THE REVIEW OF SYMBOLIC LOGIC
Volume 15, Number 4, December 2022

A MATHEMATICAL COMMITMENT WITHOUT COMPUTATIONAL STRENGTH

ANTON FREUND
Technical University of Darmstadt

Abstract. We present a new manifestation of Gödel's second incompleteness theorem and discuss its foundational significance, in particular with respect to Hilbert's program. Specifically, we consider a proper extension of Peano arithmetic (**PA**) by a mathematically meaningful axiom scheme that consists of Σ^0_2-sentences. These sentences assert that each computably enumerable (Σ^0_1-definable without parameters) property of finite binary trees has a finite basis. Since this fact entails the existence of polynomial time algorithms, it is relevant for computer science. On a technical level, our axiom scheme is a variant of an independence result due to Harvey Friedman. At the same time, the meta-mathematical properties of our axiom scheme distinguish it from most known independence results: Due to its logical complexity, our axiom scheme does not add computational strength. The only known method to establish its independence relies on Gödel's second incompleteness theorem. In contrast, Gödel's theorem is not needed for typical examples of Π^0_2-independence (such as the Paris–Harrington principle), since computational strength provides an extensional invariant on the level of Π^0_2-sentences.

§1. Summary of mathematical results. This paper consists of mathematical results and a foundational discussion. The former are summarized in the present section; the latter can be found in Section 2. In the remaining sections we provide detailed proofs of all mathematical claims.

First and foremost, our paper is based on a result by Dick de Jongh (unpublished; cf. the introduction to [35]) and Diana Schmidt [36]: The embeddability relation on finite binary trees yields a well partial order with maximal order type ε_0 (see below for an explanation). Harvey Friedman [38] has shown that this type of result yields statements of finite combinatorics that are independent of important mathematical axiom systems. Against this background, many arguments in the present paper may be considered folklore. Nevertheless we find it worthwhile to give an explicit presentation, not least because the arguments are rather sensitive with respect to quantifier complexity and the presence of parameters. At some places we provide more details than the expert may find necessary. The aim is to make the paper as accessible and self-contained as possible.

We write $\mathcal{B}$ for the set of finite binary trees. More precisely, we assume that each tree has a distinguished root node, that nodes have either zero or two children, and that left

Received: April 15, 2020.
2020 *Mathematics Subject Classification*: 03F30, 03F40, 03A05, 68R10.
Keywords and phrases: independence, computational strength, Gödel's second incompleteness theorem, Hilbert's program, Kruskal's theorem, polynomial-time algorithm.

doi:10.1017/S1755020321000265

child and right child can be distinguished. Furthermore, we identify isomorphic trees. Formally, we view $\mathcal{B}$ as the least fixed point of the following inductive clauses:

(i) There is an element $\circ \in \mathcal{B}$ (the tree that consists of a single root node).
(ii) Given s and t in $\mathcal{B}$, we obtain an element $\circ(s,t) \in \mathcal{B}$ (the tree in which the root has left subtree s and right subtree t).

For $s, t \in \mathcal{B}$ we write $s \leq_{\mathcal{B}} t$ if there is a tree embedding of s into t. Such an embedding can either map the root to the root and the immediate subtrees of s into the corresponding subtrees of t; or it maps all of s into one subtree of t. Hence we have $\circ \leq_{\mathcal{B}} t$ for any $t \in \mathcal{B}$; we have $s \leq_{\mathcal{B}} \circ$ precisely for $s = \circ$; and we have

$$\circ(s_0, s_1) \leq_{\mathcal{B}} \circ(t_0, t_1) \quad \Leftrightarrow \quad \begin{cases} s_0 \leq_{\mathcal{B}} t_0 \text{ and } s_1 \leq_{\mathcal{B}} t_1, \\ \text{or } \circ(s_0, s_1) \leq_{\mathcal{B}} t_i \text{ for some } i \in \{0,1\}. \end{cases}$$

These clauses provide a recursive definition of $\leq_{\mathcal{B}}$.

Recall that a partial order consists of a set X and a binary relation $\leq_X$ on X that is reflexive, antisymmetric and transitive. A finite or infinite sequence $x_0, x_1, \ldots$ in X is called good if there are indices $i < j$ such that we have $x_i \leq_X x_j$; otherwise, the sequence is called bad. If there is no infinite bad sequence, then $(X, \leq_X)$ is called a well partial order (wpo). Equivalently, a partial order $(X, \leq_X)$ is a wpo if, and only if, every subset $Y \subseteq X$ has a finite "basis" $a \subseteq Y$ with the following property: for any $y \in Y$ there is an $x \in a$ with $x \leq_X y$ (cf. the argument in Remark 3.1 below).

If X is a wpo, then all its linearizations are well orders (since a strictly decreasing sequence in a linearization would be a bad sequence in X). Hence the order type of each linearization is an ordinal number. The supremum of these ordinals is called the maximal order type of X. As shown by D. de Jongh and R. Parikh [8], the maximal order type of any wpo is realized by one of its linearizations (i.e., the supremum is a maximum).

Kruskal's theorem [28] implies that $(\mathcal{B}, \leq_{\mathcal{B}})$ is a well partial order. We point out that the theorem applies to arbitrary (i.e., not necessarily binary) finite trees; the "most general" version of Kruskal's theorem is investigated in [14]. Concerning the binary case, de Jongh and Schmidt have proved the finer result that $\mathcal{B}$ has maximal order type ε_0, which is the least fixed point of ordinal exponentiation with base ω (read [36, theorem II.2] in combination with the example after [36, definition I.15]). A classical result of G. Gentzen [18, 19] establishes ε_0 as the proof theoretic ordinal of Peano arithmetic (**PA**). This explains the connection with independence results.

In the present paper we consider the binary Kruskal theorem (i.e., the result that $\mathcal{B}$ is a wpo) in the context of first order arithmetic; an introduction to this setting can be found in [20]. We will be particularly interested in questions of quantifier complexity: Recall that a formula lies in the class $\Delta^0_0 = \Sigma^0_0 = \Pi^0_0$ if it only contains bounded quantifiers. Since the latter range over a finite domain, the truth of closed Δ^0_0-formulas is decidable. A Σ^0_{n+1}-formula (Π^0_{n+1}-formula) has the form $\exists_x \varphi$ (the form $\forall_x \varphi$), where φ is a Π^0_n-formula (Σ^0_n-formula). Recall that the Σ^0_1-formulas correspond to the computably enumerable relations. A relation is Δ^0_1-definable (in **PA**) if it has a Σ^0_1-definition and a Π^0_1-definition (which **PA** proves to be equivalent). The Δ^0_1-relations coincide with the decidable ones.

Working in **PA**, the elements of $\mathcal{B}$ can be represented by numerical codes for finite sets of sequences with entries from $\{0, 1\}$. Note that the relations $s \in \mathcal{B}$ and $s \leq_{\mathcal{B}} t$ are

Δ^0_1-definable in **PA**. As mentioned above, the fact that $\mathcal{B}$ is a wpo can be expressed in terms of a finite basis property. To state the latter we abbreviate

$$\exists^{\text{fin}}_a \psi(a) :\equiv \exists_a(\text{“}a \in \mathbb{N} \text{ codes a finite set”} \wedge \psi(a)).$$

In the context of **PA** it is natural to focus on definable sets. Given a formula $\varphi(s)$ with a distinguished free variable, the finite basis property for $\{s \in \mathcal{B} \mid \varphi(s)\} \subseteq \mathcal{B}$ can be formalized as

$$\mathcal{K}\varphi :\equiv \exists^{\text{fin}}_{a\subseteq\mathcal{B}}(\forall_{s\in a}\varphi(s) \wedge \forall_{t\in\mathcal{B}}(\varphi(t) \to \exists_{s\in a} s \leq_{\mathcal{B}} t)).$$

Note that the quantifiers with subscript $s \in a$ are bounded, since a is a code for a finite set (cf. [20, lemma I.1.32]); in contrast, the quantifiers with subscripts $a \subseteq \mathcal{B}$ and $t \in \mathcal{B}$ are unbounded. The symbol $\mathcal{K}$ alludes to Kruskal's theorem, which implies that all instances $\mathcal{K}\varphi$ are true (see Remark 3.1 for details). We will be most interested in the axiom scheme

$$\mathcal{K}\Sigma^-_1 := \{\mathcal{K}\varphi \mid \text{“}\varphi \text{ a } \Sigma^0_1\text{-formula with exactly one free variable”}\}.$$

The superscript of Σ^-_1 emphasizes the fact that no further free variables are allowed. This ensures that each instance of $\mathcal{K}\Sigma^-_1$ is a closed Σ^0_2-formula.

To motivate the restrictions on the quantifier complexity and the parameters, we recall the notion of computational strength: A computable function $f : \mathbb{N} \to \mathbb{N}$ is provably total in a suitable theory **T** if the latter proves $\forall_x \exists!_y\, \varphi(x, y)$ for some Σ^0_1-definition φ of the graph of f (where $\exists!$ abbreviates the existence of a unique witness). The computational strength of a theory is commonly identified with the collection of its provably total computable functions.

It is known that the computational strength of a theory does not increase when we add a true Π^0_1-sentence ψ as an axiom. Essentially, this is due to the fact that the Σ^0_1-formula $\psi \to \varphi(x, y)$ defines the same graph as $\varphi(x, y)$ (note that the definition of provably total function is extensional). A simple but fundamental observation shows that the same is true for closed Σ^0_2-axioms: It suffices to note that any true Σ^0_2-sentence $\exists_x \psi(x)$ follows from some true Π^0_1-instance $\psi(\overline{n})$ (see Proposition 3.2 for details). Note that we may not be able to compute the correct witness $n \in \mathbb{N}$; this issue will resurface at the end of the present section.

The general facts from the previous paragraph imply that $\mathbf{PA} + \mathcal{K}\Sigma^-_1$ has the same computational strength as **PA**. At this point it is crucial that we exclude parameters: If the Σ^0_1-formula φ contains further free variables, then the universal closure of $\mathcal{K}\varphi$ is a Π^0_3-formula, so that our argument does not longer apply. Note that the version with parameters can be expressed by a single Π^0_3-sentence (rather than a scheme), due to the existence of a universal computably enumerable set.

Next, we explain why $\mathbf{PA} + \mathcal{K}\Sigma^-_1$ is a proper extension of **PA**. Based on a notation system for the ordinal ε_0 (see Section 4 for details), transfinite induction can be expressed in first order arithmetic: Given a formula $\psi(\alpha)$ with a distinguished free variable, we set

$$\text{Prog}_{\varepsilon_0}(\psi) :\equiv \forall_{\gamma\prec\varepsilon_0}(\forall_{\beta\prec\gamma}\psi(\beta) \to \psi(\gamma)),$$
$$\mathcal{TI}(\varepsilon_0, \psi) :\equiv \text{Prog}_{\varepsilon_0}(\psi) \to \forall_{\alpha\prec\varepsilon_0}\psi(\alpha).$$

The scheme of parameter-free Π^0_1-induction up to ε_0 is the collection

$$\mathcal{TI}(\varepsilon_0, \Pi^-_1) := \{\mathcal{TI}(\varepsilon_0, \psi) \mid \text{“}\psi \text{ a } \Pi^0_1\text{-formula with exactly one free variable”}\}.$$

https://doi.org/10.1017/S1755020321000265 Published online by Cambridge University Press

In Section 4 we show that each instance of $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$ can be proved in $\mathbf{PA} + \mathcal{K}\Sigma_1^-$. This is a straightforward consequence of the fact that ε_0 is bounded by (and in fact equal to) the maximal order type of $\mathcal{B}$. Nevertheless we find it worthwhile to give a detailed proof, which pays attention to the quantifier complexities and the role of parameters. Gentzen [18] has used Π_1^0-induction up to ε_0 to establish the consistency of **PA**. This induction does not require parameters, as we will check in Section 5. Hence the consistency of **PA** can be proved in $\mathbf{PA} + \mathcal{K}\Sigma_1^-$. The latter must thus be a proper extension, due to Gödel's second incompleteness theorem.

In Section 6 we review the proof that $\mathcal{B}$ has maximal order type ε_0. This will allow us to show that, conversely, $\mathbf{PA} + \mathcal{TI}(\varepsilon_0, \Pi_1^-)$ proves each instance of $\mathcal{K}\Sigma_1^-$. To complete the picture, we relate transfinite induction and reflection. Let $\mathrm{Pr}_{\mathbf{PA}}(\varphi)$ be a standard formalization of the statement that the formula φ is provable in **PA** (see [20, sec. I.4(a)]; we will also write φ for $\ulcorner\varphi\urcorner$). Given a sentence φ of first order arithmetic, we put

$$\mathrm{Rfn}_{\mathbf{PA}}(\varphi) :\equiv \mathrm{Pr}_{\mathbf{PA}}(\varphi) \to \varphi.$$

The local (i.e., parameter-free) Σ_2^0-reflection principle over **PA** is the collection

$$\mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0) :\equiv \{\mathrm{Rfn}_{\mathbf{PA}}(\varphi) \mid \text{“}\varphi \text{ a closed } \Sigma_2^0\text{-formula”}\}.$$

Due to G. Kreisel and A. Lévy [27], uniform reflection over **PA** is equivalent to ε_0-induction for formulas with parameters. We will show that the proof can be adapted to the parameter free case (which can also be inferred from work of L. Beklemishev [2, 3], cf. Remark 5.2 below). This results in Theorem 7.3, which asserts

$$\mathbf{PA} + \mathcal{K}\Sigma_1^- \;\equiv\; \mathbf{PA} + \mathcal{TI}(\varepsilon_0, \Pi_1^-) \;\equiv\; \mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0). \qquad (*)$$

In view of Goryachev's theorem (see, e.g., [30, theorem IV.5]), we can conclude the following (cf. Corollary 7.4 below): Over Peano arithmetic, the Π_1^0-consequences of $\mathcal{K}\Sigma_1^-$ are precisely those of the finitely iterated consistency statements for **PA**. Due to another result of Kreisel and Lévy [27], we can also deduce that $\mathbf{PA} + \mathcal{K}\Sigma_1^-$ is not contained in any consistent extension of **PA** by a computably enumerable set of Π_2^0-sentences (see Corollary 5.4).

Concerning the structure of our paper, we point out that the inclusions $\supseteq$ from $(*)$ are proved in Sections 4 and 5, while the inclusions $\subseteq$ are proved in Sections 6 and 7. As we will see in the next section, the inclusions $\supseteq$ suffice to ensure many (but not all) of the properties that make $\mathcal{K}\Sigma_1^-$ foundationally significant. Some readers may thus prefer to skip Sections 6 and 7. For others, these sections may constitute the most interesting part of our paper.

§2. Foundational considerations. In the previous section we have presented an extension of Peano arithmetic by an axiom scheme $\mathcal{K}\Sigma_1^-$ that is related to Kruskal's theorem. The present section is concerned with the foundational significance of this extension.

Let us first recall some aspects of Hilbert's program; for a more thorough discussion and further references we refer to the introduction by R. Zach [49]. To secure the abstract methods that are central to modern mathematics, Hilbert wanted to justify them by finitist reasoning about natural numbers, which he views as “extralogical concrete objects that are intuitively present as immediate experience prior to all thought” [22, p. 171]. (All quotations from [22, 23] are translated as in [48].) The status of the natural numbers entails that certain statements about them are finitistically meaningful. This includes, first of all, statements which assert that a given tuple of numbers satisfies some primitive recursive relation. Such a statement can be verified explicitly, which explains its privileged role, but also entails—as Hilbert [22, p. 165] puts it—that it is "of no essential interest when considered by itself." In addition, one admits universal statements with verifiable instances. According to Hilbert [22, p. 173], such a statement can be accepted as "a hypothetical judgement that comes to assert something when a numeral is given." In contrast, unbounded existential statements are not seen as finitistically meaningful, as "one cannot [...] try out all numbers" [23, p. 73]. At the same time, Hilbert [23, p. 77f] emphasizes the fact that existential statements play an extremely fruitful role in abstract mathematics. One could even be tempted to say that abstract notions acquire meaning through their role in the mathematical development, a position that seems to resonate with the following statement by Hilbert [23, p. 79]:

> "To make it a universal requirement that each individual formula [...] be interpretable by itself is by no means reasonable; on the contrary, a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument."

However, such a conception of meaning is very different from the finitist one.

The extent of finitist reasoning is commonly identified with primitive recursive arithmetic (**PRA**). This identification has been justified by W. Tait [44]; in [45] he lists and refutes some objections. A quantifier-free formulation seems to be most appropriate: In a such a setting, one can only express statements that are finitistically meaningful; universal statements correspond to open formulas. We will, nevertheless, work in the usual framework of first order arithmetic with quantifiers, since the latter are needed to express our existential statements $\mathcal{K}\varphi$. Following C. Smoryński [42], we agree to identify the finitistically meaningful statements with the Π^0_1-sentences.

More specifically, then, Hilbert's program suggested to formalize all of abstract mathematics as an axiom system **T**. In order to obtain a finitist justification, one was supposed to prove the consistency of **T** in the theory **PRA**. At this point it is important to note that consistency is not merely a minimal requirement: If the consistency of a theory **T** is provable in **PRA**, then the latter proves all Π^0_1-theorems of **T**, i.e., all results that are finitistically meaningful (see [23, p. 78f]). Gödel's incompleteness theorems show that Hilbert's program cannot be carried out: It is impossible for **T** to prove its own consistency; a fortiori, the consistency of **T** cannot be established in the weaker theory **PRA**.

Despite Gödel's theorems, the aims of Hilbert's program have been achieved to an astonishing extent: A substantial part of contemporary mathematics can indeed be formalized in rather weak axiom systems (see, e.g., the work of S. Feferman [10], as well as U. Kohlenbach's proof mining program [24]). In view of these positive results, it is all the more intriguing to ask: Are there natural mathematical theorems that can be expressed but not proved in **PRA**, or in some stronger theory? To count as a natural theorem, the unprovable statement should arise from mathematical practice; it should not involve the logical notions of proof or model. In particular, consistency statements (which are unprovable by Gödel's theorem) are not seen as examples of this type.

We do have good examples of true Π^0_2-statements that are unprovable in relevant axiom systems: The Paris–Harrington principle cannot be proved in Peano arithmetic [31]; Friedman's miniaturization of Kruskal's theorem is independent of an even stronger system [38], which is associated with predicative mathematics. The situation is less satisfactory when it comes to Π^0_1-sentences, which are most important from the finitist viewpoint: The independent statement due to S. Shelah [37] involves notions from model theory, so that its status as a natural mathematical theorem can be questioned. Friedman has presented work on Π^0_1-independence from Zermelo–Fraenkel set theory (see, e.g., [15]), but his results are not yet published in final form. In the present author's opinion, the search for mathematical Π^0_1-sentences that are independent of relevant axiom systems remains one of the most interesting challenges in mathematical logic.

The axiom scheme $\mathcal{K}\Sigma^-_1$ from the previous section does not settle the challenge of natural Π^0_1-independence. The latter can, nevertheless, serve as a benchmark that helps us to assess the foundational significance of $\mathcal{K}\Sigma^-_1$. In the rest of this section we carry out such an assessment.

First, we will argue that $\mathcal{K}\Sigma^-_1$ is a natural mathematical commitment. In the previous section we have seen that $\mathcal{K}\Sigma^-_1$ is a restricted version of Kruskal's theorem. The latter is firmly established as a natural result of mathematical practice. Hence it remains to argue that the restrictions that lead to $\mathcal{K}\Sigma^-_1$ are natural as well.

In formulating $\mathcal{K}\Sigma^-_1$, we have restricted Kruskal's theorem in two ways: Firstly, we have decided to work with binary rather than arbitrary finite trees. This restriction makes it easier to determine the precise strength of $\mathcal{K}\Sigma^-_1$ (i.e., to prove the equivalence with transfinite induction and local reflection), but it is not essential: If we extend our axiom scheme to arbitrary finite trees, then it will imply the consistency of stronger axiom systems; at the same time, it will still not increase the computational strength, since it also consists of Σ^0_2-statements. The graph minor theorem of N. Robertson and P. Seymour [34] suggests a very intriguing axiom scheme that is even stronger (cf. [16]) but does not have computational strength either (for the same general reason). In summary, the restriction to binary trees is purely pragmatic and does not change the general foundational behavior. Secondly, the scheme $\mathcal{K}\Sigma^-_1$ is a restriction of Kruskal's theorem insofar as it demands a finite basis for computably enumerable—rather than arbitrary—sets of trees. In the following we give two justifications for the restriction to computably enumerable sets.

The first justification is that $\mathcal{K}\Sigma^-_1$ suffices for certain applications in computer science: Assume that P is an upward closed property of finite binary trees, which means that $P(s)$ and $s \leq_{\mathcal{B}} t$ imply $P(t)$. Often (but not always, cf. [11, theorem 3]) one will already know that P is decidable. Then P can be defined by a Σ^0_1-formula, and $\mathcal{K}\Sigma^-_1$ yields a finite $a \subseteq \mathcal{B}$ such that $P(t)$ is equivalent to $\exists_{s \in a} s \leq_{\mathcal{B}} t$. The latter can be decided in polynomial time (in the size of t). The author knows of no concrete applications in the context of trees, but the analogous argument for the graph minor relation has many applications (see, e.g., [12]).

The second justification for the restriction to computably enumerable sets is based on the idea that one can have reasons to accept $\mathcal{K}\Sigma^-_1$ but not the full Kruskal theorem for binary trees. To make this plausible we recall that $\mathcal{K}\Sigma^-_1$ is equivalent to the principle $\mathcal{TI}(\varepsilon_0, \Pi^-_1)$ of parameter-free Π^0_1-induction up to ε_0. The latter is no stronger than induction for decidable (i.e., finitistically meaningful) properties, still up to ε_0 (see, e.g., [43, lemma 4.5]). On the other hand, the full Kruskal theorem for binary trees is naturally formulated in second order arithmetic. Over the theory $\mathbf{ACA}_0$ from reverse mathematics (see [40] for an introduction), it is equivalent both to our statement $\forall_{X\subseteq\mathbb{N}}\mathcal{K}\varphi$ with $\varphi(s)\equiv s\in X$ and to the well-foundedness of ε_0 (see [33] for the crucial direction from well-foundedness to Kruskal's theorem). Now from a finitist standpoint it makes sense to accept $\mathcal{TI}(\varepsilon_0,\Pi_1^-)$ but not the second order statement that ε_0 is well-founded. Indeed, Tait [45, p. 411] states that Kreisel [25] accepts quantifier-free induction up to each ordinal below ε_0 as finitist. Also, G. Takeuti's justification of transfinite induction is supposed to "involve 'Gedankenexperimente' [thought experiments] only on clearly defined operations applied to some concretely given figures" [46, p. 97].

Next, we discuss the fact that $\mathcal{K}\Sigma_1^-$ is a scheme rather than a single statement. In the previous section we have explained that $\mathbf{PA}+\mathcal{K}\Sigma_1^-$ proves the consistency of $\mathbf{PA}$. Of course, this proof involves only finitely many instances $\mathcal{K}\varphi_1,\dots,\mathcal{K}\varphi_n$. However, we see no basis for the claim that these particular instances constitute a natural mathematical commitment—in contrast to the axiom scheme as a whole. In this sense our reference to an axiom scheme is essential. What does this entail? We think that the answer depends on our attitude toward independence phenomena.

One possibility is to think of independent statements as "unsolvable conjectures." More explicitly, one might imagine a mathematician immersed in Peano arithmetic, who is challenged to prove or refute the Paris–Harrington principle. The independence result tells us that this mathematician can never succeed. This conception of independence is clearly concerned with single statements rather than schemes. However, one can also think of independence in terms of "potential axioms." For example, one may view the principle of induction for arbitrary first order formulas as a mathematical commitment beyond the finitist standpoint. This example shows that schemes play a natural role within such a conception of independence.

A broad conception of independence may even incorporate rules, in addition to axiom schemes. In the present context it is interesting to consider the rule

$$\frac{\forall_{\gamma\prec\varepsilon_0}(\forall_{\beta\prec\gamma}\psi(\beta)\to\psi(\gamma))}{\forall_{\alpha\prec\varepsilon_0}\psi(\alpha)}$$

of Π_1^0-induction along ε_0, which allows us to infer $\forall_{\alpha\prec\varepsilon_0}\psi(\alpha)$ once we have given a proof of $\forall_{\gamma\prec\varepsilon_0}(\forall_{\beta\prec\gamma}\psi(\beta)\to\psi(\gamma))$, where $\psi(\alpha)$ can be any Π_1^0-formula without further free variables. Note that the rule does not commit us to the contrapositive of the corresponding axiom, i.e., to the least element principle. Hence the rule avoids certain existential commitments, which is well motivated in a finitist context. As shown by Beklemishev [2, theorem 3], the closure of $\mathbf{PA}$ under the rule of Π_1^0-induction along ε_0 proves the same theorems as the extension of $\mathbf{PA}$ by finitely iterated consistency statements. Note that the rule does not refer to logical notions such as proof or model. Insofar as induction up to ε_0 is a result of mathematical practice, we have a mathematical commitment on the level of Π_1^0-statements.

Let us now discuss the fact that $\mathcal{K}\Sigma_1^-$ consists of Σ_2^0-statements rather than Π_1^0-statements. At the end of the previous section we have mentioned that there is no computably enumerable set Ψ of Π_1^0-sentences (or even Π_2^0-sentences) such that $\mathbf{PA}+\Psi$ is consistent and contains $\mathbf{PA}+\mathcal{K}\Sigma_1^-$. This shows that our use of Σ_2^0-sentences is essential in a rather strong sense. How, then, do Σ_2^0- and Π_2^0-independence compare from a

https://doi.org/10.1017/S1755020321000265 Published online by Cambridge University Press

foundational perspective? To set the stage for this question, we first situate $\mathcal{K}\Sigma_1^-$ within the context of Gentzen's ordinal analysis.

Gentzen showed that each purported proof of a contradiction can be reduced to a proof with smaller ordinal label. To establish consistency, one can use this reduction in two different ways: First, it follows that a purported proof p of a contradiction leads to an infinite sequence of such proofs, labeled by a strictly decreasing sequence of ordinals. These sequences are primitive recursive with parameter p. To derive consistency, one can then invoke the primitive recursive well-foundedness of ε_0. The latter is, in fact, equivalent to uniform Σ_1^0-reflection (see [17, theorem 4.5]), which can be expressed by a single Π_2^0-statement. Secondly, one can use induction on $\alpha \prec \varepsilon_0$ to show that no proof with label α can produce a contradiction. This avoids parameters but involves a universal quantification over ordinals. As we will show, it leads to an equivalence between $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$ and $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0)$, which are axiom schemes with instances of complexity Σ_2^0.

It seems that the route via primitive recursive well-foundedness is preferred in the finitist literature. For example, Takeuti writes that the consistency proof is based on the following [46, p. 92]:

> "Whenever a concrete method of constructing decreasing sequences of ordinals is given, any such decreasing sequence must be finite."

This preference may help to explain the pre-eminence of Π_2^0-independence. As an exception, we mention that L. Beklemishev and A. Visser [4] have characterized the Σ_n^0-consequences of **PA** (and of its fragments) in terms of iterated reflection. Kreisel [26] has initiated work on finiteness theorems of complexity Σ_2^0, but here the focus is on proof-mining rather than independence.

We have seen that Gentzen's consistency proof can be concluded in two subtly different ways. These correspond to different mathematical principles that are independent of Peano arithmetic: It is well known that the strengthened finite Ramsey theorem of J. Paris and L. Harrington is equivalent to uniform Σ_1^0-reflection (see [31, theorem 3.1]) and hence to the primitive recursive well-foundedness of ε_0. The present paper complements this classical result by the equivalence between the Kruskal scheme $\mathcal{K}\Sigma_1^-$, the scheme $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0)$ of local Σ_2^0-reflection, and the transfinite induction principle $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$.

Finally, let us take up the comparison of Σ_2^0- and Π_2^0-independence. Extending Hilbert's view on Π_1^0-sentences, one could see Π_2^0-sentences as "hypothetical judgement[s]" [22, p. 173] of complexity Σ_1^0. This might suggest that Π_2^0-sentences are less abstract—in the finitist sense—than Σ_2^0-statements. From this viewpoint, the independence of $\mathcal{K}\Sigma_1^-$ would be less significant than classical independence results, such as the one by Paris and Harrington.

On the other hand, Σ_2^0-independence has particularly interesting properties with respect to provably total functions and computational strength (see the previous section for a definition). An independent Π_2^0-statement will typically add a provably total function: For the Paris–Harrington principle this is the case by [31, theorem 3.2]; the general claim is plausible in view of [17, theorems 2.24 and 4.5] and [41, theorem 5]. In contrast, we have seen that $\mathcal{K}\Sigma_1^-$ does not increase the computational strength of **PA**.

The fact that $\mathcal{K}\Sigma_1^-$ does not add provably total functions is interesting in its own right, but it becomes even more relevant in view of the following: The notion of computational strength is a relatively robust extensional invariant. Bounds on provably total functions can be established without the use of Gödel's theorem, e.g., by induction over cut-free infinite proofs (see [6]). This means that Gödel's theorem is not needed to prove that the Paris–Harrington principle is independent of **PA** (see [7] for an analogous argument with respect to Goodstein's theorem). It appears that no similar invariants are available on the level of Σ^0_2-statements. The only known proof of the fact that **PA** does not prove all instances of $\mathcal{K}\Sigma_1^-$ appeals to Gödel's theorem. In our opinion, this means that $\mathcal{K}\Sigma_1^-$ is a conceptually different and foundationally significant manifestation of mathematical independence.

§3. Analyzing the computational strength. In this section we give a detailed proof of the claim that $\mathcal{K}\Sigma_1^-$ does not increase the computational strength of **PA**. As preparation, we need to show that all instances of $\mathcal{K}\Sigma_1^-$ are true. In the following remark we argue in a strong meta theory; this will later be superseded by a proof in $\mathbf{PA} + \mathcal{TI}(\varepsilon_0, \Pi_1^-)$ (see Proposition 7.2).

Remark 3.1. As a consequence of Kruskal's theorem [28], the partial order $(\mathcal{B}, \leq_{\mathcal{B}})$ does not contain any infinite bad sequence. We will use this fact to justify an arbitrary instance

$$\mathcal{K}\varphi \equiv \exists^{\text{fin}}_{a\subseteq\mathcal{B}}(\forall_{s\in a}\varphi(s) \wedge \forall_{t\in\mathcal{B}}(\varphi(t) \to \exists_{s\in a} s \leq_{\mathcal{B}} t))$$

of the axiom scheme $\mathcal{K}\Sigma_1^-$. Aiming at a contradiction, assume that $\mathcal{K}\varphi$ is false. By a bad φ-sequence we mean a bad sequence $t_0, t_1, \ldots \subseteq \mathcal{B}$ such that $\varphi(t_i)$ holds for each i. Note that the empty sequence is a bad φ-sequence. Furthermore, each bad φ-sequence $t_0, \ldots, t_{n-1}$ can be extended into a bad φ-sequence $t_0, \ldots, t_{n-1}, t_n$. To see that this is the case, consider $a := \{t_0, \ldots, t_{n-1}\}$. As $\forall_{s\in a}\varphi(s)$ holds, the assumption that $\mathcal{K}\varphi$ is false yields an element $t_n \in \mathcal{B}$ with $\varphi(t_n)$ and $\forall_{s\in a} s \not\leq_{\mathcal{B}} t_n$. The latter ensures that $t_0, \ldots, t_{n-1}, t_n$ is still bad. By dependent choice we now get an infinite bad φ-sequence, which contradicts Kruskal's theorem.

The following result is folklore, but we provide details in order to make the paper as accessible as possible.

Proposition 3.2. *If Ψ is a set of true Σ^0_2-sentences, then the provably total functions of* $\mathbf{PA} + \Psi$ *and of* **PA** *coincide. In particular, this applies to $\Psi = \mathcal{K}\Sigma_1^-$.*

Proof. Consider a provably total function $f : \mathbb{N} \to \mathbb{N}$ of the theory $\mathbf{PA} + \Psi$. For some Σ^0_1-definition $\theta(x, y)$ of the graph of f, there are sentences $\psi_0, \ldots, \psi_{n-1} \in \Psi$ such that we have

$$\mathbf{PA} + \{\psi_0, \ldots, \psi_{n-1}\} \vdash \forall_x \exists!_y\, \theta(x, y).$$

To show that f is a provably total function of **PA**, we will define the graph of f by a modified Σ^0_1-formula $\theta'(x, y)$ such that **PA** alone proves $\forall_x \exists!_y\, \theta'(x, y)$. For this purpose we observe that the conjunction $\psi_0 \wedge \cdots \wedge \psi_{n-1}$ is equivalent to a true Σ^0_2-sentence $\exists_m \psi(m)$. Pick a number $n \in \mathbb{N}$ such that the Π^0_1-sentence $\psi(\overline{n})$ is true. Then write

$$\exists_z \theta_0(x, y, z) \equiv \psi(\overline{n}) \to \theta(x, y)$$

for a Δ^0_0-formula θ_0. Since $\psi(\overline{n})$ is true and implies each sentence ψ_i, we do have

$$f(k) = m \quad \Leftrightarrow \quad \mathbb{N} \vDash \exists_z \theta_0(\overline{k}, \overline{m}, z),$$
$$\mathbf{PA} \vdash \forall_x \exists_y \exists_z \theta_0(x, y, z).$$

However, if **PA** does not prove $\psi(\overline{n})$, then it will not prove that the value y is unique. It is well known that one can restore uniqueness by minimizing over the code of the pair $\langle y, z\rangle$. Note that minimizing over y alone would lead out of the Σ^0_1-formulas: the minimal y that satisfies $\exists_z \theta_0(x, y, z)$ is specified by a Δ^0_2-formula. To provide details we write $w = \langle y, z\rangle$ for a Δ^0_1-definition of Cantor's pairing function; recall that $w = \langle y, z\rangle$ implies $y, z \leq w$. Let $\theta'(x, y)$ be the Σ^0_1-formula

$$\exists_w (\exists_{z \leq w}(w = \langle y, z\rangle \wedge \theta_0(x, y, z)) \wedge \forall_{w' < w} \forall_{y', z' \leq w'}(w' = \langle y', z'\rangle \rightarrow \neg\theta_0(x, y', z'))).$$

It is straightforward to see that θ' defines f and that **PA** proves $\forall_x \exists!\, y\theta'(x, y)$. $\square$

§4. From the finite basis property to transfinite induction. In this section we show that $\mathbf{PA} + \mathcal{K}\Sigma^-_1$ proves each instance of $\mathcal{TI}(\varepsilon_0, \Pi^-_1)$. As we will see, it follows that $\mathbf{PA} + \mathcal{K}\Sigma^-_1$ is a proper extension of **PA**. The result of this section is a relatively straightforward consequence of the existing literature. We provide details in order to demonstrate that the argument works out with respect to formula complexity and the role of parameters.

Let us first recall the usual notation system for ordinals below ε_0. According to Cantor's normal form theorem, any ordinal α can be uniquely written as

$$\alpha = \omega^{\alpha_0} + \cdots + \omega^{\alpha_{n-1}} \quad \text{with} \quad \alpha \succeq \alpha_0 \succeq \cdots \succeq \alpha_{n-1},$$

with $\alpha = 0$ for $n = 0$. For $\alpha \prec \varepsilon_0 = \min\{\gamma \mid \omega^\gamma = \gamma\}$ we have $\alpha_0 \prec \alpha$. Recursively, this yields finite terms that represent all ordinals below ε_0 and, simultaneously, a definition of the order on the level of terms. Working in **PA**, one can develop basic ordinal arithmetic in our term system (see, e.g., [32, 43]). In the following we always refer to term representations rather than actual ordinals. The set of and the order between terms will also be denoted by ε_0 and $\prec$, respectively (in addition we write $\alpha \prec \varepsilon_0$ to express that α is one of our terms).

In the introduction we have defined a set $\mathcal{B}$ of binary trees and an embeddability relation $\leq_{\mathcal{B}}$. To establish a connection with the ordinals below ε_0, it is convenient to have a binary normal form: If $\alpha \succ 0$ has Cantor normal form as above, we write

$$\alpha =_{\mathrm{NF}} \omega^\beta + \gamma \qquad \text{for } \beta = \alpha_0 \text{ and } \gamma = \omega^{\alpha_1} + \cdots + \omega^{\alpha_{n-1}}.$$

Note that β and γ can be seen as proper subterms of α. The following construction is well-known (cf. [46, sec. 12]).

Definition 4.1 (**PA**). We construct a function $f : \varepsilon_0 \to \mathcal{B}$ by setting

$$f(\alpha) = \begin{cases} \circ & \text{if } \alpha = 0, \\ \circ(f(\beta), f(\gamma)) & \text{if } \alpha =_{\mathrm{NF}} \omega^\beta + \gamma, \end{cases}$$

which amounts to a recursion over term representations of ordinals.

Concerning the formalization in **PA**, we note that f is primitive recursive. Hence f is **PA**-provably total. In particular, the graph of f is Δ^0_1-definable in **PA**. The following folklore result shows that f satisfies the definition of a quasi embedding.

LEMMA 4.2 (**PA**). *For $\alpha, \beta \prec \varepsilon_0$, the inequality $f(\alpha) \leq_{\mathcal{B}} f(\beta)$ implies $\alpha \preceq \beta$.*

Proof. Define a height function $h : \varepsilon_0 \to \mathbb{N}$ by recursion over terms, setting

$$h(\alpha) = \begin{cases} 0 & \text{if } \alpha = 0, \\ \max\{h(\gamma), h(\delta)\} + 1 & \text{if } \alpha =_{\mathrm{NF}} \omega^\gamma + \delta. \end{cases}$$

The claim from the lemma can now be verified by induction over $h(\beta)$. For $\alpha = 0$ the implication holds because $\alpha \preceq \beta$ is true. In the remaining case we may write $\alpha =_{\mathrm{NF}} \omega^\gamma + \delta$. By the definition of $\leq_{\mathcal{B}}$, the inequality $f(\alpha) = \circ(f(\gamma), f(\delta)) \leq_{\mathcal{B}} f(\beta)$ fails for $f(\beta) = \circ$. Hence we may also assume $\beta \succ 0$, say $\beta =_{\mathrm{NF}} \omega^{\gamma'} + \delta'$. Again by the definition of $\leq_{\mathcal{B}}$, the inequality

$$f(\alpha) = \circ(f(\gamma), f(\delta)) \leq_{\mathcal{B}} \circ(f(\gamma'), f(\delta')) = f(\beta)$$

can hold for two reasons: First assume we have $f(\gamma) \leq_{\mathcal{B}} f(\gamma')$ and $f(\delta) \leq_{\mathcal{B}} f(\delta')$. In view of $h(\gamma'), h(\delta') < h(\beta)$, the induction hypothesis yields $\gamma \preceq \gamma'$ and $\delta \preceq \delta'$. By basic ordinal arithmetic we get

$$\alpha = \omega^\gamma + \delta \preceq \omega^{\gamma'} + \delta' = \beta.$$

Now assume $f(\alpha) \leq_{\mathcal{B}} f(\beta)$ holds because we have $f(\alpha) \leq_{\mathcal{B}} f(\gamma')$ or $f(\alpha) \leq_{\mathcal{B}} f(\delta')$. Inductively we get $\alpha \preceq \gamma' \preceq \omega^{\gamma'}$ or $\alpha \preceq \delta'$. Either way we have $\alpha \preceq \omega^{\gamma'} + \delta' = \beta$. □

In addition to the lemma itself, we will need the following standard consequence:

COROLLARY 4.3 (**PA**). *The function $f : \varepsilon_0 \to \mathcal{B}$ is injective.*

Proof. Consider $\alpha, \beta \prec \varepsilon_0$ with $f(\alpha) = f(\beta)$. A straightforward induction over $\mathcal{B}$ shows that $\leq_{\mathcal{B}}$ is reflexive. Hence we have $f(\alpha) \leq_{\mathcal{B}} f(\beta)$ and $f(\beta) \leq_{\mathcal{B}} f(\alpha)$. By the previous lemma this implies $\alpha \preceq \beta$ and $\beta \preceq \alpha$. Since the order relation on the ordinals is antisymmetric, we obtain $\alpha = \beta$. □

We can now show that the finite basis property implies transfinite induction. The converse implication will be established in Section 7.

PROPOSITION 4.4. *Each instance of $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$ can be proved in $\mathbf{PA} + \mathcal{K}\Sigma_1^-$.*

Proof. Working in $\mathbf{PA} + \mathcal{K}\Sigma_1^-$, we establish $\mathcal{TI}(\varepsilon_0, \psi)$ for a given Π_1^0-formula ψ with a single free variable. For this purpose we consider the formula

$$\varphi(t) :\equiv t \in \mathcal{B} \wedge \exists_{\alpha \prec \varepsilon_0}(f(\alpha) = t \wedge \neg\psi(\alpha)),$$

where $f : \varepsilon_0 \to \mathcal{B}$ is the function from Definition 4.1. Since the graph of f is Δ_1^0-definable in **PA**, we see that $\varphi(t)$ is (provably equivalent to) a Σ_1^0-formula with the single free variable t. Hence we may use $\mathcal{K}\varphi$ to get (a code for) a finite set $a \subseteq \mathcal{B}$ that satisfies

$$\forall_{s \in a}\varphi(s) \wedge \forall_{t \in \mathcal{B}}(\varphi(t) \to \exists_{s \in a} s \leq_{\mathcal{B}} t).$$

First assume that a is empty. Then $\exists_{s \in a} s \leq_{\mathcal{B}} t$ fails for all $t \in \mathcal{B}$, so that the second conjunct enforces $\forall_{t \in \mathcal{B}}\neg\varphi(t)$. Given $\alpha \prec \varepsilon_0$, it is straightforward to see that $\neg\varphi(t)$ for $t := f(\alpha) \in \mathcal{B}$ implies $\psi(\alpha)$. We thus have $\forall_{\alpha \prec \varepsilon_0}\psi(\alpha)$, which is the conclusion of $\mathcal{TI}(\varepsilon_0, \psi)$. Now assume that the finite set $a \subseteq \mathcal{B}$ is non-empty. Due to $\forall_{s \in a}\varphi(s)$, we see that a is contained in the range of f. Also recall that f is injective. By induction on the cardinality of a, one can infer that there is an ordinal $\gamma \prec \varepsilon_0$ with

$$f(\gamma) \in a \wedge \forall_{\delta \prec \gamma} f(\delta) \notin a.$$

Given an ordinal γ with this property, we now establish

$$\forall_{\beta \prec \gamma} \psi(\beta) \wedge \neg\psi(\gamma),$$

which implies that $\mathcal{TI}(\varepsilon_0, \psi)$ holds because its antecedent $\mathrm{Prog}_{\varepsilon_0}(\psi)$ fails. Aiming at the first conjunct, we consider an ordinal $\beta \prec \gamma$. If $\psi(\beta)$ was false, then $\varphi(t)$ would hold for $t := f(\beta) \in \mathcal{B}$. Since $a \subseteq \mathcal{B}$ witnesses the conclusion of $\mathcal{K}\varphi$, we would get an element $s \in a$ with $s \leq_{\mathcal{B}} t$. Writing $s = f(\delta)$ with $\delta \prec \varepsilon_0$, we could invoke Lemma 4.2 to conclude $\delta \preceq \beta \prec \gamma$. By the above this would imply $s = f(\delta) \notin a$, which yields the desired contradiction. To establish the second conjunct we observe that $f(\gamma) \in a$ implies $\varphi(f(\gamma))$. According to the definition of φ, this means that there is an ordinal $\alpha \prec \varepsilon_0$ with $f(\alpha) = f(\gamma)$ and $\neg\psi(\alpha)$. Since f is injective we get $\alpha = \gamma$ and thus $\neg\psi(\gamma)$, as required. □

According to Gentzen's ordinal analysis [18], the consistency of Peano arithmetic is provable in $\mathbf{PA} + \mathcal{TI}(\varepsilon_0, \Pi_1^-)$. A detailed proof of a stronger result can be found in the next section. Together with Proposition 4.4 and Gödel's theorem, it follows that $\mathbf{PA} + \mathcal{K}\Sigma_1^-$ is a proper extension of $\mathbf{PA}$.

§5. From transfinite induction to reflection. Working over **PA**, we show that the parameter-free induction scheme $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$ implies the local reflection principle $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0)$. The converse direction will be established in Section 7. The result is rather similar to one by Kreisel and Lévy [27], who show that induction with parameters corresponds to uniform reflection. As we will see, the connection with reflection implies that $\mathbf{PA} + \mathcal{K}\Sigma_1^-$ is not contained in any consistent extension of **PA** by a computably enumerable set of Π_2^0-sentences.

As preparation, we review the ordinal analysis of Peano arithmetic and its formalization in **PA** itself. First note that we cannot formalize the usual soundness argument by induction over formal proofs, since there is no arithmetical truth definition that would cover all relevant formulas (due to Tarski [47]). Even when we restrict attention to theorems of restricted complexity, their proofs may involve detours through more complex lemmata. The method of cut elimination aims to remove such detours in order to permit a soundness argument that is based on partial truth definitions (cf. [20, sec. I.1(d)]). However, it is not immediately possible to eliminate complex lemmata from proofs in Peano arithmetic, which may use complex instances of induction in an essential way. To resolve this problem, ordinal analysis transforms the usual finite proofs into infinite proof trees: In the realm of infinite proofs, induction can be deduced from axioms of low complexity, so that cut elimination becomes possible. Soundness can then be proved by transfinite induction over the rank of infinite proof trees.

Our ordinal analysis works with proofs in a Tait-style sequence calculus. In particular, this means that all formulas are in negation normal form, and that negation is a defined operation based on De Morgan's laws. Each node in a proof tree deduces a sequent, i.e., a finite set $\Gamma = \{\varphi_0, \dots, \varphi_{n-1}\}$ of formulas. The latter is to be interpreted as the disjunction $\bigvee \Gamma = \varphi_0 \vee \cdots \vee \varphi_{n-1}$. In the context of sequents we write Γ, φ for $\Gamma \cup \{\varphi\}$. Detours in proofs are implemented via the cut rule

$$\frac{\Gamma, \varphi \qquad \Gamma, \neg\varphi}{\Gamma},$$

which has the following intuitive significance: In order to show $\bigvee \Gamma$, it suffices to

https://doi.org/10.1017/S1755020321000265 Published online by Cambridge University Press

- prove a lemma φ (more precisely, the left premise proves $\bigvee\Gamma\vee\varphi$) and to
- prove that φ implies $\bigvee\Gamma$ (i.e., to prove $\bigvee\Gamma\vee\neg\varphi$, as in the right premise).

The crucial feature of the infinite proof system is the ω-rule

$$\frac{\Gamma,\varphi(0)\qquad\Gamma,\varphi(1)\qquad\cdots}{\Gamma,\forall_n\varphi(n)},$$

which allows to infer $\forall_n\varphi(n)$ if there is a proof of $\varphi(n)$ for each numeral n. Induction can be derived from the ω-rule, since

$$\varphi(0)\wedge\forall_m(\varphi(m)\to\varphi(m+1))\to\varphi(n)$$

has a straightforward proof for each number n. It follows that any finite proof in Peano arithmetic can be translated (or "embedded") into the infinite system.

It is not immediately clear how infinite proof trees can be formalized in Peano arithmetic. In the following we recall a very elegant approach due to Buchholz [5] (see his paper for all missing details): The idea is to work with a set $\mathbf{Z}^*$ of finite terms. Each term names an infinite proof by specifying its role in the cut elimination process. Specifically, each finite proof d in Peano arithmetic gives rise to a constant symbol $[d]\in\mathbf{Z}^*$, which denotes the translation of d into the infinite system. For each term $h\in\mathbf{Z}^*$ there is a term $Eh\in\mathbf{Z}^*$ that names the proof that results from h by a single application of cut elimination. The intermediate steps of cut elimination give rise to auxiliary function symbols. By primitive recursion over terms one can define an ordinal $\mathfrak{o}(h)\prec\varepsilon_0$ that bounds the rank of the proof tree represented by h; for example, the well-known fact that cut elimination leads to an exponential increase of the ordinal rank suggests the recursive clause $\mathfrak{o}(Eh)=\omega^{\mathfrak{o}(h)}$. Also by recursion over terms, one can determine the end sequent $\mathfrak{e}(h)$, the last rule $\mathfrak{r}(h)$, the cut rank $\mathfrak{d}(h)$, and terms $\mathfrak{s}(h,n)\in\mathbf{Z}^*$ that denote the immediate subtrees of the proof tree that is represented by h. Working in **PA** (or even in **PRA**), one can show that the term system $\mathbf{Z}^*$ is "locally correct" (see [5, theorem 3.8]); in particular this means that we have $\mathfrak{o}(\mathfrak{s}(h,n))\prec\mathfrak{o}(h)$, except when $\mathfrak{r}(h)$ signifies an axiom. To ensure "global correctness," one needs transfinite induction up to ε_0, which is not available in **PA**. In the sequel we abbreviate

$$h\vdash^\alpha_0\Gamma\quad:\Leftrightarrow\quad h\in\mathbf{Z}^*\wedge\mathfrak{o}(h)=\alpha\wedge\mathfrak{d}(h)=0\wedge\mathfrak{e}(h)\subseteq\Gamma.$$

Intuitively, this asserts that h is a cut-free infinite proof tree with rank α and end sequent Γ (note that $\bigvee\mathfrak{e}(h)$ implies $\bigvee\Gamma$). Crucially, the relation $h\vdash^\alpha_0\Gamma$ is primitive recursive and hence Δ^0_1-definable in **PA**. This implies that

$$\mathbf{Z}^*\vdash^\alpha_0\Gamma\quad:\Leftrightarrow\quad\exists_{h\in\mathbf{Z}^*}h\vdash^\alpha_0\Gamma$$

is a Σ^0_1-formula with parameters α and Γ. We can now show the promised result:

PROPOSITION 5.1. *Each instance of* $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$ *can be proved in* $\mathbf{PA}+\mathcal{TI}(\varepsilon_0,\Pi^-_1)$.

Proof. Consider a closed Σ^0_2-formula φ. Working in $\mathbf{PA}+\mathcal{TI}(\varepsilon_0,\Pi^-_1)$, we assume that we have $\mathrm{Pr}_{\mathbf{PA}}(\varphi)$. In order to establish $\mathrm{Rfn}_{\mathbf{PA}}(\varphi)$, we need to derive φ. We use Buchholz' formalization of ordinal analysis, as discussed above. By embedding and cut elimination (cf. [5, definitions 3.4 and 3.7]), the assumption $\mathrm{Pr}_{\mathbf{PA}}(\varphi)$ implies

$$\exists_{\alpha\prec\varepsilon_0}\mathbf{Z}^*\vdash^\alpha_0\{\varphi\}.$$

Write $\Gamma \subseteq \{\varphi\} \cup \Pi_1^-$ to express that Γ is a sequent that consists of Π_1^0-sentences and (possibly) the formula φ. The statement that Γ contains a true Π_1^0-sentence can be expressed by a Π_1^0-formula $\mathrm{Tr}_{\Pi_1^-}(\Gamma)$ (cf. [20, theorem I.1.75]). Aiming at a contradiction, we assume that φ is false. Under this assumption we will derive

$$\forall_{\alpha \prec \varepsilon_0} \forall_\Gamma (\Gamma \subseteq \{\varphi\} \cup \Pi_1^- \wedge \mathbf{Z}^* \vdash_0^\alpha \Gamma \to \mathrm{Tr}_{\Pi_1^-}(\Gamma)),$$

arguing by transfinite induction on $\alpha \prec \varepsilon_0$. Note that the sentence φ is represented by a fixed numeral. Hence α is the only free variable of the induction formula, and the induction is covered by the scheme $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$. Once the induction is carried out, it is straightforward to derive the desired contradiction: By the above we have $\mathbf{Z}^* \vdash_0^\alpha \{\varphi\}$ for some $\alpha \prec \varepsilon_0$. However, we cannot have $\mathrm{Tr}_{\Pi_1^-}(\{\varphi\})$, since φ was assumed to be false (note that this covers both $\varphi \in \Pi_1^0 \subseteq \Sigma_2^0$ and $\varphi \in \Sigma_2^0 \backslash \Pi_1^0$). It remains to carry out the induction. In the step we consider a sequent $\Gamma \subseteq \{\varphi\} \cup \Pi_1^-$ and assume $h \vdash_0^\alpha \Gamma$ for some $h \in \mathbf{Z}^*$. We distinguish cases according to the last rule $\mathfrak{r}(h)$. Note that this cannot be a cut, since $h \vdash_0^\alpha \Gamma$ entails $\mathfrak{d}(h) = 0$. If $\mathfrak{r}(h)$ is an axiom, then $\mathfrak{e}(h) \subseteq \Gamma$ contains a true literal (cf. [5, definition 2.2]). To complete the proof, we consider the introduction of a quantifier; the introduction of a propositional connective is similar and simpler. First assume that h ends with an ω-rule, which introduces a formula $\forall_n \theta(n) \in \Gamma$. Due to $\Gamma \subseteq \{\varphi\} \cup \Pi_1^-$ we see that $\forall_n \theta(n)$ must be a Π_1^0-sentence. Local correctness (see [5, theorem 3.8]) yields

$$\mathbf{Z}^* \vdash_0^{\mathfrak{o}(\mathfrak{s}(h,n))} \Gamma, \theta(n) \quad \text{with} \quad \mathfrak{o}(\mathfrak{s}(h,n)) \prec \mathfrak{o}(h) = \alpha,$$

for all $n \in \mathbb{N}$. The induction hypothesis implies that each sequent $\Gamma, \theta(n)$ contains a true Π_1^0-sentence. Hence we get such a sentence in Γ, or all instances $\theta(n)$ are true. In the latter case, it follows that Γ contains the true Π_1^0-sentence $\forall_n \theta(n)$. Finally, assume that $\mathfrak{r}(h)$ introduces an existential formula $\exists_n \psi(n)$. In view of $\Gamma \subseteq \{\varphi\} \cup \Pi_1^-$ we must have $\exists_n \psi(n) \equiv \varphi$ (note that [5] does not work with bounded quantifiers but treats primitive recursive relations as atomic). By local correctness there is some existential witness $k \in \mathbb{N}$ such that we have

$$\mathbf{Z}^* \vdash_0^{\mathfrak{o}(\mathfrak{s}(h,0))} \Gamma, \psi(k) \quad \text{with} \quad \mathfrak{o}(\mathfrak{s}(h,0)) \prec \mathfrak{o}(h) = \alpha.$$

The induction hypothesis yields a true Π_1^0-sentence in $\Gamma, \psi(k)$. To establish $\mathrm{Tr}_{\Pi_1^-}(\Gamma)$ it suffices to show that $\psi(k)$ cannot be true: if it was, then $\varphi \equiv \exists_n \psi(n)$ would be true as well, which contradicts our assumption. □

As mentioned in the introduction, the previous proposition can also be derived from work of Beklemishev:

REMARK 5.2. By the paragraph before Proposition 5.18 in [3], the consistency of $\mathbf{PA} + \mathrm{Con}(\mathbf{PA})$ can be derived by a single application of the (parameter-free) induction rule up to ε_0 over $\mathbf{EA}^+ + \mathrm{Con}(\mathbf{PA})$, where $\mathbf{EA}^+$ denotes the extension of elementary arithmetic by superexponentiation. As pointed out by one of the referees, the argument remains valid when the statement $\mathrm{Con}(\mathbf{PA})$ is replaced by an arbitrary Π_2^0-sentence ψ. In particular, this yields

$$\mathbf{PA} + \mathcal{TI}(\varepsilon_0, \Pi_1^-) \vdash \psi \to \mathrm{Con}(\mathbf{PA} + \psi).$$

If we take $\psi = \neg\varphi$, then the right side is the contrapositive of local reflection for φ.

The following is folklore (cf. the related result by Kreisel and Lévy [27, sec. 8]; see also [1, lemma 2] for an argument that takes the formula complexity into account).

Proposition 5.3. *There is no computably enumerable set* Ψ *of* Π^0_2*-sentences such that* $\mathbf{PA}+\Psi$ *is consistent and contains* $\mathbf{PA}+\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$.

Proof. Consider a computably enumerable set Ψ of Π^0_2-sentences such that $\mathbf{PA}+\Psi$ proves each instance of $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$. We need to show that $\mathbf{PA}+\Psi$ is inconsistent. According to [29, theorem 4], there is a single Π^0_2-sentence ψ such that $\mathbf{PA}+\psi$ is a Σ^0_2-conservative extension of $\mathbf{PA}+\Psi$. In view of conservativity, it suffices to establish the inconsistency of $\mathbf{PA}+\psi$. We have

$$\mathbf{PA}+\psi \vdash \mathrm{Pr}_{\mathbf{PA}}(\neg\psi)\to\neg\psi,$$

since the right side is an instance of $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$. Considering the contrapositive, we learn that $\mathbf{PA}+\psi$ proves its own consistency statement $\neg\,\mathrm{Pr}_{\mathbf{PA}}(\neg\psi)$, so that it is inconsistent by Gödel's theorem. □

By Propositions 4.4 and 5.1 we have

$$\mathbf{PA}+\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)\ \subseteq\ \mathbf{PA}+\mathcal{TI}(\varepsilon_0,\Pi^-_1)\ \subseteq\ \mathbf{PA}+\mathcal{K}\Sigma^-_1.$$

Hence the previous proposition has the following consequence:

Corollary 5.4. *There is no computably enumerable set* Ψ *of* Π^0_2*-sentences such that* $\mathbf{PA}+\Psi$ *is consistent and contains* $\mathbf{PA}+\mathcal{K}\Sigma^-_1$*. In particular, the latter is a proper extension of* $\mathbf{PA}$.

Since any true Σ^0_2-sentence follows from a true Π^0_1-sentence, there is a set Ξ of Π^0_1-sentences such that $\mathbf{PA}+\Xi$ is consistent and contains $\mathbf{PA}+\mathcal{K}\Sigma^-_1$. The corollary tells us that Ξ cannot be computably enumerable.

§6. A primitive recursive reification. In the rest of this paper we complete the proof that $\mathcal{K}\Sigma^-_1$, $\mathcal{TI}(\varepsilon_0,\Pi^-_1)$ and $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$ are equivalent over $\mathbf{PA}$. The present section is concerned with a technical result that will be crucial for this purpose.

Write $\mathrm{Bad}^-(\mathcal{B})$ for the set of non-empty finite bad sequences in $\mathcal{B}$. We want to construct a primitive recursive function $r:\mathrm{Bad}^-(\mathcal{B})\to\varepsilon_0$ such that we have

$$r(\langle t_0,\dots,t_n,t_{n+1}\rangle)\prec r(\langle t_0,\dots,t_n\rangle)$$

whenever $\langle t_0,\dots,t_{n+1}\rangle$ is an element of $\mathrm{Bad}^-(\mathcal{B})$, provably in $\mathbf{PA}$ (in fact in primitive recursive arithmetic). Such a function is called a reification. It ensures that $\mathcal{B}$ is a well partial order with maximal order type at most (and in fact equal to) ε_0.

As mentioned in the introduction, the result that $\mathcal{B}$ has maximal order type ε_0 is due to de Jongh and Schmidt. Experience shows that maximal order types can be witnessed by effective reifications. For the case of finite (and in particular binary) trees this has been established by M. Rathjen and A. Weiermann [33, sec. 2]. Unfortunately, we cannot simply cite their result: In [33] it is shown that $\mathbf{ACA_0}$ proves the existence of a reification; however, it is not entirely trivial to see that the constructed reification is (primitive) recursive. In the rest of this section we verify this fact in detail. Some readers may prefer to skip this verification and to continue with the applications in the next section. We point out that the following presentation is influenced by the more general construction in [21].

The reification of $\mathcal{B}$ will depend on reifications of various other orders. In the context of first order arithmetic it helps to think of these orders as types, which are represented by finite expressions.

Definition 6.1 (**PA**). The following recursive clauses generate a collection of types and a subcollection of indecomposable types:

(i) The symbols $\mathfrak{B}$ and $\mathfrak{E}$ are indecomposable types.
(ii) If A, B are types, then $A + B$ is a type.
(iii) If A, B are indecomposable types, then $A \times B$ is an indecomposable type.
(iv) If A is any type, then A^* is an indecomposable type.

On an intuitive level, one should think of $\mathfrak{B}$ as the collection $\mathcal{B}$ of binary trees and of $\mathfrak{E}$ as the empty order. The expressions $A + B$ and $A \times B$ refer to the usual notions of sum (disjoint union) and product, while A^* stands for the set of finite sequences with entries from A (with the order from Higman's lemma). Officially, the elements of these orders are represented by the terms from Definition 6.2 below, while the order relations are recovered via Definition 6.3.

Note that it is not allowed to form types such as $(A + B) \times C$, since $A + B$ is not indecomposable. This restriction corresponds to the notion of normal form from [21, definition 4.15]. We will eventually assign an (additively indecomposable) ordinal to each (indecomposable) type that does not contain $\mathfrak{B}$. In particular, the ordinals assigned to product types will be additively indecomposable. This ensures that the ordinal assignment satisfies the monotonicity property from Proposition 6.12 below (cf. the role of indecomposable ordinals in the proof of this proposition, as well as the counterexample in the paragraph after Proposition 5.10 in [21]).

As mentioned above, the elements of our orders are represented by terms of the corresponding types. To obtain primitive recursive constructions, it is crucial to work with terms of all types simultaneously. For example, it is neither possible nor necessary to construct all terms of type A before one constructs a term of type A^*. Let us point out that each term has a unique type, which can be read off by a primitive recursive function. Also recall that $\mathfrak{E}$ represents the empty order, so that no terms of this type are specified.

Definition 6.2 (**PA**). The following recursive clauses generate a collection of terms. We simultaneously specify the types of these terms:

(i) Each binary tree $t \in \mathcal{B}$ is a term of type $\mathfrak{B}$.
(ii) If a is a term of type A and B is a type, then $\iota_0^B a$ is a term of type $A + B$. If b is a term of type B and A is a type, then $\iota_1^A b$ is a term of type $A + B$.
(iii) If a and b are terms of types A and B, then $\langle a, b\rangle$ is a term of type $A \times B$.
(iv) If $a_0, \dots, a_{n-1}$ have type A, then $\langle a_0, \dots, a_{n-1}\rangle_A$ is a term of type A^*.

Note that (iii) does only apply when A and B are indecomposable.

One readily constructs a Gödel numbering # with the monotonicity properties

$$\#s, \#t < \#\circ(s,t) \text{ for } s, t \in \mathfrak{B}, \qquad \#a < \#\iota_0^B a, \qquad \#b < \#\iota_1^A b,$$
$$\#a, \#b < \#\langle a, b\rangle, \qquad \#a_0, \#\langle a_1, \dots, a_n\rangle_A < \#\langle a_0, \dots, a_n\rangle_A.$$

We will use this Gödel numbering to construct primitive recursive functions by course-of-values recursion. Binary functions can be constructed with the help of the Cantor

https://doi.org/10.1017/S1755020321000265 Published online by Cambridge University Press

pairing function, which is monotone in both components. For example, the following definition decides $a \leq_A a'$ by recursion over the code of $\langle \# a, \# a' \rangle$.

DEFINITION 6.3 (**PA**). The relation $a \leq_A a'$ between terms a and a' of the same type A is generated by the following recursive clauses (i.e., it is the smallest relation that satisfies them):

(i) If $s \leq_{\mathcal{B}} t$, then $s \leq_{\mathfrak{B}} t$.
(ii) If $a \leq_A a'$, then $\iota_0^B a \leq_{A+B} \iota_0^B a'$. If $b \leq_B b'$, then $\iota_1^A b \leq_{A+B} \iota_1^A b'$.
(iii) If $a \leq_A a'$ and $b \leq_B b'$, then $\langle a, b \rangle \leq_{A \times B} \langle a', b' \rangle$.
(iv) If there is a strictly increasing $f : \{0, \dots, m-1\} \to \{0, \dots, n-1\}$ such that $a_i \leq_A a'_{f(i)}$ holds for all $i < m$, then $\langle a_0, \dots, a_{m-1} \rangle_A \leq_{A^*} \langle a'_0, \dots, a'_{n-1} \rangle_A$.

Let us record the expected property:

LEMMA 6.4 (**PA**). *Each relation $\leq_A$ is a partial order on the terms of type A.*

Proof. First check $a \leq_A a$ by induction over $\# a$, simultaneously for all types A. Then use induction over $\# a + \# a'$ to verify that $a \leq_A a'$ and $a' \leq_A a$ imply $a = a'$. Finally, show $a \leq_A a' \,\&\, a' \leq_A a'' \Rightarrow a \leq_A a''$ by induction over $\# a + \# a' + \# a''$. $\square$

From now on we write $a \in A$ to express that a is a term of type A. Despite this notation, one should keep in mind that A is a finite expression rather than an infinite set. The following provides a substitute for the "missing" types $A \times B$.

DEFINITION 6.5 (**PA**). For arbitrary types A and B we recursively define a type $A \otimes B$ and terms $[a, b] \in A \otimes B$ for all $a \in A$ and $b \in B$: First put

$$A \otimes B = A \times B \quad \text{and} \quad [a, b] = \langle a, b \rangle \quad \text{when } A, B \text{ are indecomposable.}$$

Now consider $A = C + D$ and an arbitrary B. To save parentheses, we assume that $\otimes$ binds stronger than $+$. We then define

$$(C + D) \otimes B = C \otimes B + D \otimes B \quad \text{and} \quad [\iota_0^D c, b] = \iota_0^{D \otimes B}[c, b],\ [\iota_1^C d, b] = \iota_1^{C \otimes B}[d, b].$$

For indecomposable A and $B = C + D$ we set

$$A \otimes (C + D) = A \otimes C + A \otimes D \quad \text{and} \quad [a, \iota_0^D c] = \iota_0^{A \otimes D}[a, c],\ [a, \iota_1^C d] = \iota_1^{A \otimes C}[a, d].$$

The following is readily checked by induction on $\# a + \# a' + \# b + \# b'$.

LEMMA 6.6 (**PA**). *We have*

$$[a, b] \leq_{A \otimes B} [a', b'] \qquad \Leftrightarrow \qquad a \leq_A a' \text{ and } b \leq_B b'$$

for arbitrary terms $a, a' \in A$ and $b, b' \in B$.

For $a \in A$ we will abbreviate

$$a' \in A_a \quad :\Leftrightarrow \quad a' \in A \text{ and } a \not\leq_A a'.$$

Recall that, officially, A is not an infinite set but a finite expression that denotes a type. Similarly, $a' \in A_a$ does officially refer to a primitive recursive relation between the finite expressions a, a' and A. Informally, A_a stands for the set of all $a' \in A$ that can follow a in a bad sequence. It is known that these sets play an important role in the analysis of maximal order types. While A_a itself is not a type, it admits a quasi embedding into a type $A(a)$, as we show next. To save parentheses we agree on $A \otimes B \otimes C = (A \otimes B) \otimes C$ and $[a, b, c] = [[a, b], c]$. The following construction is similar to the one in [21, definition 5.3 and example 5.4].

DEFINITION 6.7 (**PA**). By recursion over $\#a$ we define a type $A(a)$ for each $a \in A$:

(i) We have $\mathfrak{B}(\circ) = \mathfrak{E}$ and $\mathfrak{B}(\circ(s, t)) = (\mathfrak{B}(s) + \mathfrak{B}(t))^*$.
(ii) We have $(A + B)(\iota_0^B a) = A(a) + B$ and $(A + B)(\iota_1^A b) = A + B(b)$.
(iii) We have $(A \times B)(\langle a, b\rangle) = A(a) \otimes B + A \otimes B(b)$.
(iv) We have $A^*(\langle\rangle_A) = \mathfrak{E}$ and

$$A^*(\langle a_0, \dots, a_n\rangle_A) = A(a_0)^* + A(a_0)^* \otimes A \otimes A^*(\langle a_1, \dots, a_n\rangle_A).$$

As promised, we get the following quasi embeddings:

PROPOSITION 6.8. *There is a primitive recursive function e such that* **PA** *proves the following*: *For any type A and terms $a \in A, b \in A_a$ we have $e_A(a, b) = e(a, b) \in A(a)$ (note that A can be inferred from a). Furthermore we have*

$$e_A(a, b) \leq_{A(a)} e_A(a, b') \quad \Rightarrow \quad b \leq_A b'$$

for any terms $b, b' \in A_a$.

Proof. The value $e_A(a, b)$ is defined by recursion over the code of the pair $\langle \#a, \#b\rangle$, simultaneously for all types A. Once the construction of e is complete, the second part of the proposition can be verified by induction on $\#a + \#b + \#b'$. In the following we distinguish cases according to the form of a.

First consider $a = \circ \in \mathfrak{B} = A$. Since $\circ \leq_{\mathfrak{B}} t$ is true for any $t \in \mathfrak{B}$, the set A_a is empty and there are no values to define. Now assume $a = \circ(s_0, s_1) \in \mathfrak{B} = A$. For the term $b = \circ \in \mathfrak{B}$ we put

$$e_{\mathfrak{B}}(\circ(s_0, s_1), \circ) = \langle\rangle_{\mathfrak{B}(s_0)+\mathfrak{B}(s_1)} \in (\mathfrak{B}(s_0) + \mathfrak{B}(s_1))^* = \mathfrak{B}(\circ(s_0, s_1)).$$

Now assume that we have $b = \circ(t_0, t_1) \in \mathfrak{B}$. The condition $b \in A_a$ amounts to $\circ(s_0, s_1) \not\leq_{\mathfrak{B}} \circ(t_0, t_1)$, which yields $s_0 \not\leq_{\mathfrak{B}} t_0$ or $s_1 \not\leq_{\mathfrak{B}} t_1$. Let us assume that we have $s_0 \not\leq_{\mathfrak{B}} t_0$, which amounts to $t_0 \in \mathfrak{B}_{s_0}$. We may then refer to the recursively defined value

$$e_{\mathfrak{B}}(s_0, t_0) \in \mathfrak{B}(s_0).$$

More formally, the recursive definition of $e_A(a, b)$ and the inductive verification of $e_A(a, b) \in A(a)$ should be separated. In order to do so, we can agree on a default value for the hypothetical case that the decidable property $e_{\mathfrak{B}}(s_0, t_0) \in \mathfrak{B}(s_0)$ fails; the induction shows that the default value is never called. By $\circ(s_0, s_1) \not\leq_{\mathfrak{B}} \circ(t_0, t_1)$ we also have $\circ(s_0, s_1) \not\leq_{\mathfrak{B}} t_1$, which amounts to $t_1 \in \mathfrak{B}_{\circ(s_0,s_1)}$ and provides

$$e_{\mathfrak{B}}(\circ(s_0, s_1), t_1) \in \mathfrak{B}(\circ(s_0, s_1)) = (\mathfrak{B}(s_0) + \mathfrak{B}(s_1))^*.$$

Let us agree to write $c_0 \star \langle c_1, \dots, c_n\rangle_C := \langle c_0, c_1, \dots, c_n\rangle_C \in C^*$ for terms $c_0, \dots, c_n$ of a type C. We can now state our recursive clause as

$$e_{\mathfrak{B}}(\circ(s_0, s_1), \circ(t_0, t_1)) = \begin{cases} \iota_0^{\mathfrak{B}(s_1)} e_{\mathfrak{B}}(s_0, t_0) \star e_{\mathfrak{B}}(\circ(s_0, s_1), t_1) & \text{if } s_0 \not\leq_{\mathfrak{B}} t_0, \\ \iota_1^{\mathfrak{B}(s_0)} e_{\mathfrak{B}}(s_1, t_1) \star e_{\mathfrak{B}}(\circ(s_0, s_1), t_0) & \text{otherwise.} \end{cases}$$

To explain the second case we recall that $s_1 \not\leq_{\mathfrak{B}} t_1$ must hold if $s_0 \not\leq_{\mathfrak{B}} t_0$ fails.

Before we state the other recursive clauses, let us verify that the second part of the proposition holds for $A = \mathfrak{B}$. As above we write $a = \circ(s_0, s_1)$. In the case of the term $b' = \circ$ we observe

$$e_{\mathfrak{B}}(a,b) \leq_{\mathfrak{B}(s)} e_{\mathfrak{B}}(a,b) = \langle\rangle_{\mathfrak{B}(s_0)+\mathfrak{B}(s_1)} \quad\Rightarrow\quad e_{\mathfrak{B}}(a,b) = \langle\rangle_{\mathfrak{B}(s_0)+\mathfrak{B}(s_1)}.$$

The consequent of this implication can only hold for $b = \circ$. In this case $b \leq_{\mathfrak{B}} b'$ is satisfied for any $b' \in \mathfrak{B}$. Hence it remains to consider terms of the form $b = \circ(t_0, t_1)$ and $b' = \circ(t_0', t_1')$. In general we have

$$c \star \sigma \leq_{C^*} c' \star \sigma' \quad\Leftrightarrow\quad c \star \sigma \leq_{C^*} \sigma' \text{ or } (c \leq_C c' \text{ and } \sigma \leq_{C^*} \sigma').$$

First assume that $e_{\mathfrak{B}}(s,b) \leq_{\mathfrak{B}(s)} e_{\mathfrak{B}}(s,b')$ holds because of $e_{\mathfrak{B}}(s,b) \leq_{\mathfrak{B}(s)} e_{\mathfrak{B}}(s,t_i')$. Then the induction hypothesis yields $b \leq_{\mathfrak{B}} t_i'$, which implies $b \leq_{\mathfrak{B}} \circ(t_0', t_1') = b'$. Now assume we have $e_{\mathfrak{B}}(s,b) \leq_{\mathfrak{B}(s)} e_{\mathfrak{B}}(s,b')$ because there are $i, j \in \{0,1\}$ with

$$\iota_i^{\mathfrak{B}(s_{1-i})} e_{\mathfrak{B}}(s_i, t_i) \leq_{\mathfrak{B}(s_0)+\mathfrak{B}(s_1)} \iota_j^{\mathfrak{B}(s_{1-j})} e_{\mathfrak{B}}(s_j, t_j'),$$

$$e_{\mathfrak{B}}(s, t_{1-i}) \leq_{\mathfrak{B}(s)} e_{\mathfrak{B}}(s, t_{1-j}').$$

The first inequality can only hold for $i = j$. It yields $e_{\mathfrak{B}}(s_i, t_i) \leq_{\mathfrak{B}(s_i)} e_{\mathfrak{B}}(s_i, t_i')$, which implies $t_i \leq_{\mathfrak{B}} t_i'$ by induction hypothesis. From the second inequality we can infer $t_{1-i} \leq_{\mathfrak{B}} t_{1-i}'$. Together we get $b = \circ(t_0, t_1) \leq_{\mathfrak{B}} \circ(t_0', t_1') = b'$, as desired.

Sum and product types are considerably easier to handle. We only state the recursive clauses and leave all verifications to the reader:

$$e_{A+B}(\iota_0^B a, \iota_0^B a') = \iota_0^B e_A(a, a'), \qquad e_{A+B}(\iota_0^B a, \iota_1^A b') = \iota_1^{A(a)} b',$$

$$e_{A+B}(\iota_1^A b, \iota_0^B a') = \iota_0^{B(b)} a', \qquad e_{A+B}(\iota_1^A b, \iota_1^A b') = \iota_1^A e_B(b, b').$$

$$e_{A\times B}(\langle a,b\rangle, \langle a',b'\rangle) = \begin{cases} \iota_0^{A\otimes B(b)}[e_A(a,a'), b'] & \text{if } a \not\leq_A a', \\ \iota_1^{A(a)\otimes B}[a', e_B(b,b')] & \text{otherwise.} \end{cases}$$

Finally, we consider the case of a type A^*. For $a = \langle\rangle_A \in A^*$ it suffices to observe that $(A^*)_a$ is empty, since $\langle\rangle_A \leq_{A^*} \tau$ holds for any $\tau \in A^*$. Now consider a term of the form $a = a_0 \star \sigma \in A^*$. We write $b = \langle b_0, \dots, b_{n-1}\rangle_A \in (A^*)_a$ and distinguish two cases. If we have $a_0 \not\leq_A b_i$ for all $i < n$, then we set

$$e_{A^*}(a,b) = \iota_0^{A(a_0)^*\otimes A\otimes A^*(\sigma)} \langle e_A(a_0, b_0), \dots, e_A(a_0, b_{n-1})\rangle_{A(a_0)}.$$

Note that this is an element of $A(a_0)^* + A(a_0)^* \otimes A \otimes A^*(\sigma) = A^*(a)$, as required. Otherwise we fix the smallest number $i < n$ with $a_0 \leq_A b_i$. In view of $b \in (A^*)_a$ we must have $\sigma \not\leq_{A^*} \langle b_{i+1}, \dots, b_{n-1}\rangle_A$. We can thus define $e_{A^*}(a,b)$ as

$$\iota_1^{A(a_0)^*}[\langle e_A(a_0, b_0), \dots, e_A(a_0, b_{i-1})\rangle_{A(a_0)}, b_i, e_{A^*}(\sigma, \langle b_{i+1}, \dots, b_{n-1}\rangle_A)].$$

Using the induction hypothesis, one readily checks that $e_{A^*}(a,b) \leq_{A^*(a)} e_{A^*}(a,b')$ implies $b \leq_{A^*} b'$. □

Our next aim is to iterate the previous construction along bad sequences. Given a type A, we write $\sigma \in \mathrm{Bad}(A)$ to express that σ is a finite bad sequence in A. This means that we have $\sigma = \langle a_0, \dots, a_{n-1}\rangle$ for terms $a_0, \dots, a_{n-1} \in A$ that satisfy $a_i \not\leq_A a_j$ for all $i < j < n$. If we have $\sigma \in \mathrm{Bad}(A)$ and σ is different from the empty sequence

https://doi.org/10.1017/S1755020321000265 Published online by Cambridge University Press

$\langle\rangle$, then we write $\sigma \in \mathrm{Bad}^-(A)$. For $\sigma = \langle a_0, \dots, a_{n-1}\rangle \in \mathrm{Bad}(A)$ we abbreviate $\sigma^\frown a = \langle a_0, \dots, a_{n-1}, a\rangle$ and put

$$a \in A_\sigma \qquad :\Leftrightarrow \qquad a \in A \text{ and } \sigma^\frown a \in \mathrm{Bad}^-(A).$$

Once again, note that we officially define a primitive recursive relation $a \in A_\sigma$ between finite objects, rather than an infinite set A_σ. The expressions $A(a)$ and $e_A(a, b)$ have only been explained for $a \in A$ and $b \in A_a$. We will see that the following definition does conform with these restrictions. Before this is established, we may simply assume that some default value is assigned outside of the intended domain of definition.

Definition 6.9 (**PA**). Consider a type A. For a sequence $\sigma \in \mathrm{Bad}(A)$ and a term $b \in A_\sigma$ we define $A[\sigma]$ and $\hat{e}_A(\sigma, b)$ by the recursive clauses

$$A[\langle\rangle] = A, \qquad A[\sigma^\frown a] = A[\sigma](\hat{e}_A(\sigma, a)),$$
$$\hat{e}_A(\langle\rangle, b) = b, \qquad \hat{e}_A(\sigma^\frown a, b) = e_{A[\sigma]}(\hat{e}_A(\sigma, a), \hat{e}_A(\sigma, b)).$$

In order to justify the recursion in detail, we consider $\sigma = \langle a_0, \dots, a_{n-1}\rangle$ and write $\sigma \restriction i = \langle a_0, \dots, a_{i-1}\rangle$. Then $A[\sigma \restriction i]$ and the values $\hat{e}_A(\sigma \restriction i, a_j)$ for $i \le j < n$ are constructed simultaneously by recursion on $i < n$. For $\sigma' := \sigma^\frown a_n$ with $a_n := b$ this also explains the value $\hat{e}_A(\sigma, b) = \hat{e}_A(\sigma' \restriction n, a_n)$.

Corollary 6.10 (**PA**). *If σ is a finite bad sequence in the type A, then $A[\sigma]$ is a type. For any $b \in A_\sigma$ the value $\hat{e}_A(\sigma, b)$ is a term of this type. Furthermore we have*

$$\hat{e}_A(\sigma, b) \le_{A[\sigma]} \hat{e}_A(\sigma, b') \quad \Rightarrow \quad b \le_A b'$$

for any terms $b, b' \in A_\sigma$.

Proof. We use induction on σ to verify all claims simultaneously. The case of $\sigma = \langle\rangle$ is immediate. Now assume that we have $\sigma = \sigma_0{}^\frown a$. The induction hypothesis tells us that $\hat{e}_A(\sigma_0, a)$ is a term of type $A[\sigma_0]$. In view of Definition 6.7 it follows that $A[\sigma] = A[\sigma_0](\hat{e}_A(\sigma_0, a))$ is a type. For $b \in A_\sigma$ we have $a \not\le_A b$, so that the induction hypothesis yields $\hat{e}_A(\sigma_0, a) \not\le_{A[\sigma_0]} \hat{e}_A(\sigma_0, b)$. By Proposition 6.8 we get

$$\hat{e}_A(\sigma, b) = e_{A[\sigma_0]}(\hat{e}_A(\sigma_0, a), \hat{e}_A(\sigma_0, b)) \in A[\sigma_0](\hat{e}_A(\sigma_0, a)) = A[\sigma].$$

From $\hat{e}_A(\sigma, b) \le_{A[\sigma]} \hat{e}_A(\sigma, b')$ we get $\hat{e}_A(\sigma_0, b) \le_{A[\sigma_0]} \hat{e}_A(\sigma_0, b')$, also by Proposition 6.8. Then $b \le_A b'$ follows by induction hypothesis. □

In order to obtain a reification, it remains to assign a suitable ordinal to each type. Let us write $\alpha \oplus \beta$ and $\alpha \otimes \beta$ for the natural ("Hessenberg") sum and product of ordinals $\alpha, \beta \prec \varepsilon_0$ (see, e.g., [39, sec. 4]). In contrast to the usual operations of ordinal arithmetic, the natural variants are commutative and strictly increasing in both arguments. Ordinals of the form ω^γ are additively indecomposable, in the sense that $\alpha, \beta \prec \omega^\gamma$ implies $\alpha \oplus \beta \prec \omega^\gamma$; conversely, any additively indecomposable ordinal $\delta \neq 0$ has the form $\delta = \omega^\gamma$. For $\alpha, \beta \prec \omega_2^\gamma := \omega^{(\omega^\gamma)}$ we have $\alpha \otimes \beta \prec \omega_2^\gamma$.

Definition 6.11 (**PA**). Let us say that a type is low if it does not involve the constant symbol $\mathfrak{B}$. We recursively assign an ordinal $o(A)$ to each low type A:

$$o(\mathfrak{E}) = 0, \qquad o(A + B) = o(A) \oplus o(B),$$
$$o(A \times B) = o(A) \otimes o(B), \qquad o(A^*) = \omega_2^{o(A)}.$$

The following is crucial for the construction of a reification.

https://doi.org/10.1017/S1755020321000265 Published online by Cambridge University Press

Proposition 6.12 (**PA**). *If A is a low type and $a \in A$ is a term, then $A(a)$ is a low type and we have $o(A(a)) \prec o(A)$.*

Proof. As preparation we note that $A \otimes B$ is low if the same holds for A and B. A straightforward induction shows $o(A \otimes B) = o(A) \otimes o(B)$; for example, the distributivity property from [39, lemma 4.5(8)] accounts for the inductive verification

$$o((C+D)\otimes B) = o(C\otimes B + D\otimes B) = o(C\otimes B)\oplus o(D\otimes B)$$
$$= (o(C)\otimes o(B))\oplus(o(D)\otimes o(B)) = (o(C)\oplus o(D))\otimes o(B) = o(C+D)\otimes o(B).$$

By induction on A one can show that $o(A)$ is additively indecomposable when A is an indecomposable type. The most interesting step concerns a type $A = B \times C$, where B and C are indecomposable according to Definition 6.1. Inductively we may write $o(B) = \omega^\beta$ and $o(C) = \omega^\gamma$ (unless we have $o(A) = 0$). Then

$$o(B \times C) = o(B) \otimes o(C) = \omega^\beta \otimes \omega^\gamma = \omega^{\beta\oplus\gamma}$$

is an additively indecomposable ordinal as well. The claim of the proposition can now be verified by induction over $\#a$, for all types A simultaneously. First consider the case of a term $\iota_0^B a \in A + B$. The induction hypothesis tells us that $A(a)$ is low with $o(A(a)) \prec o(A)$. Hence $(A+B)(\iota_0^B a) = A(a) + B$ is low and we have

$$o((A+B)(\iota_0^B)) = o(A(a)+B) = o(A(a)) \oplus o(B) \prec o(A) \oplus o(B) = o(A+B).$$

The case of $\iota_1^A b \in A + B$ is analogous. Now consider a term $\langle a, b\rangle \in A \times B$. In view of the above, the induction hypothesis implies that $A(a) \otimes B$ is low with ordinal

$$o(A(a)\otimes B) = o(A(a)) \otimes o(B) \prec o(A) \otimes o(B) = o(A \times B).$$

In the same way we get $o(A \otimes B(b)) \prec o(A \times B)$. In view of Definition 6.1, a type of the form $A \times B$ is always indecomposable. By the above this entails that $o(A \times B)$ is an additively indecomposable ordinal. Hence we obtain

$$o((A\times B)(\langle a,b\rangle)) = o(A(a)\otimes B + A \otimes B(b))$$
$$= o(A(a)\otimes B) \oplus o(A\otimes B(b)) \prec o(A\times B).$$

Finally, we consider the case of a type A^*. Concerning the term $\langle\rangle_A \in A^*$, we note

$$o(A^*(\langle\rangle_A)) = o(\mathfrak{E}) = 0 \prec \omega_2^{o(A)} = o(A^*).$$

Now consider a term $a \star \sigma \in A^*$ (see the proof of Proposition 6.8 for the notation). In view of $\#a, \#\sigma < \#a \star \sigma$ the induction hypothesis yields $o(A^*(\sigma)) \prec o(A^*) = \omega_2^{o(A)}$ and $o(A(a)) \prec o(A)$. The latter implies $o(A(a)^*) = \omega_2^{o(A(a))} \prec \omega_2^{o(A)}$. Since we are concerned with ordinals below ε_0, we also have $o(A) \prec \omega_2^{o(A)}$. Using the fact that $\omega_2^{o(A)}$ is additively and multiplicatively indecomposable, we can deduce

$$o(A^*(a\star\sigma)) = o(A(a)^* + A(a)^* \otimes A \otimes A^*(\sigma))$$
$$= o(A(a)^*) \oplus o(A(a)^*) \otimes o(A) \otimes o(A^*(\sigma)) \prec \omega_2^{o(A)} = o(A^*),$$

as required. □

Recall that the terms of type $\mathfrak{B}$ coincide with the finite binary trees, i.e., with the elements of $\mathcal{B}$. Below we will show that the type $\mathfrak{B}[\sigma]$ is low for any non-empty bad sequence $\sigma \in \mathrm{Bad}^-(\mathcal{B}) = \mathrm{Bad}^-(\mathfrak{B})$. To state the following definition, we simply assume that the primitive recursive function $o(\cdot)$ is extended to arbitrary arguments.

DEFINITION 6.13 (**PA**). For $\sigma \in \mathrm{Bad}^-(\mathcal{B})$ we put $r(\sigma) := o(\mathfrak{B}[\sigma])$.

Finally, we can deduce the promised result:

COROLLARY 6.14 (**PA**). *The primitive recursive function* $r : \mathrm{Bad}^-(\mathcal{B}) \to \varepsilon_0$ *is a reification, i.e., we have*

$$r(\langle t_0, \dots, t_n, t_{n+1}\rangle) \prec r(\langle t_0, \dots, t_n\rangle)$$

for any bad sequence $\langle t_0, \dots, t_n, t_{n+1}\rangle$ *in* $\mathcal{B}$.

Proof. We use induction on $\sigma \in \mathrm{Bad}^-(\mathfrak{B})$ to show that $\mathfrak{B}[\sigma]$ is a low type. For this purpose it is crucial to recall that the empty sequence was included in $\mathrm{Bad}(\mathfrak{B})$ but excluded from $\mathrm{Bad}^-(\mathfrak{B})$. Hence the base case concerns a sequence of the form $\sigma = \langle t\rangle$. In view of Definition 6.9 we have

$$\mathfrak{B}[\langle t\rangle] = \mathfrak{B}[\langle\rangle](\hat{e}_{\mathfrak{B}}(\langle\rangle, t)) = \mathfrak{B}(t).$$

Even though the type $\mathfrak{B}$ is not low, a straightforward induction on $t \in \mathfrak{B}$ shows that $\mathfrak{B}(t)$ is a low type. Now consider a sequence $\sigma^\frown t \in \mathrm{Bad}^-(\mathfrak{B})$ with $\sigma \neq \langle\rangle$. The induction hypothesis ensures that $\mathfrak{B}[\sigma]$ is a low type. According to Corollary 6.10 we have $\hat{e}_{\mathfrak{B}}(\sigma, t) \in \mathfrak{B}[\sigma]$. By (the easy part of) Proposition 6.12 we conclude that

$$\mathfrak{B}[\sigma^\frown t] = \mathfrak{B}[\sigma](\hat{e}_{\mathfrak{B}}(\sigma, t))$$

is a low type as well. The more substantial part of Proposition 6.12 yields

$$r(\sigma^\frown t) = o(\mathfrak{B}[\sigma^\frown t]) \prec o(\mathfrak{B}[\sigma]) = r(\sigma).$$

For $\sigma = \langle t_0, \dots, t_n\rangle$ and $t = t_{n+1}$ this is the claim of the corollary. □

§7. From reflection to the finite basis property.

Working over **PA**, we show that $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$ entails $\mathcal{TI}(\varepsilon_0, \Pi^-_1)$, which does in turn entail $\mathcal{K}\Sigma^-_1$. This completes our proof that all three principles are equivalent. Using Goryachev's theorem, we can deduce a characterization of the Π^0_1-sentences that are provable in $\mathbf{PA} + \mathcal{K}\Sigma^-_1$.

For the case of uniform reflection and induction with parameters, the following has been shown by Kreisel and Lévy [27].

PROPOSITION 7.1. *Each instance of* $\mathcal{TI}(\varepsilon_0, \Pi^-_1)$ *can be proved in* $\mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$.

Proof. Consider a Π^0_1-formula $\psi(x)$ with a single free variable. Arguing in the theory $\mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$, we establish $\mathcal{TI}(\varepsilon_0, \psi)$ by contraposition: Assume that the conclusion of transfinite induction fails, so that we have $\exists_{\alpha \prec \varepsilon_0} \neg\psi(\alpha)$. The latter is a Σ^0_1-formula, so that its truth can be established by an explicit verification. More formally, we invoke formalized Σ^0_1-completeness (cf. [20, theorem I.1.8]) to obtain

$$\exists_{\alpha \prec \varepsilon_0} \mathrm{Pr}_{\mathbf{PA}}(\neg\psi(\dot{\alpha})).$$

This uses Feferman's dot notation: By $\psi(\dot{\alpha})$ one denotes the closed object formula that results from $\psi(x)$ when we substitute x by the α-th numeral, where the code α is considered as a natural number (cf. the notation in [20, corollary I.1.76]). Gentzen [19] has shown that **PA** proves induction up to each fixed ordinal below ε_0. This result can itself be formalized in Peano arithmetic (and in much weaker theories, cf. [13, sec. 3]), which means that we have

$$\forall_{\alpha\prec\varepsilon_0}\,\mathrm{Pr}_{\mathbf{PA}}(\mathrm{Prog}_{\varepsilon_0}(\psi)\to\psi(\dot\alpha)).$$

Together with the above we get $\mathrm{Pr}_{\mathbf{PA}}(\neg\,\mathrm{Prog}_{\varepsilon_0}(\psi))$. By an instance of $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma^0_2)$ we obtain $\neg\,\mathrm{Prog}_{\varepsilon_0}(\psi)$, which is (provably equivalent to) a closed Σ^0_2-formula. Hence the premise of $\mathcal{TI}(\varepsilon_0,\psi)$ fails, so that our proof by contraposition is complete. □

The following is a consequence of the result that $(\mathcal{B},\leq_{\mathcal{B}})$ is a well partial order with maximal order type ε_0, which is due to de Jongh (unpublished; cf. the introduction to [35]) and Diana Schmidt (see [36, theorem II.2] in combination with the example after [36, definition I.15]). A detailed proof in our setting has been given in the previous section.

Proposition 7.2. *Each instance of $\mathcal{K}\Sigma^-_1$ can be proved in* $\mathbf{PA}+\mathcal{TI}(\varepsilon_0,\Pi^-_1)$.

Proof. Let us fix an instance $\mathcal{K}\varphi$, where φ is a Σ^0_1-formula with a single free variable. It is instructive to recall the argument from Remark 3.1, which relies on a notion of φ-sequence. If $\{n\in\mathbb{N}\,|\,\varphi(n)\}$ is computably enumerable but not decidable, then it is not decidable whether a given finite sequence is a φ-sequence. For this reason we now introduce a finer notion: Write $\varphi(x)\equiv\exists_y\theta(x,y)$ with a Δ^0_0-formula θ. As in the previous section we write $\mathrm{Bad}^-(\mathcal{B})$ for the set of non-empty finite bad sequences in $\mathcal{B}$. By a certified φ-sequence we mean a finite sequence

$$(t_0,c_0),\dots,(t_n,c_n)\subseteq\mathcal{B}\times\mathbb{N}$$

such that we have $\langle t_0,\dots,t_n\rangle\in\mathrm{Bad}^-(\mathcal{B})$ and $\theta(t_i,c_i)$ for all $i\leq n$. Note that the latter implies $\varphi(t_i)$. Since θ contains no further free variables, the notion of certified φ-sequence is defined by a Δ^0_1-formula without parameters. By picking the value $f(n)$ with minimal code, we thus obtain a (possibly partial) computable function $f:\mathbb{N}\to\mathcal{B}\times\mathbb{N}$ with the following property:

- If the sequence $\langle f(0),\dots,f(n-1)\rangle$ is defined and can be extended into a certified φ-sequence of length $n+1$, then $\langle f(0),\dots,f(n)\rangle$ is such a sequence.

Note that the relation $f(x)=y$ is Σ^0_1-definable without parameters. Working in $\mathbf{PA}$, we now assume that $\mathcal{K}\varphi$ is false and deduce that $\mathcal{TI}(\varepsilon_0,\psi)$ fails for a suitable formula ψ. The failure of $\mathcal{K}\varphi$ entails that all values $f(n)$ are defined: Inductively, we may assume that $f(m)=(t_m,c_m)$ is defined for all $m<n$; in the case of $n>0$, the construction of f ensures that $\langle(t_0,c_0),\dots,(t_{n-1},c_{n-1})\rangle$ is a certified φ-sequence. In order to deduce that $f(n)$ is defined as well, we consider the set $a:=\{t_0,\dots,t_{n-1}\}$. As $\mathcal{K}\varphi$ is false, we must have

$$\neg\forall_{s\in a}\varphi(s)\vee\exists_{t\in\mathcal{B}}(\varphi(t)\wedge\forall_{s\in a}s\not\leq_{\mathcal{B}}t).$$

For $s=t_m\in a$, the construction of f ensures $\theta(t_m,c_m)$ and thus $\varphi(s)$. Hence the second disjunct yields an element $t_n\in\mathcal{B}$ with $\varphi(t_n)$ and $t_m\not\leq_{\mathcal{B}}t_n$ for all $m<n$. The latter implies $\langle t_0,\dots,t_n\rangle\in\mathrm{Bad}^-(\mathcal{B})$. Due to $\varphi(t_n)$ we can pick a number c_n with $\theta(t_n,c_n)$. Then $\langle f(0),\dots,f(n-1),(t_n,c_n)\rangle$ is a certified φ-sequence, and $f(n)$ is defined as the smallest pair $\langle t_n,c_n\rangle$ for which this holds. We can now define a total computable function $g:\mathbb{N}\to\mathrm{Bad}^-(\mathcal{B})$ by setting

$$g(n):=\langle t_0,\dots,t_n\rangle\quad\text{with}\quad f(m)=(t_m,c_m).$$

According to Corollary 6.14, there is a primitive recursive reification

$$r : \mathrm{Bad}^{-}(\mathcal{B}) \to \varepsilon_0.$$

It follows that the total computable function $r \circ g : \mathbb{N} \to \varepsilon_0$ is strictly decreasing. This is impossible in the presence of $\mathcal{TI}(\varepsilon_0, \Pi_1^-)$. To be more precise, we consider

$$\psi(\alpha) \;:\equiv\; \forall_n\, \alpha \preceq r \circ g(n) \;\equiv\; \forall_n \forall_\sigma \forall_{\delta \prec \varepsilon_0}(g(n) = \sigma \wedge r(\sigma) = \delta \to \alpha \preceq \delta),$$

which is a Π_1^0-formula with no other free variables than α. For $\alpha = r \circ g(0) + 1 \prec \varepsilon_0$ we clearly have $\neg\psi(\alpha)$, which refutes the conclusion of $\mathcal{TI}(\varepsilon_0, \psi)$. On the other hand, the premise of this induction statement holds: To derive $\mathrm{Prog}_{\varepsilon_0}(\psi)$ by contraposition, we assume that $\psi(\gamma)$ fails, i.e., that we have $r \circ g(n) \prec \gamma$ for some $n \in \mathbb{N}$. Since $r \circ g$ is strictly decreasing, we obtain

$$r \circ g(n+1) \prec r \circ g(n) =: \beta \prec \gamma,$$

which yields $\neg\psi(\beta)$ and thus refutes $\forall_{\beta \prec \gamma}\psi(\beta)$. □

Together with Propositions 4.4, 5.1 and 7.1 we obtain the following:

THEOREM 7.3. *We have*

$$\mathbf{PA} + \mathcal{K}\Sigma_1^- \;\equiv\; \mathbf{PA} + \mathcal{TI}(\varepsilon_0, \Pi_1^-) \;\equiv\; \mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0),$$

i.e., all three theories prove the same theorems.

Let $\mathrm{Con}(\mathbf{PA} + \varphi)$ be a reasonable formalization of the statement that $\mathbf{PA} + \varphi$ is consistent. We consider the recursively generated Π_1^0-sentences

$$\mathrm{Con}_0(\mathbf{PA}) \;:\equiv\; 0 = 0,$$
$$\mathrm{Con}_{n+1}(\mathbf{PA}) \;:\equiv\; \mathrm{Con}(\mathbf{PA} + \mathrm{Con}_n(\mathbf{PA})).$$

Note that $\mathrm{Con}_1(\mathbf{PA})$ is equivalent to the usual consistency statement. As mentioned in the introduction, we obtain the following:

COROLLARY 7.4. *We have*

$$\mathbf{PA} + \mathcal{K}\Sigma_1^- \;\equiv_{\Pi_1^0}\; \mathbf{PA} + \{\mathrm{Con}_n(\mathbf{PA}) \mid n \in \mathbb{N}\},$$

i.e., the two theories prove the same Π_1^0-sentences.

Proof. Let us write $\mathrm{Rfn}_{\mathbf{PA}}$ for the full local reflection principle, i.e., the collection of all formulas $\mathrm{Pr}_{\mathbf{PA}}(\varphi) \to \varphi$, where φ can be any sentence in the language of first order arithmetic. According to Goryachev's theorem (see, e.g., [30, theorem IV.5]), any Π_1^0-theorem of $\mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}$ can be proved in $\mathbf{PA} + \{\mathrm{Con}_n(\mathbf{PA}) \mid n \in \mathbb{N}\}$. A fortiori, this applies to all Π_1^0-theorems of $\mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0) \equiv \mathbf{PA} + \mathcal{K}\Sigma_1^-$. In the other direction we have a full inclusion: The theory $\mathbf{PA} + \mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0)$ proves all theorems of $\mathbf{PA} + \{\mathrm{Con}_n(\mathbf{PA}) \mid n \in \mathbb{N}\}$, because it proves each statement $\mathrm{Con}_n(\mathbf{PA})$. For $n = 0$ this is trivial. To conclude by meta induction on n, it suffices to observe that the formula $\mathrm{Con}_n(\mathbf{PA}) \to \mathrm{Con}_{n+1}(\mathbf{PA})$ is the contrapositive of

$$\mathrm{Pr}_{\mathbf{PA}}(\neg\,\mathrm{Con}_n(\mathbf{PA})) \to \neg\,\mathrm{Con}_n(\mathbf{PA}),$$

which is an instance of $\mathrm{Rfn}_{\mathbf{PA}}(\Sigma_2^0)$. □

Note that the corollary does not extend to arbitrary formula complexity: In the theory $\mathbf{PA} + \{\mathrm{Con}_n(\mathbf{PA}) \mid n \in \mathbb{N}\}$ one cannot prove all instances of $\mathcal{K}\Sigma_1^-$, by Corollary 5.4.

Acknowledgments. I would like to thank Lev Beklemishev for our inspiring discussions and his helpful comments on a first draft. Also, I am very grateful to the anonymous referees, whose detailed feedback has helped me to improve the paper.

BIBLIOGRAPHY

[1] Beklemishev, L. (1997). Notes on local reflection principles. *Theoria*, **63**, 139–146.

[2] ———. (2004). Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, **128**(1–3), 103–123.

[3] ———. (2005). Reflection principles and provability algebras in formal arithmetic. *Russian Mathematical Surveys*, **60**(2), 197–268.

[4] Beklemishev, L., & Visser, A. (2005). On the limit existence principles in elementary arithmetic and Σ_n^0-consequences of theories. *Annals of Pure and Applied Logic*, **136**, 56–74.

[5] Buchholz, W. (1991). Notation systems for infinitary derivations. *Archive for Mathematical Logic*, **30**, 277–296.

[6] Buchholz, W., & Wainer, S. (1987). Provably computable functions and the fast growing hierarchy. In Simpson, S. G., editor. *Logic and Combinatorics. Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference 1985*. Contemporary Mathematics, Vol. 65. Providence, RI: American Mathematical Society, pp. 179–198.

[7] Cichon, E. A. (1983). A short proof of two recently discovered independence results using recursion theoretic methods. *Proceedings of the American Mathematical Society*, **87**, 704–706.

[8] de Jongh, D., & Parikh, R. (1977). Well-partial orderings and hierarchies. *Indagationes Mathematicae*, **80**(3), 195–207.

[9] Ewald, W., & Sieg, W. (eds.). (2013). *David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917–1933*. Berlin: Springer.

[10] Feferman, S. (1992). Why a little bit goes a long way: Logical foundations of scientifically applicable mathematics, *PSA*: *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*. Chicago: University of Chicago, pp. 442–455.

[11] Fellows, M. R., & Langston, M. A. (1988). Nonconstructive tools for proving polynomial-time decidability. *Journal of the Association for Computing Machinery*, **35**(3), 727–735.

[12] ———. (1992). On well-partial-order theory and its application to combinatorial problems of VLSI design. *SIAM Journal on Discrete Mathematics*, **5**(1), 117–126.

[13] Freund, A., & Pakhomov, F. (2020). Short proofs for slow consistency. *Notre Dame Journal of Formal Logic*, **61**(1), 31–49.

[14] Freund, A., Rathjen, M., & Weiermann, A. (2020). Minimal bad sequences are necessary for a uniform Kruskal theorem. Preprint, arXiv:2001.06380.

[15] Friedman, H. (2018). *Explicitly Π_1^0 Status 4/20/18*. Manuscript. Available from: https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/1/1952/files/2014/01/CMI_ExplicPi01042018-2ilxsqy.pdf (Accessed 31 March 2020).

[16] Friedman, H., Robertson, N., & Seymour, P. (1987). Metamathematics of the graph minor theorem. In Simpson, S. G., editor. *Logic and Combinatorics. Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference 1985*. Contemporary Mathematics, Vol. 65. Providence, RI: American Mathematical Society, pp. 229–261.

[17] Friedman, H., & Sheard, M. (1995). Elementary descent recursion and proof theory. *Annals of Pure and Applied Logic*, **71**, 1–45.

[18] Gentzen, G. (1936). Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, **112**, 493–565.

[19] ———. (1943). Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, **119**, 149–161.

[20] Hájek, P., & Pudlák, P. (1993). *Metamathematics of First-Order Arithmetic*. Perspectives in Mathematical Logic, vol. 3. Berlin: Springer.

[21] Hasegawa, R. (1994). Well-ordering of algebras and Kruskal's theorem. In Jones, N. D., Hagiya, M., and Sato, M., editors. *Logic, Language and Computation*. Lecture Notes in Computer Science, Vol. 792. Berlin: Springer, pp. 133–172.

[22] Hilbert, D. (1926). Über das Unendliche. *Mathematische Annalen*, **95**, 161–190, English translation in [48].

[23] ———. (1928). Die Grundlagen der Mathematik. *Abhandlungen aus dem Seminar der Hamburgischen Universität*, **6**, 65–85, reprinted in [9], English translation in [48].

[24] Kohlenbach, U. (2008). *Applied Proof Theory. Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Berlin: Springer.

[25] Kreisel, G. (1960). Ordinal logics and the characterization of informal notions of proof. In Todd, J. A., editor. *Proceedings of the International Congress of Mathematicians 1958*. Cambridge: Cambridge University Press, pp. 289–299.

[26] ———. (1982). Finiteness theorems in arithmetic: An application of Herbrand's theorem forΣ_2-formulas. In Stern, J., editor. *Proceedings of the Herbrand Symposium*. *Logic Colloquium '81*. Studies in Logic and the Foundations of Mathematics, Vol. 107. Amsterdam: North-Holland, pp. 39–55.

[27] Kreisel, G., & Lévy, A. (1968). Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **14**, 97–142.

[28] Kruskal, J. (1960). Well-quasi-ordering, the tree theorem, and Vazsonyi's conjecture. *Transactions of the American Mathematical Society*, **95**(2), 210–225.

[29] Lindström, P. (1984). On partially conservative sentences and interpretability. *Proceedings of the American Mathematical Society*, **91**(3), 436–443.

[30] ———. (1997). *Aspects of Incompleteness*. Lecture Notes in Logic, Vol. 10. Berlin: Springer.

[31] Paris, J., & Harrington, L. (1977). A mathematical incompleteness in Peano arithmetic. In Barwise, J., editor. *Handbook of Mathematical Logic*. Amsterdam: North-Holland, pp. 1133–1142.

[32] Pohlers, W. (2009). *Proof Theory. The First Step into Impredicativity*. Berlin: Springer.

[33] Rathjen, M., & Weiermann, A. (1993). Proof-theoretic investigations on Kruskal's theorem. *Annals of Pure and Applied Logic*, **60**, 49–88.

[34] Robertson, N., & Seymour, P. (2004). Graph minors. XX. Wagner's conjecture. *Journal of Combinatorial Theory, Series B*, **92**(2), 325–357.

[35] Schmidt, D. (1975). Bounds for the closure ordinals of replete monotonic increasing functions. *The Journal of Symbolic Logic*, **40**(3), 305–316.

[36] ———. (1979). *Well-Partial Orderings and Their Maximal Order Types*. Habilitationsschrift, Universität Heidelberg.

[37] Shelah, S. (1984). On logical sentences in PA. In Lolli, G., Longo, G., and Marcja, A., editors. *Logic Colloquium '82*. Studies in Logic and the Foundations of Mathematics, Vol. 122. Amsterdam: Elsevier, pp. 145–160.

[38] Simpson, S. G. (1985). Nonprovability of certain combinatorial properties of finite trees. In Harrington, L. A., Morley, M. D., Sčědrov, A., and Simpson, S. G., editors. *Harvey Friedman's Research on the Foundations of Mathematics*. Studies in Logic and the Foundations of Mathematics, Vol. 117. Amsterdam: North-Holland, pp. 87–117.

[39] ———. (1988). Ordinal numbers and the Hilbert basis theorem. *The Journal of Symbolic Logic*, **53**(3), 961–974.

[40] ———. (2009). *Subsystems of Second Order Arithmetic* (second edition). Perspectives in Logic. Cambridge: Cambridge University Press.

[41] Smith, R. L. (1985). The consistency strengths of some finite forms of the Higman and Kruskal theorems. In Harrington, L. A., Morley, M. D., Sčědrov, A., and Simpson, S. G., editors. *Harvey Friedman's Research on the Foundations of Mathematics*. Studies in Logic and the Foundations of Mathematics, Vol. 117. Amsterdam: North-Holland, pp. 119–136.

[42] Smoryński, C. (1977). The incompleteness theorems. In Barwise, J., editor. *Handbook of Mathematical Logic*. Studies in Logic and the Foundations of Mathematics, Vol. 90. Amsterdam: North-Holland, pp. 821–865.

[43] Sommer, R. (1995). Transfinite induction within Peano arithmetic. *Annals of Pure and Applied Logic* **76**, 231–289.

[44] Tait, W. (1981). Finitism. *Journal of Philosophy*, **78**, 524–546.

[45] ———. (2002). Remarks on finitism. In Sieg, W., Sommer, R., and Talcott, C., editors. *Reflections on the Foundations of Mathematics. Essays in Honor of Solomon Feferman*. Lecture Notes in Logic, Vol. 15. New York: A.K. Peters, pp. 410–419.

[46] Takeuti, G. (1987). *Proof Theory* (second edition). Studies in Logic and the Foundations of Mathematics, Vol. 81. Amsterdam: North-Holland.

[47] Tarski, A. (1936). Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, **1**, 261–405.

[48] van Heijenoort, J. (ed.). (1967). *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press.

[49] Zach, R. Hilbert's program. In Zalta, E. N., editor. *The Stanford Encyclopedia of Philosophy* (fall 2019 edition). Available from: https://plato.stanford.edu/archives/fall2019/entries/hilbert-program.

FACHBEREICH MATHEMATIK
TECHNICAL UNIVERSITY OF DARMSTADT
SCHLOSSGARTENSTR. 7
64289 DARMSTADT, GERMANY
E-mail: freund@mathematik.tu-darmstadt.de