $\langle T_0, T_1, ...\rangle$.

We show that $\mathsf{EA} + \mathsf{DS}$ proves its own consistency, whence, by Gödel's second incompleteness theorem, $\mathsf{EA} + \mathsf{DS}$ is inconsistent and hence DS is false.

Work in $\mathsf{EA} + \mathsf{DS}$. Since DS is true, it has some witness $e \colon \langle T_i \mid i \in \mathbb{N}\rangle$. We consider the sequence $e' \colon \langle T'_i \mid i \in \mathbb{N}\rangle$ that results from dropping T_0 from the sequence produced by e. More formally, we consider the sequence e' which is numerated by the Turing functional $\{e'\} : x \mapsto \{e\}(x+1)$.

Claim. *e' is provably a witness to* DS *in T_0.*

To see that e' provably witnesses the third conjunct of DS in T_0, we reason as follows.

$$\begin{aligned}
\mathsf{EA} &\vdash \forall x\ \mathsf{Pr}_{T_{x+1}}\mathsf{Con}(T_{x+2}) \text{ by } \mathsf{DS}.\\
\mathsf{EA} &\vdash \forall x\ \mathsf{Pr}_{T'_x}\mathsf{Con}(T'_{x+1}) \text{ since } T'_x = T_{x+1} \text{ by definition of } e'.\\
T_0 &\vdash \forall x\ \mathsf{Pr}_{T'_x}\mathsf{Con}(T'_{x+1}) \text{ since } T_0 \text{ extends } \mathsf{EA}.
\end{aligned}$$

To see that e' provably witnesses the second conjunct of DS in T_0, we reason as follows.

$$\begin{aligned}
\mathsf{EA} &\vdash \forall x\ \mathsf{Pr}_{T'_x}\mathsf{Con}(T'_{x+1}) \text{ as above.}\\
\mathsf{EA} &\vdash \mathsf{Pr}_{\mathsf{EA}}\big(\forall x\ \mathsf{Pr}_{T'_x}\mathsf{Con}(T'_{x+1})\big) \text{ by the } \Sigma_1 \text{ completeness of } \mathsf{EA}.\\
T_0 &\vdash \mathsf{Pr}_{\mathsf{EA}}\big(\forall x\ \mathsf{Pr}_{T'_x}\mathsf{Con}(T'_{x+1})\big) \text{ since } T_0 \text{ extends } \mathsf{EA}.
\end{aligned}$$

We now show that e' provably witnesses the first conjunct of DS in T_0. From the first conjunct of DS we infer that $\mathsf{Con}(T_0)$. It follows that T_0 is Π_1 sound. We reason as follows.

$$T_0 \vdash \mathsf{Con}(T_1) \text{ by } \mathsf{DS}.$$
$$T_0 \vdash \mathsf{Con}(T_0') \text{ since provably } T_0' = T_1.$$

We then infer that $\mathsf{Con}(T_0')$ by the Π_1 soundness of T_0. So e' is provably a witness to DS in a consistent theory. Therefore $\mathsf{EA} + \mathsf{DS}$ is consistent. ❑

Remark 3.6. Note that we just proved the non-existence of EA-*provably* descending r.e. sequences. Without the condition of EA provability such descending sequences *do* exist. H. Friedman, Smorynski, and Solovay independently proved that there is a recursive sequence $\langle T_0, T_1, ...\rangle$ of consistent extensions of EA such that for all n, $T_n \vdash \mathsf{Con}(T_{n+1})$, answering a question of Gaifman; see [26] for details.

3.4. Π_2 **soundness.** We now know that there are *no* recursive descending sequences of Π_3 sound theories with respect to the Π_3 reflection order, but there *are* recursive descending sequences of consistent theories with respect to consistency strength. In this subsection we treat the remaining case, namely, Π_2 soundness. We prove that there *is* an infinite sequences $\langle T_0, T_1, ...\rangle$ of Π_2 sound extensions of EA such that for all n, $T_n \vdash \mathsf{RFN}_{\Pi_2}(T_{n+1})$. In this sense, Theorem 3.4 is best possible.

In the section, for technical reasons it will be useful for us to impose some natural conditions on our proof predicate. We make sure that any proof in our proof system has only one conclusion, whence

$$\mathsf{EA} \vdash \forall x, y_1, y_2\Big(\big(\mathsf{Prf}_T(x, y_1) \wedge \mathsf{Prf}_T(x, y_2)\big) \to y_1 = y_2\Big).$$

Moreover, we arrange the proof system so that indices for statements are less than or equal to the indices for their proofs, i.e.,

$$\mathsf{EA} \vdash \forall x, y(\mathsf{Prf}_T(x, y) \to y \leqslant x). \tag{1}$$

Note that the conclusions of the theorems in our paper are not sensitive to the choice of proof predicate as long as the resulting provability predicates are EA-provably equivalent. And it is easy to see that even if our initial choice of $\mathsf{Prf}_T(x, \varphi)$ didn't satisfied the mentioned conditions, it is easy to modify it to satisfy the conditions, while preserving the provability predicate $\mathsf{Pr}_T(\varphi)$ up to EA-provable equivalence.

Before proving the theorem we make a few more remarks preliminary remarks. We use the symbol $\dot{-}$ to denote the truncated subtraction function, i.e., $n \dot{-} m = n - m$ if $n > m$ and 0 otherwise. We remind the reader that, provably in EA, a theory is Σ_1 sound if and only if it is Π_2 sound. We also pause to make the following remark, which will invoke in the proof of the theorem.

Remark 3.7. For any Π_2 sound extension T of EA, the theory $T + \neg\mathsf{RFN}_{\Pi_2}(T)$ is Π_2 sound. This is actually an instance Gödel's second incompleteness theorem that is applied to 1-provability rather than the ordinary provability. Recall that 1-provability predicate $1\text{-}\mathsf{Pr}_T(\varphi)$ for a theory T is

$$\exists \psi \in \Sigma_2\big(\mathsf{Tr}_{\Sigma_2}(\psi) \wedge \mathsf{Pr}_T(\psi \to \varphi)\big). \tag{2}$$

The consistency notion that corresponds to 1-provability is precisely Π_2-soundness:

$$\mathsf{EA} \vdash \forall \varphi \ \big(\neg 1\text{-}\mathsf{Pr}_T(\neg\varphi) \leftrightarrow \mathsf{RFN}_{\Pi_2}(T + \varphi)\big). \tag{3}$$

It is easy to see that 1-provability predicate for a theory T satisfies the usual Hilbrt-Bernays-Löb derivability conditions. Thus Gödel's second incompleteness theorem for it states that if a theory $T \sqsupseteq \mathsf{EA}$ is Π_2-sound, then $\mathsf{RFN}_{\Pi_2}(T)$ is not 1-provable in T. And the latter is equivalent to Π_2-soundness of $T + \neg\mathsf{RFN}_{\Pi_2}(T)$.

We are now ready for the proof of the theorem.

Theorem 3.8. *There is a recursive sequence* $(\varphi_n)_{n<\omega}$ *of* Π_2*-sound sentences such that, for each* n, $\mathsf{EA} + \varphi_n \vdash \mathsf{RFN}_{\Pi_2}(\mathsf{EA} + \varphi_{n+1})$.

Proof. For each $n \in \mathbb{N}$, we define the sentence φ_n as follows:

$$\varphi_n := \exists\psi \in \Sigma_1 \exists p \Big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \wedge \neg\mathsf{True}_{\Sigma_1}(\psi) \wedge \mathsf{RFN}_{\Pi_2}\big(\mathbf{R}^{p \dot{-} n}_{\Pi_2}(\mathsf{EA})\big)\Big)$$

That is, φ_n expresses "$\mathsf{I}\Sigma_2$ proves a false Σ_1 sentence via a proof p, and Π_2 reflection for EA can be iterated up to $p \dot{-} n$."

The motivation for picking that individual formula is as follows: To find a descending sequence, we will iterate Π_2 reflection up to some non-standard number. So we need to make sure that our formula forces a certain number to be non-standard but without implying any false Π_2 sentences. The way we do that is by saying that $\mathsf{I}\Sigma_2$ proves a false Σ_1 sentence. This has (we will show) no false Π_2 consequences. However, (the code of) any proof witnessing a failure of Σ_1 soundness in $\mathsf{I}\Sigma_2$ must be non-standard. We find our descending sequence by iterating Π_2 reflection up to this non-standard number.

Now the formal details start. We need to check that φ_n is Π_2 sound for each n, and that $\mathsf{EA} + \varphi_n \vdash \mathsf{RFN}_{\Pi_2}(\mathsf{EA} + \varphi_{n_1})$.

Claim. $\mathsf{EA} + \varphi_n$ *is* Π_2 *sound for each* n.

The first thing to note is that

$$\mathsf{I}\Sigma_2 \vdash \forall x\ \mathsf{RFN}_{\Pi_2}\big(\mathbf{R}^{x}_{\Pi_2}(\mathsf{EA})\big), \tag{4}$$

where the $\mathsf{I}\Sigma_2$-proof is the induction on x. Recall that $\mathsf{I}\Sigma_2 \equiv I\Pi_2$ and $\forall x\ \mathsf{RFN}_{\Pi_2}\big(\mathbf{R}^{x}_{\Pi_2}(\mathsf{EA})\big)$ is a Π_2-formula, hence $\mathsf{I}\Sigma_2$ could formalize the necessary induction. Also it is known that $\mathsf{I}\Sigma_2 \sqsupseteq \mathsf{I}\Sigma_1 \equiv \mathsf{EA} + \mathsf{RFN}_{\Pi_3}(\mathsf{EA})$ and that

$$\mathsf{EA} + \mathsf{RFN}_{\Pi_3}(\mathsf{EA}) \vdash \psi \rightarrow \mathsf{RFN}_{\Pi_2}(\mathsf{EA} + \psi),$$

for any Π_2-formula ψ. This allows us to verify the base and step of the induction in $\mathsf{I}\Sigma_2$.

The second thing to note is that, since Π_2 reflection is provably equivalent (in EA) to Σ_1 reflection, it follows that:

$$\mathsf{I}\Sigma_2 + \neg\mathsf{RFN}_{\Pi_2}(\mathsf{I}\Sigma_2) \vdash \exists\psi \in \Sigma_1 \exists p\big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \wedge \neg\mathsf{True}_{\Sigma_1}(\psi)\big) \tag{5}$$

Putting these two observations together, we infer that, for each standard $n \in \mathbb{N}$,

$$\mathsf{I}\Sigma_2 + \neg\mathsf{RFN}_{\Pi_2}(\mathsf{I}\Sigma_2) \vdash \exists\psi \in \Sigma_1 \exists p\big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \wedge \neg\mathsf{True}_{\Sigma_1}(\psi) \wedge \mathsf{RFN}^{p \dot{-} n}_{\Pi_2}(\mathsf{EA})\big) \tag{6}$$

which is just to say that for each standard $n \in \mathbb{N}$, $\mathsf{I}\Sigma_2 + \neg\mathsf{RFN}_{\Pi_2}(\mathsf{I}\Sigma_2) \vdash \varphi_n$. Thus, to see that $\mathsf{EA} + \varphi_n$ is Π_2 sound, it suffices to observe that $\mathsf{I}\Sigma_2 + \neg\mathsf{RFN}_{\Pi_2}(\mathsf{I}\Sigma_2)$ is Π_2 sound. The latter claim follows immediately from Remark 3.7.

Before checking that $\mathsf{EA}+\varphi_n \vdash \mathsf{RFN}_{\Pi_2}(\mathsf{EA}+\varphi_{n+1})$, we will establish the following lemma:

Lemma 3.9. *For all standard* $n \in \mathbb{N}$,

$$\mathsf{EA} \vdash \forall p \forall\psi \in \Sigma_1\Big(\big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \wedge \neg\mathsf{True}_{\Sigma_1}(\psi)\big) \rightarrow p > n\Big).$$

Proof. The first thing to note is that (by the Σ_1 soundness of $\mathsf{I}\Sigma_2$ and the Σ_1 completeness of EA) for any $\psi \in \Sigma_1$, if $\mathsf{I}\Sigma_2 \vdash \psi$ then also $\mathsf{EA} \vdash \psi$. Now, for any standard $p \in \mathbb{N}$, EA can check whether p constitutes an $\mathsf{I}\Sigma_2$ proof of a Σ_1 sentence ψ, and if p does constitute such a proof, then EA will prove ψ as well. That is, for each standard $p \in \mathbb{N}$:

$$\mathsf{EA} \vdash \forall \psi \in \Sigma_1 \Big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \to \mathsf{True}_{\Sigma_1}(\psi) \Big)$$

It follows that for each standard $n \in \mathbb{N}$:

$$\mathsf{EA} \vdash \forall p \leqslant n \forall \psi \in \Sigma_1 \Big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \to \mathsf{True}_{\Sigma_1}(\psi) \Big)$$

Whence for each standard $n \in \mathbb{N}$:

$$\mathsf{EA} \vdash \forall p \forall \psi \in \Sigma_1 \Big(\big(\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) \wedge \neg \mathsf{True}_{\Sigma_1}(\psi) \big) \to p > n \Big)$$

This completes the proof of the lemma. ❑

With the lemma on board, we are now ready to verify the following claim:

Claim. *For each* $n \in \mathbb{N}$,

$$\mathsf{EA} + \varphi_n \vdash \mathsf{RFN}_{\Pi_2}(\mathsf{EA} + \varphi_{n+1}).$$

Let's fix an $n \in \mathbb{N}$ and **reason in** $\mathsf{EA} + \varphi_n$**:**

According to φ_n, there is an $\mathsf{I}\Sigma_2$ proof p of a false Σ_1 sentence ψ and $\mathsf{RFN}_{\Pi_2}(\mathbf{R}_{\Pi_2}^{p \dot{-} n}(\mathsf{EA}))$ is Π_2-sound. From Lemma 3.9 we infer that $p > n$. It follows that $p \dot{-} n > 0$, whence $p \dot{-} n = (p \dot{-} (n+1)) + 1$. Hence

$$\mathbf{R}_{\Pi_2}^{p \dot{-} n}(\mathsf{EA}) \equiv \mathbf{R}_{\Pi_2}^{(p \dot{-} (n+1))+1}(\mathsf{EA}) \equiv \mathsf{EA} + \mathsf{RFN}_{\Pi_2}\big(\mathbf{R}_{\Pi_2}^{p \dot{-} (n+1)}(\mathsf{EA})\big). \tag{7}$$

Thus

$$\mathsf{RFN}_{\Pi_2}\Big(\mathsf{EA} + \mathsf{RFN}_{\Pi_2}\big(\mathbf{R}_{\Pi_2}^{p \dot{-} (n+1)}(\mathsf{EA})\big)\Big)$$

Since $\mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi)$ is a true Σ_1 sentence and ψ is a false Σ_1 sentence we infer that

$$\mathsf{RFN}_{\Pi_2}\Big(\mathsf{EA} + \mathsf{Prf}_{\mathsf{I}\Sigma_2}(p, \psi) + \neg \mathsf{True}_{\Sigma_1}(\psi) + \mathsf{RFN}_{\Pi_2}\big(\mathbf{R}_{\Pi_2}^{p \dot{-} (n+1)}(\mathsf{EA})\big)\Big)$$

Which straightforwardly implies $\mathsf{RFN}_{\Pi_2}(\mathsf{EA} + \varphi_{n+1})$. This completes the proof of the theorem. ❑

Question 3.10. *In Theorem 3.4 and Theorem 3.8 we studied how strong reflection principles should be to guarantee that there are no* recursive *descending sequences in the corresponding reflection order. It is natural to ask how this result could be generalized to higher Turing degrees.*

Let n *be a natural number. For which* m *is there a sequence* $\langle T_i \mid i \in \mathbb{N} \rangle$ *recursive in* $0^{(n)}$ *such that all* T_i *are* Π_m *sound extensions of* EA *and* $T_i \vdash \mathsf{RFN}_{\Pi_m}(T_{i+1})$, *for all* i*? The same question for* Σ_m*?*

4. Iterated reflection and conservation

In this section we prove a number of conservation theorems relating iterated reflection and transfinite induction. These results are inspired by the following theorem, which is often known as *Schmerl's formula* [22]. For an ordinal notation system α, ω_n^α is the result of n-applications of ω-exponentiation (see §2.2), starting with α, i.e., $\omega_0^\alpha = \alpha$ and $\omega_{n+1}^\alpha = \omega^{\omega_n^\alpha}$.

Theorem 4.1 (Schmerl)**.** *Let n, m be natural numbers. In EA^+, for any notation system α,*

$$\mathbf{R}^{\alpha}_{\Pi^0_{n+m}}(\mathsf{EA}^+) \equiv_{\Pi^0_n} \mathbf{R}^{\omega_m(\alpha)}_{\Pi^0_n}(\mathsf{EA}^+).$$

Schmerl's formula is a useful tool for calculating the proof-theoretic ordinals of first-order theories. In this section we will develop tools in the mold of Schmerl's formula for calculating the proof-theoretic ordinals of second-order theories. Throughout this section we will rely on the following analogue of Theorem 4.1 that is also due to Schmerl [23].

Theorem 4.2 (Schmerl)**.** *Provably in EA^+, for any ordinal notation α,*

$$\mathbf{R}^{\alpha}_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X)) \equiv_{\mathbf{\Pi}^0_n} \mathbf{R}^{\varepsilon_\alpha}_{\mathbf{\Pi}^0_n}(\mathsf{EA}^+(X)).$$

Note that the versions of Schmerl's formulas that we give above aren't exactly what Schmerl proved, but rather versions of the formulas that are natural given the notation of our paper. And they could be proved by either application of Schmerl's technique or Beklemishev's technique [2]. In fact in a early preprint of this paper [19, §6.2] we provided a proof of Theorem 4.2, however since the technique that we used wasn't new and the result is just a slight variation of [23] we removed it from the paper.

Here is a roadmap for the rest of this section. In §4.1 we prove Theorem 4.9 that states that

$$\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \equiv_{\mathbf{\Pi}^0_\infty} \mathbf{R}^{1+\alpha}_{\mathbf{\Pi}^0_3}(\mathsf{EA}^+(X)).$$

In §4.3 we use this result to prove Theorem 1.6, i.e., that

$$\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0) \equiv_{\Pi^1_1(\Pi^0_3)} \mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0).$$

In §5 we will combine Theorem 1.6 with the results from §3 (especially Theorem 3.2 and Theorem 3.3) to establish connections between iterated reflection and ordinal analysis. In particular, we will use iterated reflection principles to calculate the proof-theoretic ordinals of a wide range of theories.

Before continuing, we alert the reader that many of the proofs in this section use Schmerl's technique of reflexive induction. For a description of this technique, please see §2.4.

4.1. **Iterated reflection and recursive comprehension.** Recall that there are no descending chains in the $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}$ ordering of $\Pi^1_1(\Pi^0_3)$ sound extensions of RCA_0 (this is Theorem 3.3). In this subsection we investigate iterated $\Pi^1_1(\Pi^0_3)$ reflection over the theory RCA_0. The main result of this subsection is that $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)$ is Π^1_1 conservative over $\mathbf{R}^{1+\alpha}_{\mathbf{\Pi}^0_3}(\mathsf{EA}^+(X))$. This result will be used in the next section to calculate proof-theoretic ordinals of subsystems of second-order arithmetic.

Before proving the theorem we prove a few lemmas. These lemmas concern proof-theoretic properties of theories that are closed under an inference rule that we call the $\mathbf{\Delta}^0_1$ *substitution rule.*

Definition 4.3. Suppose φ and $\theta(x)$ are $\mathbf{\Pi}^0_\infty$ formulas that may have other free variables. We denote by $\varphi[\theta(x)]$ the result of substituting the formula $\theta(x)$ in for the free set variable X, i.e. to obtain $\varphi[\theta(x)]$ we first rename all the bounded variables of φ in order to ensure that there are no clashes with free variables of θ and then replace each atomic subformula of φ of the form $t \in X$ with $\theta(t)$.

Definition 4.4. We write $\mathsf{Subst}_{\mathbf{\Delta}^0_1}[\varphi]$ to denote the formula

$$\forall\theta_1(x)\forall\theta_2(x)\Big(\forall y\big(\mathsf{Tr}_{\mathbf{\Pi}^0_1}(\theta_1(y)) \leftrightarrow \mathsf{Tr}_{\mathbf{\Sigma}^0_1}(\theta_2(y))\big) \to \varphi[\mathsf{Tr}_{\mathbf{\Pi}^0_1}(\theta_1(x))]\Big).$$

A theory T is closed under the $\mathbf{\Delta}^0_1$ *substitution rule* if, for any formula $\psi(X)$, whenever $T \vdash \psi(X)$ then $T \vdash \mathsf{Subst}_{\mathbf{\Delta}^0_1}[\psi]$.

Recall that there is a translation $\varphi(X) \longmapsto \forall X\ \varphi(X)$ from the set of $\mathbf{\Pi}^0_\infty$ sentences to the set of sentences of the language of second order arithmetic. Recall also that we are regarding the pseudo-Π^1_1 language as a sublanguage of the language of second order arithmetic by identifying each pseudo Π^1_1 sentence with its translation.

Lemma 4.5. (EA^+) *For each $\mathbf{\Pi}^0_\infty$ sentence $\varphi(X)$ the following are equivalent.*

(1) $\mathsf{RCA}_0 + \forall X\ \varphi(X)$ is $\mathbf{\Pi}^0_\infty$ conservative over $\mathsf{I\Sigma}_1(X) + \varphi(X)$.
(2) $\mathsf{I\Sigma}_1(X) + \varphi(X)$ is closed under the $\mathbf{\Delta}^0_1$ substitution rule.
(3) $\mathsf{I\Sigma}_1(X) + \varphi(X)$ proves $\mathsf{Subst}_{\mathbf{\Delta}^0_1}[\varphi]$.

Proof. We work in EA^+ and consider a $\mathbf{\Pi}^0_\infty$ sentence $\varphi(X)$.

(1) → (2): Suppose that $\mathsf{RCA}_0+\forall X\ \varphi(X)$ is Π^1_1 conservative over $\mathsf{I\Sigma}_1(X)+\varphi(X)$. Suppose that $\mathsf{I\Sigma}_1(X) + \varphi(X) \vdash \psi(X)$. Then $\mathsf{RCA}_0 + \forall X\ \varphi(X) \vdash \psi(X)$. Applying recursive comprehension, we derive $\mathsf{RCA}_0 + \forall X\ \varphi(X) \vdash \mathsf{Subst}_{\mathbf{\Delta}^0_1}[\psi]$. Hence, by Π^1_1 conservativity, $\mathsf{I\Sigma}_1(X) + \varphi(X) \vdash \mathsf{Subst}_{\mathbf{\Delta}^0_1}[\psi]$.

(2) → (3): By application of the $\mathbf{\Delta}^0_1$ substitution rule to φ.

(3) → (1): Suppose that $\mathsf{I\Sigma}_1(X) + \varphi(X)$ proves $\mathsf{Subst}_{\mathbf{\Delta}^0_1}[\varphi]$. We recall the well-known ω-interpretation of RCA_0 into $\mathsf{I\Sigma}_1(X)$ wherein we interpret sets by indices for X-recursive sets; see, e.g., [24, §IX.1]. The image of the sentence $\forall X\ \varphi(X)$ under this interpretation is the sentence $\mathsf{Subst}_{\mathbf{\Delta}^0_1}[\varphi]$. This latter sentence is provable in $\mathsf{I\Sigma}_1(X)+\varphi(X)$ by assumption. Thus, this interpretation actually interprets $\mathsf{RCA}_0 + \forall X\ \varphi(X)$ in $\mathsf{I\Sigma}_1(X)+\varphi(X)$. Therefore, for any sentence $\psi(X)$, if $\mathsf{RCA}_0+\forall X\ \varphi(X)$ proves $\forall X\ \psi(X)$, then $\mathsf{I\Sigma}_1(X) + \varphi(X)$ proves $\mathsf{Subst}_{\mathbf{\Delta}^0_1}[\psi]$, which is the image of $\forall X\ \psi(X)$ under the interpretation. Obviously, $\mathsf{I\Sigma}_1(X) + \varphi(X) \vdash \mathsf{Subst}_{\mathbf{\Delta}^0_1}[\psi] \to \psi(X)$, for any $\mathbf{\Pi}^0_\infty$ formula $\psi(X)$. Therefore, $\mathsf{RCA}_0 + \forall X\ \varphi(X)$ is $\mathbf{\Pi}^0_\infty$ conservative over $\mathsf{I\Sigma}_1(X) + \varphi(X)$. ❑

Question 4.6. *Combining Theorem 3.3 and Lemma 4.5 it is easy to observe that the restriction of the order $\prec_{\mathbf{\Pi}^0_3}$ to $\mathbf{\Pi}^0_3$-sound r.e. extensions of $\mathsf{I\Sigma}_1(X)$ that are closed under the $\mathbf{\Delta}^0_1$-substitution rule is well-founded. Could we drop the condition on closure under the $\mathbf{\Delta}^0_1$-substitution rule? For which n is the restriction of the order $\prec_{\mathbf{\Pi}^0_n}$ to $\mathbf{\Pi}^0_n$-sound r.e. extensions of $\mathsf{I\Sigma}_1(X)$ well-founded?*

Remark 4.7. We recall that $\mathsf{I\Sigma_1} \equiv \mathsf{EA}^+ + \mathsf{RFN}_{\Pi_3}(\mathsf{EA}^+)$. See, e.g., [3]. The same argument could be used to show that $\mathsf{I\Sigma_1}(X) \equiv \mathsf{EA}^+(X) + \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathsf{EA}^+(X))$.

Lemma 4.8. (EA^+) *If EA^+ proves "$T \sqsupseteq \mathsf{I\Sigma}_1(X)$ is closed under the $\mathbf{\Delta}^0_1$ substitution rule," then $\mathsf{EA}^+(X) + \mathsf{RFN}_{\mathbf{\Pi^0_3}}(T)$ is closed under the $\mathbf{\Delta}^0_1$ substitution rule.*

Proof. Suppose that EA^+ proves "$T \sqsupseteq \mathsf{I\Sigma}_1(X)$ is closed under the $\mathbf{\Delta}^0_1$ substitution rule." Let us use the name U for the theory $\mathsf{EA}^+(X) + \mathsf{RFN}_{\mathbf{\Pi}^0_3}(T)$. We want to show that U is closed under the $\mathbf{\Delta}^0_1$ substitution rule. Note that, by Remark 4.7,

U contains $\mathsf{I}\boldsymbol{\Sigma}_{\mathbf{1}}(X)$. That is, $U \equiv \mathsf{I}\boldsymbol{\Sigma}_{\mathbf{1}}(X) + \mathsf{RFN}_{\boldsymbol{\Pi}^0_3}(T)$. Over $\mathsf{EA}(X)$, the reflection schema $\mathsf{RFN}_{\boldsymbol{\Pi}^0_3}(T)$ is equivalent to

$$\forall \varphi \in \boldsymbol{\Pi}^0_3 \Big(\mathsf{Pr}_T \big(\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \big) \to \mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \Big).$$

Thus, by Lemma 4.5, it suffices to show that U proves

$$\mathsf{Subst}_{\boldsymbol{\Delta}^0_1}[\forall \varphi \in \boldsymbol{\Pi}^0_3 \Big(\mathsf{Pr}_T \big(\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \big) \to \mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \Big)].$$

But since the formula $\mathsf{Pr}_T(\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi))$ doesn't contain occurences of X, we could push $\mathsf{Subst}_{\boldsymbol{\Delta}^0_1}$ under the quantifier, i.e., it will be sufficient to show that

$$U \vdash \forall \varphi \in \boldsymbol{\Pi}^0_3 \Big(\mathsf{Pr}_T \big(\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \big) \to \mathsf{Subst}_{\boldsymbol{\Delta}^0_1}[\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi)] \Big).$$

Observe that $\mathsf{Subst}_{\boldsymbol{\Delta}^0_1}[\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi)]$ is equivalent to a $\boldsymbol{\Pi}^0_3$ formula over $\mathsf{EA}(X)$. We reason as follows.

$U \vdash$ "T is closed under the $\boldsymbol{\Delta}^0_1$ substitution rule," by assumption.

$U \vdash \forall \varphi \in \boldsymbol{\Pi}^0_3 \Big(\mathsf{Pr}_T \big(\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \big) \to \mathsf{Pr}_T(\mathsf{Subst}_{\boldsymbol{\Delta}^0_1}[\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi)]) \Big)$

$U \vdash \forall \varphi \in \boldsymbol{\Pi}^0_3 \Big(\mathsf{Pr}_T \big(\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi) \big) \to \mathsf{Subst}_{\boldsymbol{\Delta}^0_1}[\mathsf{Tr}_{\boldsymbol{\Pi}^0_3}(\varphi)] \Big)$ by $\mathsf{RFN}_{\boldsymbol{\Pi}^0_3}(T)$.

This concludes the proof of the lemma. ❑

With these lemmas on board we are ready for the proof of the main theorem of this subsection.

Theorem 4.9. (EA^+) *For any ordinal notation α,*

$$\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \equiv_{\boldsymbol{\Pi}^0_\infty} \mathbf{R}^{1+\alpha}_{\boldsymbol{\Pi}^0_3}(\mathsf{EA}^+(X)).$$

Proof. We prove the claim by reflexive induction. We reason in EA^+ and assume the reflexive induction hypothesis: provably in EA^+, for any $\beta \prec \alpha$,

$$\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \equiv_{\boldsymbol{\Pi}^0_\infty} \mathbf{R}^{1+\beta}_{\boldsymbol{\Pi}^0_3}(\mathsf{EA}^+(X)).$$

Of course, since RCA_0 contains EA^+, this also implies that,

$$\mathsf{RCA}_0 \vdash \forall \beta \prec \alpha \Big(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \equiv_{\boldsymbol{\Pi}^0_\infty} \mathbf{R}^{1+\beta}_{\boldsymbol{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big) \Big)$$

If RCA_0 proves mutual Γ conservation of two theories T and U, then $\mathsf{RFN}_\Gamma(T)$ and $\mathsf{RFN}_\Gamma(U)$ are equivalent over RCA_0. Thus, we immediately infer

$$\mathsf{RCA}_0 \vdash \forall \beta \prec \alpha \Big(\mathsf{RFN}_{\boldsymbol{\Pi}^0_3} \big(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \big) \leftrightarrow \mathsf{RFN}_{\boldsymbol{\Pi}^0_3} \big(\mathbf{R}^{1+\beta}_{\boldsymbol{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big)\big) \Big) \tag{8}$$

We now reason as follows.

$$\begin{aligned} \mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) &\equiv \mathsf{RCA}_0 + \bigcup_{\beta \prec \alpha} \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)} \big(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \big) \text{ by definition.} \\ &\equiv_{\boldsymbol{\Pi}^0_\infty} \mathsf{RCA}_0 + \bigcup_{\beta \prec \alpha} \mathsf{RFN}_{\boldsymbol{\Pi}^0_3} \big(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \big) \\ &\equiv \mathsf{RCA}_0 + \bigcup_{\beta \prec \alpha} \mathsf{RFN}_{\boldsymbol{\Pi}^0_3} \Big(\mathbf{R}^{1+\beta}_{\boldsymbol{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big) \Big) \text{ by (8).} \end{aligned}$$

Since $\mathbf{R}^1_{\mathbf{\Pi}^0_3}(\mathsf{EA}^+(X)) \equiv \mathsf{I\Sigma}_1(X)$, we are able to show that

$$\mathbf{R}^{1+\alpha}_{\mathbf{\Pi^0_3}}\big(\mathsf{EA}^+(X)\big) \equiv \mathsf{I\Sigma}_1(X) + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_3}\Big(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}\big(\mathsf{EA}^+(X)\big)\Big),$$

by the following argument:

$$\begin{aligned}\mathbf{R}^{1+\alpha}_{\mathbf{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big) &\equiv \mathbf{R}^1_{\mathbf{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big) + \mathbf{R}^{1+\alpha}_{\mathbf{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big) \text{ since } 1 \leqslant 1+\alpha.\\ &\equiv \mathsf{I\Sigma}_1(X) + \mathbf{R}^{1+\alpha}_{\mathbf{\Pi^0_3}}\big(\mathsf{EA}^+(X)\big) \text{ since } \mathbf{R}^1_{\mathbf{\Pi}^0_3}\big(\mathsf{EA}^+(X)\big) \equiv \mathsf{I\Sigma}_1(X).\\ &\equiv \mathsf{I\Sigma}_1(X) + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_3}\Big(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}\big(\mathsf{EA}^+(X)\big)\Big) \text{ by definition.}\end{aligned}$$

Hence in order to finish the proof of the lemma it will be enough to show that

$$\mathsf{I\Sigma}_1(X) + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X))) \equiv_{\mathbf{\Pi}^0_\infty} \mathsf{RCA}_0 + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi}^0_3}(\mathsf{EA}^+(X))),$$

which, by Lemma 4.5, can be achieved by proving that

$$\mathsf{I\Sigma}_1(X) + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X)))$$

is closed under the $\mathbf{\Delta}^0_1$ substitution rule. We will prove this closedness in the rest of the proof.

By a usual compactness argument, it will be enough to show that $\mathsf{I\Sigma}_1(X)$ is closed under the $\mathbf{\Delta}^0_1$ substitution rule and that for each $\beta \prec \alpha$ the theories $\mathsf{I\Sigma}_1(X) + \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X)))$ are closed under the $\mathbf{\Delta}^0_1$ substitution rule. Closure of $\mathsf{I\Sigma}_1(X)$ under the $\mathbf{\Delta}^0_1$ substitution rule follows directly from Lemma 4.5.

By Lemma 4.5, we infer that, for each $\beta \prec \alpha$, $\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X))$ is EA^+ provably closed under the $\mathbf{\Delta}^0_1$ substitution rule. Thus, by Lemma 4.8, we infer that for each $\beta \prec \alpha$,

$$\mathsf{EA}^+(X) + \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X)))$$

is closed under the $\mathbf{\Delta}^0_1$ substitution rule. Since $\mathsf{EA}^+(X)+\mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X))) \sqsupseteq \mathsf{I\Sigma}_1(X)$, the theory $\mathsf{I\Sigma}_1(X) + \mathsf{RFN}_{\mathbf{\Pi}^0_3}(\mathbf{R}^{1+\beta}_{\mathbf{\Pi^0_3}}(\mathsf{EA}^+(X)))$ is closed under the $\mathbf{\Delta}^0_1$ substitution rule. This concludes the proof of the lemma. ❑

4.2. **Iterated reflection and arithmetical comprehension.** In this subsection we investigate the relationship between iterated Π^1_1 reflection over ACA_0 and iterated $\Pi^1_1(\Pi^0_3)$ reflection over RCA_0. The main theorem of this subsection is that $\mathbf{R}^\alpha_{\Pi^1_1}(\mathsf{ACA}_0)$ is $\Pi^1_1(\Pi^0_3)$ conservative over $\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)$. The proof of the main theorem of this subsection is similar to the proof of Theorem 4.9. For our first step towards this result, we establish a conservation theorem relating extensions of ACA_0 with extensions of $\mathsf{PA}(X)$.

There is a standard semantic argument that ACA_0 is conservative over PA (see, e.g., [24, Section IX.1]). We will present a version of this argument for extensions of ACA_0 by Π^1_1 sentences. Moreover we ensure that this conservation result is provable in ACA_0. Before presenting the argument, we will say a bit about how we will formalize model theory within ACA_0 for the purposes of our argument.

We will reason in ACA_0 and use the formalization of model theory from [24, Section II.8]. Recall that according to formalization from [24, Section II.8] a model

$\mathfrak{M}$ essentially is a set that encodes the domain of $\mathfrak{M}$ (which is by necessity a subset of $\mathbb{N}$) and the full satisfaction relation for $\mathfrak{M}$ (the latter essentially is the elementary diagram of the model $\mathfrak{M}$). Note that if one would require $\mathfrak{M}$ contain information only about the satisfaction of atomic formulas, rather then all formulas, the resulting notion of a model would be weaker. This is due to the fact that in ACA_0, unlike in stronger theories, it is not always possible to recover the elementary diagram of a model from its atomic diagram.

Due to this limitation, in ACA_0 it is sometimes (including in our proof) useful to employ weak models [24, Definition II.8.9]. A *weak model* $\mathfrak{M}$ of a theory T is a set that encodes the domain of $\mathfrak{M}$ and a partial satisfaction relation for $\mathfrak{M}$ that is defined only on Boolean combinations of subformulas of formulas used in axioms of T such that all the axioms of T are according to this satisfaction relation. The key fact that we use is that ACA_0 proves that any theory that has a weak model is consistent [24, Theorem II.8.10].

Lemma 4.10. (ACA_0) *Let* $\varphi(X), \psi(X)$ *be* $\mathbf{\Pi}^0_\infty$. *If* $\mathsf{ACA}_0 + \forall X\ \varphi(X) \vdash \forall X\ \psi(X)$ *then* $\mathsf{PA}(X) + \{\varphi[\theta] : \theta(x) \text{ is } \mathbf{\Pi}^0_\infty\} \vdash \psi(X)$, *where* θ *could contain additional variables.*

Proof. We reason in ACA_0. We denote by U the theory $\mathsf{PA}(X) + \{\varphi[\theta] : \theta \text{ is } \mathbf{\Pi}^0_\infty\}$. Let us consider any $\psi(X)$ such that $U \nvdash \psi(X)$. To prove the lemma we need to show that $\mathsf{ACA}_0 + \forall X\ \varphi(X) \nvdash \forall X\ \psi(X)$.

There is a model $\mathfrak{M}$ of $U + \neg\psi(X)$. Note that here X is just a unary predicate. We enrich $\mathfrak{M}$ by adding, as the family $\mathcal{S}$ of second-order objects, all the sets defined in $\mathfrak{M}$ by $\mathbf{\Pi}^0_\infty$ formulas that may contain additional parameters from the model.

Let us first show how we could finish the proof without ensuring that our argument could be formalized in ACA_0 and only then indicate how to carry out the formalization. Indeed, it is easy to see that the second-order structure $(\mathfrak{M}, \mathcal{S})$ satisfies $\mathsf{ACA}_0 + \forall X\ \varphi(X)$: the presence of the full induction schema in U guarantees that $(\mathfrak{M}, \mathcal{S})$ satisfies set induction, our definition of $\mathcal{S}$ guarantees that arithmetical comprehension holds in $(\mathfrak{M}, \mathcal{S})$, and the fact that we had axioms $\{\varphi[\theta] : \theta \text{ is } \mathbf{\Pi}^0_\infty\}$ in U guarantees that $\forall X\ \varphi(X)$ holds in $(\mathfrak{M}, \mathcal{S})$. And since $\psi(X)$ failed in $\mathfrak{M}$, the sentence $\forall X\ \psi(X)$ fails in $(\mathfrak{M}, \mathcal{S})$. Therefore, $\mathsf{ACA}_0 + \forall X\ \varphi(X) \nvdash \forall X\ \psi(X)$.

Now let us show how to formalize the latter argument in ACA_0. We want to show that we could extend $(\mathfrak{M}, \mathcal{S})$ to a weak model of $\mathsf{ACA}_0 + \forall X\ \varphi(X)$. From the satisfaction relation for $\mathfrak{M}$ we can trivially construct the partial satisfaction relation for $(\mathfrak{M}, \mathcal{S})$ that covers all Π^0_∞ formulas with parameters from $(\mathfrak{M}, \mathcal{S})$. And since we are working in ACA_0, using arithmetical comprehension for every (externally) fixed n we could expand the latter partial satisfaction relation to all the formulas constructed from Π^0_∞ formulas by arbitrary use of propositional connectives and with introduction of at most n quantifier alternations. For $n = 2$ this expanded partial satisfaction relation covers all the axioms of $\mathsf{ACA}_0 + \forall X\ \varphi(X) + \neg\forall X\ \psi(X)$. Now after we constructed this satisfaction relation we could proceed as in the paragraph above and show that in this partial satisfaction realtion all the axioms of $\mathsf{ACA}_0 + \forall X\ \ \varphi(X) + \neg\forall X\ \ \psi(X)$ are true. Hence we have a weak model of $\mathsf{ACA}_0 + \forall X\ \varphi(X)$. Therefore, $\mathsf{ACA}_0 + \forall X\ \varphi(X) \nvdash \forall X\ \psi(X)$. ❑

Remark 4.11. Although we don't provide a proof here, we note that with some additional care it is possible to establish Lemma 4.10 in EA^+ by appealing to the Π_2-conservativity of $\mathsf{WKL}^\star_0$ + "super-exponentiation is total" over EA^+, see [25] for the Π_2-conservativity of $\mathsf{WKL}_0^\star$ over EA. But it isn't possible to prove this result in EA since ACA_0 enjoys non-elementary speed-up over PA.

Definition 4.12. We say that a pseudo Π^1_1 theory $T(X)$ is *closed under substitution* if whenever $T \vdash \varphi(X)$ then also $T \vdash \varphi[\theta(x)]$ for any $\mathbf{\Pi}^0_\infty$ formula θ.

Lemma 4.13. *If a theory T proves every substitution variant of its own axioms, then T is closed under substitution.*

Proof. Suppose that T proves every substitution variant of its own axioms. Let θ be a $\mathbf{\Pi}^0_\infty$ formula and let $\varphi(X)$ be a theorem of T. Since $\varphi(X)$ is a theorem of T, there is some finite conjunction $A_T(X)$ of axioms of T such that the sentence

$$A_T(X) \to \varphi(X)$$

is a theorem of pure logic. Since pure logic is closed under substitution, the sentence

$$A_T[\theta(x)] \to \varphi[\theta(x)]$$

is also a theorem of pure logic. Since T proves every substitution variant of its own axioms, T proves $A_T[\theta(x)]$, whence T proves $\varphi[\theta(x)]$. ❑

Lemma 4.14. $\mathsf{PA}(X) + \mathbf{R}^\alpha_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))$ *is closed under substitution.*

Proof. We prove the claim by reflexive induction. We reason within EA^+ and assume the reflexive induction hypothesis: provably in EA^+, for all $\beta \prec \alpha$, $\mathsf{PA}(X) + \mathbf{R}^\beta_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))$ is closed under substitution. First we note that

$$\mathbf{R}^\alpha_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big) \equiv \mathsf{PA}(X) + \bigcup_{\beta \prec \alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\Big(\mathbf{R}^\beta_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)\Big).$$

For $\beta \prec \alpha$ let us denote by S_β the theory

$$\mathsf{PA}(X) + \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\Big(\mathbf{R}^\beta_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)\Big).$$

To prove that $\mathbf{R}^\alpha_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)$ is closed under substitution it suffices to prove that, for every $\beta \prec \alpha$, S_β is closed under substitution.

By Lemma 4.13, to prove that S_β is closed under substitution, it suffices to show that S_β proves every substitution-variant of its own axioms. Let us use the name U_β to denote the theory $\mathbf{R}^\beta_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)$. An axiom of the theory S_β is either an axiom of $\mathsf{PA}(X)$ or is a sentence of the form $\forall \vec{y}\big(Pr_{U_\beta}\big(\varphi(X,\vec{y})\big) \to \varphi(X,\vec{y})\big)$. Already the theory $\mathsf{PA}(X)$ proves every substitutional instance of its own axioms. By the reflexive induction hypothesis, U_β is provably closed under substitution. So S_β proves $\forall \vec{y}\big(Pr_{U_\beta}\big(\varphi(X,\vec{y})\big) \to \varphi(\theta,\vec{y})\big)$ for any formula θ. This is to say that S_β proves every substitution instance of its axioms. ❑

Remark 4.15. It follows from the lemma that the theories $\mathsf{PA}(X) + \{\mathbf{R}^\alpha_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))[\theta] : \theta \in \mathbf{\Pi}^0_\infty\}$ and $\mathsf{PA}(X) + \mathbf{R}^\alpha_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))$ are equivalent. We will make use of this observation in the proof of Lemma 4.16.

Most of the work towards proving the main theorem of this section is contained in the proof of the following key lemma.

Lemma 4.16. $\mathbf{R}^\alpha_{\Pi^1_1}(\mathsf{ACA}_0)$ *is* $\mathbf{\Pi}^0_\infty$ *conservative over* $\mathbf{R}^\alpha_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))$.

Proof. We prove the claim by reflexive induction. We reason within ACA_0 and assume the reflexive induction hypothesis: provably in ACA_0, for all $\beta \prec \alpha$, $\mathbf{R}^{\beta}_{\Pi^1_1}(\mathsf{ACA}_0)$ is $\mathbf{\Pi}^0_\infty$ conservative over $\mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))$. This means that, provably in ACA_0, for any $\beta \prec \alpha$, $\mathbf{\Pi}^0_\infty$ reflection over $\mathbf{R}^{\beta}_{\Pi^1_1}(\mathsf{ACA}_0)$ is equivalent to $\mathbf{\Pi}^0_\infty$ reflection over $\mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))$. That is:

$$\mathsf{ACA}_0 \vdash \forall \beta \prec \alpha \Big(\mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\big(\mathbf{R}^{\beta}_{\Pi^1_1}(\mathsf{ACA}_0)\big) \leftrightarrow \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\big(\mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X))\big)\Big) \tag{9}$$

We reason as follows.

$$\begin{aligned}
\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0) &\equiv \mathsf{ACA}_0 + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\Pi^1_1}\big(\mathbf{R}^{\beta}_{\Pi^1_1}(\mathsf{ACA}_0)\big) \text{ by definition.}\\
&\equiv_{\mathbf{\Pi}^0_\infty} \mathsf{ACA}_0 + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\big(\mathbf{R}^{\beta}_{\Pi^1_1}(\mathsf{ACA}_0)\big)\\
&\equiv \mathsf{ACA}_0 + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\Big(\mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)\Big) \text{ by (9).}\\
&\equiv \mathsf{ACA}_0 + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\Big(\mathsf{PA}(X) + \mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)\Big) \text{ by definition.}\\
&\equiv_{\mathbf{\Pi}^0_\infty} \mathsf{PA}(X) + \bigcup_{\beta\prec\alpha}\bigcup_{\theta\in\mathbf{\Pi}^0_\infty} \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\Big(\mathsf{PA}(X) + \mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)\Big)[\theta] \text{ by Lemma 4.10.}\\
&\equiv_{\mathbf{\Pi}^0_\infty} \mathsf{PA}(X) + \bigcup_{\beta\prec\alpha} \mathsf{RFN}_{\mathbf{\Pi}^0_\infty}\Big(\mathsf{PA}(X) + \mathbf{R}^{\beta}_{\mathbf{\Pi}^0_\infty}\big(\mathsf{PA}(X)\big)\Big) \text{ by Remark 4.15.}\\
&\equiv_{\mathbf{\Pi}^0_\infty} \mathbf{R}^{\alpha}_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X)) \text{ by definition.}
\end{aligned}$$

This concludes the proof. ❑

The proof of the the main theorem of this section is now straightforward, given Theorem 4.9 and Lemma 4.16.

Theorem 4.17. $\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)$ *is* $\Pi^1_1(\Pi^0_3)$ *conservative over* $\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)$.

Proof. We reason as follows.

$$\begin{aligned}
\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0) &\equiv_{\mathbf{\Pi}^0_\infty} \mathbf{R}^{\alpha}_{\mathbf{\Pi}^0_\infty}(\mathsf{PA}(X)) \text{ by Lemma 4.16.}\\
&\equiv_{\mathbf{\Pi}^0_3} \mathbf{R}^{\varepsilon_\alpha}_{\mathbf{\Pi}^0_3}(\mathsf{EA}^+(X)) \text{ by Theorem 4.2.}\\
&\equiv_{\mathbf{\Pi}^0_3} \mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0) \text{ by Theorem 4.9.}
\end{aligned}$$

Note for each $\Pi^1_1(\Pi^0_3)$ sentence φ we could find a $\mathbf{\Pi}^0_3$ sentence φ' such that RCA_0 proves the equivalence of φ and (the translation into the second order language of) φ'. Thus moreover we have

$$\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0) \equiv_{\Pi^1_1(\Pi^0_3)} \mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0).$$

This completes the proof of the theorem. ❑

5. Reflection ranks and proof-theoretic ordinals

In this section we introduce the notion of *reflection rank*. We then use the results from the previous section to establish connections between reflection ranks and proof-theoretic ordinals.

5.1. **Reflection ranks.** Recall that the reflection order $\prec_{\Pi^1_1}$ on r.e. extensions of ACA_0 is:

$$T_1 \prec_{\Pi^1_1} T_2 \overset{\text{def}}{\iff} T_2 \vdash \mathsf{RFN}_{\Pi^1_1}(T_1).$$

For a theory $T \sqsupseteq \mathsf{ACA}_0$ we define the *reflection rank* $|T|_{\mathsf{ACA}_0} \in \mathbf{On} \cup \{\infty\}$ as the rank of T in the order $\prec_{\Pi^1_1}$.

Remark 5.1. We recall that as usual the rank function $\rho\colon A \to \mathbf{On} \cup \{\infty\}$ for a binary relation $(A, \lhd)$ is the only function such that $\rho(a) = \sup\{\rho(b) + 1 \mid b \lhd a\}$. Here the linear order $<$ on ordinals is extended to the class $\mathbf{On} \cup \{\infty\}$ by puting $\boldsymbol{\alpha} < \infty$, for all $\boldsymbol{\alpha} \in \mathbf{On}$. The operation $\boldsymbol{\alpha} \mapsto \boldsymbol{\alpha}+1$ is extended to the class $\mathbf{On} \cup \{\infty\}$ and puting $\infty + 1 = \infty$. Note that $\rho(a) \in \mathbf{On}$ iff the cone $\{b \mid b \lhd a\}$ is well-founded with respect to $\lhd$.

Recall that Theorem 3.2 states that $|T|_{\mathsf{ACA}_0} \in \mathbf{On}$, for Π^1_1-sound T.

We will also consider the more general notion of reflection rank with respect to some other base theories. For second-order theories $U \sqsupseteq \mathsf{RCA}_0$ we consider the reflection order $\prec_{\Pi^1_1(\Pi^0_3)}$:

$$U_1 \prec_{\Pi^1_1(\Pi^0_3)} U_2 \overset{\text{def}}{\iff} U_2 \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(U_1).$$

Let us consider some base theory $T_0 \sqsupseteq \mathsf{RCA}_0$. We define the set $\mathcal{E}\text{-}T_0$ of all theories U such that EA proves that $U \sqsupseteq T_0$. For $U \in \mathcal{E}\text{-}T_0$ we denote by $|U|_{T_0}$ the rank of U in the order $(\mathcal{E}\text{-}T_0, \prec_{\Pi^1_1(\Pi^0_3)})$. Note that $\Pi^1_1(\Pi^0_3)$-sound extensions of T_0 have a well-founded rank in this ordering by Theorem 3.3.

Remark 5.2. For a theory T_0 given by a finite list of axioms the set $\mathcal{E}\text{-}T_0$ coincides with the set of all U such that $U \sqsupseteq T_0$. Indeed, for any T_0 given by a finite list of axioms we have a Σ_1 formula in EA that expresses $U \sqsupseteq T_0$ with U as a parameter (the Σ_1 formula states that there is a U-proof of the conjunction of all the axioms of T_0).

Remark 5.3. The definition of the rank $|T|_{\mathsf{ACA}_0}$ given in the beginning of the section coincides with the more general definition of rank, since in ACA_0 each Π^1_1 formula is equivalent to a $\Pi^1_1(\Pi^0_3)$-formula and hence for any $T \sqsupseteq \mathsf{ACA}_0$,

$$\mathsf{ACA}_0 \vdash \mathsf{RFN}_{\Pi^1_1}(T) \leftrightarrow \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(T).$$

Straightforwardly from Theorem 3.3 we get the following.

Corollary 5.4. *If $U \sqsupseteq \mathsf{RCA}_0$ is $\Pi^1_1(\Pi^0_3)$-sound, then the rank $|U|_{\mathsf{RCA}_0} \in \mathbf{On}$. Hence for each $T_0 \sqsupseteq \mathsf{RCA}_0$ and $\Pi^1_1(\Pi^0_3)$-sound theory $U \in \mathcal{E}\text{-}T_0$ we have $|U|_{T_0} \in \mathbf{On}$.*

Remark 5.5. The converse of Corollary 5.4 is not true, there are $\Pi^1_1(\Pi^0_3)$ unsound theories whose rank is an ordinal. In particular, for each consistent theory $T_0 \sqsupseteq \mathsf{RCA}_0$, we have $|T_0 + \neg\mathsf{Con}(T_0)|_{T_0} = 0$. Indeed, assume $T_0 + \neg\mathsf{Con}(T_0) \vdash \mathsf{RFN}_{\Pi^1_1}(U)$, for some $U \in \mathcal{E}\text{-}T_0$. Then

$$\begin{aligned} T_0 + \neg\mathsf{Con}(T_0) &\vdash \mathsf{RFN}_{\Pi^1_1}(T_0) \\ &\vdash \mathsf{Con}(T_0) \\ &\vdash \bot. \end{aligned}$$

But by Gödel's Second Incompleteness Theorem $T_0 + \neg\mathsf{Con}(T_0)$ is consistent. This is to say that, though $T_0 + \neg\mathsf{Con}(T_0)$ is not Π^1_1 sound, $|T_0 + \neg\mathsf{Con}(T_0)|_{T_0} \in \mathbf{On}$.

Note that later we will introduce a notion of robust reflection rank that enjoys much better behavior and, in particular, satisfies the converse of Corollary 5.4.

Recall that for an ordinal notation α we denote by $|\alpha| \in \mathbf{On} \cup \{\infty\}$ the rank of the ordinal notation α in the order $\prec$.

The main proposition proved in this subsection is the following:

Proposition 5.6. *For each $\Pi^1_2(\Pi^0_2)$-sound theory T_0 and ordinal notation α:*

$$|\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)|_{T_0} = |\alpha|.$$

In order to prove the proposition we first prove the following lemma.

Lemma 5.7. *If $|U|_{T_0} > |\alpha|$ then there is a true $\Sigma^1_1(\Pi^0_2)$ sentence φ such that*

$$U + \varphi \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)). \tag{10}$$

Proof. We prove the lemma by transfinite induction on $|\alpha|$. Since $|U|_{T_0} > |\alpha|$, there is a $V \in \mathcal{E}\text{-}T_0$ such that $U \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(V)$ and $|V|_{T_0} \geqslant |\alpha|$. By the induction hypothesis there are true $\Sigma^1_1(\Pi^0_2)$ sentences φ_β, for all $\beta \prec \alpha$, such that

$$V + \varphi_\beta \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0)).$$

We now formalize the latter fact by a single $\Sigma^1_1(\Pi^0_2)$ sentence φ, which states that there is a sequence of Π^0_2 formulas $\langle \psi_\beta(Y) \mid \beta \prec \alpha\rangle$ without free variables other that Y and sequence of sets $\langle S_\beta \mid \beta \prec \alpha\rangle$ such that

- for all β, the formula $\psi_\beta(Y)$ holds on $Y = S_\beta$;
- for all β, we have $V + \exists Y\ \psi_\beta(Y) \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0))$.

It is easy to see that indeed we could form a $\Sigma^1_1(\Pi^0_2)$ sentence φ constituting the desired formalization.

Now let us show that φ is true. Without loss of generality, we may assume that each φ_β is of the form $\exists Y \theta_\beta(Y)$, where all $\theta_\beta(Y)$ are Π^0_2-formulas. We put each ψ_β to be θ_β and for each $\beta \prec \alpha$ we choose S_β so that $\theta_\beta(Y)$ holds on $Y = S_\beta$. Thus we see that φ is true.

We establish (10) by reasoning in $U+\varphi$ and showing that the theory $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)$ is $\Pi^1_1(\Pi^0_3)$-sound. It is enough for us to establish the $\Pi^1_1(\Pi^0_3)$-soundness of each finite subtheory of $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)$, i.e., each theory

$$T_0 + \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0)),$$

for $\beta \prec \alpha$. We know (from U) that V is $\Pi^1_1(\Pi^0_3)$-sound. And also (from φ) we have a Π^0_2-formula $\psi_\beta(Y)$ such that

$$V + \exists Y\ \psi_\beta(Y) \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0))$$

and a set S_β such that $\psi_\beta(S_\beta)$ holds. From the $\Pi^1_1(\Pi^0_3)$-soundness of V we infer the $\Pi^1_1(\Pi^0_3)$-soundness of $V + \exists Y\ \psi_\beta(Y)$. Therefore $T_0 + \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0))$ is $\Pi^1_1(\Pi^0_3)$-sound. ❑

We are nearly in a position to prove Proposition 5.6. Before doing so, we pause to state two lemmas, the truth of which may easily be verified.

Lemma 5.8 (RCA_0)**.**
If T is $\Pi^1_2(\Pi^0_2)$ sound and α is a well-ordering, then $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T)$ is $\Pi^1_1(\Pi^0_3)$ sound.

Lemma 5.9 (RCA_0)**.** *If T is $\Pi^1_1(\Pi^0_3)$ sound and φ is a true $\Sigma^1_1(\Pi^0_2)$ formula, then $T+\varphi$ is $\Pi^1_1(\Pi^0_3)$ sound.*

Proof. First let us notice that $|\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)|_{T_0} \geqslant |\alpha|$. Indeed this inequality holds since there is a homomorphism $\beta \mapsto \mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0)$ of the low $\prec$-cone of α (the order $(\{\beta \mid \beta \preceq \alpha\}, \prec)$) to the low $\prec_{\Pi^1_1(\Pi^0_3)}$-cone of $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)$ in $\mathcal{E}$-T_0.

Now assume for a contradiction that $|\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)|_{T_0} > |\alpha|$. In this case by Lemma 5.7 we have

$$\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0) + \varphi \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)),$$

for some true $\Sigma^1_1(\Pi^0_2)$ sentence φ. We derive

$$\begin{aligned}\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0) + \varphi &\vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)+\varphi)\\ &\vdash \mathsf{Con}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)+\varphi).\end{aligned}$$

So $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)+\varphi$ is inconsistent by Gödel's Second Incompleteness Theorem. Yet by Lemma 5.8, since T_0 is $\Pi^1_2(\Pi^0_2)$ sound, $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)$ is $\Pi^1_1(\Pi^0_3)$ sound. Thus, by Lemma 5.9, $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)+\varphi$ is consistent. This is a contradiction. ❑

5.2. **Proof-theoretic ordinals.** For a theory $T \sqsupseteq \mathsf{RCA}_0$ we write $|T|_{\mathsf{WO}}$ to denote the *proof-theoretic ordinal* of T, which we define as the supremum of the ranks $|\alpha|$ of ordinal notations α such that $T \vdash \mathsf{WO}(\alpha)$. The formula $\mathsf{WO}(\alpha)$ is

$$\forall X((\exists \beta \prec \alpha)\ \beta \in X \to (\exists \beta \prec \alpha)(\beta \in X \wedge (\forall \gamma \prec \beta)\gamma \notin X)).$$

Remark 5.10. One may also define $|T|_{\mathsf{WO}}$ for second-order theories in terms of primitive recursive well-orders (alternatively recursive well-orders), i.e., $|T|_{\mathsf{WO}}$ then would be defined as the supremum of order types of primitive recursive (T-provably recursive) binary relations $\rhd$ such that $T \vdash \mathsf{WO}(\rhd)$. If T proves the well-orderedness of an ill-founded relation then this supremum by definition is ∞. We note that our definition coincides with the definitions above for $T \sqsupseteq \mathsf{RCA}_0$. The connection between presentations of ordinals of various degrees of "niceness" is extensively discussed in M. Rathjen's survey [21], and the equivalence under consideration could be proved by a slight extension of the proof of [21, Proposition 2.19(i)].[3]

Theorem 5.11. $|\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)|_{\mathsf{WO}} = |\varepsilon_\alpha|$.

In order to prove the theorem we first establish the following lemma:

Lemma 5.12. *For each α*

1. *the theory ACA_0 proves $\mathsf{WO}(\alpha) \to \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0))$;*
2. *the theory ACA_0^+ proves $\mathsf{WO}(\alpha) \to \mathsf{RFN}_{\Pi^1_1}(\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0))$.*

We will derive Lemma 5.12 from the more general Lemma 5.13.

We will follow Simpson's formalization of countable coded models of the language of second-order arithmetic [24, Definition VII.2.1]. Under this definition a countable coded ω-model $\mathfrak{M}$ is a code for a countable family $W_0, W_1, \ldots$ of subsetes of $\mathbb{N}$, where $\{W_0, W_1, \ldots\}$ is the $\mathfrak{M}$-domain for sets of naturals. We note that the property "$\mathfrak{M}$

[3]The proof of [21, Proposition 2.19(i)] implicitly uses Σ_1-collection inside the theory T, although the claim is stated for all T containing PRA. But this issue doesn't affect the theories that we are interested in since $\mathsf{RCA}_0 \vdash \mathsf{B}\Sigma_1$

is a countable coded ω-model" is arithmetical. The expression $X \in \mathfrak{M}$ denotes the natural Σ^0_2 formula that expresses the fact that the set X is coded in a model $\mathfrak{M}$ (i.e. it is one of $X = W_i$, for some i). For each fixed second-order formula $\varphi(X_1, \ldots, X_n, x_1, \ldots, x_n)$ the expression $\mathfrak{M} \models \varphi(X_1, \ldots, X_n, x_1, \ldots, x_n)$ denotes the natural second-order formula that expresses that $\mathfrak{M}$ is a countable coded ω-model, sets $X_1, \ldots, X_n$ are coded in $\mathfrak{M}$, and $\varphi(X_1, \ldots, X_n, x_1, \ldots, x_n)$ is true in $\mathfrak{M}$. We express the fact that that $\varphi(X_1, \ldots, X_n, x_1, \ldots, x_n)$ is true in $\mathfrak{M}$ by relativizing second-order quatifiers $\forall X$ and $\exists X$ to $\forall X \in \mathfrak{M}$ and $\exists X \in \mathfrak{M}$. Note that the latter quantifiers are in fact just first-order quantifiers. Hence $\mathfrak{M} \models \varphi(\vec{X}, \vec{x})$ is equivalent to a Π^0_m-formula, where m depends only on the depth of quantifier alternations in φ. For a fixed theory T given by a finite list of axioms, by $\mathfrak{M} \models T$ we mean the formula $\mathfrak{M} \models \varphi$, where φ is the conjunction of all the axioms of T.

For each theory $T_0 \sqsupseteq \mathsf{RCA}_0$ given by a finite list of axioms we denote by T_0^+ the theory $T_0 + \mathsf{ACA}_0 +$ "every set is contained in an ω-model of T_0." We use this notation by analogy with ACA_0^+. We note that for $T_0 = \mathsf{ACA}_0$ the theory T_0^+ is just ACA_0^+ and for $T_0 = \mathsf{RCA}_0$ the theory T_0^+ is just ACA_0.

Lemma 5.13. *For each T_0 given by a finite list of axioms*

$$T_0^+ \vdash \forall \alpha \left(\mathsf{WO}(\alpha) \to \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}\big(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0)\big)\right).$$

Proof. We reason in T_0^+. We assume $\mathsf{WO}(\alpha)$ and claim $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0))$.

Note that it suffices to show that $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0))$ is true in all the ω-models of T_0. Indeed, since $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0))$ is a $\Pi^1_1(\Pi^0_3)$ sentence, if it fails, this fact is witnessed by some set X and hence $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0))$ fails in all the ω-models of T_0 containing X.

Now let us consider an ω-model $\mathfrak{M}$ of T_0 and show $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(T_0))$. We note that, for some fixed k, all the facts of the form $\mathfrak{M} \models \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0))$ are Π^0_k. In order to finish the proof it suffices to show $\mathfrak{M} \models \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0))$, for all $\beta \preceq \alpha$ by transfinite induction on $\beta \preceq \alpha$. By the induction hypothesis we know that $\mathfrak{M}$ is a model of $\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0)$. Since $\mathfrak{M}$ is an ω-model we need to show that for all the (standard) proofs p of a $\Pi^1_1(\Pi^0_3)$-sentence φ in $\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(T_0)$ the sentence φ is true in $\mathfrak{M}$. We consider some proof p of this form and apply the cut-elimination theorem for predicate calculus to make sure that all the intermediate formulas in the proof are of the complexity $\Pi^1_n(\Pi^0_m)$ for some externally fixed n and m (depending only on the complexity of the axioms of T_0). We proceed by showing by induction on formulas in the proof that all of them are true in the model $\mathfrak{M}$; we can do this since the satisfaction relation for $\Pi^1_n(\Pi^0_m)$-formulas in $\mathfrak{M}$ is arithmetical. ❑

Lemma 5.14.

$$\mathsf{RCA}_0 \vdash \forall \alpha \left(\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}\big(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)\big) \to \mathsf{WO}(\alpha)\right)$$

Proof. We prove the lemma by reflexive induction on α in RCA_0. We reason in RCA_0 and assume the reflexive induction hypothesis

$$\forall \beta \prec \alpha \ \mathsf{Pr}_{\mathsf{RCA}_0}\left(\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}\big(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)\big) \to \mathsf{WO}(\beta)\right).$$

We need to show that:

$$\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}\big(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)\big) \to \mathsf{WO}(\alpha) \tag{11}$$

So assume the antecedent of (11). From the reflexive induction hypothesis we see that for each individual $\beta \prec \alpha$ the theory $\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)$ proves $\mathsf{WO}(\beta)$. Since $\mathsf{WO}(\beta)$ is a $\Pi^1_1(\Pi^0_3)$-formula, we infer from the antecedent of (11) that $\forall \beta \prec \alpha\, \mathsf{WO}(\beta)$. Thus $\mathsf{WO}(\alpha)$. ❑

Now we are ready to prove Theorem 5.11

Proof. From Theorem 4.17 we know that

$$\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0) \equiv_{\Pi^1_1(\Pi^0_3)} \mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0).$$

From Lemma 5.14 we see that $\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)$ proves $\mathsf{WO}(\beta)$ for each $\beta \prec \varepsilon_\alpha$ and thus $|\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)|_{\mathsf{WO}} \geqslant |\varepsilon_\alpha|$.

In order to prove $|\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)| \leqslant |\varepsilon_\alpha|$ let us assume that for some β the theory $\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)$ proves $\mathsf{WO}(\beta)$ and then show that $|\beta| < |\varepsilon_\alpha|$. Indeed, by Lemma 5.12 the theory $\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)$ proves $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\beta}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0))$. Hence

$$|\beta| < |\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)|_{\mathsf{RCA}_0} = |\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)|_{\mathsf{RCA}_0}.$$

And Proposition 5.6 gives us

$$|\varepsilon_\alpha| = |\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)|_{\mathsf{RCA}_0} > |\beta|.$$

This completes the proof. ❑

5.3. **Extensions of ACA_0^+.** It is usually attributed to Kreisel that for extensions $T \sqsupseteq \mathsf{ACA}_0$ the proof-theoretic ordinal $|T|_{\mathsf{WO}} = |T+\varphi|_{\mathsf{WO}}$, for any true Σ^1_1-sentence φ (see [20, Theorem 6.7.4,6.7.5]). We note that our notion of reflection rank $|T|_{\mathsf{ACA}_0}$ does not enjoy the same property.

Remark 5.15. Let us consider an ordinal notation system α for some large recursive ordinal, for example the Bachmann-Howard ordinal. Now we modify α to define pathological ordinal notation α'. The order $\prec_{\alpha'}$ is the restriction of $\prec_\alpha$ to numbers m such that $\forall x \leqslant m\ \neg \mathsf{Prf}_{\mathsf{ACA}_0}(x, 0=1)$. And α' corresponds to the same element of the domain of $\prec_\alpha$ as α (note that since ACA_0 is consistent this element is in the domain of $\prec_{\alpha'}$ as well). We see externally that α' is isomorphic to α, since ACA_0 is consistent. Let us denote by Iso the true Σ^1_1-sentence that expresses the fact that α and α' are isomorphic. Clearly,

$$\mathsf{ACA}_0 + \mathsf{WO}(\alpha') + \mathsf{Iso} \sqsupseteq \mathsf{ACA}_0 + \mathsf{WO}(\alpha),$$

$$|\mathsf{ACA}_0 + \mathsf{WO}(\alpha') + \mathsf{Iso}|_{\mathsf{ACA}_0} \geqslant |\mathsf{ACA}_0 + \mathsf{WO}(\alpha)|_{\mathsf{ACA}_0}$$

and under our choice of α the rank $|\mathsf{ACA}_0 + \mathsf{WO}(\alpha)|_{\mathsf{ACA}_0}$ will be equal to the Bachmann-Howard ordinal. At the same time, the theory $\mathsf{ACA}_0 + \neg\mathsf{Con}(\mathsf{ACA}_0)$ proves that α' is isomorphic to some finite order and hence

$$\mathsf{ACA}_0 + \neg\mathsf{Con}(\mathsf{ACA}_0) \vdash \mathsf{WO}(\alpha').$$

Hence

$$|\mathsf{ACA}_0 + \mathsf{WO}(\alpha')|_{\mathsf{ACA}_0} \leqslant |\mathsf{ACA}_0 + \neg\mathsf{Con}(\mathsf{ACA}_0)|_{\mathsf{ACA}_0} = 0,$$

the latter equality follows from Remark 5.5. And thus

$$|\mathsf{ACA}_0 + \mathsf{WO}(\alpha')|_{\mathsf{WO}} < |\mathsf{ACA}_0 + \mathsf{WO}(\alpha') + \mathsf{Iso}|_{\mathsf{WO}}.$$

Accordingly, Iso is a true Σ^1_1 sentence that alters the reflection rank of the theory $\mathsf{ACA}_0 + \mathsf{WO}(\alpha')$.

We address this problem with two different results. First in Theorem 5.16 we show that for any extension $T \sqsupseteq \mathsf{ACA}_0^+$, $|T|_{\mathsf{ACA}_0} = |T|_{\mathsf{WO}}$. Second we introduce the notion of robust reflection rank $|\cdot|^\star_{\mathsf{ACA}_0}$ that enjoys a number of nice properties and at the same time coincides with reflection rank $|\cdot|_{\mathsf{ACA}_0}$, for many natural theories T (in particular, for any any T such that $T \equiv_{\Pi^1_1} \mathbf{R}^\alpha_{\Pi^1_1}(\mathsf{ACA}_0)$, for some ordinal notation α).

Theorem 5.16. *Suppose $T \sqsupseteq \mathsf{ACA}_0^+$ then*

$$|T|_{\mathsf{WO}} = |T|_{\mathsf{ACA}_0}.$$

We prove the following general theorem

Theorem 5.17. *Suppose $\Pi^1_2(\Pi^0_2)$-sound theory T_0 is given by a finite list of axioms. Then for each $U \sqsupseteq T_0^+$ we have*

$$|U|_{\mathsf{WO}} = |U|_{T_0}.$$

Proof. Combining Lemma 5.13 and Proposition 5.6 we see that $|U|_{\mathsf{WO}} \leqslant |U|_{T_0}$. In order to show that $|U|_{\mathsf{WO}} \geqslant |U|_{T_0}$ we prove that for each $\boldsymbol{\alpha} < |U|_{T_0}$ we have $\boldsymbol{\alpha} < |U|_{\mathsf{WO}}$. We consider $\boldsymbol{\alpha} < |U|_{T_0}$. From Lemma 5.7 we see that there is an ordinal notation α and a true $\Sigma^1_1(\Pi^0_2)$-sentence φ such that $|\alpha| = \boldsymbol{\alpha}$ and

$$U + \varphi \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0)).$$

Since $T_0 \sqsupseteq \mathsf{RCA}_0$, we have

$$U + \varphi \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)).$$

And hence by Lemma 5.14 we have $U + \varphi \vdash \mathsf{WO}(\alpha)$. Thus

$$\boldsymbol{\alpha} = |\alpha| < |U + \varphi|_{\mathsf{WO}} = |U|_{\mathsf{WO}}.$$

This completes the proof of the theorem. ❑

5.4. **Robust reflection rank.** The *robust reflection rank* $|U|^\star_{T_0}$ of a theory $U \in \mathcal{E}\text{-}T_0$ over a theory $T_0 \sqsupseteq \mathsf{RCA}_0$ is defined as follows:

$$|U|^\star_{T_0} = \sup\{|U + \varphi|_{T_0} \colon \varphi \text{ is a true } \Sigma^1_1(\Pi^0_2)\text{-sentence}\}.$$

Proposition 5.18. *For theories $T_0 \sqsupseteq \mathsf{RCA}_0$ and $U \in \mathcal{E}\text{-}T_0$ the robust reflection rank $|U|^\star_{T_0}$ is an ordinal iff U is $\Pi^1_1(\Pi^0_3)$-sound.*

Proof. If U is $\Pi^1_1(\Pi^0_3)$-sound then for any true $\Sigma^1_1(\Pi^0_2)$-sentence φ the theory $U+\varphi$ is $\Pi^1_1(\Pi^0_3)$-sound. Thus, by Corollary 5.4 each rank $|U+\varphi|_{T_0} \in \mathbf{On}$ and so $|U|^\star_{T_0} \in \mathbf{On}$.

If U is not $\Pi^1_1(\Pi^0_3)$-sound then there is a false $\Pi^1_1(\Pi^0_3)$ sentence φ that U proves. Let ψ be a true $\Sigma^1_1(\Pi^0_2)$-sentence that is RCA_0-provably equivalent to $\neg\varphi$. Clearly, $U+\psi$ is inconsistent, so $U+\psi \prec_{\Pi^1_1(\Pi^0_3)} U+\psi$ and hence $\infty = |U+\psi|_{T_0} = |U|^\star_{T_0}$. ❑

Proposition 5.19. *Suppose $T_0 \sqsupseteq \mathsf{RCA}_0$ is $\Pi^1_2(\Pi^0_3)$-sound, $U \in \mathcal{E}\text{-}T_0$, and for some ordinal notation α we have $U \equiv_{\Pi^1_1(\Pi^0_3)} \mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0)$. Then*

$$|U|^\star_{T_0} = |U|_{T_0} = |\alpha|.$$

Proof. We use Proposition 5.6 and see that

$$|U|^\star_{T_0} \geqslant |U|_{T_0} = |\alpha|.$$

Let us assume for a contradiction that $|U|^\star_{T_0} > |\alpha|$. In this case from Lemma 5.7 there is a true $\Sigma^1_1(\Pi^0_2)$ sentence φ such that

$$U + \varphi \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U)).$$

Of course, this implies that

$$U + \vdash \varphi \to \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U)).$$

Note that $\varphi \to \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U))$ is a $\Pi^1_1(\Pi^0_3)$ sentence. Thus, from the assumption that $U \equiv_{\Pi^1_1(\Pi^0_3)} \mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0)$, it follows that:

$$\begin{aligned} \mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0) &\vdash \varphi \to \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U)) \\ \mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0) + \varphi &\vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U)) \\ &\vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U) + \varphi) \text{ by Lemma 5.9.} \\ &\vdash \mathsf{Con}(\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(U) + \varphi)) \end{aligned}$$

Thus, $\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0) + \varphi$ is inconsistent by Gödel's Second Incompleteness Theorem. On the other hand, since T_0 is $\Pi^1_2(\Pi^0_3)$ sound, $\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0)$ is $\Pi^1_1(\Pi^0_3)$ sound by Lemma 5.8. Thus, $\mathbf{R}^\alpha_{\Pi^1_1(\Pi^0_3)}(T_0) + \varphi$ is consistent by Lemma 5.9. This is a contradiction. ❑

Finally we connect the notions of robust reflection rank $|\cdot|^\star_{\mathsf{ACA}_0}$ and proof-theoretic ordinal $|\cdot|_{\mathsf{WO}}$:

Theorem 5.20. *For any theory $T \in \mathcal{E}$-ACA_0 with robust reflection rank $|T|^\star_{\mathsf{ACA}_0} = \boldsymbol{\alpha}$ we have $|T|_{\mathsf{WO}} = \boldsymbol{\varepsilon_\alpha}$ (here by definition we put $\boldsymbol{\varepsilon}_\infty = \infty$).*

Proof. First let us show that $|T|_{\mathsf{WO}} \geqslant \boldsymbol{\varepsilon_\alpha}$. We break into cases based on whether $\boldsymbol{\alpha} = \infty$ or $\boldsymbol{\alpha} \in \mathbf{On}$

Assume $\boldsymbol{\alpha} = \infty$. Then by Proposition 5.18 there is false Π^1_1 sentence φ that is provable in T. Now we could construct an ordinal notation α such that $\mathsf{WO}(\alpha)$ is ACA_0-provably equivalent to φ: we put φ in the tree normal form [24, Lemma V.1.4] and take α to be the Kleene-Brouwer order on the tree. Clearly, $T \vdash \mathsf{WO}(\alpha)$ and $|\alpha| = \infty$. Thus $|T|_{\mathsf{WO}} = \infty = \boldsymbol{\varepsilon_\alpha}$.

Now assume that $\boldsymbol{\alpha} \in \mathbf{On}$. Let us consider some $\boldsymbol{\beta} < \boldsymbol{\varepsilon_\alpha}$ and show that $|T|_{\mathsf{WO}} > \boldsymbol{\beta}$. From the definition of robust reflection rank it is easy to see that we could find some true $\Sigma^1_1(\Pi^0_2)$ sentence φ such that $\boldsymbol{\beta} < \boldsymbol{\varepsilon}_{\,|T+\varphi|_{\mathsf{ACA}_0}}$. Since $|T + \varphi|_{\mathsf{ACA}_0}$ is the rank of a Σ^0_1 binary relation, $|T + \varphi|_{\mathsf{ACA}_0} < \boldsymbol{\omega_1^{CK}}$. Thus we could choose an ordinal notation γ such that $|\gamma| < |T + \varphi|_{\mathsf{ACA}_0}$ but $\boldsymbol{\beta} < \boldsymbol{\varepsilon}_{|\gamma|+1}$. From Lemma 5.7 we infer that there is a true $\Sigma^1_1(\Pi^0_2)$-sentence φ' such that $T + \varphi + \varphi' \vdash \mathsf{RFN}_{\Pi^1_1}(\mathbf{R}^\gamma_{\Pi^1_1}(\mathsf{ACA}_0))$. We find a $\beta \prec \varepsilon_{\gamma+1}$ such that $|\beta| = \boldsymbol{\beta}$. By the same reasoning as in the proof of Theorem 5.11 we infer that $\mathbf{R}^{\gamma+1}_{\Pi^1_1}(\mathsf{ACA}_0) \vdash \mathsf{WO}(\beta)$. Thus $T + \varphi + \varphi' \vdash \mathsf{WO}(\beta)$. Hence $|T + \varphi + \varphi'|_{\mathsf{WO}} > \boldsymbol{\beta}$. From Kreisel's Theorem about proof-theoretic ordinals of extensions of ACA_0 we infer that $|T|_{\mathsf{WO}} = |T + \varphi + \varphi'|_{\mathsf{WO}} > \boldsymbol{\beta}$.

Now let us show that $|T|_{\mathsf{WO}} \leqslant \boldsymbol{\varepsilon_\alpha}$. Assume, for the sake of contradiction, that $|T|_{\mathsf{WO}} > \boldsymbol{\varepsilon_\alpha}$. Then there is an ordinal notation β with $|\beta| = \boldsymbol{\varepsilon_\alpha}$ such that $|T|_{\mathsf{WO}} \vdash \mathsf{WO}(\beta)$. Let us fix some ordinal notation α such that $|\alpha| = \boldsymbol{\alpha}$. Clearly, there is an isomorphism between β and ε_α. Let us denote by Iso the natural $\Sigma^1_1(\Pi^0_2)$-sentence expressing the latter fact. We see that $T + \mathsf{Iso} \vdash \mathsf{WO}(\varepsilon_\alpha)$. Thus by Lemma 5.12 we see that

$$T + \mathsf{Iso} \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\varepsilon_\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0)).$$

From Theorem 4.17 we conclude that

$$T + \mathsf{Iso} \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathbf{R}^{\alpha}_{\Pi^1_1(\Pi^0_3)}(\mathsf{ACA}_0)).$$

Since over ACA_0 every Π^1_1-formula is equivalent to a $\Pi^1_1(\Pi^0_3)$-formula,

$$T + \mathsf{Iso} \vdash \mathsf{RFN}_{\Pi^1_1}(\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)).$$

Therefore

$$|T|^{\star}_{\mathsf{ACA}_0} \geqslant |T + \mathsf{Iso}|_{\mathsf{ACA}_0} > |\mathbf{R}^{\alpha}_{\Pi^1_1}(\mathsf{ACA}_0)|_{\mathsf{ACA}_0} = |\alpha| = \boldsymbol{\alpha},$$

but $|T|^{\star}_{\mathsf{ACA}_0} = \boldsymbol{\alpha}$, a contradiction. ❑

6. Ordinal notation systems based on reflection principles

In this section we turn to ordinal notation systems based on reflection principles, like the one Beklemishev introduced in [3]. We will formally describe such a notation system momentarily, but, roughly, the elements of such notation systems are theories axiomatized by reflection principles and the ordering on them is given by consistency strength. Beklemishev endorsed the use of such notation systems as an approach to the well-known *canonicity problem* of ordinal notation systems. Since then, such notation systems have been intensively studied; see [12] for a survey of these notation systems and their properties.

We will consider ordinal notation systems based on the calculus RC^0 due to Beklemishev [6]. In earlier works, e.g. [3] on modal logic based ordinal analysis, ordinal notation systems arose from fragments of the polymodal provability logic GLP. However, this application of polymodal provability logic didn't required the full expressive power of GLP. Thus, starting from a work of Dashkov [10], strictly positive modal logics have been isolated that yield the same ordinal notation system as the logic GLP, but are much simpler from a technical point of view.

The set of formulas of RC^0 is given by the following inductive definition:

$$F ::= \top \mid F \wedge F \mid \Diamond_n F, \text{ where } n \text{ ranges over } \mathbb{N}.$$

An RC^0 sequent is an expression $A \vdash B$, where A and B are RC^0-formulas. The axioms and rules of inference of RC^0 are:

1. $A \vdash A$; $A \vdash \top$; if $A \vdash B$ and $B \vdash C$ then $A \vdash C$;
2. $A \wedge B \vdash A$; $A \wedge B \vdash B$; if $A \vdash B$ and $A \vdash C$ then $A \vdash B \wedge C$;
3. if $A \vdash B$ then $\Diamond_n A \vdash \Diamond_n B$, for all $n \in \mathbb{N}$;
4. $\Diamond_n \Diamond_n A \vdash \Diamond_n A$, for every $n \in \mathbb{N}$;
5. $\Diamond_n A \vdash \Diamond_m A$, for all $n > m$;
6. $\Diamond_n A \wedge \Diamond_m B \vdash \Diamond_n(A \wedge \Diamond_m B)$, for all $n > m$.

Let us describe the intended interpretation of RC^0-formulas in $\mathcal{L}_1$-sentences. The interpretation $\top^*$ of $\top$ is $0 = 0$. The interpretation $(A \wedge B)^*$ is $A^* \wedge B^*$. The interpretation $(\Diamond_n A)^*$ is $\mathsf{RFN}_{\Sigma_n}(A^*)$. A routine check by induction on the length of RC^0-derivations shows that if $A \vdash B$ then $\mathsf{EA} + A^* \vdash B^*$, for any RC^0-formulas A and B.

For a more extensive coverage of positive provability logic see [7].

We denote by $\mathcal{W}$ the set of all RC^0 formulas. The binary relation $<_n$, and the natural equivalence relation $\sim$ are given by

$$A <_n B \overset{\text{def}}{\iff} B \vdash \Diamond_n A, \quad A \sim B \overset{\text{def}}{\iff} B \vdash A \text{ and } A \vdash B.$$

The Beklemishev ordinal notation system for ε_0 is the structure $(\mathcal{W}/{\sim}, <_0)$.

The following result is due to Beklemishev (see [5, 6]):

Theorem 6.1. $(\mathcal{W}/{\sim}, <_0)$ *is a well-ordering with the order type* ε_0.

The transitivity of $(\mathcal{W}/{\sim}, <_0)$ is trivial. The linearity of $(\mathcal{W}/{\sim}, <_0)$ is provable by a purely syntactical argument within the system RC^0. But Beklemishev's proof of the well-foundedness of $(\mathcal{W}/{\sim}, <_0)$ was based on the construction of an isomorphism with Cantor's ordinal notation system for ε_0, i.e., Cantor normal forms.

Here we will give a proof of the well-foundedness part of Theorem 6.1 by providing an alternative interpretation of the $\Diamond_n$'s by reflection principles in *second*-order arithmetic and then applying the results of §3 to derive well-foundedness.

Theorem 6.2. $(\mathcal{W}, <_0)$ *is a well-founded relation.*

Proof. We prove that the set $\mathcal{W}$ of RC^0-formulas is well-founded with respect to $<_0$.

We give an alternative interpretation of RC^0. According to this interpretation, the image $\top^*$ of $\top$ is $0 = 0$, $(A \wedge B)^*$ is $A^* \wedge B^*$, and $(\Diamond_n A)^*$ is $\mathsf{RFN}_{\Pi^1_{n+1}}(\mathsf{ACA}_0 + A^*)$.

We note that if $A \vdash B$ is a derivable RC^0-sequent then $\mathsf{ACA}_0 + A^* \vdash B^*$. This can be checked by a straightforward induction on RC^0-derivations. Also from the definition it is clear that for any A the theory $\mathsf{ACA}_0 + A^*$ is Π^1_1-sound (and in fact true A^* is true).

Now assume for a contradiction that there is an infinite descending chain $A_0 >_0 A_1 >_0 \ldots$ of RC^0-formulas. Then $A_0^*, A_1^*, \ldots$ is an infinite sequence of sentences such that $\mathsf{ACA}_0 + A_i^* \vdash \mathsf{RFN}_{\Pi^1_1}(\mathsf{ACA}_0 + A_{i+1}^*)$. Henceforth we have a $\prec_{\Pi^1_1}$-descending chain of Π^1_1-sound extensions of ACA_0, contradicting Theorem 3.2. ❑

The key fact that we have used in this proof is that all the theories A_i^* are Π^1_1-sound. In fact all the theories under consideration are subtheories of ACA and hence the proof is naturally formalizable in $\mathsf{ACA}_0 + \mathsf{RFN}_{\Pi^1_1}(\mathsf{ACA})$.[4]

Now we show that the same kind of argument could be carried in ACA_0 itself.

Theorem 6.3. *For each* $A \in \mathcal{W}$*, the theory* ACA_0 *proves that* $(\{B \in \mathcal{W} \mid B <_0 A\}, <_0)$ *is well-founded.*

Proof. Note that in RC^0 any formula A follows from formulas $\Diamond_n \top$ such that, for all $\Diamond_m$ that occur in A, $m < n$; this fact could be proved by a straightforward induction on length of A. Clearly, for any such n, the set $\{B \in \mathcal{W} \mid B <_0 A\}$ is a subset of $\{B \in \mathcal{W} \mid B <_0 \Diamond_n \top\}$. Thus, without loss of generality, we may conside only the case of A being of the form $\Diamond_n \top$.

Now we reason in ACA_0. We assume for a contradiction that there is an infinite descending chain $\Diamond_n \top >_0 A_0 >_0 A_1 >_0 \ldots$ of RC^0-formulas.

[4]The fact that $\mathsf{ACA} \equiv_{\Pi^1_\infty} \mathsf{RFN}_{\Pi^1_\infty}(\mathsf{ACA}_0)$ could be proved by a standard technique going back to Kreisel and Lévy [16]. A study of the exact correspondence between restrictions of the schemes of reflection and induction in the setting of second order arithmetic has been recently performed by Frittaion [14].

We construct a countably-coded ω-model $\mathfrak{M}$ of RCA_0 that contains this chain. Note that using arithmetical comprehension we could construct a (set encoding) partial satisfaction relation for $\mathfrak{M}$ that the sentence RCA_0 (conjunction of all axioms from some natural finite axiomatization of RCA_0) and all $\Pi^1_{n+1}(\Pi^0_3)$ formulas. We want to show that if RCA_0 proves some $\Pi^1_{n+1}(\Pi^0_3)$ sentence φ then φ is true in $\mathfrak{M}$. For this we consider any cut-free proof p of the sequent $\neg\mathsf{RCA}_0, \varphi$. And next by induction on subproofs of p show that all sequents in p are valid in $\mathfrak{M}$ (according to the partial satisfaction relation that we constructed above). Hence the principle $\mathsf{RFN}_{\Pi^1_{n+1}(\Pi^0_3)}(\mathsf{RCA}_0)$ holds in $\mathfrak{M}$.

We again define an alternative interpretation of RC^0. The interpretation $\top^*$ is $0 = 0$, the interpretations $(A \wedge B)^*$ are $A^* \wedge B^*$, and the intepretations $(\Diamond_k A_i)^*$ are $\mathsf{RFN}_{\Pi^1_{k+1}(\Pi^0_3)}(\mathsf{RCA}_0 + A^*_i)$. From the previous paragraph we see that $\mathfrak{M} \models (\Diamond_n \top)^*$. And since $\Diamond_n \top >_0 A_0$, we have $\mathfrak{M} \models (\Diamond_0 A_0)^*$, i.e., $\mathfrak{M} \models \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0 + A^*_0)$. Thus in $\mathfrak{M}$ there is an infinite sequence of theories $\mathsf{RCA}_0 + A^*_0, \mathsf{RCA}_0 + A^*_1, \ldots$ such that $\mathsf{RCA}_0 + A^*_i \vdash \mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0 + A^*_{i+1})$ and $\mathsf{RFN}_{\Pi^1_1(\Pi^0_3)}(\mathsf{RCA}_0 + A^*_0)$. Since $\mathfrak{M}$ is a model of RCA_0, by Theorem 3.3 we reach a contradiction. ❑

References

[1] Lev Beklemishev. Iterated local reflection versus iterated consistency. *Annals of Pure and Applied Logic*, 75(1-2):25–48, 1995.

[2] Lev Beklemishev. Proof-theoretic analysis by iterated reflection. *Archive for Mathematical Logic*, 42(6):515–552, 2003.

[3] Lev Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128(1-3):103–123, 2004.

[4] Lev Beklemishev. Reflection principles and provability algebras in formal arithmetic. *Russian Mathematical Surveys*, 60(2):197, 2005.

[5] Lev Beklemishev. Veblen hierarchy in the context of provability algebras. In P. Hájek, L. Valdés-Villanueva, and D. Westerståhl, editors, *Logic, Methodology and Philosophy of Science, Proceedings of the Twelfth International Congress*, pages 65–78. Kings College Publications, London, 2005. Preprint: Logic Group Preprint Series 232, Utrecht University, June 2004.

[6] Lev Beklemishev. Calibrating provability logic: From modal logic to reflection calculus. *Advances in modal logic*, 9:89–94, 2012.

[7] Lev Beklemishev. Positive provability logic for uniform reflection principles. *Annals of Pure and Applied Logic*, 165(1):82–105, 2014.

[8] Lev Beklemishev. Reflection calculus and conservativity spectra. *arXiv preprint arXiv:1703.09314*, 2017.

[9] Lev Beklemishev and Fedor Pakhomov. Reflection algebras and conservation results for theories of iterated truth. *arXiv preprint arXiv:1908.10302*, 2019.

[10] Evgenij Dashkov. On the positive fragment of the polymodal provability logic GLP. *Mathematical Notes*, 91(3-4):318–333, 2012.

[11] Ali Enayat and Fedor Pakhomov. Truth, disjunction, and induction. *Archive for Mathematical Logic*, 58(5):753–766, Aug 2019.

[12] David Fernández-Duque. Worms and spiders: Reflection calculi and ordinal notation systems. *arXiv preprint arXiv:1605.08867*, 2016.

[13] Harvey Friedman. Uniformly defined descending sequences of degrees. *The Journal of Symbolic Logic*, 41(2):363–367, 1976.

[14] Emanuele Frittaion. Uniform reflection in second order arithmetic. 2019.

[15] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*, volume 3. Cambridge University Press, 2017.

[16] Georg Kreisel and Azriel Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Mathematical Logic Quarterly*, 14(7-12):97–142, 1968.

[17] Per Lindström. *Aspects of incompleteness*, volume 10. Cambridge University Press, 2017.

[18] Patrick Lutz and James Walsh. Incompleteness and jump hierarchies. *arXiv preprint arXiv:1909.10603*, 2019.
[19] Fedor Pakhomov and James Walsh. Reflection ranks and ordinal analysis. *arXiv preprint arXiv:1805.02095v1*, 2018.
[20] Wolfram Pohlers. *Proof theory: The first step into impredicativity*. Springer Science, 2008.
[21] Michael Rathjen. The realm of ordinal analysis. In *Sets and Proofs*, pages 219–279. Cambridge University Press, 1999.
[22] Ulf R. Schmerl. A fine structure generated by reflection formulas over primitive recursive arithmetic. *Studies in Logic and the Foundations of Mathematics*, 97:335–350, 1979.
[23] Ulf R. Schmerl. Iterated reflection principles and the ω-rule. *The Journal of Symbolic Logic*, 47(4):721–733, 1982.
[24] Stephen G. Simpson. *Subsystems of second order arithmetic*, volume 1. Cambridge University Press, 2009.
[25] Stephen G. Simpson and Rick L. Smith. Factorization of polynomials and $\Sigma 10$ induction. *Annals of Pure and Applied Logic*, 31(2):289–306, 1986.
[26] Craig Smorynski. *Self-reference and modal logic*. Springer Science & Business Media, 2012.
[27] John Steel. Descending sequences of degrees. *The Journal of Symbolic Logic*, 40(1):59–61, 1975.

Steklov Mathematical Institute, Moscow
Institute of Mathematics of the Czech Academy of Sciences
Email address: pakhfn@mi-ras.ru

Group in Logic and the Methodology of Science, University of California, Berkeley
Email address: walsh@math.berkeley.edu