FUNDAMENTA
MATHEMATICAE
152 (1997)

Interpreting reflexive theories in finitely many axioms

by

V. Yu. Shavrukov (Utrecht)

Abstract. For finitely axiomatized sequential theories F and reflexive theories R, we give a characterization of the relation 'F interprets R' in terms of provability of restricted consistency statements on cuts. This characterization is used in a proof that the set of Π_1 (as well as Σ_1) sentences π such that GB interprets $\mathrm{ZF} + \pi$ is Σ^0_3-complete.

0. Introduction. Relative interpretability among formal theories has been particularly well studied in two special cases: that of finitely axiomatized sequential theories (see Smoryński [14], Pudlák [11], Visser [16] etc.), and of reflexive, esp. essentially reflexive, theories (see Lindström [7], [8] etc.). There are nice characterizations of the interpretability relation between a pair of theories from the same one of the two indicated classes. These mostly involve provability of restricted consistency statements on cuts, versions of Π_1 conservativity relativized to cuts, and provability in weak theories of relative (restricted) consistency. For essentially reflexive theories it has been shown by Solovay and Lindström that the relation of relative interpretability among these is Π^0_2-complete. More specifically, the set of Σ_1 sentences σ such that PA interprets $\mathrm{PA}+\sigma$ already is Π^0_2-complete (Lindström [7]; see also Jeroslow [6]). As for the other case, interpretability of finitely axiomatized theories is clearly a Σ^0_1 matter.

In the present paper we shall be looking at a crossway case—we consider interpretability of a reflexive theory R in a finitely axiomatized sequential F. In Section 2 we shall give a characterization of this interpretability relation in terms similar to those of earlier characterizations for other pairs of classes of theories. In fact, the relevant ingredients of the proof have been lying around for some time scattered through a number of papers by Pudlák and Visser.

1991 *Mathematics Subject Classification*: Primary 03F25, 03D35; Secondary 03F30.

Relying on this characterization, we show in Section 3 that the set of Π_1 (as well as Σ_1) sentences π such that F interprets $\mathrm{R} + \pi$ is, under an auxiliary sufficient strength condition on R, $\varSigma^0_3$-complete. Earlier progress in this direction included an unpublished proof by Švejdar that the set of (Π_2, I believe) sentences γ such that GB interprets $\mathrm{ZF} + \gamma$ is both $\varSigma^0_2$- and $\varPi^0_2$-hard. We shall also mention an example to the effect that the sufficient strength condition is undroppable.

Due to the fact that cuts, which are generally not closed under exponentiation, are instrumental in dealing with interpretability in finitely axiomatized theories, weak arithmetic turns out to be the ambient context for our arguments. (This may be seen as somewhat ironic, for the main objective of the present paper is the strengthening of the above-mentioned Švejdar's result on GB and ZF that are, by all accounts, fairly strong theories.) Consequently, we need to recall many weak-arithmetical lemmas and some other background material. We do that in Section 1, where we also establish proper setting for developments of the paper.

Our $\varSigma^0_3$-completeness result shows for the first, to my knowledge, time some sophistication on the part of Π_1 sentences with respect to interpretability, for the set of Π_1 sentences π such that T interprets $\mathrm{T} + \pi$ is r.e. both for T finitely axiomatized and for T essentially reflexive. Note also that $\varSigma^0_3$-completeness shows that relative interpretability among formal theories in finite languages is generally just as arithmetically complex as it sounds (there is an interpretation ... for every T_2-proof ... there is a T_1-proof).

I would like to thank Lev Beklemishev, Alessandro Berarducci, Richard Sommer, Vítězslav Švejdar, Rineke Verbrugge, and Albert Visser for helpful discussions and sound suggestions.

1. Preliminaries. All theories in this paper are understood to be classical r.e. consistent ones and to speak a first-order language with at most finitely many predicate and function symbols. We refer the reader to Hájek & Pudlák [5, Chapter V] for the individual theories that we consider ($\mathrm{I}\Delta_0$, $\mathrm{I}\Delta_0 + \Omega_1$ etc.).

Δ_0 is the class of formulas in the orthodox arithmetical language $(0, S, +, \times)$ with all quantifiers bounded. We write $\Delta_0(\omega_1)$ if we allow the function $\omega_1(x) = x^{\log x}$ into bounding terms as well as matrices of formulas. To terms of the extended language we refer as ω_1*-terms*. Polynomial-length transformations of syntactical objects correspond to transformations of gödelnumbers bounded by suitable ω_1-terms. Recall that $\mathrm{I}\Delta_0 + \Omega_1$ proves induction for $\Delta_0(\omega_1)$ formulas; cf. Hájek & Pudlák [5, Proposition V.1.3]. Proposition V.1.4 of the same book says that whenever $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, \exists y\, \delta(x, y)$ with $\delta(x, y)$ in $\Delta_0(\omega_1)$, there is an ω_1-term $f(x)$ such that $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, \exists y \le f(x)\ \delta(x, y)$. It is easily verified that $\mathrm{I}\Delta_0 + \Omega_1$ proves every $\forall\Delta_0(\omega_1)$

formula equivalent to a $\forall\Delta_0$ one. We shall ambiguously call both these classes Π_1. $\exists\Delta_0$ (or $\exists\Delta_0(\omega_1)$) formulas are Σ_1. The subclasses $\Sigma^{\rm b}_1(\omega_1)$ and $\Pi^{\rm b}_1(\omega_1)$ of $\Delta_0(\omega_1)$ are the same as $\Sigma^{\rm b}_1$ and $\Pi^{\rm b}_1$ of Buss [3, §1.6] respectively, but with ω_1 instead of Buss' # function. Those formulas that are $({\rm I}\Delta_0+\Omega_1)$-equivalent both to a $\Sigma^{\rm b}_1(\omega_1)$ and a $\Pi^{\rm b}_1(\omega_1)$ formula are $\Delta^{\rm b}_1(\omega_1)$.

Theories T considered in this paper are supposed to come equipped with a fixed translation t of the arithmetical language into that of T. We denote by Γ the class of those formulas of T that are translations of arithmetical formulas from Γ, for any of the formula classes Γ introduced above, and provide no notational distinction between arithmetical formulas proper and those translated into the language of T. By saying that T *contains* an arithmetical theory S we mean that t is an interpretation of S in T.

We use dyadic numerals throughout the paper as in Buss [3, §2.1]. If $\varphi(x)$ is an arithmetical formula and $\overline{n}$ is a numeral then $\varphi(\overline{n})$ translated into the language of T is either of the formulas $\exists x\,(`x=\overline{n}\text{'}\wedge\varphi(x))$, $\forall x\,(`x=\overline{n}\text{'}\to\varphi(x))$, where

$$`x=\overline{n}\text{'} \;\equiv\; \exists y_{\log n},\dots,y_0\Big(x=y_{\log n}\wedge y_0=\varepsilon_0\wedge \bigwedge_{\log n>i\geq 0} y_{i+1}=2y_i+\varepsilon_i\Big)$$

and ε_i are appropriate bits in the binary expansion of n (cf. Visser [17, 7.3]). By $\log n$ we actually mean $\lfloor\log_2(\max\{1,n\})\rfloor$ (so that, for $n>0$, one has $m\leq\log n$ iff $2^m\leq n$). Let $\ulcorner\cdot\urcorner$ denote gödelnumbers. The set $\{\ulcorner\varphi(\overline{n})\urcorner\mid \text{all } n\}$ is $\Delta^{\rm b}_1(\omega_1)$-definable because $\{\ulcorner`x=\overline{n}\text{'}\urcorner\mid\text{all } n\}$ is defined by induction on construction of dyadic numerals. The function $n\mapsto\ulcorner\varphi(\overline{n})\urcorner$ is $\Sigma^{\rm b}_1(\omega_1)$-definable for a similar reason (see Buss [3, §§7.2–3]). Hence substitution of $\ulcorner\varphi(\overline{n})\urcorner$ for a variable in a $\Delta^{\rm b}_1(\omega_1)$ or more complex formula does not push up its complexity (Buss [3, §2.3]). We shall be consistently omitting the numeral bar, confusion between numerals and variables being unlikely.

We proceed to list prerequisite facts and lemmas. In Paris *et al.* [10, Theorem 6] (the ω_1-less variant of) the following proposition is supplied with a highly mysterious proof.

1.1. Proposition. *Let $x\sqsubseteq y$ be a $\Delta_0(\omega_1)$ relation on a domain specified by a $\Delta_0(\omega_1)$ formula $\delta(z)$. Suppose ${\rm I}\Delta_0+\Omega_1$ proves that $\sqsubseteq$ is reflexive and transitive. Then ${\rm I}\Delta_0+\Omega_1$ proves that for each z there is a $\sqsubseteq$-minimal element among $\{u\leq z\mid\delta(u)\}$ unless this set is empty.*

Proof $({\rm I}\Delta_0+\Omega_1)$. This follows by straightforward induction on z. ∎

Counting logarithmically small amounts of elements of a Δ_0 set is a well-known procedure (see Paris & Wilkie [9]). In the sequel we shall need to have simple properties of the counting formulas verifiable in a weak theory.

1.2. Proposition. *Let $\delta(x,y)$ be a Δ_0 formula. There is then a Δ_0 formula $\#_z\{x<v\mid\delta(x,y)\}=u$ in variables v, z, y, and u, which, under*

the condition $v \le \log z$, *holds true if and only if the cardinality of the set* $\{x < v \mid \delta(x,y)\}$ *is equal to* u. *Moreover*, *this formula can be chosen so that* $\mathrm{I}\Delta_0$ *proves*:

(a) $v \le \log z \to \exists! u\, (\#_z\{x < v \mid \delta(x,y)\} = u)$;

(b) $v \le \log z \wedge \forall x < v\, (\delta(x,y_1) \to \delta(x,y_2))$
$$\to.\ \#_z\{x < v \mid \delta(x,y_1)\} \le \#_z\{x < v \mid \delta(x,y_2)\};$$

(c) $v_1 \le v_2 \le \log z \to \#_z\{x < v_1 \mid \delta(x,y)\} \le \#_z\{x < v_2 \mid \delta(x,y)\}$;

(d) $v \le \log z \to \#_z\{x < v \mid \delta(x,y)\} \le v$;

(e) $v \le \log z_1 \wedge v \le \log z_2$
$$\to.\ \#_{z_1}\{x < v \mid \delta(x,y)\} = \#_{z_2}\{x < v \mid \delta(x,y)\}.$$

Reference. The formula in question is described in e.g. Berarducci & D'Aquino [1]. All clauses can be easily inferred from Lemma 2.4 of that paper. ∎

The reader may consult Hájek & Pudlák [5, V.3(c)] for a Δ_0 definition of the relation $y = 2^x$ and its simple properties as known to $\mathrm{I}\Delta_0$ that in the sequel will be taken for granted. A prominent role in our exposition will be played by the superexponential function:

1.3. Proposition. *There exists a* Δ_0 *formula* $z = 2^y_x$ *such that* $\mathrm{I}\Delta_0$ *proves*:

(a) $z = 2^y_x \wedge w = 2^y_x \to.\ z = w$;
(b) $2^x_0 = x$;
(c) $z = 2^y_{x+1} \leftrightarrow \exists w \le z\, (w = 2^y_x \wedge z = 2^w)$;
(d) $z = 2^y_{x+t} \leftrightarrow \exists w \le z\, (w = 2^y_x \wedge z = 2^w_t)$;
(e) $z = 2^y_x \wedge u \le y \to.\ \exists w \le z\, (w = 2^u_x)$;
(f) $z = 2^u_x \wedge w = 2^v_x \to.\ z \le w \leftrightarrow u \le v$;
(g) $z = 2^y_x \to z \ge y$.

Reference. See e.g. D'Aquino [4, 3.1] or Wilkie [19, section 3], where iteration of Δ_0-definable fast growing functions is handled in greater generality. ∎

1.4. Convention. Consider the following formulas:

$$\begin{aligned} 2^y_x \le z \ &\equiv\ \exists w \le z\, (w = 2^y_x), \\ `z \le 2^y_x\text{'} \ &\equiv\ \forall w < z\, (w \ne 2^y_x), \\ `2^u_x \le 2^v_y\text{'} \ &\equiv\ (x \le y \wedge `u \le 2^v_{y-x}\text{'}) \vee (x \ge y \wedge 2^u_{x-y} \le v), \end{aligned}$$

and analogous formulas for $<$ instead of $\le$. Note that all these are Δ_0 formulas. They allow one to speak of the value of the superexponential function without, in the second and the third formula, any commitment as to the existence of this value other than that implied by the context. This circumstance is stressed by the use of quotes, although we shall dispense with this practice after the next lemma. Henceforth, it is understood that we are reasoning in terms of these formulas whenever we consider within a formal theory the value of superexponential function whose convergence we do not claim.

The next proposition spells out the coherence conditions between the 'real' and the 'imaginary' values of 2^y_x.

1.5. PROPOSITION ($\mathrm{I}\Delta_0$). (a) $x \le y \leftrightarrow$ '$2^z_x \le 2^z_y$';

(b) $y \le z \leftrightarrow$ '$2^y_x \le 2^z_x$';

(c) '$2^u_x \le 2^v_y$' $\leftrightarrow$ '$2^v_y \not< 2^u_x$';

(d) $z = 2^u_x \rightarrow.\ 2^v_y \le z \leftrightarrow$ '$2^v_y \le 2^u_x$', *and analogously for* $<$ *in place of* $\le$;

(e) $z = 2^u_x \rightarrow.$ '$z \le 2^v_y$' $\leftrightarrow$ '$2^u_x \le 2^v_y$', *also for* $<$;

(f) '$2^y_x \le 2^y_x$';

(g) '$2^u_x \le 2^v_y$' $\wedge$ '$2^v_y \le 2^w_z$' $\rightarrow.$ '$2^u_x \le 2^w_z$';

(h) '$2^u_x \le 2^v_y$' $\wedge\, y < x \rightarrow.\ u < v$.

Proof ($\mathrm{I}\Delta_0$). (a) follows from 1.3(g).

(b) and, for that matter, (f) are immediate.

(c) holds on the strength of the obvious '$t < 2^w_z$' $\vee\, t = 2^w_z \vee t > 2^w_z$.

(d) Let us handle one particular case of the ($\leftarrow$) direction by way of example. Suppose $z = 2^u_x$, '$2^v_y \le 2^u_x$', and $x \ge y$, which implies '$v \le 2^u_{x-y}$'. By 1.3(d) there is a $w \le z$ such that $v \le w = 2^u_{x-y}$ and $z = 2^w_y$. By 1.3(e) one finds a $t \le z$, $t = 2^v_y$, so that $2^v_y \le z = 2^u_x$. If, on top of that, we had '$2^v_y < 2^u_x$', i.e. '$v < 2^u_{x-y}$', then this would result in $v < w$, which would, by 1.3(f), entail $2^v_y < 2^w_y = z$.

(e) follows at once from clauses (c) and (d).

(g) The proof splits into six cases according to the relative ordering of x, y, and z. We only treat one case. Suppose $x \le z \le y$, so that '$u \le 2^v_{y-x}$' and $2^v_{y-z} \le w$. We have to show '$u \le 2^w_{z-x}$'. Assume, for a contradiction, $u > 2^w_{z-x} = t$. Then we have '$2^w_{z-x} \ge 2^v_{y-x}$' as implied by '$2^w_z \ge 2^v_y$', and hence, by clause (d), $t \ge 2^v_{y-x} \ge u$, contradicting $u > t$. Therefore '$u \le 2^w_{z-x}$'.

(h) One considers 2^v_x and argues by contraposition using clauses (a), (b), (g), and (c). ∎

The depth of quantifier alternations in a formula φ is denoted by $\varrho(\varphi)$. For a proof p we set $\varrho(p) = \varrho(\alpha)$ if α is a formula occurring in p with ϱ the largest among such (cf. Visser [18, 2.4]).

1.6. PROPOSITION. *Suppose* F *is a finitely axiomatized theory. Then there is a constant* H *such that for every formula* $\varphi(x)$ *and every* $n \in \omega$, *if* $\mathrm{F} \vdash \varphi(n)$ *then* F *proves* $\varphi(n)$ *by a proof* p *with* $\varrho(p) \le \varrho(\varphi(x)) + H$.

Comment. This is similar to Lemma 2.6 in Pudlák [11]. As concerns free variables, recall that, by our conventions, $\varrho(\varphi(n)) \le \varrho(\varphi(x)) + 1$. ∎

We restrict the notion of proof predicates $x : \Box\,\varphi$ (see Buss [3, §7.3]) to $\Sigma^{\mathrm{b}}_1(\omega_1)$ ones satisfying $\mathrm{I}\Delta_0 + \Omega_1 \vdash x : \Box\,\varphi \leftrightarrow \exists y \le x\,(y : \Box\,\varphi)$, so that the x in $x : \Box\,\varphi$ should be thought of as an upper bound rather than a code of a proof. The underlying formula $\alpha(x)$ specifying the set of non-logical axioms is always presupposed to be $\Delta^{\mathrm{b}}_1(\omega_1)$. One can effectively associate such a proof predicate to any effective presentation of an r.e. theory, although this may involve specifying an axiom set different from (but equivalent to) the original one. If we are speaking of a finitely axiomatized theory then $\alpha(x)$ is assumed to list all the finitely many non-logical axioms:

$$\alpha(x) \equiv \bigvee\!\!\!\!\bigvee_{\text{axioms } \alpha \text{ of T}} x = \ulcorner \alpha \urcorner .$$

Analogously, we introduce a restricted version $x : \Box_k\,\varphi$ which is the same as $x : \Box\,\varphi$ except that it only accepts proofs with $\varrho \le k + K$, where K is some fixed constant which will only be specified when relevant. $\Box\,\varphi$ means $\exists x\,(x : \Box\,\varphi)$ and $\Diamond$ stands for $\neg\,\Box\,\neg$. Similarly for $\Box_k$.

$\mathrm{I}\Delta_0 + \Omega_1$ verifies simple properties of these predicates, like e.g. the closure of $\Box$ under first-order deductive rules and the closure of $\Box_k$ under propositional logic.

1.7. Proposition. *Let $\Box$ be a proof predicate of a theory* T *containing* $\mathrm{I}\Delta_0 + \Omega_1$.

(a) *Let $\sigma(x)$ be a $\Sigma^{\mathrm{b}}_1(\omega_1)$ formula. There is a $k \in \omega$ such that* $\mathrm{I}\Delta_0 + \Omega_1 \vdash \sigma(x) \to \Box_k\,\sigma(x)$.

(b) *Let $\delta(x)$ be a $\Delta_0(\omega_1)$ formula. There is a $k \in \omega$ such that*

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \delta(x) \wedge \exists y\,(y = 2^{2^x}) \to\!.\ \Box_k\,\delta(x).$$

Comment. (a) Buss [3, §7.4] and Wilkie & Paris [20, Theorem 6.4] show that for some finite subtheory S of $\mathrm{I}\Delta_0 + \Omega_1$ it is provable in $\mathrm{I}\Delta_0 + \Omega_1$ that if $\sigma(x)$ holds then S proves $\sigma(x)$ by a proof whose ϱ does not exceed $\varrho(\sigma(x))$ by more than a constant. When dealing with interpreted arithmetic, recall that to each S-proof p there corresponds a T-proof q with $\varrho(q)$ linear in $\varrho(p)$ and, since S is finitely axiomatized, with length of q polynomial in that of p.

(b) This is established by an argument similar to (a) using (the proof of) Proposition 2.10 of Berarducci & Verbrugge [2]. ∎

All *cuts* in this paper are understood to be closed under ω_1 (cf. Hájek & Pudlák [5, III.3(c)]). If $\varphi(x)$ is an arithmetical formula, let $\varphi^{\mathcal{J}}(x)$ be the result of relativizing all unbounded quantifiers in $\varphi(x)$ to a T-cut $\mathcal{J}$. If T contains $\mathrm{I}\Delta_0$ then for every instance ι of Δ_0 induction axiom one has

$\mathrm{T} \vdash \iota^{\mathcal{J}}$ because ι is (equivalent to) a Π_1 sentence. Thus we have the whole of $\mathrm{I}\Delta_0 + \Omega_1$ on any cut.

1.8. PROPOSITION. *Let $\mathcal{J}$ be a cut in a theory* T *containing* $\mathrm{I}\Delta_0$. *For every $k \in \omega$ there exists a cut $\mathcal{K}$ such that* $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x \in \mathcal{K}\, \exists y \in \mathcal{J}\, (y = 2^x_k)$.

Reference. Hájek & Pudlák [5, Theorem III.3.5]. ∎

Let $\mathrm{Cut}\,\mathcal{K}$ be the formula expressing that $\mathcal{K}$ is a cut closed under ω_1.

1.9. PROPOSITION. *Let $\Box$ be a proof predicate of a theory* T *containing* $\mathrm{I}\Delta_0 + \Omega_1$.

(a) *There is a constant D such that*

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall k, \mathcal{J}\, (\Box_k(\mathrm{Cut}\,\mathcal{J}) \to \forall x\, \Box_{k+D}(x \in \mathcal{J})).$$

(b) $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, (\exists y\, (y = 2^x) \to \forall z\, \Box\, \exists w\, (w = 2^z_x))$.

Reference. (a) Visser [17, Lemma 3.4.2].
(b) Pudlák [12, Lemma 2.2] or Visser [17, Fact 3.4.4]. ∎

Let $x : \Box_{\mathrm{tab}}\, \varphi$ be a $\Sigma^{\mathrm{b}}_1(\omega_1)$ predicate formalizing provability by tableaux proofs $\le x$ in a theory T (see Wilkie & Paris [20, 8.9]). The following proposition due to Visser is a formalized variant of cut-elimination.

1.10. PROPOSITION. *Let $\Box$ be a proof predicate of a theory* T. *There are constants N and C such that* $\mathrm{I}\Delta_0 + \Omega_1 \vdash x : \Box_k\, \varphi \wedge y = 2^x_{C+k\cdot N} \to\!. \; y : \Box_{\mathrm{tab}}\, \varphi$ *for every sentence φ.*

Reference. Visser [18, Remark 2.4.8]. ∎

Sequential theories are those that contain $\mathrm{I}\Delta_0 + \Omega_1$ and are capable of reasoning about finite sequences of arbitrary objects, which enables them to construct satisfaction predicates (cf. Hájek & Pudlák [5, III.1(b), III.3(c)]).

1.11. PROPOSITION. *Let* F *be a finitely axiomatized sequential theory and $\Box$ a proof predicate for* T.

(a) *There is an* F*-cut $\mathcal{I}$ such that* $\mathrm{F} \vdash \neg\, \Box^{\mathcal{I}}_{\mathrm{tab}}\, \bot$.

(b) *For every $k \in \omega$ there is an* F*-cut $\mathcal{K}$ such that* $\mathrm{F} \vdash \Diamond^{\mathcal{K}}_k\, \top$.

Comment. (a) See Pudlák [11, Corollary 3.2(i)] or Visser [16, Fact 5.6.6].
(b) follows from (a) and Propositions 1.10 and 1.8. ∎

2. Interpretability

2.1. PROPOSITION. *Suppose a sequential theory* T *interprets a theory* S *containing* $\mathrm{I}\Delta_0$. *Then there is a* T*-cut $\mathcal{I}$ such that* $\mathrm{T} \vdash \pi^{\mathcal{I}}$ *whenever* $\mathrm{S} \vdash \pi$ *and π is a Π_1 sentence.*

Comment. This follows from a lemma of Pudlák saying that an interpretation of S in T gives rise to an isomorphism between a T-cut and an initial segment of the interpreted S-numbers (see Pudlák [11, Lemma 3.3], Visser [16, 5.8], or Visser [18, 2.5.1]). ■

Consider a provability predicate $\triangle$ of a theory S. All the conventions concerning provability predicates are the same for $\triangle$ as for $\Box$, except for the meaning of $\triangle_k$. Namely, use of this notation implies that $\triangle_k$ stands for provability in a finite subtheory of S whose axiom set is singled out by a $\Delta_1^{\mathrm{b}}(\omega_1)$ formula $\alpha_k(x)$ (k a free variable) which satisfies

- for all $k \in \omega$ there is an $m_k \in \omega$ such that $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, (\alpha_k(x) \to x \leq m_k)$;
- $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall k, x\, (\alpha_k(x) \to \alpha_{k+1}(x))$;
- $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall \varphi\, (\triangle\, \varphi \leftrightarrow \exists k\; \triangle_k\, \varphi)$.

$\triangledown$ stands for $\neg \triangle \neg$. Similarly for $\triangledown_k$.

2.2. Proposition. *Suppose there is a cut $\mathcal{I}$ in a theory* T *containing* $\mathrm{I}\Delta_0$ *such that* $\mathrm{T} \vdash \triangledown_k^{\mathcal{I}} \top$ *for all $k \in \omega$, where $\triangle$ is a provability predicate for a theory* S*. Then* T *interprets* S.

Comment. This is established by a variant of the Feferman construction found in Visser [18, Lemma 3.2.1] and Visser [16, 6.2.2.1]. ■

A theory R is *reflexive* if it contains $\mathrm{I}\Delta_0 + \Omega_1$ and $\mathrm{R} \vdash \triangledown_k \top$ for all $k \in \omega$, where $\triangle$ is a provability predicate for R. This property is easily seen to be independent of the particular choice of $\triangle$ complying with our conventions (although it does depend on the way R contains $\mathrm{I}\Delta_0 + \Omega_1$, i.e. on the choice of interpretation of $\mathrm{I}\Delta_0 + \Omega_1$ in R). It is equivalently expressed by saying that R proves consistency of every one of its finite subtheories.

2.3. Theorem. *Let* F *be a finitely axiomatized sequential and* R *a reflexive theory with provability predicates $\Box$ and $\triangle$ respectively. The following are then equivalent*:

(i) F *interprets* R;

(ii) *There is an* F*-cut $\mathcal{I}$ such that* $\mathrm{F} \vdash \pi^{\mathcal{I}}$ *whenever* $\mathrm{R} \vdash \pi$ *for any* Π_1 *sentence* π;

(iii) *There is an* F*-cut $\mathcal{I}$ such that* $\mathrm{F} \vdash \triangledown_n^{\mathcal{I}} \top$ *for all $n \in \omega$*;

(iv) *There is an $m \in \omega$ such that* $\mathrm{I}\Delta_0 + \Omega_1 \vdash \Diamond_m \top \to \triangledown_n \top$ *for all $n \in \omega$.*

Proof. (i)⇒(ii) is immediate by Proposition 2.1.

(ii)⇒(iii) because R is reflexive.

(iii)⇒(i) by Proposition 2.2.

(iii)⇒(iv). Let $\mathcal{I}$ be an F-cut for which we have $\mathrm{F} \vdash \triangledown_n^{\mathcal{I}} \top$ for all $n \in \omega$. Use Propositions 1.6, 1.7(a), and 1.9(a) to find an $m \in \omega$ such that F proves

$\triangledown^{\mathcal{I}}_n \top$ for all $n \in \omega$ by proofs p_n with $\varrho(p_n) \le m+K$, $\mathrm{I}\Delta_0+\Omega_1 \vdash x:\triangle_n \bot \to \Box_m(x:\triangle_n \bot)$, and $\mathrm{I}\Delta_0+\Omega_1 \vdash \forall x\, \Box_m(x \in \mathcal{I})$. Fix an $n \in \omega$ and reason in $\mathrm{I}\Delta_0+\Omega_1$:

Suppose $x:\triangle_n \bot$. Then $\Box_m(x:\triangle_n \bot)$ and $\Box_m(x \in \mathcal{I})$, and so $\Box_m \triangle^{\mathcal{I}}_n \bot$. (Why? Well, let's assume that we are taking $x:\triangle_n \bot$ in the form $\exists y\,(\text{`}y = \overline{x}\text{'} \wedge y:\triangle_n \bot)$ and $x \in \mathcal{I}$ as $\forall y\,(\text{`}y = \overline{x}\text{'} \to \mathcal{I}(y))$. From these two formulas one infers $\exists y\,(\mathcal{I}(y) \wedge y:\triangle_n \bot)$, i.e. $\triangle^{\mathcal{I}}_n \bot$ by a proof whose ϱ does not exceed that of the premises.) On the other hand we have $\Box_m \triangledown^{\mathcal{I}}_n \top$. Thus $\Box_m \bot$ because $\triangledown^{\mathcal{I}}_n \top$ is the negation of $\triangle^{\mathcal{I}}_n \bot$.

Therefore, $\mathrm{I}\Delta_0+\Omega_1 \vdash \triangle_n \bot \to \Box_m \bot$ as was required to show.

(iv)⇒(iii). Immediate by Proposition 1.11(b). ∎

Theorem 2.3 suggests that the case of F = GB and R = ZF is not entirely representative for the general case, since by Corollary 4.3 of Pudlák [11] there is a GB-cut $\mathcal{I}$ such that GB $\vdash \mathrm{Con}^{\mathcal{I}}_{\mathrm{ZF}}$, which appears to be rather stronger than clause (iii) of 2.3. Indeed, if interpretability of F in R implied a similar condition for all pairs (F, R) of theories as above, then the relation of relative interpretability between such theories would be r.e. This will be shown not to be the case in the next section.

3. Σ^0_3-completeness

3.1. Theorem. *Suppose a consistent finitely axiomatized sequential theory* F *interprets a reflexive theory* R. *Then*

(a) *The set* $\{\sigma \in \Sigma_1 \mid \mathrm{F}$ *interprets* $\mathrm{R}+\sigma\}$ *is* Σ^0_3*-complete.*

(b) *If* R *contains* $\mathrm{I}\Delta_0+\mathrm{Exp}$ *then* $\{\pi \in \Pi_1 \mid \mathrm{F}$ *interprets* $\mathrm{R}+\pi\}$ *is also* Σ^0_3*-complete.*

Most of the rest of the present section is devoted to the proof of this theorem. We do (a) and (b) in essentially one go.

3.2. Conventions. (a) We construct a provability predicate $\triangle$ for the theory R which satisfies, apart from our earlier conventions, two additional conditions:

- if $\mathrm{I}\Delta_0+\Omega_1 \vdash \varphi$ then $\triangle_0\, \varphi$ holds;
- $\mathrm{I}\Delta_0+\Omega_1 \vdash \forall n\, \triangle_n(\forall m < n\, \triangledown_m \top)$.

First we define a certain natural number Z and possibly replace the distinguished interpretation of $\mathrm{I}\Delta_0+\Omega_1$ in R by a different one without, however, violating any assumptions on R.

Let $\alpha(x)$ be a $\Delta^{\mathrm{b}}_1(\omega_1)$ formula specifying an axiom set for the theory R. There are two cases:

Case 1: R contains $\mathrm{I}\Delta_0+\mathrm{Exp}$. Since $\mathrm{I}\Delta_0+\mathrm{Exp}$ is a finitely axiomatizable theory (see Hájek & Pudlák [5, Theorem V.5.6]), there is a number Z such that the axioms of R (as specified by $\alpha(x)$) needed to prove translations of the finitely many axioms for $\mathrm{I}\Delta_0 + \mathrm{Exp}$ all have gödelnumber $\leq Z$.

Case 2: R does not contain $\mathrm{I}\Delta_0 + \mathrm{Exp}$. Since R contains $\mathrm{I}\Delta_0 + \Omega_1$ and there exists a finite subtheory S of $\mathrm{I}\Delta_0 + \Omega_1$ that interprets $\mathrm{I}\Delta_0 + \Omega_1$ by relativization to a cut $\mathcal{K}$ and the identical translation of arithmetical operations (see Hájek & Pudlák [5, V.5(c)]), we can find such an interpretation of $\mathrm{I}\Delta_0 + \Omega_1$ in finitely many axioms of R. Let us adopt this cut $\mathcal{K}$ as the distinguished interpretation of $\mathrm{I}\Delta_0 + \Omega_1$ in R and note that R is still reflexive because the consistency of any finite subtheory of R on $\mathcal{K}$ follows from that in the original natural number domain of R. Let Z be the largest among the gödelnumbers of the finite set of axioms of R needed to prove the relativization to $\mathcal{K}$ of $\mathrm{I}\Delta_0 + \Omega_1$.

We use the number Z just constructed to self-referentially define a formula $\alpha_k(x)$ on which the predicates $\triangle_k$ will be based:

$$\alpha_k(x) \;\equiv\; (\alpha(x) \wedge x \leq Z + k) \vee \exists n < k\, (x = \ulcorner \triangledown_n \top \urcorner).$$

The disjunct $\exists n < k\, (x = \ulcorner \triangledown_n \top \urcorner)$ is equivalent to

$$\exists n\, (x = \ulcorner \triangledown_n \top \urcorner) \wedge \forall m \leq x\, (x = \ulcorner \triangledown_m \top \urcorner \to m < k).$$

Since $\exists n\, (x = \ulcorner \triangledown_n \top \urcorner)$ is equivalent to a $\Delta^{\mathrm{b}}_1(\omega_1)$ formula, $\alpha_k(x)$ is also $\Delta^{\mathrm{b}}_1(\omega_1)$. All the conditions $\alpha_k(x)$ that have been promised to satisfy are now easily checked. In particular, that the theory corresponding to $\triangle_k$ is, for standard k, a subtheory of R is established using reflexivity of R and (external) induction on k.

Define $\triangle\,\varphi$ as $\exists k\ \triangle_k\,\varphi$. Caution: $\triangle$ is generally not provably equivalent to the provability predicate based on $\alpha(x)$.

(b) We now select a provability predicate $\Box$ for the theory F. Since F is finitely axiomatized, all the freedom left by our conventions is the choice of the constant K such that $\Box_k$ only accepts proofs with $\varrho \leq k + K$. Along with K, we fix an exhaustive sequence $(\mathcal{J}_i)_{i\in\omega}$ of F-cuts with the function $k \mapsto \ulcorner \mathcal{J}_k \urcorner$ $\Sigma^{\mathrm{b}}_1(\omega_1)$-definable. The following conditions have to be satisfied:

- $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x, n, \varphi\, (x : \triangle_n\, \varphi \to \Box_0 (x : \triangle_n\, \varphi))$;
- $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x, k\ \Box_k\, x \in \mathcal{J}_k$;
- if $\mathrm{F} \vdash \triangledown^{\mathcal{J}_k}_m\, \varphi$ then F proves $\triangledown^{\mathcal{J}_k}_m\, \varphi$ with $\varrho \leq k + K$, for all k, m, φ.

Since $x : \triangle_n\, \varphi$ is $\Delta^{\mathrm{b}}_1(\omega_1)$, the first condition is, in view of Proposition 1.7(a), satisfied by simply taking K sufficiently large. By Proposition 1.9(a), to secure the second condition it suffices to provably have $\forall k\ \Box_{k-D}\, \mathrm{Cut}\, \mathcal{J}_k$ for a certain constant D. Let $\mathcal{N}$ be the trivial cut: $x \in \mathcal{N} \equiv x = x$. Suppose F proves $\mathrm{Cut}\, \mathcal{N}$ by a proof with $\varrho \leq Q$. Choosing $K \geq Q + D$ will guarantee $\forall x\ \Box_0 (x \in \mathcal{N})$. Now we can arrange the sequence $(\mathcal{J}_i)_{i\in\omega}$ by delaying the enumeration of a particular cut $\mathcal{J}$ and patching the sequence with repetitions of $\mathcal{N}$ until a proof of $\mathrm{Cut}\,\mathcal{J}$ with a suitable ϱ appears. Clearly, a procedure like that which eventually enumerates every F-cut can be chosen to result in a $\Sigma^{\mathrm{b}}_1(\omega_1)$-definable function $k \mapsto \ulcorner \mathcal{J}_k \urcorner$. Note that we necessarily have $\varrho(\mathcal{J}_k) \leq k + E$ for some constant E. With these provisions for $(\mathcal{J}_i)_{i\in\omega}$ the third condition is, in view of Proposition 1.6, satisfied by possibly increasing the value of K still further.

3.3. Definition. Consider an arbitrary unary r.e. predicate $S(n)$. One can effectively associate to it a Δ_0 formula $S(n)\!\downarrow\, \leq x$, where n and x are free variables, with the property that $S(n)$ holds true just in case there is an x for which $S(n)\!\downarrow\, \leq x$ (see Hájek & Pudlák [5, V.4(c)]). Let $S(n)\!\downarrow\, > x$ be its negation. To this formula we relate a collection of self-referentially defined formulas:

$$\begin{aligned}
&\textit{pre-trouble}_{\Box}(x) \;\equiv\; \exists k, m \leq \log x\, (\#_x\{n < m \mid S(n)\!\downarrow\, > x\} \geq k \\
&\hspace{18em} \wedge\, x : \Box_k \nabla^{\mathcal{J}_k}_m \pi), \\
&\textit{pre-trouble}_{\triangle}(x) \;\equiv\; \exists m \leq \log x\, (x : \triangle_m \neg\sigma), \\
&\textit{pre-trouble}(x) \;\equiv\; \textit{pre-trouble}_{\Box}(x) \vee \textit{pre-trouble}_{\triangle}(x), \\
&\textit{min-trouble}(x) \;\equiv\; \textit{pre-trouble}(x) \wedge \forall y < x\, \neg \textit{pre-trouble}(y), \\
&\textit{trouble}_{\Box}(x,k) \;\equiv\; k \leq \log x \wedge \exists z \leq x\, \exists m \leq \log x\, (\textit{min-trouble}(z) \\
&\hspace{10em} \wedge\, \#_x\{n < m \mid S(n)\!\downarrow\, > z\} \geq k \wedge x : \Box_k \nabla^{\mathcal{J}_k}_m \pi), \\
&\textit{trouble}_{\triangle}(x,k) \;\equiv\; k \leq \log x \wedge \exists z \leq x\, \exists m \leq \log x\, (\textit{min-trouble}(z) \\
&\hspace{12em} \wedge\, \#_x\{n < m \mid S(n)\!\downarrow\, > z\} \leq k \wedge x : \triangle_m \neg\sigma), \\
&(\textit{trouble}_{\Box} \not\preceq \textit{trouble}_{\triangle})(u) \;\equiv\; \forall x, k\, (2^x_{k\cdot N} \leq u \wedge \textit{trouble}_{\Box}(x,k) \\
&\hspace{14em} \rightarrow\!.\; \exists y, l\, (2^y_{l\cdot N} < 2^x_{k\cdot N} \wedge \textit{trouble}_{\triangle}(y,l))), \\
&(\textit{trouble}_{\triangle} \prec \textit{trouble}_{\Box})(u) \;\equiv\; \exists x, k\, (2^x_{k\cdot N} \leq u \wedge \textit{trouble}_{\triangle}(x,k) \\
&\hspace{10em} \wedge\, \forall y, l\, (2^y_{l\cdot N} \leq u \wedge \textit{trouble}_{\Box}(y,l) \rightarrow\!.\; 2^x_{k\cdot N} < 2^y_{l\cdot N})), \\
&\pi \;\equiv\; \forall u\, (\textit{trouble}_{\Box} \not\preceq \textit{trouble}_{\triangle})(u), \\
&\sigma \;\equiv\; \exists v\, \exists u \leq \log\log v\, (\textit{trouble}_{\triangle} \prec \textit{trouble}_{\Box})(u).
\end{aligned}$$

Here the constant N is the one corresponding to the predicate $\Box$ by Proposition 1.10. For future reference we also fix the other constant C of the same proposition.

Observe that all the formulas introduced are $\Delta_0(\omega_1)$ with the exception of the last two sentences. The sentence π is clearly Π_1 and σ is Σ_1. Moreover, σ is (equivalent to) an $\exists\Delta^{\mathrm{b}}_1(\omega_1)$ sentence, for all quantifiers in $(\textit{trouble}_{\triangle} \prec \textit{trouble}_{\Box})(u)$ can be bounded by $\log v$ for almost all values of the quantified variables. This is because $u \leq \log\log v$ and any ω_1-term $f(u)$ is eventually dominated by $\log v$.

3.4. Lemma $(\mathrm{I}\Delta_0 + \Omega_1)$. $\sigma \to \pi$.

Proof $(\mathrm{I}\Delta_0 + \Omega_1)$. This is because $(\mathit{trouble}_\triangle \prec \mathit{trouble}_\square)(w)$ implies $(\mathit{trouble}_\triangle \prec \mathit{trouble}_\square)(u)$ for all $u \geq w$, entailing $(\mathit{trouble}_\square \not\preceq \mathit{trouble}_\triangle)(u)$, which in turn implies $(\mathit{trouble}_\square \not\preceq \mathit{trouble}_\triangle)(v)$ for all $v \leq u$. ■

3.5. Definition. We introduce some convenient abbreviations:

$$\begin{aligned}
&\mathit{trouble}(x,k) \equiv \mathit{trouble}_\square(x,k) \vee \mathit{trouble}_\triangle(x,k),\\
&\mathit{first\text{-}trouble}_\square\lceil w(x,k) \equiv x \leq w \wedge \mathit{trouble}_\square(x,k)\\
&\qquad\qquad \wedge \forall y \leq w \, \forall l \leq \log y \, (\mathit{trouble}(y,l) \to 2^x_{k\cdot N} \leq 2^y_{l\cdot N}),\\
&\mathit{first\text{-}trouble}_\triangle\lceil w(x,k) \equiv x \leq w \wedge \mathit{trouble}_\triangle(x,k)\\
&\qquad\qquad \wedge \forall y \leq w \, \forall l \leq \log y \, (\mathit{trouble}_\square(y,l) \to 2^x_{k\cdot N} < 2^y_{l\cdot N})\\
&\qquad\qquad \wedge \forall y \leq w \, \forall l \leq \log y \, (\mathit{trouble}_\triangle(y,l) \to 2^x_{k\cdot N} \leq 2^y_{l\cdot N}),\\
&(\mathit{trouble}_\square \preceq \mathit{trouble}_\triangle)\lceil w \equiv \exists x \leq w \, \exists k \leq \log x \, \mathit{first\text{-}trouble}_\square\lceil w(x,k),\\
&(\mathit{trouble}_\triangle \prec \mathit{trouble}_\square)\lceil w \equiv \exists x \leq w \, \exists k \leq \log x \, \mathit{first\text{-}trouble}_\triangle\lceil w(x,k).
\end{aligned}$$

Observe that in view of Convention 1.4 all these formulas are $\Delta_0(\omega_1)$.

3.6. Lemma. $(\mathrm{I}\Delta_0 + \Omega_1)$. (a) $\mathit{pre\text{-}trouble}(x) \to \exists y \leq x \ \mathit{min\text{-}trouble}(y)$;
(b) $\mathit{min\text{-}trouble}(x) \to \exists k \leq \log x \ \mathit{trouble}(x,k)$.

Proof. Easy. ■

3.7. Lemma $(\mathrm{I}\Delta_0 + \Omega_1)$. *We have*

$$\mathit{trouble}(x,k) \wedge x \leq u \to. \ (\mathit{trouble}_\square \preceq \mathit{trouble}_\triangle)\lceil u \vee (\mathit{trouble}_\triangle \prec \mathit{trouble}_\square)\lceil u.$$

Proof $(\mathrm{I}\Delta_0 + \Omega_1)$. Consider the following formula:

$$\begin{aligned}
l \sqsubseteq_u m \equiv \ &\forall z \leq u \, (m \leq \log z \wedge \mathit{trouble}(z,m)\\
&\qquad \to. \ \exists y \leq u \, (l \leq \log y \wedge \mathit{trouble}(y,l) \wedge 2^y_{l\cdot N} \leq 2^z_{m\cdot N})).
\end{aligned}$$

In view of Proposition 1.5(f) and (g), $\sqsubseteq_u$ is a reflexive transitive $\Delta_0(\omega_1)$ relation on numbers $\leq \log u$, hence by Proposition 1.1 there is an $n \leq \log u$ which is a minimum in this preordering. Both $n \sqsubseteq_u k$ and $k \not\sqsubseteq_u n$ imply that there is a (smallest) $v \leq u$ with $\mathit{trouble}(v,n)$ and $2^v_{n\cdot N} \leq 2^x_{k\cdot N}$. If we have $\mathit{trouble}_\square(w,p)$ for some $w \leq u$, $p \leq \log w$ such that $2^w_{p\cdot N} = 2^v_{n\cdot N}$ then, clearly, $\mathit{first\text{-}trouble}_\square\lceil u(w,p)$ holds. Otherwise one has $\mathit{first\text{-}trouble}_\triangle\lceil u(v,n)$. In either case the conclusion of the lemma is satisfied. ■

3.8. Lemma $(\mathrm{I}\Delta_0 + \Omega_1)$. (a) $\pi \wedge \mathit{trouble}(x,k) \wedge u \geq 2^x_{k\cdot N} \to. \ (\mathit{trouble}_\triangle \prec \mathit{trouble}_\square)\lceil u$;

(b) $\mathit{trouble}(x,k) \wedge \exists u, v \, (u \geq 2^x_{k\cdot N} \wedge v = 2^{2^u} \wedge (\mathit{trouble}_\triangle \prec \mathit{trouble}_\square)\lceil u) \to. \ \sigma$.

Proof $(\mathrm{I}\Delta_0+\Omega_1)$. (a) By Lemma 3.7, we only have to exclude $(\mathit{trouble}_\square \preceq \mathit{trouble}_\triangle)\lceil u$ because, by Proposition 1.3(g), $x \leq u$. So suppose we had

$\mathit{first\text{-}trouble}_{\Box}\lceil u(y,l)$. We would then also have $2^y_{l\cdot N} \le 2^x_{k\cdot N}$. Consider $w = 2^y_{l\cdot N} \le u$. Since there is no $z \le w$, $m \le \log z$ such that $\mathit{trouble}_{\triangle}(z,m)$ and $2^z_{m\cdot N} < 2^y_{l\cdot N} = w$, we have $\neg(\mathit{trouble}_{\Box} \not\preceq \mathit{trouble}_{\triangle})(w)$, and hence $\neg\pi$. This contradiction proves $(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u$.

(b) Consider $y \le u$, $l \le \log y$ such that $\mathit{first\text{-}trouble}_{\triangle}\lceil u(y,l)$. Since $2^y_{l\cdot N} \le 2^x_{k\cdot N} \le u$, one clearly has $(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})(u)$ and hence σ, for 2^{2^u} exists. ∎

3.9. Lemma ($\mathrm{I}\Delta_0 + \Omega_1$). *We have*

$$\pi \wedge \mathit{first\text{-}trouble}_{\Box}\lceil y(y,l) \wedge \exists v\,(v = 2^y_{2+l\cdot N}) \wedge \mathit{min\text{-}trouble}(z)$$
$$\wedge\, p \le \log y \wedge \#_y\{n < p \mid S(n){\downarrow} > z\} \ge l \rightarrow\!.\ \exists q < p\,(\triangle_q \bot).$$

Proof ($\mathrm{I}\Delta_0 + \Omega_1$). Suppose the antecedent of the above statement holds and consider $u = 2^y_{l\cdot N} \ge y$. Since we have $\mathit{trouble}(y,l)$, there follows $(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u$ by Lemma 3.8(a). $(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u$ says that there are $x \le u$ and $k \le \log x$ such that $\mathit{first\text{-}trouble}_{\triangle}\lceil u(x,k)$. This implies that we cannot have $2^y_{l\cdot N} < 2^x_{k\cdot N}$, for this together with $\mathit{trouble}_{\Box}(y,l)$ contradicts $\mathit{first\text{-}trouble}_{\triangle}\lceil u(x,k)$. Therefore $2^x_{k\cdot N} \le 2^y_{l\cdot N}$. Also, since $u \ge y$, $\mathit{first\text{-}trouble}_{\triangle}\lceil u(x,k)$ and $\mathit{first\text{-}trouble}_{\Box}\lceil y(y,l)$ would conflict unless $x > y$. There is at most one w with $\mathit{min\text{-}trouble}(w)$ and therefore $\mathit{trouble}_{\triangle}(x,k)$ says that $\#_x\{n < q \mid S(n){\downarrow} > z\} \le k$ for some $q \le \log x$ such that $\triangle_q \neg\sigma$. By Proposition 1.5(h) we have $k < l$. Taking into account that $l \le \#_y\{n < p \mid S(n){\downarrow} > z\}$, this entails

$$\#_x\{n < q \mid S(n){\downarrow} > z\} \le k < l \le \#_y\{n < p \mid S(n){\downarrow} > z\}$$

whence $q < p$ follows by Proposition 1.2(c) and (e).

Since $v = 2^{2^u}$ exists we have σ by Lemma 3.8(a). Therefore $\triangle_q \sigma$ by Proposition 1.7(a), which together with $\triangle_q \neg\sigma$ gives $\triangle_q \bot$. ∎

3.10. Lemma ($\mathrm{I}\Delta_0 + \Omega_1$). *We have*

$$\mathit{trouble}(x,k) \wedge \exists w\,(w = 2^x_{3+C+k\cdot N})$$
$$\rightarrow\!.\ \Box_{\mathrm{tab}} \bot \vee (\triangle \bot \wedge \forall m\,(\triangle_m \neg\sigma \rightarrow \triangle_m \bot)).$$

Proof ($\mathrm{I}\Delta_0 + \Omega_1$). Consider $u = 2^x_{k\cdot N} \le w$. Since we have $\mathit{trouble}(x,k)$, there holds, by Lemma 3.7,

$$(\mathit{trouble}_{\Box} \preceq \mathit{trouble}_{\triangle})\lceil \quad \text{or} \quad (\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u.$$

Case 1: $(\mathit{trouble}_{\Box} \preceq \mathit{trouble}_{\triangle})\lceil u$. For some y, l such that $2^y_{l\cdot N} \le 2^x_{k\cdot N} = u$ we have $\mathit{first\text{-}trouble}_{\Box}\lceil u(y,l)$ and therefore $\mathit{first\text{-}trouble}_{\Box}\lceil y(y,l)$, whence $y : \Box_l \nabla_p^{\mathcal{J}_l} \pi$ for some $p \le \log y$ with $\#_y\{n < p \mid S(n){\downarrow} > z\} \ge l$, where z is such that $\mathit{min\text{-}trouble}(z)$.

As $2^l \le y$ exists, so does $2^{2+l\cdot N}$, hence by Proposition 1.9(b) and Convention 3.2(a) there is a standard ω_1-term c such that $c(y) : \triangle_p \exists v\,(v = 2^y_{2+l\cdot N})$.

By Proposition 1.7(b) for some standard ω_1-term d we have

$$d(2^{2^y}) : \triangle_p(\mathit{first\text{-}trouble}_{\Box}\lceil y(y,l) \wedge \mathit{min\text{-}trouble}(z) \\ \wedge\, p \le \log y \wedge \#_y\{n < p \mid S(n)\!\downarrow\, > z\} \ge l).$$

Hence by Lemma 3.9 there is a standard ω_1-term e such that $e(2^{2^y}) : \triangle_p(\pi \to \exists q < p\,(\triangle_q \bot))$ and so by Convention 3.2(a), $t = f(2^{2^y}) : \triangle_p \neg\pi$, where f also is a standard ω_1-term.

By Convention 3.2(b) for some standard ω_1-terms g and h we have $g(f(2^{2^y})) : \Box_l(t : \triangle_p \neg\pi)$ and $h(f(2^{2^y})) : \Box_l(t \in \mathcal{J}_l)$. Therefore $i(2^{2^y}) : \Box_l \triangle_p^{\mathcal{J}_l} \neg\pi$ for some standard ω_1-term i. Recalling that $y : \Box_l \nabla_p^{\mathcal{J}_l} \pi$, we have yet another standard ω_1-term j with $j(2^{2^y}) : \Box_l \bot$. By Proposition 1.10 this leads to $2^{j(2^{2^y})}_{C+l\cdot N} : \Box_{\mathrm{tab}} \bot$ once we know that $2^{j(2^{2^y})}_{C+l\cdot N}$ exists. Since $l \le \log y$ and $2^{2^{2^y}} \ge j(2^{2^y})$ for all but standard-finitely many y, this follows from the existence of $2^{2^{2^{2^y}}}_{C+l\cdot N} = 2^y_{3+C+l\cdot N} \le 2^x_{3+C+k\cdot N}$. So $\Box_{\mathrm{tab}} \bot$.

Case 2: $(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u$. Since $v = 2^{2^u}$ exists, we have $\triangle_0(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u$ and $\triangle_0\, \mathit{trouble}(x,k)$ by Proposition 1.7(b), for $x \le u$. Hence there follows $\triangle_0\, \sigma$ by Lemma 3.8(b). Now $(\mathit{trouble}_{\triangle} \prec \mathit{trouble}_{\Box})\lceil u$ implies $y : \triangle \neg\sigma$. Therefore $\triangle \bot$. If we, on top of that, had $\triangle_m \neg\sigma$ for some m then we would also have $\triangle_m \bot$.

Thus the two cases correspond to the two disjuncts of the conclusion of the present lemma. ∎

3.11. Lemma. *pre-trouble*(a) *holds for no* $a \in \omega$.

Proof. By Lemma 3.6, *pre-trouble*(a) leads, via *min-trouble*(b) for some $b \le a$, to *trouble*(b,n) for an appropriate $n \le \log b$. By Lemma 3.10 this results in the inconsistency of either F or R, which we have assumed not to be the case. ∎

Since F interprets R we can by Theorem 2.3 fix a cut $\mathcal{I}$ such that $\mathrm{F} \vdash \nabla^{\mathcal{I}}_n \top$ for all $n \in \omega$. By Proposition 1.11(a) we may also assume $\mathrm{F} \vdash \Diamond^{\mathcal{I}}_{\mathrm{tab}} \top$.

3.12. Lemma. (a) *Let* $m \in \omega$ *and* $\#\{n < m \mid S(n)\!\uparrow\} \le k$. *Consider a cut* $\mathcal{K}$ *such that* $\mathrm{F} \vdash \forall x \in \mathcal{K}\, \exists y \in \mathcal{I}\, (y = 2^x_{3+C+k\cdot N})$. *We have* $\mathrm{F} \vdash \nabla^{\mathcal{K}}_m \sigma$.

(b) *Let* $m \in \omega$ *and* $\#\{n < m \mid S(n)\!\uparrow\} \ge k$. *Then* $\mathrm{F} \nvdash \nabla^{\mathcal{J}_k}_m \pi$.

Proof. (a) Since $\#\{n < m \mid S(n)\!\uparrow\} \le k$, there is an $a \in \omega$ such that $\#\{n < m \mid S(n)\!\downarrow\, > a\} = \#\{n < m \mid S(n)\!\uparrow\} \le k$. By Lemma 3.11, $\neg$*pre-trouble*(a). Next reason in F:

Suppose $x \in \mathcal{K}$ and $x : \triangle_m \neg\sigma$. We clearly can assume $x \ge a$ and $\log x \ge k, m$, so *pre-trouble*$_{\triangle}(x,k)$. By Lemma 3.6(a) there is a $z \le x$ with *min-trouble*(z). Since $\neg$*pre-trouble*(a), we have $z > a$. Therefore

$$\#_{2^m}\{n < m \mid S(n){\downarrow} > z\} \le \#_{2^m}\{n < m \mid S(n){\downarrow} > a\} \le k$$

by Proposition 1.2(b), and *trouble*$_\triangle(x,k)$ holds. Hence by Lemma 3.10, $\Box^{\mathcal{I}}_{\mathrm{tab}}\,\bot$ or $\triangle^{\mathcal{I}}_m\,\bot$, for $2^x_{3+C+k\cdot N} \in \mathcal{I}$. This, however, contradicts the choice of the cut $\mathcal{I}$.

Thus $\mathrm{F} \vdash \nabla^{\mathcal{K}}_m\,\sigma$ as required.

(b) Suppose $\mathrm{F} \vdash \nabla^{\mathcal{J}_k}_m\,\pi$. Then by Convention 3.2(b), the theory F proves the same sentence by a proof p such that $\varrho(p) \le k + K$, i.e. $p : \Box_k\,\nabla^{\mathcal{J}_k}_m\,\pi$. As $\#\{n < m \mid S(n){\uparrow}\} \ge k$, we also have $\#\{n < m \mid S(n){\downarrow} > p\} \ge k$. Hence *pre-trouble*$_\Box(p)$, contrary to Lemma 3.11. So $\mathrm{F} \nvdash \nabla^{\mathcal{J}_k}_m\,\pi$. ∎

3.13. Lemma. (a) $\mathrm{R} + \sigma$ *is reflexive*, *i.e.* $\mathrm{R} \vdash \sigma \to \nabla_n\,\sigma$ *for all* $n \in \omega$.

(b) *If* R *contains* $\mathrm{I}\Delta_0 + \mathrm{Exp}$ *then* $\mathrm{R} + \pi$ *is also reflexive.*

Proof. (a) Since σ is $\exists\Sigma^{\mathrm{b}}_1(\omega_1)$ we have $\mathrm{R} \vdash \sigma \to \triangle_n\,\sigma$ by Proposition 1.7(a). Hence $\mathrm{R} \vdash \sigma \to \nabla_n\,\sigma$, for $\mathrm{R} \vdash \nabla_n\,\top$.

(b) Fix an $n \in \omega$ and reason in R:

Assume π and $\triangle_n\,\neg\pi$ and reason towards a contradiction. By Lemma 3.4 one has $x{:}\triangle_n\,\neg\sigma$ for some x such that $\log x{\ge}n$. We then have *pre-trouble*$_\triangle(x)$ and hence *min-trouble*(z) for some $z \le x$. Now $\#_{2^n}\{m < n \mid S(m){\downarrow} > z\} \le n$ by Proposition 1.2(d) and so *trouble*$_\triangle(x,n)$. Therefore, by π, we have $($*trouble*$_\triangle \prec$ *trouble*$_\Box)\lceil 2^x_{n\cdot N}$ and so σ by Lemma 3.8 since exponentiation is available. Hence $\triangle_n\,\sigma$ and $\triangle_n\,\pi$ by Lemma 3.4. But $\triangle_n\,\pi$ and $\triangle_n\,\neg\pi$ result in $\triangle_n\,\bot$, which contradicts the reflexivity of R.

Thus $\mathrm{R} \vdash \pi \to \nabla_n\,\pi$ for any $n \in \omega$. ∎

3.14. Lemma. (a) *Suppose* $\{n \in \omega \mid S(n){\downarrow}\}$ *is cofinite. Then* F *interprets both* $\mathrm{R} + \sigma$ *and* $\mathrm{R} + \pi$.

(b) *Suppose* $\{n \in \omega \mid S(n){\downarrow}\}$ *is not cofinite. Then* F *does not interpret* $\mathrm{R} + \sigma$.

(c) *If* $\{n \in \omega \mid S(n){\downarrow}\}$ *is not cofinite and* R *contains* $\mathrm{I}\Delta_0 + \mathrm{Exp}$ *then* F *does not interpret* $\mathrm{R} + \pi$.

Proof. (a) If $\{n \in \omega \mid S(n){\downarrow}\}$ is cofinite then there is a $k \in \omega$ such that $\#\{n < m \mid S(n){\uparrow}\} \le k$ for any $m \in \omega$. By Proposition 1.8 pick a cut $\mathcal{K}$ satisfying the condition of Lemma 3.12(a). We then have $\mathrm{F} \vdash \nabla^{\mathcal{K}}_m\,\sigma$ for any $m \in \omega$. By Proposition 2.2 this means that F interprets $\mathrm{R} + \sigma$. Therefore F interprets $\mathrm{R} + \pi$ by Lemma 3.4.

(b) and (c). Assume $\{n \in \omega \mid S(n){\downarrow}\}$ is not cofinite and consider an arbitrary F-cut $\mathcal{J}$. By our assumptions we have $\mathcal{J} = \mathcal{J}_k$ for some $k \in \omega$. Since $\{n \in \omega \mid S(n){\uparrow}\}$ is infinite, there exists an $m \in \omega$ such that $\#\{n < m \mid S(n){\uparrow}\} \ge k$. By Lemma 3.12(b) this implies that $\mathrm{F} \nvdash \nabla^{\mathcal{J}}_m\,\pi$. Hence

$\mathrm{F} \nvdash \triangledown^{\mathcal{J}}_{m}\, \sigma$ by Lemma 3.4. So by Theorem 2.3 and Lemma 3.13(a), F does not interpret $\mathrm{R} + \sigma$.

If, on top of that, R contains $\mathrm{I}\Delta_0 + \mathrm{Exp}$ then by Lemma 3.13(b), $\mathrm{R} + \pi$ is reflexive and hence, by Theorem 2.3, F does not interpret $\mathrm{R} + \pi$ either. ■

3.15. Proof of Theorem 3.1. Definition 3.3 provides an effective way to construct a Π_1 sentence π_S and a Σ_1 sentence σ_S from an index S of an r.e. set. The set of S such that $\{n \in \omega \mid S(n)\downarrow\}$ is cofinite is known to be Σ^0_3-complete. We have:

(a) F interprets $\mathrm{R} + \sigma_S$ iff $\{n \in \omega \mid S(n)\downarrow\}$ is cofinite by Lemma 3.14(a) and (b).

(b) Assuming $\mathrm{R} \vdash \mathrm{I}\Delta_0 + \mathrm{Exp}$, F interprets $\mathrm{R} + \pi_S$ iff $\{n \in \omega \mid S(n)\downarrow\}$ is cofinite by Lemma 3.14(a) and (c).

The theorem follows. ■

3.16. Remarks. Note that both $\mathrm{R} + \sigma_S$ and $\mathrm{R} + \pi_S$ are, in view of Lemma 3.12(a), locally interpretable in F regardless of the behaviour of S.

Švejdar [15] constructs a Π_1 sentence π such that GB interprets $\mathrm{ZF} + \pi$ but neither GB interprets $\mathrm{GB} + \pi$ nor ZF interprets $\mathrm{ZF} + \pi$. Observe that since a Σ^0_3-complete set cannot be the union of two sets of lower complexity, our theorem provides a supplement to Švejdar's result in that the sentence π can also be chosen Σ_1. (Not that this could not be obtained by Švejdar's methods, though.)

In Theorem 3.1(b), instead of requiring that R contain $\mathrm{I}\Delta_0 + \mathrm{Exp}$, we could have imposed the condition that R be Σ_1-*essentially reflexive*, i.e. $\mathrm{R} \vdash \triangle_n\, \sigma \to \sigma$ for all $n \in \omega$ and Σ_1 sentences σ, since Lemma 3.13(b) is the only point where Exp is needed. Also, the $\triangle_0(\mathrm{I}\Delta_0 + \Omega_1)$ clause of Convention 3.2(a) would have to be weakened in that $\mathrm{I}\Delta_0 + \Omega_1$ would have to be replaced by a finite part of that theory sufficiently large for our arguments. Σ_1-essential reflexivity and Exp are independent of one another for reflexive theories, although they are both implied by *uniform* Σ_1*-essential reflexivity*: $\mathrm{R} \vdash \triangle_n\, \sigma(x) \to \sigma(x)$, for all $n \in \omega$ and Σ_1 formulas $\sigma(x)$.

Finally, we present an example to the effect that the unpleasant restriction on R in Theorem 3.1(b) cannot be completely removed.

3.17. Example. *There exists a finitely axiomatized sequential theory* F *interpreting a reflexive theory* R *with* $\{\pi \in \Pi_1 \mid \mathrm{F}$ *interprets* $\mathrm{R} + \pi\}$ *r.e.*

Consider $\mathrm{F} = \mathrm{I}\Delta_0 + \mathrm{Superexp}$, which is finitely axiomatized sequential, and $\mathrm{R} = (\mathrm{I}\Delta_0 + \Omega_1)_\omega = \mathrm{I}\Delta_0 + \Omega_1 + \triangledown \top + \triangledown \triangledown \top + \ldots$, where $\triangle$ is the provability predicate of $\mathrm{I}\Delta_0 + \Omega_1$. Clearly, $(\mathrm{I}\Delta_0 + \Omega_1)_\omega$ is reflexive. By Theorem 3.3 of Sieg [13], $\mathrm{I}\Delta_0 + \mathrm{Superexp}$ proves Σ_1 reflection for $\mathrm{I}\Delta_0 + \mathrm{Exp}$ and hence for $\mathrm{I}\Delta_0+\Omega_1$ as well. Therefore $\mathrm{I}\Delta_0+\mathrm{Superexp}$ proves every theorem of $(\mathrm{I}\Delta_0+\Omega_1)_\omega$ and thus interprets $(\mathrm{I}\Delta_0+\Omega_1)_\omega$.

The following claim shows that $\{\pi\in\Pi_1 \mid \mathrm{I}\Delta_0+\mathrm{Superexp}$ interprets $(\mathrm{I}\Delta_0+\Omega_1)_\omega+\pi\}$ is r.e.

Claim. *Let* π *be a* Π_1 *sentence.* $\mathrm{I}\Delta_0+\mathrm{Superexp}$ *interprets* $(\mathrm{I}\Delta_0+\Omega_1)_\omega+\pi$ *iff there is an* $(\mathrm{I}\Delta_0+\mathrm{Superexp})$*-cut* $\mathcal{J}$ *such that* $\mathrm{I}\Delta_0+\mathrm{Superexp}\vdash\pi^{\mathcal{J}}$.

$\mathrm{I}\Delta_0+\mathrm{Superexp}$ proves every theorem of $(\mathrm{I}\Delta_0+\Omega_1)_\omega$ relativized to any $(\mathrm{I}\Delta_0+\mathrm{Superexp})$-cut because the only non-Π_1 axiom of $(\mathrm{I}\Delta_0+\Omega_1)_\omega$ is Ω_1 and we have assumed that all cuts are closed under ω_1. If $\pi^{\mathcal{J}}$ is also proved then $\mathcal{J}$ defines an interpretation of $(\mathrm{I}\Delta_0+\Omega_1)_\omega+\pi$ in $\mathrm{I}\Delta_0+\mathrm{Superexp}$.

Conversely, if $\mathrm{I}\Delta_0+\mathrm{Superexp}$ interprets $(\mathrm{I}\Delta_0+\Omega_1)_\omega+\pi$ then $\mathrm{I}\Delta_0+\mathrm{Superexp}\vdash\pi^{\mathcal{J}}$ for an appropriate $\mathcal{J}$ by Proposition 2.1. ∎

References

[1] A. Berarducci and P. D'Aquino, Δ_0*-complexity of the relation* $y=\prod_{i\le n}F(i)$, Ann. Pure Appl. Logic 75 (1995), 49–56.

[2] A. Berarducci and R. Verbrugge, *On the provability logic of bounded arithmetic*, ibid. 61 (1993), 75–93.

[3] S. R. Buss, *Bounded Arithmetic*, Bibliopolis, Napoli, 1986.

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

[5] P. Hájek and P. Pudlák, *Metamathematics of First-Order Arithmetic*, Springer, Berlin, 1993.

[6] R. G. Jeroslow, *Non-effectiveness in S. Orey's arithmetical compactness theorem*, Z. Math. Logik Grundlagen Math. 17 (1971), 285–289.

[7] P. Lindström, *Some results on interpretability*, in: Proc. 5th Scandinavian Logic Sympos., F. V. Jensen, B. H. Mayoh and K. K. Møller (eds.), Aalborg Univ. Press, 1979, 329–361.

[8] —, *On partially conservative sentences and interpretability*, Proc. Amer. Math. Soc. 91 (1984), 436–443.

[9] J. Paris and A. Wilkie, *Counting* Δ_0 *sets*, Fund. Math. 127 (1986), 67–76.

[10] J. B. Paris, A. J. Wilkie and A. R. Woods, *Provability of the pigeonhole principle and the existence of infinitely many primes*, J. Symbolic Logic 53 (1988), 1235–1244.

[11] P. Pudlák, *Cuts, consistency statements and interpretations*, ibid. 50 (1985), 423–441.

[12] —, *On the length of proofs of finitistic consistency statements in first order theories*, in: Logic Colloquium '84, J. B. Paris, A. J. Wilkie and G. M. Wilmers (eds.), North-Holland, Amsterdam, 1986, 165–196.

[13] W. Sieg, *Fragments of arithmetic*, Ann. Pure Appl. Logic 28 (1985), 33–71.

[14] C. Smoryński, *Nonstandard models and related developments*, in: Harvey Friedman's Research on the Foundations of Mathematics, L. A. Harrington, M. D. Morley, A. Ščedrov and S. G. Simpson (eds.), North-Holland, Amsterdam, 1985, 179–229.

[15] V. Švejdar, *A sentence that is difficult to interpret*, Comment. Math. Univ. Carolin. 22 (1981), 661–666.

[16] A. Visser, *Interpretability logic*, in: Mathematical Logic, P. P. Petkov (ed.), Plenum Press, New York, 1990, 175–209.

[17] —, *An inside view of* EXP; *or, the closed fragment of the provability logic of* I$\Delta_0 + \Omega_1$ *with a propositional constant for* EXP, J. Symbolic Logic 57 (1992), 131–165.

[18] —, *The unprovability of small inconsistency*, *A study of local and global interpretability*, Arch. Math. Logic 32 (1993), 275–298.

[19] A. J. Wilkie, *On sentences interpretable in systems of arithmetic*, in: Logic Colloquium '84, J. B. Paris, A. J. Wilkie and G. M. Wilmers (eds.), North-Holland, Amsterdam, 1986, 329–342.

[20] A. J. Wilkie and J. B. Paris, *On the scheme of induction for bounded arithmetic formulas*, Ann. Pure Appl. Logic 35 (1987), 261–302.

Department of Philosophy
Utrecht University
Heidelberglaan 8
3584 CS Utrecht, The Netherlands
E-mail: volodya@phil.ruu.nl

Received 15 August 1995;
in revised form 22 October 1996