HIERARCHICAL INCOMPLETENESS RESULTS FOR ARITHMETICALLY DEFINABLE EXTENSIONS OF FRAGMENTS OF ARITHMETIC

RASMUS BLANCK

University of Gothenburg

Abstract. There has been a recent interest in hierarchical generalizations of classic incompleteness results. This paper provides evidence that such generalizations are readily obtainable from suitably formulated hierarchical versions of the principles used in the original proofs. By collecting such principles, we prove hierarchical versions of Mostowski's theorem on independent formulae, Kripke's theorem on flexible formulae, Woodin's theorem on the universal algorithm, and a few related results. As a corollary, we obtain the expected result that the formula expressing "T is Σ_n-ill" is a canonical example of a Σ_{n+1} formula that is Π_{n+1}-conservative over T.

§1. Introduction. There has been a recent interest in hierarchical generalizations of classic incompleteness results [5, 23, 27, 40]. A sample result, generalizing the Gödel–Rosser incompleteness theorem, and independently proved by both Kikuchi and Kurahashi [23] and Salehi and Seraji [40], is:

THEOREM 1.1. *Let* T *be a* Σ_{n+1}-*definable,* Σ_n-*sound extension of* PA. *Then there is a* Π_{n+1} *sentence that is undecidable in* T.

In this paper I argue that such hierarchical generalizations can often be obtained from the original proofs by replacing certain principles used in the proofs by appropriately formulated hierarchical versions, while the essence of the arguments remains the same. The hierarchical principles, once appropriately formulated, are in turn often provable by appropriate generalizations of the core concepts employed in the proofs of the ordinary ones, but even so, there is no single source to which to turn for them. Both Smoryński [45] and Beklemishev [1] give good partial accounts of the syntactical side, and Poizat [35] gives a hierarchical perspective on the basic model theory of arithmetic, including model-theoretic proofs of hierarchical versions of Gödel's first and second incompleteness theorems. Still, I find certain aspects lacking. With this in mind, one aim of this paper is to collect a number of principles that may be useful to the reader who herself wishes to prove hierarchical incompleteness results without having to reinvent the wheel.

These principles are then put to use to prove a number of general incompleteness results for arithmetically definable extensions of fragments of PA. The goal is not to

---

prove the sharpest or most general results (in fact some of the results follow from each other), but rather to exemplify how the hierarchical principles enter into more or less well-known proof methods. Even so, the results presented here improve on some results of Chao and Seraji [5], Kikuchi and Kurahashi [23], and Salehi and Seraji [40], and sharpen some of Blanck [3], Hamkins [18], Lindström [28], and Woodin [48]. These sharpenings are in terms of gauging the amount of induction needed for the proofs, bringing the (in this particular sub-field largely ignored) fragments-of-arithmetic perspective to attention.

In order to state the results in a more general form, I have chosen to consider only extensions of $I\Delta_0 + \exp$, although under the somewhat unusual name $I\Sigma_0 + \exp$. This allows for formulating results for extensions of, e.g., $I\Sigma_n + \exp$, ensuring that partial satisfaction predicates are well behaved even for $n = 0$. While it is sometimes possible to push the background theory below $I\Sigma_0 + \exp$, I have refrained from doing so, to instead focus on the more general hierarchical picture.

§2. **Notation and conventions.** The expressions $\exists x \le t\phi(x)$ and $\forall x \le t\phi(x)$ are used as shorthand for $\exists x(x \le t \wedge \phi(x))$ and $\forall x(x \le t \rightarrow \phi(x))$, where t is some term in the language of arithmetic. The initial quantifiers of these formulae are *bounded* and a formula containing only bounded quantifiers is a *bounded formula*. Let $\Delta_0 = \Sigma_0 = \Pi_0$ be the class of bounded formulae.

The arithmetical hierarchy is defined as follows. A formula is Σ_{n+1} iff it is of the form $\exists x_1 \ldots x_m \pi(x_1, \ldots, x_m)$ where π is a Π_n formula (that may contain other variables than $x_1, \ldots, x_m$). Similarly, a formula is Π_{n+1} iff it is of the form $\forall x_1 \ldots x_m \sigma(x_1, \ldots, x_m)$ where σ is a Σ_n formula. $\Delta_n(\mathcal{M})$ ($\Delta_n(T)$) is the set of Σ_n formulae that are equivalent to Π_n formulae in a given model $\mathcal{M}$ (theory T). Throughout the paper Γ denotes either Σ_{n+1} or Π_{n+1}, and we always assume only that $n \ge 0$.

Theories are understood as sets of sentences, thought of as the set of nonlogical axioms of the theory. $I\Sigma_n$ is the theory obtained by adding induction for Σ_n formulae to Robinson's arithmetic Q, while $I\Sigma_0 + \exp$ is Q plus Σ_0-induction plus an axiom stating that the exponentiation function is total. We assume that all theories denoted T, etc., are consistent, arithmetically definable, extensions of $I\Sigma_0 + \exp$.

T is Γ-*sound* iff for all Γ sentences γ, if $T \vdash \gamma$, then $\mathbb{N} \models \gamma$. The converse implication is sometimes known as Γ-completeness: hence T is Γ-*complete* iff for all Γ sentences γ, if $\mathbb{N} \models \gamma$, then $T \vdash \gamma$. S is Γ-*conservative* over T iff for all Γ sentences γ, if $S \vdash \gamma$, then $T \vdash \gamma$.

We rely on a coding of finite sets and sequences in $I\Sigma_0 + \exp$, as developed by Hájek & Pudlák [17, chap. I.1]. The set $\Sigma_0(X)$ is obtained by adding atomic formulae of the form $t \in X$ (where t is any term) and closing under propositional connectives and bounded quantifiers. The set $\Sigma_1(X)$ is obtained from $\Sigma_0(X)$ in the usual manner.

Let ε be Ackermann's membership relation: $n\varepsilon a$ expresses "the nth bit of the binary expansion of a is 1." In other words, a can be regarded as a code for the set consisting of all n such that $n\varepsilon a$. Then a is a *z-piece of* $\phi(x)$ if $\forall n < z(n\varepsilon a \leftrightarrow \phi(n))$.

If $\phi(x)$ is any formula, $\ulcorner\phi(x)\urcorner$ denotes the numeral for the Gödel number of $\phi(x)$ under some fixed Gödel numbering, but we make no typographical distinction between natural numbers and the corresponding numerals. We use Feferman's dot notation $\ulcorner\phi(\dot{x})\urcorner$ to represent the Gödel number of the sentence obtained by replacing the variable x with the actual value of x; hence x is free in $\ulcorner\phi(\dot{x})\urcorner$. The notation := is used

to express equality between formulae. Let $\top := 0 = 0$ and $\bot := \neg\top$. Let $\phi^0 := \neg\phi$ and $\phi^1 := \phi$. [1]

A relation X is *numerated* in T by a formula ϕ iff $X = \{\langle k_1, \dots, k_n \rangle \in \omega^n : \text{T} \vdash \phi(k_1, \dots, k_n)\}$, and *binumerated* by ϕ in T if $\neg\phi$ also numerates the complement of X. A relation X is *correctly numerated* by ϕ if ϕ numerates X and for all $k_1, \dots, k_n$, $\text{T} \vdash \phi(k_1, \dots, k_n)$ iff $\phi(k_1, \dots, k_n)$ is true. A function f is *strongly representable* in T iff there is a formula $\phi(x_1, \dots, x_n, y)$ that numerates $f(x_1, \dots, x_n) = y$ in T, and moreover, if $f(k_1, \dots, k_n) = m$, then $\text{T} \vdash \forall y(\phi(k_1, \dots, k_n, y) \rightarrow y = m)$.

Given a formula $\tau(z)$, let $\text{Prf}_\tau(x, y)$ be a formula expressing "y is a proof of the formula x from the set of sentences satisfying $\tau(z)$." Let $\text{Pr}_\tau(x)$ be the formula $\exists y \text{Prf}_\tau(x, y)$, and let Con_τ be the sentence $\neg\text{Pr}_\tau(\ulcorner\bot\urcorner)$. Whenever $\tau(z)$ is Σ_{n+1}, $\text{Pr}_\tau(x)$ is equivalent to a Σ_{n+1} formula in the real world. For any formula $\tau(x)$, the notation $(\tau + z)(x)$ is used as shorthand for $\tau(x) \vee x = z$. This convention is used in expressions such as $\text{Prf}_{\tau+z}(x)$.

Models of arithmetic are denoted $\mathcal{M}$, etc., while the respective domains are denoted M, etc. The standard model is denoted $\mathbb{N}$ and its domain is ω. If $\mathcal{M}$ is non-standard, then the *standard system* of $\mathcal{M}$, $\text{SSy}(\mathcal{M})$, is the collection of sets $X \subseteq \omega$ such that for some $a \in M$, $X = \{n \in \omega : \mathcal{M} \models n \varepsilon a\}$. Then X is *coded* in $\mathcal{M}$, and a is a *code for X*.

A relation $X \subseteq M^n$ is Γ-*definable in* $\mathcal{M}$ (with parameters) iff there is a tuple $\overline{b} \in M$ and a formula $\phi(x_1, \dots, x_n, \overline{y}) \in \Gamma$ such that $X = \{\langle m_1, \dots, m_n \rangle \in M^n : \mathcal{M} \models \phi(m_1, \dots, m_n, \overline{b})\}$. Whenever this terminology is used without specifying a model $\mathcal{M}$, it is assumed that $\mathcal{M} = \mathbb{N}$. Thus, in particular, a Γ-*definition* of a theory T is a Γ formula $\tau(x)$ such that $\text{T} = \{\phi : \mathbb{N} \models \tau(\ulcorner\phi\urcorner)\}$.

For each Γ, $\text{Th}_\Gamma(\mathcal{M})$ is the set of Γ sentences true in $\mathcal{M}$, that is, the set $\{\phi \in \Gamma : \mathcal{M} \models \phi\}$. If $\mathcal{M}$ is a submodel of $\mathcal{N}$, and for all $\overline{a} \in M$ and $\gamma(\overline{x}) \in \Gamma$, $\mathcal{M} \models \gamma(\overline{a})$ iff $\mathcal{N} \models \gamma(\overline{a})$, then $\mathcal{N}$ is a Γ-*elementary extension of* $\mathcal{M}$. If $\mathcal{M} \models \sigma$ for some Σ_{n+1} sentence σ, and $\mathcal{N}$ is a Σ_n-elementary extension of $\mathcal{M}$, then $\mathcal{N} \models \sigma$. $\mathcal{N}$ is an *end-extension of* $\mathcal{M}$ (or, equivalently, $\mathcal{M}$ is an *initial segment of* $\mathcal{N}$) iff $\mathcal{M}$ is a submodel of $\mathcal{N}$, and whenever $a \in M$ and $b \in N$, and $\mathcal{N} \models b < a$, then $b \in M$.

Let $\emptyset^{(n)}$ denote the nth Turing jump of the empty set (see, e.g., [39, p. 254]). For each n, let $\langle \varphi_i^n : i \in \omega \rangle$ be an acceptable (in the sense of Rogers [39, exercise 2.10]) enumeration of the functions that are recursively enumerable (r.e.) in $\emptyset^{(n)}$; these are the partial n-recursive functions. For each partial n-recursive function φ_e^n, let the eth n-r.e. set W_e^n be the domain of φ_e^n. If f is a function such that $f \simeq \varphi_e^n$ for some e, then e is an *n-index for f*.

§3. **Preliminary principles.** This section presents a number of suitably formulated principles that are useful in proving hierarchical incompleteness results of the kind given in Theorem 1.1. The basic versions of these principles can be found scattered across the literature, and none of them should be too surprising to the reader familiar with, e.g., Hájek and Pudlák [17], Kaye [22], Lindström [29], and Smoryński [45].

The essence of the generalizations presented in this paper is that r.e. theories are replaced by Σ_{n+1}-definable ones, and that the base theory is pushed down as far as it

---

[1] The literature is not in total agreement about this convention. We follow Hájek and Pudlák [17], while for example Lindström [29] has it the other way around. The mnemonic here is that the superscript 1 signals that ϕ occurs positively.

will go below PA. On some occasions the theories have to satisfy additional constraints such as Σ_n-soundness or Π_n-completeness for the generalization to go through, and whenever this is the case that will be pointed out explicitly.

3.1. Basics, representability, recursion theory. Many textbooks in metamathematics rely on some version of the fact that Robinson's arithmetic Q is Σ_1-complete: it proves all true Σ_1 sentences. This has the consequence that every r.e. set can be numerated in Q by a Σ_1 formula, which in turn allows for representing the theorem set of Q in Q and proving the first incompleteness theorem. Many proofs of these facts rely on Δ_0 being closed under bounded quantification, and Q deciding all Δ_0 sentences.

Since the aim of this paper is to generalize incompleteness results to Σ_{n+1}-definable theories rather than r.e. ones, there is a need for establishing a similar correspondence between Σ_{n+1}-definable sets and theories sufficient to represent them. On these higher levels, the roles of Δ_0 and Q are played by the classes $\Sigma_0(\Sigma_n)$ and the theories $I\Sigma_n + Th_{\Pi_n}(\mathbb{N})$, respectively. Establishing this relationship is the goal of this subsection.

The following four observations (see, e.g., [43, lemma 0.3], [16, lemma 2.2], [1, lemma 2.9], [23, proposition 3.7]) serve as the starting point for the analogy between Q and $I\Sigma_n + Th_{\Pi_n}(\mathbb{N})$.

FACT 3.2 (Basic properties of soundness and completeness).

1. T *is consistent iff* T *is Σ_0-sound.*
2. T *is Π_n-complete iff* T *is Σ_{n+1}-complete.*
3. T *is Σ_n-sound iff* T *is Π_{n+1}-sound iff* T $+ Th_{\Sigma_{n+1}}(\mathbb{N})$ *is consistent.*
4. *If* T *is consistent and Π_n-complete, then* T *is Σ_n-sound.*

While the usual formulation of Gödel's theorem pertains to r.e., consistent theories, the hierarchically formulated Theorem 1.1 is stated for Σ_{n+1}-definable, Σ_n-sound theories. In light of the preceding fact, consistency and Σ_0-soundness are equivalent, so the ordinary statement fits nicely into the hierarchical statement of the theorem. However, inspection of the published proofs of Theorem 1.1 reveals that Π_n-complete theories enter the argument in an indispensable way. In the proof by Kikuchi & Kurahashi [23], Π_n-completeness of T can be explicitly assumed, since in the case where T is not Π_n-complete, it fails to prove all true Π_n sentences and can hardly prove all true Π_{n+1} sentences. By contrast, the proof by Salehi & Seraji [40] bypasses assuming Π_n-completeness of T by instead constructing a Π_{n+1} sentence that is independent of the Π_n-complete T $+ Th_{\Pi_n}(\mathbb{N})$, a theory whose consistency is guaranteed by the Σ_n-soundness of T. That sentence is, a fortiori, also independent of T.

The most conservative generalization of "r.e., consistent" is therefore "Σ_{n+1}-definable, Π_n-complete, and consistent" rather than "Σ_{n+1}-definable, Σ_n-sound." As is clear from Theorem 1.1, the assumption of Π_n-completeness of T is sometimes excessively strong, and mere Σ_n-soundness is indeed enough for some further applications as well. In those cases, the proofs rely on the consistency of T $+ Th_{\Pi_n}(\mathbb{N})$, as in the proof by Salehi & Seraji [40]. In other cases, however, the Π_n-completeness is indispensable for the generalization to go through.[2]

---

[2] One of the referees pointed out that even Π_n-completeness is sometimes not enough for a straightforward hierarchical generalization to hold: see, e.g., Theorem 11 of [27] for an example.

The analogy with regard to soundness and completeness therefore goes as follows. Since Q is Π_0-complete, it is also Σ_1-complete, and since it is also consistent it follows that Q is Σ_0-sound and therefore also Π_1-sound. Now consider $\mathrm{Th}_{\Pi_n}(\mathbb{N})$: This theory is Π_n-complete and therefore also Σ_{n+1}-complete; it is consistent and therefore Π_{n+1}-sound.

The role of $\mathrm{I}\Sigma_n$ in the analogy between Q and $\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ becomes clear through the next fact. Its proof will take up most of the remainder of this subsection.

FACT 3.3.

1. *Every Σ_n- (or Π_n-) definable relation on ω is binumerated by a Σ_n (or Π_n) formula in $Q + \mathrm{Th}_{\Pi_n}(\mathbb{N})$.*
2. *Every Σ_{n+1}-definable relation on ω is numerated by a Σ_{n+1} formula in $\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N})$.*
3. *Every function from ω^k to ω that is recursive in $\emptyset^{(n)}$ is strongly representable by a Σ_{n+1} formula (with particularly nice properties) in $\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N})$.[3]*

The first item is immediately seen to be true, since $Q + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ is consistent, Σ_{n+1}-complete, and Π_{n+1}-sound. The second and third items are elaborated on below, but first we need a few more stepping stones to help out in the constructions.

FACT 3.4 (Parametric diagonal lemma).

1. *For every Γ formula $\gamma(x, y)$, we can effectively find a Γ formula $\xi(x)$ such that*

$$Q \vdash \forall x (\xi(x) \leftrightarrow \gamma(x, \ulcorner \xi \urcorner)).$$

2. *For every Γ formula $\gamma(x, y)$, we can effectively find a Γ formula $\xi(x)$ such that, for each k,*

$$Q \vdash \xi(k) \leftrightarrow \gamma(k, \ulcorner \xi(k) \urcorner).$$

BIBLIOGRAPHICAL REMARK. Item 1. of the above is essentially due to Montague [33, lemma 1]. The proof by Ehrenfeucht and Feferman [10, lemma 1] of 2. actually suffices to prove 1.; see also [44] for a discussion of the development of the diagonal lemma.

FACT 3.5 (Generalized Craig's trick). *For any Σ_{n+1}-definable theory T, defined by a Σ_{n+1} formula $\sigma(z)$, there is a deductively equivalent theory S, defined by a formula $\pi(x)$ that is Π_n in $\mathrm{I}\Sigma_n$. Moreover, $\mathrm{I}\Sigma_{n+1} \vdash \forall x (\mathrm{Pr}_\sigma(x) \leftrightarrow \mathrm{Pr}_\pi(x))$.[4]*

BIBLIOGRAPHICAL REMARK. Craig's [7] formulation pertains to r.e. theories, for which there exist deductively equivalent theories with primitive recursive definitions. An early hierarchical generalization is due to Grzegorczyk et al. [14, theorem 2.2.C]. The hierarchical formulation presented here follows by inspection of a proof by Kurahashi [27, proposition 9].

---

[3] As is well known, $\mathrm{I}\Sigma_1$ proves the totality of the exponential function, so the additional axiom exp is only required in the case $n = 0$ to make sure that the partial satisfaction predicates used to establish the particularly nice properties are well-behaved. A similar remark applies also to many of the following facts, and to the statements of some of the theorems in Section 4.

[4] For $n = 0$, it is known that $\mathrm{I}\Sigma_1$ suffices to show that the Craigified theory is deductively equivalent to the original one [17, remark III.2.30]. By contrast, this is not true of $\mathrm{I}\Sigma_0 + \exp$ [47].

CONVENTION. *In light of Craig's trick, we may assume that any Σ_{n+1}-definable theory is in fact Π_n-defined. In what follows, we write $\mathrm{Prf}_T(x, y)$ to denote (ambiguously) any formula $\mathrm{Prf}_\tau(x, y)$ where τ is a Π_n binumeration of T in $I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N})$, and moreover, this formula can be assumed to be Π_n. Consequently, $\mathrm{Pr}_T(x)$ is Σ_{n+1} and Con_T is Π_{n+1} in $I\Sigma_n$.*

FACT 3.6 [17, lemmata I.2.9 and I.2.14 and theorem I.2.25].

1. *In $I\Sigma_n$, both Σ_n and Π_n are closed under bounded quantifiers.*[5]
2. $I\Sigma_n \vdash I\Sigma_0(\Sigma_n)$.
3. *Each $\Sigma_0(\Sigma_n)$ formula is Δ_{n+1} in $I\Sigma_n$.*

All the pieces are now in place to prove a, sometimes useful, lemma from which Fact 3.3(2) follows immediately. The proof highlights the steps where induction and the additional truth from the standard model is required. For the statement of the lemma, recall that a relation X is correctly numerated by ϕ in T if ϕ numerates X in T, and for all $k_1, \ldots, k_n$, $T \vdash \phi(k_1, \ldots, k_n)$ iff $\phi(k_1, \ldots, k_n)$ is true.

LEMMA 3.7. *Let T be any Σ_{n+1}-definable, Π_n-complete, and consistent extension of $I\Sigma_n$ and let $R(x_1, \ldots, x_m)$ be any Σ_{n+1}-definable relation. There is a Σ_{n+1} formula $\phi(x_1, \ldots, x_m)$ that correctly numerates R in T.*

Proof. Let $\rho(x_1, \ldots, x_m)$ define R. We may assume that $\rho(x_1, \ldots, x_m)$ is of the form $\exists z \pi(x_1, \ldots, x_m, z)$ with $\pi \in \Pi_n$. Let $\phi(x_1, \ldots, x_m)$ be such that, for all $k_1, \ldots, k_m$, $I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ proves

$$\phi(k_1, \ldots, k_m) \leftrightarrow \exists z(\pi(k_1, \ldots, k_m, z) \wedge \forall y \leq z \neg \mathrm{Prf}_T(\ulcorner \phi(k_1, \ldots, k_m) \urcorner, y)).$$

Recall that if T is Σ_{n+1}-definable, then there is a deductively equivalent Π_n definition of T, and $\mathrm{Prf}_T(x, y)$ is therefore equivalent to a Π_n formula in $I\Sigma_n$. Then

$$\pi(k_1, \ldots, k_m, z) \wedge \forall y \leq z \neg \mathrm{Prf}_T(\ulcorner \phi(k_1, \ldots, k_m) \urcorner, y)$$

is a $\Sigma_0(\Sigma_n)$ formula, and is therefore, by Fact 3.6, equivalent to a Δ_{n+1} formula in $I\Sigma_n$. It follows that ϕ is equivalent to a Σ_{n+1} formula in $I\Sigma_n$.

Suppose $R(k_1, \ldots, k_m)$. Then $\mathbb{N} \models \rho(k_1, \ldots, k_m)$, so there is then an i such that $\mathbb{N} \models \pi(k_1, \ldots, k_m, i)$. Since π is Π_n and $\mathrm{Th}_{\Pi_n}(\mathbb{N})$ is Π_n-complete, it follows that $I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \pi(k_1, \ldots, k_m, i)$. Suppose, for a contradiction, that $T \nvdash \phi(k_1, \ldots, k_m)$. By Fact 3.3(1),

$$I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \neg \mathrm{Prf}_T(\ulcorner \phi(k_1, \ldots, k_m) \urcorner, p),$$

for all p. It follows that

$$I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \forall y \leq i \neg \mathrm{Prf}_T(\ulcorner \phi(k_1, \ldots, k_m) \urcorner, y),$$

so $I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \phi(k_1, \ldots, k_m)$, and $T \vdash \phi(k_1, \ldots, k_m)$.

Conversely, suppose $T \vdash \phi(k_1, \ldots, k_m)$, and let p be a proof of $\phi(k_1, \ldots, k_m)$ in T. Then $I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \mathrm{Prf}_T(\ulcorner \phi(k_1, \ldots, k_m) \urcorner, p)$. It follows that

$$I\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \forall y \leq z \neg \mathrm{Prf}_T(\ulcorner \phi(k_1, \ldots, k_m) \urcorner, y) \rightarrow z < p.$$

Suppose, for a contradiction, that $\neg R(k_1, \ldots, k_m)$. Then for all i,

---

[5] In fact, only the weaker principle of Σ_n-*collection* is required for this first item. However, collection plays no prominent role elsewhere in this paper.

$$\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \neg\pi(k_1, \dots, k_m, i).$$

It follows that $\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \neg\exists z < p\,\pi(k_1, \dots, k_m, z)$. We get

$$\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \neg\exists z(\pi(k_1, \dots, k_m, z) \wedge \forall y \leq z \neg\mathrm{Prf}_{\mathrm{T}}(\ulcorner\phi(k_1, \dots, k_m)\urcorner, y)),$$

whence $\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \neg\phi(k_1, \dots, k_m)$, and $\mathrm{T} \vdash \neg\phi(k_1, \dots, k_m)$.

It remains to show that $\mathrm{T} \vdash \phi(k_1, \dots, k_m)$ iff $\phi(k_1, \dots, k_m)$ is true. The implication from right to left is trivial, since ϕ is Σ_{n+1} in $\mathrm{I}\Sigma_n$ and $\mathrm{Th}_{\Pi_n}(\mathbb{N})$ is Σ_{n+1}-complete. For the other direction, suppose $\mathrm{T} \vdash \phi(k_1, \dots, k_m)$, and let p be the least such proof. Then

$$\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \forall y \leq z \neg\mathrm{Prf}_{\mathrm{T}}(\ulcorner\phi(k_1, \dots, k_m)\urcorner, y) \rightarrow z < p.$$

Suppose further that there is no $i < p$ such that $\pi(k_1, \dots, k_m, i)$. Then, as before, $\mathrm{T} \vdash \neg\phi(k_1, \dots, k_m)$, a contradiction. Thus there is an $i < p$ such that $\pi(k_1, \dots, k_m, i)$ is true. Since p is minimal, $\forall y \leq i \neg\mathrm{Prf}_{\mathrm{T}}(\ulcorner\phi(k_1, \dots, k_m)\urcorner, y)$, so

$$\exists z(\pi(k_1, \dots, k_m, z) \wedge \forall y \leq z \neg\mathrm{Prf}_{\mathrm{T}}(\ulcorner\phi(k_1, \dots, k_m)\urcorner, y))$$

is true, and therefore $\phi(k_1, \dots, k_m)$ is true, as desired. $\square$

BIBLIOGRAPHICAL REMARK. The correct representability of Σ_1 relations in Q is due to Ehrenfeucht and Feferman [10]. The proof above mimics Lindström's [29] version of Shepherdson's [42] proof of the same result.

Fact 3.3(2) follows directly from Lemma 3.7, and it only remains to give an argument for Fact 3.3(3). For this we need two more facts, the first being a version of Post's theorem.

FACT 3.8 [36]. *A relation is Σ_{n+1} iff it is r.e. in $\emptyset^{(n)}$.*

FACT 3.9 (The selection theorem). *For each Σ_{n+1} formula ϕ with exactly the variables $x_1, \dots, x_k$ free, there is a Σ_{n+1} formula $\mathrm{Sel}\{\phi\}$ with exactly the same free variables, such that*:

1. $\mathrm{I}\Sigma_n \vdash \forall x_1, \dots, x_k(\mathrm{Sel}\{\phi\}(x_1, \dots, x_k) \rightarrow \phi(x_1, \dots, x_k))$;
2. $\mathrm{I}\Sigma_n \vdash \forall x_1, \dots, x_k, z(\mathrm{Sel}\{\phi\}(x_1, \dots, x_k) \wedge \mathrm{Sel}\{\phi\}(x_1, \dots, x_{k-1}, z) \rightarrow x_k = z)$;
3. $\mathrm{I}\Sigma_n \vdash \forall x_1, \dots, x_{k-1}(\exists x_k \phi(x_1, \dots, x_k) \rightarrow \exists x_k \mathrm{Sel}\{\phi\}(x_1, \dots, x_k))$.

Proof sketch. We may assume that any given Σ_{n+1} formula $\phi(x_1, \dots, x_k)$ is on the form $\exists w \pi(x_1, \dots, x_k, w)$ with $\pi \in \Pi_n$. Let $(w)_i$ denote the ith element of the ordered pair coded by w, and define $\mathrm{Sel}\{\phi\}(x_1, \dots, x_k)$ to be the formula

$$\exists w(\pi(x_1, \dots, x_{k-1}, (w)_0, (w)_1) \wedge \forall u \leq w \neg\pi(x_1, \dots, x_{k-1}, (u)_0, (u)_1) \wedge (w)_0 = x_k).$$

By Fact 3.6, this formula is equivalent in $\mathrm{I}\Sigma_n$ to a Σ_{n+1} formula. Using Σ_n-induction, it is easy to check that it has the three desired properties. $\square$

BIBLIOGRAPHICAL REMARK. Smoryński [45, theorem 0.6.9] provides a proof of the selection theorem for $n = 0$, and the generalization is straightforward. The numeration below of partial recursive functions is also based on the treatment in Section 0 of his book.

Fact 3.3(3) can now be established as follows: Let f be a k-ary partial n-recursive function; then the relation $f(x_1, \dots, x_k) = y$ is Σ_{n+1} by Fact 3.8. By Fact 3.3(2), this relation is correctly numerated in $\mathrm{I}\Sigma_n + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ by some Σ_{n+1} formula $\psi(x_1, \dots, x_k, y)$. Finally, by Fact 3.9, f is strongly represented in $\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ by the Σ_{n+1} formula $\mathrm{Sel}\{\mathrm{Sat}_{\Sigma_{n+1}}\}(\ulcorner\psi\urcorner, x_1, \dots, x_k, y)$. This concludes the proof of Fact 3.3.

CONVENTION. *The partial n-recursive function with index e, φ_e^n, can now be defined to be the function whose graph is defined by* $\mathrm{Sel}\{\mathrm{Sat}_{\Sigma_{n+1}}\}(e, y_1, \dots, y_k, z)$ *in* $\mathbb{N}$. *The resulting enumeration is acceptable in the sense of Rogers* [39]. *Whenever convenient,* $\varphi_e^n(m_1, \dots, m_i) = k$ *is used as a shorthand for* $\mathrm{Sel}\{\mathrm{Sat}_{\Sigma_{n+1}}\}(e, m_1, \dots, m_i, k)$.

To wrap up this subsection, we sketch a proof of a formalized version of the second recursion theorem.

FACT 3.10 (Formalized second recursion theorem). *Let $f : \omega^2 \to \omega$ be n-recursive. There is an $e \in \omega$ such that*

$$\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \varphi_e^n(y) \simeq f(e, y).$$

Proof. Let, by Fact 3.3(3), $\psi(x, y, z)$ be a Σ_{n+1} formula strongly representing f in $\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N})$. Let, by Fact 3.4, $\gamma(y, z)$ be a formula such that

$$\mathrm{I}\Sigma_0 + \exp \vdash \gamma(y, z) \leftrightarrow \mathrm{Sel}\{\mathrm{Sat}_{\Sigma_{n+1}}\}(\ulcorner\psi\urcorner, \ulcorner\gamma\urcorner, y, z).$$

Then $e = \ulcorner\gamma\urcorner$ is as desired. $\square$

The recursion theorem is usually deployed in the following manner. Define an n-recursive function $f(z, x)$ in stages, using z as a parameter; the resulting function may differ depending on the choice of z. By the recursion theorem, there is then an n-index e such that $\varphi_e^n(x)$ coincides with $f(e, x)$. This legitimates self-referential constructions where an index of f is being used in the construction of f itself.

BIBLIOGRAPHICAL REMARK. The recursion theorem is due to Kleene [24, theorem XXVII]. Smoryński [45, theorem 0.6.12] shows how it can be formalized in (essentially) $\mathrm{I}\Sigma_0 + \exp$ for (0-)recursive functions.

3.2. Strong provability predicates. A central piece in the proof of Gödel's incompleteness theorem for r.e. theories T is the use of formal provability predicates $\mathrm{Pr}_T(x)$, expressing "x is provable in T." In the current setting, with Σ_{n+1}-definable theories in focus, the corresponding *strong provability predicates* $\mathrm{Pr}_{T,\Sigma_{n+1}}(x)$ express "x is provable in T from true Σ_{n+1} sentences."[6] To define these predicates, we first need to introduce partial satisfaction predicates.

FACT 3.11 (Partial satisfaction predicates). *For each k and Γ, there is a $k + 1$-ary Γ formula* $\mathrm{Sat}_\Gamma(x, x_1, \dots, x_k)$ *such that for every Γ formula $\phi(x_1, \dots, x_k)$,*

$$\mathrm{I}\Sigma_0 + \exp \vdash \forall x_1, \dots, x_k(\phi(x_1, \dots, x_k) \leftrightarrow \mathrm{Sat}_\Gamma(\ulcorner\phi\urcorner, x_1, \dots, x_k)).$$

Hence there is also a Γ formula $\mathrm{Tr}_\Gamma(x)$ such that for every Γ sentence ϕ,

$$\mathrm{I}\Sigma_0 + \exp \vdash \phi \leftrightarrow \mathrm{Tr}_\Gamma(\ulcorner\phi\urcorner).$$

---

[6] Similar provability predicates, $\mathrm{Pr}_{T,\Pi_n}(x)$ appear in the literature under the name strong provability, n-provability, and oracle provability [1, 20, 25, 47]. The difference is usually only a matter of taste, since the theories $\mathrm{T} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ and $\mathrm{T} + \mathrm{Th}_{\Sigma_{n+1}}(\mathbb{N})$ are deductively equivalent, and under reasonable assumptions this is reflected also in the formalized notions $\mathrm{Pr}_{T,\Pi_n}(x)$ and $\mathrm{Pr}_{T,\Sigma_{n+1}}(x)$. However, the relationship between the related notions $\mathrm{Pr}_{T,\Pi_n}^k(x)$ and $\mathrm{Pr}_{T,\Sigma_{n+1}}^k(x)$ (introduced below) is not as immediate, since there k bounds the length of the proof of x and the Gödel number of the additional true Π_n or Σ_{n+1} sentence used in the proof. Even though every Σ_{n+1}-provable sentence has a Π_n-proof, this proof might be much longer than the original one. We opt for the Σ_{n+1} versions, since it helps in the proof of Fact 3.23 below, and since it makes (some of) the indices align nicely.

BIBLIOGRAPHICAL REMARK. Modern proofs of this fact are due to Hájek and Pudlák [17] and Kaye [22]. The use of partial satisfaction predicates, however, goes back to Hilbert and Bernays [19].

DEFINITION 3.12. Let, for each Γ, $\mathrm{Prf}_{T,\Gamma}(x, y)$ be the formula

$$\exists z(z \in \Gamma \wedge \mathrm{Tr}_\Gamma(z) \wedge \mathrm{Prf}_{T+z}(x, y)).$$

Let $\mathrm{Pr}_{T,\Gamma}(x) := \exists y \mathrm{Prf}_{T,\Gamma}(x, y)$, and let $\mathrm{Con}_{T,\Gamma} := \neg \mathrm{Pr}_{T,\Gamma}(\ulcorner \bot \urcorner)$.

We are mainly interested in the predicates $\mathrm{Pr}_{T,\Sigma_{n+1}}(x)$. These provability predicates have many similarities with the usual provability predicate. Firstly, if T is Σ_{n+1}-definable, then $\mathrm{Pr}_{T,\Sigma_{n+1}}(x)$ is Σ_{n+1}, and $\mathrm{Con}_{T,\Sigma_{n+1}}$ consequently Π_{n+1}. Secondly, they satisfy provable Σ_{n+1}-completeness and Löb's derivability conditions. These notions also have their roots with Hilbert and Bernays [19]; see also [11, theorem 5.4 and corollary 5.5]. The versions presented here are extracted from [1, propositions 2.10 and 2.11] and [45, lemma 3.3.7].

FACT 3.13 (Provable Σ_{m+1}-completeness). *Let* T *be a* Σ_{n+1}-*definable theory, and let* $\sigma(x_1, \ldots, x_k)$ *be any* Σ_{m+1} *formula. Then*

$$\mathrm{I}\Sigma_0 + \exp \vdash \forall x_1, \ldots, x_k (\sigma(x_1, \ldots, x_k) \to \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \sigma(\dot{x}_1, \ldots, \dot{x}_k) \urcorner)).$$

FACT 3.14 (Löb conditions). *If* T *is a* Σ_{n+1}-*definable extension of* $\mathrm{I}\Sigma_0 + \exp$, *then, for each* $m \geq 0$, *and for all sentences* ϕ, ψ,

L1. *if* $T \vdash \phi$, *then* $\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi \urcorner)$;
L2. $\mathrm{I}\Sigma_0 + \exp \vdash \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi \urcorner) \wedge \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi \to \psi \urcorner) \to \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \psi \urcorner)$;
L3. $\mathrm{I}\Sigma_0 + \exp \vdash \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi \urcorner) \to \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi \urcorner) \urcorner)$.

Similar statements also hold for formulae:

L1'. *if* $T \vdash \forall x \phi(x)$, *then* $\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \forall x \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi(\dot{x}) \urcorner)$;
L2'. $\mathrm{I}\Sigma_0 + \exp \vdash \forall x (\mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi(\dot{x}) \urcorner) \wedge \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi(\dot{x}) \to \psi(\dot{x}) \urcorner)$
$\to \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \psi(\dot{x}) \urcorner))$;
L3'. $\mathrm{I}\Sigma_0 + \exp \vdash \forall x (\mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi(\dot{x}) \urcorner) \to \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \mathrm{Pr}_{T,\Sigma_{m+1}}(\ulcorner \phi(\dot{x}) \urcorner) \urcorner))$.

The stronger background theory used in items L1. and L1'. is enough to ensure the numerability of the Σ_{n+1}-definable theory T. For r.e. theories T, $\mathrm{I}\Sigma_0 + \exp$ suffices.

We now turn our attention to the *bounded provability predicates* that feature prominently in the sequel. Consider again the Σ_{n+1} formula $\mathrm{Pr}_{T,\Sigma_{n+1}}(x)$ for a Σ_{n+1}-definable T. We may assume that there is a Π_n formula $\pi(x, y)$ such that

$$\mathrm{I}\Sigma_n \vdash \mathrm{Pr}_{T,\Sigma_{n+1}}(x) \leftrightarrow \exists y \pi(x, y).$$

Let $\mathrm{Pr}_{T,\Sigma_{n+1}}^k(x)$ be the formula $\exists z \leq k \pi(x, z)$.[7] Then, by Fact 3.6, this formula is Π_n in $\mathrm{I}\Sigma_n$, and therefore decidable in $\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n}(\mathbb{N})$.[8] As a consequence, we have strong reflection properties for the bounded proof predicates.

FACT 3.15 (Uniform small reflection). *Let* T *be a* Σ_{n+1}-*definable, consistent extension of* $\mathrm{I}\Sigma_0 + \exp$. *For each* $\phi(x)$ *and* k, *we have*:

---

[7] In the notation of, e.g., [30], $\mathrm{Pr}_{T,\Sigma_{n+1}}^k(x)$ would be written $k : \mathrm{Pr}_{T,\Sigma_{n+1}}(x)$.
[8] The additional axiom exp is again only required for $n = 0$, to handle the partial truth definition occurring in $\mathrm{Pr}_{T,\Sigma_{n+1}}$. Remarks of this type will hereafter be omitted.

$$\mathrm{I}\Sigma_n + \exp + \mathrm{Th}_{\Pi_n(\mathbb{N})} \vdash \forall x(\mathrm{Pr}^k_{\mathrm{T},\Sigma_{n+1}}(\ulcorner\phi(\dot{x})\urcorner) \to \phi(x)).$$

BIBLIOGRAPHICAL REMARK. A forerunner to this reflection principle is proved by Feferman [12, lemma 2.18]. The generalization to Σ_{n+1}-definable theories is straightforward; see also [17, lemma III.4.40] for some middle ground: reflection for proofs from true Σ_{n+1} sentences for r.e. theories.

FACT 3.16 (Formalized small reflection). *Let* T *be a* Σ_{n+1}-*definable, consistent extension of* $\mathrm{I}\Sigma_{n+1}$. *Then we have*:

$$\mathrm{I}\Sigma_0 + \exp \vdash \forall\phi\forall z\,\mathrm{Pr}_{\mathrm{T},\Sigma_{n+1}}(\ulcorner\mathrm{Pr}^{\dot{z}}_{\mathrm{T},\Sigma_{n+1}}(\ulcorner\phi\urcorner) \to \phi\urcorner).$$

BIBLIOGRAPHICAL REMARK. Verbrugge & Visser [46] show how the small reflection principle can be formalized in (theories weaker than) $\mathrm{I}\Sigma_0 + \exp$. I am grateful to one of the referees for pointing out a crucial error in an earlier statement of this fact.

3.3. Model theory of arithmetic.

The remainder of this section concerns the model theory of arithmetic, building up to a characterization of Π_n-conservativity in the spirit of Orey, Hájek, Guaspari, and Lindström. A first step toward that goal is the following miniaturization of the arithmetized completeness theorem in the style of McAloon [31, theorems 1.7 and 2.2].

FACT 3.17 (The arithmetized completeness theorem). *Fix* $m \le n$. *If* $\mathcal{M} \models \mathrm{I}\Sigma_{n+1}$, *and* T *is a theory that is* Σ_{n+1}-*definable in* $\mathcal{M}$ *such that* $\mathcal{M} \models \mathrm{Con}_{\mathrm{T},\Pi_m}$, *then there is a* Σ_m-*elementary end-extension of* $\mathcal{M}$ *satisfying* T.

The proof of the arithmetized completeness theorem rests on the following version of the low basis theorem by Hájek & Pudlák [17, corollary I.3.10(1)]:

FACT 3.18 (Low basis theorem). *Provably in* $\mathrm{I}\Sigma_{n+1}$, *each dyadic unbounded* Δ_{n+1} *tree has an unbounded* LL_{n+1} *branch*.

Here, a *tree* is a set of finite binary sequences that is closed under taking initial segments. A *branch* through a tree is a subtree that is linearly ordered under the relation "being an initial segment of." See [17, chap. I.3(b)] for more details. A definition of the class LL_{n+1} used in the statement of the low basis theorem can be found in [17, chap. I.2(d)]. The only properties of LL_{n+1} sets that are used in the proof of the arithmetized completeness theorem are that $\mathrm{I}\Sigma_{n+1}$ proves induction for $\Sigma_1(LL_{n+1})$ sets, and that every set recursive in an LL_{n+1} set is itself LL_{n+1} [17, I.2.78–79]).

For the proof of the arithmetized completeness theorem, we also rely on $\mathrm{I}\Sigma_{n+1}$ being able to define formalized versions of syntactic and semantic notions such as "formula," "term," "theory," "satisfaction," and "model," so that the relevant constructions can be carried out within $\mathrm{I}\Sigma_{n+1}$ itself. The reader is again referred to [17], especially Chapter I.4, for a detailed development of these concepts. The generalization to Σ_{n+1}-definable theories is straightforward, and fits safely within $\mathrm{I}\Sigma_{n+1}$.

Proof of the arithmetized completeness theorem. Fix $m \le n$. Let $\mathcal{M} \models \mathrm{I}\Sigma_{n+1}$ and let T be a theory Σ_{n+1}-definable in $\mathcal{M}$ such that $\mathcal{M} \models \mathrm{Con}_{\mathrm{T},\Pi_m}$. Reason in $\mathcal{M}$:

Since T is Σ_{n+1} and we have $\mathrm{I}\Sigma_{n+1}$, we may assume T to be Π_n-defined and Henkinized. Let $\psi_0, \psi_1, \ldots$ be an enumeration of all sentences. Define a dyadic tree T by

$s \in T$ iff there is no $p \le s$ such that p is a proof of contradiction in
T from the true Π_m sentences plus $\{\psi_i^{(s)_i} : i < l(s)\}$.

The proof relation for $T + \mathrm{Th}_{\Pi_m}(\mathcal{M})$ is Π_n in $\mathcal{M}$, so the tree T is at most Δ_{n+1} in $\mathcal{M}$. Then $I\Sigma_{n+1}$ suffices to show that T is unbounded, so by Fact 3.18 there is an unbounded LL_{n+1} branch $B = \{b_0, b_1, \dots\}$ through T.

Let $\hat{T} = \{\psi_i : b_i = 1\}$. Then $\hat{T}$ is recursive in the LL_{n+1} branch B, and is therefore LL_{n+1} itself. A term model $\mathcal{K}$ satisfying $T + \mathrm{Th}_{\Pi_m}(\mathcal{M})$ can then be read off $\hat{T}$ in the usual way. Finally, note that $\mathcal{K}$ is recursive in $\hat{T}$ and therefore LL_{n+1}. Using induction for $\Sigma_1(LL_{n+1})$ (provided by $I\Sigma_{n+1}$), we can now define an LL_{n+1} embedding f of $\mathcal{M}$ onto an initial segment of $\mathcal{K}$ by letting $f(0) = 0^{\mathcal{K}}$ and $f(x+1) =$ the $\mathcal{K}$-successor of $f(x)$. $\square$

The next few facts are used in the proof of the generalized Orey–Hájek characterization.

FACT 3.19 (Overspill). *Let $\mathcal{M} \models I\Sigma_{n+1}$. Suppose that $a \in M$ and that $\phi(x, y)$ is a $\Sigma_0(\Sigma_{n+1})$ formula such that $\mathcal{M} \models \phi(k, a)$ for all $k \in \omega$. Then there is a $b \in M \setminus \omega$ such that $\mathcal{M} \models \forall x \le b\, \phi(x, a)$.*

BIBLIOGRAPHICAL REMARK. The notion of overspill is originally due to Robinson [38], while the hierarchical version stated here is from [17, corollary IV.1.16].

FACT 3.20. *If $\mathcal{M}$ is a non-standard model of $I\Sigma_{n+1}$, and $\phi(x)$ is a Σ_{n+1} formula which may include parameters from $\mathcal{M}$, then $\{k \in \omega : \mathcal{M} \models \phi(k)\}$ is coded in $\mathcal{M}$.*

It follows that if $\mathcal{M} \models I\Sigma_{n+1}$, then $\mathrm{Th}_{\Sigma_{n+1}}(\mathcal{M})$ is coded in $\mathcal{M}$.

FACT 3.21 [32], cf. [8]. *If $\mathcal{M}$ is a countable non-standard model of $I\Delta_0$, and T is Σ_1-sound, then for every non-standard $c \in M$, there is a non-standard initial segment of $\mathcal{M}$ below c that is a model of T.*

FACT 3.22 (Refined Friedman embedding theorem). *If $\mathcal{M}$ and $\mathcal{N}$ are countable non-standard models of $I\Sigma_{n+1}$ then the following are equivalent:*

1. *$\mathcal{M}$ is embeddable as a Σ_n-elementary initial segment of $\mathcal{N}$;*
2. *$\mathrm{SSy}(\mathcal{M}) = \mathrm{SSy}(\mathcal{N})$ and $\mathrm{Th}_{\Sigma_{n+1}}(\mathcal{M}) \subseteq \mathrm{Th}_{\Sigma_{n+1}}(\mathcal{N})$.*

BIBLIOGRAPHICAL REMARK. This refinement of Friedman's [13] embedding theorem for $n = 0$ is due to Ressayre [37, theorem 1.I] and Dimitracopoulos and Paris [9, corollary 2.4], independently. The hierarchical generalization is straightforward, and has been worked out by Cornaros [6, corollary 15].

We are now ready to prove the final fact: an excerpt of a generalization of the Orey–Hájek–Guaspari–Lindström characterization of interpretability. For extensions of PA, the equivalence of 1. and 3. is due to Guaspari [15, theorem 6.5(1)]. The equivalence of 1. and 2. for finitely axiomatizable theories seems to have been known to experts for some time, while the equivalence of 2. and 3. for r.e. extensions of fragments of PA is stated without proof by Blanck and Enayat [4, theorem 2.11]. With the previous facts of this section in place, the generalization to Σ_{n+1}-definable theories presents no further difficulties.

FACT 3.23 (OHGL characterization). *Let S and T be Σ_{n+1}-definable, consistent extensions of $I\Sigma_{n+1}$, and suppose that S is also Π_n-complete. The following are equivalent:*

1. *S is Π_{n+1}-conservative over T;*
2. *for all $k \in \omega$, $T \vdash \mathrm{Con}^k_{S, \Sigma_{n+1}}$;*
3. *every countable model $\mathcal{M}$ of T with $S \in \mathrm{SSy}(\mathcal{M})$ has a Σ_n-elementary extension to a model of S.*

Proof. 1. $\Rightarrow$ 2. Suppose that S is Π_{n+1}-conservative over T. By Fact 3.15 and Π_n-completeness of S, $S \vdash \mathrm{Con}^k_{S,\Sigma_{n+1}}$ for all $k \in \omega$. But $\mathrm{Con}^k_{S,\Sigma_{n+1}}$ is at most Π_{n+1}, so $T \vdash \mathrm{Con}^k_{S,\Sigma_{n+1}}$ for all $k \in \omega$.

2. $\Rightarrow$ 3. Suppose $T \vdash \mathrm{Con}^k_{S,\Sigma_{n+1}}$ for all $k \in \omega$. Let $\mathcal{M}$ be a countable model of T and suppose that $S \in \mathrm{SSy}(\mathcal{M})$. Since T extends $I\Sigma_{n+1}$, it follows that $\mathrm{Con}^k_{S,\Sigma_{n+1}}$ is at most Π_{n+1}, and we can use overspill to get $\mathcal{M} \models \mathrm{Con}^c_{S,\Sigma_{n+1}}$ for some non-standard $c \in M$. By Fact 3.21, there is a submodel $\mathcal{M}_0 \models \mathrm{PA}$ of $\mathcal{M}$, all of whose elements are below c, and there is some non-standard a below c that codes $\mathrm{Th}_{\Sigma_{n+1}}(\mathcal{M})$. This ensures that $\mathcal{M}_0 \models \mathrm{Con}_{S+\{m:m\varepsilon a\}}$, so Fact 3.17 guarantees the existence of an end-extension $\mathcal{K}$ of $\mathcal{M}_0$, satisfying $S + \{m : m\varepsilon a\}$ and therefore also $S + \mathrm{Th}_{\Sigma_{n+1}}(\mathcal{M})$.

At this point, the situation is that $\mathrm{SSy}(\mathcal{M}) = \mathrm{SSy}(\mathcal{M}_0) = \mathrm{SSy}(\mathcal{K})$, $\mathcal{M}$ and $\mathcal{K}$ are countable, and $\mathrm{Th}_{\Sigma_{n+1}}(\mathcal{M}) \subseteq \mathrm{Th}_{\Sigma_{n+1}}(\mathcal{K})$. Then Fact 3.22 ensures that $\mathcal{M}$ can be embedded as a Σ_n-elementary initial segment of $\mathcal{K}$.

3. $\Rightarrow$ 1. Prove the contrapositive statement by assuming that S is not Π_{n+1}-conservative over T. Then there is a Π_{n+1} sentence π such that $S \vdash \pi$ but $T + \neg\pi$ is consistent. Let $\mathcal{M} \models T + \neg\pi$. If $\mathcal{K} \models S$ were a Σ_n-elementary extension of $\mathcal{M}$, then $\mathcal{K}$ would satisfy both π and $\neg\pi$, a contradiction. $\qquad\square$

§4. Applications. The goal of this section is to prove a handful of hierarchical incompleteness results, using the tools we reviewed in the previous one. The first such result stems from Mostowski [34, theorem 2], who proved that whenever $\{T_i : i \in \omega\}$ is an r.e. family of consistent, r.e. theories extending Q, then there is a Π_1 formula that is simultaneously independent over these theories. Here, we understand the concept of an independent formula in the following way:

DEFINITION 4.24. A formula $\xi(x)$ is *independent* over T if, for every $g : \omega \to \{0, 1\}$, the theory $T + \{\xi(k)^{g(k)}\}$ is consistent. Recall that $\xi(k)^0 = \neg\xi(k)$ and $\xi(k)^1 = \xi(k)$.

While one of Mostowski's accomplishments was the simultaneous independence over a whole r.e. family of theories, this aspect of his result is deliberately ignored here. Instead, we focus on how to construct formulae independent over Σ_{n+1}-definable theories.

THEOREM 4.25. *Let* T *be a* Σ_{n+1}-*definable,* Σ_n-*sound extension of* $I\Sigma_n + \exp$. *Then there is a* Σ_{n+1} *formula* $\xi(x)$ *that is independent over* T.

Proof. Define a function $f(x)$ by the stipulation that $f(m) = k$ iff there is a proof p of $\varphi^n_m(m) \neq k$ in T, and for each $q < p$ and $k_0 \leq k$, q is not a proof of $\varphi^n_m(m) \neq k_0$ in T.

Since T is Σ_{n+1}-definable, there is a deductively equivalent Π_n-definition of T. Hence, the relation $f(x) = y$ is r.e. in $\emptyset^{(n)}$, and therefore partial n-recursive. Let e be an n-index for f, and let $\xi(x)$ be the formula

$$\exists z(\varphi^n_e(e) = z \wedge \mathrm{Sat}_{\Sigma_{n+1}}(z, x)).$$

Since T extends $I\Sigma_n$, we may assume that both $\varphi^n_x(y) = z$ and $\xi(x)$ are equivalent in T to Σ_{n+1} formulae. The proof that $\xi(x)$ is as desired has two parts. The first part shows that $T + \varphi^n_e(e) = k$ is consistent for each $k \in \omega$.

Suppose, for a contradiction, that $T + \varphi^n_e(e) = k$ is inconsistent for some $k \in \omega$. We may assume that k is the least such number. Then $T \vdash \varphi^n_e(e) \neq k$ with some

minimal proof p, so $f(e) = \varphi_e^n(e) = k$ by definition. With T extending $I\Sigma_n$, we have $T + \text{Th}_{\Pi_n}(\mathbb{N}) \vdash \varphi_e^n(e) = k$ by Fact 3.3(3). But then $T + \text{Th}_{\Pi_n}(\mathbb{N})$ is inconsistent, which by Fact 3.2 contradicts the assumption that T is Σ_n-sound. Hence the theory $T + \varphi_e^n(e) = k$ is consistent for any choice of $k \in \omega$.

In this final part of the proof, we show that $\xi(x)$ is independent over T. Let g be any function from ω to $\{0, 1\}$ and let $X = \{\xi(k)^{g(k)} : k \in \omega\}$. Let Y be any finite subset of X, and let Z be the set $\{k : \xi(k) \in Y\}$. Let $\zeta(x) := x\varepsilon a$, where a is a code for the finite set Z; then ζ binumerates Z in Q.

Reason in the consistent theory $T + \varphi_e^n(e) = \ulcorner\zeta\urcorner$:

> If $\xi(x)$, then $\varphi_e^n(e) = z \wedge \text{Sat}_{\Sigma_{n+1}}(z, x)$ for some z. But z is unique and $\varphi_e^n(e) = \ulcorner\zeta\urcorner$, so $\text{Sat}_{\Sigma_{n+1}}(\ulcorner\zeta\urcorner, x)$ and therefore $\zeta(x)$.
> Conversely, observe that since $\varphi_e^n(e) = \ulcorner\zeta\urcorner$, $\xi(x)$ follows from $\zeta(x)$ by Fact 3.11 and $\exists$-introduction.

Hence the theory $T + \forall x(\xi(x) \leftrightarrow \text{Sat}_{\Sigma_{n+1}}(\ulcorner\zeta\urcorner, x))$ is consistent. If $k \in Z$, then $T \vdash \text{Sat}_{\Sigma_{n+1}}(\ulcorner\zeta\urcorner, k)$, so $T + \forall x(\xi(x) \leftrightarrow \text{Sat}_{\Sigma_{n+1}}(\ulcorner\zeta\urcorner, x)) \vdash \xi(k)$, and similarly for $k \notin Z$. Hence the consistent theory $T + \forall x(\xi(x) \leftrightarrow \text{Sat}_{\Sigma_{n+1}}(\ulcorner\zeta\urcorner, x))$ proves all the sentences in Y, so $T + Y$ is consistent. By compactness, it follows that $T + X$ is consistent, and therefore $\xi(x)$ is independent over T.   □

The Gödel–Rosser incompleteness Theorem 1.1 for arithmetically definable theories follows immediately from the result above. While the generalization to arithmetically definable theories is new, the basic idea of this proof is due to Kripke [26, corollary 1.1], who used it to rederive Mostowski's result from his own theorem on the existence of flexible formulae. Here, we understand flexibility in the following sense:

DEFINITION 4.26. A formula $\gamma(x)$ is *flexible* for Γ over T if, for every $\delta(x) \in \Gamma$, the theory $T + \forall x(\gamma(x) \leftrightarrow \delta(x))$ is consistent.

The definitions used by Kripke obscure the original content of his theorem, but, in hindsight, his proof yields that for every consistent, r.e. extension T of $I\Sigma_0 + \exp$, there is a Σ_{n+1} formula that is flexible for Σ_{n+1} over T. Striving for some unification, we derive a hierarchical version of Kripke's theorem by generalizing a result of Lindström's [28, proposition 2]; which in turn is a generalization of both Mostowski's and Kripke's results, as well as of Scott's famous lemma used to realize countable Scott sets as standard systems of models of PA [41].

THEOREM 4.27. *Let T be a Σ_{n+1}-definable extension of $I\Sigma_m + \exp$, with $m \geq n$. For every Σ_m formula $\phi(x)$, there is a Σ_{m+1} formula $\gamma(x)$ such that for every $g : \omega \to \{0, 1\}$, if*

$$T_g = T + \{\phi(k)^{g(k)} : k \in \omega\}$$

is Σ_n-sound, then $\gamma(x)$ is flexible for Σ_{m+1} over T_g.

Proof. Fix n and let $\phi(x) \in \Sigma_m$, with $m \geq n$. Let $f(s, \ulcorner\eta\urcorner) = \ulcorner\sigma\urcorner$ iff the following holds:

1. s is binary sequence of length $k + 1$;
2. there is a proof p of $\neg\forall x(\eta(x) \leftrightarrow \sigma(x))$ in $T + \phi(0)^{(s)_0} + \cdots + \phi(k)^{(s)_k}$;
3. for every $q < p$ and any $\ulcorner\sigma_0\urcorner \leq \ulcorner\sigma\urcorner$, q is not a proof of $\neg\forall x(\eta(x) \leftrightarrow \sigma_0(x))$ in $T + \phi(0)^{(s)_0} + \cdots + \phi(k)^{(s)_k}$.

Here $(s)_k$ denotes the kth element of the sequence s.

Since T is Σ_{n+1}-definable, there is a deductively equivalent Π_n-definition of T. Hence, the relation $f(x, y) = z$ is r.e. in $\emptyset^{(n)}$, and therefore partial n-recursive. Let e be an n-index for f.

Let $\mathrm{Seq}_\phi(x)$ be the formula

$$\forall y < l(x)(y \varepsilon x \leftrightarrow \phi(y)),$$

where $l(x)$ denotes the length of x (this is the formula "x is a $l(x)$-piece of ϕ"). Whenever $\phi(x)$ is Σ_m, $\mathrm{Seq}_\phi(x)$ is $\Sigma_0(\Sigma_m)$, and since $T \vdash I\Sigma_m$, it is Δ_{m+1} in T. Let, by Fact 3.4, $\gamma(x)$ be such that

$$T \vdash \forall x(\gamma(x) \leftrightarrow \exists s \exists z(\mathrm{Seq}_\phi(s) \wedge \varphi_e^n(s, \ulcorner \gamma \urcorner) = z \wedge \mathrm{Sat}_{\Sigma_{m+1}}(z, x))).$$

Since $T \vdash I\Sigma_m$ and $m \geq n$, the formula strongly representing f in $T + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ is equivalent to a Σ_{n+1} formula in T. It follows that $\gamma(x)$ is equivalent to a Σ_{m+1} formula in T.

Suppose, for a contradiction, that there is a $g : \omega \to \{0, 1\}$ and a $\sigma(x) \in \Sigma_{m+1}$ such that T_g is Σ_n-sound, but $T_g + \forall x(\gamma(x) \leftrightarrow \sigma(x))$ is inconsistent. If there are more than one such $\sigma(x)$ for a given g, consider the one with the least Gödel number. There is then an initial subsequence s of g, of length $k + 1$ for some k, such that p is a proof of $\neg \forall x(\gamma(x) \leftrightarrow \sigma(x))$ in $T + \phi(0)^{(s)_0} + \cdots + \phi(k)^{(s)_k}$. Let s be the initial subsequence of g corresponding to the least such p.

It is now clear that neither p nor any $q < p$ can be a proof of $\neg \forall x(\gamma(x) \leftrightarrow \sigma_0(x))$ in $T + \phi(0)^{(s)_0} + \cdots + \phi(k)^{(s)_k}$ for any formula $\sigma_0(x)$ whose Gödel number is less than $\ulcorner \sigma \urcorner$. Hence, by definition, $f(s, \ulcorner \gamma \urcorner) = \ulcorner \sigma \urcorner$, and by Fact 3.3,

$$T + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \varphi_e^n(s, \ulcorner \gamma \urcorner) = \ulcorner \sigma \urcorner.$$

By choice of s, $T_g \vdash \mathrm{Seq}_\phi(\ulcorner s \urcorner)$, so $T_g + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \forall x(\gamma(x) \leftrightarrow \sigma(x))$ by an argument similar to that in the proof of Theorem 4.25. Then $T_g + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ is inconsistent, contradicting the assumption that T_g was Σ_n-sound. $\qquad\square$

By choosing $\phi(x)$ as $\top$ in the construction above, we obtain the expected hierarchical version of Kripke's theorem. A similar, but not entirely correct, claim is made by Blanck [3, theorem 4.8].

COROLLARY 4.28. *Let* T *be a* Σ_{n+1}-*definable,* Σ_n-*sound extension of* $I\Sigma_n$ + exp. *For all* $m \geq n$, *there is a* Σ_{m+1} *formula* $\gamma(x)$ *that is flexible for* Σ_{m+1} *over* T.

Mostowski's theorem for r.e. extensions of $I\Sigma_0$ + exp then follows immediately by using the method described in the proof of Theorem 4.25. A similar argument also yields Scott's lemma.

The next objective is to show how the hierarchical version of Kripke's theorem can be formalized in Π_n-complete extensions of $I\Sigma_{n+1}$. A similar, but not entirely correct, claim is made by Blanck [3, theorem 5.8]. The present proof is a minor modification of an argument of Blanck [3, theorem 5.1].

THEOREM 4.29. *Let* S *be a* Π_n-*complete, consistent extension of* $I\Sigma_{n+1}$, *and let* T *be* Σ_{n+1}-*definable. For all* $m \geq n$, *there is a* Σ_{m+1} *formula* $\gamma(x)$ *such that*:

1. $I\Sigma_{n+1} \vdash Con_{T,\Sigma_{n+1}} \rightarrow \forall x \neg \gamma(x)$;
2. *if* $\sigma(x) \in \Sigma_{m+1}$, *then every model of* $S + Con_{T,\Sigma_{n+1}}$ *has a* Σ_n-*elementary extension to a model of* $T + \forall x(\gamma(x) \leftrightarrow \sigma(x))$.

Proof. Fix n, and let $\phi(x, z)$ be the Σ_{n+1} formula $Pr_{T,\Sigma_{n+1}}(\ulcorner \varphi^n_{\dot{x}}(\dot{x}) \neq \dot{z} \urcorner)$. Let e be the Gödel number of $\phi(x, z)$.

Recall that $\varphi^n_x(y) = z$ is shorthand for $\text{Sel}\{Sat_{\Sigma_{n+1}}\}(x, y, z)$. Therefore, by Fact 3.9(1) we have

$$I\Sigma_{n+1} \vdash \forall z(\varphi^n_e(e) = z \rightarrow Sat_{\Sigma_{n+1}}(e, e, z)), \tag{1}$$

so by construction of $\phi(x, z)$ and choice of e,

$$I\Sigma_{n+1} \vdash \forall z(\varphi^n_e(e) = z \rightarrow Pr_{T,\Sigma_{n+1}}(\ulcorner \varphi^n_e(e) \neq \dot{z} \urcorner)). \tag{2}$$

By Fact 3.13,

$$I\Sigma_{n+1} \vdash \forall z(\varphi^n_e(e) = z \rightarrow Pr_{T,\Sigma_{n+1}}(\ulcorner \varphi^n_e(e) = \dot{z} \urcorner)), \tag{3}$$

so (2) and (3) together with Fact 3.14 give

$$I\Sigma_{n+1} \vdash \forall z(Con_{T,\Sigma_{n+1}} \rightarrow \varphi^n_e(e) \neq z). \tag{4}$$

Now, observe that

$$I\Sigma_{n+1} \vdash \exists z \neg Con_{T+\varphi^n_e(e)=\dot{z},\Sigma_{n+1}} \leftrightarrow \exists z Pr_{T,\Sigma_{n+1}}(\ulcorner \varphi^n_e(e) \neq \dot{z} \urcorner). \tag{5}$$

By construction of $\phi(x, z)$, the right hand side of the equivalence of (5) is identical to $\exists z \phi(e, z)$, and by Fact 3.11, we have

$$I\Sigma_{n+1} \vdash \exists z \phi(e, z) \leftrightarrow \exists z Sat_{\Sigma_{n+1}}(e, e, z). \tag{6}$$

Therefore, by Fact 3.9(3) and the convention on φ^n_e, (5) gives

$$I\Sigma_{n+1} \vdash \exists z \neg Con_{T+\varphi^n_e(e)=\dot{z},\Sigma_{n+1}} \leftrightarrow \exists z(\varphi^n_e(e) = z). \tag{7}$$

Together with (4), this implies

$$I\Sigma_{n+1} \vdash \forall z(Con_{T,\Sigma_{n+1}} \rightarrow Con_{T+\varphi^n_e(e)=\dot{z},\Sigma_{n+1}}). \tag{8}$$

Then the Σ_{m+1} formula $\gamma(x) := \exists z(\varphi^n_e(e) = z \wedge Sat_{\Sigma_{m+1}}(z, x))$ is as desired, and the first part of the theorem follows directly from (4). For the second part, let $\mathcal{M}$ be any model of $S + Con_{T,\Sigma_{n+1}}$, and let $\sigma(x)$ be any Σ_{m+1} formula. By (8), we immediately get $\mathcal{M} \models Con_{T+\varphi^n_e(e)=\ulcorner\sigma\urcorner,\Sigma_{n+1}}$. Moreover, since S is a Π_n-complete extension of $I\Sigma_{n+1}$, and T is Σ_{n+1}-definable, the theory $T + \varphi^n_e(e) = \ulcorner \sigma \urcorner$ is Σ_{n+1}-definable in $\mathcal{M}$, using Craig's trick. By Fact 3.17, there is then a Σ_n-elementary end-extension $\mathcal{K}$ of $\mathcal{M}$ that satisfies $T + \varphi^n_e(e) = \ulcorner \sigma \urcorner$. Since $\mathcal{K}$ satisfies $\varphi^n_e(e) = \ulcorner \sigma \urcorner$, it follows that $\mathcal{K} \models \forall x(\gamma(x) \leftrightarrow \sigma(x))$, as desired. □

COROLLARY 4.30. *Let* S *be a* Σ_{n+1}-*definable,* Π_n-*complete, and consistent extension of* $I\Sigma_{n+1}$, *and let* $\gamma(x)$ *be as in the proof of Theorem 4.29. Then, for each* $\sigma(x) \in \Sigma_{m+1}$ *with* $m \geq n$, $S + \forall x(\gamma(x) \leftrightarrow \sigma(x))$ *is* Π_{n+1}-*conservative over* $S + Con_{S,\Sigma_{n+1}}$.

The next theorem has a different flavor than the earlier ones, and is a generalization of Woodin's theorem on the universal algorithm [48]; see also [4, theorem 3.1]. A version for r.e. extensions of PA is independently due to Hamkins [18, theorem 18], and the proof presented here uses a method that I learned from Shavrukov. The particular

Solovay-style construction used in the proof is similar to the ones used by Berarducci [2] and Japaridze [21].

THEOREM 4.31. *Let* T *be a* Σ_{n+1}*-definable,* Π_n*-complete, and consistent extension of* $I\Sigma_{n+1}$. *There is a* Σ_{n+1}*-definable set* W_e *such that*:

1. $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash$ *"*W_e *is finite"*;
2. $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \mathrm{Con}_{T,\Sigma_{n+1}} \to W_e = \emptyset$;
3. *for each countable model* $\mathcal{M} \models T$, *if* s *is an* $\mathcal{M}$*-finite set such that* $\mathcal{M} \models W_e \subseteq s$, *then there is a* Σ_n*-elementary extension of* $\mathcal{M}$ *satisfying* $T + W_e = s$.

Proof. The set W_e is defined (in $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ and in the real world) as follows, using the formalization of the recursion theorem (Fact 3.10). At the same time, an auxiliary function $r(x)$ is defined.

Stage 0: Set $W_{e,0} = \emptyset$, and $r(0) = \infty$.[9]

Stage $x + 1$: Suppose $r(x) = m$. There are two cases:

> **Case A**: s is a finite set such that $s \supseteq W_{e,x}$, $k < m$, and x witnesses a Σ_{n+1} formula $\sigma(s)$ such that k is a proof in $T + \mathrm{Th}_{\Sigma_{n+1}}(\mathbb{N})$ of $\forall t(\sigma(t) \to W_e \neq t)$. Should there be more than one eligible candidate for either k or s, then choose the least such k, and then the least s corresponding to that k. Then set $W_{e,x+1} = s$ and $r(x + 1) = k$.
>
> **Case B**: Otherwise, set $W_{e,x+1} = W_{e,x}$ and $r(x + 1) = m$.

Let $W_e = \bigcup_x W_{e,x}$.

To prove 1., reason as follows: Since the proof relation for $T + \mathrm{Th}_{\Sigma_{n+1}}(\mathbb{N})$ is Δ_{n+1} in $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$, W_e is r.e. in $\emptyset^{(n)}$, and therefore Σ_{n+1} by Fact 3.8. Provably in $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$, we have that $W_{e,x+1} \supseteq W_{e,x}$, and $r(x + 1) \leq r(x)$, so by the Σ_{n+1}-least number principle (which is available thanks to Σ_{n+1}-induction), there is a limit $R = \lim_x r(x)$. For each x with $W_{e,x+1} \neq W_{e,x}$, $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$ proves $r(x + 1) < r(x)$, whence there can only be finitely many such x. So $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash$ "W_e is finite."

Note also that $T \vdash R > k$ for all $k \in \omega$. To show this, fix $k \in \omega$ and argue in T:

> Suppose $R \leq k$. Let y be minimal such that $r(y + 1) = R$. Then $W_e = W_{e,y+1} = s$ for some s such that R is a proof in $T + \mathrm{Th}_{\Sigma_{n+1}}(\mathbb{N})$ of $\forall t(\sigma(t) \to W_e \neq t)$, where $\sigma(s)$ is a true Σ_{n+1} formula.

But, by Fact 3.15,

> since $\forall t(\sigma(t) \to W_e \neq t)$ is proved from a true Σ_{n+1} sentence with a proof not exceeding k, it must be true. Since $\sigma(s)$ is true, $W_e \neq s$ is also true, and the contradiction proves $R > k$.

To prove 2., argue for the contrapositive statement in $I\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$:

> If $W_e = s \neq \emptyset$, then $\mathrm{Pr}^m_{T,\Sigma_{n+1}}(\ulcorner \forall t(\sigma(t) \to W_e \neq t) \urcorner)$ for some m and some true Σ_{n+1} formula $\sigma(s)$. The relation $s \subseteq W_{e,x}$ is Δ_{n+1} by Fact 3.6(3), so $\mathrm{Pr}_{T,\Sigma_{n+1}}(\ulcorner \dot{s} \subseteq W_e \urcorner)$ follows by Fact 3.13. Now reason inside $\mathrm{Pr}_{T,\Sigma_{n+1}}$:

---

[9] Here ∞ is a formal symbol that by definition is greater than all the natural numbers.

There is some $u = W_e$ with $u \supseteq s$, so by construction, $\sigma'(u)$ is true, and $\mathrm{Pr}_{\mathrm{T},\Sigma_{n+1}}^k(\ulcorner \forall t(\sigma'(t) \to W_e \neq t) \urcorner)$ for some $k \leq m$ and some Σ_{n+1} formula $\sigma'(u)$.

Apply Fact 3.16, and continue reasoning inside $\mathrm{Pr}_{\mathrm{T},\Sigma_{n+1}}$:

Then $\forall t(\sigma'(t) \to W_e \neq t)$ and $\sigma'(u)$, so $W_e \neq u$.

Then $\mathrm{Pr}_{\mathrm{T},\Sigma_{n+1}}(\ulcorner \exists u(W_e = u \wedge W_e \neq u) \urcorner)$, so Fact 3.14 gives $\neg\mathrm{Con}_{\mathrm{T},\Sigma_{n+1}}$ as desired.

To prove 3., first fix $m \in \omega$. By Fact 3.15, there is a proof k in T of

$$\forall t(\mathrm{Pr}_{\mathrm{T},\Sigma_{n+1}}^m(\ulcorner W_e \neq \dot{i} \urcorner) \to W_e \neq t).$$

Now reason in T:

Consider any finite $s \supseteq W_e$, and suppose x is a proof $\leq m$ of $W_e \neq s$ in $\mathrm{T} + \mathrm{Th}_{\Sigma_{n+1}}(\mathbb{N})$. Then $s \supseteq W_{e,x+1}$, and therefore $r(x + 1) \leq k$ by construction of $r(x + 1)$: here $\mathrm{Pr}_{\mathrm{T},\Sigma_{n+1}}^m(\ulcorner W_e \neq \dot{s} \urcorner)$ is $\mathrm{I}\Sigma_{n+1}$-equivalent to a true Σ_{n+1} sentence playing the role of $\sigma(s)$. But $k < R \leq r(x + 1)$, and the contradiction proves $\mathrm{Con}_{\mathrm{T}+W_e=\dot{s},\Sigma_{n+1}}^m$.

Therefore for all $m \in \omega$, $\mathrm{T} \vdash \forall s \supseteq W_e\ \mathrm{Con}_{\mathrm{T}+W_e=\dot{s},\Sigma_{n+1}}^m$.

For the final part of the proof, let $\mathcal{M}$ be any countable model of T, and let s be any $\mathcal{M}$-finite set such that $\mathcal{M} \models W_e \subseteq s$. Since T is a Σ_{n+1}-definable, Π_n-complete extension of $\mathrm{I}\Sigma_{n+1}$, Facts 3.11 and 3.20 imply that $\mathrm{T} + \mathrm{Th}_{\Sigma_{n+1}}(\mathcal{M}) + W_e = s \in \mathrm{SSy}(\mathcal{M})$. Since $\mathrm{T} \vdash \mathrm{Con}_{\mathrm{T}+W_e=\dot{s},\Sigma_{n+1}}^m$ for all $m \in \omega$, Fact 3.23 guarantees the existence of a Σ_n-elementary extension of $\mathcal{M}$ satisfying $\mathrm{T} + W_e = s$, which concludes the proof of the theorem. $\square$

COROLLARY 4.32. *With T as in Theorem 4.31, $\neg\mathrm{Con}_{\mathrm{T},\Sigma_{n+1}}$ is Π_{n+1}-conservative over T.*

Proof. Every countable model of T has a Σ_n-elementary extension satisfying $W_e \neq \emptyset$, and therefore also $\mathrm{T} + \neg\mathrm{Con}_{\mathrm{T},\Sigma_{n+1}}$ by the theorem. By Fact 3.23, the conclusion follows.[10] $\square$

The set W_e defined in Theorem 4.31 can be used to prove results of a more Kripkean variety, by using the information contained in W_e as codes for other sets. The next result is of this kind, and improves on Theorem 7.21 of [3] by generalizing to arithmetically definable theories. A version for r.e. extensions of PA is independently due to Hamkins [18, theorem 22(1)], who also noted that there is a very short proof of it from Theorem 4.31.

THEOREM 4.33. *Let T be a Σ_{n+1}-definable, Π_n-complete, and consistent extension of $\mathrm{I}\Sigma_{n+1}$. For all $m \geq n$, there is a Σ_{m+2} formula $\gamma(x)$ such that:*

1. $\mathrm{I}\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \mathrm{Con}_{\mathrm{T},\Sigma_{n+1}} \to \forall x \neg\gamma(x)$;
2. *for every $\sigma(x) \in \Sigma_{m+2}$, every countable model of T has a Σ_n-elementary extension satisfying $\mathrm{T} + \forall x(\gamma(x) \leftrightarrow \sigma(x))$.*

---

[10] As pointed out by one of the referees, adapting Kreisel's original proof of the Π_1-conservativity of $\neg\mathrm{Con}_{\mathrm{T}}$ over T is a simpler way to establish Corollary 4.32 than going via Theorem 4.31.

Proof sketch. It is straightforward to adapt the construction in the proof of Theorem 4.31 to produce an $\mathcal{M}$-finite binary sequence S_e, rather than a set [4, 18, 48]. Assume an enumeration of Σ_{m+2} formulae in which every Σ_{m+2} formula occurs infinitely often, and that every finite binary sequence codes such a formula. Let $\gamma(x)$ be the formula $\exists z(S_e = z \wedge \mathrm{Sat}_{\Sigma_{m+2}}(z, x))$. Since $S_e = z$ is at most Σ_{n+2} and $m \geq n$, it follows that $\gamma(x)$ is Σ_{m+2}.

Pick any $\sigma(x) \in \Sigma_{m+2}$, let $\mathcal{M}$ be any countable model of T and let s be S_e as calculated within $\mathcal{M}$. By assumption on the enumeration of Σ_{m+2} formulae, there is an $\mathcal{M}$-finite sequence $t \supseteq s$ such that t codes $\ulcorner\sigma(x)\urcorner$. By the sequence version of Theorem 4.31, there is a Σ_n-elementary extension $\mathcal{K}$ of $\mathcal{M}$ in which $S_e = t$. Then $\gamma(x)$ coincides with $\sigma(x)$ in $\mathcal{K}$, and therefore is as desired. $\qquad\square$

The question remains to which extent $m + 2$ can be replaced by $m + 1$ in the statement of Theorem 4.33. Some partial answers are already available: Theorem 4.31 gives a positive answer restricted to Σ_{m+1} formulae $\sigma(x)$ whose extension is $\mathcal{M}$-finite and for which $\mathcal{M} \models \forall x(\gamma(x) \rightarrow \sigma(x))$, while Theorem 4.29 can be seen as giving a partial positive answer that is restricted to models of $\mathrm{T} + \mathrm{Con}_{\mathrm{T},\Sigma_{n+1}}$. Blanck [3, chap. 7.4] lists several other partial answers to this question in a setting where T is an r.e. extension of PA and $n = 0$. By using the principles of Section 3 of the present paper, those constructions can be easily modified to give equally unsatisfactory answers in the present setting. The salient remaining question is as follows:

QUESTION. *Let* T *be a* Σ_{n+1}-*definable,* Π_n-*complete, and consistent extension of* $\mathrm{I}\Sigma_{n+1}$. *Is there a* Σ_{n+1} *formula* $\gamma(x)$ *such that*:

1. $\mathrm{I}\Sigma_{n+1} + \mathrm{Th}_{\Pi_n}(\mathbb{N}) \vdash \mathrm{Con}_{\mathrm{T},\Sigma_{n+1}} \rightarrow \forall x \neg\gamma(x)$;
2. *for every* $\sigma(x) \in \Sigma_{n+1}$, *every countable model of* $\mathrm{T} + \forall x(\gamma(x) \rightarrow \sigma(x))$ *has a* Σ_n-*elementary extension satisfying* $\mathrm{T} + \forall x(\gamma(x) \leftrightarrow \sigma(x))$?

§5. Discussion.

By inspecting the results proved in Section 4, we see two classes of Σ_{n+1}-definable theories emerging:

1. Σ_n-sound extensions of $\mathrm{I}\Sigma_n + \exp$; and
2. Π_n-complete, consistent extensions of $\mathrm{I}\Sigma_{n+1}$.

As suggested by Theorem 1.1, theories in the first class are strong enough for some applications. These include the results of Salehi & Seraji [40] and Kikuchi & Kurahashi [23], together with Theorems 4.25 and 4.27 (and their corollaries) of the present paper. This success relies on the fact that Σ_n-soundness of T guarantees the consistency of $\mathrm{T} + \mathrm{Th}_{\Pi_n}(\mathbb{N})$, in which the n-recursive functions can be strongly represented by a formula that is Σ_{n+1} in the presence of Σ_n-induction.

The second class of theories is required to prove results on Σ_n-elementary extensions of models of Σ_{n+1}-definable theories, for example results on partial conservativity via the OHGL characterization (Theorem 4.29 and onward). In these cases, Π_n-completeness of T ensures that every model $\mathcal{M}$ of T is a Σ_n-elementary extension of the standard model, which in the presence of Σ_{n+1}-induction suffices for T to be Σ_{n+1}-definable in $\mathcal{M}$ by using Craig's trick. Σ_{n+1}-induction is also used to prove the arithmetized completeness theorem for Σ_{n+1}-definable theories, which is indispensable for constructing the Σ_n-elementary extensions.

The facts listed in Section 3 should be enough to derive hierarchical generalizations for arithmetically definable extensions of fragments of PA of many of the theorems in, e.g., Lindström's classic *Aspects of Incompleteness* [29]. As suggested by the results in the present paper, some of these generalizations would apply only to Π_n-complete theories, while in other cases mere Σ_n-soundness would do. Others might not be prone to such generalizations at all, as shown by Kurahashi [27, theorem 11] and pointed out to me by one of the referees. In fact, it would be interesting to see which of the results in, say, the first five chapters of *Aspects* (where the results do not depend on T being essentially reflexive) that are prone to such generalizations, using these principles.

REFERENCES

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

[2] Berarducci, A. (1990). The interpretability logic of Peano arithmetic. *The Journal of Symbolic Logic*, **55**(3), 1059–1089.

[3] Blanck, R. (2017). *Contributions to the Metamathematics of Arithmetic*: *Fixed Points*, *Independence*, *and Flexibility*. Ph.D. Thesis, University of Gothenburg, Gothenburg.

[4] Blanck, R., & Enayat, A. (2017). Marginalia on a theorem of Woodin. *The Journal of Symbolic Logic*, **82**(1), 359–374.

[5] Chao, C., & Seraji, P. (2018). Gödel's second incompleteness theorem for Σ_n-definable theories. *Logic Journal of the IGPL*, **26**(2), 255–257.

[6] Cornaros, C. (20XX). Versions of Friedman's theorem for fragments of PA. Unpublished manuscript.

[7] Craig, W. (1953). On axiomatizability within a system. *The Journal of Symbolic Logic*, **18**(1), 30–32.

[8] D'Aquino, P. (1993). A sharpened version of McAloon's theorem on initial segments of models of $I\Delta_0$. *Annals of Pure and Applied Logic*, **61**, 49–62.

[9] Dimitracopoulos, C., & Paris, J. (1988). A note on a theorem of H. Friedman. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **34**, 13–17.

[10] Ehrenfeucht, A., & Feferman, S. (1960). Representability of recursively enumerable sets in formal theories. *Archiv für mathematische Logik und Grundlagenforschung*, **5**(1–2), 37–41.

[11] Feferman, S. (1960). Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, **49**, 35–92.

[12] ———, (1962). Transfinite recursive progressions of axiomatic theories. *The Journal of Symbolic Logic*, **27**(3), 259–316.

[13] Friedman, H. (1973). Countable models of set theories. In Mathias, A. R. D. and Rogers, H., editors. *Cambridge Summer School in Mathematical Logic 1971*. Springer Lecture Notes in Mathematics, Vol. 337. New York: Springer, pp. 539–573.

[14] Grzegorczyk, A., Mostowski, A., & Ryll-Nardzewski, C. (1958). The classical and the ω-complete arithmetic. *The Journal of Symbolic Logic*, **23**(2), 188–206.

[15] Guaspari, D. (1979). Partially conservative extensions of arithmetic. *Transactions of the American Mathematical Society*, **254**, 47–68.

[16] Hájek, P. (1977). Experimental logics and Π_3^0 theories. *The Journal of Symbolic Logic*, **42**(4), 515–522.

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

[18] Hamkins, J. D. (2018). The modal logic of arithmetic potentialism and the universal algorithm. Preprint, arXiv:1801.04599.

[19] Hilbert, D., & Bernays, P. (1934–1939). *Grundlagen der Mathematik*, Vols. 1–2. Berlin, Germany: Springer.

[20] Ignatiev, K. N. (1993). On strong provability predicates and the associated modal logics. *The Journal of Symbolic Logic*, **58**(1), 249–290.

[21] Japaridze, G. (1994). A simple proof of arithmetical completeness for Π_1-conservativity logic. *Notre Dame Journal of Formal Logic*, **35**(3), 346–354.

[22] Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford Logic Guides, Vol. 15. Oxford, UK: Clarendon Press.

[23] Kikuchi, M., & Kurahashi, T. (2017). Generalizations of Gödel's incompleteness theorems for Σ_n-definable theories of arithmetic. *The Review of Symbolic Logic*, **10**(4), 603–616.

[24] Kleene, S. C. (1952). *Introduction to Metamathematics*. Amsterdam, Netherlands: North-Holland.

[25] Kolmakov, E., & Beklemishev, L. (2019). Axiomatization of provable n-provability. *The Journal of Symbolic Logic*, **84**(2), 849–869.

[26] Kripke, S. A. (1962). "Flexible" predicates of formal number theory. *Proceedings of the American Mathematical Society*, **13**(4), 647–650.

[27] Kurahashi, T. (2018). On partial disjunction properties of theories containing Peano arithmetic. *Archive for Mathematical Logic*, **57**(7), 953–980.

[28] Lindström, P. (1984). A note on independent formulas. In *Notes on Formulas with Prescribed Properties in Arithmetical Theories*. Philosophical Communications, Red Series, Vol. 25. Gothenburg: Göteborgs Universitet, pp. 1–5.

[29] ———, (2003). *Aspects of Incompleteness* (second edition). Lecture Notes in Logic, Vol. 10. Natick, MA: A. K. Peters.

[30] Lindström, P., & Shavrukov, V. Yu. (2008). The $\forall\exists$ theory of Peano Σ_1 sentences. *Journal of Mathematical Logic*, **8**(2), 251–280.

[31] McAloon, K. (1978). Completeness theorems, incompleteness theorems and models of arithmetic. *Transactions of the American Mathematical Society*, **239**, 253–277.

[32] ———, (1982). On the complexity of models of arithmetic. *The Journal of Symbolic Logic*, **47**(2), 403–415.

[33] Montague, R. (1962). Theories incomparable with respect to relative interpretability. *The Journal of Symbolic Logic*, **27**(2), 195–211.

[34] Mostowski, A. (1961). A generalization of the incompleteness theorem. *Fundamenta Mathematicae*, **49**(2), 205–232.

[35] Poizat, B. (2000). *A Course in Model Theory*. New York: Springer.

[36] Post, E. L. (1948). Degrees of recursive unsolvability. *Bulletin of the American Mathematical Society*, **54**(7), 641–642.

[37] Ressayre, J.-P. (1987). Nonstandard universes with strong embeddings, and their finite approximations. In Simpson, S., editor. *Logic and Combinatorics*. Contemporary Mathematics, Vol. 65. Providence, RI: American Mathematical Society, pp. 333–358.

[38] Robinson, A. (1963). On languages which are based on non-standard arithmetic. *Nagoya Mathematical Journal*, **22**, 83–117.

[39] Rogers, H., Jr. (1967). *Theory of Recursive Functions and Effective Computability*. New York: McGraw-Hill.

[40] Salehi, S., & Seraji, P. (2017). Gödel–Rosser's incompleteness theorem, generalized and optimized for definable theories. *Journal of Logic and Computation*, **27**(5), 1391–1397.

[41] Scott, D. (1962). Algebras of sets binumerable in complete extensions of arithmetic. In Dekker, J. C. E., editor. *Recursive Function Theory*. Proceedings of Symposia in Pure Mathematics, Vol. 5. Providence, RI: American Mathematical Society, pp. 117–121.

[42] Shepherdson, J. C. (1961). Representability of recursively enumerable sets in formal theories. *Archiv für mathematische Logik und Grundlagenforschung*, **5**(3), 119–127.

[43] Smoryński, C. (1977). ω-consistency and reflection. In *Colloque International de Logique*: *Clermont-Ferrand*, *18–25 Juillet 1975*. Colloques internationaux du Centre National de la Recherche Scientifique, Vol. 249. Paris: Editions du C.N.R.S, pp. 167–181.

[44] ———, (1981). Fifty years of self-reference in arithmetic. *Notre Dame Journal of Formal Logic*, **22**(4), 357–375.

[45] ———, (1985). *Self-Reference and Modal Logic*. New York: Springer.

[46] Verbrugge, R., & Visser, A. (1994). A small reflection principle for bounded arithmetic. *The Journal of Symbolic Logic*, **59**(3), 785–812.

[47] Visser, A. (2015). Oracle bites theory. In Gosh, S, and Szymanik, J., editors. *The Facts Matter. Essays on Logic and Cognition in Honour of Rineke Verbrugge*. London, UK: College Publications, pp. 133–147.

[48] Woodin, W. H. (2011). A potential subtlety concerning the distinction between determinism and nondeterminism. In Heller, M., and Woodin, W. H., editors. *Infinity*: *New Research Frontiers*. Cambridge: Cambridge University Press, pp. 119–129.

DEPARTMENT OF PHILOSOPHY, LINGUISTICS AND THEORY OF SCIENCE
UNIVERSITY OF GOTHENBURG
BOX 200, SE-405 30, GOTHENBURG, SWEDEN
E-mail: rasmus.blanck@gu.se