Institute for Logic, Language and Computation

A SMART CHILD OF PEANO'S

V.Yu. Shavrukov
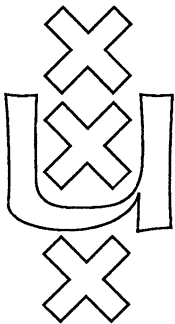
University of Amsterdam

The ILLC Prepublication Series

1990

Logic, Semantics and Philosophy of Language

LP-90-01	Jaap van der Does	A Generalized Quantifier Logic for Naked Infinitives
LP-90-02	Jeroen Groenendijk, Martin Stokhof	Dynamic Montague Grammar
LP-90-03	Renate Bartsch	Concept Formation and Concept Composition
LP-90-04	Aarne Ranta	Intuitionistic Categorial Grammar
LP-90-05	Patrick Blackburn	Nominal Tense Logic
LP-90-06	Gennaro Chierchia	The Variablity of Impersonal Subjects
LP-90-07	Gennaro Chierchia	Anaphora and Dynamic Logic
LP-90-08	Herman Hendriks	Flexible Montague Grammar
LP-90-09	Paul Dekker	The Scope of Negation in Discourse, towards a Flexible Dynamic Montague grammar
LP-90-10	Theo M.V. Janssen	Models for Discourse Markers
LP-90-11	Johan van Benthem	General Dynamics
LP-90-12	Serge Lapierre	A Functional Partial Semantics for Intensional Logic
LP-90-13	Zhisheng Huang	Logics for Belief Dependence
LP-90-14	Jeroen Groenendijk, Martin Stokhof	Two Theories of Dynamic Semantics
LP-90-15	Maarten de Rijke	The Modal Logic of Inequality
LP-90-16	Zhisheng Huang, Karen Kwast	Awareness, Negation and Logical Omniscience
LP-90-17	Paul Dekker	Existential Disclosure, Implicit Arguments in Dynamic Semantics

Mathematical Logic and Foundations

ML-90-01	Harold Schellinx	Isomorphisms and Non-Isomorphisms of Graph Models
ML-90-02	Jaap van Oosten	A Semantical Proof of De Jongh's Theorem
ML-90-03	Yde Venema	Relational Games
ML-90-04	Maarten de Rijke	Unary Interpretability Logic
ML-90-05	Domenico Zambella	Sequences with Simple Initial Segments
ML-90-06	Jaap van Oosten	Extension of Lifschitz' Realizability to Higher Order Arithmetic, and a Solution to a Problem of F. Richman
ML-90-07	Maarten de Rijke	A Note on the Interpretability Logic of Finitely Axiomatized Theories
ML-90-08	Harold Schellinx	Some Syntactical Observations on Linear Logic
ML-90-09	Dick de Jongh, Duccio Pianigiani	Solution of a Problem of David Guaspari
ML-90-10	Michiel van Lambalgen	Randomness in Set Theory
ML-90-11	Paul C. Gilmore	The Consistency of an Extended NaDSet

Computation and Complexity Theory

CT-90-01	John Tromp, Peter van Emde Boas	Associative Storage Modification Machines
CT-90-02	Sieger van Denneheuvel, Gerard R. Renardel de Lavalette	A Normal Form for PCSJ Expressions
CT-90-03	Ricard Gavaldà, Leen Torenvliet, Osamu Watanabe, José L. Balcázar	Generalized Kolmogorov Complexity in Relativized Separations
CT-90-04	Harry Buhrman, Edith Spaan, Leen Torenvliet	Bounded Reductions
CT-90-05	Sieger van Denneheuvel, Karen Kwast	Efficient Normalization of Database and Constraint Expressions
CT-90-06	Michiel Smid, Peter van Emde Boas	Dynamic Data Structures on Multiple Storage Media, a Tutorial
CT-90-07	Kees Doets	Greatest Fixed Points of Logic Programs
CT-90-08	Fred de Geus, Ernest Rotterdam, Sieger van Denneheuvel, Peter van Emde Boas	Physiological Modelling using RL
CT-90-09	Roel de Vrijer	Unique Normal Forms for Combinatory Logic with Parallel Conditional, a case study in conditional rewriting

Other Prepublications

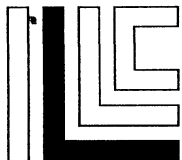|		
X-90-01	A.S. Troelstra	Remarks on Intuitionism and the Philosophy of Mathematics, Revised Version
X-90-02	Maarten de Rijke	Some Chapters on Interpretability Logic
X-90-03	L.D. Beklemishev	On the Complexity of Arithmetical Interpretations of Modal Formulae
X-90-04		Annual Report 1989
X-90-05	Valentin Shehtman	Derived Sets in Euclidean Spaces and Modal Logic
X-90-06	Valentin Goranko, Solomon Passy	Using the Universal Modality: Gains and Questions
X-90-07	V.Yu. Shavrukov	The Lindenbaum Fixed Point Algebra is Undecidable
X-90-08	L.D. Beklemishev	Provability Logics for Natural Turing Progressions of Arithmetical Theories
X-90-09	V.Yu. Shavrukov	On Rosser's Provability Predicate
X-90-10	Sieger van Denneheuvel, Peter van Emde Boas	An Overview of the Rule Language RL/1
X-90-11	Alessandra Carbone	Provable Fixed points in $I\Delta_0+\Omega_1$, revised version
X-90-12	Maarten de Rijke	Bi-Unary Interpretability Logic
X-90-13	K.N. Ignatiev	Dzhaparidze's Polymodal Logic: Arithmetical Completeness, Fixed Point Property, Craig's Property
X-90-14	L.A. Chagrova	Undecidable Problems in Correspondence Theory
X-90-15	A.S. Troelstra	Lectures on Linear Logic

1991

Logic, Semantics and Philosophy of Langauge

LP-91-01	Wiebe van der Hoek, Maarten de Rijke	Generalized Quantifiers and Modal Logic
LP-91-02	Frank Veltman	Defaults in Update Semantics
LP-91-03	Willem Groeneveld	Dynamic Semantics and Circular Propositions
LP-91-04	Makoto Kanazawa	The Lambek Calculus enriched with Additional Connectives
LP-91-05	Zhisheng Huang, Peter van Emde Boas	The Schoenmakers Paradox: Its Solution in a Belief Dependence Framework
LP-91-06	Zhisheng Huang, Peter van Emde Boas	Belief Dependence, Revision and Persistence
LP-91-07	Henk Verkuyl, Jaap van der Does	The Semantics of Plural Noun Phrases
LP-91-08	Víctor Sánchez Valencia	Categorial Grammar and Natural Reasoning
LP-91-09	Arthur Nieuwendijk	Semantics and Comparative Logic
LP-91-10	Johan van Benthem	Logic and the Flow of Information

Mathematical Logic and Foundations

ML-91-01	Yde Venema	Cylindric Modal Logic
ML-91-02	Alessandro Berarducci, Rineke Verbrugge	On the Metamathematics of Weak Theories
ML-91-03	Domenico Zambella	On the Proofs of Arithmetical Completeness for Interpretability Logic
ML-91-04	Raymond Hoofman, Harold Schellinx	Collapsing Graph Models by Preorders
ML-91-05	A.S. Troelstra	History of Constructivism in the Twentieth Century
ML-91-06	Inge Bethke	Finite Type Structures within Combinatory Algebras
ML-91-07	Yde Venema	Modal Derivation Rules
ML-91-08	Inge Bethke	Going Stable in Graph Models
ML-91-09	V.Yu. Shavrukov	A Note on the Diagonalizable Algebras of PA and ZF
ML-91-10	Maarten de Rijke, Yde Venema	Sahlqvist's Theorem for Boolean Algebras with Operators
ML-91-11	Rineke Verbrugge	Feasible Interpretability
ML-91-12	Johan van Benthem	Modal Frame Classes, revisited

A SMART CHILD OF PEANO'S

V.Yu. Shavrukov
Department of Mathematics and Computer Science
University of Amsterdam

A Smart Child of Peano's

V. Yu. Shavrukov

Department of Mathematics & Computer Science

University of Amsterdam

Plantage Muidergracht 24

1018 TV Amsterdam

the Netherlands

volodya@fwi.uva.nl

December 8, 1992

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. $\mathrm{I}\Sigma_x$ 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 gödelsentences for $\triangle$, is also considered.

0. 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 $(\mathrm{PA}{\restriction}n)_{n\in\omega}$ with $\mathrm{PA}{\restriction}n{+}1$ extending $\mathrm{PA}{\restriction}n$, and $\mathrm{PA} = \bigcup_{n\in\omega}\mathrm{PA}{\restriction}n$.

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. $\mathrm{PA}{\restriction}x$ is consistent and $\mathrm{PA}{\restriction}x \vdash \varphi$".

The sequence $(\mathrm{PA}{\restriction}n)_{n\in\omega}$ 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 $\mathrm{PA}{\restriction}n$ 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

1

relative interpretability became apparent in Feferman [2], Orey [15] and Hájek [5]. The celebrated fixed point of Lindström [8, Theorem B] and Švejdar [21, Lemma 4.5] 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 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 $PA\lceil n$ is fairly unimportant (see Orey [15], Lindström [8], Švejdar [21], Montagna [12] or Berarducci [1]). One is therefore usually contented with the traditional choice

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

The relation of $PA\lceil n$ to $PA\lceil n+1$ becomes then dependent on tiny intimate details of gödelnumbering of sentences and thus the only feasibly available properties of $(PA\lceil n)_{n \in \omega}$ 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 5, whose validity is dependent on the exact choice of $(PA\lceil n)_{n \in \omega}$. He also shows that a more specific choice

$$PA\lceil n = I\Sigma_n$$

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

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

$$I\Sigma_{n+1} \vdash \text{'for every } \Pi_{n+2}\text{-sentence } \pi, \text{ if } I\Sigma_n \vdash \pi \text{ then } \pi \text{ is true'} \qquad \blacksquare$$

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

0.2. FACT ($I\Sigma_1$). *For all* $m, n \in \omega$ *PA proves uniform* Π_m*-reflection for* $I\Sigma_n$. $\qquad \blacksquare$

The property of PA expressed by Fact 0.2 is known under the name of *essential reflexivity*.

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

A number of other sequences of theories is known to enjoy properties similar to Fact 0.1 but we shall not strive for more generality. In this paper we stick almost exclusively with the definition of $\triangle$ based on $(I\Sigma_n)_{n \in \omega}$ and construct the joint provability

2

logic of $\square$ and $\triangle$. The peculiarity of $\triangle$ in provability-logical context is that the modal operator corresponding to this predicate asks for a Kripke semantics encorporating 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 (Fact 0.1) and in fact reflection keeps this lack to a minimum by providing a new modal principle approaching transitivity.

In Section 1 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 2. Section 3 proves the arithmetical completeness theorem for **LF**. Finally, in Sections 4 and 5 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.

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.

1. LF

1.1. DEFINITION. The language of the system **LF** is the propositional language with two unary modal operators $\square$ and $\triangle$. Formulae in this language will be referred to as ($\square\triangle$-)*formulae*. Shorthand for $\neg\triangle\neg$ is $\triangledown$. The operator $\square$ 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) $\square\varphi \to \square\triangle\varphi$

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

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

(F5) $\triangledown\top$

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

3

The only new rule is the $\triangle$-*necessitation rule* $\dfrac{\varphi}{\triangle\varphi}$.

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 $\square\triangle$-formulas for the same propositional variable in another $\square\triangle$-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 $\square\varphi \to \square\square\varphi$. In **LF** the operator $\triangle$ also enjoys a weak analogue of Löb's schema $\square(\square\varphi \to \varphi) \to \square\varphi$ which, for a change, follows now from weak transitivity (in fact, the weakened versions are interderivable):

1.2. PROPOSITION. *For* $\{\psi_i\}_{i\in I}$ *a finite collection of* $\square\triangle$-*formulas and* φ *an arbitrary* $\square\triangle$-*formula one has*

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

PROOF. First we prove $\mathbf{LF} \vdash \triangle(\square\bot \vee \triangle\varphi \to.\ \varphi) \to \triangle\varphi$.

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

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

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

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

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

$\qquad\qquad\qquad\qquad\qquad \to \square(\square\varphi \to \varphi)$ (by (F4))

$\qquad\qquad\qquad\qquad\qquad \to \square\varphi$ (by Löb's axiom)

$\qquad\qquad\qquad\qquad\qquad \to \triangle\square\varphi$ (by (F3))

$\qquad\qquad\qquad\qquad\qquad \to \triangle\varphi$ (by (5))   q.e.d.

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

4

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

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

$$\text{(by (S))}$$

$$\to \triangle\Big((\triangle\psi \to \psi) \vee \triangle\big(\psi \wedge ((\triangle\psi \to \psi) \to \varphi)\big)\Big)$$

$$\text{(by (6))}$$

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

$$\to \triangle(\varphi \vee \triangle\varphi) \qquad\qquad \text{(by (8))}$$

$$\to \triangle\varphi \qquad\qquad \text{(by (7))} \quad \text{q.e.d.}$$

Finally, suppose that $\square\triangle$-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 $\mathbf{LF} \vdash \triangle\big((\chi \wedge \theta) \vee \triangle\varphi \to\!. \ \varphi\big) \to \triangle\varphi$ for φ arbitrary. Here is the derivation:

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

(10) $\mathbf{LF} \vdash \triangle\big(\chi \vee \triangle(\theta \vee \triangle\varphi \to\!. \ \varphi) \to\!. \ (\theta \vee \triangle\varphi \to\!. \ \varphi)\big) \to \triangle(\theta \vee \triangle\varphi \to\!. \ \varphi)$ \hfill (by IH)

$\mathbf{LF} \vdash \triangle\big((\chi \wedge \theta) \vee \triangle\varphi \to\!. \ \varphi\big) \to \triangle\big((\chi \vee \triangle\varphi) \wedge (\theta \vee \triangle\varphi) \to\!. \ \varphi\big)$

$$\text{(by propositional logic)}$$

$$\to \triangle\big(\chi \vee \triangle\varphi \to\!. \ (\theta \vee \triangle\varphi \to\!. \ \varphi)\big)$$

$$\to \triangle\big(\chi \vee \triangle(\theta \vee \triangle\varphi \to\!. \ \varphi) \to\!. \ (\theta \vee \triangle\varphi \to\!. \ \varphi)\big)$$

$$\text{(by (9))}$$

$$\to \triangle(\theta \vee \triangle\varphi \to\!. \ \varphi) \qquad\qquad \text{(by (10))}$$

$$\to \triangle\varphi \qquad\qquad \text{(by (9))} \quad \text{q.e.d.} \quad \blacksquare$$

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

1.3. PROPOSITION. *For $\{\psi_i\}_{i \in I}$ a finite collection of $\square\triangle$-formulas one has*

$$\mathbf{LF} \vdash \nabla\Big(\square\bot \wedge \bigwedge_{i \in I}(\triangle\psi_i \to \psi_i)\Big)$$

PROOF. We shall prove that the negation of this formula implies $\bot$ in $\mathbf{LF}$, that is $\mathbf{LF} \vdash \triangle\neg\big(\square\bot \wedge \bigwedge_{i \in I}(\triangle\psi_i \to \psi_i)\big) \to \bot$:

$$\mathbf{LF} \vdash \triangle\neg\Big(\square\bot \wedge \bigwedge_{i \in I}(\triangle\psi_i \to \psi_i)\Big) \to \triangle\Big(\big(\square\bot \wedge \bigwedge_{i \in I}(\triangle\psi_i \to \psi_i)\big) \vee \triangle\bot \to\!. \ \bot\Big)$$

$$\text{(by (F5))}$$

$$\to \triangle\bot \qquad\qquad \text{(by Proposition 1.2)}$$

$$\to \bot \qquad\qquad \text{(by (F5))} \quad \text{q.e.d.} \quad \blacksquare$$

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

1.4. DEFINITION. A function $^\circ$ assigning arithmetic sentences to $\Box\Delta$-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) $(\Delta\varphi)^\circ$ is 'there exists an x s.t. PA$\lceil x$ is consistent and PA$\lceil x \vdash \varphi^\circ$' (recall that we have agreed that PA$\lceil x = I\Sigma_x$ unless otherwise specified).

1.5. PROPOSITION. *For any gf-interpretation $^\circ$, **LF** $\vdash \varphi$ implies 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 PA$\lceil n = I\Sigma_n$.

We therefore only check (S): $\Delta\varphi \to \Delta\big((\Delta\psi \to \psi)\vee\Delta\varphi\big)$. Suppose $\psi^\circ \in \Pi_n$. Reason in PA:

Assume $(\Delta\varphi)^\circ$.

Case 1. PA *is consistent.*

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

Case 2. PA *is inconsistent.*

Let μ be the maximal x s.t. $I\Sigma_x$ is consistent. (Note that by Fact 0.2, $\mu + 1 \geq n$.) We then have $I\Sigma_\mu \vdash \varphi^\circ$ and we shall prove $\big((\Delta\psi \to \psi) \vee \Delta\varphi\big)^\circ$ inside $I\Sigma_\mu$:

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

Case 2.2. $I\Sigma_{\bar\mu}$ *is inconsistent.*
By Fact 0.1 the theory $I\Sigma_{\bar\mu-1}$ is consistent and hence for any arithmetical sentence γ one has $\Delta\gamma$ iff $I\Sigma_{\bar\mu-1} \vdash \gamma$. But then we have $(\Delta\psi \to \psi)^\circ$ by Fact 0.1, for ψ° is a Π_n- and hence a $\Pi_{\bar\mu+1}$-sentence and provability in $I\Sigma_{\bar\mu-1}$ testifies to the truth of such sentences.

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

Thus in either case $\big(\Delta((\Delta\psi \to \psi) \vee \Delta\varphi)\big)^\circ$ q.e.d. ∎

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 Δ-*gödelsentences*, i.e. fixed points of the form $\gamma \leftrightarrow \neg\Delta\gamma$ for the Feferman predicate based on $(I\Sigma_n)_{n\in\omega}$. Relying on Proposition 1.5 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.

1.6. PROPOSITION.

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

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

(1) $\quad \mathbf{LF} \vdash \Box(\neg q \to \triangle q) \to \Box\Big(\neg q \to \triangle\big((\triangle p \to p) \lor \triangle q\big)\Big)$ $\qquad$ (by (S))

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

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

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

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

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (by (2) and (3))

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

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

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

1.7. EXERCISE. Show that all $\triangle$-gödelsentences are provably in PA equivalent to the sentence

$$\forall x\, (I\Sigma_x \vdash `I\Sigma_x \vdash \bot{}' \to I\Sigma_x \vdash \bot) \qquad\qquad\qquad\qquad ∎$$

We shall return to the subject of gödelsentences in Section 5.

2. Models for LF

2.1. DEFINITION. A *marshmallow frame* is a tuple $\boldsymbol{W} = (W, R, S)$ 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 $x\,S\,y$ and $x\,R\,z$ then $x\,R\,y$,

 (vi)   If $x\,S\,y\,S\,z$ then $x\,S\,z$ or $y\,S\,y$,

(vii)   For each x there is an y s.t. $x\,S\,y$.

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

$$T_W = \{\, a \in W \mid a \, R \, b \text{ for no } b \in W \,\}$$
$$D_W = \{\, a \in W \mid a \, S \, a \,\}$$

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

2.2. LEMMA. (a) *If $x \in W - T$ then $(x \, R \, y$ iff $x \, S \, y)$.*

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

(c) $D \subseteq T$.

(d) *For all $x \in W$ there is an $y \in W$ s.t. $x \, S \, y \in D$.*

PROOF. (a) follows from (ii) and (v) of Definition 2.1.

(b) Immediate from (iv) of Definition 2.1.

(c) follows from (a) and (i) of Definition 2.1.

(d) Consider an $x \in W$. Suppose that $y \in D$ held for no $y \in W$ s.t. $x \, S \, y$. Then by iterating (vii) of Definition 2.1 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 $x \, S \, x_i \, S \, x_j$ for all $i < j$. On the other hand, W is finite so we have $x \, S \, x_i = x_j$ for some $i < j$ which is a contradiction for one then has $x \, S \, x_i \in D$. ∎

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

2.4. PROPOSITION. *For any $\square\triangle$-formula φ and any marshmallow model $\mathcal{W}$, if $\mathbf{LF} \vdash \varphi$ then $\mathcal{W} \Vdash \varphi$.* ∎

Our aim is to reverse the implication of Proposition 2.4. 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.

2.5. DEFINITION. A finite set α of $\square\triangle$-formulas is *adequate* if

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

(ii) α is closed under subformulas,

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

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

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

8

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

2.6. DEFINITION. Let α be adequate and define W_α to be the set of all sets w of $\square\triangle$-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 2.6 is finite, for if $w \in W_\alpha$ then 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,

2.7. LEMMA. *Each set of $\square\triangle$-formulas satisfying (ii) and (iii) of Definition 2.6 has a superset in W_α.*

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

2.8. DEFINITION. Let α be adequate. We define relations R_α and S_α on W_α. Put

$$v \, R_\alpha \, w \quad \text{iff} \quad (\text{for any formula } \varphi, \text{ if } \square\varphi \in v \text{ then } \varphi, \triangle\varphi, \triangle\triangle\varphi, \ldots \in w)$$

$$\text{and there exists a formula } \square\psi \in w - v$$

Recall that in marshmallow frames the set T of R-topmost elements is defined exclusively in terms of the relation R (cf. Definition 2.1). 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

$$v \, S_\alpha \, w \quad \text{iff} \quad v \, R_\alpha \, w \quad \text{or}$$

$$\big(v, w \in T, \, (\varphi \in w \text{ whenever } \triangle\varphi \in v)$$

$$\text{and (for any } \psi, \chi, \text{ if } \triangle\psi \in v \cap \alpha \text{ and } \triangle\chi \in w \cap \alpha$$

$$\text{then } \triangle\psi \in w \text{ or } \chi \in w)\big)$$

2.9. LEMMA. *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 2.1 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

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

Clause (vi) comes first. Suppose $u \, S_\alpha \, v \, S_\alpha \, w$ and not $v \, S_\alpha \, v$. 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 $v\ S_\alpha\ v$ means that we can fix a formula $\triangle\psi \in v \cap \alpha$ s.t. $\psi \notin v$. Let us check $u\ S_\alpha\ w$. Suppose $\triangle\varphi \in u$ so that by $u\ S_\alpha\ v$ 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 $v\ S_\alpha\ w$, 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 $\chi \in w$ q.e.d.

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

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

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

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

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

$$\vdash \square\bot \wedge \bigwedge_{\triangle\varphi \in d} \triangle\varphi \to \bot \quad \text{(by Proposition 1.3)}$$

which would deny v membership in W_α. Furthermore, we show that w_0 satisfies (ii) of Definition 2.6: 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 2.7 a set w with $w_0 \subseteq w \in W_\alpha$. Moreover, since $\square\bot \in w_0 \subseteq w$, 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 $v\ S_\alpha\ w$. ∎

2.10. DEFINITION. The marshmallow model $\mathcal{W}_\alpha = (\boldsymbol{W_\alpha}, \Vdash)$ is defined by putting

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

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

2.11. LEMMA. *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 $\square$ case. We turn to $\triangle$ under the assumption $w \in T$.

(if): Suppose $\triangle\varphi \in w \cap \alpha$. Then for all v with $w\ S_\alpha\ v$ 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

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

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

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

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

(by Δ-necessitation and (F3))

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

Thus $\Delta\varphi$ could not escape being in w contrary to assumptions. Therefore u_0 is consistent and, as in the proof of Lemma 2.9, it is seen that u_0 satisfies (ii) of Definition 2.6. Hence there exists an u with $u_0 \subseteq u \in W_\alpha$. We leave it to the reader to check $w\, S_\alpha\, u$ so that $w \not\Vdash \Delta\varphi$ because, by IH, $u \not\Vdash \varphi$. ∎

2.12. THEOREM. $\mathbf{LF} \vdash \varphi$ iff for any marshmallow model $\mathcal{W}$ one has $\mathcal{W} \Vdash \varphi$.

PROOF. (only if) is Proposition 2.4.

(if): Suppose $\mathbf{LF} \not\vdash \varphi$. Let then α be an adequate set containing $\neg\varphi$. By Proposition 2.7 there exists a $w \in W_\alpha$ with $\neg\varphi \in w$. Hence by Lemma 2.11 we have $w \not\Vdash \varphi$ and so $\mathcal{W}_\alpha \not\Vdash \varphi$ as required. ∎

2.13. EXERCISE (de Jongh). Consider the logic $\mathbf{F}$ 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 nonempty W and satisfying (vi) and (vii) of Definition 2.1. Prove $\mathbf{F}$ to be complete w.r.t. marshmallows.

(b) Show that $\mathbf{LF}$ is conservative over $\mathbf{F}$.

HINT. (a) Weed the proof of Theorem 2.12.

(b) Observe that any marshmallow can be represented as the set T of an appropriate marshmallow model. ∎

3. A Solovay function

In this Section we prove the arithmetical completeness theorem for $\mathbf{LF}$. As usual, we shall do so by constructing a suitable Solovay function (cf. Solovay [20], Smoryński [18, Chapter 3] 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 $\mathbf{W} = (W, R, S)$ which however will have to satisfy the following two conditions:

(i) $\mathbf{W}$ has a bottom node, that is there is a node $0 \in W$ s.t. $0\, R\, a$ 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 $F : \omega \rightarrow W$ proceeds parallel to that of an auxiliary function $G : \omega \rightarrow \omega \cup \{\infty\}$ which stores some information relevant for F's locomotion. The two bear a close relationship to the pair appearing in Visser [23, section 10] 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.

3.1. DEFINITION ($I\Sigma_1$). Define primitive recursive functions F and G by simultaneous recursion:

$$F(0) = 0; \quad G(0) = \infty$$

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

Case A. $F(x) \, R \, a$ 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. $F(x) \, S \, b \notin D$, $F(x) \in T$ and x is an $I\Sigma_{G(x)}$-*proof of* $L \neq b$.

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

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

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

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

$$F(x + 1) = F(x); \quad G(x + 1) = G(x)$$

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

The mute point of this definition is what happens if $G(x) = 0$ and instructions of Case C tell G to decrease still lower. One way to get around this is to stipulate that $I\Sigma_{-1}$ is a theory that proves no theorems whatsoever. This however would unnecessarily overcomplicate our definition. Another way is to leave the matter up in the air but to take account of the fact, to be shortly seen, that for every x the theory $I\Sigma_x$ proves that G can not decrease below $x - 1$. In this case some of the lemmas stated below

will require $I\Sigma_2$ instead of the indicated $I\Sigma_1$ to formalize, but this does not affect the validity of our results.

3.2. LEMMA (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 2.2(b) it can not go back. Moreover, one can, just as in Lemma 2.2(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. ∎

3.3. DEFINITION ($I\Sigma_1$). 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.

3.4. LEMMA ($I\Sigma_1$). (a) *If $L = a \neq 0$ then* PA $\vdash \bigvee_{aRb} L = b$.

(b) *If $L = a$ then for any b with $a\,R\,b$* PA *is consistent with $L = b$.*

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 L should by Fact 0.2 be actually unequal to b.

Thus PA is indeed inconsistent.

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

3.5. LEMMA ($I\Sigma_1$). *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 3.4 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 2.1 there is a b s.t. $L\,S\,b$. 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 $I\Sigma_{\bar\mu}$-proof. But then, since $L \neq a$ is a Π_2-sentence, L can not, by Fact 0.1, be equal to a. A contradiction.

$I\Sigma_{\mu+1}$ is therefore inconsistent q.e.d. ■

3.6. LEMMA ($I\Sigma_1$). *Suppose $L = a \in T$.*

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

(b) *If $a\ S\ b$ 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 $a\ S\ c$. 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 3.1 involves a decrease in the value of G and therefore the proof that brings F to c is an $I\Sigma_{\bar\mu-1}$-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 3.5. ■

3.7. LEMMA ($I\Sigma_1$). (a) *If $L = a \neq 0$ then $\triangle \bigvee_{aSb} L = b$.*

(b) *If $L = a\ S\ b$ then $\triangledown L = b$.*

PROOF. For $L \notin T$ this reduces to Lemma 3.4. For $L \in T$ use Lemmas 3.5–6. ■

3.8. LEMMA. $L = 0$ *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 3.4. ■

3.9. THEOREM. $\mathbf{LF} \vdash \varphi$ *iff for any gf-interpretation $^\circ$,* PA $\vdash \varphi^\circ$.

PROOF. We have already proved (only if) in Proposition 1.5.

(if) is proved in the standard manner (cf. Solovay [20] or Smoryński [18]): Suppose $\mathbf{LF} \nvdash \varphi$. Then by Theorem 2.12 we have a marshmallow model $\mathcal{W} = (W, R, S, \Vdash)$ with $\mathcal{W} \nVdash \varphi$. Append a new bottom 0 to $\mathcal{W}$ and apply Definition 3.1 to the result. Define a gf-interpretation $^\circ$ by putting

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

14

'for propositional variables p occurring in φ. Use Lemmas 3.4 and 3.7 to show by induction on the structure of φ that

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

so that if $W \ni a \not\Vdash \varphi$ then $\text{PA} \vdash \varphi^\circ$ would imply $\text{PA} \vdash L \neq a$ contradicting Lemma 3.8. Conclude $\text{PA} \nvdash \varphi^\circ$. ∎

Theorem 3.9 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 $^\circ$ rendering a $\Box\triangle$-formula φ provable by ensuring that $(\Box\varphi)^\circ$ is true.

3.10. DEFINITION. For φ a $\Box\triangle$-formula define the formula $R(\varphi)$ to be

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

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

COMMENT. This Proposition enjoys a proof similar to that of Theorem 3.9 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 $\text{PA} \vdash L \neq 0 \to (L \Vdash \Box\varphi \to \Box L \Vdash \varphi)$ of Theorem 3.9. (Consult Solovay [20] or Smoryński [18, Chapter 3] for a similar situation.) ∎

3.12. REMARK. Theorem 3.9 and Proposition 3.11 can be adapted to the logic **F** on $\Box$-free formulas and marshmallows $\mathbf{W} = (W, S)$ of Exercise 2.13. 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. $0 \, S \, a$ 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.

For a centered marshmallow with a center 0 the construction of the Solovay function can be visualized as follows: Make the marshmallow into a marshmallow frame by adjoining a new R-bottom 0_R below the whole of the marshmallow. After that apply Definition 3.1 to the resulting frame and then glue 0_R to 0.

The nice point about this construction is that here we get more satisfactory commutation in that one can prove $\text{I}\Sigma_1 \vdash L = 0 \to \nabla L = 0$ and hence drop the syntactical precondition that 0 forces a certain formula in the analogue of Proposition 3.11.

3.13. EXERCISE. Consider logics $\mathbf{LF}^\omega$ and $\mathbf{F}^\omega$, where $\mathbf{LF}^\omega$ is obtained from $\mathbf{LF}$ by adding the schema $\Box\varphi \to \varphi$ and the logic $\mathbf{F}^\omega$ results from $\mathbf{F}$ 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.

15

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

HINT. (a) Show that $\mathbf{F}^\omega$ is complete w.r.t. centers of centered marshmallows.

(b) Use Proposition 3.11.   ∎

The last exploit in this Section of the Solovay function F introduced in Definition 3.1 is to illustrate that the validity of the axiom (S) w.r.t. gf-interpretations might actually fail under a choice of $(\text{PA}\lceil n)_{n\in\omega}$, the base sequence for $\triangle$, different from $(I\Sigma_n)_{n\in\omega}$. Let $\operatorname{con} T$ denote the sentence expressing the consistency of a finitely axiomatized theory T and consider the following sequence:

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

We write $\triangle^*$ for the Feferman predicate based on this sequence of theories. 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 3.1.

3.14. LEMMA ($I\Sigma_1$). *Suppose* $L \in T - D$.

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

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

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

PROOF. (a) Let $L = a \notin D$ and reason in $I\Sigma_\mu$:
If $L = b \in D$ then G must have decreased its value when getting to b and therefore $\mu < \bar\mu$. By Lemma 3.5 the theory $I\Sigma_{\bar\mu}$ is then inconsistent.
Conversely, if $L \notin D$ then either Case C takes place while F travels from a to L in which case $I\Sigma_{\bar\mu-1} \vdash L \neq \bar L$ which, by reflection, is absurd, or Case C does not occur during this period implying $\mu = \bar\mu$. By Lemma 3.5 it follows that the theory $I\Sigma_{\bar\mu}$ is consistent.

(b) By Lemma 3.6(b) $I\Sigma_\mu$ is consistent with $L = b$. By (a) of the present Lemma it therefore has to be consistent with $\operatorname{con} I\Sigma_{\bar\mu}$.

(c) Immediate from Lemma 3.6(a) and clause (a) of the present Lemma.   ∎

3.15. PROPOSITION. *There exists an arithmetical sentence γ s.t.*

$$\text{PA} \not\vdash \nabla^*(\triangle^*\gamma \to \gamma)$$

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

$$
\begin{array}{c}
2 \\
1 \;\; S \longrightarrow \bigcirc \;\; S \quad 3 \\
\bigcirc \xleftarrow{\;S\;} \longrightarrow \bigcirc \;\; S \\
R,S \quad R,S \quad R,S \\
\bigcirc \\
0
\end{array}
$$
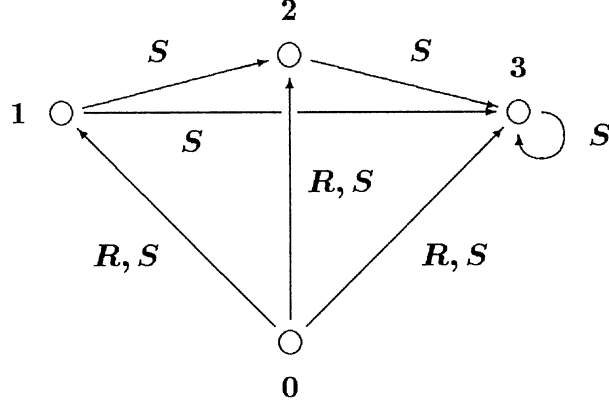
Observe:

$$I\Sigma_1 \vdash L = 2 \to. \; L \neq 3 \wedge (I\Sigma_\mu \vdash L = 3) \qquad \text{(by Lemma 3.6(a))}$$

$$\to. \; L \neq 3 \wedge \Delta^* L = 3 \qquad \text{(by Lemma 3.5)}$$

$$I\Sigma_1 \vdash L = 1 \to. \; \mathrm{con}\,(I\Sigma_\mu + \mathrm{con}\,I\Sigma_{\bar\mu}) \wedge (I\Sigma_\mu + \mathrm{con}\,I\Sigma_{\bar\mu} \vdash L = 2)$$

$$\text{(by Lemma 3.14(b) and (c))}$$

$$\to \Delta^* L = 2 \quad .$$

This combines to give

$$I\Sigma_1 \vdash L = 1 \to \Delta^*(L \neq 3 \wedge \Delta^* L = 3)$$

and therefore by Lemma 3.8

$$\mathrm{PA} \not\vdash \nabla^*(\Delta^* L = 3 \to L = 3)$$

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

4. Some fixed points

Many qualitative and quantitative aspects of fixed points of arithmetical formulae corresponding to modal $\Box\Delta$-formulae under interpretations similar to our gf-interpretation are discussed in Visser [22]. In this Section we first observe that Theorem 3.11 affords a corollary classifying the quantitative behaviour of fixed points of gf-interpretations of $\Box\Delta$-formulae.

4.1. PROPOSITION. *Let $\varphi(p)$ be a $\Box\Delta$-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*

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

is PA-provably unique, i.e.

$$\text{PA} \vdash \square\big(\gamma_1 \leftrightarrow \varphi^\circ(\gamma_1)\big) \wedge \square\big(\gamma_2 \leftrightarrow \varphi^\circ(\gamma_2)\big) \rightarrow. \square(\gamma_1 \leftrightarrow \gamma_2)$$

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 3.9 we have

$$\mathbf{LF} \not\vdash \square\big(p \leftrightarrow \varphi(p)\big) \wedge \square\big(q \leftrightarrow \varphi(q)\big) \rightarrow. \square(p \leftrightarrow q)$$

Therefore by Theorem 2.12 there exists a marshmallow model $\mathcal{W} = (W, R, S, \Vdash)$ whose bottom 0 forces $\square\big(p \leftrightarrow \varphi(p)\big) \wedge \square\big(q \leftrightarrow \varphi(q)\big)$ but not $\square(p \leftrightarrow q)$. 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 (cf. Exercise 2.13) with a forcing relation $\mathcal{T} = (T, S{\restriction}T, \Vdash{\restriction}T)$ (here $\restriction$ stands for restriction) and construct a marshmallow model $\mathcal{V}_N$ which consists of two copies of $\mathcal{T}$ 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 $\mathcal{T}$ 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 $\mathcal{V}_N$. An easy application of the pigeon-hole principle shows that we can choose the value of N so that

$$R\Big(\square\big(p \leftrightarrow \varphi(p)\big) \wedge \square\big(q \leftrightarrow \varphi(q)\big) \wedge \square\big(r \leftrightarrow \varphi(r)\big) \rightarrow.$$

$$\rightarrow. \square(p \leftrightarrow q) \vee \square(p \leftrightarrow r) \vee \square(q \leftrightarrow r)\Big)$$

is forced at the bottom of $\mathcal{V}_N$. We then apply Proposition 3.11 to obtain a gf-interpretation $^\circ$ under which the formulas

$$\square\big(p \leftrightarrow \varphi(p)\big), \quad \square\big(q \leftrightarrow \varphi(q)\big) \quad \text{and} \quad \square\big(r \leftrightarrow \varphi(r)\big)$$

are true and the formulas

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

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 $\square\triangle$-formula, namely sentences γ satisfying

$$\text{PA} \vdash \gamma \leftrightarrow \triangle\gamma$$

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 4.1. Note that both sentences $\bot$ and $\top$ are Σ_1. In fact, any $\triangle$-henkinsentence produced by a direct application of our Proposition 3.11 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 $\text{PA} + \gamma$ and $\text{PA} + \neg\gamma$ 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

4.2. FACT (Orey [15], Hájek [5] and Guaspari [3]; see Berarducci [1] for a concise presentation). *For arithmetical sentences γ and δ the following are equivalent:*

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

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

  (iii) *Every model of* $\text{PA} + \delta$ *has an endextension modelling* $\text{PA} + \gamma$.　　　　■

(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).

4.3. PROPOSITION. *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 3.9, and resorting to this method to construct just one individual sentence almost amounts to a breach of the code of honour. Nevertheless, I believe that this format might help the reader to better visualize the proof thus enhancing the digestability of the argument.

The scene is set by the following Kripke frame:
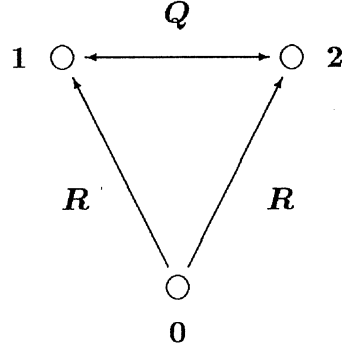
The new name Q for the auxiliary relation suggests that it should be treated in a way different from the one the relation S of marshmallow frames was treated in in Section 3.

We define the two relevant functions:

$$f(0) = 0; \quad g(0) = \infty$$

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

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

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

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

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

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

Case D. *None of the previous cases apply.*

$$f(x + 1) = f(x); \quad g(x + 1) = g(x)$$

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 of 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 g and f reach their limit values. Call the ε-terms naming these limits M and ℓ 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 $I\Sigma_{\bar{M}-1}$ of $\ell \neq a$ which can not exist since $\ell = a$. Hence $\ell = 1$.

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

20

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 $I\Sigma_{\bar{M}} \vdash \ell \neq \bar{\ell}$.

So,

(1) $I\Sigma_1 \vdash \ell = 1 \rightarrow \Delta\ell = 1$ (by Claims 1 and 2)

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

(3) $I\Sigma_1 \vdash \Delta\ell = 1 \rightarrow \Delta\ell \neq 2$

$\quad\quad\quad\quad \rightarrow \Box\ell \neq 2$

$\quad\quad\quad\quad \rightarrow \ell \neq 0$

$\quad\quad\quad\quad \rightarrow. \ell = 1 \vee \ell = 2$

$\quad\quad\quad\quad \rightarrow \ell = 1$ (by (2) and (3))

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

> By Fact 0.2 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

(4) PA $\vdash \ell = 1 \rightarrow \mathrm{con}\,(I\Sigma_n + \ell = 2)$

$\quad\quad\quad\quad \rightarrow \mathrm{con}\,(I\Sigma_n + \ell \neq 1)$

On the other hand,

$\quad$ PA $\vdash \ell \neq 1 \rightarrow. \ell = 0 \vee \ell = 2$

$\quad\quad\quad\quad \rightarrow. \mathrm{con}\,(\mathrm{PA} + \ell = 1) \vee \mathrm{con}\,(I\Sigma_n + \ell = 1)$ (symmetric to (4))

$\quad\quad\quad\quad \rightarrow \mathrm{con}\,(I\Sigma_n + \ell = 1)$

Thus by (ii) $\Rightarrow$ (i) of Fact 4.2, $\ell = 1$ is an oreysentence and satisfies the statement of the Proposition ($\ell = 2$ would have done just as well). $\quad\quad\blacksquare$

4.4. REMARK. Proposition 4.3 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 PA$\ulcorner g(x)$ can reflect on the proof x.

5. More gödelsentences

In this Section we shall produce one more example highlighting the sensitivity of the provability-logical behaviour of the Feferman predicate to the choice of the base sequence of theories. With Proposition 3.15 we have already experienced the fragility of the axiom schema (S) and Proposition 1.3 in this respect.

Here we assault Proposition 1.6 which asserted the provable uniqueness of Δ-gödel-sentences for Feferman predicates satisfying (S). In other words, we set out to produce a sequence of theories whose Feferman predicate possesses inequivalent gödelsenten-ces. The uniqueness question for gödelsentences of Feferman predicates was raised by Montagna [11]. Smoryński [19] discovered uniqueness of gödelsentences of Δ'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 & Solovay [4] show that provable uniqueness of gödelsentences of this predicates (=rossersentences) depends on otherwise insignificant details of gödelnumbering of proofs.

Oddly enough, our main tool for this task are the Solovay function F together with its accomplice G launched in Definition 3.1 that were originally called to life to prove the completeness theorem for the provability logic of a Δ with provably unique gödelsentences.

Consider the predicate Δ' whose base sequence is $\big(I\Sigma_{n+1}(I\Sigma_n)\big)_{n\in\omega}$, where the theory $I\Sigma_{n+1}(I\Sigma_n)$ is 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 Δ' with the construction of Definition 3.1 and 3.3 and in what follows we adopt some of the notation from Section 3.

5.1. LEMMA $(I\Sigma_1)$. *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 $\Delta'\gamma$ iff $I\Sigma_\mu(I\Sigma_{\bar\mu-1}) \vdash \gamma$.*

PROOF. By Lemma 3.5 $I\Sigma_{\mu+1}$ is inconsistent and hence so is $I\Sigma_{\mu+1}(I\Sigma_{\bar\mu})$. In the other direction, if $I\Sigma_\mu \vdash \text{`}I\Sigma_{\bar\mu-1} \vdash \perp\text{'}$, then one by reflection has $I\Sigma_\mu \vdash \perp$ which would contradict Lemma 3.5. ∎

5.2. LEMMA $(I\Sigma_2)$. *Suppose $L = a \in T$.*

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

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

(c) *If $a\,S^2\,b$ 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 which we shall call c.
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 $I\Sigma_{\bar\mu-2}$ reflecting whereupon we get $L \neq b$.
This way we have shown $I\Sigma_\mu \vdash \text{`}I\Sigma_{\bar\mu-1} \vdash L \neq b\text{'}$ which proves the claim.

(b) If for a node b s.t. $a\ S\ b$ 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 3.6(b).

(c) Let c be a node s.t. $a\ S\ c\ S\ b$. 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 Fact 0.1 the theory $I\Sigma_{\bar\mu-1}$ is consistent. Therefore by Lemma 3.5 $\mu \geq \bar\mu - 1$ and so $I\Sigma_\mu \vdash L \neq b$. By Lemma 3.6(b) we can not then have $L = c$. But $I\Sigma_\mu \vdash L \neq c$ contradicts Lemma 3.6(b) q.e.d. ∎

5.3. LEMMA ($I\Sigma_2$). (a) If $l = a \neq 0$ then $\Delta' \bigvee_{a(S \cup S^2)b} L = b$.

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

PROOF. If $L \notin T$ then PA is consistent and Δ' is the same as $\square$. Lemma 3.4 settles this case. For $L \in T$ use Lemmas 5.1–2. ∎

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

PROOF. Consider the predicate Δ' studied in Lemmas 5.1–3. Let us apply Lemma 5.3 to the marshmallow frame obtained from the marshmallow depicted below by adjoining an R-bottom 0.
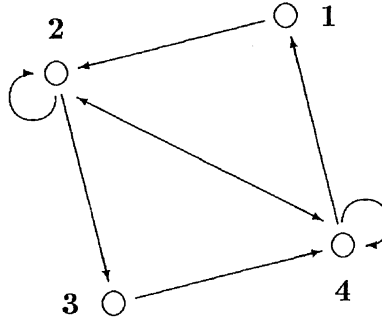
An easy verification using that lemma shows that

(1) $I\Sigma_2 \vdash L = 1 \leftrightarrow \neg\Delta' L = 1$

and, in perfect symmetry,

(2) $I\Sigma_2 \vdash L = 3 \leftrightarrow \neg\Delta' L = 3$

So $L = 1$ and $L = 3$ are Δ'-gödelsentences that are clearly inequivalent.

We would however be cheating were we to declare the sequence $\big(I\Sigma_{n+1}(I\Sigma_n)\big)_{n\in\omega}$ and the predicate Δ' 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 $I\Sigma_{n+1}(I\Sigma_n)$ 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 $L = 1$ and $L = 3$ by γ_0 and γ_1 respectively. Consider the following fixed point equation on the formula $\xi(x)$:

$$I\Sigma_1 \vdash \forall x \left(\xi(x) \leftrightarrow \forall y \left(\bigwedge_{0 \le i \le 1} \left((I\Sigma_x \vdash_{\overline{y}} \xi(x) \to \gamma_i) \to \gamma_i \right) \to. \right. \right.$$

$$\left. \left. \to. \bigwedge_{0 \le i \le 1} \left((I\Sigma_{x+1} \vdash_{\overline{y}} {}^\backprime I\Sigma_x \vdash \gamma_i{}') \to \gamma_i \right) \right) \right)$$

where $T \vdash_{\overline{y}} \delta$ stands for the formula expressing that δ has a T-proof of gödelnumber $\le y$. One can mimick the proof of Lindström [9, Lemma 2, Case 1] to the effect that for $0 \le i \le 1$ 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 both i's we get

(3) $\quad I\Sigma_1 \vdash \forall x \left((I\Sigma_x \vdash \xi(x) \to \gamma_i) \leftrightarrow I\Sigma_{x+1} \vdash {}^\backprime I\Sigma_x \vdash \gamma_i{}' \right)$

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

$$I\Sigma_1 \vdash \forall x \left((I\Sigma_x \vdash \xi(x) \to \gamma_i) \to I\Sigma_{x+1} \vdash {}^\backprime I\Sigma_x \vdash \gamma_i{}' \right) \qquad \text{(by (3))}$$

$$\to I\Sigma_{x+1} \vdash \gamma_i) \qquad \text{(by reflection)}$$

$$\to I\Sigma_x \vdash {}^\backprime I\Sigma_{x+1} \vdash \gamma_i{}')$$

$$\to (I\Sigma_x \vdash I\Sigma_{x+1} \to \gamma_i))$$

(4) $\qquad\qquad \to (I\Sigma_x \vdash I\Sigma_{x+1} \vee \xi(x) \to. \gamma_i))$

The other direction $I\Sigma_1 \vdash \forall x \left((I\Sigma_x \vdash I\Sigma_{x+1} \vee \xi(x) \to. \gamma_i) \to (I\Sigma_x \vdash \xi(x) \to \gamma_i) \right)$ is immediate.

We now claim that the sequence of theories $\left(I\Sigma_n + (I\Sigma_{n+1} \vee \xi(n)) \right)_{n \in \omega}$ together with the Feferman predicate $\tilde{\Delta}$ 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 note that since $I\Sigma_0 \vdash \neg(\gamma_0 \wedge \gamma_1)$, we have

$$I\Sigma_1 \vdash \forall x \left(\neg con \left(I\Sigma_x + (I\Sigma_{x+1} \vee \xi(x)) \right) \leftrightarrow I\Sigma_x + (I\Sigma_{x+1} \vee \xi(x)) \vdash \gamma_0 \wedge \gamma_1 \right)$$

$$\leftrightarrow I\Sigma_{x+1}(I\Sigma_x) \vdash \gamma_0 \wedge \gamma_1 \right) \qquad \text{(by (3))}$$

(5) $\qquad\qquad\qquad \leftrightarrow \neg con\, I\Sigma_{x+1}(I\Sigma_x) \right)$

so that $I\Sigma_{n+1}(I\Sigma_n)$ and $I\Sigma_n + (I\Sigma_{n+1} \vee \xi(n))$ are provably equiconsistent.

24

Finally, we demonstrate that the γ_i's we have constructed are $\tilde{\Delta}$-gödelsentences as well as Δ'- ones:

$$I\Sigma_2 \vdash \neg\gamma_i \leftrightarrow \Delta'\gamma_i \qquad \text{(by (1) and (2))}$$

$$\leftrightarrow \exists x \left(\mathrm{con}\left(I\Sigma_{x+1}(I\Sigma_x) \right) \wedge \left(I\Sigma_{x+1}(I\Sigma_x) \vdash \gamma_i \right) \right)$$

$$\leftrightarrow \exists x \left(\mathrm{con}\left(I\Sigma_x + (I\Sigma_{x+1} \vee \xi(x)) \right) \wedge \left(I\Sigma_x + (I\Sigma_{x+1} \vee \xi(x)) \vdash \gamma_i \right) \right)$$

$$\text{(by (3), (4) and (5))}$$

$$\leftrightarrow \tilde{\Delta}\gamma_i$$

Thus γ_0 and γ_1 are inequivalent $\tilde{\Delta}$-gödelsentences q.e.d. ∎

Another fact of a distinctly similar flavour that contrasts Proposition 1.6 is the following

5.5. EXERCISE (Visser). Show that there are infinitely many inequivalent sentences γ satisfying

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

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

References

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

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

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

[4] D. Guaspari & R. M. Solovay. Rosser sentences. *Annals of Mathematical Logic* 16 (1979) 81–99.

[5] P. Hájek. On interpretability in set theories. *Commentationes Mathematicae Universitatis Carolinae* 12 (1971) 73–79.

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

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

[8] P. Lindström. Some results on interpretability. In: *Proceedings from 5th Scandinavian Logic Symposium* (F.V.Jensen, B.H.Mayoh & K.K.Møller, eds.) Aalborg University Press, Aalborg 1979, 329–361.

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

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

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

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

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

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

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

[16] J. B. Paris & L. A. S. Kirby. Σ_n-collection schemas in arithmetic. In: *Logic Colloquium '77* (A.Macintyre, L.Pacholski & J.Paris, eds.) North-Holland, Amsterdam 1978, 199–209.

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

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

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

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

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

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

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

The ILLC Prepublication Series