Notre Dame Journal of Formal Logic
Volume 35, Number 2, Spring 1994

A Smart Child of Peano's

V. Yu. SHAVRUKOV

Abstract We calculate the provability logic of a special form of the Feferman provability predicate together with the usual provability predicate of Peano Arithmetic. In other words, we construct a bimodal system with the intended interpretation of the expression $\Box\varphi$ being as usual the formalization of "φ is provable in PA" and the new modal operator $\triangle$ standing, when applied to φ, for "there exists an x s.t. $I\Sigma_x$ is consistent and proves φ". The new system is called **LF**. We construct a Kripke semantics for **LF** and prove the arithmetical completeness theorem for this system. A small number of other issues concerning the Feferman predicate, such as uniqueness of gödel-sentences for $\triangle$, is also considered.

1 Introduction The Feferman provability predicate for reflexive recursively axiomatized theories emerged for the first time in Feferman's paper [2]. Starting with a reflexive theory, Peano Arithmetic PA being the conventional example, one chooses a sequence of finitely axiomatized theories $(\mathrm{PA}\lceil n)_{n\in\omega}$ with $\mathrm{PA}\lceil n+1$ extending $\mathrm{PA}\lceil n$, and $\mathrm{PA} = \bigcup_{n\in\omega} \mathrm{PA}\lceil n$.

Outside modal-logical contexts, let us write $\triangle$ for the Feferman predicate reserving the shorthand $\Box$ for the usual provability predicate. $\triangle\varphi$ is then defined as the formalization of

"there exists an $x \in \omega$ s.t. $\mathrm{PA}\lceil x$ is consistent and $\mathrm{PA}\lceil x \vdash \varphi$".

The sequence $(\mathrm{PA}\lceil n)_{n\in\omega}$ is called the *base sequence* for this $\triangle$.

The *reflexivity property* of PA translates as saying that for all $n \in \omega$ PA proves that $\mathrm{PA}\lceil n$ is consistent. This was first established by Mostowski [13] and is crucial for practically all applications of $\triangle$.

The first use of $\triangle$ was to illustrate the relevance of the Hilbert-Bernays derivability conditions to Gödel's Second Incompleteness Theorem. The close connection of $\triangle$ to relative interpretability became apparent in Feferman [2], Orey

Received June 30, 1993; revised February 17, 1994

[15], and Hájek [5]. The celebrated fixed point of Theorem B of Lindström [8] and Lemma 4.5 of Švejdar [21] also makes implicit use of $\triangle$.

In Montagna [11], Visser [22], and Smoryński [19], $\triangle$ is treated as an object rather than just as a tool of study. Such also is the approach of the present paper. Motivation is discussed in detail in the second of the three aforementioned papers. This shift of $\triangle$'s status necessitates a closer look at its definition. For most applications the exact content of the theories $\mathrm{PA}\lceil n$ is fairly unimportant (see Orey [15], Lindström [8], Švejdar [21], Montagna [12], or Berarducci [1]) and one is therefore usually contented with the traditional choice

$\mathrm{PA}\lceil n$ = the theory axiomatized by the axioms of PA of gödelnumber $\leq n$.

The relation of $\mathrm{PA}\lceil n$ to $\mathrm{PA}\lceil n+1$ becomes then dependent on tiny intimate details of gödelnumbering of sentences and thus the only feasibly available properties of $(\mathrm{PA}\lceil n)_{n\in\omega}$ turn out to be the ones that we have already mentioned.

Smoryński [19] provides an example of a property of $\triangle$ whose proof and, as we shall see in Section 6, whose validity is dependent on the exact choice of $(\mathrm{PA}\lceil n)_{n\in\omega}$. He also shows that a more specific choice

$$\mathrm{PA}\lceil n = \mathrm{I}\Sigma_n$$

(see Paris and Kirby [16], Sieg [17], or Chapter 10 of Kaye [6]) can make questions about $\triangle$ much more malleable thus providing a better controllable $\triangle$. The key property of the theories $\mathrm{I}\Sigma_n$ is

Proposition 1.1 ($\mathrm{I}\Sigma_1$) *For all $n \in \omega$ the theory $\mathrm{I}\Sigma_{n+1}$ proves uniform Π_{n+2}-reflection for $\mathrm{I}\Sigma_n$, that is*:

$\mathrm{I}\Sigma_{n+1} \vdash$ 'for every Π_{n+2}-sentence π, if $\mathrm{I}\Sigma_n \vdash \pi$ then π is true'.

While the fact itself is widely known (see Leivant [7] or Ono [14]), its formalizability in $\mathrm{I}\Sigma_1$ has, as far as I know, never been explicitly stated, but it is not difficult to trace down the proofs of Corollary 4.4 of Sieg [17] or Exercise 10.8 of Kaye [6]. As an immediate corollary we have:

Corollary 1.2 ($\mathrm{I}\Sigma_1$) *For all $m,n \in \omega$ PA proves uniform Π_m-reflection for $\mathrm{I}\Sigma_n$.*

The property of PA expressed by Corollary 1.2 is known under the name of *essential reflexivity.*

Although the $\triangle$ based on $(\mathrm{I}\Sigma_n)_{n\in\omega}$ is not, strictly speaking, a Feferman predicate for $\mathrm{I}\Sigma_0$ is, most likely, not finitely axiomatizable, this discrepancy need not deter us for, provably in PA, it is only the tail of the sequence $(\mathrm{PA}\lceil n)_{n\in\omega}$ that matters as far as $\triangle$ is concerned. Alternatively, one can replace $\mathrm{I}\Sigma_0$ by $\mathrm{I}\Sigma_0 + \exp$.

A number of other sequences of theories is known to enjoy properties similar to Proposition 1.1 but we shall not strive for more generality. In this paper we stick almost exclusively with the definition of $\triangle$ based on $(\mathrm{I}\Sigma_n)_{n\in\omega}$ and construct the joint provability logic of $\Box$ and $\triangle$. The peculiarity of $\triangle$ in provability-logical context is that the modal operator corresponding to this predicate asks for a Kripke semantics incorporating a non-transitive relation S between nodes

of a Kripke frame, for $\triangle\varphi \to \triangle\triangle\varphi$ is not generally valid. This is the situation encountered in Visser [22] where the author treats a provability predicate akin to $\triangle$. The property of Visser's provability predicate relied on to overcome the difficulties caused by the failure of transitivity is completeness, which means that in the absence of transitivity every node enjoys a unique S-successor.

Our circumstances are slightly different. We show that lack of transitivity can be effectively compensated by reflection (Proposition 1.1) and in fact reflection keeps this lack to a minimum by providing a new modal principle approaching transitivity.

In Section 2 we introduce, acquire some experience with derivations in, and relate to formalized provability the modal system **LF** whose Kripke semantics is dealt with in Section 3. Section 4 proves the arithmetical completeness theorem for **LF**. Finally, in Sections 5 and 6 we answer two earlier questions concerning $\triangle$.

The extended Introduction should not lead the reader to hope for particularly detailed proofs. While the author takes, in matters of exposition, full advantage of his privileged position on a giant's shoulders, the reader may occasionally need to refresh his/her knowledge of some background material, for which purpose (Solovay [20] or Smoryński [18, Part I]) and Visser [22] should be highly beneficial. Those are also the sources that the reader will whenever possible be referred to for an omitted (part of a) proof.

2 LF

Definition 2.1 The language of the system **LF** is the propositional language with two unary modal operators $\Box$ and $\triangle$. Formulae in this language will be referred to as ($\Box\triangle$-)*formulae*. Shorthand for $\neg\triangle\neg$ is ∇. The operator $\Box$ abides by the laws of the logic **L** (cf. Solovay [20, **L** = G], Smoryński [18, **L** = **PRL**], or Visser [22, **L** = (L1)–(L4)]) and here are the axiom schemas for $\triangle$:

(F1) $\triangle(\varphi \to \psi) \to (\triangle\varphi \to \triangle\psi)$

(F2) $\Box\varphi \to \Box\triangle\varphi$

(F3) $\Box\varphi \to \triangle\Box\varphi$

(F4) $\Box\varphi \leftrightarrow (\triangle\varphi \vee \Box\bot)$

(F5) $\nabla\top$

(S) $\triangle\varphi \to \triangle\big((\triangle\psi \to \psi) \vee \triangle\varphi\big)$.

The only new rule is the $\triangle$-necessitation rule $\varphi/\triangle\varphi$.

The nearest relative of **LF** in the literature is the system BMF of Visser [22]. The axiom schema (F1) together with the $\triangle$-necessitation rule yield the *substitution property* for **LF**: the results of substituting two **LF**-equivalent $\Box\triangle$-formulas for the same propositional variable in another $\Box\triangle$-formula are **LF**-equivalent. We shall use this property throughout this and the next Section without special notice.

Note that (S) is a weakened analogue of the transitivity schema $\Box\varphi \to \Box\Box\varphi$. In **LF** the operator $\triangle$ also enjoys a weak analogue of Löb's schema $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ which, for a change, follows now from weak transitivity (in fact, the weakened versions are interderivable):

Proposition 2.2 *For $\{\psi_i\}_{i\in I}$ a finite collection of $\Box\triangle$-formulas and φ an arbitrary $\Box\triangle$-formula one has*

$$\mathbf{LF} \vdash \triangle\left(\left(\Box\bot \wedge \bigwedge_{i\in I}(\triangle\psi_i \to \psi_i)\right) \vee \triangle\varphi \to .\ \varphi\right) \to \triangle\varphi.$$

Proof: First we prove $\mathbf{LF} \vdash \triangle(\Box\bot \vee \triangle\varphi \to .\ \varphi) \to \triangle\varphi$.

(1) $\mathbf{LF} \vdash \triangle(\Box\bot \vee \triangle\varphi \to .\ \varphi) \to \triangle(\triangle\varphi \to \varphi)$ (by $\triangle$-necessitation and (F1))

(2) $\mathbf{LF} \vdash \triangle(\Box\bot \vee \triangle\varphi \to .\ \varphi) \to \triangle(\Box\bot \to \varphi)$

(3) $\mathbf{LF} \vdash \triangle(\neg\Box\bot \wedge \Box\varphi \to .\ \triangle\varphi)$ (by (F4))

(4) $\mathbf{LF} \vdash \triangle(\Box\bot \vee \triangle\varphi \to .\ \varphi) \to \triangle(\neg\Box\bot \wedge \Box\varphi \to .\ \varphi)$ (by (1) and (3))

(5) $\to \triangle(\Box\varphi \to \varphi)$ (by (2) and (4))

$\to \Box(\Box\varphi \to \varphi)$ (by (F4))

$\to \Box\varphi$ (by Löb's axiom)

$\to \triangle\Box\varphi$ (by (F3))

$\to \triangle\varphi$ (by (5)).

Second, one shows $\mathbf{LF} \vdash \triangle((\triangle\psi \to \psi) \vee \triangle\varphi \to .\ \varphi) \to \triangle\varphi$ for any formula ψ.

(6) $\mathbf{LF} \vdash \triangle((\triangle\psi \to \psi) \vee \triangle\psi)$ (by propositional logic and $\triangle$-necessitation)

(7) $\mathbf{LF} \vdash \triangle((\triangle\psi \to \psi) \vee \triangle\varphi \to .\ \varphi) \to \triangle(\triangle\varphi \to \varphi)$

(8) $\mathbf{LF} \vdash \triangle((\triangle\psi \to \psi) \vee \triangle\varphi \to .\ \varphi) \to \triangle((\triangle\psi \to \psi) \to \varphi)$

$\to \triangle((\triangle\psi \to \psi) \vee \triangle((\triangle\psi \to \psi) \to \varphi))$ (by (S))

$\to \triangle((\triangle\psi \to \psi) \vee \triangle(\psi \wedge ((\triangle\psi \to \psi) \to \varphi)))$ (by (6))

$\to \triangle((\triangle\psi \to \psi) \vee \triangle\varphi)$

$\to \triangle(\varphi \vee \triangle\varphi)$ (by (8))

$\to \triangle\varphi$ (by (7)).

Finally, suppose that $\Box\triangle$-formulas χ and θ are s.t. for any formula τ one has $\mathbf{LF} \vdash \triangle(\chi \vee \triangle\tau \to .\ \tau) \to \triangle\tau$ and $\mathbf{LF} \vdash \triangle(\theta \vee \triangle\tau \to .\ \tau) \to \triangle\tau$. The Proposition will clearly be subject to a straightforward inductive proof once we establish that $\mathbf{LF} \vdash \triangle((\chi \wedge \theta) \vee \triangle\varphi \to .\ \varphi) \to \triangle\varphi$ for φ arbitrary. Here is the derivation:

(9) $\mathbf{LF} \vdash \triangle(\theta \vee \triangle\varphi \to .\ \varphi) \to \triangle\varphi$ (by IH)

(10) $\mathbf{LF} \vdash \triangle(\chi \vee \triangle(\theta \vee \triangle\varphi \rightarrow . \varphi) \rightarrow . (\theta \vee \triangle\varphi \rightarrow . \varphi)) \rightarrow \triangle(\theta \vee \triangle\varphi \rightarrow . \varphi)$ (by IH)

$\mathbf{LF} \vdash \triangle((\chi \wedge \theta) \vee \triangle\varphi \rightarrow . \varphi) \rightarrow \triangle((\chi \vee \triangle\varphi) \wedge (\theta \vee \triangle\varphi) \rightarrow . \varphi)$ (by propositional logic)

$\rightarrow \triangle(\chi \vee \triangle\varphi \rightarrow . (\theta \vee \triangle\varphi \rightarrow . \varphi))$

$\rightarrow \triangle(\chi \vee \triangle(\theta \vee \triangle\varphi \rightarrow . \varphi) \rightarrow . (\theta \vee \triangle\varphi \rightarrow . \varphi))$ (by (9))

$\rightarrow \triangle(\theta \vee \triangle\varphi \rightarrow . \varphi)$ (by (10))

$\rightarrow \triangle\varphi$ (by (9)).

The following Proposition will enable a slight shortcut in the Kripke model developments of the next Section.

Proposition 2.3 *For $\{\psi_i\}_{i \in I}$ a finite collection of $\Box\triangle$-formulas one has*

$$\mathbf{LF} \vdash \nabla\left(\Box\bot \wedge \bigwedge\!\!\!\bigwedge_{i \in I} (\triangle\psi_i \rightarrow \psi_i)\right).$$

Proof: We shall prove that the negation of this formula implies $\bot$ in **LF**, that is $\mathbf{LF} \vdash \triangle\neg(\Box\bot \wedge \bigwedge\!\!\!\bigwedge_{i \in I}(\triangle\psi_i \rightarrow \psi_i)) \rightarrow \bot$:

$$\mathbf{LF} \vdash \triangle\neg\left(\Box\bot \wedge \bigwedge\!\!\!\bigwedge_{i \in I} (\triangle\psi_i \rightarrow \psi_i)\right) \rightarrow \triangle\left(\left(\Box\bot \wedge \bigwedge\!\!\!\bigwedge_{i \in I} (\triangle\psi_i \rightarrow \psi_i)\right) \vee \triangle\bot \rightarrow . \bot\right)$$ (by (F5))

$\rightarrow \triangle\bot$ (by Proposition 1.2)

$\rightarrow \bot$ (by (F5)).

Next we formally define the provability interpretation of **LF**.

Definition 2.4 A function $^\circ$ assigning arithmetic sentences to $\Box\triangle$-formulas is a *gf-interpretation* if

(i) $^\circ$ distributes over propositional connectives,

(ii) $(\Box\varphi)^\circ$ is the formalization of "φ° is provable in PA" and

(iii) $(\triangle\varphi)^\circ$ is 'there exists an x s.t. $\mathrm{PA}\lceil x$ is consistent and $\mathrm{PA}\lceil x \vdash \varphi^\circ$' (recall that we have agreed that $\mathrm{PA}\lceil x = \mathrm{I}\Sigma_x$ unless otherwise specified).

Proposition 2.5 *For any gf-interpretation $^\circ$, $\mathbf{LF} \vdash \varphi$ implies $\mathrm{PA} \vdash \varphi^\circ$.*

Proof: The correctness of all elements of **LF** with respect to gf-interpretations for arbitrary Feferman predicates, except for the schema (S), is verified in Montagna [11] and Visser [22]. (S) is the only axiom which depends on our convention $\mathrm{PA}\lceil n = \mathrm{I}\Sigma_n$.

We therefore only check (S): $\triangle\varphi \rightarrow \triangle((\triangle\psi \rightarrow \psi) \vee \triangle\varphi)$. Suppose $\psi^\circ \in \Pi_n$. Reason in PA:
Assume $(\triangle\varphi)^\circ$.

Case 1. PA *is consistent.* In this case $(\triangle\varphi)^\circ$ is a synonym for $(\Box\varphi)^\circ$ and we have $(\Box\triangle\varphi)^\circ$ by (F2)° whence $(\Box((\triangle\psi\to\psi)\vee\triangle\varphi))^\circ$ trivially follows. Since PA is consistent, this is the same as $(\triangle((\triangle\psi\to\psi)\vee\triangle\varphi))^\circ$ as required.

Case 2. PA *is inconsistent.* Let μ be the maximal x s.t. $\mathrm{I}\Sigma_x$ is consistent. (Note that by Corollary 1.2, $\mu+1\geq n$.) We then have $\mathrm{I}\Sigma_\mu\vdash\varphi^\circ$ and we shall prove $((\triangle\psi\to\psi)\vee\triangle\varphi)^\circ$ inside $\mathrm{I}\Sigma_\mu$:

Case 2.1. $\mathrm{I}\Sigma_{\bar\mu}$ *is consistent.* This one is easy: $\mathrm{I}\Sigma_{\bar\mu}\vdash\varphi^\circ$ and therefore $(\triangle\varphi)^\circ$.

Case 2.2. $\mathrm{I}\Sigma_{\bar\mu}$ *is inconsistent.* By Proposition 1.1 the theory $\mathrm{I}\Sigma_{\bar\mu-1}$ is consistent and hence for any arithmetical sentence γ one has $\triangle\gamma$ iff $\mathrm{I}\Sigma_{\bar\mu-1}\vdash\gamma$. But then we have $(\triangle\psi\to\psi)^\circ$ by Proposition 1.1, for ψ° is a Π_n- and hence a $\Pi_{\bar\mu+1}$-sentence and provability in $\mathrm{I}\Sigma_{\bar\mu-1}$ testifies to the truth of such sentences.

Thus in either case $((\triangle\psi\to\psi)\vee\triangle\varphi)^\circ$.

Thus in either case $(\triangle((\triangle\psi\to\psi)\vee\triangle\varphi))^\circ$.

To close the Section we present yet another derivation in **LF**. The topic comes from Smoryński [19] where the author shows PA-provable uniqueness of $\triangle$-*gödelsentences*, i.e., fixed points of the form $\gamma\leftrightarrow\neg\triangle\gamma$ for the Feferman predicate based on $(\mathrm{I}\Sigma_n)_{n\in\omega}$. Relying on Proposition 2.5 we are now able to give a purely modal proof of a formalization of this result. Our proof appears to be different from that of Smoryński.

Proposition 2.6

$$\mathbf{LF}\vdash\Box(p\leftrightarrow\neg\triangle p)\wedge\Box(q\leftrightarrow\neg\triangle q)\to.\ \Box(p\leftrightarrow q).$$

Proof: We shall only prove one half of the implication, namely we simplify the succedent to $\Box(p\to q)$.

(1) $\mathbf{LF}\vdash\Box(\neg q\to\triangle q)\to\Box(\neg q\to\triangle((\triangle p\to p)\vee\triangle q))$ (by (S))

$\mathbf{LF}\vdash\Box(\neg q\leftrightarrow\triangle q)\to\Box(\neg q\to\triangle\neg\triangle q)$ (by (F2))

(2) $\to\triangle(\triangle p\to p))$ (by (1))

(3) $\mathbf{LF}\vdash\Box(\neg p\to\triangle p)\to\Box\triangle(\neg\triangle p\to p)$ (by (F2))

$\mathbf{LF}\vdash\Box(\neg p\leftrightarrow\triangle p)\wedge\Box(\neg q\leftrightarrow\triangle q)$

$\to.\ \Box(\neg q\to.\ \triangle(\triangle p\to p)\wedge\triangle(\neg\triangle p\to p))$ (by (2) and (3))

$\to.\ \Box(\neg q\to\triangle p)$

$\to.\ \Box(\neg q\to\neg p)$

$\to.\ \Box(p\to q).$

Exercise 2.7 Show that all $\triangle$-gödelsentences are provably in PA equivalent to the sentence

$$\forall x(\mathrm{I}\Sigma_x \vdash \text{`}\mathrm{I}\Sigma_x \vdash \bot\text{'} \to \mathrm{I}\Sigma_x \vdash \bot).$$

We shall return to the subject of gödelsentences in Section 6.

3 Models for LF

Definition 3.1 A *marshmallow frame* is a tuple $\mathbb{W} = (W, R, S)$ where W is a nonempty finite set, and R and S are binary relations on W s.t.

(i) R is transitive and irreflexive,

(ii) $R \subseteq S$,

(iii) $R \circ S \subseteq R$,

(iv) $S \circ R \subseteq R$,

(v) If xSy and xRz then xRy,

(vi) If $xSySz$ then xSz or ySy,

(vii) For each x there is an y s.t. xSy.

We refer to property (vi) as *skew-transitivity*. For $\mathbb{W}$ a marshmallow frame we define

$$T_W = \{a \in W \mid aRb \text{ for no } b \in W\}$$
$$D_W = \{a \in W \mid aSa\}.$$

and agree to always omit the subscripts. In this notation we have

Lemma 3.2 (a) *If $x \in W - T$ then (xRy iff xSy).*

(b) *If xSy and $x \in T$ then $y \in T$ as well.*

(c) *$D \subseteq T$.*

(d) *For all $x \in W$ there is a $y \in W$ s.t. $xSy \in D$.*

Proof: (a) follows from (ii) and (v) of Definition 3.1.

(b) Immediate from (iv) of Definition 3.1.

(c) follows from (a) and (i) of Definition 3.1.

(d) Consider an $x \in W$. Suppose that $y \in D$ held for no $y \in W$ s.t. xSy. Then by iterating (vii) of Definition 3.1 we could get an infinite sequence $x = x_0 S x_1 S x_2 \ldots$ By induction on j one shows using skew-transitivity that $xSx_i Sx_j$ for all $i < j$. On the other hand, W is finite, so we have $xSx_i = x_j$ for some $i < j$ which is a contradiction for one then has $xSx_i \in D$.

Definition 3.3 A *marshmallow model* is a pair $\mathfrak{W} = (\mathbb{W}, \Vdash)$, where $\mathbb{W} = (W, R, S)$ is a marshmallow frame and $\Vdash$ is a *forcing relation* (cf. Visser [22]) between elements of W and $\Box\triangle$-formulae, R and S being the accessibility relations for $\Box$ and $\triangle$, respectively. One writes $\mathfrak{W} \Vdash \varphi$ if $a \Vdash \varphi$ for all $a \in W$.

Proposition 3.4 *For any* $\Box\triangle$*-formula* φ *and any marshmallow model* $\mathfrak{W}$, *if* **LF** $\vdash \varphi$ *then* $\mathfrak{W} \Vdash \varphi$.

Our aim is to reverse the implication of Proposition 3.4. In doing so we shall follow faithfully closely the presentation of Visser [22,5.3] where lolly-models, of which our marshmallow models are strongly reminiscent, are handled.

Definition 3.5 A finite set α of $\Box\triangle$-formulas is *adequate* if

(i) $\Box\bot \in \alpha$,

(ii) α is closed under subformulas,

(iii) If $\varphi \in \alpha$ is not of the form $\neg\psi$ then $\neg\varphi \in \alpha$,

(iv) $\Box\varphi \in \alpha$ iff $\triangle\varphi \in \alpha$,

(v) If $\triangle\varphi, \triangle\psi \in \alpha$ then $(\triangle\psi \to \psi) \vee \triangle\varphi \in \alpha$.

Clearly, any finite set of $\Box\triangle$-formulas is a subset of an adequate set.

Definition 3.6 Let α be adequate and define W_α to be the set of all sets w of $\Box\triangle$-formulas satisfying

(i) If $\varphi, \neg\varphi \in \alpha$ then $\varphi \in w$ or $\neg\varphi \in w$,

(ii) If $\varphi \in w - \alpha$ then φ is of the form $\triangle\psi$ and both ψ and $\triangle\triangle\psi$ are in w,

(iii) **LF** does not refute the conjunction of any finite subset of w.

The set W_α of Definition 3.6 is finite, for if $w \in W_\alpha$ then, due to clause (ii), formulas from w that are outside α come in chunks, each one of which can be traced down to a different formula in α. Moreover,

Lemma 3.7 *Each set of* $\Box\triangle$*-formulas satisfying* (ii) *and* (iii) *of Definition* 3.6 *has a superset in* W_α.

Proof: Starting with such a set, keep on adding appropriate elements of α until (i) is also satisfied.

Definition 3.8 Let α be adequate. We define relations R_α and S_α on W_α. Put

$vR_\alpha w$ iff (for any formula φ, if $\Box\varphi \in v$ then $\varphi, \triangle\varphi, \triangle\triangle\varphi, \ldots \in w$) and there exists a formula $\Box\psi \in w - v$.

Recall that in marshmallow frames the set T of R-topmost elements is defined exclusively in terms of the relation R (cf. Definition 3.1). Even though we do not yet know whether we are dealing with a marshmallow frame, we shall still have that definition of T in mind. Let

$vS_\alpha w$ iff $vR_\alpha w$ or

$v, w \in T$, $\varphi \in w$ whenever $\triangle\varphi \in v$, and for any ψ, χ, if $\triangle\psi \in v \cap \alpha$ and $\triangle\chi \in w \cap \alpha$ then $\triangle\psi \in w$ or $\chi \in w$.

Lemma 3.9 *For an adequate set* α *the frame* $\boldsymbol{W}_\alpha = (W_\alpha, R_\alpha, S_\alpha)$ *is a marshmallow frame.*

Proof: We shall only prove clauses (vi) and (vii) of Definition 3.1 of marshmallow frames, referring the reader to Visser [22] for the rest (or rather the beginning) of the proof as well as for the fact that

$$\text{for any } w \in W_\alpha, \qquad w \in T \quad \text{iff} \quad \Box\bot \in w.$$

Clause (vi) comes first. Suppose $uS_\alpha vS_\alpha w$ and not $vS_\alpha v$. We stipulate that all the three elements u, v, and w are in T for otherwise the claim can be derived from the properties of marshmallow frames considered to be already established for $\boldsymbol{W}_\alpha$. The failure of $vS_\alpha v$ means then that we can fix a formula $\triangle\psi \in v \cap \alpha$ s.t. $\psi \notin v$. Let us check $uS_\alpha w$. Suppose $\triangle\varphi \in u$ so that by $uS_\alpha v$ one has $\varphi \in v$. If $\triangle\varphi \in \alpha$ then recall that we also have $\triangle\psi \in v \cap \alpha$ so $\psi \in v$ or $\triangle\varphi \in v$ must hold. The first is not the case, therefore $\triangle\varphi \in v$. If $\triangle\varphi \notin \alpha$ then $\triangle\triangle\varphi \in u$ and hence $\triangle\varphi \in v$ all the same. Since $vS_\alpha w$, there holds $\varphi \in w$ and if, in addition, $\triangle\varphi \in \alpha$ and $\triangle\chi \in w \cap \alpha$ then $\triangle\varphi \in w$ or $\chi \in w$.

We turn now to clause (vii). Let $v \in W_\alpha$. We look for a $w \in W_\alpha$ with $vS_\alpha w$. We only consider the case $v \in T$ so that $\Box\bot \in v$ for if $v \notin T$ then there is a $z \in W_\alpha$ with $vR_\alpha z$ implying $vS_\alpha z$. Consider the set

$$w_0 = \{\Box\bot\} \cup \{\varphi \mid \triangle\varphi \in v\} \cup \{\triangle\psi \to \psi \mid \triangle\psi \in \alpha\}.$$

No conjunction of any finite subset of w_0 is refuted in **LF** for otherwise for a certain finite $d \subseteq \{\varphi \mid \triangle\varphi \in v\}$ one has

$$\mathbf{LF} \vdash \Box\bot \wedge \bigwedge_{\triangle\varphi\in d} \varphi \to .\ \neg \bigwedge_{\triangle\psi\in\alpha} (\triangle\psi \to \psi)$$

$$\vdash \Box\bot \wedge \bigwedge_{\triangle\varphi\in d} \triangle\varphi \to .\ \triangle\neg \bigwedge_{\triangle\psi\in\alpha} (\triangle\psi \to \psi) \qquad \text{(by } \triangle\text{-necessitation and (F3))}$$

$$\vdash \Box\bot \wedge \bigwedge_{\triangle\varphi\in d} \triangle\varphi \to \bot \qquad \text{(by Proposition 2.3)}$$

which would deny v membership in W_α. Furthermore, we show that w_0 satisfies (ii) of Definition 3.6: Suppose $\triangle\varphi \in w_0 - \alpha$. Then $\triangle\triangle\varphi \in v - \alpha$, hence $\triangle\varphi, \triangle\triangle\triangle\varphi \in v$, hence $\varphi, \triangle\triangle\varphi \in w_0$. There is therefore by Lemma 3.7 a set w with $w_0 \subseteq w \in W_\alpha$. Moreover, since $\Box\bot \in w_0 \subseteq w$, one has $w \in T$. Finally, one has $\psi \in w$ whenever $\triangle\psi \in w \cap \alpha$ because $\{\triangle\psi \to \psi \mid \triangle\psi \in \alpha\} \subseteq w$. So $vS_\alpha w$.

Definition 3.10 The marshmallow model $\mathfrak{W}_\alpha = (\boldsymbol{W}_\alpha, \Vdash)$ is defined by putting

$$w \Vdash p \text{ iff } p \in w$$

for propositional variables $p \in \alpha$ and $w \in W_\alpha$.

Lemma 3.11 *For $\varphi \in \alpha$ and $w \in W_\alpha$ one has $w \Vdash \varphi$ iff $\varphi \in w$.*

Proof: The lemma is proved by induction on the structure of φ. See Visser [22] for the induction step in the $\Box$ case. We turn to $\triangle$ under the assumption $w \in T$.
(if): Suppose $\triangle\varphi \in w \cap \alpha$. Then for all v with $wS_\alpha v$ one has $\varphi \in v$, that is, by IH, $v \Vdash \varphi$ and hence $w \Vdash \triangle\varphi$.
(only if): Suppose $\triangle\varphi \in \alpha - w$. Consider the set

$$u_0 = \{\neg\varphi\} \cup \{\Box\bot\} \cup \{\psi \mid \triangle\psi \in w\} \cup \{(\triangle\chi \to \chi) \vee \triangle\theta \mid \triangle\chi \in \alpha, \triangle\theta \in w \cap \alpha\}.$$

We show that u_0 is consistent with **LF**. Suppose it is not. Then for some finite set $d \subseteq \{\psi \mid \Delta\psi \in w\}$ there holds

$$\mathbf{LF} \vdash \Box\bot \wedge \bigwedge\!\!\!\bigwedge_{\Delta\psi\in d} \psi \wedge \bigwedge\!\!\!\bigwedge_{\substack{\Delta\chi\in\alpha \\ \Delta\theta\in w\cap\alpha}} ((\Delta\chi \to \chi) \vee \Delta\theta) \to . \varphi$$

$$\vdash \Box\bot \wedge \bigwedge\!\!\!\bigwedge_{\Delta\psi\in d} \Delta\psi \wedge \bigwedge\!\!\!\bigwedge_{\substack{\Delta\chi\in\alpha \\ \Delta\theta\in w\cap\alpha}} \Delta((\Delta\chi \to \chi) \vee \Delta\theta) \to . \Delta\varphi$$

(by Δ-necessitation and (F3))

$$\vdash \Box\bot \wedge \bigwedge\!\!\!\bigwedge_{\Delta\psi\in d} \Delta\psi \wedge \bigwedge\!\!\!\bigwedge_{\Delta\theta\in w\cap\alpha} \Delta\theta \to . \Delta\varphi \qquad \text{(by (S))}.$$

Thus $\Delta\varphi$ could not escape being in w contrary to assumptions. Therefore u_0 is consistent and, as in the proof of Lemma 3.9, it is seen that u_0 satisfies (ii) of Definition 3.6. Hence there exists a u with $u_0 \subseteq u \in W_\alpha$. We leave it to the reader to check $wS_\alpha u$ so that $w \Vdash \Delta\varphi$ because, by IH, $u \nVdash \varphi$.

Theorem 3.12 **LF** $\vdash \varphi$ *iff for any marshmallow model* $\mathfrak{W}$ *one has* $\mathfrak{W} \Vdash \varphi$.

Proof: (only if) is Proposition 3.4.
(if): Suppose **LF** $\nvdash \varphi$. Let then α be an adequate set containing $\neg\varphi$. By Lemma 3.7 there exists a $w \in W_\alpha$ with $\neg\varphi \in w$. Hence by Lemma 3.11 we have $w \nVdash \varphi$ and so $\mathfrak{W}_\alpha \nVdash \varphi$ as required.

Exercise 3.13 (de Jongh) Consider the logic **F** on $\Box$-free $\Box\Delta$-formulas axiomatized by (F1), (F5), (S), and Δ-necessitation.

(a) Consider *marshmallows* (W, S), S being a binary relation on a finite non-empty W and satisfying (vi) and (vii) of Definition 3.1. Prove **F** to be complete w.r.t. marshmallows.

(b) Show that **LF** is conservative over **F**.

Hints: (a) Weed the proof of Theorem 3.12.
(b) Observe that any marshmallow can be represented as the set T of an appropriate marshmallow model.

4 A Solovay function In this Section we prove the arithmetical completeness theorem for **LF**. As usual, we shall do so by constructing a suitable Solovay function (cf. Solovay [20], Chapter 3 of Smoryński [18], or Visser [22]) climbing up a Kripke frame, namely one of the marshmallow frames constructed in the previous Section. We describe the construction of a Solovay function for an almost arbitrary marshmallow frame $\boldsymbol{W} = (W, R, S)$ which however will have to satisfy the following two conditions:

(i) $\boldsymbol{W}$ has a bottom node, that is there is a node $0 \in W$ s.t. $0Ra$ or $0 = a$ for all $a \in W$ and

(ii) There is a node $a \in W$ distinct from 0.

Note that by appending a new bottom node R- and S-below any given marshmallow model one obtains another one satisfying both (i) and (ii).

The construction of the Solovay function $F: \omega \to W$ proceeds parallel to that of an auxiliary function $G: \omega \to \omega \cup \{-1, \infty\}$ which stores some information relevant for F's locomotion. The two bear a close relationship to the pair appearing in Section 10 of Visser [23] that are, in turn, a variation on the Solovay function of Berarducci [1].

To make things work smoothly we have to require that the formalization of proofs in arithmetic is reasonable and uniform so that the following is known to $I\Sigma_1$:

(i) If x happens to be a proof in PA of a sentence φ then x is not a proof in PA of any sentence distinct from φ,

(ii) If x is a PA-proof of φ then x also is a proof of the same sentence in $I\Sigma_y$ for some y,

(iii) If x is a proof of φ in $I\Sigma_y$ then x is also a proof of φ in PA as well as in $I\Sigma_z$ for all $z \geq y$,

(iv) If φ is provable in $I\Sigma_y$ then there are arbitrarily large $I\Sigma_y$-proofs of φ.

Definition 4.1 ($I\Sigma_1$) Define primitive recursive functions F and G by simultaneous recursion and the Recursion Theorem:

$$F(0) = 0; \qquad G(0) = \infty.$$

The value of $F(x+1)$ and $G(x+1)$ is defined by cases.

Case A: $F(x)Ra$ and x is a PA-proof of $L \neq a$.

$$F(x+1) = a; \quad G(x+1) = \begin{cases} \text{any } y \text{ s.t. } x \text{ is an } I\Sigma_y\text{-proof of } L \neq a & \\ & \text{if } a \in T, \\ \infty & \text{otherwise.} \end{cases}$$

Case B: $F(x)Sb \notin D$, $F(x) \in T$ and x is an $I\Sigma_{G(x)}$-proof of $L \neq b$.

$$F(x+1) = b; \qquad G(x+1) = G(x).$$

Case C: $F(x)Sc \in D$, $F(x) \in T$ and x is an $I\Sigma_{G(x)}$-proof of $L \neq c$.

$$F(x+1) = c; \qquad G(x+1) = G(x) - 1.$$

Case D: None of Cases A–C is the case.

$$F(x+1) = F(x); \qquad G(x+1) = G(x).$$

Finally, $L \neq d$ is the formula expressing that d is not the limit value of F.

By "any y" in Case A we mean, of course, any y chosen in a primitive-recursive way. In Cases B and C, we stipulate that no x is an $I\Sigma_{-1}$-proof of anything.

Lemma 4.2 (The Limit Lemma) ($I\Sigma_1$) *Both F and G are eventually constant.*

Proof: It is easily established by induction on the argument that the function G is monotonously decreasing and hence reaches a limit value.

Were F to stay forever nomadic, it would sooner or later reach T wherefrom, by Lemma 3.2(b) it cannot go back. Moreover, one can, just as in Lemma 3.2(d),

prove that it would have to infinitely often come to nodes in D, each time decreasing, according to Case C, the value of G by one. This clearly contradicts the fact that G has a limit. Thus F reaches a limit as well.

Definition 4.3 ($I\Sigma_1$) The Limit Lemma allows one to define the following two ε-terms (that is, these are definitions of names rather than of values):

$$L = \lim_{x\to\infty} F(x); \qquad \mu = \lim_{x\to\infty} G(x).$$

Let us further agree for the remainder of the paper that $I\Sigma_\infty = PA$.

Lemma 4.4 ($I\Sigma_1$) *(a) If $L = a \neq 0$ then $\Box \mathbb{W}_{aRb} L = b$.*
(b) *If $L = a$ then for no b with aRb does one have $\Box L \neq b$.*

Proof: (a) In the case that $a \notin T$ the old proof (cf. Solovay [20] or Smoryński [18, Chapter 3]) goes through for F can only leave a by an R-arrow whereafter it is only allowed to go along R- and S-arrows. Fortunately, we have both $R \circ R \subseteq R$ and $R \circ S \subseteq R$.

Suppose now $a \in T$ so that we have to show that PA is inconsistent. Let x be such that $G(x) < \infty$ and $I\Sigma_{G(x)} \vdash L \neq a$. Reason in PA:

> By the Limit Lemma, there is a b s.t. $L = b$. Moreover, after F had reached T, the function G could only decrease and therefore the proof of $L \neq b$ that brought F to b is a proof in $I\Sigma_y$ with $y \leq G(x)$ and hence in $I\Sigma_{G(x)}$. But if $L \neq b$ is proved in $I\Sigma_{G(x)}$ then, since by Corollary 1.2 we have reflection for $I\Sigma_{G(x)}$, L should be actually unequal to b.

Thus PA is indeed inconsistent.

(b) is proved exactly as before (cf. Solovay [20] or Smoryński [18]).

Lemma 4.5 ($I\Sigma_1$) *μ is the largest x s.t. $I\Sigma_x$ is consistent. Therefore for any arithmetical sentence γ one has $\triangle\gamma$ iff $I\Sigma_\mu \vdash \gamma$.*

Proof: If $L \notin T$ then by Lemma 4.4 the whole of PA ($= I\Sigma_\infty = I\Sigma_\mu$) is consistent. Assume $L \in T$ which implies $\mu < \infty$. We have to show that $I\Sigma_\mu$ is consistent whereas $I\Sigma_{\mu+1}$ is not.

Suppose $I\Sigma_\mu$ were not consistent. By clause (vii) of Definition 3.1 there is a b s.t. LSb. We then have $I\Sigma_\mu \vdash L \neq b$. If $b \notin D$ then $b \neq L$ and by Case B the function F would have to walk away from its limit every time an $I\Sigma_\mu$-proof of $L \neq b$ occurs. If $b \in D$ then on encountering an $I\Sigma_\mu$-proof of $L \neq b$ the function G would, by Case C, have to decrease below its limit. Thus $I\Sigma_\mu$ has to be consistent and, moreover, consistent with $L = b$.

To show that $I\Sigma_{\mu+1}$ is inconsistent reason in $I\Sigma_{\mu+1}$:

> By the Limit Lemma, there is an a s.t. $L = a$. Since the function G can only decrease, the proof of $L \neq a$ that brings F to a is an $I\Sigma_{\tilde{\mu}}$-proof. But then, since $L \neq a$ is a Π_2-sentence, L cannot, by Proposition 1.1, be equal to a. A contradiction.

$I\Sigma_{\mu+1}$ is therefore inconsistent.

Lemma 4.6 ($I\Sigma_1$) *Suppose* $L = a \in T$.

(a) $I\Sigma_\mu \vdash \mathbb{W}_{aSb} L = b$.

(b) *If* aSb *then* $I\Sigma_\mu$ *is consistent with* $L = b$.

Proof: (a) Suppose $L = a \in T$. Consider a node c s.t. $a \neq c$ and c does not satisfy aSc. Reason in $I\Sigma_\mu$:

> To get to c the function F has to leave a and after that, by skew-transitivity, arrive at and leave a node in D. This by Case C of Definition 4.1 involves a decrease in the value of G and therefore the proof that brings F to c is an $I\Sigma_{\tilde{\mu}-1}$-proof of $L \neq c$. Reflecting on that we get $L \neq c$.

Furthermore, in the case that $a \notin D$, we must have had an $I\Sigma_\mu$-proof of $L \neq a$ to get to a in the first place.

Thus for every node c that is not S-accessible from a we have $I\Sigma_\mu \vdash L \neq c$. The claim follows.

(b) See the first part of the proof of Lemma 4.5.

Lemma 4.7 ($I\Sigma_1$) (*a*) *If* $L = a \neq 0$ *then* $\triangle \mathbb{W}_{aSb} L = b$.

(b) *If* $L = aSb$ *then* $\nabla L = b$.

Proof: For $L \notin T$ this reduces to Lemma 4.4. For $L \in T$ use Lemmas 4.5–6.

Lemma 4.8 $L = 0$ *is true and for any* $a \in W$, *the sentence* $L = a$ *is consistent with* PA.

Proof: Consult Solovay [20] or Smoryński [18] or use Lemma 4.4.

Theorem 4.9 **LF** $\vdash \varphi$ *iff for any gf-interpretation* °, PA $\vdash \varphi^\circ$.

Proof: We have already proved (only if) in Proposition 2.5.
(if) This is proved in the standard manner (cf. Solovay [20] or Smoryński [18]): Suppose **LF** $\nvdash \varphi$. Then by Theorem 3.12 we have a marshmallow model $\mathfrak{W} = (W, R, S, \Vdash)$ with $\mathfrak{W} \nVdash \varphi$. Append a new bottom 0 to $\mathfrak{W}$ and apply Definition 4.1 to the result. Define a gf-interpretation ° by putting

$$p^\circ = \bigvee_{W \ni a \Vdash p} L = a$$

for propositional variables p occurring in φ. Use Lemmas 4.4 and 4.7 to show by induction on the structure of φ that

$$\mathrm{Pa} \vdash L \neq 0 \to (\varphi^\circ \leftrightarrow L \Vdash \varphi)$$

so that if $W \ni a \nVdash \varphi$ then PA $\vdash \varphi^\circ$ would imply PA $\vdash L \neq a$ contradicting Lemma 4.8. Conclude PA $\nvdash \varphi^\circ$.

Theorem 4.9 is very useful for constructing gf-interpretations under which a particular $\Box\triangle$-formula is not provable. As usual it is accompanied by a similar result which allows to construct gf-interpretations ° rendering a $\Box\triangle$-formula φ provable by ensuring that $(\Box\varphi)^\circ$ is true.

Definition 4.10 For φ a $\Box\triangle$-formula define the formula $R(\varphi)$ to be

$$\neg\Box\bot \wedge \bigwedge \{\Box\psi \to \psi \mid \Box\psi \text{ or } \triangle\psi \text{ is a subformula of } \varphi\}.$$

Proposition 4.11 *Suppose φ is a $\Box\triangle$-formula and the bottom 0 of a marshmallow model $\mathfrak{W}$ forces $R(\varphi)$. Then there exists a gf-interpretation $^\circ$ s.t. (φ° is true if and only if $0 \Vdash \varphi$).*

This Proposition enjoys a proof similar to that of Theorem 4.9 for assuming $R(\varphi)$ is forced at the bottom 0 of a marshmallow model we can drop the antecedent $L \neq 0$ in the key inductive step $\mathrm{PA} \vdash L \neq 0 \to (L \Vdash \Box\varphi \to \Box L \Vdash \varphi)$ of Theorem 4.9. (Consult Solovay [20] or Chapter 3 of Smoryński [18] for a similar situation.)

Remark 4.12 Theorem 4.9 and Proposition 4.11 can be adapted to the logic **F** on $\Box$-free formulas and marshmallows $\mathcal{W} = (W, S)$ of Exercise 3.13. In this case the requirements (i) and (ii) on the marshmallow frames handled in our construction can be replaced by the single condition

(iii) There is a node $0 \in W$ s.t. $0Sa$ for all $a \in W$.

Nodes 0 that satisfy this property are called *centers* and marshmallows of this kind are *centered*. Any marshmallow can clearly be made centered by adding a center without disturbing the forcing relation at the old nodes.

The construction of the Solovay function for a marshmallow $\mathfrak{W}$ with a center 0 can be visualized as follows: Make the marshmallow into a marshmallow frame by adjoining a new R-bottom 0_R below the whole of $\mathfrak{W}$. After that apply Definition 4.1 to the resulting frame. However, when proving the analogue of Lemma 4.7, 0 and 0_R should be treated as a single node, that is, one defines the sentence $L =$ "$0 + 0_R$" as $L = 0 \vee L = 0_R$, with the understanding that the $\triangle$-forcing relation at "$0 + 0_R$" coincides with that at 0.

The nice point about this construction is that here we get more satisfactory commutation in that one can prove $\mathrm{I}\Sigma_1 \vdash L =$ "$0 + 0_R$" $\to \nabla L =$ "$0 + 0_R$" and hence drop the precondition that 0 forces a certain formula in the analogue of Proposition 4.11. This amounts to an embedding of the (finite) $\triangle$-algebra corresponding to $\mathfrak{W}$ into that of PA.

The progress achieved so far casts doubt on the conclusions of a discussion in Chapter 4 of Smoryński [18] drawn from investigations into the structure of finite fixed point algebras. Smoryński argues that extensional arithmetical formulae subject to successful finite algebraic or relational modeling ought to bear a much closer similarity to $\Box$ than does the $\triangle$ of the present paper. The point here is that a finite $\triangle$-algebra never, except for the trivial case, generates a fixed point algebra the way diagonalizable (= $\Box$-) algebras do, but this does not prevent finite $\triangle$-algebras from being applicable to the study of the predicate $\triangle$.

Exercise 4.13 Consider logics $\mathbf{LF}^\omega$ and $\mathbf{F}^\omega$, where $\mathbf{LF}^\omega$ is obtained from **LF** by adding the schema $\Box\varphi \to \varphi$ and the logic $\mathbf{F}^\omega$ results from the logic **F** from Exercise 3.13 by adding the schema $\triangle\varphi \to .\ \varphi \wedge \triangle\triangle\varphi$ with the usual restriction that whatever is derived with the help of the new schemas is in both cases subject neither to $\Box$- nor to $\triangle$-necessitation.

(a) Prove that $\mathbf{LF}^\omega \vdash \mathbf{F}^\omega$ and that $\mathbf{LF}^\omega$ is conservative over $\mathbf{F}^\omega$.

(b) Prove that for any $\Box\triangle$-formula φ one has $\mathbf{LF}^\omega \vdash \varphi$ iff $\omega \vDash \varphi^\circ$ for any gf-interpretation $^\circ$ (and hence the same holds for $\mathbf{F}^\omega$ and $\Box$-free φ).

Hint: (a) Show that $\mathbf{F}^{\omega}$ is complete w.r.t. centers of centered marshmallows. For the "difficult" direction observe that if for a node a of a marshmallow $\mathcal{M}$ and a $\Box$-free formula φ one has

$$a \Vdash \bigwedge\!\!\!\bigwedge \left\{ \triangle\psi \to .\ \psi \wedge \bigwedge\!\!\!\bigwedge_{2 \le n \le N} \triangle^n \psi \mid \psi \text{ a subformula of } \varphi \right\}$$

with N safely larger than the largest number of nested occurrences of $\triangle$'s in φ, then one can, preserving the forcing at a of subformulas of φ, transform $\mathcal{M}$ into a marshmallow of which a is a center.

(b) Use Proposition 4.11.

The last exploit in this Section of the Solovay function F introduced in Definition 4.1 is to illustrate that the validity of the axiom (S) w.r.t. gf-interpretations might actually fail under a choice of $(\mathrm{PA}\lceil n)_{n\in\omega}$, the base sequence for $\triangle$, different from $(\mathrm{I}\Sigma_n)_{n\in\omega}$. Let con T denote the sentence expressing the consistency of a finitely axiomatized theory T and consider the following sequence:

$$\mathrm{PA}\lceil n = \begin{cases} \mathrm{I}\Sigma_{n/2} & \text{if } n \text{ is even,} \\ \mathrm{I}\Sigma_{(n-1)/2} + \mathrm{con}\,\mathrm{I}\Sigma_{(n-1)/2} & \text{otherwise.} \end{cases}$$

A similar, although shorter, sequence was considered by Counterexample 3.1 of Smoryński [19]. We write $\triangle^*$ for the Feferman predicate based on this sequence of theories. Note that since the base sequence just introduced is a refinement of the base sequence $(\mathrm{I}\Sigma_n)_{n\in\omega}$ corresponding to $\triangle$, we have $\mathrm{I}\Sigma_1 \vdash \triangle\gamma \to \triangle^*\gamma$ for all sentences γ.

We shall show that (S) and, indeed, its consequence, the modal schema $\nabla(\triangle\varphi \to \varphi)$, are not valid for $\triangle^*$. First we need to know more about the limits of the functions F and G of Definition 4.1.

Lemma 4.14 ($\mathrm{I}\Sigma_1$) *Suppose* $L \in T - D$.

(a) $\mathrm{I}\Sigma_\mu \vdash L \in D \leftrightarrow \neg\mathrm{con}\,\mathrm{I}\Sigma_{\bar\mu}$.

(b) *If there is a node* b *with* $LSb \notin D$ *then* $\mathrm{I}\Sigma_\mu$ *is consistent with* $\mathrm{con}\,\mathrm{I}\Sigma_{\bar\mu}$.

(c) $\mathrm{I}\Sigma_\mu + \mathrm{con}\,\mathrm{I}\Sigma_{\bar\mu} \vdash \bigvee_{aSb\notin D} L = b$.

Proof: (a) Let $L = a \in T - D$ and reason in $\mathrm{I}\Sigma_\mu$:

If $L = b \in D$ then G must have decreased its value when getting to b and therefore $\mu < \bar\mu$. By Lemma 4.5 the theory $\mathrm{I}\Sigma_{\bar\mu}$ is then inconsistent. Conversely, if $L \notin D$ then either Case C takes place while F travels from a to L in which case $\mathrm{I}\Sigma_{\bar\mu-1} \vdash L \neq \bar L$ which, by reflection, is absurd, or Case C does not occur during this period implying $\mu = \bar\mu$. By Lemma 4.5 it follows that the theory $\mathrm{I}\Sigma_{\bar\mu}$ is consistent.

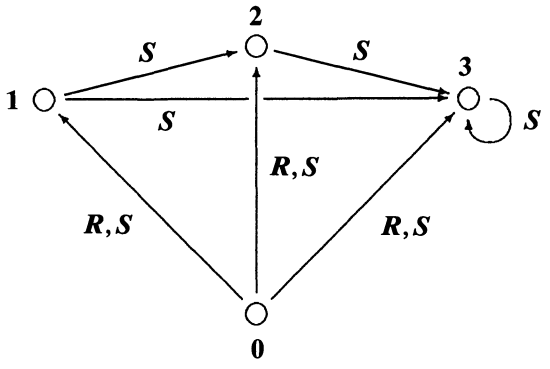
(b) By Lemma 4.6(b) $\mathrm{I}\Sigma_\mu$ is consistent with $L = b$. By (a) of the present Lemma it therefore has to be consistent with $\mathrm{con}\,\mathrm{I}\Sigma_{\bar\mu}$.

(c) Immediate from Lemma 4.6(a) and clause (a) of the present Lemma.

Proposition 4.15 *There exists an arithmetical sentence γ s.t.*

$$\mathrm{PA} \nvdash \nabla^*(\triangle^*\gamma \to \gamma).$$

Proof: Consider the functions F and G operating on the following marshmallow frame:



Observe:

$$\mathrm{I}\Sigma_1 \vdash L = 2 \to .\ L \neq 3 \wedge (\mathrm{I}\Sigma_\mu \vdash L = 3) \qquad \text{(by Lemma 4.6(a))}$$

$$\to .\ L \neq 3 \wedge \triangle^* L = 3 \qquad \text{(by Lemma 4.5)}$$

$$\mathrm{I}\Sigma_1 \vdash L = 1 \to .\ \mathrm{con}\,(\mathrm{I}\Sigma_\mu + \mathrm{con}\,\mathrm{I}\Sigma_{\bar{\mu}}) \wedge (\mathrm{I}\Sigma_\mu + \mathrm{con}\,\mathrm{I}\Sigma_{\bar{\mu}} \vdash L = 2)$$
(by Lemma 4.14(b) and (c))

$$\to \triangle^* L = 2.$$

This combines to give

$$\mathrm{I}\Sigma_1 \vdash L = 1 \to \triangle^*(L \neq 3 \wedge \triangle^* L = 3)$$

and therefore by Lemma 4.8

$$\mathrm{PA} \nvdash \nabla^*(\triangle^* L = 3 \to L = 3).$$

Thus putting γ to be $L = 3$ we are done.

5 Some fixed points Many qualitative and quantitative aspects of fixed points of arithmetical formulae corresponding to modal $\Box\triangle$-formulae under interpretations similar to our gf-interpretation are discussed in Visser [22]. In this Section we first observe that Theorem 4.11 affords a corollary classifying the quantitative behavior of fixed points of gf-interpretations of $\Box\triangle$-formulae. The argument is adapted from Application 6.11 of Visser [22].

Proposition 5.1 *Let $\varphi(p)$ be a $\Box\triangle$-formula where no propositional variable other than p occurs and s.t. every occurrence of that variable takes place within the scope of a modal operator. Then we have either*

(i) *The arithmetical sentence γ satisfying*

$$\mathrm{PA} \vdash \gamma \leftrightarrow \varphi^\circ(\gamma)$$

is PA*-provably unique, i.e.,*

$$\mathrm{PA} \vdash \Box(\gamma_1 \leftrightarrow \varphi^\circ(\gamma_1)) \wedge \Box(\gamma_2 \leftrightarrow \varphi^\circ(\gamma_2)) \rightarrow . \Box(\gamma_1 \leftrightarrow \gamma_2)$$

for all arithmetical sentences γ_1 *and* γ_2*, or*

(ii) *There exist infinitely many inequivalent fixed points of* $\varphi^\circ(x)$.

Proof (sketch): Suppose (i) is not the case. Then by Theorem 4.9 we have

$$\mathbf{LF} \nvdash \Box(p \leftrightarrow \varphi(p)) \wedge \Box(q \leftrightarrow \varphi(q)) \rightarrow . \Box(p \leftrightarrow q).$$

Therefore by Theorem 3.12 there exists a marshmallow model $\mathfrak{W} = (W, R, S, \Vdash)$ whose bottom 0 forces $\Box(p \leftrightarrow \varphi(p)) \wedge \Box(q \leftrightarrow \varphi(q))$ but not $\Box(p \leftrightarrow q)$. Note that we must then have a node $a \in T \subseteq W$ which does not force $p \leftrightarrow q$ for otherwise by induction on R-depth of elements of W it could be shown that $p \leftrightarrow q$ is forced everywhere. We isolate the subset T of W in the form of a marshmallow with a forcing relation: $\mathfrak{T} = (T, S\lceil T, \Vdash\lceil T)$ (here $\lceil$ stands for restriction) and construct a marshmallow model $\mathfrak{V}_N$ which consists of two copies of $\mathfrak{T}$ with N new linearly R-ordered nodes appended R-below those two copies. The exact value of $N \in \omega$ will be specified later. We define the forcing of a new propositional variable r in the two copies of $\mathfrak{T}$ so as to coincide with the forcing of p in the first copy and with that of q in the second. Further define forcing at the new nodes of the propositional variables p, q and r by R-downward induction in the unique way that makes $p \leftrightarrow \varphi(p)$, $q \leftrightarrow \varphi(q)$ and $r \leftrightarrow \varphi(r)$ forced everywhere in $\mathfrak{V}_N$. An easy application of the pigeon-hole principle shows that we can choose the value of N so that

$$R\big(\Box(p \rightarrow \varphi(p)) \wedge \Box(q \leftrightarrow \varphi(q)) \wedge \Box(r \leftrightarrow \varphi(r)) \rightarrow . \Box(p \leftrightarrow q) \vee \Box(p \leftrightarrow r) \vee \Box(q \leftrightarrow r)\big)$$

is forced at the bottom of $\mathfrak{V}_N$. We then apply Proposition 4.11 to obtain a gf-interpretation $^\circ$ under which the formulas

$$\Box(p \leftrightarrow \varphi(p)),\ \Box(q \leftrightarrow \varphi(q)),\ \text{and}\ \Box(r \leftrightarrow \varphi(r))$$

are true and the formulas

$$\Box(p \leftrightarrow q),\ \Box(p \leftrightarrow r),\ \text{and}\ \Box(q \leftrightarrow r)$$

are false so that we have obtained three fixed points of $\varphi^\circ(x)$ that are not provably equivalent in PA.

The reader will easily see how to generalize this proof to obtain arbitrarily finitely many PA-inequivalent fixed points of $\varphi^\circ(x)$.

Next we address fixed points of a particular $\Box\triangle$-formula, namely sentences γ satisfying

$$\mathrm{PA} \vdash \gamma \leftrightarrow \triangle\gamma.$$

These are called $\triangle$*-henkinsentences.* The two such fixed points that surrender most promptly to an eager quest are $\bot$ and $\top$ which shows that the modal formula $\triangle p$ falls into the second category of Proposition 5.1. Note that both sen-

tences $\bot$ and $\top$ are Σ_1. In fact, any $\triangle$-henkinsentence produced by a direct application of our Proposition 4.11 will be Σ_1. Visser [22] asks whether this is characteristic of all $\triangle$-henkinsentences.

We shall show that this is not the case, namely we exhibit a $\triangle$-henkinsentence which is not provably equivalent to any Σ_1 sentence. The $\triangle$-henkinsentence that we construct is actually an *oreysentence*, that is a sentence γ s.t. the theories $\mathrm{PA} + \gamma$ and $\mathrm{PA} + \neg\gamma$ are relatively interpretable in one another, or, equivalently, both are interpretable in PA. This may be viewed as somewhat unexpected since oreysentences originally used to be constructed as fixed points of the formula $\neg\triangle p$ rather than $\triangle p$ (see Švejdar [21] and Lindström [8]). To deduce the non-Σ_1-ness of an oreysentence we lean on the following

Proposition 5.2 *For arithmetical sentences γ and δ the following are equivalent*:

(i) $\mathrm{PA} + \gamma$ *is relatively interpretable in* $\mathrm{PA} + \delta$,

(ii) *For every* $n \in \omega$, $\mathrm{PA} \vdash \delta \to \mathrm{con}(\mathrm{I}\Sigma_n + \gamma)$,

(iii) *Every model of* $\mathrm{PA} + \delta$ *has an endextension modeling* $\mathrm{PA} + \gamma$.

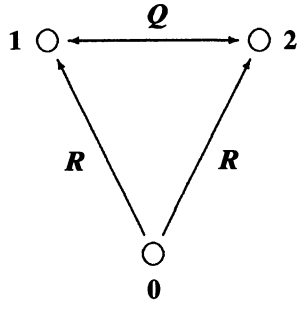
(Orey [15], Hájek [5], and Guaspari [3]; see Berarducci [1] for a concise presentation. This fact also serves as an excuse for not giving the definition of relative interpretability here.)

Note that for γ an oreysentence one can by iteratedly applying (i) $\Rightarrow$ (iii) construct an infinite sequence $(\mathcal{M}_i)_{i\in\omega}$ of models of PA s.t. $\mathcal{M}_{i+1}$ is an endextension of $\mathcal{M}_i$ and $\mathcal{M}_{i+1} \models \gamma$ iff $\mathcal{M}_i \not\models \gamma$. Thus γ exhibits a mutability alien not just to Σ_1 sentences, but also to any boolean combination of those (see Manevitz & Stavi [10] for more details).

Proposition 5.3 *There exists an oreysentence which also is a $\triangle$-henkinsentence.*

Proof: We shall construct the promised orey- and $\triangle$-henkinsentence by means of another Solovay function. These functions have traditionally only been used to obtain results of mind-sweeping generality, such as exemplified by our Theorem 4.9, and resorting to this method to construct just one individual sentence almost amounts to a breach of the code of honor. Nevertheless, I believe that this format might help the reader to better visualize the proof thus enhancing the digestibility of the argument.

The scene is set by the following Kripke frame:



The new name Q for the auxiliary relation suggests that it is going to be treated in a way different from the one the relation S of marshmallow frames was treated in Section 4.

We define the two relevant functions:

$$f(0) = 0; \qquad g(0) = \infty.$$

Case A: $0 = f(x)Ra$ *and* x *is a* PA*-proof of* $\ell \neq a$.

$$f(x+1) = a; \qquad g(x+1) = \text{any } y \text{ s.t. } x \text{ is an } I\Sigma_y\text{-proof of } \ell \neq a.$$

Case B: $f(x)Qb$, *and* x *is an* $I\Sigma_{g(x)-1}$*-proof of* $\ell \neq b$.

$$f(x+1) = b; \qquad g(x+1) = g(x).$$

Case C: $f(x) \neq 0$ *and* x *is a proof in* $I\Sigma_{g(x)}$ *of* $\perp$.

$$f(x+1) = f(x); \qquad g(x+1) = g(x) - 1.$$

Case D: None of the previous cases apply.

$$f(x+1) = f(x); \qquad g(x+1) = g(x).$$

The sentence $\ell \neq d$ expresses, as before, that d is not the limit value of f.

First note that g is a decreasing function. Second, two occurrences of Case B at the same value g guarantee at least one occurrence of Case C, and hence a decrease in the value of g. Therefore $I\Sigma_1$ proves that both f and g reach their limit values. Call the ε-terms naming these limits ℓ and M, respectively.

Claim 1 ($I\Sigma_1$): *If* $\ell = 1$ *then* $I\Sigma_M \vdash \ell = 1$.
Assume $\ell = 1$. Reason in $I\Sigma_M$:

> If f moves out of 1 and reaches a limit value a, then this can only be due to a proof in $I\Sigma_{\overline{M}-1}$ of $\ell \neq a$ which cannot exist since $\ell = a$. Hence $\ell = 1$.

Claim 2 ($I\Sigma_1$): $I\Sigma_M$ *is consistent.*
If $\ell = 0$ then $I\Sigma_M = \text{PA}$ is clearly consistent. If $\ell \neq 0$ then the claim is immediate on inspection of Case C.

Claim 3 ($I\Sigma_1$): *If* $\ell \neq 0$ *then* $I\Sigma_{M+1}$ *is inconsistent.*
If g arrives at M by Case C this is clear. If it does so by Case A then this can, as usual, be established by reflection, for $I\Sigma_{M+1}$ knows that $I\Sigma_{\overline{M}} \vdash \ell \neq \bar{\ell}$.

(In addition to Claims 1–3, the reader may amuse him/herself by showing that, provably in $I\Sigma_1$, any occurrence of Case A or B is eventually succeeded by an occurrence of Case C.)

So,

(1) $I\Sigma_1 \vdash \ell = 1 \to \triangle \ell = 1$ (by Claims 1 and 2)

(2) $I\Sigma_1 \vdash \ell = 2 \to \triangle \ell = 2$ (symmetric to (1))

(3) $I\Sigma_1 \vdash \triangle \ell = 1 \to \triangle \ell \neq 2$

$$\to \Box \ell \neq 2$$

$$\to \ell \neq 0$$

$$\to .\ \ell = 1 \vee \ell = 2$$

$$\to \ell = 1 \qquad \text{(by (2) and (3)).}$$

Thus $\ell = 1$ is indeed a $\triangle$-henkinsentence. Now we establish that it is an oreysentence as well. Consider an $n \in \omega$. Reason in PA:

> By Corollary 1.2 and Claim 3 there holds $n < M$ and hence, assuming $\ell = 1$, a proof in $I\Sigma_n$ of $\ell \neq 2$ would by Case B move f to 2 and decrease g below its limit.

Therefore

$$\text{(4) } \mathrm{PA} \vdash \ell = 1 \to \mathrm{con}(I\Sigma_n + \ell = 2) \to \mathrm{con}(I\Sigma_n + \ell \neq 1).$$

On the other hand,

$$\mathrm{PA} \vdash \ell \neq 1 \to .\ \ell = 0 \vee \ell = 2 \to .\ \mathrm{con}(\mathrm{PA} + \ell = 1) \vee \mathrm{con}(I\Sigma_n + \ell = 1) \quad \text{(symmetric to (4))} \to \mathrm{con}(I\Sigma_n + \ell = 1).$$

Thus by (ii) $\Rightarrow$ (i) of Proposition 5.2, $\ell = 1$ is an oreysentence and satisfies the statement of the Proposition ($\ell = 2$ would have done just as well).

Remark 5.4 Proposition 5.3 also holds for Feferman predicates other than the one based on $(I\Sigma_n)_{n\in\omega}$. One only has to modify the construction so as to ensure that in Case B $\mathrm{PA}\lceil g(x)$ can reflect on the proof x.

6 More gödelsentences In this Section we shall produce one more example highlighting the sensitivity of the provability-logical behavior of the Feferman predicate to the choice of the base sequence of theories. With Proposition 4.15 we have already experienced the fragility of the axiom schema (S) and Proposition 2.3 in this respect.

Here we assault Proposition 2.6 which asserted the provable uniqueness of $\triangle$-gödelsentences for Feferman predicates satisfying (S). In other words, we set out to produce a sequence of theories whose Feferman predicate possesses inequivalent gödelsentences. The uniqueness question for gödelsentences of Feferman predicates was raised by Montagna [11]. Smoryński [19] discovered uniqueness of gödelsentences of $\triangle$'s based on $(I\Sigma_n)_{n\in\omega}$ and similar sequences. The same paper also ponders the question whether this uniqueness might fail under a weirder choice of the base sequence. By settling this we show that the situation is not unlike that with Rosser predicates: Guaspari and Solovay [4] demonstrate that provable uniqueness of gödelsentences for these predicates (= rossersentences) depends on certain details of gödelnumbering of proofs that have not yet been seen to make clear proof-theoretic sense in other contexts.

Oddly enough, our main tool for this task is the Solovay function F launched in Section 4 that was originally called to life to prove the completeness theorem for the provability logic of a $\triangle$ with provably unique gödelsentences.

Consider the predicate $\triangle'$ whose base sequence is $(I\Sigma_{n+1}(I\Sigma_n))_{n\in\omega}$, where the theories $I\Sigma_{n+1}(I\Sigma_n)$ are defined by the requirement that

$$I\Sigma_{n+1}(I\Sigma_n) \vdash \gamma \quad \text{iff} \quad I\Sigma_{n+1} \vdash \text{`}I\Sigma_n \vdash \gamma\text{'}$$

for all sentences γ. While the reader will easily recognize the theory $I\Sigma_{n+1}(I\Sigma_n)$ as proving exactly the same theorems as $I\Sigma_n$, we shall see that gödelsentences are confused by this disguise to the point of losing their unique identity.

The next three lemmas investigate the interrelations of $\triangle'$ with the construction of Definitions 4.1 and 4.3 and in what follows we adopt some of the notation from Section 4.

Lemma 6.1 ($I\Sigma_1$) *If $\mu < \infty$ then μ is the largest x s.t. $I\Sigma_x(I\Sigma_{x-1})$ is consistent. Therefore for any arithmetical sentence γ one has $\triangle'\gamma$ iff $I\Sigma_\mu(I\Sigma_{\bar{\mu}-1}) \vdash \gamma$.*

Proof: By Lemma 4.5 $I\Sigma_{\mu+1}$ is inconsistent and hence so is $I\Sigma_{\mu+1}(I\Sigma_{\bar{\mu}})$. In the other direction, if $I\Sigma_\mu \vdash$ '$I\Sigma_{\bar{\mu}-1} \vdash \perp$', then one by reflection has $I\Sigma_\mu \vdash \perp$ which would contradict Lemma 4.5.

Lemma 6.2 ($I\Sigma_2$) *Suppose $L = a \in T$.*

(a) $I\Sigma_\mu(I\Sigma_{\bar{\mu}-1}) \vdash \mathbb{W}_{a(S\cup S^2)b} L = b$.

(b) *If aSb then $I\Sigma_\mu(I\Sigma_{\bar{\mu}-1})$ is consistent with $L = b$.*

(c) *If aS^2b then $I\Sigma_\mu(I\Sigma_{\bar{\mu}-1})$ is consistent with $L = b$.*

Proof: (a) Take a node b which does not satisfy $a(S \cup S^2)b$. Reason in $I\Sigma_\mu$:

Since $L \neq a$

(recall that there is an $I\Sigma_\mu$-proof of $L \neq a$ since in real life L is equal to a),

the function F will have to abandon a for a different node. Reason in $I\Sigma_{\bar{\mu}-1}$:

> Were F to reach the node b it would, after fleeing a, be compelled to arrive at and leave two nodes in D, b being at least that far away. Therefore a hypothetical proof of $L \neq b$ that finally transports F to b is a proof in $I\Sigma_{\bar{\bar{\mu}}-2}$ reflecting whereupon we get $L \neq b$.

This way we have shown $I\Sigma_\mu \vdash$ '$I\Sigma_{\bar{\mu}-1} \vdash L \neq b$' which proves the claim.

(b) If for a node b s.t. aSb we had $I\Sigma_\mu \vdash$ '$I\Sigma_{\bar{\mu}-1} \vdash L \neq b$', we could conclude $I\Sigma_\mu \vdash L \neq b$ by reflection, which contradicts Lemma 4.6(b).

(c) Let c be a node s.t. $aScSb$. Suppose for a contradiction that $I\Sigma_\mu \vdash$ '$I\Sigma_{\bar{\mu}-1} \vdash L \neq b$' and reason in $I\Sigma_\mu$:

> Recall that by Proposition 1.1 the theory $I\Sigma_{\bar{\mu}-1}$ is consistent. Therefore by Lemma 4.5 $\mu \geq \bar{\mu} - 1$ and so $I\Sigma_\mu \vdash L \neq b$. By Lemma 4.6(b) we cannot then have $L = c$.

But $I\Sigma_\mu \vdash L \neq c$ contradicts Lemma 4.6(b).

The reason why Lemma 6.2 takes $I\Sigma_2$ instead of the usual $I\Sigma_1$ to formalize is rather trivial: $I\Sigma_2$ proves that $\mu - 1 \geq 0$ so that $I\Sigma_{\mu-1}$ can be trusted to carry out the innermost argument in the proof of clause (a).

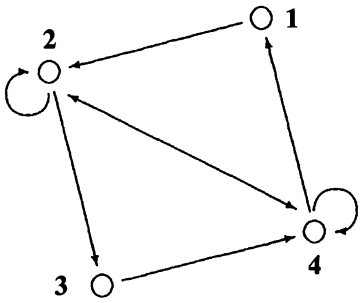
Lemma 6.3 ($I\Sigma_2$) *(a) If $L = a \neq 0$ then $\triangle'\mathbb{W}_{a(S\cup S^2)b} L = b$.*

(b) *If $L = a(S \cup S^2)b$ then $\triangledown' L = b$.*

Proof: If $L \notin T$ then PA is consistent and $\triangle'$ is the same as $\Box$. Lemma 4.4 settles this case. For $L \in T$ use Lemma 6.1 and Lemma 6.2.

Proposition 6.4 *There exists a base sequence of subtheories of* PA *s.t. the Feferman predicate based on it possesses inequivalent gödelsentences.*

Proof: Consider the predicate $\triangle'$ studied in Lemmas 6.1–3. Let us apply Lemma 6.3 to the marshmallow frame obtained from the marshmallow depicted below by adjoining an R-bottom 0.



An easy verification using that lemma shows that

(1) $\mathrm{I}\Sigma_2 \vdash L \neq 1 \leftrightarrow \neg\triangle' L \neq 1$

and, in perfect symmetry,

(2) $\mathrm{I}\Sigma_2 \vdash L \neq 3 \leftrightarrow \neg\triangle' L \neq 3$.

So $L \neq 1$ and $L \neq 3$ are $\triangle'$-gödelsentences that are clearly inequivalent.

We would be cheating were we to declare the sequence $(\mathrm{I}\Sigma_{n+1}(\mathrm{I}\Sigma_n))_{n\in\omega}$ and the predicate $\triangle'$ to be a solution to our problem. For while it is easily seen that each theory in this sequence extends its predecessor and the union of the sequence is equal to PA, the question, as addressed from within PA, whether the theories $\mathrm{I}\Sigma_{n+1}(\mathrm{I}\Sigma_n)$ are finitely axiomatizable appears to be much more complicated. Rather than go into that we shall indicate how to modify the sequence so as to obtain one of finitely axiomatized theories which would still preserve the pluralistic effect on gödelsentences. Here we use an adaptation of a technique of Lindström [9] for constructing a finite axiomatization of bounded-complexity consequences of a theory.

For brevity, denote the earlier constructed sentences $L \neq 1$ and $L \neq 3$ by γ_0 and γ_1, respectively, and let further γ_2 denote $\bot$. Consider the following fixed point equation on the formula $\xi(x)$:

$$\mathrm{I}\Sigma_1 \vdash \forall x\Big(\xi(x) \leftrightarrow \forall y\Big(\bigwedge_{0\le i\le 2} ((\mathrm{I}\Sigma_x \vdash_y \xi(x) \to \gamma_i) \to \gamma_i) \to\cdot \bigwedge_{0\le i\le 2} (\mathrm{I}\Sigma_{x+1} \vdash_y \text{`}\mathrm{I}\Sigma_x \vdash \gamma_i\text{'}) \to \gamma_i)\Big)\Big),$$

where $\mathrm{T} \vdash_y \delta$ stands for the formula expressing that δ has a T-proof of gödelnumber $\le y$. One can mimick the proof of Lemma 2, Case 1 of Lindström [9]

to the effect that for $0 \leq i \leq 2$ and any n the theory $I\Sigma_n + \xi(n)$ proves γ_i if and only if $I\Sigma_{n+1}(I\Sigma_n)$ does. Moreover, this proof is straightforwardly formalizable so that for all three i's we get

(3) $I\Sigma_1 \vdash \forall x\big((I\Sigma_x \vdash \xi(x) \to \gamma_i) \leftrightarrow I\Sigma_{x+1} \vdash \text{`}I\Sigma_x \vdash \gamma_i\text{'}\big)$.

Furthermore, we show that one can do with the formula $I\Sigma_{x+1} \vee \xi(x)$ in place of $\xi(x)$ as well.

$$
\begin{aligned}
I\Sigma_1 \vdash \forall x\big((I\Sigma_x \vdash \xi(x) \to \gamma_i) &\to I\Sigma_{x+1} \vdash \text{`}I\Sigma_x \vdash \gamma_i\text{'}\big) && \text{(by (3))}\\
&\to I\Sigma_{x+1} \vdash \gamma_i) && \text{(by reflection)}\\
&\to I\Sigma_x \vdash \text{`}I\Sigma_{x+1} \vdash \gamma_i\text{'})\\
&\to (I\Sigma_x \vdash I\Sigma_{x+1} \to \gamma_i))\\
&\to (I\Sigma_x \vdash I\Sigma_{x+1} \vee \xi(x) \to .\ \gamma_i)). && \text{(4)}
\end{aligned}
$$

The other direction $I\Sigma_1 \vdash \forall x((I\Sigma_x \vdash I\Sigma_{x+1} \vee \xi(x) \to .\ \gamma_i) \to (I\Sigma_x \vdash \xi(x) \to \gamma_i))$ is immediate. Note that for $i = 2$ this means that $I\Sigma_{n+1}(I\Sigma_n)$ and $I\Sigma_n + (I\Sigma_{n+1} \vee \xi(n))$ are provably equiconsistent.

We now claim that the sequence of theories $(I\Sigma_n + (I\Sigma_{n+1} \vee \xi(n)))_{n\in\omega}$ together with the Feferman predicate $\tilde{\triangle}$ based on this sequence satisfy all requirements of the present Proposition. First, each theory is finitely axiomatized and

$$I\Sigma_{n+1} \vdash I\Sigma_n + (I\Sigma_{n+1} \vee \xi(n)) \vdash I\Sigma_n$$

so that the sequence monotonously increases in strength and exhausts the whole of PA.

Next we demonstrate that for $0 \leq i \leq 1$ the γ_i's we have constructed are $\tilde{\triangle}$-gödelsentences as well as $\triangle'$-ones:

$$
\begin{aligned}
I\Sigma_2 \vdash \neg\gamma_i &\leftrightarrow \triangle'\gamma_i && \text{(by (1) and (2))}\\
&\leftrightarrow \exists x\big(\mathrm{con}\big(I\Sigma_{x+1}(I\Sigma_x)\big) \wedge \big(I\Sigma_{x+1}(I\Sigma_x) \vdash \gamma_i\big)\big)\\
&\leftrightarrow \exists x\big(\mathrm{con}\big(I\Sigma_x + (I\Sigma_{x+1} \vee \xi(x))\big) \wedge \big(I\Sigma_x + (I\Sigma_{x+1} \vee \xi(x)) \vdash \gamma_i\big)\big)\\
& && \text{(by (3) and (4))}\\
&\leftrightarrow \tilde{\triangle}\gamma_i.
\end{aligned}
$$

Thus γ_0 and γ_1 are inequivalent $\tilde{\triangle}$-gödelsentences.

Another fact of a distinctly similar flavor that contrasts Proposition 2.6 is the following:

Exercise 6.5 (Visser) Show that there are infinitely many inequivalent sentences γ satisfying

$$\mathrm{PA} \vdash \gamma \leftrightarrow \neg\triangle\triangle\gamma$$

for the Feferman predicate $\triangle$ based on $(I\Sigma_n)_{n\in\omega}$.

Acknowledgments The present paper has been inspired by the joint effect of Visser [22] and Smoryński [19]. The author would like to thank Dick de Jongh, Rineke Verbrugge, and Albert Visser, who extradited a collection of parasites from an earlier draft of the

present paper and suggested a large number of improvements and simplifications. Conversations with these people also led the author to produce some of the material that he considered appropriate to include here. Further errors were spotted by a referee.

REFERENCES

[1] Berarducci, A., "The interpretability logic of Peano Arithmetic," *The Journal of Symbolic Logic*, vol. 55 (1990), pp. 1059–1089.

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

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

[4] Guaspari, D. and R. M. Solovay, "Rosser sentences," *Annals of Mathematical Logic*, vol. 16 (1979), pp. 81–99.

[5] Hájek, P., "On interpretability in set theories," *Commentationes Mathematicae Universitatis Carolinae*, vol. 12 (1961), pp. 73–79.

[6] Kaye, R., *Models of Peano Arithmetic*, Clarendon Press, Oxford, 1991.

[7] Leivant, D., "The optimality of induction as an axiomatization of arithmetic," *The Journal of Symbolic Logic*, vol. 48 (1983), pp. 182–184.

[8] Lindström, P., "Some results on interpretability," pp. 329–361 in *Proceedings from 5th Scandinavian Logic Symposium*, edited by F. V. Jensen, B. H. Mayoh, and K. K. Møller, Aalborg University Press, Aalborg, 1979.

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

[10] Manevitz, L. and J. Stavi, "Δ_2^0 operators and alternating sentences in arithmetic," *The Journal of Symbolic Logic*, vol. 45 (1980), pp. 144–154.

[11] Montagna, F., "On the algebraization of a Feferman's predicate (the algebraization of theories which express Theor; X)," *Studia Logica*, vol. 37 (1978), pp. 221–236.

[12] Montagna, F., "Provability in finite subtheories of PA and relative interpretability: a modal investigation," *The Journal of Symbolic Logic*, vol. 52 (1987), pp. 494–511.

[13] Mostowski, A., "On models of axiomatic systems," *Fundamenta Mathematicae*, vol. 39 (1952), pp. 133–158.

[14] Ono, H., "Reflection principles in fragments of Peano arithmetic," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 33 (1987), pp. 317–333.

[15] Orey, S., "Relative interpretations," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 7 (1961), pp. 146–153.

[16] Paris, J. B. and L. A. S. Kirby, "Σ_n-collection schemas in arithmetic," pp. 199–209 in *Logic Colloquium '77*, edited by A. Macintyre, L. Pacholski, and J. Paris, North-Holland, Amsterdam, 1978.

[17] Sieg, W., "Fragments of arithmetic," *Annals of Pure and Applied Logic*, vol. 28 (1985), pp. 33–71.

[18] Smoryński, C., *Self-Reference and Modal Logic*, Springer-Verlag, New York, 1985.

[19] Smoryński, C., "Arithmetic analogues of McAloon's unique Rosser sentences," *Archive for Mathematical Logic*, vol. 28 (1989), pp. 1–21.

[20] Solovay, R. M., "Provability interpretations of modal logic," *Israel Journal of Mathematics*, vol. 25 (1976), pp. 287–304.

[21] Švejdar, V., "Degrees of interpretability," *Commentationes Mathematicae Universitatis Carolinae*, vol. 19 (1978), pp. 789–813.

[22] Visser, A., "Peano's smart children: a provability logical study of systems with built-in consistency," *Note Dame Journal of Formal Logic*, vol. 30 (1989), pp. 161–196.

[23] Visser, A., "The formalization of interpretability," *Logic Group Preprint Series No. 47*, Department of Philosophy, University of Utrecht, 1989.

Department of Mathematics and Computer Science
University of Amsterdam, Plantage Muidergracht 24
1018 TV Amsterdam, The Netherlands
e-mail: volodya@fwi.uva.nl