THE JOURNAL OF SYMBOLIC LOGIC
Volume 89, Number 1, March 2024

# AN INCOMPLETENESS THEOREM VIA ORDINAL ANALYSIS

JAMES WALSH

**Abstract.** We present an analogue of Gödel's second incompleteness theorem for systems of second-order arithmetic. Whereas Gödel showed that sufficiently strong theories that are $\Pi^0_1$-sound and $\Sigma^0_1$-definable do not prove their own $\Pi^0_1$-soundness, we prove that sufficiently strong theories that are $\Pi^1_1$-sound and $\Sigma^1_1$-definable do not prove their own $\Pi^1_1$-soundness. Our proof does not involve the construction of a self-referential sentence but rather relies on ordinal analysis.

**§1. Introduction.** The motivation for this project come from two sources: Gödel's second incompleteness theorem and Gentzen's consistency proof of arithmetic. These results are complementary in many ways. In the first place, they jointly form a complicated and ambiguous resolution of Hilbert's problem of proving the consistency of arithmetic. Moreover, Gentzen's proof refines Gödel's result by exhibiting the first example of a non-meta-mathematical arithmetic statement—namely, the statement that $\varepsilon_0$ lacks primitive recursive descending sequences—that is not provable from the Peano axioms. Though Gödel's result is highly general, his proof relies on self-reference, rendering it opaque and mysterious [6, 15, 16]. By contrast, Gentzen's proof is concrete but his results are specific to the case of Peano arithmetic. In this paper we prove a version of the second incompleteness theorem that is general like Gödel's but with a proof that is concrete like Gentzen's; in particular, we use the methods of ordinal analysis and do not rely on diagonalization or self-reference.

Let's start by giving a typical statement Gödel's second incompleteness theorem:

THEOREM 1.1 (Gödel). *No consistent and recursively axiomatizable extension of elementary arithmetic proves its own consistency.*

Recursive axiomatizability is equivalent to $\Sigma^0_1$-definability by Craig's Trick. Moreover, consistency is provably equivalent (in elementary arithmetic) to $\Pi^0_1$-soundness. Hence, we may restate Gödel's Theorem as follows:

THEOREM 1.2 (Gödel). *If $T$ is a $\Pi^0_1$-sound and $\Sigma^0_1$-definable extension of elementary arithmetic, then $T$ does not prove its own $\Pi^0_1$-soundness.*

We prove the following analogous result for systems of second-order arithmetic:

THEOREM 1.3. *If $T$ is a $\Pi^1_1$-sound and $\Sigma^1_1$-definable extension of $\Sigma^1_1$-$\mathsf{AC}_0$, then $T$ does not prove its own $\Pi^1_1$-soundness.*

Received September 21, 2021.
2020 *Mathematics Subject Classification.* 03F15, 03F40.
*Key words and phrases.* proof theory, second-order arithmetic, ordinal analysis, incompleteness.


0022-4812/24/8901-0004
DOI:10.1017/jsl.2022.64

$\Pi^1_1$-soundness is a strictly stronger condition than $\Pi^0_1$-soundness. However, $\Sigma^1_1$-definability is a strictly weaker condition than $\Sigma^0_1$-definability. Hence, Theorem 1.3 is neither weaker nor stronger than Gödel's Theorem but incomparable with it.

Let's take a brief look at the ideas motivating the proof. In what follows $\mathsf{WF}(\prec)$ is a sentence expressing the well-foundedness of $\prec$:

$$\mathsf{WF}(\prec) := \forall X \big(\exists x \in X \to \exists x \in X\ \forall y \in X\ \neg y \prec x\big);$$

$\mathsf{RFN}_{\Pi^1_1}(T)$ is a sentence naturally expressing the $\Pi^1_1$-soundness of $T$:

$$\mathsf{RFN}_{\Pi^1_1}(T) := \forall \varphi \in \Pi^1_1 \big(\mathsf{Pr}_T(\varphi) \to \mathsf{True}_{\Pi^1_1}(\varphi)\big);$$

and the proof-theoretic ordinal $|T|_{\mathsf{AN}}$ of a theory $T$ is the supremum of the ordinals $\alpha$ for which there is some $\Sigma^1_1$ presentation $\prec$ of $\alpha$ such that $T \vdash \mathsf{WF}(\prec)$.

Assuming that $T$ is $\Pi^1_1$-sound and $\Sigma^1_1$-definable, Spector's $\Sigma^1_1$-bounding theorem implies that $|T|_{\mathsf{AN}}$ is strictly less than $\omega_1^{\mathsf{CK}}$, whence $|T|_{\mathsf{AN}}$ has some $\Sigma^1_1$ presentation. For any $\Sigma^1_1$-presentation $\prec$ of $|T|_{\mathsf{AN}}$, the following is true by definition:

$$T \nvdash \mathsf{WF}(\prec). \tag{1}$$

We then need to show that there is *at least one* $\Sigma^1_1$-presentation $\prec$ of $|T|_{\mathsf{AN}}$ such that:

$$T \vdash \mathsf{RFN}_{\Pi^1_1}(T) \to \mathsf{WF}(\prec). \tag{2}$$

For then, from claims (1) and (2), we infer that $T \nvdash \mathsf{RFN}_{\Pi^1_1}(T)$.

This proof is analogous to a folklore proof of a different version of the second incompleteness theorem, namely, that no $\Pi^0_2$-sound and $\Sigma^0_1$-definable theory $T$ proves its own $\Pi^0_2$-soundness. In this folklore proof one first defines a recursive function $f_T$ that is not provably total in $T$ by diagonalizing against the set of provably total recursive functions of $T$; one then shows that the totality of $f_T$ is $T$-provably equivalent to the $\Pi^0_2$-soundness of $T$. See [3] for a detailed proof. Just as the class of the provably total recursive functions of $T$ is the canonical "invariant" measuring the $\Pi^0_2$-strength of $T$, the proof-theoretic ordinal of $T$ is the canonical "invariant" measuring the $\Pi^1_1$-strength of $T$. And just as the non-provability of $\Pi^0_2$-soundness is derived by defining a total recursive function in terms of the canonical $\Pi^0_2$-invariant, we derive Theorem 1.3 by defining a well-ordering in terms of the canonical $\Pi^1_1$-invariant.

A slight modification of our proof of Theorem 1.3 delivers a stronger result:

THEOREM 1.4. *There is no sequence $(T_n)_{n<\omega}$ of $\Pi^1_1$-sound and $\Sigma^1_1$-definable extensions of $\Sigma^1_1$-$\mathsf{AC}_0$ such that for each $n$, $T_n \vdash \mathsf{RFN}_{\Pi^1_1}(T_{n+1})$.*

To see that Theorem 1.4 implies Theorem 1.3, note that if $T$ were a counter-example to Theorem 1.3 then we would get a counter-example to Theorem 1.4 by letting $T = T_n$ for each $n$. Theorem 1.4 extends earlier work [8, 9] of Pakhomov and the author, who proved the following:

THEOREM 1.5 (Pakhomov–W.). *There is no sequence $(T_n)_{n<\omega}$ of $\Pi^1_1$-sound and $\Sigma^0_1$-definable extensions of $\mathsf{ACA}_0$ such that for each $n$, $T_n \vdash \mathsf{RFN}_{\Pi^1_1}(T_{n+1})$.*

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

Pakhomov and the author proved Theorem 1.5 to provide an explanation for the apparent pre-well-ordering of natural theories by proof-theoretic strength; see [17] for a discussion of this phenomenon. Theorem 1.4 extends this explanation to the new setting of $\Sigma^1_1$-definable theories. Whereas the proof of Theorem 1.5 appeals to the second incompleteness theorem and makes no mention of proof-theoretic ordinals, our proof of Theorem 1.4 uses ordinal analysis and does not appeal to any version of the second incompleteness theorem. Note that Theorem 1.4 is neither stronger nor weaker than Theorem 1.5; the former requires the stronger hypothesis that $T$ extend $\Sigma^1_1$-$\mathsf{AC}_0$ but only the weaker hypothesis that $T$ be $\Sigma^1_1$-definable.

The main tool that we use to derive Theorems 1.3 and 1.4 is Spector's $\Sigma^1_1$-bounding theorem. Though the standard proofs (e.g., [14, Chapter 1, Corollary 5.5]) of Spector's theorem rely on diagonalization, there is an alternate diagonalization-free proof due to Beckmann and Pohlers [2]. This latter proof derives $\Sigma^1_1$-bounding from an analysis of cut-free infinitary derivations. In particular, $\Sigma^1_1$-bounding is derived from a result known as "the boundedness theorem," which roughly states that for arithmetically definable well-orders $\prec$, the order-type of $\prec$ cannot exceed the depth of the shortest proof of the well-foundedness of $\prec$ in $\omega$-logic. Versions of the boundedness theorem are already implicit in Gentzen's proof [5] that $\mathsf{PA}$ does not prove the primitive recursive well-foundedness of $\varepsilon_0$. Note that Gentzen's proof of this independence result does *not* appeal to Gödel's second incompleteness theorem and does not rely on self-reference but rather involves a combinatorial analysis of proofs in $\mathsf{PA}$.

We also rely on a *formalized* version of Spector's theorem. Roughly, the formalized version says that for any $\Sigma^1_1$ predicate $H$, if $\mathsf{ACA}_0$ proves "$H$ is a set of recursive ordinals," then for some $e$, $\mathsf{ACA}_0$ proves "$e$ is a recursive ordinal but $\neg H(e)$." The standard proof of the formalized version of $\Sigma^1_1$-bounding uses the recursion theorem to define a recursive function. Though this is not exactly the construction of a self-referential sentence, definitions using the recursion theorem share the opacity of constructions of sentences using the fixed point lemma. Accordingly, we provide a new proof of this formalized version of Spector's theorem. The new proof uses the same techniques Gentzen used to prove the boundedness theorem and that Beckmann and Pohlers used to prove the $\Sigma^1_1$-bounding theorem. Thus, we do not rely on self-reference or diagonalization in any form.

Here is our plan for the rest of the paper. In Section 2 we cover some preliminary material, including definitions and notation; we also discuss two folklore results that we use to prove the main theorems. In Section 3 we provide proofs of Theorems 1.3 and 1.4. In Section 4 we provide an alternate proof using infinitary derivations of the formalized version of Spector's $\Sigma^1_1$-bounding theorem. Finally, in Section 5 we present some open problems concerning the optimality of Theorem 1.3.

**§2. Preliminaries.** One of the central concepts in proof theory is that of a proof-theoretic ordinal. To say what proof-theoretic ordinals are, we must say what a presentation of an ordinal is.

Definition 2.1. For a syntactic complexity class $\Gamma$, a $\Gamma$ *presentation* of an ordinal $\alpha$ is a $\Gamma$ formula that defines an ordering of order-type $\alpha$ over the standard structure $\big(\mathbb{N}, \mathcal{P}(\mathbb{N})\big)$.

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

We now present two definitions of "proof-theoretic ordinal," both of which are necessary for our proof.

DEFINITION 2.2. Let $|T|_{\mathsf{RE}}$ be the supremum of the ordinals $\alpha$ for which there is some $\Sigma^0_1$ presentation $\prec$ of $\alpha$ such that $T \vdash \mathsf{WF}(\prec)$.

DEFINITION 2.3. Let $|T|_{\mathsf{AN}}$ be the supremum of the ordinals $\alpha$ for which there is some $\Sigma^1_1$ presentation $\prec$ of $\alpha$ such that $T \vdash \mathsf{WF}(\prec)$.

The AN in the notation $|T|_{\mathsf{AN}}$ means *analytic*. Indeed, $|T|_{\mathsf{AN}}$ is the supremum of the $T$-provably well-founded *analytic* linear orders, where analytic means lightface $\Sigma^1_1$.

By definition, $|T|_{\mathsf{RE}} \leqslant |T|_{\mathsf{AN}} \leqslant \omega_1^{\mathsf{CK}}$. However, we can say more about the relationship between these three values if we make some assumptions about $T$. In the following subsections we will describe these results, which belong to mathematical folklore.

**2.1. The first folklore result.** If $T$ is $\Pi^1_1$-sound and $\Sigma^1_1$-definable, we can say more about the relationship between $|T|_{\mathsf{AN}}$ and $\omega_1^{\mathsf{CK}}$.

THEOREM 2.4 (Folklore). *If $T$ is $\Pi^1_1$-sound and $\Sigma^1_1$-definable, then $|T|_{\mathsf{RE}} < \omega_1^{\mathsf{CK}}$.*

Theorem 2.4 follows immediately from Spector's $\Sigma^1_1$-bounding theorem:

THEOREM 2.5 (Spector). *For any $\Sigma^1_1$ presentation $\prec$ of an ordinal*, $\mathsf{otyp}(\prec) < \omega_1^{\mathsf{CK}}$.

Standard proofs of Spector's $\Sigma^1_1$-bounding appeal to the fact that Kleene's $\mathcal{O}$ is not $\Sigma^1_1$-definable. Note that the latter is typically proved using an ordinary diagonalization argument; see, e.g., [14, Chapter 1, Theorem 5.4].

However, there is an alternate proof due to Beckmann and Pohlers [2] of Spector's Theorem that does *not* use diagonalization. The Beckmann–Pohlers proof proceeds by analyzing the structure of infinitary cut-free derivations. Beckmann and Pohlers derive $\Sigma^1_1$-bounding a result known as "the boundedness lemma," which they claim is essentially implicit in Gentzen's proof of the PA non-derivability of $\varepsilon_0$-induction. Accordingly, when we appeal to Theorem 2.4, we are appeal to a result that has a Gentzen-style non-diagonalization proof.

**2.2. The second folklore result.** We can say more about the relationship between $|T|_{\mathsf{RE}}$ and $|T|_{\mathsf{AN}}$ for all $\Pi^1_1$-sound $T$ that extend $\Sigma^1_1$-$\mathsf{AC}_0$. Recall that $\Sigma^1_1$-$\mathsf{AC}_0$ is the theory whose axioms are those of $\mathsf{ACA}_0$ plus each instance of the schema:

$$\forall n \exists X \varphi(n, X) \rightarrow \exists Y \forall n \varphi\big(n, (Y)_n\big),$$

where $\varphi(n, X)$ is a $\Sigma^1_1$ formula in which $Y$ does not occur and where

$$(Y)_n = \{i \mid (i, n) \in Y\}.$$

THEOREM 2.6 (Folklore). *If $T$ is a $\Pi^1_1$-sound extension of $\Sigma^1_1$-$\mathsf{AC}_0$, then* $|T|_{\mathsf{RE}} = |T|_{\mathsf{AN}}$.

The main tool for proving Theorem 2.6 is a *formalized* version of Spector's theorem. To state this formalized result, we first introduce some notation.

DEFINITION 2.7. Let $Rec := \{e \in \mathbb{N} \mid e \text{ is an index of a total recursive function}\}$. With each $e \in Rec$ there is an associated relation $\prec_e$ where

$$n \prec_e m :\Leftrightarrow \{e\}(\langle n, m\rangle) = 0,$$

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

where $\langle , \rangle$ is a primitive recursive pairing function.

Definition 2.8. Let $\mathfrak{W}_{Rec} := \{e \in \mathbb{N} \mid e \in Rec \wedge \mathsf{WO}(\prec_e)\}$.

In [13] (see Proposition 2.19), Rathjen derives Theorem 2.6 from the following lemma, which is a formalized version of Spector's $\Sigma^1_1$-bounding theorem:

Lemma 2.9 (Rathjen). *Suppose $H(x)$ is a $\Sigma^1_1$ formula such that*:

$$\mathsf{ACA}_0 \vdash \forall x\big(H(x) \rightarrow x \in \mathfrak{W}_{Rec}\big).$$

*Then, for some $e \in Rec$*,

$$\mathsf{ACA}_0 \vdash e \in \mathfrak{W}_{Rec} \wedge \neg H(e).$$

Theorem 2.6 is straightforwardly derived from Lemma 2.9. Rathjen provides a proof of Lemma 1.1 in [12]. Note that this proof of Lemma 2.9 makes use of the recursion theorem. In Section 4 we will present an alternative proof of Lemma 2.9 that does not make any use of the recursion theorem or other diagonalization.

**2.3. Remarks.** Before continuing, let's highlight some features of these folklore results and their relationship to the main theorems.

First, note the role that $\Sigma^1_1$-bounding plays in the proofs of the folklore results. We will not mention $\Sigma^1_1$-bounding explicitly in the proofs of the main theorems, but we will still rely on it insofar as it is used to prove these folklore results. We feel that the role of $\Sigma^1_1$-bounding is so important that it is worth explicitly highlighting where it is being used.

Second, note that in the proofs of both folklore results, we must appeal to the $\Pi^1_1$-soundness of $T$. We will not appeal to $\Pi^1_1$-soundness explicitly in the proofs of the main theorems; we will only rely on it insofar as we invoke these folklore results.

One can find proofs of Theorems 2.4 and 2.6 in [13].

**§3. The main theorems.** In this section we prove our main theorem, an analogue of Gödel's second incompleteness theorem. We start by introducing two formulas and make a remark about their syntactic complexity. We will use these formulas and appeal to the remark many times, so it is worth isolating them here.

Definition 3.1. For a binary formula $\lhd$, let $\mathsf{LO}(\lhd)$ stand for the conjunction of the following clauses:

(1) $\neg\exists x \mathsf{True}_{\Sigma^0_1}(x \lhd x)$.
(2) $\forall x \forall y\big(\mathsf{True}_{\Sigma^0_1}(x \lhd y) \vee \mathsf{True}_{\Sigma^0_1}(y \lhd x) \vee x = y\big)$.
(3) $\forall x \forall y \forall z\Big(\big(\mathsf{True}_{\Sigma^0_1}(x \lhd y) \wedge \mathsf{True}_{\Sigma^0_1}(y \lhd z)\big) \rightarrow \mathsf{True}_{\Sigma^0_1}(x \lhd z)\Big)$.

Definition 3.2. For a binary formula $\lhd$, let $\mathsf{WF}(\lhd)$ stand for

$$\forall X\big(\exists x \in X \rightarrow \exists x \in X\ \forall y \in X\ \neg\mathsf{True}_{\Sigma^0_1}(y \lhd x)\big).$$

Remark 3.3. Note the use of the $\Sigma^0_1$ truth-predicate in Definitions 3.1 and 3.2. Thus, for any formula $\lhd$, $\mathsf{LO}(\lhd)$ is arithmetic and $\mathsf{WF}(\lhd)$ is $\Pi^1_1$. Of course, $\mathsf{LO}(\lhd)$ and $\mathsf{WF}(\lhd)$ will make the most sense when applied to $\Sigma^0_1$ formulas or in quantified statements about $\Sigma^0_1$ formulas. We shall use it in the latter way.

**3.1. The key lemma.** The key to the proofs of Theorems 1.3 and 1.4 is the following lemma:

LEMMA 3.4. *If $T$ is $\Pi^1_1$-sound and $\Sigma^1_1$-definable, then there is a $\Sigma^1_1$ presentation $\prec_T$ of $|T|_{\mathsf{RE}}$ such that $\Sigma^1_1$-$\mathsf{AC}_0 \vdash \mathsf{RFN}_{\Pi^1_1}(T) \rightarrow \mathsf{WF}(\prec_T)$.*

PROOF. Let $T$ be $\Pi^1_1$-sound and $\Sigma^1_1$-definable. By Theorem 2.4, $|T|_{\mathsf{RE}} < \omega_1^{\mathsf{CK}}$, whence there is some $\Sigma^0_1$-definable $\prec_\star$ such that $|T|_{\mathsf{RE}} = \mathsf{otyp}(\prec_\star)$.

We are now going to define an alternate presentation $\prec_T$ of $|T|_{\mathsf{RE}}$. Informally, the formula $\alpha \prec_T \beta$ says that $\alpha$ is less than $\beta$ in an initial segment of the $\prec_\star$ ordering that embeds into a $\Sigma^0_1$-definable linear order $\lhd$ such that $T$ proves the well-foundedness of $\lhd$. More formally, we define $\alpha \prec_T \beta$ as the conjunction of

(1) $\alpha \prec_\star \beta$,
(2) $\exists \lhd \in \Sigma^0_1\ \exists f\ \Big(\mathsf{Emb}(f, \prec_\star\restriction \beta, \lhd)\ \wedge \mathsf{LO}(\lhd)\ \wedge\ \mathsf{Pr}_T\big(\mathsf{WF}(\lhd)\big)\Big)$,

where $\mathsf{Emb}(f, \prec_\star\restriction \beta, \lhd)$ stands for

$$\forall x \forall y \Big((y \preceq_\star \beta \wedge x \prec_\star y) \rightarrow \mathsf{True}_{\Sigma^0_1}\big(f(x) \lhd f(y)\big)\Big),$$

and where $\mathsf{LO}(\lhd)$ and $\mathsf{WF}(\lhd)$ are as in Definitions 3.1 and 3.2.

CLAIM. *$\prec_T$ is $\Sigma^1_1$-$\mathsf{AC}_0$-provably equivalent to a $\Sigma^1_1$ formula.*

Clearly $\alpha \prec_\star \beta$ is $\Sigma^0_1$. Now let's look at the second conjunct of $\alpha \prec_T \beta$. Note that $\mathsf{Emb}(f, \prec_\star\restriction \beta, \lhd)$ and $\mathsf{LO}(\lhd)$ are both arithmetic. On the other hand, $\mathsf{Pr}_T\big(\mathsf{WF}(\lhd)\big)$ is $\Sigma^1_1$, since $T$ is $\Sigma^1_1$-definable. So the conjunction

$$\mathsf{Emb}(f, \prec_\star\restriction \beta, \lhd)\ \wedge \mathsf{LO}(\lhd)\ \wedge\ \mathsf{Pr}_T\big(\mathsf{WF}(\lhd)\big)$$

is provably equivalent in $\Sigma^1_1$-$\mathsf{AC}_0$ to a $\Sigma^1_1$ formula. Thus, the second conjunct of $\alpha \prec_T \beta$ is given by an existential number quantifier before an existential set quantifier before a (formula that is $\Sigma^1_1$-$\mathsf{AC}_0$-provably equivalent to a) $\Sigma^1_1$ formula. It follows that $\prec_T$ is $\Sigma^1_1$-$\mathsf{AC}_0$-provably equivalent to a $\Sigma^1_1$ formula.

CLAIM. *$\prec_T$ is a presentation of $|T|_{\mathsf{RE}}$.*

$\mathsf{otyp}(\prec_T) \leqslant |T|_{\mathsf{RE}}$: The first conjunct in the definition of $\prec_T$ ensures that $\mathsf{otyp}(\prec_T) \leqslant \mathsf{otyp}(\prec_\star)$. To finish the argument, recall that $\mathsf{otyp}(\prec_\star) = |T|_{\mathsf{RE}}$.

$\mathsf{otyp}(\prec_T) \geqslant |T|_{\mathsf{RE}}$: Let $\alpha < |T|_{\mathsf{RE}} = \mathsf{otyp}(\prec_\star)$. We need to see that $\alpha < \mathsf{otyp}(\prec_T)$. Since $\alpha < \mathsf{otyp}(\prec_\star)$ and $\alpha < |T|_{\mathsf{RE}}$, there is an embedding of an initial segment of $\prec_\star$ that includes the $\prec_\star$ representation of $\alpha$ into a $\Sigma^0_1$-definable well-order that is $T$-provably well-founded. It is then immediate from the definition of $\prec_T$ that $\alpha < \mathsf{otyp}(\prec_T)$.

CLAIM. $\Sigma^1_1$-$\mathsf{AC}_0 \vdash \mathsf{RFN}_{\Pi^1_1}(T) \rightarrow \mathsf{WF}(\prec_T)$.

Reason in $\Sigma^1_1$-$\mathsf{AC}_0$: Suppose that $\prec_T$ is ill-founded. Then, by the definition of $\prec_T$, there is some infinite descending sequence in $\prec_\star$ that embeds into a $\Sigma^0_1$-definable linear order $\lhd$ such that $T \vdash \mathsf{WF}(\lhd)$. Since $\lhd$ embeds an ill-founded linear order, $\lhd$ is ill-founded. So $T$ proves a false $\Pi^1_1$ sentence, namely, $\mathsf{WF}(\lhd)$. ⊣

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

**3.2. An incompleteness theorem.** We now present a proof of Theorem 1.3, restated here:

THEOREM 3.5. *If $T$ is a $\Pi^1_1$-sound and $\Sigma^1_1$-definable extension of* $\Sigma^1_1\text{-}\mathsf{AC}_0$, *then $T$ does not prove its own $\Pi^1_1$-soundness.*

PROOF. By Lemma 3.4, there is a $\Sigma^1_1$ presentation $\prec_T$ of $|T|_{\mathsf{RE}}$ such that

$$\Sigma^1_1\text{-}\mathsf{AC}_0 \vdash \mathsf{RFN}_{\Pi^1_1}(T) \to \mathsf{WF}(\prec_T).$$

Since $T$ extends $\Sigma^1_1\text{-}\mathsf{AC}_0$, we infer that

$$T \vdash \mathsf{RFN}_{\Pi^1_1}(T) \to \mathsf{WF}(\prec_T). \tag{3}$$

CLAIM. $T \nvdash \mathsf{WF}(\prec_T)$.

Since $T$ extends $\Sigma^1_1\text{-}\mathsf{AC}_0$, by Theorem 2.6, $|T|_{\mathsf{RE}} = |T|_{\mathsf{AN}}$. Moreover, since $T$ extends $\Sigma^1_1\text{-}\mathsf{AC}_0$, we infer that $\prec_T$ is $T$-provably equivalent to a $\Sigma^1_1$ formula. So $\prec_T$ is a presentation of $|T|_{\mathsf{AN}}$ that is $T$-provably equivalent to a $\Sigma^1_1$ formula, whence $T \nvdash \mathsf{WF}(\prec_T)$.

It follows immediately from (3) and from the claim that $T \nvdash \mathsf{RFN}_{\Pi^1_1}(T)$. ⊣

**3.3. Well-foundedness.** In this subsection we prove a strengthening of Theorem 3.5 that is of independent interest. The following result is proved in [8, 9]:

THEOREM 3.6 (Pakhomov–W.). *There is no sequence $(T_n)_{n<\omega}$ of $\Pi^1_1$-sound and $\Sigma^0_1$-definable extensions of* $\mathsf{ACA}_0$ *such that for each $n$, $T_n \vdash \mathsf{RFN}_{\Pi^1_1}(T_{n+1})$.*

Pakhomov and the author proved Theorem 3.6 to provide an explanation of the apparent pre-well-ordering of natural theories by proof-theoretic strength; see [17] for a discussion of this phenomenon. In [8, 9], Theorem 3.6 is proved using Gödel's second incompleteness theorem. In particular, we show that the theory $\mathsf{ACA}_0 + \varphi$, where $\varphi$ states that Theorem 3.6 is false, proves its own consistency. In [7] it is claimed that such a result "could be proved by showing that a descending sequence $(T_n)_{n<\omega}$ of theories would induce a descending sequence in the ordinals (namely, the associated sequence of proof-theoretic ordinals)." We now present such a proof (though for $\Sigma^1_1$-definable extensions of $\Sigma^1_1\text{-}\mathsf{AC}_0$ rather than for $\Sigma^0_1$-definable extensions of $\mathsf{ACA}_0$).

What follows is a restatement of Theorem 1.4:

THEOREM 3.7. *There is no sequence $(T_n)_{n<\omega}$ of $\Pi^1_1$-sound and $\Sigma^1_1$-definable extensions of* $\Sigma^1_1\text{-}\mathsf{AC}_0$ *such that for each $n$, $T_n \vdash \mathsf{RFN}_{\Pi^1_1}(T_{n+1})$.*

PROOF. Suppose that there is such a sequence $(T_n)_{n<\omega}$. From Lemma 3.4 we infer that, for each $n$, there is a $\Sigma^1_1$ presentation $\prec_{T_n}$ of $|T_n|_{\mathsf{RE}}$ such that

$$\Sigma^1_1\text{-}\mathsf{AC}_0 \vdash \mathsf{RFN}_{\Pi^1_1}(T_n) \to \mathsf{WF}(\prec_{T_n}).$$

Since each $T_n$ extends $\Sigma^1_1\text{-}\mathsf{AC}_0$, Theorem 2.6 entails that, for each $n$, there is a $\Sigma^1_1$ presentation $\prec_{T_n}$ of $|T_n|_{\mathsf{AN}}$ such that

$$\Sigma^1_1\text{-}\mathsf{AC}_0 \vdash \mathsf{RFN}_{\Pi^1_1}(T_n) \to \mathsf{WF}(\prec_{T_n}).$$

Since each $T_n$ extends $\Sigma^1_1$-$\mathsf{AC}_0$, for each $n$,

$$T_n \vdash \mathsf{RFN}_{\Pi^1_1}(T_{n+1}) \to \mathsf{WF}(\prec_{T_{n+1}}).$$

By assumption, for each $n$, $T_n \vdash \mathsf{RFN}_{\Pi^1_1}(T_{n+1})$, so we infer that, for each $n$,

$$T_n \vdash \mathsf{WF}(\prec_{T_{n+1}}).$$

Whence $|T_n|_{\mathsf{AN}} > |T_{n+1}|_{\mathsf{AN}}$ for each $n$. Yet $|T_n|_{\mathsf{AN}}$ is an ordinal for each $n$. So $(|T_n|_{\mathsf{AN}})_{n<\omega}$ is a descending sequence in the ordinals. ⊣

Note that Theorem 3.7 entails Theorem 3.5. Indeed, if $T$ were a counter-example to Theorem 3.5 then we would get a counter-example to Theorem 3.7 by letting $T = T_n$ for each $n$.

**§4. Avoiding diagonalization.** Considering the motivations outlined in Section 1, it is desirable to avoid diagoanlization in the proofs of Theorems 2.4 and 2.6.

As discussed in Section 2.1, the standard proofs of Spector's $\Sigma^1_1$-bounding theorem rely on diagonalization. However, there is already an alternate proof of Spector's theorem due to Beckmann and Pohlers [2] that uses Gentzen's methods and avoids diagonalization.

Theorem 2.6, on the other hand, relies on Lemma 2.9, which is a formalized version of $\Sigma^1_1$-bounding. Rathjen's proof of Lemma 2.9 uses the recursion theorem to formalize the standard proof of $\Sigma^1_1$-bounding. In this section we develop an alternate proof of Lemma 2.9. Rather than an attempt to formalize the diagonalization proof of $\Sigma^1_1$-bounding in a different way, we instead formalize the Beckmann–Pohlers proof.

**4.1. Infinitary derivations.** The Beckmann–Pohlers proof of $\Sigma^1_1$-bounding involves the analysis of derivations in a cut-free infinitary proof system. We provide here a standard definition of such a proof system; for other discussion of such proof systems, see [10, 11]. Note that this proof system is a version of the Tait calculus. Thus, our proof system deals with formulas within which negation is only appended to atomic formulas; this is possible due to the normal form theorems available in classical logic. In what follows, let $\mathsf{Diag}(\mathbb{N})$ be the atomic diagram of $\mathbb{N}$ in the signature $(0, 1, +, \times)$.

DEFINITION 4.1. We define $\vdash^\alpha \Delta$ inductively by the following clauses:

(AxM) If $\Delta \cap \mathsf{Diag}(\mathbb{N}) \neq \emptyset$, then $\vdash^\alpha \Delta$ for all ordinals $\alpha$.
(AxL) If $t^{\mathbb{N}} = s^{\mathbb{N}}$, then $\vdash^\alpha \Delta, s \notin X, t \in X$ for all ordinals $\alpha$.
($\bigwedge$) If $\vdash^{\alpha_i} \Delta, A_i$ and $\alpha_i < \alpha$ for $i = 1, 2$, then $\vdash^\alpha \Delta, A_1 \wedge A_2$.
($\bigvee$) If $\vdash^{\alpha_i} \Delta, A_i$ and $\alpha_i < \alpha$ for some $i \in \{1, 2\}$, then $\vdash^\alpha \Delta, A_1 \vee A_2$.
($\forall$) If $\vdash^{\alpha_i} \Delta, A(i)$ and $\alpha_i < \alpha$ for all $i \in \mathbb{N}$, then $\vdash^\alpha \Delta, \forall x A(x)$.
($\exists$) If $\vdash^{\alpha_i} \Delta, A(i)$ and $\alpha_i < \alpha$ for some $i \in \mathbb{N}$, then $\vdash^\alpha \Delta, \exists x A(x)$.

The relation $\vdash^\alpha \Delta$ is to be read that there is an infinite proof tree of $\bigvee \Delta$ whose depth is bounded by the ordinal $\alpha$.

Let's briefly record two lemmas that we will make use of. First we state the "monotonicity lemma," which follows immediately from the definition of $\vdash^\alpha \Delta$:

LEMMA 4.2. *If* $\vdash^\alpha \Delta$, $\alpha \leqslant \beta$, *and* $\Delta \subseteq \Gamma$, *then* $\vdash^\beta \Gamma$.

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

Second, we state the $\wedge$-inversion rule. For a proof of the $\wedge$-inversion rule, see [10, Theorem 10.7].

LEMMA 4.3. *If* $\vdash^{\alpha} \Delta, \bigwedge\{A_i \mid i \in I\}$ *then, for all* $i \in I$, $\vdash^{\alpha} A_i$.

The infinitary proof calculus is sound and complete for $\Pi^1_1$ sentences of arithmetic:

THEOREM 4.4. *For any* $\Pi^1_1$ *sentence* $\forall \vec{X} \varphi(\vec{X})$,

$$\mathbb{N} \vDash \forall \vec{X} \varphi(\vec{X}) \Leftrightarrow \exists \alpha < \omega_1^{\mathsf{CK}} \vdash^{\alpha} F(\vec{X}).$$

In fact, there is a sharp restricted version of the completeness half of the theorem relating consequences of $\mathsf{ACA}_0$ and proofs of height less than $\varepsilon_0$:

THEOREM 4.5. *For any* $\Pi^1_1$ *sentence* $\forall \vec{X} \varphi(\vec{X})$,

$$\mathsf{ACA}_0 \vdash \forall \vec{X} \varphi(\vec{X}) \Rightarrow \exists \alpha < \varepsilon_0 \vdash^{\alpha} \varphi(\vec{X}).$$

Note that the definition of infinitary derivations makes use of transfinite recursion and is beyond the scope $\mathsf{ACA}_0$. Nevertheless, there are many methods for formalizing infinitary derivations in such a way that appropriate versions of Lemmas 4.2, 4.3 and Theorem 4.5 are provable in $\mathsf{ACA}_0$, all without recourse to the fixed point lemma or recursion theorem. For present purposes, we will need to formalize only those infinitary derivations whose depth is less than $\varepsilon_0$, a rather meager class of infinitary derivations. We will turn to the specifics in the next subsection.

**4.2. Formalizing infinitary derivations.** In this subsection we turn to the task of formalization in $\mathsf{ACA}_0$. This task has two components. First, we must describe how it is that we *define* infinitary proofs in $\mathsf{ACA}_0$. Second, we must describe how it is that we *reason* about infinitary proofs in $\mathsf{ACA}_0$. A necessary pre-condition for completing both tasks is fixing an ordinal notation system.

REMARK 4.6. We *fix* a nice ordinal notation system for ordinals up to and including (at least) $2^{\varepsilon_0} + 1$. We use the symbols $\{<, >, \leqslant, \geqslant\}$ for this ordinal notation system. In the remainder of this section of the paper, when we use these symbols we are using them to refer to this fixed ordinal notation system.

The basic idea behind our definition of infinitary proofs in $\mathsf{ACA}_0$ is that infinitary proofs are $\omega$-branching trees. Each node in the proof is *tagged* with a sequent, ordinal notation, and a rule:

DEFINITION 4.7. Let $\mathsf{SEQ}$ be the set of finite sequents, i.e., sets of formulas in (Tait calculus) normal form in the signature $(0, 1, +, \times)$. Let

$$\mathsf{RULE} = \{\mathsf{AxM}, \mathsf{AxL}, \wedge, \vee, \forall, \exists, \mathsf{CUT}, \mathsf{REP}\}.$$

We demand that the trees satisfy *local correctness conditions*. The local correctness conditions merely say that if a node is tagged with a sequent $\Delta$ and rule $R$, then the premises of that node are tagged with sequents that are correct for the rule $R$. Buchholz essentially introduces these local correctness conditions (changed only slightly here) in [4, Definitions 2.1–2.3].

DEFINITION 4.8. Let $(\Delta, R) \subseteq \mathsf{SEQ} \times \mathsf{RULE}$ and let $(\Delta)_{i \in I}$ be a sequence of sequents (the premises of $\Delta$). We say that $(\Delta, R)$ and $(\Delta)_{i \in I}$ jointly satisfy the local correctness conditions if each of the following holds:

(AxM) If $R = \mathsf{AxM}$ then $\Delta \cap \mathsf{Diag}(\mathbb{N}) \neq \emptyset$.

(AxL) If $R = \mathsf{AxL}$ then there are $t^{\mathbb{N}} = s^{\mathbb{N}}$ such that $s \notin X, t \in X \in \Delta$.

($\wedge$) If $R = \wedge$ then $I = \{1,2\}$ and for some $A_1$ and $A_2$:

$$A_1 \wedge A_2 \in \Delta \text{ and for all } i \in \{1,2\}, \Delta_i \subseteq \Delta, A_i.$$

($\wedge$) If $R = \vee$ then $I \subseteq \{1,2\}$ and for some $A_1$ and $A_2$,

$$A_1 \vee A_2 \in \Delta \text{ and for some } i \in \{1,2\}, \Delta_i \subseteq \Delta, A_i.$$

($\forall$) If $R = \forall$ then $I = \mathbb{N}$ and for some $\forall x A(x)$,

$$\forall x A(x) \in \Delta \text{ and for all } i \in \mathbb{N}, \Delta_i \subseteq \Delta, A(i).$$

($\exists$) If $R = \exists$ then $I \subseteq \mathbb{N}$ and for some $\exists x A(x)$,

$$\exists x A(x) \in \Delta \text{ and for some } i \in \mathbb{N}, \Delta_i \subseteq \Delta, A(i).$$

(CUT) If $R = \mathsf{CUT}$ then $I = \{1,2\}$ and for some $A$,

$$\Delta_1 \subseteq \Delta, A \text{ and } \Delta_2 \subseteq \Delta, \neg A.$$

(REP) If $R = \mathsf{REP}$ then $I = \{1\}$ and $\Delta_1 = \Delta$.

Definition 4.9 ($\mathsf{ACA}_0$). An *infinitary proof* is an $\omega$-branching tree where each node is labeled by a triple $(\Delta, R, \alpha)$ consisting of a sequent $\Delta$, rule $R$, and ordinal notation $\alpha$ such that:

(1) The ordinal labels strictly descend from the root towards the axioms.
(2) The local correctness conditions from Definition 4.8 are satisfied.

Note that (1) does not force the ordinal tags to be exact but merely to give bounds.

We are particular interested in those infinitary proofs in which the rule $\mathsf{CUT}$ is not applied. We write $\vdash^{\alpha}_{\Delta}$ if the sequent $\Delta$ has such an infinitary proof wherein the root has ordinal tag $\alpha$.

Now we turn to the task of formalizing reasoning about infinitary derivations in $\mathsf{ACA}_0$. One particularly elegant way of formalizing such reasoning is due to Buchholz [4]. We fix a standard embedding $f$ of proofs of $\Pi^1_1$ statements in $\mathsf{ACA}_0$ into infinitary derivations in $\omega$-logic. Using our ordinal notation system $<$ that includes a representation of $\varepsilon_0$, we can define a term system for those infinitary derivations that arise from $f$. The term of a proof in this term system encodes the information in its root, i.e., its sequent, the rule it was inferred with, and an ordinal bound. The definition of this term system for infinitary derivations uses primitive recursion but does not use the fixed point lemma or recursion theorem.

Remark 4.10. An infinitary proof is coded by the label of its root. Buchholz shows that there are primitive recursive functions that can be used to compute, from the code of (the root of) a proof $P$, the codes of $P$'s subtrees. Accordingly, we can use $\mathsf{ACA}_0$ (and even $\mathsf{RCA}_0$) to construct a proof tree from its code. Moreover, we will be able to prove in $\mathsf{ACA}_0$ that the defined tree satisfies the definition of an infinitary proof by the way Buchholz sets up his term system for infinitary derivations.

In Definition 4.9 we did not require that the ordinal tags are exact but merely that they are bounds. Hence, the new version of Lemma 4.2 is trivial:

LEMMA 4.11. *For any* $\alpha < \varepsilon_0$, $\mathsf{ACA}_0$ *proves "if* $\vdash^\alpha \Delta$, $\alpha \leqslant \beta$, *and* $\Delta \subseteq \Gamma$, *then* $\vdash^\beta \Gamma$."

For the analogue of Lemma 4.3, we refer the reader to the proof of Theorem 10.7 in [10]. Note that Pohlers proves Theorem 10.7 by induction along $\alpha$. In the following we can follow suit since we are assuming $\alpha < \varepsilon_0$.

LEMMA 4.12. *For any* $\alpha < \varepsilon_0$, $\mathsf{ACA}_0$ *proves "if* $\vdash^\alpha \Delta, \bigwedge\{A_i \mid i \in I\}$ *then, for all* $i \in I, \vdash^\alpha A_i$."

Finally, we note that the following version of Theorem 4.5 follows easily given how we have set things up:

THEOREM 4.13. *For any* $\Pi^1_1$ *sentence* $\forall \vec{X} \varphi(\vec{X})$, *if* $\mathsf{ACA}_0 \vdash \forall \vec{X} \varphi(\vec{X})$ *then for some* $\alpha < \varepsilon_0$, $\mathsf{ACA}_0$ *proves that* $\vdash^\alpha \varphi(\vec{X})$.

Indeed, if $\mathsf{ACA}_0 \vdash \forall \vec{X} \varphi(\vec{X})$ then, through the usual embedding $f$ of $\mathsf{ACA}_0$ proofs into $\omega$-logic, $\mathsf{ACA}_0$ can prove that there is an infinitary derivation with height $< \varepsilon_0$ of $\varphi(\vec{X})$. We get a term for this proof in Buchholz's term system and then, by Remark 4.10, we use it to construct an $\omega$-proof of height $\alpha$ of $\varphi(\vec{X})$, all in $\mathsf{ACA}_0$.

Before continuing, we want to note that Lemmas 4.11 and 4.12 and Theorem 4.13 are all proved without recourse to the recursion theorem or self-reference.

**4.3. The boundedness lemma.** We need to check that a version of what is called "the boundedness lemma" is provable in $\mathsf{ACA}_0$. Beckmann and Pohlers claim that a version of the boundedness lemma is already implicit in Gentzen's [5] proof of the $\mathsf{PA}$ non-derivability of $\varepsilon_0$-induction. To state this result, let us recall one definition that occurs frequently in the work of Pohlers.[1]

DEFINITION 4.14. The truth complexity $\mathsf{tc}\big(\forall \vec{X} \varphi(\vec{X})\big)$ of a $\Pi^1_1$ statement $\forall \vec{X} \varphi(\vec{X})$ is the least $\alpha$ such that $\vdash^\alpha \varphi(\vec{X})$.

Before continuing we will also fix some notation. We let

$$\mathsf{field}(\prec) := \{x \mid \exists y (x \prec y \vee x \prec y)\},$$
$$\mathsf{Prog}(\prec, X) := \forall x \Big( \big(x \in \mathsf{field}(\prec) \wedge \forall y (y \prec x \rightarrow y \in X)\big) \rightarrow x \in X \Big),$$
$$\mathsf{TI}(\prec) := \forall X \big(\mathsf{Prog}(\prec, X) \rightarrow \forall x \in \mathsf{field}(\prec)\ x \in X\big).$$

Note that $\mathsf{TI}(\prec)$ expresses transfinite induction along $\prec$ and for arithmetically definable $\prec$ the sentence $\mathsf{TI}(\prec)$ is $\Pi^1_1$. An upshot of the boundedness lemma is the boundedness theorem, which establishes a tight connection between $\mathsf{otyp}(\prec)$ and $\mathsf{tc}\big(\mathsf{TI}(\prec)\big)$. Beckmann and Pohlers attribute the following consequence of the boundedness theorem to Gentzen:

THEOREM 4.15 (Gentzen). *For any arithmetic well-ordering* $\prec$,

$$\mathsf{otyp}(\prec) \leqslant 2^{\mathsf{tc}\left(\mathsf{TI}(\prec)\right)}.$$

[1]Note that sometimes (e.g., in [11]) a different definition is given and this definition is stated as a theorem. Elsewhere, as in [2], the definition given here is used.

Beckmann [2] has sharpened Gentzen's result to show that $\mathsf{otyp}(\prec) \leqslant \mathsf{tc}\big(\mathsf{TI}(\prec)\big)$, which he derives from a sharp version of the boundedness lemma. For present purposes, we will not need the sharp version. To state the version that we will need, we need to cover some definitions.

DEFINITION 4.16. A formula $\varphi$ is *$X$-positive* if $\varphi$ has no occurrences of $X$ of the form $t \notin X$.

DEFINITION 4.17. If $\varphi$ is a formula, then $\varphi[X \mapsto \psi]$ is the set of formulas we get by replacing each occurrence of $t \in X$ in $\varphi$ with $\psi(t)$. If $\Delta = \{\varphi_1, \dots, \varphi_n\}$ is a set of formulas, then $\Delta[X \mapsto \psi] = \{\varphi_1[X \mapsto \psi], \dots, \varphi_n[X \mapsto \psi]\}$.

DEFINITION 4.18. For any well-ordering $\prec$:

(1) $|n|_\prec$ is the rank of $n$ in $\prec$.
(2) $\prec_\alpha = \{n \mid |n|_\prec < \alpha\}$.

REMARK 4.19. Note that $y \in X[X \mapsto \prec_\alpha] = |y|_\prec < \alpha$.

The following lemma—the boundedness lemma—is a version of Lemma 13.9 in [10]. Our proof is essentially the same as that in Pohlers, except that Pohlers relies on some notions that are not formalizable in $\mathsf{ACA}_0$. In particular, we are careful to use partial truth-predicates rather than speak of satisfaction in $\mathbb{N}$.

LEMMA 4.20 ($\mathsf{ACA}_0$). *Let $\alpha < \varepsilon_0$ be well-founded. Let $\prec$ be an arithmetic well-ordering. Let $\Delta$ be a finite set of $X$-positive formulas. Suppose that*

$$\vdash^\alpha \neg\mathsf{Prog}(\prec, X), t_1 \notin X, \dots, t_n \notin X, \Delta.$$

*Then it follows that*

$$\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_\gamma]\big)\Big),$$

*where $\gamma = \beta + 2^\alpha$ and $\beta = \max\{|t_1|_\prec, \dots, |t_n|_\prec\}$.*

PROOF. We prove the claim by induction on $\alpha$; note that this is licit since we are assuming that $\alpha$ is well-founded. We split into cases based on the final inference in the derivation that yields $\vdash^\alpha \neg\mathsf{Prog}(\prec, X), t_1 \notin X, \dots, t_n \notin X, \Delta$. Note that we do not have to consider the inference CUT since the derivation is cut-free. Note that we also do not have to consider the inference REP; if the given proof ends with repetition we simply look at some smaller proof of the same sequent that does not end with repetition.

CASE 1: *The sequent $\neg\mathsf{Prog}(\prec, X), t_1 \notin X, \dots, t_n \notin X, \Delta$ is an axiom according to* (AxM). The set $\Delta$ contains a true atomic formula $\varphi$. Then $\varphi = \varphi[X \mapsto \prec_\gamma]$. So $\Delta[X \mapsto \prec_\gamma]$ contains a true formula, namely $\varphi = \varphi[X \mapsto \prec_\gamma]$.

CASE 2: *The sequent $\neg\mathsf{Prog}(\prec, X), t_1 \notin X, \dots, t_n \notin X, \Delta$ is an axiom according to* (AxL). $\Delta$ contains a formula $t_i \in X$ for some $i \leqslant n$. If $\beta_i = |t_i|_\prec$, then $\beta_i \leqslant \beta < \gamma$ and $\mathsf{True}_{\Pi^1_1}\big((t_i \in X)[X \mapsto \prec_\gamma]\big)$ since $\beta_i < \gamma$. Hence $\mathsf{True}_{\Pi^1_1}\big(\bigvee \Delta[X \mapsto \prec_\gamma]\big)$.

CASE 3: *The final inference yields* $\Delta$. Assume that the main formula of the final inference belongs to $\Delta$. Then we have the premises

$$\vdash^{\alpha_i} \neg\mathsf{Prog}(\prec, X), t_1 \notin X, \dots, t_n \notin X, \Delta_i,$$

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

where $\Delta_i$ contains only $X$-positive formulas. From the induction hypothesis we infer that $\forall i\ \mathsf{True}_{\Pi^1_1}\big(\forall X(\bigvee \Delta_i[X \mapsto \prec_{\gamma_i}])\big)$ where $\gamma_i = \beta + 2^{\alpha_i}$. Lemma 4.11, i.e., the monotonicity lemma, delivers

$$\forall i\ \mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta_i[X \mapsto \prec_{\gamma}]\big)\Big).$$

Appealing to Lemma 4.11 once again we infer that

$$\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{\gamma}]\big)\Big),$$

since validity is preserved by all inferences.
CASE 4: *The final inference yields* $\neg\mathsf{Prog}(\prec, X)$. The main formula of the final inference is

$$\exists x\Big(\big(x \in \mathsf{field}(\prec) \wedge \forall y(y \prec x \rightarrow y \in X)\big) \wedge x \notin X\Big).$$

Then we have the premise:

$$\vdash^{\alpha_0} \neg\mathsf{Prog}(\prec, X), t \in \mathsf{field}(\prec) \wedge \forall y(\neg y \prec t \vee y \in X) \wedge t \notin X, t_1 \notin X, \dots, t_n \notin X, \Delta.$$

By $\wedge$-inversion, i.e., Lemma 4.12, we obtain

$$\vdash^{\alpha_0} \neg\mathsf{Prog}(\prec, X), t \in \mathsf{field}(\prec), \forall y(\neg y \prec t \vee y \in X), t_1 \notin X, \dots, t_n \notin X, \Delta \tag{4}$$

and also

$$\vdash^{\alpha_0} \neg\mathsf{Prog}(\prec, X), t \notin X, t_1 \notin X, \dots, t_n \notin X, \Delta. \tag{5}$$

Assume towards a contradiction that $\neg\mathsf{True}_{\Pi^1_1}\big(\forall X(\bigvee \Delta[X \mapsto \prec_{\gamma}])\big)$.

Applying the induction hypothesis to (4) we obtain

$$\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{2^{\gamma_0}}] \vee \forall y(y \prec t \rightarrow y \in X)[X \mapsto \prec_{2^{\gamma_0}}]\big)\Big), \tag{6}$$

where $\gamma_0 = \beta + 2^{\alpha_0}$.

By Lemma 4.11

$$\neg\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{\gamma}]\big)\Big) \text{ entails } \neg\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{2^{\gamma_0}}]\big)\Big).$$

By (6) we then obtain that $y \in \prec_{2^{\gamma_0}}$ for all $y \prec t$, i.e., $|t|_{\prec} \leqslant 2^{\gamma_0}$. Letting $\beta_0 := \mathsf{max}\{|t|_{\prec}, \beta\}$, then we have $\beta_0 \leqslant \gamma_0$. Applying the induction hypothesis to (5), we obtain

$$\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{\beta_0 + 2^{\alpha_0}}]\big)\Big).$$

Note that $\beta_0 \leqslant \beta + 2^{\alpha_0}$ and also $2^{\alpha_0} + 2^{\alpha_0} \leqslant 2^{\alpha}$. Hence,

$$\beta_0 + 2^{\alpha_0} \leqslant \beta + 2^{\alpha_0} + 2^{\alpha_0} \leqslant \beta + 2^{\alpha} = \gamma.$$

Lemma 4.11 then yields

$$\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{\gamma}]\big)\Big),$$

contradicting our initial assumption. ⊣

For the purposes of the present paper, we will appeal only to the following special case of the previous lemma:

Corollary 4.21 ($\mathsf{ACA}_0$). *Let $\alpha < \varepsilon_0$ be well-founded. Let $\prec$ be an arithmetic well-ordering. Let $\Delta$ be a finite set of $X$-positive formulas. Suppose that*

$$\vdash^{\alpha} \neg\mathsf{Prog}(\prec, X), \Delta.$$

*Then it follows that*

$$\mathsf{True}_{\Pi^1_1}\Big(\forall X\big(\bigvee \Delta[X \mapsto \prec_{2^\alpha}]\big)\Big).$$

**4.4. Formalizing $\Sigma^1_1$-bounding.** We are now ready to provide a diagonalization-free proof of Lemma 2.9, restated here:

Lemma 4.22 (Rathjen). *Suppose $H(x)$ is a $\Sigma^1_1$ formula such that*

$$\mathsf{ACA}_0 \vdash \forall x\big(H(x) \to x \in \mathfrak{W}_{Rec}\big).$$

*Then, for some $e \in Rec$,*

$$\mathsf{ACA}_0 \vdash e \in \mathfrak{W}_{Rec} \wedge \neg H(e).$$

Proof. Let $H(x)$ be a $\Sigma^1_1$ formula satisfying the hypothesis of the lemma. Then $H(x)$ is of the form $\exists Y\theta(x, Y)$ for some arithmetic formula $\theta$. For an $x \in \mathfrak{W}_{Rec}$, let $\prec^x$ be the well-ordering encoded by $x$.

We reason as follows:

$$\begin{aligned}
&\mathsf{ACA}_0 \vdash \forall x\big(\exists Y\theta(x, Y) \to x \in \mathfrak{W}_{Rec}\big),\\
&\mathsf{ACA}_0 \vdash \forall x\big(\neg\exists Y\theta(x, Y) \vee \forall X\mathsf{TI}(\prec^x, X)\big),\\
&\mathsf{ACA}_0 \vdash \forall x\Big(\neg\exists Y\theta(x, Y) \vee \forall X\big(\neg\mathsf{Prog}(\prec^x, X) \vee \forall y \in \mathsf{field}(\prec^x)\ y \in X\big)\Big),\\
&\mathsf{ACA}_0 \vdash \forall X\forall Y\forall x\big(\neg\theta(x, Y) \vee \neg\mathsf{Prog}(\prec^x, X) \vee \forall y \in \mathsf{field}(\prec^x)\ y \in X\big).
\end{aligned}$$

By Theorem 4.13, there is an $\alpha < \varepsilon_0$ such that the following is provable in $\mathsf{ACA}_0$:

$$\vdash^{\alpha} \neg\theta(x, Y), \neg\mathsf{Prog}(\prec^x, X), \forall y \in \mathsf{field}(\prec^x)\ y \in X. \tag{7}$$

We now switch to *reasoning in* $\mathsf{ACA}_0$. Suppose that $H(n)$ holds. That is,

$$\exists Y\theta(n, Y). \tag{8}$$

Applying Corollary 4.21 to (7) we infer that

$$\forall Y\big(\neg\theta(n, Y) \vee \forall y \in \mathsf{field}(\prec^x)\ y \in \prec^n_{2^\alpha}\big).$$

Which, by definition of $\prec^x_{2^\alpha}$, is just to say

$$\forall Y\big(\neg\theta(n, Y) \vee \forall y \in \mathsf{field}(\prec^n)\ y \in \{k \mid |k|_{\prec^n} < 2^\alpha\}\big).$$

Which is just to say that

$$\forall Y\big(\neg\theta(n, Y) \vee \forall y \in \mathsf{field}(\prec^n)\ |y|_{\prec^n} < 2^\alpha\big). \tag{9}$$

Combining (8) and (9) we see that $\forall y \in \mathsf{field}(\prec^n)\ |y|_{\prec^n} < 2^\alpha$.

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

This is just to say that $\mathsf{otyp}(\prec^n) < 2^\alpha$. We infer that

$$\sup\{\mathsf{otyp}(\prec^x) \mid \mathsf{True}_{\Sigma^1_1}(H(x))\} < 2^\alpha < 2^\alpha + 1 < \varepsilon_0.$$

Whence $2^\alpha + 1 \in \mathfrak{W}_{Rec}$ but $\neg H(2^\alpha + 1)$. ⊣

**§5. Open problems.** We will conclude with open problem concerning the sharpness of these theorems. That is, can any of the hypotheses in the statement of these theorems be weakened? There are three hypotheses that can be tweaked in interesting ways. First, there is the question of relaxing the definability condition.

QUESTION 5.1. *Is there a $\Pi^1_1$-sound and $\Pi^1_1$-definable extension of $\Sigma^1_1$-$\mathsf{AC}_0$ that proves its own $\Pi^1_1$-soundness?*

Note that a positive answer to this question would imply that Theorem 3.5 is sharp, at least along the dimension of the descriptive complexity of $T$.

If we do not demand that the theory extends $\Sigma^1_1$-$\mathsf{AC}_0$, then we can get a positive answer of sorts by considering the set of $\Pi^1_1$ truths. This theory is $\Pi^1_1$-sound and definable by the $\Pi^1_1$ predicate $\mathsf{True}_{\Pi^1_1}(x)$, which says that $x$ encodes a sentence that is $\mathsf{ACA}_0$-provably equivalent to a $\Pi^1_1$ sentence. This theory proves its own $\Pi^1_1$-reflection *statement*:

$$\forall\varphi \in \Pi^1_1\big(\mathsf{True}_{\Pi^1_1}(\varphi) \rightarrow \mathsf{True}_{\Pi^1_1}(\varphi)\big),$$

which is a logical truth. However, note that this depends on treating $\Pi^1_1$-reflection as a single statement, which is arguably inappropriate for a theory that does not extend $\mathsf{ACA}_0$. If we treated $\Pi^1_1$-reflection as a schema

$$\left\{\forall\vec{x}\Big(\mathsf{True}_{\Pi^1_1}(\varphi(\vec{x})) \rightarrow \varphi(\vec{x})\Big) \mid \varphi(\vec{x}) \in \Pi^1_1\right\},$$

then the theory in question might not prove instances of this schema since $\mathsf{ACA}_0$ is required to transform arbitrary $\Pi^1_1$ statements into normal form.

If we define $T$ as the union of $\Sigma^1_1$-$\mathsf{AC}_0$ with the set of all $\Pi^1_1$ truths, then the $\Pi^1_1$-reflection statement

$$\forall\varphi \in \Pi^1_1\big(\mathsf{Pr}_T(\varphi) \rightarrow \mathsf{True}_{\Pi^1_1}(\varphi)\big)$$

is not a $\Pi^1_1$ statement, since the antecedent is $\Pi^1_1$, so it does not trivially follow from $T$.

Second, there is the question of relaxing the soundness condition.

QUESTION 5.2. *Is there a $\Sigma^1_1$-sound and* $\Sigma^1_1$-definable *extension of $\Sigma^1_1$-$\mathsf{AC}_0$ that proves its own $\Pi^1_1$-soundness?*

If we demand in addition that the theory proves Theorem 3.5 and *provably* extends $\Sigma^1_1$-$\mathsf{AC}_0$, then we get a strong negative answer. Suppose that:

(1) $T$ is definable by a $\Sigma^1_1$ formula $\tau$.
(2) $T$ proves that $\tau$ extends $\Sigma^1_1$-$\mathsf{AC}_0$.
(3) $T$ proves the $\Pi^1_1$-soundness of $\tau$.

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press

Then, since $T$ proves Theorem 3.5, $T$ proves that $\tau$ is not $\Pi^1_1$-sound. Hence, $T$ proves both that $\tau$ is and is not $\Pi^1_1$-sound, i.e., $T$ is inconsistent, whence $T$ is not $\Sigma^1_1$-sound.

Finally, there is the question of relaxing the condition that $T$ extend $\Sigma^1_1$-$\mathsf{AC}_0$.

QUESTION 5.3. *Is there a $\Pi^1_1$-sound and $\Sigma^1_1$-definable* extension of $\mathsf{ACA}_0$ *that proves its own $\Pi^1_1$-soundness*?

Regarding this question there are reasons to expect a negative answer. In a recent preprint [1], Aguilera and Pakhomov have introduced $|T|_{\Pi^1_2}$, the $\Pi^1_2$ norm of $T$. $|T|_{\Pi^1_2}$ is a dilator associated with $T$ that is in some ways analogous to the proof-theoretic ordinal of $T$, except that it measures the $\Pi^1_2$ consequences of $T$. Aguilera and Pakhomov prove that $\Pi^1_2$-reflection for $T$ is equivalent to the statement "$|T|_{\Pi^1_2}$ is a dilator" [1, Theorem 7]; this is an analogue of Lemma 3.4. The important point is that this result is proved for $T$ extending $\mathsf{ACA}_0$. An appropriate reformulation of their proof of Theorem 7 may deliver a negative answer to Question 5.3, which would strengthen Theorem 3.5.

**Acknowledgments.** Thanks to Dan Appel, Reid Dale, and Paolo Mancosu for discussion. Thanks also to an anonymous referee for helpful comments and suggestions. Special thanks to Anton Freund for discussion of Section 4 and to Antonio Montalbán for discussion of Section 3.

## REFERENCES

[1] J. P. AGUILERA and F. PAKHOMOV, *The $\Pi^1_2$ consequences of a theory*, preprint, 2021, arXiv:2109.11652.

[2] A. BECKMANN and W. POHLERS, *Applications of cut-free infinitary derivations to generalized recursion theory*. ***Annals of Pure and Applied Logic***, vol. 94 (1998), nos. 1–3, pp. 7–19.

[3] L. D. BEKLEMISHEV, *Induction rules, reflection principles, and provably recursive functions*. ***Annals of Pure and Applied Logic***, vol. 85 (1997), no. 3, pp. 193–242.

[4] W. BUCHHOLZ, *Notation systems for infinitary derivations*. ***Archive for Mathematical Logic***, vol. 30 (1991), nos. 5–6, pp. 277–296.

[5] G. GENTZEN, *Provability and nonprovability of restricted transfinite induction in elementary number theory*, ***The Collected Papers of Gerhard Gentzen*** (M. E. Szabo, editor), North-Holland, Amsterdam, 1969, pp. 287–308.

[6] H. KOTLARSKI, *The incompleteness theorems after 70 years*. ***Annals of Pure and Applied Logic***, vol. 126 (2004), nos. 1–3, pp. 125–138.

[7] P. LUTZ and J. WALSH, *Incompleteness and jump hierarchies*. ***Proceedings of the American Mathematical Society***, vol. 148 (2020), no. 11, pp. 4997–5006.

[8] F. PAKHOMOV and J. WALSH, *Reflection ranks and ordinal analysis*, this JOURNAL, vol. 86 (2021), pp. 1350–1384.

[9] ———, *Reflection ranks via infinitary derivations*, preprint, 2021, arXiv:2107.03521.

[10] W. POHLERS, ***Proof Theory: An Introduction***, Lecture Notes in Mathematics, vol. 1407 Springer, Berlin–Heidelberg, 1989, p. VI-213.

[11] ———, *Subsystems of set theory and second order number theory*. ***Handbook of Proof Theory***, (S. Buss, editor) vol. 137 (1998), pp. 209–335.

[12] M. RATHJEN, *The role of parameters in bar rule and bar induction*, this JOURNAL, vol. 56 (1991), no. 2, pp. 715–730.

[13] ———, *The realm of ordinal analysis*, ***Sets and Proofs*** (S. Barry Cooper and J. K. Truss, editors), London Mathematical Society Lecture Note Series, Cambridge University Press, New York, 1999, pp. 219–280.

[14] G. E. Sacks, ***Higher Recursion Theory, vol. 2***, Cambridge University Press, Cambridge University Press, New York, 2017.

[15] S. Salehi. *On the diagonal lemma of Gödel and Carnap*. ***Bulletin of Symbolic Logic***, vol. 26 (2020), no. 1, pp. 80–88.

[16] A. Visser, *From Tarski to Gödel—Or how to derive the second incompleteness theorem from the undefinability of truth without self-reference*. ***Journal of Logic and Computation***, vol. 29 (2019), no. 5, pp. 595–604.

[17] J. Walsh, *On the hierarchy of natural theories*, preprint, 2021, arXiv:2106.05794.

SAGE SCHOOL OF PHILOSOPHY
CORNELL UNIVERSITY
ITHACA, NY 14850, USA
*E-mail*: jameswalsh@cornell.edu

https://doi.org/10.1017/jsl.2022.64 Published online by Cambridge University Press