Peano's Smart Children

A provability logical study of systems with built-in consistency

Albert Visser

Department of philosophy
University of Utrecht

Logic Group
Preprint Series
No.14

ILLUSTRA NOS SOL IUSTITIAE

Department of Philosophy
University of Utrecht

Peano's Smart Children

A provability logical study of systems with built-in consistency

Albert Visser

Department of Philosophy
University of Utrecht

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

Peano's Smart Children

A provability logical study of systems with built-in consistency

Albert Visser

1 Introduction

Consistency can be built into a system in various ways. The two best known constructions are Rosser's and Feferman's. Both construction's take a given formal system in the usual sense as initial data. Consider for example Peano Arithmetic (PA). A proof in the Peano System will count as a proof in the Rosser System based on PA, if there is no shorter Peano proof of the negation of its conclusion. The Feferman System can be described in various interesting ways -modulo provable equivalence in PA of the formulas defining the set of theorems. One such way is: a proof in the Peano System will count as a proof in the Feferman System based on PA, if the finite set of arithmetical Peano axioms smaller than or equal to the largest arithmetical Peano axiom used in the proof is consistent.

The reasons such constructions occur in the literature are various:

i) They serve as counterexamples in the study of the relation between Gödel's first and second Incompleteness Theorem (see Feferman [1960]).
ii) They serve as didactical examples in philosophical discussions, like the debate on intensionality in Mathematics (see e.g. Auerbach[1985]) and the discussion on the possible bearing of the Incompleteness Theorems on the Minds & Machines problem (see e.g. Lucas[1961], Webb[1980], Bowie[1982]).
iii) Rosser's construction is used to sharpen Gödel's first Incompleteness Theorem.
iv) Feferman's construction is an important tool in the study of Relative Interpretability (see Feferman[1960], Orey[1961]).

The main objects of study in the present paper are certain variants of both Rosser's an Feferman's construction. My motivations are closely related to (i)-(iv) above:

a) There is much interest in the study of bimodal systems in the current literature on Provability Logic (see e.g. Montagna[1984] and Smoryński[1985]). There are two directions of research: first there is the pure study of arithmetical selfreference, secondly there is the study of arithmetical selfreference as a tool to unify selfreferential arguments in Arithmetic (see Smoryński[1985], chapter 7). In the first line of study one aims at characterizing the modal logic for a certain 'given' class of interpretations. There is no objection here to have 'few' interpretations and strong modal systems. In the second line one looks primarily for a modal system which is sound for as many interpretations as possible, but is still rich enough to carry out the proofs of the arithmetical arguments under study. The distinction between the two lines described here is not precisely that between pure and applied. The first line also has its typical applications: Solovay-style completeness results yield a powerful machinery to produce arithmetical sentences with rich but controlled properties; these sentences can be used to prove various incompleteness and other results (for an example, see §9 of this paper).

The contribution of this paper is to the first line. I provide an example of a rich modal logic of not too standard sort, valid for two different arithmetical interpretations. This example can be used to test conjectures concerning the conditions for uniqueness and explicit definability of fixed points (see Smoryński[198?] for a discussion of these matters) . Questions of uniqueness and explicit definability generalize the problem of the precise connection between the first and second Incompleteness Theorem; in this sense (a) generalizes (i). Also the logic can be used to illustrate the point that one can simulate the results of intensional selfreference (like e.g. Rosser's Theorem) to quite an extent by applying provably extensional selfreference -the cost being an increase in the complexity (modulo provable equivalence) of the sentences involved.

b) The modal derivability conditions are an improvement in the presentation of systems with built-in consistency in the discussions mentioned under (ii) above.
c) The methods developped have as spin-off an application to Relative Interpretability: I answer a question of Per Lindström for the case of PA.

2 Prerequisites

Knowledge of Feferman[1960] & Smoryński[1985] should bring the reader a long way.

3 Acknowledgements

I thank Johan van Benthem, Dirk van Dalen, Karst Koymans, Henryk Kotlarski and Fer-Jan de Vries for stimulating discussions. Erik Krabbe carefully read parts of an early draft of this paper. I am grateful. I especially thank George Kreisel without whose interest and questions the paper probably never would have seen the light of day.

4 Contents of the paper

In §5 the necessary notions and notations are introduced. §6 is a step by step introduction to the construction of the systems that are central in this paper. This § also illustrates the powers of provably extensional selfreference and contains a discussion of the problem of uniqueness and explicitness of the Gödel- and Henkinsenteces of the various systems considered. §7 treats the bimodal principles valid for the two central systems; of course a Kripke model completeness theorem is proved. §8 has a partial result on embedding Kripke models for our modal system into Arithmetic. In §9 this embedding result is applied to a problem about Relative Interpretability.

5 Conventions, notions & elementary facts

5.1 Point

All the arithmetical results in the next sections will be stated for Peano Arithmetic. Of course PA is just a convenient peg to hang the discussion on: mostly any RE theory into which PA, minus induction, plus Σ_2-induction, can be interpreted would do. Where results on Relative Interpretability appear one must also demand that the theories considered are essentially reflexive.

5.2 $\Box$ and Δ (in different contexts)

Let Proof(x,y) be the Δ_0 arithmetical formula representing the relation: x is the Gödelnumber of a PA-proof of the formula with Gödelnumber y. We assume for convenience that: $PA \vdash \forall x \exists! y\ Proof(x,y)$. Let $Prov(y) := \exists x Proof(x,y)$.

We write par abus de langage 'Proof(x, $A(x_1,...,x_n)$)' for: $Proof(x, \ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner)$, here:

i) all free variables of A are among those shown.
ii) $\ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner$ is the "Gödelterm" for $A(x_1,...,x_n)$ as defined in Smoryński [1985], p43.

The modal operators $\Box$ and Δ will appear both in the context of modal logic and in the context of arithmetic. '$\Box A(x_1,...,x_n)$' will stand for : $Prov(\ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner)$. In arithmetical contexts '$\Delta A(x_1,...x_n)$' will stand for: $B(\ulcorner A(\dot{x}_1,...,\dot{x}_n)\urcorner)$, where B(x) is the arithmethization of theoremhood in the particular system with built-in consistency that we are considering at the place of occurrence of '$\Delta A(x_1,...,x_n)$'. In case confusion is possible we will use: Δ^R, Δ^K, etc. . To differentiate arithmetical from modal contexts we use: A, B, ... for arithmetical formulas and : ϕ, ψ, ... for modal propositional formulas.

If t is a term for a provably recursive function we will have (supposing that t is substitutable for x in A): $PA \vdash (\Box A(x))[t/x] \leftrightarrow \Box A(t)$. We will only employ terms for provably recursive functions, so we may indeed treat $x_1,...,x_n$ in $\Box A(x_1,...,x_n)$ simply as free variables. Similarly for Δ.

'$\Diamond$' will stand for: $\neg\Box\neg$, and '∇' for: $\neg\Delta\neg$.

When we want to consider systems with other axiom sets than PA, we write: Proof_{α}, Prov_{α}, $\Box_{\alpha}$, etc. , where α is a formula that represents the axiom set of the system under consideration in an intensionally correct way in PA. We fix a formula π correctly representing the axiom set of PA. Thus our notation '$\Box$' is just short for: $\Box_{\pi}$.

5.3 $\Box\restriction x$ and $\Box^*$

Define: $\pi\restriction x(y) \;:\Leftrightarrow\; \pi(y) \wedge y \leq x$,
$\Box\restriction x\ A \;:\Leftrightarrow\; \Box_{\pi\restriction x} A$
$\Diamond\restriction x\ A \;:\Leftrightarrow\; \neg\Box\restriction x \neg A$,
$\Box^* A \;:\Leftrightarrow\; \exists x\ \Box\restriction x\ A$.

Of course $\mathrm{PA} \vdash \Box A \leftrightarrow \Box^* A$, but the difference in form will be of some importance when Rosser-orderings come into play. (The usefulness of $\Box^*$ in this connection was discovered by Švejdar, see Švejdar[1983].)

5.4 Witnessing and the Rosser-ordering

Let A be of the form $\exists x\ A_0(x)$. Define: $t \text{ wit } A \;:\Leftrightarrow\; A_0(t)$. Here we assume that bound variables in A_0 are renamed -if necessary- to make t substitutable for x in A_0.

Let A be of the form $\exists x\ A_0(x)$ and B of the form $\exists x\ B_0(x)$. The Rosser-orderings between A and B are defined as follows:
$A \leq B \;:\Leftrightarrow\; \exists x\ (\ A_0(x) \wedge \forall y{<}x\ \neg B_0(y)\ )$,
$A < B \;:\Leftrightarrow\; \exists x\ (\ A_0(x) \wedge \forall y{\leq}x\ \neg B_0(y)\ )$.
We will always apply witnessing and the Rosser-ordering to the precise forms in which the relevant arithmetical formulas are introduced.

In connection with the Feferman System we will consider formulas of the form $\Box^*C{<}\Box^*D$. These formulas are of the more general form $A{<}B$, where A is $\exists x \exists y\ A_0(x,y)$ with A_0 in Δ_0 and where B is $\exists x\exists y\ B_0(x,y)$ with B_0 in Δ_0. It is of some interest to know the complexity of such formulas $A{<}B$. Prima facie $A{<}B$ is Σ_2. We have:
$\mathrm{PA}\vdash \exists x(\ \exists y A_0(x,y) \wedge \forall z{\leq}x \forall u \neg B_0(z,u)\ ) \leftrightarrow \forall u \exists x(\ \exists y A_0(x,y) \wedge \forall z{\leq}x \neg B_0(z,u)\ )$.
The "$\rightarrow$" side is trivial, for the "$\leftarrow$" side reason in PA:

> Suppose $\forall u \exists x(\exists y A_0(x,y) \wedge \forall z{\leq}x \neg B_0(z,u))$. It follows that $\exists x \exists y A_0(x,y)$. Let x_0 be the smallest such x. Consider any u. Pick an x such that $\exists y A_0(x,y)$ and $\forall z{\leq}x \neg B_0(z,u)$. Clearly $x_0{\leq}x$ and hence $\forall z{\leq}x_0 \neg B_0(z,u)$. Conclude: $\exists y A_0(x_0,y) \wedge \forall z{\leq}x_0 \neg B_0(z,u)$. □

Both Švejdar (see Švejdar[1978]) and Lindström (see Lindström[1979]) show that in every degree of Relative Interpretability over PA there is a sentence of the form $A<B$ where A and B are as above. Thus every degree of Relative Interpretability contains a Δ_2 sentence.

5.5 Relative Interpretability

'$A \lhd B$' stands for the arithmetization of: PA+A is relatively interpretable in PA+B. By a result of Orey and Hájek: $PA \vdash A \lhd B \leftrightarrow \forall x \Box(B \rightarrow \Diamond \lceil x A)$. We list a number of principles valid for $\Box$ and $\lhd$:

I1 $PA \vdash \Box(B \rightarrow A) \rightarrow A \lhd B$

I2 $PA \vdash (A \lhd B \wedge B \lhd C) \rightarrow A \lhd C$

I3 $PA \vdash (A \lhd B \wedge A \lhd C) \rightarrow A \lhd (B \vee C)$

I4 $PA \vdash A \lhd B \rightarrow (\Diamond B \rightarrow \Diamond A)$

I5 $PA \vdash \Diamond A \lhd B \rightarrow \Box(B \rightarrow \Diamond A)$

I6 $PA \vdash A \lhd \Diamond A$

I7 $PA \vdash A \lhd B \rightarrow (A \wedge \Box C) \lhd (B \wedge \Box C)$

The principle I7 is new and is due to Franco Montagna. We prove I4, I5 and I7. First we treat I4 and I5. Given I1 and I2 it is easily seen that I4 is equivalent to:

I4' $PA \vdash \bot \lhd B \rightarrow \Box \neg B$

Thus it is sufficient to prove:

J1 for all P in Π_1: $PA \vdash P \lhd B \rightarrow \Box(B \rightarrow P)$

First note that for every n: $PA \vdash \forall x \Box(B \rightarrow \Diamond \lceil x A) \leftrightarrow \forall x > \underline{n}\ \Box(B \rightarrow \Diamond \lceil x A)$. Pick q so big that $\Box \lceil q$ contains Robinson's Arithmetic. We have for S in Σ_1: $PA \vdash \forall x > \underline{q}\ \Box(\Box \lceil x S \leftrightarrow S)$, ergo for P in Π_1: $PA \vdash \forall x > \underline{q}\ \Box(\Diamond \lceil x P \leftrightarrow P)$. Hence:

$$PA \vdash P \lhd B \leftrightarrow \forall x > \underline{q}\ \Box(B \rightarrow \Diamond \lceil x P)$$
$$\leftrightarrow \Box(B \rightarrow P)$$

We turn to I7. We prove:

J2 for all S in Σ_1: $PA \vdash A \lhd B \rightarrow (A \wedge S) \lhd (B \wedge S)$

Suppose S is Σ_1. Let q be as above. Note that:

$$PA \vdash \forall x > \underline{q}\ \Box(\ S \rightarrow \Box \lceil x((D \wedge S) \leftrightarrow D)\ )$$

It follows that:

$$PA \vdash \forall x \Box(B \rightarrow \Diamond \lceil x A) \rightarrow \forall x > \underline{q} \Box(\ (B \wedge S) \rightarrow \Diamond \lceil x(A \wedge S)\ )$$
$$\rightarrow \forall x \Box(\ (B \wedge S) \rightarrow \Diamond \lceil x(A \wedge S)\ )$$

For further information see: Švejdar[1983].

5.6 On systems

Philosophically it is -I think- best to make the whole apparatus for generating theorems part of the identity conditions of systems. For our purposes it is however more convenient to confuse the systems considered with the arithmetical predicates that codify theoremhood in the system in an intensionally correct way. I will say that a system with associated arithmetical predicate A is a *variant* of a system with predicate B if $PA \vdash \forall x(A(x) \leftrightarrow B(x))$.

The notion of 'system' is kept more or less open in this paper. The usual formal systems are still paradigms of systemhood. The systems we consider here are in some sense derived from the usual systems:

they use the proofs of formal systems as data. A second point is that the systems considered can be seen to be extensionally equal to the formal systems on which they are based, given the information that the original systems are consistent.

6 Systems with built-in consistency, an introduction.

This section serves several purposes. First it exhibits various ways of 'loading' systems with desired 'modal' properties. Secondly it contains brief discussions of the various systems with built-in consistency that can be found in the literature. Thirdly the problems of uniqueness and of explicitness of Gödel- and Henkinsentences of the systems introduced are considered. (The rationale behind the attention for these specific problems is that these problems were historically at the crib of provability logic for the $\Box$, and secondly that these problems turn out to be a quite pleasant starting point when one wants to get acquainted with the systems studied here.) In the fourth place I give examples of the powers and possibilities of provably extensional selfreference. Specifically I show how to use provably extensional selfreference to construct four non-equivalent Oreysentences.

In this section '$\vdash$' stands for: PA$\vdash$. A,B,C stand for *formulas* of the language of PA. Note that by our conventions we have: $\vdash A(x) \Rightarrow \vdash \forall x A(x)$, but not: $\vdash \Box A(x) \rightarrow \Box \forall x A(x)$.

For the record we state here the usual principles valid in PA for $\Box$.

P The Peano System

The provability principles of PA are:

L1	$\vdash A \Rightarrow \vdash \Box A$
L2	$\vdash \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$
L3	$\vdash \Box A \rightarrow \Box\Box A$
L4	$\vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$

We will use these principles without explicit mention.

R The Rosser System

The Rosser System is defined as follows: $\Delta A :\Leftrightarrow \Box A < \Box\neg A$.

Some principles valid for the Rosser System in PA are:

1	$\vdash A \Rightarrow \vdash \Delta A$	
2	$\vdash \neg\Delta\bot$	
3	$\vdash \Delta A \rightarrow \Delta\Box A$	
4	$\vdash \neg\Box\bot \rightarrow (\Delta A \leftrightarrow \Box A)$	

Some direct consequences of 1-4 are:

5	$\vdash \Delta A \rightarrow \Box A$	(4)
6	$\vdash \Box A \rightarrow \Box\Delta A$	(3,4)

It is perhaps worth noting that the set of theorems of the Rosser

System is *provably* infinite. Reason in PA: in case $\neg\Box\bot$ this is trivial. In case $\Box\bot$ for any A clearly one of A, $\neg A$, $\neg\neg A$, $\neg\neg\neg A$, ... will be Rosser-provable.

In the Rosser System we have two explicit but non-unique Henkinsentences:

7 $\quad \vdash \top \leftrightarrow \Delta\top$ (1)

8 $\quad \vdash \bot \leftrightarrow \Delta\bot$ (2)

Consider a Gödelsentence of the Rosser System, i.e. a sentence G such that:

9 $\quad \vdash G \leftrightarrow \neg\Delta G$

We have:

10 $\quad \vdash \Box G \to (\Box\Delta G \wedge \Box\neg\Delta G) \to \Box\bot$ (6,9)

Of course we can also prove:

11 $\quad \vdash \Box\neg G \to \Box\bot$

But *not* from the modal principles collected up to now: we have to go back to the underlying Rosser-ordering. A slight change in the definition of Δ removes this defect as we will see under K.

Uniqueness or non-uniqueness of Gödelsentences in the Rosser System is still an open problem, Guaspari and Solovay show that if one allows *variants* of Prov in the definition of Δ, the answer can be yes and can be no (see Guaspari-Solovay[1979]).

I'm not aware of an argument that there are no explicit Gödelsentences for Δ.

OPEN QUESTION : Are there explicit Gödelsentences for Δ^R ?

OPEN QUESTION : If one allows Σ_1 variants (with one existential quantifier in front of the Δ) of Prov in the definition of Δ^R, can there be explicit Gödelsentences for Δ^R ?

K Kreisel's symmetrized Rosser System

Kreisel's variation on the Rosser System is reported in Hilbert-Bernays [1970], p298-302.

Define: $\Delta A :\Leftrightarrow \exists x[\mathrm{Proof}(x,A) \wedge \forall u,v,b,c \leq x((\mathrm{Proof}(u,b) \wedge \mathrm{Proof}(v,c)) \to c \neq \mathrm{neg}(b))]$

Clearly ΔA is Σ_1. The Kreisel System satisfies the principles 1-4 and the additional principle:

12 $\quad \vdash \neg(\Delta A \wedge \Delta\neg A)$

We can now prove 11 modally:

$$\begin{aligned} \vdash \Box\neg G &\to \Box\Delta G \wedge \Box\Delta\neg G && (6,9)\\ &\to \Box(\Delta G \wedge \Delta\neg G)\\ &\to \Box\bot && (12) \quad \Box \end{aligned}$$

Note: $\vdash \Box\bot \rightarrow$ "$\{A \mid \Delta A\}$ is finite".

R' A minor variation of Rosser's System

Yet another defect of the Rosser System is that we have no appropriate bimodal counterpart for the underlying principle:

$$\vdash \Box\bot \rightarrow (\Box A < \Box\neg A \vee \Box\neg A \leq \Box A)$$

This can be easily repaired. Define:

$$R'_0 := \emptyset$$

$$R'_{n+1} := \begin{cases} R'_n \cup \{A\} & \text{if Proof}(n,A) \text{ and } (\neg A) \notin R'_n \\ R'_n & \text{otherwise} \end{cases}$$

Let ΔA be the arithmetization of: $\exists n\ A \in R'_n$. Clearly ΔA is in Σ_1. We have: $\vdash \Delta A \leftrightarrow \Box A < \Delta\neg A$, and even: $\vdash \forall x(x \text{ wit } \Delta A \leftrightarrow x \text{ wit } \Box A < \Delta\neg A)$. This last fact happens to characterize $\Delta^{R'}$.

The principles 1-4 hold for Δ, plus the additional:

13 $\quad \vdash \Box\bot \rightarrow (\Delta A \vee \Delta\neg A)$

A direct consequence is:

14 $\quad \vdash \Box A \rightarrow (\Delta A \vee \Delta\neg A)$ $\qquad$ (4,13)

Finally note that: $\vdash \Delta^{R} A \rightarrow \Delta^{R'} A$.

BM The Bernardi-Montagna System

We now jump up directly to a system which is richer -from the modal point of view- both than Δ^K and $\Delta^{R'}$. This system was discovered by Claudio Bernardi and Franco Montagna (see Bernardi-Montagna[1982]). Define:

$$BM_0 := \emptyset$$

$$BM_{n+1} := \begin{cases} BM_n \cup \{A\} & \text{if Proof}(n,A) \text{ and } BM_n \cup \{A\} \text{ is consistent in propositional logic} \\ BM_n & \text{otherwise} \end{cases}$$

Let ΔA be the arithmetization of: $\exists n\ A \in BM_{n+1}$. Clearly ΔA is Σ_1. The principles 1-4 and 13 hold for Δ. We also have by elementary reasoning:

15 $\quad \vdash \Delta(A \rightarrow B) \rightarrow (\Delta A \rightarrow \Delta B)$

15 in combination with 2 entails 12, so the principles valid for Δ^{BM} comprise those valid for Δ^K and $\Delta^{R'}$ (at least in so far as we have found such principles).

mBM The modified Bernardi-Montagna System

For our purposes we want the following additional Principle:

16 $\vdash \Box A \to \Delta\Box A$

It is not plausible that one could prove 16 for the BM System without additional assumptions about the order of the proofs of PA. E.g. given $\Box\bot$, why would one have $\Delta\Box\bot$ rather than $\Delta\neg\Box\bot$? However we can modify the BM System in such a way that we get 16.

Let $\vdash_{Prop}$ stand for derivability in Propositional Logic. Define:

$$mBM_0 := \emptyset$$

$$mBM_{n+1} := \begin{cases} mBM_n \cup \{A\} & \text{if } Proof(n,A) \text{ and } mBM_n \cup \{A\} \nvdash_{Prop} \neg\Box\bot \\ mBM_n & \text{otherwise} \end{cases}$$

Let ΔA be the arithmetization of: $\exists n\; A \in mBM_{n+1}$. Clearly ΔA is Σ_1.

It is easily seen that 1-3 and 15 are valid. We check 4: reason in PA:

> Trivially $\Delta A \to \Box A$. Suppose $\neg\Box\bot$ and $\Box A$, say $Proof(x,A)$. The only reason A could be left out of mBM_{x+1} is that $mBM_x \cup \{A\} \vdash_{Prop} \neg\Box\bot$. But then $\Box\neg\Box\bot$ and hence $\Box\bot$. Contradiction. Conclude: $A \in mBM_{x+1}$ and thus ΔA. □

We prove 13: reason in PA:

> Suppose $\Box\bot$. For certain x and y we will have: $Proof(x,A)$ and $Proof(y,\neg A)$. Let $z := max(x,y)$. If $\neg\Delta A$ and $\neg\Delta\neg A$ we will have: $mBM_{z+1} \cup \{A\} \vdash_{Prop} \neg\Box\bot$ and $mBM_{z+1} \cup \{\neg A\} \vdash_{Prop} \neg\Box\bot$. Hence $mBM_{z+1} \vdash_{Prop} \neg\Box\bot$. Quod non. □

Finally we turn to 16. Clearly:

17 $\vdash \neg\Delta\neg\Box\bot$

It follows that:

$\vdash \Box\bot \to \Delta\Box\bot$ (13,17)

Moreover:

$\vdash \Delta\Box\bot \to \Delta\Box A$ (1,15)

Hence:

$\vdash \Box\bot \to (\Box A \to \Delta\Box A)$

Also:

$\vdash \neg\Box\bot \to (\Delta\Box A \leftrightarrow \Box\Box A)$ (4)

Thus:

$\vdash \neg\Box\bot \to (\Box A \to \Delta\Box A)$

So we may conclude that 16 holds. (Conversely one can derive 17 from 4, 12 and 16.)

Let us list for convenience the principles valid for Δ^{mBM} with brand-new names:

B1 $\vdash A \Rightarrow \vdash \Delta A$
B2 $\vdash \Delta(A \to B) \to (\Delta A \to \Delta B)$
B3 $\vdash \neg \Delta \bot$
B4 $\vdash \Box A \to \Delta \Box A$
B5 $\vdash \Delta A \to \Box \Delta A$
B6 $\vdash \neg \Box \bot \to (\Delta A \leftrightarrow \Box A)$
B7 $\vdash \Box \bot \to (\Delta A \vee \Delta \neg A)$

We note some important consequences of these principles. First a strengthening of Löb's Axiom:

18 $\vdash \Delta(\Box A \to A) \to \Delta A$

Proof:

$$\begin{aligned} \vdash \Delta(\Box A \to A) &\to \Box(\Box A \to A) && (5) \\ &\to \Box A && \\ &\to \Delta \Box A && (B4) \\ &\to \Delta A && (B2) \end{aligned}$$
□

The second consequence is the principle of provable extensionality:

19 $\vdash \Box(A \leftrightarrow B) \to \Box(\Delta A \leftrightarrow \Delta B)$

Proof:

$$\begin{aligned} \vdash \Box(A \leftrightarrow B) &\to \Box\Delta(A \leftrightarrow B) && (6) \\ &\to \Box(\Delta A \leftrightarrow \Delta B) && (B2) \end{aligned}$$
□

The next principle is an immediate consequence of 12 and B7:

20 $\vdash \Box \bot \to (\Delta A \leftrightarrow \nabla A)$

Define: $\Box^0 \bot := \bot$, $\Box^{n+1} \bot := \Box \Box^n \bot$, $\Box^\omega \bot := \top$. We will say that an arithmetical formula A is *modally closed* if A is built up from $\top, \bot$ with the propositional connectives and $\Box, \Delta$ (in other words: if A is an interpreted sentence of the closed fragment of the bimodal propositional logic with operators $\Box$ and Δ).

21 Suppose A is modally closed. There is an $\alpha \in \{0,\ldots,\omega\}$ such that: $\vdash \Delta A \leftrightarrow \Box^\alpha \bot$.

Proof: Consider B built from $\top, \bot$ with the propositional connectives and $\Box$. First there is the familiar fact that:

$$\vdash (B \wedge \Box B) \leftrightarrow \Box^\beta \bot \text{ , for some } \beta \in \{0,\ldots,\omega\}.$$

Hence:

$$\begin{aligned} \vdash \Delta B &\leftrightarrow \Delta(B \wedge \Box B) \\ &\leftrightarrow \Delta \Box^\beta \bot \end{aligned}$$

Secondly we have: $\vdash \Delta \Box^0 \bot \leftrightarrow \Box^0 \bot$, and $\vdash \Delta \Box^{1+\gamma} \bot \leftrightarrow \Box^{2+\gamma} \bot$ (as is easily seen by considering the cases that $\Box \bot$ and that $\neg \Box \bot$ separately). Combining we see that $\vdash \Delta B \leftrightarrow \Box^\delta \bot$, for some $\delta \in \{0,\ldots,\omega\}$.

21 follows with a trivial induction on A. (Note that we didn't use B7 in the argument.) □

We turn to the Henkinsentences of Δ. We have already seen that $\bot$ and $\top$ are explicit Henkinsentences (7,8). For Δ^{mBM} we can show that they are the only explicit Henkinsentences. Consider H satisfying:

22 $\quad \vdash H \leftrightarrow \Delta H$

If H is explicit, it follows that H is modally closed. Hence by 21:

$$\vdash H \leftrightarrow \Box^{\alpha}\bot \text{ , for some } \alpha\in\{0,\ldots,\omega\}.$$

If $\alpha\neq 0$, $\alpha\neq\omega$, it follows that for some $n\in\{1,2,\ldots\}$:

$$\vdash \Box^{n}\bot \leftrightarrow \Delta\Box^{n}\bot$$
$$\leftrightarrow \Box^{n+1}\bot$$

Quod non. □

OPEN QUESTION: Are there non-explicit Henkinsentences of Δ^{mBM}?

Next we turn to the Gödelsentences of Δ. Under R and K we have seen that these have the Rosser Property (10,11). We show that they are non-explicit and non-unique.

Consider G satisfying 9. If G were explicit, G would be modally closed. Hence by 21: $\vdash G \leftrightarrow \neg\Box^{\alpha}\bot$ $(\alpha\in\{0,\ldots,\omega\})$. If $\alpha\neq 0$, we see:

$$\vdash \Delta G \leftrightarrow \Delta\neg\Box^{\alpha}\bot \quad \text{(B1,B2)}$$
$$\leftrightarrow \Delta\bot \quad \text{(18)}$$
$$\leftrightarrow \bot \quad \text{(B3)}$$

So by 9: $\vdash G$. Thus $\alpha=0$. Contradiction. If $\alpha=0$, we have: $\vdash G$ and thus: $\vdash \Delta G$, i.e. by 9: $\vdash \neg G$. Contradiction. Hence G cannot be explicit.

To see that G is not unique we show that ∇G is also a Gödelsentence and that ∇G is not provably equivalent to G. First we show:

23 $\quad \vdash \nabla G \leftrightarrow \neg\Delta\nabla G$

To prove 23 it is clearly sufficient to show:

a $\quad \vdash \neg\Box\bot \rightarrow \nabla G$

b $\quad \vdash \neg\Box\bot \rightarrow \neg\Delta\nabla G$

c $\quad \vdash \Box\bot \rightarrow (\nabla G\leftrightarrow\neg\Delta\nabla G)$

We prove a by contraposition:

$$\vdash \Delta\neg G \rightarrow \Box\neg G \quad (5)$$
$$\rightarrow (\Box\Delta G\wedge\Box\Delta\neg G) \quad (6,9)$$
$$\rightarrow \Box\bot \quad (12)$$

To prove b, we show first:

$$\vdash \Box\Box\bot \rightarrow (\Delta\nabla G \rightarrow \Box\nabla G \quad (5)$$
$$\rightarrow \Box\Delta G \quad (20)$$
$$\rightarrow \Box\neg G \quad (9)$$
$$\rightarrow \Box\bot\ ) \quad \text{(as in the proof of a)}$$

Hence:

$$\vdash \Delta\nabla G \rightarrow (\Box\Box\bot\rightarrow\Box\bot)$$

Thus:

$$\vdash \Delta\nabla G \rightarrow \Box\Delta\nabla G \quad \text{(B5)}$$
$$\rightarrow \Box(\Box\Box\bot\rightarrow\Box\bot)$$
$$\rightarrow \Box\Box\bot$$

Combining:

$$\vdash \Delta\nabla G \rightarrow \Box\bot.$$

As to c:

$$
\begin{array}{lll}
\vdash \Box\bot & \to \Delta\Box\bot & \text{(B4)} \\
& \to \Delta(\nabla G \leftrightarrow \Delta G) & \text{(B1,B2,20)} \\
& \to (\Delta\nabla G \leftrightarrow \Delta\Delta G) & \text{(B2)} \\
& \to (\Delta\nabla G \leftrightarrow \Delta\neg G) & \text{(9,B1,B2)} \\
& \to (\neg\Delta\nabla G \leftrightarrow \nabla G) &
\end{array}
$$

Next we show that ∇G is not provably equivalent to G. It is clearly sufficient to prove:

24 $\vdash \Box(G \leftrightarrow \nabla G) \to \Box\bot$

Clearly:

$$
\begin{array}{lll}
\vdash (G \wedge \nabla G) & \to (\neg\Delta G \wedge \neg\Delta\neg G) & \\
& \to \neg\Box\bot & \text{(B7)}
\end{array}
$$

And:

$$
\begin{array}{lll}
\vdash (\neg G \wedge \neg\nabla G) & \to (\Delta G \wedge \Delta\neg G) & \\
& \to \bot & \text{(12)}
\end{array}
$$

Combining:

$$
\begin{array}{ll}
\vdash \Box(G \leftrightarrow \nabla G) & \to \Box(\, (G \wedge \nabla G) \vee (\neg G \wedge \neg\nabla G)\,) \\
& \to \Box\neg\Box\bot \\
& \to \Box\bot
\end{array}
$$

(Another way to prove 24 is by noting that:

$$
\begin{array}{lll}
\vdash \Box\Box\bot \to (\, \Box(G \leftrightarrow \nabla G) & \to \Box(G \leftrightarrow \Delta G) & \text{(20)} \\
& \to \Box\bot \quad) &)
\end{array}
$$

We leave it to the reader to verify that our procedure yields no further independent Gödelsentences, i.e. :

25 $\vdash G \leftrightarrow \nabla\nabla G$

In §7 we will see that as far as our modal principles are concerned we cannot show more than: if there is a Gödelsentence of Δ then there is a second non-equivalent one.

OPEN QUESTION : Are there three pairwise non-equivalent Gödel-sentences of Δ^{mBM} ?

THE R, K, R', BM and mBM Systems are all Σ_1, yet they escape the second Incompleteness Theorem. By a well known result of Feferman (see Feferman[1960]) these systems cannot be *provably* closed under the axioms and rules of predicate logic, in other words we do not have: $\vdash \Delta A \leftrightarrow \Box_\Delta A$. If Δ is one of Δ^{BM}, Δ^{mBM} we can say a bit more. Let Δ be one of Δ^{BM}, Δ^{mBM}. Let Q be the conjunction of the axioms of Robinson's Arithmetic. We clearly have: $\vdash Q$ and hence: $\vdash \Delta Q$ and thus: $\vdash \Box_\Delta Q$. It immediately follows from the provable Σ_1-completeness of Robinson's Arithmetic that:

26 $\vdash \Delta A \to \Box_\Delta \Delta A$

Let G be a Gödelsentence of Δ. We have:

d $\vdash \Box\bot \to (\Delta G \vee \Delta\neg G)$ (B7)

e $\vdash \Delta G \to \Box_\Delta \Delta G$ (26)

f	$\vdash \Delta G \to \Box_\Delta G$		
	$\to \Box_\Delta \neg\Delta G$		(9)
g	$\vdash \Delta G \to \Box_\Delta \bot$		(e,f)
h	$\vdash \Delta\neg G \to \Box_\Delta \Delta\neg G$		(26)
	$\to \Box_\Delta \neg\Delta G$		(B1,12)
i	$\vdash \Delta\neg G \to \Box_\Delta \neg G$		
	$\to \Box_\Delta \Delta G$		(9)
j	$\vdash \Delta\neg G \to \Box_\Delta \bot$		(h,i)
k	$\vdash \Box\bot \to \Box_\Delta \bot$		(d,g,j)
l	$\vdash \Box\bot \to (\Box_\Delta A \leftrightarrow \Box A)$		(k)
m	$\vdash \neg\Box\bot \to (\Box_\Delta A \leftrightarrow \Box_\Box A$		
	$\leftrightarrow \Box A)$		
27	$\vdash \Box_\Delta A \leftrightarrow \Box A$		(l,m)

So Δ^{BM} and Δ^{mBM} are provably axiom sets for the theorems of PA. The same thing can be proved for $\Delta^{R'}$ by a slightly refined variant of the above argument. What about Δ^{R} and Δ^{K}? I don't know, but *yes* in the first case and *no* in the second seems to me a good guess.

We now turn to systems that are provably closed under the axioms and rules of predicate logic.

F The Feferman System

The Feferman System was invented by Feferman (see Feferman[1960]) as an illustration in the study of the conditions for Gödel's second Incompleteness Theorem. Orey discovered important applications of its provability predicate in the theory of Relative Interpretability (see Feferman[1960], Orey[1961]). A modal study of this provability predicate was made by Montagna (Montagna[1978]).

Let me start by giving two rather different intuitive descriptions of the Feferman System (or, to be faithful to the conventions of 5.6, I should rather say: let me describe two variants of the Feferman System).

Suppose the arithmetical axioms of PA are enumerated as: $A_1, A_2, A_3, \ldots$, in the order of their Gödelnumbers (i.e. : $i<j \Rightarrow \ulcorner A_i \urcorner < \ulcorner A_j \urcorner$). We call a set X of arithmetical axioms of PA *initial* if: $A_i \in X$ and $j<i \Rightarrow A_j \in X$.

The Feferman System is simply the first order system in the language of PA axiomatized by: $F := \bigcup\{X \mid X$ is a finite, initial, consistent set of arithmetical axioms of PA$\}$.

Clearly from the extensional point of view F coincides with the usual axiom set of PA. The Feferman System can be viewed as a system, where to be licened to use axiom A_i one needs the *external information* that $\{A_j \mid j \le i\}$ is consistent.

The second way to introduce the Feferman System is as follows: suppose we enumerate the *proofs* in the system PA by: $\pi_1, \pi_2, \pi_3, \ldots$. As soon as we hit upon a proof π_i of $\perp$, we extract the axiom A_j with largest Gödelnumber from π_i . We backtrack and scratch out all the proofs employing axioms A_k with $k \ge j$. Then we go on enumerating proofs, skipping those employing axioms A_k with $k \ge j$. As soon as we meet another proof of $\perp$ we repeat the procedure. We call a proof *stable* if it occurs in our enumeration and is never scratched out. The stable proofs are the proofs of the Feferman System.

Under this last description the Feferman System can be seen as a fully effective procedure, that will eventually yield all stable proofs. The catch here is of course that someone who does not know the consistency of PA will not be able to predict -at least not prima facie- when a proof is stable. In fact the situation is more subtle: someone knowing PA, but not its consistency, will ipso facto not know that *all proofs* are stable, but he will know *of every proof* that it is stable.

We turn to the formal definition of the Feferman System. Define:

$\pi^*(x) \;:\Leftrightarrow\; \pi(x) \wedge \Diamond\!\upharpoonright\! x\top$

$\Delta A \;:\Leftrightarrow\; \Box_{\pi^*} A$

We give a few equivalents of ΔA.

28 $\quad \vdash \Delta A \leftrightarrow \exists x\, (\Box\!\upharpoonright\! x A \wedge \Diamond\!\upharpoonright\! x\top)$

Let f be a primitive recursive function with:

$$f(n) := \begin{cases} \text{the largest of the Gödel numbers of the arithmetical axioms occurring in } \pi \text{ if } n = \ulcorner \pi \urcorner \text{ for some proof } \pi \\ 0 \text{ otherwise} \end{cases}$$

We have:

29 $\quad \vdash \Delta A \leftrightarrow \exists x\, (\mathrm{Proof}(x,A) \wedge \Diamond\!\upharpoonright\! f(x)\top)$

Remember that $\Box^* A :\Leftrightarrow \exists x \Box\!\upharpoonright\! x A$. We have:

30 $\quad \vdash \Delta A \leftrightarrow \Box^* A < \Box^* \perp$

31 $\quad \vdash \Delta A \leftrightarrow \Box^* A < \Box^* \neg A$

31 brings out the similarity between the Feferman System and the Rosser System. By 5.4 and 30 or 31 we see that Δ is Δ_2 .

B1-B6 are valid for Δ. In Feferman[1960] all of these except B4 are mentioned. In Montagna[1978] a modal study is made of B1, B2, B3, B5, B6. The validity of B2, B3 and B6 is immediate. To prove the other principles we will use the well known fact that PA is provably essentially reflexive, and hence:

$$\vdash \forall x \Box(\Box{\restriction}x A \rightarrow A)$$

It follows that:

$$\vdash \forall x \Box\Diamond{\restriction}x \top$$

Ad B1:

$$\begin{aligned} \vdash A &\Rightarrow \text{for some } n \vdash \Box{\restriction}\underline{n}A \\ &\Rightarrow \text{for some } n \vdash \Box{\restriction}\underline{n}A \wedge \Diamond{\restriction}\underline{n}\top \\ &\Rightarrow \vdash \Delta A \end{aligned}$$

Ad B4:

We will prove the stronger principle:

32 Let S be Σ_1, then: $\vdash S \rightarrow \Delta S$

Let $\Box_Q$ stand for provability in Robinson's Arithmetic. For some q PA$\vdash \Box_Q A \rightarrow \Box{\restriction}\underline{q}A$. Let S be Σ_1. We have:

$$\begin{aligned} \vdash S &\rightarrow \Box_Q S \\ &\rightarrow \Box{\restriction}\underline{q}S \\ &\rightarrow (\Box{\restriction}\underline{q}S \wedge \Diamond{\restriction}\underline{q}\top) \\ &\rightarrow \Delta S \end{aligned}$$

Note that we *do* have B4 for Δ^{mBM}, but not 32. (In fact *assuming* 32 for Δ^{mBM} quickly leads to the inconsistency of PA)

Ad B5:

It is clearly sufficient to prove 6, i.e. : $\vdash \Box A \rightarrow \Delta\Box A$. To do this we formalize the reasoning for B1:

$$\begin{aligned} \vdash \Box A &\rightarrow \exists x \Box\Box{\restriction}x A \\ &\rightarrow \exists x \Box(\Box{\restriction}x A \wedge \Diamond{\restriction}x \top) \\ &\rightarrow \Box\Delta A \end{aligned}$$

Just as Δ^{mBM}, Δ^F has precisely two non-equivalent *explicit* Henkin-sentences. I will now show that Δ^F has in fact infinitely many pairwise non-equivalent Henkinsentences. First we need to know a bit about Σ-minded sentences. A sentence A is *Σ-minded* if both $A \wedge \Box A$ and $\neg A \wedge \Box\neg A$ are provably equivalent (in PA) to a Σ_1 formula. A good example of a Σ-minded sentence is the ordinary Σ_1 Rossersentence. We have:

33 if A is Σ-minded then: $\vdash \Delta A \leftrightarrow (\Box A \wedge (\Box\bot \rightarrow A))$

Proof: Suppose A is Σ-minded.

"$\rightarrow$" Clearly $\vdash \Delta A \rightarrow \Box A$. Moreover:

$$\vdash (\Delta A \wedge \Box\bot) \to (\neg A \to (\neg A \wedge \Box\neg A) \to \Delta(\neg A \wedge \Box\neg A) \to \Delta\bot \to \bot)$$

(B1,B2,32)
(B1,B2)
(B3)

"←" We have:

$$\vdash (\Box A \wedge (\Box\bot \to A)) \to (\Box\bot \to (A \wedge \Box A) \to \Delta(A \wedge \Box A) \to \Delta A)$$

(B1,B2,32)
(B1,B2)

Moreover:

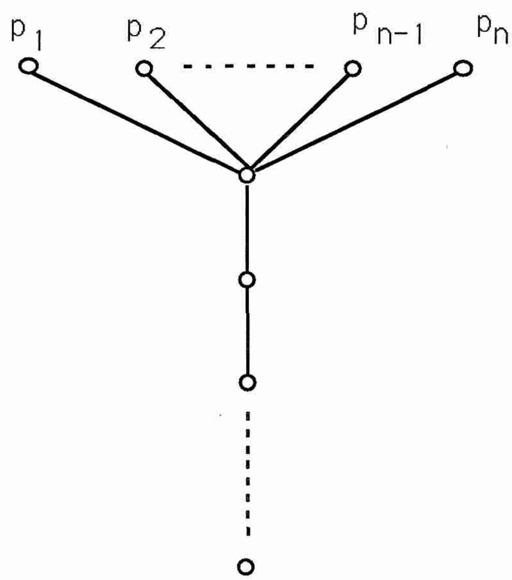
$$\vdash (\Box A \wedge (\Box\bot \to A)) \to (\neg\Box\bot \to \Delta A) \quad \text{(B6)} \quad \Box$$

Note that our proof only uses B1, B2, B3, B6, 32. 33 is an example of the phenomenon of *reduction* : an arithmetical predicate takes a simple, uncharacteristic form on some restricted set of formulas. A further, more involved example of reduction will be given in section 9 (see 9.1).

To prove that there are infinitely many pairwise non-equivalent Henkin-sentences I have to borrow some material and definitions of Visser [1984]. The reader not familiar with this paper can at least get the essential idea of the argument by considering the ordinary Σ_1 Rossersentence R (i.e. any sentence satisfying: $\vdash R \leftrightarrow \Box\neg R < \Box R$) and $S := \Box R \leq \Box\neg R$, and by proving for her/himself that R and S are Σ-minded and satisfy: $\vdash R \leftrightarrow (\Box R \wedge (\Box\bot \to R))$, $\vdash S \leftrightarrow (\Box S \wedge (\Box\bot \to S))$.

Consider a tailmodel K . We write: $[\![\phi]\!]$, for the set of nodes that force ϕ ; $[\phi]$, for $[\phi](K,PA)$, and: $\langle\phi\rangle$, for $\langle\phi\rangle(K,PA)$. Note that $[\![\phi \wedge \Box\phi]\!]$ is upwards closed and that $PA \vdash ([\phi] \wedge \Box[\phi]) \leftrightarrow [\phi \wedge \Box\phi]$. It follows that $\vdash ([\phi] \wedge \Box[\phi]) \leftrightarrow \exists x\, h(x) \in [\![\phi \wedge \Box\phi]\!]$, and hence that $[\phi] \wedge \Box[\phi]$ is provably equivalent to a Σ_1 sentence. Combining this with the fact that $\vdash \neg[\phi] \leftrightarrow [\neg\phi]$, we find that $[\phi]$ is Σ-minded.

Now consider the tailmodel pictured below:



Here the p_i are only forced as shown. Clearly $\Vdash p_i \leftrightarrow (\Box p_i \wedge (\Box\bot \rightarrow p_i))$, hence by the Embedding Lemma and the fact that $[p_i]$ is Σ-minded: $\vdash [p_i] \leftrightarrow \Delta[p_i]$. On the other hand for $i \neq j$: $\Vdash (p_i \leftrightarrow p_j) \rightarrow \neg\Box\bot$, hence: $\vdash [\ (p_i \leftrightarrow p_j) \rightarrow \neg\Box\bot\ ]$ and so: $\vdash ([p_i] \leftrightarrow [p_j]) \rightarrow \neg\Box\bot$. Thus the $[p_i]$ are pairwise non-equivalent Henkinsentences of Δ. Since n can be freely chosen it follows that there are infinitely many pairwise non-equivalent Henkinsenteces of Δ.

We state two open problems:

OPEN PROBLEM : Are there Henkinsentences of Δ^F that are not provably equivalent to Σ_1 sentences?

OPEN PROBLEM : What are the possible truthvalues of the *literal* Henkinsentences of Δ^F?

We turn to Gödelsentences. Let G satisfy 9. Just as in the case of Δ^{mBM} G is non explicit. (We can, using the observations about tail models above, also see this "vom höheren Standpunkt", for consider the 'minimal' tailmodel, i.e. the linear one. The 'propositions' of this model correspond precisely to the closed fragment of Löb's Logic. Clearly the interpretations of this closed fragment are going to be closed under Δ (modulo provable equivalence). It follows that the modally closed sentences are provably equivalent to arithmetical interpretations of elements of the closed fragment of Löb's Logic. In the model the 'equation' ($\phi \leftrightarrow \neg(\Box\phi \wedge (\Box\bot \rightarrow \phi))$) has no solution, hence no modally closed sentence solves the equation in PA !)

The argument for the non-uniqueness of the Gödelsentences of Δ^{mBM} depended upon B7, so we can't use it here. The problem of the uniqueness of G thus remains open. This problem was first posed in Montagna[1978].

MONTAGNA'S PROBLEM : Is G unique?

The Gödelsentence of Δ^F is an Oreysentence. Before defining what an Oreysentence is, I want to note that the fact that G is such a sentence only depends upon B1, B2, B3, 6 and 32. Let's call a Δ that satisfies these principles *precocious* .

An *Oreysentence* is a sentence A with the property $A \lhd \top$ and $\neg A \lhd \top$. (Strictly speaking my usage is at variance with the tradition: e.g. Gödelsentences and Rossersentences are sentences that solve certain fixed equations; on the other hand we say of a sentence A satisfying $\vdash (\Box A \vee \Box\neg A) \rightarrow \Box\bot$, that *it has the Rosser Property* . Rossersentences have the Rosser Property, but there are others. So the more correct usage would be: sentence with the Orey Property.) Trivially the negation of an Oreysentence is again an Oreysentence.

We show that the Gödelsentence of any precocious Δ is an Oreysentence. Suppose Δ is precocious. First we prove:

34 $\quad \vdash A \lhd \nabla A$

Proof:

$$\begin{aligned} \vdash \forall x\, \Box(\Box_{\ulcorner}x B \to B) &\to \forall x\, \Box\Delta(\Box_{\ulcorner}x B \to B) && (6)\\ &\to \forall x\, \Box(\Delta\Box_{\ulcorner}x B \to \Delta B) && (B2)\\ &\to \forall x\, \Box(\Box_{\ulcorner}x B \to \Delta B) && (32) \end{aligned}$$

□

Secondly one easily proves using I1, I2, I3:

35 $\quad \vdash A \lhd \neg A \to A \lhd \top$

We have:

$$\begin{aligned} \vdash G &\lhd \nabla G && (34)\\ &\lhd \Delta G && (B1,B2,B3,I1,I2)\\ &\lhd \neg G && (9,I1,I2) \end{aligned}$$

Hence by 35:

$$\vdash G \lhd \top$$

Moreover:

$$\begin{aligned} \vdash \neg G &\lhd \nabla\neg G && (34)\\ &\lhd \neg\Delta G && (B1,B2,I1,I2)\\ &\lhd G && (9,I1,I2) \end{aligned}$$

Hence by I1, I2 and 35:

$$\vdash \neg G \lhd \top$$

A curious fact is that the Gödelsentence of Δ^F is *precisely* the Oreysentence discovered independently by Lindström and Švejdar (see Lindström[1979] and Švejdar[1978]); by 5.4 *this* Oreysentence is Δ_2. In §9 we will see that there are infinitely many non-equivalent Oreysentences.

Before leaving the subject of Oreysentences, I want to note that Oreysentences are Σ_1- and Π_1-flexible and that they are Kentsentences.

Let Γ be a set of formulas. A formula A is *Γ-flexible* if for all B in Γ: $\vdash \Box\neg(A \leftrightarrow B) \to \Box\bot$. A sentence A is a *Kentsentence* if $(A \wedge \Box A)$ is not provably equivalent to a Σ_1 sentence. I show that an Oreysentence is a Kentsentence and leave the proof that Oreysentences are Σ_1- and Π_1-flexible to the reader. Suppose A is an Oreysentence and suppose for a reductio that A is a Kentsentence. Then clearly $(\neg A \vee \Diamond\neg A)$ is provably equivalent to a Π_1 sentence and hence:

$$\begin{aligned} \vdash (A \lhd \top \wedge \neg A \lhd \top) &\to (\neg A \vee \Diamond\neg A) \lhd \top && (I1,I2)\\ &\to \Box(\neg A \vee \Diamond\neg A) && (I1,I2,J1)\\ &\to \Box(A \to \Diamond\neg A) && \\ &\to \Diamond\neg A \lhd \top && (I1,I2)\\ &\to \Box\Diamond\neg A && (I5)\\ &\to \Box\bot && \end{aligned}$$

□

mF The modified Feferman System

The modified Feferman system is a modification both of the Feferman system and of the BM system. Define:

$$mF_0 := \emptyset$$

$$mF_{n+1} := \begin{cases} mF_n \cup \{A\} & \text{if Proof}(n,A) \text{ and } mF_n \cup \{A\} \text{ is consistent} \\ mF_n & \text{otherwise} \end{cases}$$

Let $\Delta_x A$ be the arithmetization of: $A \in mF_{x+1}$. Define further:

$$\Delta A \ :\Leftrightarrow\ \exists x\ \Delta_x A$$

$$\Delta^* A :\Leftrightarrow\ \exists x\ \Box_{\Delta x} A$$

It is easily seen that: $\vdash \Delta A \leftrightarrow \Box_\Delta A$, and hence that: $\vdash \Delta A \leftrightarrow \Delta^* A$.
Moreover we have: $\vdash \Delta A \leftrightarrow \Box A < \Delta^* \neg A$, and even:

$$\vdash \forall x\ (\ x \text{ wit } \Delta A \leftrightarrow x \text{ wit } \Box A < \Delta^* A\ ).$$

This last observation brings out the Rosserlike character of Δ.

We claim that Δ satisfies B1-B7. The argument for the validity of B1-B6 is similar to the one for the case of Δ^F. We treat for example B5.

Define: $\Box_x A :\Leftrightarrow \exists y \leq x \text{ proof}(x,A)$. Clearly $\vdash \forall x\ \Box \neg \Box_{\Box x} A$, and hence by induction on x in PA: $\vdash \forall x\ \Box(\Delta_x A \leftrightarrow \Box_x A)$. It follows that:

$$\begin{aligned} \vdash \Delta A &\rightarrow \Box A \\ &\rightarrow \exists x \Box \Box_x A \\ &\rightarrow \Box \Delta A \end{aligned}$$

The argument for B7 is similar to the one for the case of Δ^{mBM}. Just as Δ^F, Δ^{mF} satisfies 32, i.e. Δ^{mF} is provably Σ_1-complete.

Prima facie Δ is Σ_2. It is seen to be Δ_2 by the following observation:

35 $\quad \vdash \neg \Delta A \leftrightarrow (\ \neg \Box A \vee (\Box \bot \wedge \Delta \neg A)\ ) \quad$ (B6,B7)

About the Henkinsentences of Δ^{mF} the same remarks can be made as for Δ^F. Just as Δ^{mBM}, Δ^{mF} has at least two non-equivalent Gödelsentences. Clearly Δ^{mF} is precocious. It is now easy to see that the two non-equivalent Gödelsentences and their negations give us four pairwise non-equivalent Oreysentences. In §8 we will show that Δ^{mF} has in fact infinitely many pairwise non-equivalent Gödelsentences; thus there are infinitely many pairwise non-equivalent Oreysentences.

Δ^{mF} is our final system and the main object of study of this paper. In §7 we will study the principles B1-B7 from the modal point of view. In §8 we give a partial result on embedding Kripke models for our modal system into Arithmetic. In §9 we apply the result of §8 to Relative Interpretability.

§ 7 The system BMF

7.1 Description of the system

BMF is the smallest system, containing the tautologies of propositional logic, closed under Modus Ponens and the following axioms and rules:

L1	$\vdash \varphi \Rightarrow \vdash \Box \varphi$
L2	$\vdash \Box(\varphi \to \psi) \to (\Box \varphi \to \Box \psi)$
L3	$\vdash \Box \varphi \to \Box \Box \varphi$
L4	$\vdash \Box(\Box \varphi \to \varphi) \to \Box \varphi$
B1	$\vdash \varphi \Rightarrow \vdash \Delta \varphi$
B2	$\vdash \Delta(\varphi \to \psi) \to (\Delta \varphi \to \Delta \psi)$
B3	$\vdash \neg \Delta \bot$
B4	$\vdash \Box \varphi \to \Delta \Box \varphi$
B5	$\vdash \Delta \varphi \to \Box \Delta \varphi$
B6	$\vdash \neg \Box \bot \to (\Delta \varphi \leftrightarrow \Box \varphi)$
B7	$\vdash \Box \bot \to (\Delta \varphi \vee \Delta \neg \varphi)$

This list is very long and rather redundant. A more economical list is: B1, B2, B3, B5, B7 and the principles:

B8	$\vdash \Box \varphi \leftrightarrow (\Delta \varphi \vee \Box \bot)$
B9	$\vdash \Delta(\Box \varphi \to \varphi) \to \Delta \varphi$

B8 is easily derived from L1, L2, B6. B9 follows from L1, L2, L4, B2, B4, B6.

Let me briefly indicate how to derive the long list from the short one: L1 follows from B1, B8; L2 from B2, B8. L4 is proved from B8, B9; B6 from B8. We show how to derive B4 by a familiar trick:

a)	$\vdash \varphi \to (\Box(\varphi \wedge \Box \varphi) \to (\varphi \wedge \Box \varphi))$	L1
b)	$\vdash \Delta \varphi \to \Delta(\Box(\varphi \wedge \Box \varphi) \to (\varphi \wedge \Box \varphi))$	a, B1, B2
c)	$\vdash \Delta \varphi \to \Delta(\varphi \wedge \Box \varphi)$	b, B9
d)	$\vdash \Delta \varphi \to \Delta \Box \varphi$	c, B1, B2
e)	$\vdash \neg \Box \bot \to (\Box \varphi \to \Delta \Box \varphi)$	d, B6
f)	$\vdash \Box \bot \to (\Delta \Box \bot \vee \Delta \neg \Box \bot)$	B7
g)	$\vdash \neg \Delta \neg \Box \bot$	B3, B9
h)	$\vdash \Box \bot \to \Delta \Box \bot$	f, g

i)	$\vdash \Box \bot \to \Box \varphi$	L1, L2
j)	$\vdash \Delta \Box \bot \to \Delta \Box \varphi$	i, B1, B2
h)	$\vdash \Box \bot \to \Delta \Box \varphi$	h, j
l)	$\vdash \Box \varphi \to \Delta \Box \varphi$	e, h

Finally L3 follows from B4, B8.

We list a few further convenient consequences of BMF:

B10	$\vdash \Delta \varphi \to \Box \varphi$	
B11	$\vdash \Box \varphi \to \Box \Delta \varphi$	
B12	$\vdash \Box \bot \to (\Delta(\varphi \vee \psi) \leftrightarrow (\Delta \varphi \vee \Delta \psi))$	
B13	$\vdash \Box(\varphi \leftrightarrow \psi) \to \Box(\Delta \varphi \leftrightarrow \Delta \psi)$	(Provable Extensionality)
S	$\vdash \Box \varphi \to (\psi \leftrightarrow \chi) \Rightarrow$	
	$\vdash \Box \varphi \to (\nu[\psi / p] \leftrightarrow \nu[\chi / p])$	(Substitution Rule)

7.2 Non Uniqueness & Non-Expliciteness in BMF

Clearly the discussion on non uniqueness and non expliciteness of Henkin- and Gödelsentences of § 6 under mBM can be carried out in BMF.
E.g. one can show:

$$\vdash \Box (p \leftrightarrow \neg \Delta p) \to \Box(\nabla p \leftrightarrow \neg \Delta \nabla p)$$

$$\vdash \Box (p \leftrightarrow \neg \Delta p) \to (\Box(p \leftrightarrow \nabla p) \to \Box \bot)$$

7.3 Kripke Semantics for BMF

7.3.1 Definition

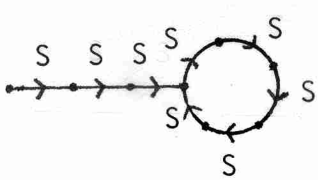
a) Let K be a finite, non empty set. Let S be a binary relation on K. The structure <K,S> is called *a lolly* if:

i) for each k,k' in K: $k S^T k'$. Here S^T is the transitive, symmetric, reflexive closure of S.

ii) for each k in K, there is *precisely one* k' in K with k S k'. We will call k' with k S k': the *S-successor* of k.

iii) there is at most one k in K such that for no k' in K: k' S k.

Clearly a lolly looks like this:



b) A lolly such that for every k in K there is a k' in K with $k'Sk$ is also called *a circle*.

c) A structure $<K,R,S>$ is called *a lolly-frame* if K is non empty, R and S are binary relations in K and:

i) R is transitive

ii) R is upwards wellfounded

Let $K_0 := \{k \text{ in } K \mid \text{for no } k' \text{ in } K\ kRk'\}$

$K_1 := K \setminus K_0$

$S_0 := S \upharpoonright K_0$

$S_0^T :=$ the transitive, symmetric, reflexive closure of S_0

iii) $k \in K_1 \Rightarrow (kSk' \Leftrightarrow kRk')$

iv) Suppose $k \in K_0$. Let $[k] := \{k' \mid k' S_0^T k\}$. Then $<[k], S_0 \upharpoonright [k]>$ is a lolly. Moreover if $k'Rk$, then $k'Rk''$ for all k'' in $[k]$.

v) $k \in K_0$ and $kSk' \Rightarrow k' \in K_0$

d) A *lolly-model* is a structure $<K,R,S,\Vdash>$, where $<K,R,S>$ is a lolly-frame and $\Vdash$ is a relation between elements of K and formulas of the language of BMF, satisfying:

i) $k \Vdash \top$

ii) $k \nVdash \bot$

iii) $k \Vdash (\varphi \wedge \psi) \Leftrightarrow (k \Vdash \varphi$ and $k \Vdash \psi)$

iv) $k \Vdash (\varphi \vee \psi) \Leftrightarrow (k \Vdash \varphi$ or $k \Vdash \psi)$

v) $k \Vdash (\varphi \to \psi) \Leftrightarrow (k \Vdash \varphi \Rightarrow k \Vdash \psi)$

vi) $k \Vdash \neg\varphi \Leftrightarrow k \nVdash \varphi$

vii) $k \Vdash \Box\varphi \Leftrightarrow$ for all k' with kRk' $k' \Vdash \varphi$

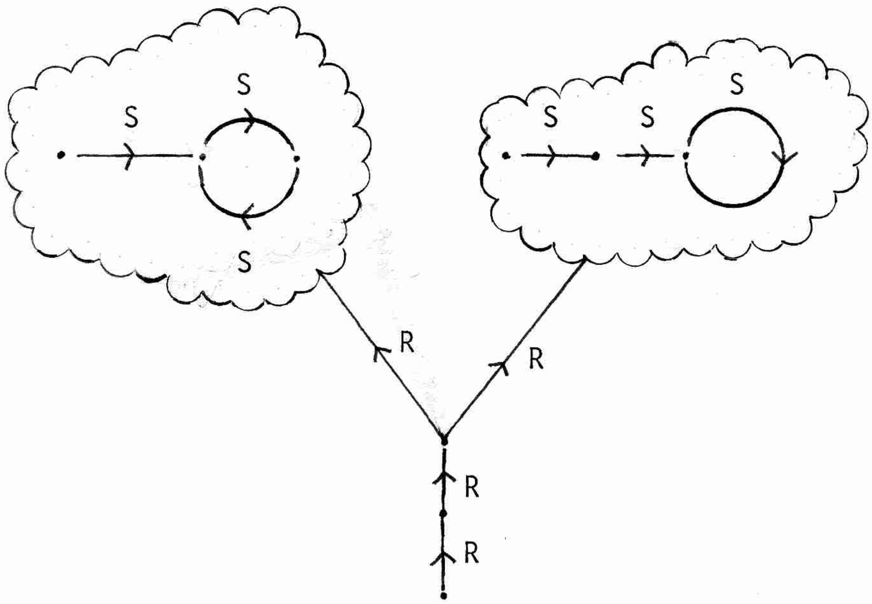
viii) $k \Vdash \Delta\varphi \Leftrightarrow$ for all k' with kSk' $k' \Vdash \varphi$

7.3.2 Remark

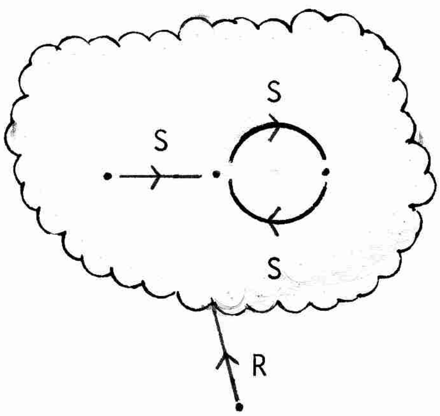
It is easy to verify that lolly-frames also satisfy:

$kRk'Sk'' \Rightarrow kRk''$, and $kSk'Rk'' \Rightarrow kRk''$.

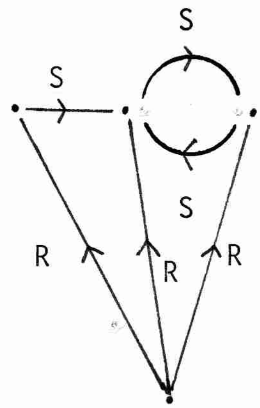
A lolly-frame is best visualized as a conventional frame for provability logic where the topnodes are blown up to lollies:



Here e.g.



means:



Note that we don't draw the arrows to exhibit the transitivity of R. Also because $R \subseteq S$, we don't write 'S' next to R-arrows.

7.3.3 Soundness

Consider any lolly-model $\mathcal{K} = \langle K, R, S, \Vdash \rangle$. We write $\mathcal{K} \Vdash \varphi$ for: for all $k \in K$ $k \Vdash \varphi$. We have: BMF $\vdash \varphi \Rightarrow \mathcal{K} \Vdash \varphi$.

Proof: entirely routine. □

7.3.4 Completeness

Suppose BMF $\nvdash \varphi$, then there is a finite lolly-model K such that $K \nVdash \varphi$.

We proceed with some preliminaries for the proof of 7.3.4.

7.3.5 Definition

Let Γ and Δ be sets of formulas of the language of BMF.

a) $\Gamma \vdash \Delta :\Leftrightarrow$ there are finite $\Gamma_0 \subseteq \Gamma$, $\Delta_0 \subseteq \Delta$ such that BMF $\vdash \bigwedge\Gamma_0 \to \bigvee\Delta_0$.

(The empty conjunction is: $\top$, the empty disjunction: $\bot$)

b) Let X be a set of formulas. Γ is *X-saturated* if: Γ is consistent and for each $\Delta \subseteq X$: $\Gamma \vdash \Delta \Rightarrow$ there is a $\varphi \in \Delta$ such that $\varphi \in \Gamma$.

7.3.6 Lemma

Suppose $\Gamma \nvdash \Delta$. Let X be a set of formulas. There is a set $Y \subseteq X$, such that $Y \cup \Gamma$ is X-saturated and $Y, \Gamma \nvdash \Delta$.

Proof: entirely routine. □

7.3.7 The Henkin Construction

Let X be a finite set of formulas that is closed under subformulas, such that: $\neg\Box\bot \in X$ and $(\Delta\varphi \in X \Leftrightarrow \Box\varphi \in X)$.

We construct a Kripke Model.

K, the set of nodes of the Kripke model we are constructing, consists of those sets of formulas y such that:

i) y is X-saturated

ii) if φ is in y and not in X, then φ is of the form $\Delta\psi$ and both ψ and $\Delta\Delta\psi$ are in y.

Clearly y consists of elements of X plus for certain χ in $X \cap y$ also $\Delta\chi$, $\Delta\Delta\chi$, $\Delta\Delta\Delta\chi$, ... As is easily seen K is finite and non-empty (by 7.3.6). For $x, y \in k$ we define:

$x R y :\Leftrightarrow (\Box\varphi \in x \Rightarrow \varphi, \Delta\varphi, \Delta^2\varphi, \ldots \in y)$ and (there is a $\Box\psi \in y$ with $\Box\psi \notin x$)

$x\,S\,y :\Leftrightarrow ((\neg\Box\bot)\in x$ and $x\,R\,y)$ or $(\Box\bot\in X, \Box\bot\in y\; x$ and $(\Delta\varphi\in X \Rightarrow \varphi\in y)$ and $((\Delta\varphi\notin x$ and $\Delta\varphi\in X) \Rightarrow \varphi\notin y))$

Finally we define: $x \Vdash p_i :\Leftrightarrow p_i \in x$.

Claim 1: R is transitive and irreflexive (and hence upwards wellfounded).

Claim 2: $x\,R\,y\,S\,z \Rightarrow x\,R\,z$

The simple proofs of Claims 1 & 2 are left to the reader.

Claim 3: For $\varphi\in X$: $x \Vdash \varphi \Leftrightarrow \varphi\in x$

Proof of Claim 3: We prove this claim by induction on φ in X. The only non trivial cases are when φ is of the form $\Box\psi$ or $\Delta\psi$.

- Suppose $\varphi = \Box\psi$.
  - Suppose $\Box\psi\in x$ and $x\,R\,y$. Then $\psi\in y$, hence by IH: $y\Vdash\psi$. Therefore $x\Vdash\Box\psi$.
  - Suppose $\Box\psi\notin x$. Let $x^R := \{\chi, \Delta\chi, \Delta^2\chi, .. \mid \Box\chi\in x\}$. We claim: $x^R, \Box\psi \nvdash \psi$. Otherwise: $\{\Box\chi, \Box\Delta\chi, \ldots \mid \Box\chi\in x\} \vdash \Box(\Box\psi\to\psi)$. Hence $\{\Box\chi \mid \Box\chi\in x\}\vdash\Box\psi$ and thus $\Box\psi\in x$. Quod non. By 7.3.6 there is a set $x_0\subseteq X$ such that $x_0\cup x^R\cup\{\Box\psi\}$ is X-saturated and $x_0\cup x^R\cup\{\Box\psi\}\nvdash\psi$. As is easily seen $x_0\cup x^R\cup\{\Box\psi\}\in K$. Define $y := x_0\cup x^R\cup\{\Box\psi\}$. Clearly $x\,R\,y$ and $\psi\notin y$ and so by IH $y\nVdash\psi$. Hence $x\nVdash\Box\psi$.

- Suppose $\varphi = \Delta\psi$. In case $(\neg\Box\bot)\in x$, this reduces to the previous case. So we assume $\Box\bot\in x$.
  - Suppose $\Delta\psi\in x$ and $x\,S\,y$. Then $\psi\in y$, hence by IH: $y\Vdash\psi$. Conclude: $x\Vdash\Delta\psi$.
  - Suppose $\Delta\psi\notin x$. Let $x^S := \{\chi \mid \Delta\chi\in x\}\cup\{\Box\bot\}$ and $x_S := \{\chi \mid \Delta\chi\notin x, \Delta\chi\in X\}$. We claim: $x^S \nvdash x_S$. Otherwise: $x\vdash\Delta\,W x_S$, ergo (by B12, using the fact that $\Box\bot\in x)$: $x\vdash W\{\Delta\chi \mid \chi\in x_S\}$. Hence $x\vdash\Delta\chi$ for *some* χ in x_S. Contradiction. By 7.3.6 there is an $x_0\subseteq X$ such that $x_0\cup x^S$ is X-saturated and $x_0\cup x^S\nvdash x_S$. Let $y := x_0\cup x^S$. We show: $y\in K$. Suppose $\nu\in y$ and $\nu\notin X$. Clearly $\nu\in x^S$, hence $\Delta\nu\in x$. $\Delta\nu\notin X$, so ν and $\Delta\Delta\nu$ are in x. $\nu\in x$ and $\nu\notin X$, so ν must be of the form $\Delta\rho$. Conclude that ρ and $\Delta\Delta\rho$ are in x^S. Next we show $x\,S\,y$. We have: $\Box\bot\in x$, $\Box\bot\in y$ and $\Delta\chi\in x\Rightarrow\chi\in y$. If $\Delta\chi\notin x$ and $\Delta\chi\in X$, then $\chi\in x_S$, so $\chi\notin y$.

Conclude $x\,S\,y$.

Since $\psi \in x_S$, we have $\psi \notin y$. So by IH : $y \nVdash \psi$ and hence $x \nVdash \Delta\psi$.

□

Claim 4: (There is a y with $x\,R\,y$) $\Leftrightarrow$ $(\neg\Box\bot) \in x$

Claim 5: for every x there is a y with $x\,S\,y$

We leave the simple proofs to the reader. (For the proof of claim 5, note that $\Delta\bot \in X$.)

The model we constructed is not quite a lolly-model yet, so a small transmutation is due.

Consider any x such that $\Box\bot \in x$. Clearly we can produce a sequence $x =: x_0\,S\,x_1\,S \ldots S\,x_{n+1}$, where $x_i = x_{n+1}$ for some $i < n+1$ and where if $k < j$ and $x_k = x_j$, then $k = i$ and $j = n+1$. We define a small lolly model L_x as follows: $\langle\{x_0, \ldots, x_n\}, R', S', \Vdash'\rangle$, where:

- R' is empty
- $y\,S'\,z :\Leftrightarrow y = x_j$ and $z = x_{j+1}$ for some $j \in \{0, \ldots, n\}$
- $y \Vdash' p_i \;:\Leftrightarrow p_i \in y$

We claim for $y \in \{x_0, \ldots, x_n\}$ and $\varphi \in X : y \Vdash' \varphi \Leftrightarrow y \Vdash \varphi$.

Proof: by induction on φ in X for all x_j simultaneously.

The atomic case and the cases of $\wedge, \vee, \neg, \rightarrow$ are trivial. If φ is $\Box\psi$ it is sufficient to note that, since R' is empty, $x_j \Vdash' \Box\psi$ and that on the other hand $\Box\bot \in x_j$ for each x_j. Hence by claim 3: $x_j \Vdash \Box\bot$. So $x_j \Vdash \Box\psi$.

Suppose $\varphi = \Delta\psi$. Note that: $x_i \Vdash \Delta\psi \Rightarrow x_{i+1} \Vdash \psi$, and $x_i \nVdash \Delta\psi \Rightarrow x_{i+1} \nVdash \psi$. Hence: $x_i \Vdash' \Delta\psi \Leftrightarrow x_{i+1} \Vdash' \psi \overset{IH}{\Leftrightarrow} x_{i+1} \Vdash \psi \Leftrightarrow x_i \Vdash \Delta\psi$.

With each x such that $\Box\bot \in x$ we associate a small lolly-model L_x as above. Define:

$K^* := \{x \in K \mid \neg\Box\bot \in x\} \cup \{\langle y, x\rangle \mid \Box\bot \in x \in K \text{ where } y \text{ is in the domain of } L_x\}$

$K_0{}^*:= \{\langle y,x\rangle \mid \Box\bot \in x \in K$ and y is in the domain of $L_x\}$

$K_1{}^*:= \{x \in K \mid \neg\Box\bot \in x\}$

$u\,R^*\,v :\Leftrightarrow$ (u,v are in $K_1{}^*$ and $u\,R\,v$) or

(u is in $K_1{}^*$, v is in $K_0{}^*$, where $v = \langle y,x\rangle$ and $u\,R\,x$)

$u\,S^*\,v :\Leftrightarrow$ $u\,R^*\,v$ or u is in $K_0{}^*$, v is in $K_0{}^*$, $u = \langle y,x\rangle$, $v = \langle z,x\rangle$ and

$y\,S'\,z$, where S' is the relevant relation of L_x.)

$u \Vdash^* p_i :\Leftrightarrow (u \in K_1{}^*$ and $p_i \in u)$ or $(u \in K_0{}^*$, $u = \langle y,x\rangle$ and $p_i \in y)$

$\mathcal{K}^* := \langle K^*, R^*, S^*, \Vdash^*\rangle$.

We claim:

A) $\mathcal{K}^*$ is a lolly-model

B) If $u \in K_1{}^*$, then $(u \Vdash^* \varphi \Leftrightarrow u \Vdash \varphi)$ for $\varphi \in X$. If $u \in K_0{}^*$, $u = \langle y,x\rangle$, then $(u \Vdash^* \varphi \Leftrightarrow y \Vdash \varphi)$ for $\varphi \in X$.

Proof of A: One easily verifies that K^* is finite and that R^* is transitive, irreflexive and hence upwards wellfounded. Moreover: $u \in K_1{}^* \Leftrightarrow$ there is a v in K^* such that $u\,R^*\,v$. The definition of S^* implies: $u \in K_1{}^* \Rightarrow (u\,S^*\,v \Leftrightarrow u\,R^*\,v)$. We leave to the reader the easy verification that $\langle [\langle y,x\rangle], S_0{}^* \upharpoonright [\langle y,x\rangle]\rangle$ is isomorphic to the lolly-frame part of L_x (for $\langle y,x\rangle$ in $K_0{}^*$). Clearly if $u\,R^*\,\langle y,x\rangle$ then $u\,R^*\,\langle z,x\rangle$ for all z in the domain of L_x. Also if $u \in K_0{}^*$ and $u\,S^*\,v$, then $v \in K_0{}^*$. □

Proof of B: by induction on φ in X, simultaneously for all u in K^*.

- If $u \in K_0{}^*$, $u = \langle y,x\rangle$: $u \Vdash^* \varphi \Leftrightarrow y \Vdash' \varphi$

$\Leftrightarrow y \Vdash \varphi$

(The first equivalence is by a completely trivial induction).

- Suppose $u \in K_1{}^*$. The case that φ is atomic is trivial and so are the cases of $\wedge, \vee, \rightarrow$ and $\neg$.

  - Suppose $\varphi = \Box\psi$

    - Suppose $u \Vdash \Box\psi$ and $u\,R^*\,v$. If v is in $K_1{}^*$, we have $u\,R\,v$, hence $v \Vdash \psi$, so by IH: $v \Vdash^* \psi$. If v is in $K_0{}^*$, say $v = \langle y,x\rangle$, we have uRx. Using claim 2, we show: uRy. It follows that $y \Vdash \psi$. Hence by IH: $v \Vdash^* \psi$. Conclude $u \Vdash^* \Box\psi$.

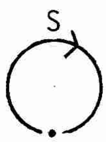
- Suppose $u \Vdash^* \Box\psi$ and uRy. If $(\neg\Box\bot) \in y$, we have uR^*y; hence $y \Vdash^* \psi$, so by IH: $y \Vdash \psi$. If $\Box\bot \in y$, then $uR^* \langle y,y\rangle$ and $\langle y,y\rangle \Vdash^* \psi$. Hence by IH: $y \Vdash \psi$. Conclude $u \Vdash \Box\psi$.

- The case that $\varphi = \Delta\psi$ is similar. □

Proof of 7.3.4: Suppose BMF $\nvdash \varphi$. Let X_0 be the smallest set that is closed under subformulas and contains φ and $\neg\Box\bot$. Let $X := X_0 \cup \{\Delta\psi \mid \Box\psi \in X_0\} \cup \{\Box\psi \mid \Delta\psi \in X_0\}$. Construct a finite lolly-model K^* as in 7.3.7 for X. By 7.3.6 there is an X-saturated $x_0 \subseteq X$ such that $x_0 \nvdash \varphi$. x_0 will correspond to a node of K^*, say u, and $u \nVdash^* \varphi$. □

7.3.8 Application

In 6 under mBM we showed that neither in BMF nor in PA there is an explicit Gödelsentence for Δ, where in PA Δ is interpreted as Δ^{mBM} or Δ^F or Δ^{mF}. For the case of BMF this fact can be easily shown by considering the following Kripke Model:



7.4 Conservativity of Fixed Point Equations over BMF

7.4.1 Definition

Let L be the language of BMF. We will consider extensions of L with finitely many constants. If $C = \{c_1, \ldots, c_n\}$, then L_C is the extension of L with $c_1, \ldots, c_n$, closed under the usual formation rules.

A constant c or a variable p is *modalized* in φ if *all* occurences of c or p are in the scope of Δ or $\Box$.

A list of equations E for $C = \{c_1, \ldots, c_n\}$ is a list of axioms of the form:

$$\vdash c_1 \leftrightarrow \varphi_1,$$
$$\vdots$$
$$\vdash c_n \leftrightarrow \varphi_n,$$

where the φ_i are in L_C and where the c_j are modalized in each of $\varphi_1, \ldots, \varphi_n$.

7.4.2 <u>Theorem</u>

Let E be a list of equations for C. Then for each φ in L:
$\text{BMF} + E \vdash \varphi \Rightarrow \text{BMF} \vdash \varphi$. In other words BMF + E is *conservative* over BMF w.r.t. L.

The rest of 7.4 will be devoted to the proof of 7.4.2.

7.4.3 <u>Definition</u>

Let K and K' be Kripke Models for L_C. We define:

$K \approx K'$:$\Leftrightarrow$ there is an $E \subseteq K \times K'$ such that:

i) for every $k \in K$ there is a $k' \in K'$ s.t. $k E k'$; for every $k' \in K'$ there is a $k \in K$ s.t. $k E k'$.

ii) $k_0 R k_1, k_0 E k_0' \Rightarrow$ there is a k_1' s.t. $k_1 E k_1'$ and $k_0' R' k_1'$.
$k_0' R' k_1', k_0 E k_0' \Rightarrow$ there is a k_1 s.t. $k_1 E k_1'$ and $k_0 R k_1$.

iii) similarly as in (ii) for S

iv) $k E k' \Rightarrow (k \Vdash p_i \Leftrightarrow k' \Vdash' p_i)$

Clearly $\approx$ is an equivalence relation.

7.4.4 Lemma

Consider K, K' with $K \approx K'$ and an E as in 7.4.3. Then we have: for each φ in L: $k E k' \Rightarrow (k \Vdash \varphi \Leftrightarrow k' \Vdash' \varphi)$.

Proof: entirely routine. □

To prove 7.4.2 it is clearly sufficient to provide for every lolly-model K for L a model K' for L_C that satisfies E, with $K \approx K'$. (Here we interpret K as a model for L_C by making e.g. all c_i false on K)

Let us first note that *if* we could solve the equations of E on the circles of a given lolly-model, *then* we can solve the equations on the rest of the model. Simply work downwards! Let S_T be the transitive closure of S: if the forcing relation is defined for C on all k' with $k S_T k'$, we set: $k \Vdash c_i :\Leftrightarrow k \Vdash \varphi_i$. This works because the c_j occur modalized in φ_i. Note that the procedure doesn't work *on* the circles!

So it is sufficient to solve the equations on a circle. This we cannot do in general, e.g. $\vdash c \leftrightarrow \neg\Delta c$, has no solution on:  S

What we can do is to blow up a given circle C to a new one C' in such a way that $C \approx C'$ and such that the equations of E are solved on C'. (I use 'circle' here as short for: lolly-model with as frame a circle.)

7.4.5 Definitions

i) For φ in L_C we define $n_C(\varphi)$, the nesting depth of the c_i's in φ as follows:

$n_C(\varphi) := 0$ if φ is an atom

$n_C(\psi \circ \chi) := \max(n_C(\psi); n_C(\chi))$ if $\circ \in \{\wedge, \vee, \rightarrow\}$

$n_C(\neg\psi) := n_C(\psi)$

$$n_C(\Delta\psi) := \begin{cases} 0 \text{ if no } c_i \text{ occurs in } \psi \\ n_C(\psi) + 1 \text{ otherwise} \end{cases}$$

$$n_C(\Box\psi) := \begin{cases} 0 \text{ if no } c_i \text{ occurs in } \psi \\ 1 \text{ otherwise} \end{cases}$$

ii) Consider a frame $F = \langle M,T,U \rangle$, where $T \subseteq U$ and $(m\,T\,m'\,U\,m'' \Rightarrow m\,T\,m'')$.

A *path* π between m,m' is a sequence: $m = m_1\,U\,m_2\,U\,m_3 \ldots U\,m_n = m'$. The *length* of this path is n-1.

The asymmetric distance $d : M \times M \to \{0,1, \ldots, \infty\}$ is defined by:

$d(m,m') := \inf\{\text{length of } \pi \mid \pi \text{ is a path between } m,m'\}$

Note: $d(m,m) = 0$, $d(m,m'') \leq d(m,m') + d(m',m'')$ (assuming: $n + \infty = \infty + n = \infty + \infty = \infty$).

$B(m,N) := \{m' \in M \mid d(m,m') \leq N\}$.

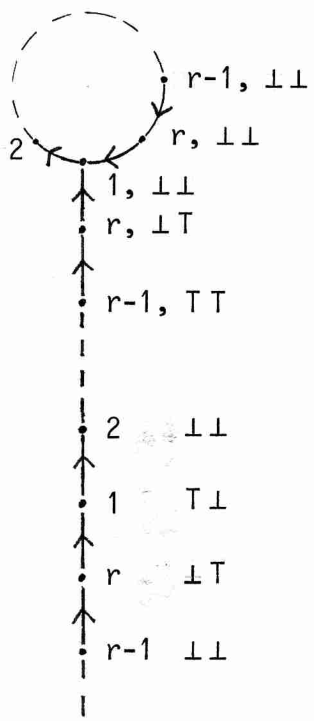
7.4.6 Seeing Lemma

This lemma refines 7.4.4 a bit. Consider $M = \langle M,T,U, \Vdash \rangle$ and $M' = \langle M',T',U', \Vdash' \rangle$, where $\langle M,T,U \rangle$ and $\langle M',T',U' \rangle$ satisfy the conditions of 7.4.5 (ii). Consider φ in L_C with $n_C(\varphi) = N$, and m_0, m_0' respectively in M,M'. Suppose there is a relation E satisfying the conditions of 7.4.3. Suppose E is such that $m_0\,E\,m_0'$ and for all $m_1 \in B(m_0,N)$, for all $m_1' \in B(m_0',N)$ with $m_1\,E\,m_1'$, we have $(m_1 \Vdash c_i \Leftrightarrow m_1' \Vdash' c_i)$ $(i = 1, \ldots, n)$. Then: $m_0 \Vdash \varphi \Leftrightarrow m_0' \Vdash' \varphi$.

Proof: induction on φ. Suppose e.g. $\varphi = \Box\psi$ and some c_i occurs in ψ. Suppose $m_0 \Vdash \Box\psi$ and $m_0'\,T'm_1'$. There is an m_1 such that $m_1\,E\,m_1'$, and $m_0\,T\,m_1$. Let $n_C(\psi) = M$. Clearly $B(m_1,M) \subseteq B(m_0,1)$ and $B(m_1',M) \subseteq B(m_0',1)$. So we apply the IH to ψ and get: $m_1 \Vdash \psi \Leftrightarrow m_1' \Vdash' \psi$. Conclude $m_1' \Vdash' \psi$. Hence $m_0' \Vdash' \Box\psi$. The rest is easier or similar. □

We turn to the construction of C' from C. Let a circle C be given with a forcing relation for L. We stipulate arbitralily that the c_i are false on C. Suppose the domain of $C = \{1, \ldots, r\}$. We construct an auxiliary model $\mathcal{D}$ with $\mathcal{D} \approx C$, as follows: we hang a tail of copies w.r.t. the forcing relation on (the atoms of) L of the nodes of C under C, repeating over and over again the order of the nodes of C.

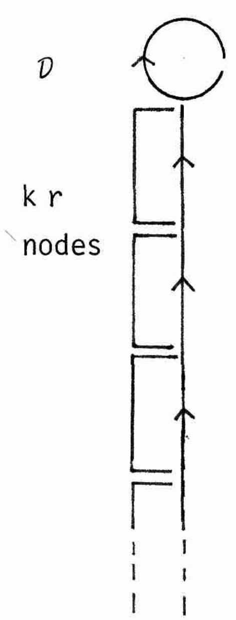
$\mathcal{D}$



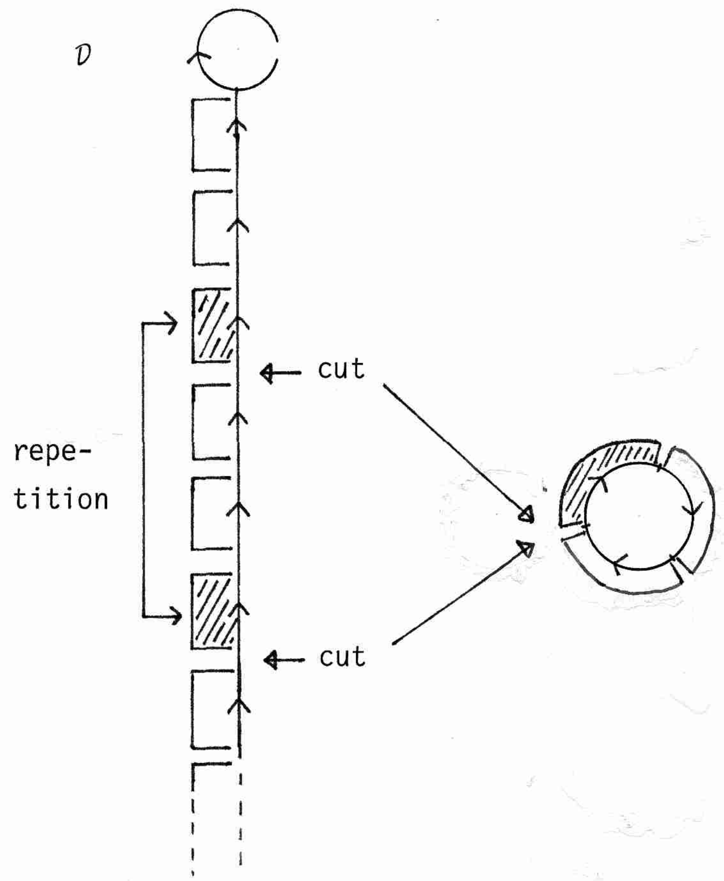
The truth values of c_1, c_2 are shown at each node.

We assign values to the c_i on the nodes of the tail in the obvious way: if we have given values to the c_i in all nodes above α we set: $\alpha \Vdash c_i :\Leftrightarrow \alpha \Vdash \varphi_i$. This definition is sensible because the c_j occur modalized in φ_i.

Let $N := \max \{n_C(\varphi_i) \mid i = 1, \ldots, n\}$ and let k be the smallest number such that $kr \geq N$. We look at stretches of k copies of $1, \ldots, r$ in the tail as in the next picture:



On each such stretch we will have a certain sequence of truthvalue assignments for the $c_1, \ldots, c_n$. Of course at a certain point (going downwards) we will meet a sequence we have already seen. Now we cut the tail directly *below* the first stretch that is repeated and directly *below* its first repetition. See the next picture. Then we connect the ends, leaving the assignments to $c_1, \ldots, c_n$ intact. The new model so obtained will be our desired model C'.



Clearly $C' \approx C$. Moreover consider the relation E between the nodes of $\mathcal{D}$ and those of C' that matches nodes of the stretch that was cut out to the corresponding nodes of C' and is then uniquely extended to satisfy the demands of definition 7.4.3. As is easily seen: for each φ_i and for each α, α', where α is on the stretch that was cut out and $\alpha E \alpha'$, the hypothesis of the Seeing Lemma is fulfilled, so $\alpha \Vdash \varphi_i \Leftrightarrow \alpha' \Vdash' \varphi_i$.

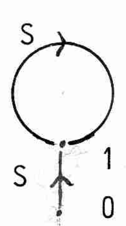
We conclude:

$\alpha' \Vdash' c_i \Leftrightarrow \alpha \Vdash c_i \Leftrightarrow \alpha \Vdash \varphi_i \Leftrightarrow \alpha' \Vdash' \varphi_i$.

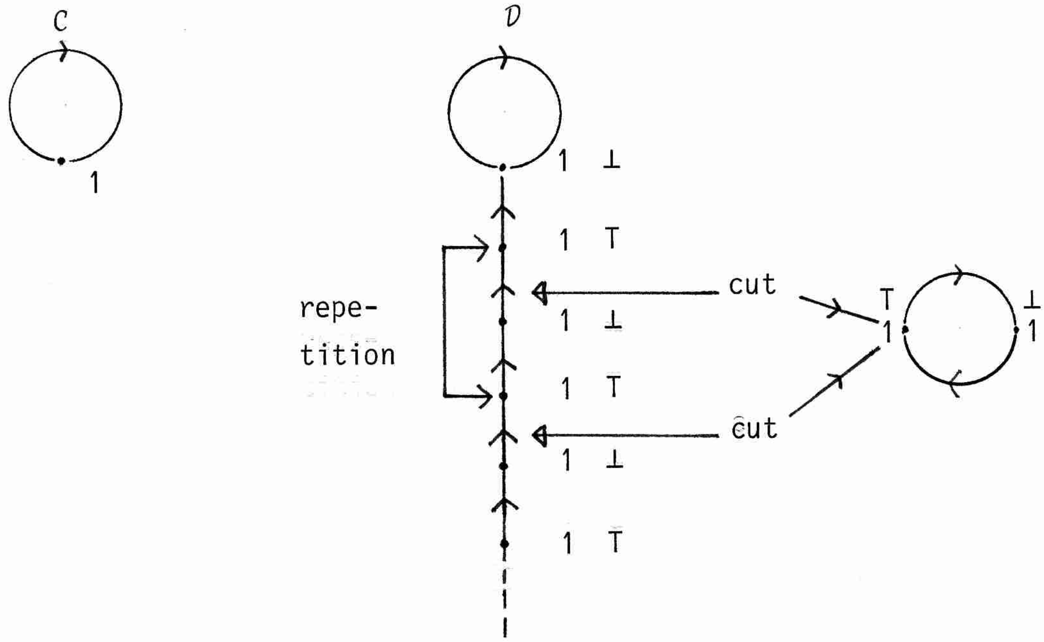
Hence $\alpha' \Vdash' c_i \leftrightarrow \varphi_i$.

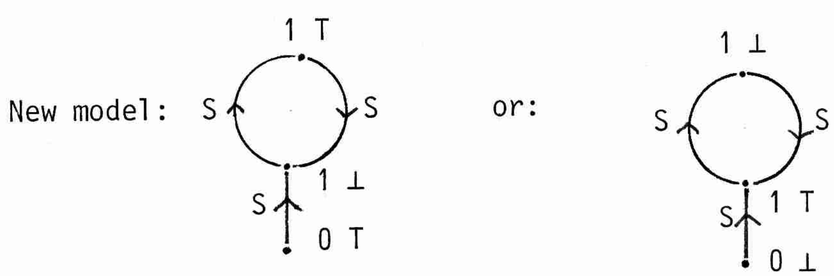
Our final construction is this. We start with lolly-model K. We blow up every circle C in K to a new circle C' with $C \approx C'$ in which the equations of E are solved. We attach the new circles C' to their sticks at some point in C' corresponding via some E that witnesses $C \approx C'$ to the original point to which the stick of C was attached. We extend R in the obvious way and we extend the forcing relation to the c_i as described above. Clearly our new model K' satisfies: $K' \approx K$. Moreover the equations of E are solved in K'.

Consider for example E with as sole member: $\vdash c \rightarrow \neg \Delta c$, and consider the model:



We have N = r = k = 1. Thus the construction is as follows:





8 Embedding circle-tail models in Arithmetic

We would like to generalize the result of Solovay[1976] to the logic BMF, interpreting Δ as Δ^{MF}. To do this we must embed lolly-models in Arithmetic. This program however meets with a difficulty I couldn't solve: in a nutshell the problem is how to handle the sticks of the lollies. It turns out that *if the sticks are absent* a straightforward embedding is possible. For the record I state the obvious open problem:

STICK PROBLEM: Can lolly-models be embedded in Arithmetic?

Even if we do not achieve arithmetical completeness for BMF it seems to me that the partial result proved here is of interest: the Embedding Theorem gives us a powerful machinery to construct arithmetical sentences (see also §9). Moreover the methods employed add to our experience with Solovay style arguments: we have the first example here of an embedding of structures that are not (completely) upwards well-founded.

In this § I follow the presentation of Visser[1984].

To get a true embedding of circle-models in Arithmetic we must add a tail to the circle-models: consider a finite circle-model. We hang a down-going $\omega+1$-tail (in R) under it as in the picture below. We can arrange it so that the nodes of the finite model at the top are numbered: 1,...,N , and the nodes of the down-going tail except bottom: N+1,N+2,... ; finally the bottom: 0 . The nodes numbered N+1,N+2,... 0 will be called: tailelements. We stipulate that at each of the nodes only finitely many atoms are forced and that on all elements of the tail including 0 the same atoms are forced.

We call the resulting models circle-tailmodels. Clearly a circle-tailmodel *is* a circle-model.

An immediate consequence of our definition is:

8.1 Tail Lemma

$0 \Vdash \phi \Leftrightarrow$ there is a k such that for all m>k $m \Vdash \phi$
$0 \nVdash \phi \Leftrightarrow$ there is a k such that for all m>k $m \nVdash \phi$

Proof: a simple induction on ϕ. □

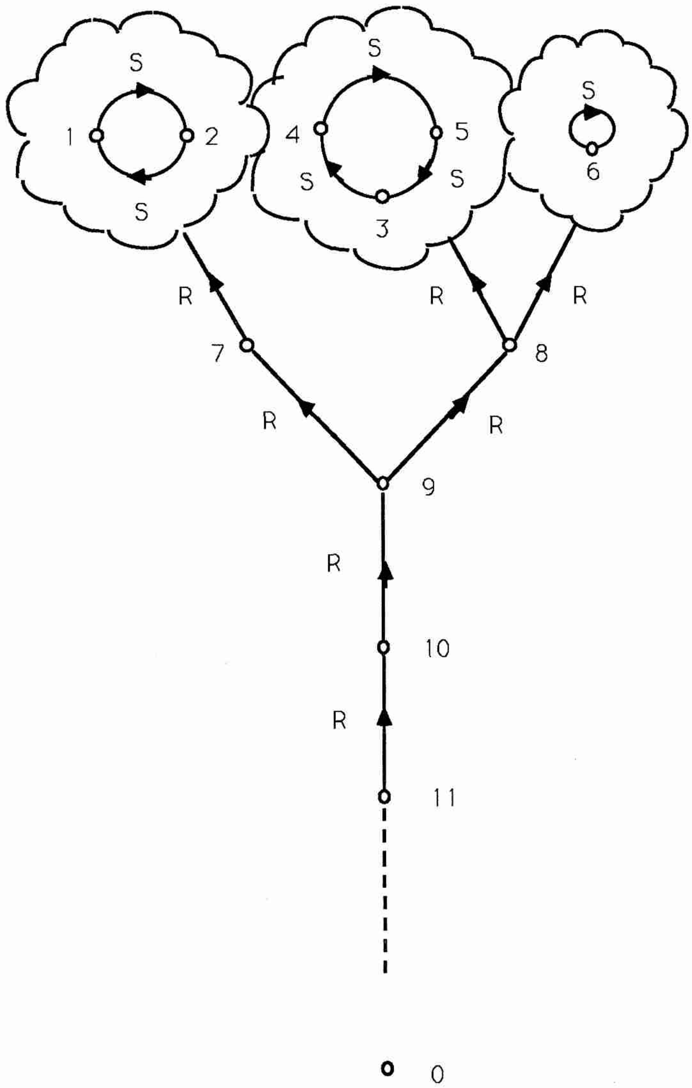
Let $[\![\phi]\!] := \{k \mid k \Vdash \phi\}$. Then by the Tail Lemma $[\![\phi]\!]$ is either finite or cofinite.

Note that circle-tailmodels satisfy the principle:

C $\vdash \Box(\Box\bot \rightarrow \Delta\phi) \rightarrow \Box(\Box\bot \rightarrow \phi)$

I would be very surprised if C were arithmetically valid. A lolly-model to refute C is easily found.

OPEN PROBLEM: Is C arithmetically valid when Δ is interpreted as Δ^{mF}?



For the rest of this §:

i) We fix a circle-tailmodel K .

ii) We assume that K is suitably described in Arithmetic. Specifically we assume that R and S are given Δ_0 definitions in such a way that all their simple properties are verifiable.

iii) We interpret Δ as Δ^{mF} in arithmetical contexts.

iv) '$\vdash$' stands for: PA$\vdash$.

v) We assume that 'Proof' satisfies the following plausible assumptions:

$$\vdash \Box A \rightarrow \forall x \exists y{>}x\,\mathrm{Proof}(y,A)$$
$$\vdash \forall u \forall v(\ \mathrm{Proof}(u,v) \rightarrow v{<}u\ )$$

We define a variant of Solovay's reluctant function: the function that dares not go anywhere for fear of having to stay. Our variant will not dare to go anywhere for fear of coming there too often.

Define by the Recursion Theorem:

$$\mathrm{COF}a \quad :\Leftrightarrow \quad \forall x \exists y{>}x \; hy=a$$

$$h0 \quad := \quad 0$$

$$h(k+1) \quad := \quad \begin{cases} a & \text{if for some } a \text{ with } hkRa \; \mathrm{Proof}(k,\neg\mathrm{COF}a) \\ a & \text{if for some } a \text{ with } hkSa \; \mathrm{Proof}(k,\neg\mathrm{COF}a) \\ & \text{and } (\neg\mathrm{COF}a)\epsilon mF_{k+1} \\ hk & \text{otherwise} \end{cases}$$

It is easy to see that the arithmetization of '$(\neg\mathrm{COF}a)\epsilon mF_{k+1}$' is Δ_2 and hence that h is Δ_2 .

An important difference with Solovay's original construction is that we use 'COFa' instead of '1=a'. Later we will see that $\vdash$ COFa $\leftrightarrow$ 1=a ; but to show this we need that h is defined using 'COF' rather than '1'.

We prove a sequence of lemmas about h .

8.2 Lemma

Let S* be the transitive reflexive closure of S, then:

$$\vdash \forall x \forall y \; (\; x \leq y \rightarrow hxS^*hy \;)$$

Proof: by a trivial induction on z with x+z=y. □

8.3 Lemma

i) $\vdash (\; (\forall z{<}x \; hz\epsilon K_1) \wedge hx=y \;) \rightarrow \Delta hx=y$

ii) $\vdash \forall x \forall y \; (\; hx=y \rightarrow \Box hx=y \;)$

Proof:

i) Reason in PA:

The proof is by induction on x. The case that $x=\underline{0}$ is trivial. Suppose x=u+1, $\forall z{<}x \; hz\epsilon K_1$, hx=y and hu=v. There are three possibilities:

a) hx was computed by the first clause of the definition of h. We have vRy and Proof(u,¬COFy). Hence: ΔvRy and ΔProof(u,¬COFy) By the Induction Hypothesis: Δhu=v. Conclude Δhx=y.

b) hx was computed by the second clause of the definition of h. We have: vSy and Proof(u,¬COFy). Because $v\epsilon K_1$, we also have: vRy. So hx was also computed by the first clause.

c) hx was computed by the third clause. Clearly y=v. Either for no w Proof(u,¬COFw) or for some w Proof(u,¬COFw) and not uRw. (We may ignore the second clause by the reasoning under b.) Hence we have $\Delta\forall w{<}u\neg$Proof(u,¬COFw) or (ΔProof(u,¬COFw) and $\Delta\neg$vRw). By the Induction Hypothesis: Δhu=v. Conclude:

$\Delta hx=y$. □(i)

ii) The proof is by an easy induction in PA, using the fact that:

$\vdash \forall u \forall v\ (\ u \in mF_v \rightarrow \square(u \in mF_v)\ )$ □(ii)

We define LIMa $:\Leftrightarrow \exists x\ hx=a \wedge \forall x \forall y(\ (hx=a \wedge x \leq y) \rightarrow hy=a\ )$.

8.4 Lemma

$\vdash \forall a\ (\ COFa \rightarrow LIMa\ )$

Proof: By 8.2 clearly: $\vdash \forall a\ (\ (COFa \wedge a \in K_1) \rightarrow LIMa\ )$, so it is sufficient to show: $\vdash \forall a\ (\ (COFa \wedge a \in K_0) \rightarrow LIMa\ )$. Reason in PA:

Suppose COFa and $a \in K_0$. Assume a is on the circle C with, say, $a=\underline{a}_1 S \underline{a}_2 S ... S \underline{a}_n S \underline{a}_{n+1}=a$. Let x_0 be the unique number such that $hx_0 \in K_1$ and $h(x_0+1) \in K_0$. Clearly $h(x_0+1)=\underline{a}_j$ for some j. By 8.3(i): $\Delta h(x_0+1)=\underline{a}_j$. Conclude: $\Delta \bigvee \{COF\underline{a}_i \mid i=1,...,n\}$.

Now suppose for a reductio that ¬LIMa. Clearly by 8.2: $COF\underline{a}_1$, $COF\underline{a}_2$, ... , $COF\underline{a}_n$. It follows from the definition of h that: $\Delta\neg COF\underline{a}_1$, $\Delta\neg COF\underline{a}_2$, ... , $\Delta\neg COF\underline{a}_n$ (or how else could h move on and on?) . Hence: $\Delta \bigvee \{COF\underline{a}_i \mid i=1,...,n\}$ and $\Delta \bigwedge \{\neg COF\underline{a}_i \mid i=1,...n\}$, ergo $\Delta\bot$ and thus $\bot$. Conclude LIMa. □

8.5 Lemma

$\vdash \exists a\ LIMa$

Proof: It is easily seen that $\vdash \exists a\ COFa$. □

8.6 Lemma

$\vdash \exists x\ hx \in K_0 \leftrightarrow \square\bot$

Proof: Reason in PA:

"←" Trivial.

"→" Suppose $hx=\underline{a}_1$, where $\underline{a}_1$ is on the circle C, given by:

$\underline{a}_1 S \underline{a}_2 S ... S \underline{a}_n S \underline{a}_{n+1}=\underline{a}_1$.

We have $\square hx=\underline{a}_1$ by 8.3(ii), hence: $\square \bigvee \{COF\underline{a}_i \mid i=1,...,n\}$. h moved up to $\underline{a}_1$ by the first or by the second clause. In either case we have: $\square\neg COF\underline{a}_1$.

We show for k=0,...,n-1: $\bigwedge \{\square\Delta^k\neg COF\underline{a}_j \mid j=1,...,k+1\}$, by (external) induction on k. The case that k=0 is simply $\square\neg COF\underline{a}_1$. Suppose: $\bigwedge \{\square\Delta^k\neg COF\underline{a}_j \mid j=1,...k+1\}$. By B11: $\bigwedge \{\square\Delta^{k+1}\neg COF\underline{a}_j \mid j=1,...,k+1\}$. We show: $\square\Delta^{k+1}\neg COF\underline{a}_{k+2}$. Clearly:

$\square(\ (hx=\underline{a}_1 \wedge \bigwedge \{\neg COF\underline{a}_j \mid j=1,...,k+1\}) \rightarrow \exists y \geq x\ hy=\underline{a}_{k+2}\ )$.

Hence:

$\Box(\,(hx=\underline{a}_1 \wedge \bigwedge\{\neg COF\underline{a}_j \mid j=1,...,k+1\}) \rightarrow \Delta\neg COF\underline{a}_{k+2}\,)$.

Conclude using L1, L2, B1, B2, B11:

$(\Box\Delta^k hx=\underline{a}_1 \wedge \bigwedge\{\Box\Delta^k\neg COF\underline{a}_j \mid j=1,...k+1\}) \rightarrow \Box\Delta^{k+1}\neg COF\underline{a}_{k+2}$.

Moreover by B11 we have from $\Box hx=a_1$: $\Box\Delta^k hx=\underline{a}_1$. So finally:

$\Box\Delta^{k+1}\neg COF\underline{a}_{k+2}$.

We have found: $\bigwedge\{\Box\Delta^{n-1}\neg COF\underline{a}_j \mid j=1,...,n\}$. On the other hand we have: $\Box\bigvee\{COF\underline{a}_j \mid j=1,...,n\}$, hence: $\Box\Delta^{n-1}\bigvee\{COF\underline{a}_j \mid j=1,...,n\}$. Combining we find: $\Box\Delta^{n-1}\bot$ and hence: $\bot$. □

Consider i in K_0. We call the S-successor of i: σi , and the S-predecessor: πi.

8.7 Lemma

i) $\vdash (COFu \wedge uSv) \rightarrow \nabla COFv$

ii) $\vdash (y\in K_0 \wedge COFy) \rightarrow \Delta COF\sigma y$

iii) $\vdash (y\in K_0 \wedge \Box\bot) \rightarrow (\Delta COF\sigma y \rightarrow COFy)$

iv) $\vdash (y\in K_0 \wedge \Box\bot) \rightarrow (COFy \leftrightarrow \Delta COF\sigma y)$

Proof:

i) Reason in PA:

Suppose COFu, uSv and $\Delta\neg COFv$. By 8.4: LIMu. Suppose hx=u and for all $y\geq x$ hx=u. For some z $(\neg COFv)\in mF_z$. Consider w with w>x, w>z and Proof(w,¬COFv). Clearly $(\neg COFv)\in mF_{w+1}$. Hence h would move up to v at z+1. Quod non. Conclude $\neg\Delta\neg COFv$. □(i)

ii) Immediate from (i) using $\vdash (y\in K_0 \wedge COFy) \rightarrow \Box\bot$, which follows directly from 8.6, and $\vdash \Box\bot \rightarrow (\nabla A \leftrightarrow \Delta A)$. □(ii)

iii) Reason in PA:

Suppose $y\in K_0$, $\Box\bot$ and $\Delta COF\sigma y$. From $\Box\bot$ we have by 8.6 that for some $z\in K_0$: COFz. By (ii): $\Delta COF\sigma z$. Hence by 8.4, B1, B2 and $\Delta COF\sigma y$: $\Delta y=z$ and thus y=z. (We have Π_1-Reflection for Δ!) □(iii)

iv) By (ii) and (iii). □(iv)

8.8 Definition

i) Let f be a function from the propositional variables of the language of BMF to the sentences of PA. We define $(\ )^f$ from the formulas of the language of BMF as follows:

- $(p_i)^f := f(p_i)$
- $(\ )^f$ commutes with the propositional connectives (including $\top,\bot$)
- $(\Box\phi)^f := \Box(\phi)^f$ (Note that '□' shifts its meaning!)
- $(\Delta\phi)^f := \Delta(\phi)^f$

ii) Consider ϕ in the language of BMF. If $[\![\phi]\!]$ is finite, we set:

$[\phi] := \bigvee\{COF\underline{i} \mid i\Vdash\phi\}$ (We take $\bigvee\emptyset := (\underline{0}=\underline{1})$.)

If $[\![\phi]\!]$ is cofinite, we set:

$[\phi] := \bigwedge\{\neg COF\underline{i} \mid i \nVdash \phi\}$ (We take $\bigwedge\emptyset := (\underline{0}=\underline{0}.)$)

Note that $[\phi]$ is simply an arithmetization of: $\exists x \in [\![\phi]\!]\ COFx$.

iii) Define $Fp_i := [p_i]$, and $\langle\phi\rangle := (\phi)^F$.

8.9 Embedding Theorem

$$\vdash \langle\phi\rangle \leftrightarrow [\phi]$$

Proof: It is clearly sufficient to show in PA that [ ] 'commutes' with the logical constants including $\Box$ and Δ. The cases of the propositional constants are trivial (using 8.4). We show:

i) $\vdash [\Box\psi] \leftrightarrow \Box[\psi]$

ii) $\vdash [\Delta\psi] \leftrightarrow \Delta[\psi]$

Proof of (i): In case $\{i \mid i \Vdash \Box\psi\}$ is infinite, we have: $[\![\Box\psi]\!]=[\![\psi]\!]=\omega$, hence $[\Box\psi]=[\psi]=(\underline{0}=\underline{0})$. It follows that $\vdash [\Box\psi] \leftrightarrow \Box[\psi]$. Suppose $\{i \mid i \Vdash \psi\}$ is finite. Reason in PA:

"$\leftarrow$" Let $\underline{j}_1,\ldots,\underline{j}_s$ be the complete set of nodes such that $\underline{j}_k \Vdash \Box\psi$ and $\underline{j}_k \nVdash \psi$. Suppose $\Box[\psi]$. Clearly $\Box\neg COF\underline{j}_k$. Suppose $Proof(p,\neg COF\underline{j}_k)$ and $hp=y$. There are two possibilities:

Case 1: $yR\underline{j}_k$. If $yR\underline{j}_k$, clearly $h(p+1)=\underline{j}_k$.

Case 2: $\neg yR\underline{j}_k$. It follows that if $COFx$, then $\neg xR\underline{j}_k$, for if we had $yS^*xR\underline{j}_k$, it would follow that $yR\underline{j}_k$.

In both cases: $COFx \to \neg xR\underline{j}_k$.

On the other hand it is easily seen that if $x \nVdash \Box\psi$ then $xR\underline{j}_k$ for some k. Hence: $COFx \to x \Vdash \Box\psi$. Conclude $[\Box\psi]$

"$\to$" Suppose $COF\underline{i}$ for some $\underline{i}$ with $\underline{i} \Vdash \Box\psi$. Because $\underline{i} \neq \underline{0}$, h must have moved up to $\underline{i}$ at a certain point by clause 1 or clause 2 of the definition of h. In either case we have $\Box\neg COF\underline{i}$. Suppose $hx=\underline{i}$. By 8.3(ii): $\Box hx=\underline{i}$. If $\underline{i} \in K_0$ we have by 8.6: $\Box\bot$, and hence $\Box[\psi]$. If $\underline{i} \in K_1$ we see that $\Box\neg COF\underline{i}$ and $\Box hx=\underline{i}$ imply: $\Box\forall y(COFy \to \underline{i}Ry)$. Conclude: $\Box[\psi]$. □(i)

Proof of (ii): Clearly $\vdash [\neg\Box\bot \to (\Delta\psi \leftrightarrow \Box\psi)]$, hence $\vdash \neg\Box\bot \to ([\Delta\psi] \leftrightarrow \Box[\psi])$ by the fact that [ ] 'commutes' with the propositional connectives and $\Box$. Also: $\vdash \neg\Box\bot \to (\Delta[\psi] \leftrightarrow \Box[\psi])$. So we may conclude: $\vdash \neg\Box\bot \to ([\Delta\psi] \leftrightarrow \Delta[\psi])$.

To complete the argument we show: $\vdash \Box\bot \to ([\Delta\psi] \leftrightarrow \Delta[\psi])$:

$$\begin{aligned}
\vdash \Box\bot \to ([\Delta\psi] &\leftrightarrow ([\Delta\psi] \wedge \Box\bot) \\
&\leftrightarrow [\Delta\psi \wedge \Box\bot] \\
&\leftrightarrow \bigvee\{COF\underline{j} \mid j \Vdash \Delta\psi \wedge \Box\bot\} \\
&\leftrightarrow \bigvee\{COF\pi\underline{i} \mid i \Vdash \psi \wedge \Box\bot\} \\
&\leftrightarrow \bigvee\{\Delta COF\underline{i} \mid i \Vdash \psi \wedge \Box\bot\} && (8.7(iv)) \\
&\leftrightarrow \Delta\bigvee\{COF\underline{i} \mid i \Vdash \psi \wedge \Box\bot\} && (B12) \\
&\leftrightarrow \Delta[\psi \wedge \Box\bot] \\
&\leftrightarrow \Delta[\psi]\ ) && (B1,B2,B4)\quad \Box(ii)
\end{aligned}$$

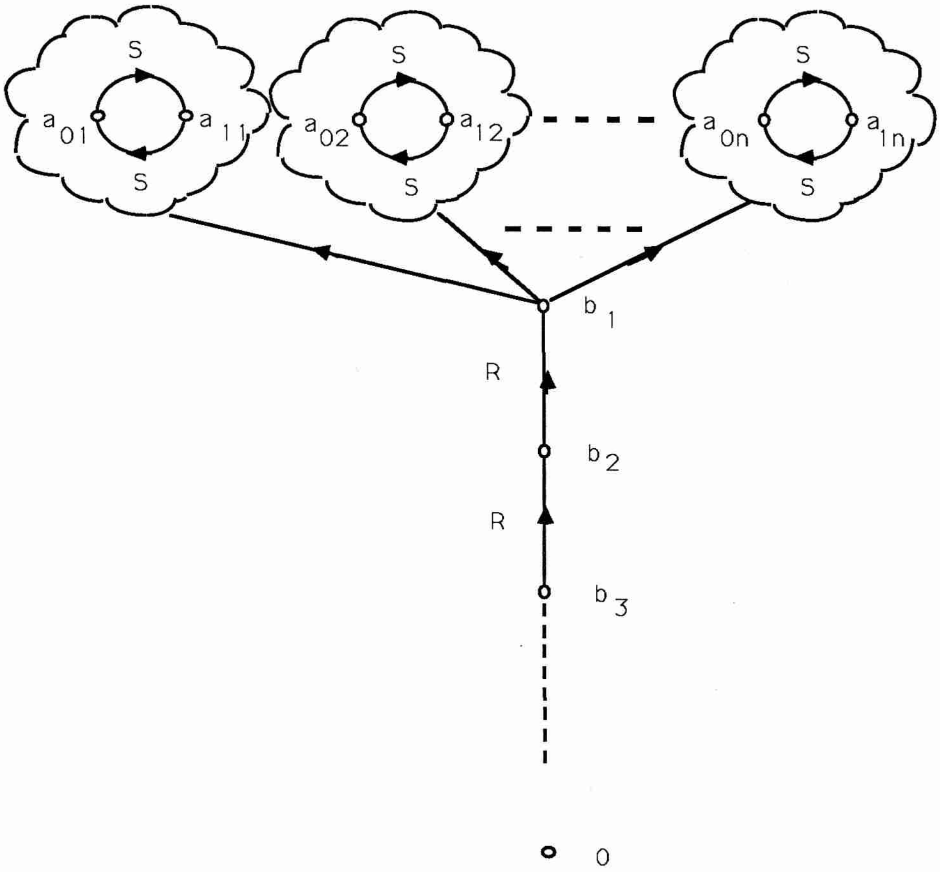
8.10 Remark

The reduction result proved as 33 in §6 clearly applies to Δ^{mF}. It implies that for the arithmetical embedding of *traditional* tailmodels we have: $\vdash \Delta[\phi] \leftrightarrow (\Box[\phi] \wedge (\Box\bot\rightarrow[\phi]))$. We can now understand this result in a new way: the arithmetical embedding of traditional tailmodels is similar to the arithmetical embedding of circle-tailmodels *which have just singleton circles* ! (This point will become even clearer in the light of lemma 9.3.)

8.11 Application

There are infinitely many non-equivalent Gödelsentences for Δ^{mF}.

Proof: It is clearly sufficient to prove that for any n there are n non-equivalent Gödelsentences for Δ. Consider the following circle-tailmodel:



Let s be a sequence $c_1c_2...c_n$ of o's and 1's. Consider an atom p_s. Let:

$a_{0i}\Vdash p_s \;:\Leftrightarrow\; c_i=0$

$a_{1i}\Vdash p_s \;:\Leftrightarrow\; c_i=1$

$b_j\Vdash p_s$ for all j

$0\Vdash p_s$

Define: $G_s := [p_s]$

It follows immediately from the Embedding Theorem that:

$$\vdash G_s \leftrightarrow \neg\Delta G_s$$

Moreover if $s \neq s'$:

$$\vdash \Box(G_s \leftrightarrow G_{s'}) \rightarrow \Box\bot$$

Because Gödelsentences of Δ are Oreysentences it follows that there are infinitely many non-equivalent Oreysentences.

9 Δ^{mF} meets Relative Interpretability

In this § 'Δ' will stand for 'Δ^{mF}' in arithmetical contexts. '$\vdash$' will stand for 'PA$\vdash$'. We fix a circle-tailmodel K .

For convenience we repeat the derivability conditions we collected for Relative Interpretability in 5.5 here:

- I1 $\vdash \Box(B\rightarrow A) \rightarrow A\lhd B$
- I2 $\vdash (A\lhd B \wedge B\lhd C) \rightarrow A\lhd C$
- I3 $\vdash (A\lhd B \wedge A\lhd C) \rightarrow A\lhd(B\vee C)$
- I4 $\vdash A\lhd B \rightarrow (\Diamond B\rightarrow\Diamond A)$
- I5 $\vdash \Diamond A\lhd B \rightarrow \Box(B\rightarrow\Diamond A)$
- I6 $\vdash A\lhd\Diamond A$
- I7 $\vdash A\lhd B \rightarrow (A\wedge\Box C)\lhd(B\wedge\Box C)$
- J1 for all P in Π_1: PA$\vdash$ $P\lhd B \rightarrow \Box(B\rightarrow P)$
- J2 for all S in Σ_1: PA$\vdash$ $A\lhd B \rightarrow (A\wedge S)\lhd(B\wedge S)$

We add the for our purposes essential 34 of §6:

- J3 $\vdash A\lhd\nabla A$

Note that I5, I6, I7 are redundant in our present list.

We list some immediate consequences of our list:

- J4 $\vdash A\lhd\Delta A$ (B1,B2,B3,I1,J3,I2)

Define: $A\equiv B :\Leftrightarrow A\lhd B\wedge B\lhd A$

- J5 $\vdash (A\equiv A'\wedge B\equiv B') \rightarrow (A\lhd B\leftrightarrow A'\lhd B')$ (I2)
- J6 $\vdash \bigwedge\{A_i\lhd B_i \mid i=1,\dots,n\} \rightarrow \bigvee\{A_i \mid i=1,\dots,n\}\lhd\bigvee\{B_i \mid i=1,\dots,n\}$ (I1,I2,I3)
- J7 $\vdash B\equiv(B\vee\Diamond B)$ (I1,I6,I3)
- J8 If $P\in\Pi_1$, then: $\vdash \Box((B\vee\Diamond B)\leftrightarrow P) \rightarrow (\,B\lhd C \leftrightarrow \Box(C\rightarrow(B\vee\Diamond B))\,)$ (I1,J5,J7,J1)

We want to take a closer look at the interaction between $\lhd$ and the sentences $[\phi]$ constructed in §8. The classes of sentences $[\phi]$, constructed for different circle-tailmodels, are too poor to refute all modal principles in the language with $\Box$ and $\lhd$ *not* valid in PA. For example Per Lindström has shown that there is a Σ_1 sentence A such that

$$\nvdash A\lhd\top \rightarrow \Box(A\lhd\top)$$

On the other hand we will see that:

$$\vdash [\phi]\lhd\top \rightarrow \Box([\phi]\lhd\top)$$

This weakness however turns out to be a strength: $[\phi]\lhd\top$ *reduces* to a simpler formula. (We encountered the phenomenon of reduction before in connection with Feferman's Predicate.)

We define an ad hoc modal operator $(\ )^*$ as follows: $[\![\phi^*]\!]$ is the smallest set X such that $[\![\phi]\!]\subseteq X$ and if $j\in X\cap K_0$, then $\sigma j\in X$. In other words $[\![\phi^*]\!]$ is obtained by adding all circles C such that $C\cap[\![\phi]\!]\neq\emptyset$ to $[\![\phi]\!]$.

9.1 Reduction Theorem

$$\vdash [\phi]\lhd A \leftrightarrow \Box(A\rightarrow([\phi^*]\vee\Diamond[\phi^*]))$$

To prove 9.1 we need a few lemmas.

9.2 definition

We define a *recursive* function h_0 as follows:

$$h_0 0 := 0$$

$$h_0(k+1) := \begin{cases} a & \text{if for some a with } h_0kRa \text{ Proof}(k,\neg\text{COF}a) \\ hk & \text{otherwise} \end{cases}$$

Here 'COF' is as in §8. Note that COF is based on h and not on h_0 !

9.3 Lemma

i) $\vdash (\forall z{<}x\ hz\in K_1) \rightarrow hx=h_0x$

ii) Let S* be the transitive, reflexive closure of S, then:
$\vdash \forall x\ h_0xS^*hx$

Proof: The proof is in both cases by a simple induction on x in PA. These inductions are much like the proof of 8.3(i). □

9.4 Corollary

$[\phi]$ is Δ_2.

Proof: It is clearly sufficient to show that sentences of the form $\text{COF}\underline{i}$ are Δ_2. In case $i\in K_0$ we have by 8.9 : $\vdash \text{COF}\underline{i} \leftrightarrow \Delta\text{COF}\sigma\underline{i}$, hence $\text{COF}\underline{i}$ is in Δ_2. In case $i\in K_1$ we have by 8.4 and 9.3:

$$\vdash \text{COF}\underline{i} \leftrightarrow (\ \exists x h_0x=\underline{i} \wedge \forall x\forall y((h_0x=\underline{i}\wedge x\leq y)\rightarrow h_0y=\underline{i})\ ) \quad \Box$$

9.5 Definition

Consider $X\subseteq K$. We call X *upwards persistent* if: $(i\in X$ and $iSj) \Rightarrow j\in X$.

9.6 Lemma

Suppose $[\![\phi]\!]$ is upwards persistent then $[\phi]$ is provably equivalent to a Σ_1 sentence.

Proof: In case $[\![\phi]\!]$ is infinite this is trivial. Suppose $[\![\phi]\!]$ is finite, we show:

$$\vdash [\phi] \leftrightarrow \bigvee\{\exists x h_0 x=\underline{i} \mid i \Vdash \phi\}$$

Reason in PA:

"$\leftarrow$" Suppose $h_0 x=\underline{i}$ for $\underline{i}\in[\![\phi]\!]$. $\underline{i}S^*hx$ by 9.3(ii). Hence by the upwards persistence of $[\![\phi]\!]$: $hx\in[\![\phi]\!]$. Thus: $\forall z>x\ hz\in[\![\phi]\!]$. Conclude: $[\phi]$.

"$\rightarrow$" Suppose $COF\underline{i}$ for $\underline{i}\in[\![\phi]\!]$. In case $\underline{i}\in K_1$ we have by 9.3(i): $\exists x h_0 x=\underline{i}$. Suppose $\underline{i}\in K_0$. Say $\underline{i}$ is on circle C . Clearly there is a u on C and a y such that $hy=u$ and for all $z<y$ $hz\in K_1$. By 9.3(i) $h_0 y=u$. $[\![\phi]\!]$ is upwards persistent, $\underline{i}$ is in $[\![\phi]\!]$, $\underline{i}$ is on C , hence $C \subseteq[\![\phi]\!]$. Conclude $u\in[\![\phi]\!]$. Ergo: $\exists y\ h_0 y\in[\![\phi]\!]$. □

9.7 Lemma

Suppose $\underline{i}$ is on circle C . Then: $\vdash COF\underline{i} \lhd \bigvee\{COF\underline{j} \mid j\in C\}$.

Proof: Reason in PA:

By 8.7(ii) we have: $\Box(COF\pi\underline{i}\rightarrow\Delta COF\underline{i})$. Hence by I1: $(\Delta COF\underline{i})\lhd COF\pi\underline{i}$.

By J4 and I2:

$$COF\underline{i} \lhd COF\pi\underline{i}.$$

Similarly we have:

$$COF\pi\underline{i} \lhd COF\pi^2\underline{i}$$

.

.

.

$$COF\pi^{n-2}\underline{i} \lhd COF\pi^{n-1}\underline{i}$$

Here we suppose that n is the number of elements of C . By I1, I2 and the above we have:

$$COF\underline{i} \lhd COF\underline{i}$$

$$COF\underline{i} \lhd COF\pi\underline{i}$$

.

.

.

$$COF\underline{i} \lhd COF\pi^{n-1}\underline{i}$$

Hence by I3:

$$COF\underline{i} \lhd \bigvee\{COF\pi^k\underline{i} \mid 0\leq k<n\}$$

In other words:

$$COF\underline{i} \lhd \bigvee\{COF\underline{j} \mid j\in C\}$$ □

9.8 Lemma

$$\vdash [\phi]\equiv[\phi^*]$$

Proof: It is immediate that $\vdash [\phi^*]\lhd[\phi]$. We show: $\vdash [\phi]\lhd[\phi^*]$. Reason in PA:

First note that by 8.9: $\Box([\phi] \leftrightarrow ([\phi\wedge\Box\bot]\vee[\phi\wedge\neg\Box\bot]))$. Hence by J6 and I1: $[\phi\wedge\Box\bot]\lhd[\phi^*\wedge\Box\bot] \rightarrow [\phi]\lhd[\phi^*]$. It follows that we may restrict ourselves to ϕ with $[\![\phi]\!]\subseteq K_0$. So suppose $[\![\phi]\!]\subseteq K_0$. Clearly $[\![\phi^*]\!]$ consists precisely of those circles C with $[\![\phi]\!]\cap C$ not empty. We have by 9.7 and J6:

$$\mathbb{W}\{COF\underline{i} \mid i\Vdash\phi\}\lhd\mathbb{W}\{\mathbb{W}\{COF\underline{j} \mid j\in C\} \mid C\cap[\![\phi]\!]\neq\emptyset\}$$

In other words: $[\phi]\lhd[\phi^*]$. □

9.9 Lemma

$([\phi^*]\vee\Diamond[\phi^*])$ is provably equivalent to a Π_1 sentence.

Proof: Note that: $\vdash \neg([\phi^*]\vee\Diamond[\phi^*]) \leftrightarrow [\neg\phi^*\wedge\Box\neg\phi^*]$. Moreover as is easily seen $[\![\neg\phi^*\wedge\Box\neg\phi^*]\!]$ is upwards persistent. Apply 9.6. □

Proof of 9.1:

We have :

$$\vdash [\phi]\lhd A \leftrightarrow [\phi^*]\lhd A \qquad (9.8,J5)$$
$$\leftrightarrow \Box(A\rightarrow([\phi^*]\vee\Diamond[\phi^*])) \qquad (9.9,J8)$$
□

9.10 Corollary

$$\vdash [\phi]\lhd\top \leftrightarrow \Box(\Box\bot\rightarrow[\phi^*])$$

Proof: We leave it as an excercise to the reader to show that:

$$\vdash \Box(A\vee\Diamond A) \leftrightarrow \Box(\Box\bot\rightarrow A)$$
□

9.11 On a question of Lindström

Per Lindström asks: for which sets Γ of propositional formulas in the variables $p_1,...,p_n$ are there arithmetical sentences $B_1,...,B_n$ such that: $\Gamma=\{\phi \mid \phi(B_1,...,B_n)\lhd\top\}$? (See Lindström[1982] p6; actually Lindström poses his question for arbitrary essentially reflexive theories T. I think that inspection of the argument of this paper shows that the answer given here applies to consistent essentially reflexive RE theories T into which PA restricted to Σ_2-induction can be translated.)

Let's say that $\{\phi \mid \phi(B_1,...,B_n)\lhd\top\}$ is *the interpretability class* of $B_1,...,B_n$. A moment's reflection shows that interpretability classes Γ should satisfy:

i) $\top\in\Gamma$

ii) $\bot\notin\Gamma$

iii) $\phi\vdash_{Prop}\psi$ and $\phi\in\Gamma \Rightarrow \psi\in\Gamma$

We will show that conversely every Γ of propositional formulas in

$p_1,\ldots,p_n$ satisfying (i), (ii), (iii) is an interpretability class.

Proof: Let Γ be a class of propositional formulas in $p_1,\ldots p_n$ satisfying (i), (ii), (iii). The plan of the proof is to construct a circle-tailmodel K and to take $B_i := [p_i]$. 9.10 tells us that what happens below the circles is really irrelevant, so we start by stipulating an arbitrary tail, say $b_0 \ldots b_3 R b_2 R b_1$, where no atom is true at the nodes b_j. We proceed to construct the circles.

$\Gamma^C := \{\phi \mid \phi$ is a propositional formula in the variables $p_1,\ldots,p_n$ and $\phi \notin \Gamma\}$. Note that Γ and Γ^C are both closed under provable equivalence in propositional logic (in the language based on $p_1,\ldots p_n$). Let $\phi_1,\ldots,\phi_k$ be representatives of the equivalence classes of Γ and let $\psi_1,\ldots,\psi_m$ be representatives of the equivalence classes of Γ^C. Define:

$K_0 := \{\langle i,j\rangle \mid 1 \leq i \leq k, 1 \leq j \leq m\}$

$\langle i,j\rangle S \langle i',j'\rangle :\Leftrightarrow j=j'$ and $((1 \leq i < k$ and $i'=i+1)$ or $(i=k$ and $i'=1))$

Let's say that the nodes $\langle i,j\rangle$ for fixed j form a circle C_j.

Consider a node $\langle i,j\rangle$. Clearly $\phi_i \nvdash_{Prop} \psi_j$. So there is an assignment f of truthvalues to $p_1,\ldots,p_n$ under which ϕ_i is true and ψ_j is false. Pick such an assignment f and put:

$\langle i,j\rangle \Vdash p_s :\Leftrightarrow fp_s = \top$

Clearly on every circle C_j there is a node $\langle i,j\rangle$ such that $\langle i,j\rangle \Vdash \phi_i$. Hence $(\Box\bot \rightarrow \phi_i^*)$ and $\Box(\Box\bot \rightarrow \phi_i^*)$ are forced everywhere in the model.

On the other hand no node $\langle i,j\rangle$ on C_j forces ψ_j, hence $\langle i,j\rangle \nVdash \Box\bot \rightarrow \psi_j^*$. It follows that $(\Box(\Box\bot \rightarrow \psi_j^*) \rightarrow \Box\bot)$ is forced everywhere in the model.

Put $B_s := [p_s]$. Note that for any propositional formula χ in $p_1,\ldots,p_n$ we have: $\chi(B_1,\ldots,B_n) = \langle\chi\rangle$. We have by 8.9:

$\vdash \Box(\Box\bot \rightarrow [\phi_i^*])$	$\Rightarrow$	(9.10)
$\vdash [\phi_i] \lhd \top$	$\Rightarrow$	(8.9,I1,J5)
$\vdash \langle\phi_i\rangle \lhd \top$	$\Rightarrow$	
$\vdash \phi_i(B_1,\ldots,B_n) \lhd \top$	$\Rightarrow$	(Reflection Principle)
$\phi_i(B_1,\ldots,B_n)$		

Moreover by 8.9:

$\vdash \Box(\Box\bot \rightarrow [\psi_j^*]) \rightarrow \Box\bot$	$\Rightarrow$	(9.10)
$\vdash [\psi_j] \lhd \top \rightarrow \Box\bot$	$\Rightarrow$	(8.9,I1,J5)
$\vdash \langle\psi_j\rangle \lhd \top \rightarrow \Box\bot$	$\Rightarrow$	
$\vdash \psi_j(B_1,\ldots,B_n) \lhd \top \rightarrow \Box\bot$	$\Rightarrow$	(Reflection Principle)
$\neg\psi_j(B_1,\ldots,B_n)$		

Note that the uses of the Reflection Principle are eliminable here: we could just have proved the necessary lemmas externally, i.e. in non-formalized form.

It follows immediately that $\Gamma=\{\phi \mid \phi(B_1,\ldots,B_n)\lhd\top\}$. □

9.12 Remark

Note that in the proof of 9.11 it would have sufficed to consider representatives ϕ_i of the equivalence classes in Γ that are *minimal* in the implication ordering. Similarly we need only consider representatives ψ_j of the equivalence classes of Γ^C that are *maximal* in the implication ordering.

REFERENCES

D.D. Auerbach, 1985, Intentionality and the Gödel Theorems, *Philosophical Studies 48*, 337-351.

C. Bernardi & F. Montagna, 1982, Equivalence relations induced by extensional formulae: classification by means of a new fixed point property, *Rapporto matematico 63*, Dipartimento di Matematica, Via del Capitano 15, 53100 Siena.

G.L. Bowie, 1982, Lucas' number is finally up, *Journal of Philosophical Logic 11*, 279-285.

S. Feferman, 1960, Arithmetization of metamathematics in a general setting, *Fundamenta Mathematica XLIX*, 35-92.

D. Guaspari & R.M. Solovay, 1979, Rosser sentences, *Annals of Mathematical Logic 16*, 81-99.

D. Hilbert & P. Bernays, 1970, *Grundlagen der Mathematik II*, Springer-Verlag.

P. Lindström, 1979, Some results on interpretability, *Proceedings of the 5th Scandinavian Logic Symposium*, Aalborg 1979, 329-361.

P. Lindström, 1982, More on partially conservative sentences and interpretability, *Philosophical Comunications, Red Series, No 17*, Department of Philosophy, University of Göteborg.

J.R. Lucas, 1961, Minds, Machines and Gödel, *Philosophy 36*, 120-124.

F. Montagna, 1978, On the Algebraization of a Feferman's Predicate, *Studia Logica 37*, 221-263.

F. Montagna, 1984, Provability in finite Subtheories of PA and Relative Interpretability: a Modal Investigation, *Rapporto Matematico 118*, Dipartimento di Matematica, Via del Capitano 15, 53100 Siena.

S. Orey, 1961, Relative interpretations, *Zeitschrifft für Mathematische Logik und Grunlagen der Mathematik 7*, 146-153.

C. Smoryński, 198?, Quantified Modal Logic and Selfreference.

C. Smoryński, 1985, *Selfreference and Modal Logic*, Springer-Verlag.

R.M. Solovay, Provability interpretations of modal logic, *Israel Journal of Mathematics 25*, 287-304.

V. Švejdar, 1978, Degrees of interpretability, Commentationes Mathematicae Universitatis Carolinae 19, 789-813.

V. Švejdar, 1983, Modal analysis of generalized Rosser sentences, *JSL 48* , 986-999.

A. Visser, 1984, The provability logics of recursively enumerable theories extending Peano Arithmetic at arbitrary theories extending Peano Arithmetic, *Journal of Philosophical Logic 13* , 79-113.

J. Webb, 1980, *Mechanism, Mentalism and Metamathematics* , Reidel.