Notre Dame Journal of Formal Logic
Volume 30, Number 2, Spring 1989

Peano's Smart Children: A Provability Logical Study of Systems with Built-in Consistency

ALBERT VISSER*

Abstract The systems studied in this article prove the same theorems (from the "extensional" point of view) as Peano Arithmetic, but are equipped with a self-correction procedure. These systems prove their own consistency and thus escape Gödel's second theorem. Here, the provability logics of these systems are studied. An application of the results obtained turns out to be the solution to a problem of Orey on relative interpretability.

1 Introduction Consistency can be built into a system in various ways. The two best known constructions are Rosser's and Feferman's, both of which 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 this: 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 relations between Gödel's first and second Incompleteness Theorems (see [4]).
(ii) They serve as didactic examples in philosophical discussions, like the

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

Received September 14, 1987

debate on intensionality in mathematics (e.g. [1]) and the discussion on the possible bearing of the Incompleteness Theorems on the Minds & Machines problem (see [10], [21], [3]).

(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 [4], [13]).

The main objects of study in the present article are certain variants of both Rosser's and Feferman's constructions. 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. [12] and [16]). There are two directions of research: The pure study of arithmetical self-reference and the study of arithmetical self-reference as a tool for unifying self-referential arguments in arithmetic (see [6], 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 having '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 for producing arithmetical sentences with rich but controlled properties. These sentences can be used to prove various incompleteness and other kinds of results (for an example, see Section 7 of this paper).

The contribution of this study lies along the first line. I provide an example of a rich modal logic of not too standard a 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 [15] 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 Theorems; in this sense (a) generalizes (i). The logic can also be used to illustrate the point that one can simulate the results of intensional self-reference (as with Rosser's Theorem) to quite an extent by applying provably extensional self-reference—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 developed have as a spin-off an application to relative interpretability: I answer a question posed by Orey for the case of PA.

As far as prerequisites go, knowledge of [4] and [16] should bring the reader a long way.

2 Contents of the article In Section 3 the necessary notions and notations are introduced. Section 4 is a step-by-step introduction to the construction of the systems that are central in this article. This section also illustrates the powers of provably extensional self-reference and contains a discussion of the problem of uniqueness and explicitness of the Gödel and Henkin sentences of the various systems considered. Section 5 treats the bimodal principles valid for the two central systems; a Kripke model completeness theorem is proved. Section 6 has a partial result on embedding Kripke models for our modal system into arithmetic. In Section 7 this embedding result is applied to a problem about relative interpretability.

3 Conventions, notions, and elementary facts

3.1 Point All the arithmetical results in the next sections will be stated for Peano Arithmetic. PA, of course, is just a convenient peg to hang the discussion on; almost 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 be essentially reflexive.

3.2 $\Box$ and $\triangle$ (in different contexts) Let $\mathrm{Proof}(x,y)$ be the $\triangle_0$ arithmetical formula representing the relation: x is the Gödel number of a PA-proof of the formula with Gödel number y. We assume for convenience that $\mathrm{PA} \vdash \forall x \exists ! y \mathrm{Proof}(x,y)$. Let $\mathrm{Prov}(y) = \exists x \mathrm{Proof}(x,y)$. We write, *par abus de langage*, '$\mathrm{Proof}(x, A(x_1,\ldots,x_n))$' for $\mathrm{Proof}(x, \ulcorner A(\dot{x}_1,\ldots,\dot{x}_n) \urcorner)$, where:

(i) all free variables of A are among those shown
(ii) $\ulcorner A(\dot{x}_1,\ldots,\dot{x}_n) \urcorner$ is the "Gödel term" for $A(x_1,\ldots,x_n)$ as defined on p. 43 of [16].

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

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

'$\Diamond$' will be an abbreviation for $\neg\Box\neg$, and '∇' for $\neg\triangle\neg$.

When we want to consider systems with other axiom sets than PA, we will 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 $\Box_\pi$.

3.3 $\Box\restriction x$ *and* $\Box^*$ Define:

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

It is clear that $\mathrm{PA} \vdash \Box A \leftrightarrow \Box^*A$, 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 [18].)

3.4 Witnessing and the Rosser-ordering Let A be of the form $\exists xA_0(x)$. Define t 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 $\exists xA_0(x)$ and B of the form $\exists xB_0(x)$. 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 $\exists x\exists yA_0(x,y)$ with A_0 in Δ_0 and where B is $\exists x\exists yB_0(x,y)$ 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 the following theorem:

Theorem $\mathrm{PA} \vdash \exists x(\exists yA_0(x,y) \wedge \forall z \leq x\forall u\neg B_0(z,u)) \leftrightarrow \forall u\exists x(\exists yA_0(x,y) \wedge \forall z \leq x\neg B_0(z,u))$.

Proof: The "$\rightarrow$" side is trivial. For the "$\leftarrow$" side reason in PA as follows: Suppose that $\forall u\exists x(\exists yA_0(x,y) \wedge \forall z \leq x\neg B_0(z,u))$. It follows that $\exists x\exists yA_0(x,y)$. Let x_0 be the smallest such x. Consider any u. Pick an x such that $\exists yA_0(x,y)$ and $\forall z \leq x\neg B_0(z,u)$. Clearly $x_0 \leq x$ and hence $\forall z \leq x_0\neg B_0(z,u)$. We then conclude that $\exists yA_0(x_0,y) \wedge \forall z \leq x_0\neg B_0(z,u)$.

Both Švejdar [18] and Lindström [9] 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.

3.5 Relative interpretability '$A \lhd B$' stands for the arithmetization of: PA + A is relatively interpretable in PA + B. $\mathrm{PA} \vdash A \lhd B \leftrightarrow \forall x\Box(B \rightarrow \Diamond\restriction xA)$ is a result due to Orey and Hájek. We list a number of principles valid for $\Box$ and $\lhd$:

(I1) $\mathrm{PA} \vdash \Box(B \rightarrow A) \rightarrow A \lhd B$
(I2) $\mathrm{PA} \vdash (A \lhd B \wedge B \lhd C) \rightarrow A \lhd C$
(I3) $\mathrm{PA} \vdash (A \lhd B \wedge A \lhd C) \rightarrow A \lhd (B \vee C)$
(I4) $\mathrm{PA} \vdash A \lhd B \rightarrow (\Diamond B \rightarrow \Diamond A)$

(I5) $\mathrm{PA} \vdash \Diamond A \lhd B \to \Box(B \to \Diamond A)$
(I6) $\mathrm{PA} \vdash A \lhd \Diamond A$
(I7) $\mathrm{PA} \vdash A \lhd B \to (A \wedge \Box C) \lhd (B \wedge \Box C)$.

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

(I4′) $\mathrm{PA} \vdash \bot \lhd B \to \Box\neg B$.

Thus it is sufficient to prove

(J1) For all P in Π_1, $\mathrm{PA} \vdash P \lhd B \to \Box(B \to P)$.

First note that for every n, $\mathrm{PA} \vdash \forall x \Box(B \to \Diamond\!\restriction\! xA) \leftrightarrow \forall x > n\, \Box(B \to \Diamond\!\restriction\! xA)$. Pick q so big that $\Box\!\restriction\! q$ contains Robinson's Arithmetic. We then have for S in Σ_1 that $\mathrm{PA} \vdash \forall x > q\, \Box(\Box\!\restriction\! xS \leftrightarrow S)$, so for P in Π_1 $\mathrm{PA} \vdash \forall x > q\, \Box(\Diamond\!\restriction\! xP \leftrightarrow P)$. Hence

$$\begin{aligned}\mathrm{PA} \vdash P \lhd B &\leftrightarrow \forall x > q\, \Box(B \to \Diamond\!\restriction\! xP)\\ &\leftrightarrow \Box(B \to P).\end{aligned}$$

We now turn to (I7). We prove

(J2) For all S in Σ_1, $\mathrm{PA} \vdash A \lhd B \to (A \wedge S) \lhd (B \wedge S)$.

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

$$\mathrm{PA} \vdash \forall x > q\, \Box(S \to \Box\!\restriction\! x((D \wedge S) \leftrightarrow D)).$$

It follows that

$$\begin{aligned}\mathrm{PA} \vdash \forall x \Box(B \to \Diamond\!\restriction\! xA) &\to \forall x > q \Box((B \wedge S) \to \Diamond\!\restriction\! x(A \wedge S))\\ &\to \forall x \Box((B \wedge S) \to \Diamond\!\restriction\! x(A \wedge S)).\end{aligned}$$

For further details see [19].

3.6 On systems Philosophically, I think it is best to make the whole apparatus for generating theorems part of the identity conditions of systems. For our purposes however, it is 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 if a *variant* of a system with predicate B if $\mathrm{PA} \vdash \forall x(A(x) \leftrightarrow B(x))$.

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 may be seen to be extensionally equal to the formal systems on which they are based, given the information that the original systems are consistent.

4 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 prob-

lems of uniqueness and of explicitness of Gödel and Henkin sentences of the systems introduced are considered. (The rationale behind the attention to these specific problems is that these problems were historically at the crib of provability logic for $\Box$, and also 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 self-reference. Specifically, I show how to use provably extensional self-reference to construct four nonequivalent Orey sentences.

In this section '$\vdash$' stands for PA $\vdash$, and A, B, C stand for *formulas* of the language of PA. Note that by our conventions we have that $\vdash A(x) \Rightarrow \vdash \forall x A(x)$, 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: $\triangle A \Leftrightarrow \Box A < \Box \neg A$.

Some principles valid for the Rosser System in PA are:

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

Some direct consequences of (1)–(4) are:

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

It is perhaps worth noting that the set of theorems of the Rosser System is *provably* infinite. Reason in PA as follows: in case $\neg \Box \bot$ this is trivial. In case $\Box \bot$ for any A, clearly one of A, $\neg A$, $\neg\neg A$, $\neg\neg\neg A, \ldots$, will be Rosser-provable.

In the Rosser System we have two explicit but nonunique Henkin sentences:

(7) $\vdash \top \leftrightarrow \triangle \top$ (1)
(8) $\vdash \bot \leftrightarrow \triangle \bot$. (2)

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

(9) $\vdash G \leftrightarrow \neg \triangle G$.

We have

(10) $\vdash \Box G \rightarrow (\Box \triangle G \wedge \Box \neg \triangle G) \rightarrow \Box \bot$. (6),(9)

Of course we can also prove

(11) $\vdash \Box\neg G \rightarrow \Box\bot$

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 $\triangle$ removes this defect as we will see under K.

The uniqueness or nonuniqueness of Gödel sentences in the Rosser System is still an open problem. Guaspari and Solovay show that if one allows *variants* of Prov in the definition of $\triangle$, the answer may be yes and may be no (see [5]). I am not aware of any argument that there are no explicit Gödel sentences for $\triangle$.

Open question Are there explicit Gödel sentences for $\triangle^R$?

Open question If one allows Σ_1 variants (with one existential quantifier in front of the $\triangle$) of Prov in the definition of $\triangle^R$, can there be explicit Gödel sentences for $\triangle^R$?

K (Kreisel's symmetrized Rosser System) Kreisel's variation on the Rosser System is reported in [7], pp. 298–302.

Define $\triangle A \Leftrightarrow \exists x[\text{Proof}(x,A) \wedge \forall u,v,b,c \leq x((\text{Proof}(u,b) \wedge \text{Proof}(v,c)) \rightarrow c \neq \text{neg}(b))]$. Clearly, $\triangle A$ is Σ_1. The Kreisel System satisfies principles (1)–(4) and the additional principle

(12) $\vdash \neg(\triangle A \wedge \triangle\neg A)$.

We can now prove (11) modally:

$$\begin{aligned} \vdash \Box\neg G &\rightarrow \Box\triangle G \wedge \Box\triangle\neg G && (6),(9)\\ &\rightarrow \Box(\triangle G \wedge \triangle\neg G)\\ &\rightarrow \Box\bot. && (12)\end{aligned}$$

Note also that $\vdash \Box\bot \rightarrow$ "$\{A \mid \triangle A\}$ 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

$$\mathrm{R}'_0 = \varnothing$$

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

Let $\triangle A$ be the arithmetization of $\exists n A \in \mathrm{R}'_{n+1}$. Clearly, $\triangle A$ is Σ_1. We have that $\vdash \triangle A \leftrightarrow \Box A < \triangle\neg A$, and even $\vdash \forall x(x \text{ wit } \triangle A \leftrightarrow x \text{ wit } \Box A < \triangle\neg A)$. This last fact happens to characterize $\triangle^{R'}$.

Principles (1)–(4) hold for $\triangle$, plus the additional principle

(13) $\vdash \Box\bot \rightarrow (\triangle A \vee \triangle\neg A)$.

A direct consequence is

(14) $\vdash \Box A \to (\triangle A \vee \triangle \neg A)$. (4),(13)

Finally, note that $\vdash \triangle^{R} A \to \triangle^{R'} A$.

BM (The Bernardi-Montagna System) We now jump up directly to a system which is richer, from the modal point of view, than both $\triangle^{K}$ and $\triangle^{R'}$. This system was discovered by Claudio Bernardi and Franco Montagna (see [2]). Define

$$\mathrm{BM}_0 = \varnothing$$

$$\mathrm{BM}_{n+1} = \begin{cases} \mathrm{BM}_n \cup \{A\}, & \text{if } \mathrm{Proof}(n,A) \text{ and } \mathrm{BM}_n \cup \{A\} \text{ is consistent} \\ & \text{in propositional logic} \\ \mathrm{BM}_n, & \text{otherwise.} \end{cases}$$

Let $\triangle A$ be the arithmetization of $\exists n A \in \mathrm{BM}_{n+1}$. Clearly, $\triangle A$ is Σ_1. Also, principles (1)–(4) and (13) hold for $\triangle$. We also have, by elementary reasoning,

(15) $\vdash \triangle (A \to B) \to (\triangle A \to \triangle B)$.

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

mBM (The modified Bernardi-Montagna System) For our purposes we want the following additional principle

(16) $\vdash \triangle A \to \triangle \Box A$.

It is not plausible for one to 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 $\triangle \Box \bot$ rather than $\triangle \neg \Box \bot$? We can, however, modify the BM System in such a way that we get (16).

Let $\vdash_{\mathrm{Prop}}$ stand for derivability in propositional logic. Define

$$\mathrm{mBM}_0 = \varnothing$$

$$\mathrm{mBM}_{n+1} = \begin{cases} \mathrm{mBM}_n \cup \{A\}, & \text{if } \mathrm{Proof}(n,A) \text{ and } \mathrm{mBM}_n \cup \{A\} \nvdash_{\mathrm{prop}} \neg \Box \bot \\ \mathrm{mBM}_n, & \text{otherwise.} \end{cases}$$

Let $\triangle A$ be the arithmetization of $\exists n A \in \mathrm{mBM}_{n+1}$. Clearly, $\triangle A$ is Σ_1.

It is easily seen that (1)–(3) and (15) are valid. We may verify (4) in PA as follows:

> Trivially, $\triangle A \to \Box A$. Suppose that $\neg \Box \bot$ and $\Box A$, say $\mathrm{Proof}(x,A)$. The only reason A could be left out of mBM_{x+1} is that $\mathrm{mBM}_x \cup \{A\} \vdash_{\mathrm{Prop}} \neg \Box \bot$. But then $\Box \neg \Box \bot$ and hence $\Box \bot$, which is a contradiction. So we conclude that $A \in \mathrm{mBM}_{x+1}$ and thus $\triangle A$.

(13) can be proved in PA as follows:

> Suppose that $\Box \bot$. For certain x and y we will have that $\mathrm{Proof}(x,A)$ and $\mathrm{Proof}(y,\neg A)$. Let $z = \max(x,y)$. If $\neg \triangle A$ and $\neg \triangle \neg A$ we will have that

$\mathrm{mBM}_{z+1} \cup \{A\} \vdash_{\mathrm{Prop}} \neg\Box\bot$ and $\mathrm{mBM}_{z+1} \cup \{\neg A\} \vdash_{\mathrm{Prop}} \neg\Box\bot$. Hence $\mathrm{mBM}_{z+1} \vdash_{\mathrm{Prop}} \neg\Box\bot$.

Finally, we turn to (16). Clearly,

(17) $\vdash \neg\triangle\neg\Box\bot$

from which it follows that

$$\vdash \Box\bot \to \triangle\Box\bot \qquad (13),(17)$$

moreover

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

hence

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

also

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

thus

$$\vdash \neg\Box\bot \to (\Box A \to \triangle\Box A).$$

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

Let us list for the sake of convenience the principles valid for $\triangle^{\mathrm{mBM}}$ with brand new names:

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

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

(18) $\vdash \triangle(\Box A \to A) \to \triangle A$.

Proof: $\vdash \triangle(\Box A \to A) \to \Box(\Box A \to A)$ (5)
$\to \Box A$
$\to \triangle\Box A$ (B4)
$\to \triangle A$. (B2)

The second consequence is the principle of provable extensionality

(19) $\vdash \Box(A \leftrightarrow B) \to \Box(\triangle A \leftrightarrow \triangle B)$.

Proof: $\vdash \Box(A \leftrightarrow B) \to \Box\triangle(A \leftrightarrow B)$ (6)
$\to \Box(\triangle A \leftrightarrow \triangle B)$. (B2)

The next principle is an immediate consequence of (12) and (B7)

(20) $\vdash \Box\bot \rightarrow (\triangle A \leftrightarrow \nabla A)$.

Let us 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 $\Box,\triangle$ (in other words, if A is an interpreted sentence of the closed fragment of the bimodal propositional logic with operators $\Box$ and $\triangle$).

(21) Suppose that A is modally closed. Then there is an $\alpha \in \{0,\ldots,\omega\}$ such that $\vdash \triangle A \leftrightarrow \Box^\alpha\bot$.

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

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

Secondly we have that $\vdash \triangle\Box^0\bot \leftrightarrow \Box^0\bot$, and $\vdash \triangle\Box^{1+\gamma}\bot \leftrightarrow \Box^{2+\gamma}\bot$ (as is easily seen by considering the cases $\Box\bot$ and $\neg\Box\bot$ separately). Combining these results we see that $\vdash \triangle B \leftrightarrow \Box^\delta\bot$, for some $\delta \in \{0,\ldots,\omega\}$. (21) follows by a trivial induction on A. (Note that we didn't use (B7) in the argument.)

We now turn to the Henkin sentences of $\triangle$. We have already seen in (7) and (8) that $\bot$ and $\top$ are explicit Henkin sentences. For $\triangle^{\mathrm{mBM}}$ we can show that they are the only explicit Henkin sentences. Consider H satisfying:

(22) $\vdash H \leftrightarrow \triangle H$.

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 $n \in \{1,2,\ldots\}$

$$\begin{aligned}\vdash \Box^n\bot &\leftrightarrow \triangle\Box^n\bot\\ &\leftrightarrow \Box^{n+1}\bot.\end{aligned}$$

Open question Are there nonexplicit Henkin sentences of $\triangle^{\mathrm{mBM}}$?

Next we turn to the Gödel sentences of $\triangle$. Under R and K we have seen in (10) and (11) that these have the Rosser Property. We show that they are nonexplicit and nonunique.

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 get that

$$\begin{aligned}\vdash \triangle G &\leftrightarrow \triangle\neg\Box^\alpha\bot && \text{(B1),(B2)}\\ &\leftrightarrow \triangle\bot && \text{(18)}\\ &\leftrightarrow \bot. && \text{(B3)}\end{aligned}$$

So, by (9), $\vdash G$. Thus $\alpha = 0$, which is a contradiction. If $\alpha = 0$, we have that $\vdash G$ and thus $\vdash \triangle G$, i.e., by (9), $\vdash \neg G$, another contradiction. Hence G cannot be explicit.

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

(23) $\vdash \nabla G \leftrightarrow \neg\triangle\nabla G$.

To prove (23) it is clearly sufficient to show

(a) $\vdash \neg\Box\bot \rightarrow \nabla G$
(b) $\vdash \neg\Box\bot \rightarrow \neg\triangle\nabla G$
(c) $\vdash \Box\bot \rightarrow (\nabla G \leftrightarrow \neg\triangle\nabla G)$.

We prove (a) by contraposition

$$\begin{aligned} \vdash \triangle\neg G &\rightarrow \Box\neg G && (5)\\ &\rightarrow (\Box\triangle G \wedge \Box\triangle\neg G) && (6),(9)\\ &\rightarrow \Box\bot. && (12) \end{aligned}$$

To prove (b), we show first

$$\begin{aligned} \vdash \Box\Box\bot \rightarrow (\triangle\nabla G &\rightarrow \Box\nabla G) && (5)\\ &\rightarrow \Box\triangle G) && (20)\\ &\rightarrow \Box\neg G) && (9)\\ &\rightarrow \Box\bot). && \text{(as in the proof of (a))} \end{aligned}$$

Hence

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

thus

$$\begin{aligned} \vdash \triangle\nabla G &\rightarrow \Box\triangle\nabla G && (B5)\\ &\rightarrow \Box(\Box\Box\bot \rightarrow \Box\bot)\\ &\rightarrow \Box\Box\bot \end{aligned}$$

and so, combining,

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

The proof of (c) goes as follows:

$$\begin{aligned} \vdash \Box\bot &\rightarrow \triangle\Box\bot && (B4)\\ &\rightarrow \triangle(\nabla G \leftrightarrow \triangle G) && (B1),(B2),(20)\\ &\rightarrow (\triangle\nabla G \leftrightarrow \triangle\triangle G) && (B2)\\ &\rightarrow (\triangle\nabla G \leftrightarrow \triangle\neg G) && (9),(B1),(B2)\\ &\rightarrow (\neg\triangle\nabla G \leftrightarrow \nabla G). \end{aligned}$$

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

(24) $\vdash \Box(G \leftrightarrow \nabla G) \rightarrow \Box\bot$.

Clearly,

$$\begin{aligned} \vdash (G \wedge \nabla G) &\rightarrow (\neg\triangle G \wedge \neg\triangle\neg G)\\ &\rightarrow \neg\Box\bot && (B7) \end{aligned}$$

and

$$\vdash (\neg G \wedge \neg \nabla G) \rightarrow (\triangle G \wedge \triangle \neg G) \rightarrow \bot. \tag{12}$$

Combining, we get that

$$\begin{aligned} \vdash \Box (G \leftrightarrow \nabla G) &\rightarrow \Box ((G \wedge \nabla G) \vee (\neg G \wedge \neg \nabla G)) \\ &\rightarrow \Box \neg \Box \bot \\ &\rightarrow \Box \bot. \end{aligned}$$

Another way to prove (24) is by noting that

$$\vdash \Box \Box \bot \rightarrow (\Box (G \leftrightarrow \nabla G) \rightarrow \Box (G \leftrightarrow \triangle G) \rightarrow \Box \bot). \tag{20}$$

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

(25) $\vdash G \leftrightarrow \nabla \nabla G$.

In Section 5 we will see that as far as our modal principles are concerned we cannot show more than this: if there is a Gödel sentence of $\triangle$ then there is a second, nonequivalent one.

Open question Are there three pairwise nonequivalent Gödel sentences of $\triangle^{\mathrm{mBM}}$?

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 [4]) these systems cannot be *provably* closed under the axioms and rules of predicate logic; in other words, we do not have $\vdash \triangle A \leftrightarrow \Box_{\triangle} A$. If $\triangle$ is one of $\triangle^{\mathrm{BM}}, \triangle^{\mathrm{mBM}}$ we can say a bit more. Let $\triangle$ be one of $\triangle^{\mathrm{BM}}, \triangle^{\mathrm{mBM}}$. Let Q be the conjunction of the axioms of Robinson's Arithmetic. We clearly have $\vdash Q$, and hence $\vdash \triangle Q$, and thus $\vdash \Box_{\triangle} Q$. It immediately follows from the provable Σ_1-completeness of Robinson's Arithmetic that

(26) $\vdash \triangle A \rightarrow \Box_{\triangle} \triangle A$.

Let G be a Gödel sentence of $\triangle$. We have

(d)	$\vdash \Box \bot \rightarrow (\triangle G \vee \triangle \neg G)$	(B7)
(e)	$\vdash \triangle G \rightarrow \Box_{\triangle} \triangle G$	(26)
(f)	$\vdash \triangle G \rightarrow \Box_{\triangle} G \rightarrow \Box_{\triangle} \neg \triangle G$	(9)
(g)	$\vdash \triangle G \rightarrow \Box_{\triangle} \bot$	(e),(f)
(h)	$\vdash \triangle \neg G \rightarrow \Box_{\triangle} \triangle \neg G \rightarrow \Box_{\triangle} \neg \triangle G$	(26) (B1),(12)
(i)	$\vdash \triangle \neg G \rightarrow \Box_{\triangle} \neg G \rightarrow \Box_{\triangle} \triangle G$	(9)
(j)	$\vdash \triangle \neg G \rightarrow \Box_{\triangle} \bot$	(h),(i)
(k)	$\vdash \Box \bot \rightarrow \Box_{\triangle} \bot$	(d),(g),(j)
(l)	$\vdash \Box \bot \rightarrow (\Box_{\triangle} A \leftrightarrow \Box A)$	(k)
(m)	$\vdash \neg \Box \bot \rightarrow (\Box_{\triangle} A \leftrightarrow \Box_{\Box} A \leftrightarrow \Box A)$	
(27)	$\vdash \Box_{\triangle} A \leftrightarrow \Box A$.	(l),(m)

So $\triangle^{BM}$ and $\triangle^{mBM}$ are provably axiom sets for the theorems of PA. The same thing can be proved for $\triangle^{R'}$ by a slightly refined variant of the above argument. What about $\triangle^{R}$ and $\triangle^{K}$? I'm not certain, but a good guess is that it would be so for $\triangle^{R}$ but not for $\triangle^{K}$.

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 [4]) 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 [4], [13]). A modal study of this provability predicate was made by Montagna [11].

Let us start by giving two rather different intuitive descriptions of the Feferman System (or, to be faithful to the conventions of Section 3.6, I should say: let us describe two variants of the Feferman System).

Suppose the arithmetical axioms of PA are enumerated by $A_1, A_2, A_3, \ldots$, in the order of their Gödel numbers (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 $\mathrm{F} = \bigcup \{X \mid X \text{ is a finite, initial, consistent set of arithmetical axioms of PA}\}$.

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 licensed to use axiom A_i one needs the *external information* that $\{A_j \mid j \leq i\}$ 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 $\bot$, we extract the axiom A_j with largest Gödel number from π_i. We backtrack and scratch out all the proofs employing axioms A_k with $k \geq j$. Then we go on enumerating proofs, skipping those employing axioms A_k with $k \geq j$. As soon as we meet another proof of $\bot$ 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 even subtler: 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 now turn to the formal definition of the Feferman System. Define

$$\pi^*(x) \Leftrightarrow \pi(x) \wedge \Diamond \ulcorner x \top$$
$$\triangle A \Leftrightarrow \Box_{\pi^*} A.$$

The following are equivalents of $\triangle A$:

(28) $\quad \vdash \triangle A \leftrightarrow \exists x(\Box \ulcorner xA \wedge \Diamond \ulcorner x1\top)$.

where f is a primitive recursive function with

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

we have

(29) $\vdash \triangle A \leftrightarrow \exists x(\mathrm{Proof}(x,A) \wedge \Diamond \ulcorner f(x) \top)$.

Remembering that $\Box^* A \Leftrightarrow \exists x \Box \ulcorner xA$, we have

(30) $\vdash \triangle A \leftrightarrow \triangle^* A < \Box^* \bot$
(31) $\vdash \triangle A \leftrightarrow \Box^* A < \Box^* \neg A$.

(31) brings out the similarity between the Feferman System and the Rosser System. By 3.4 and (30) or (31) we see that $\triangle$ is $\triangle_2$.

(B1)–(B6) are valid for $\triangle$. In [4] all of these except (B4) are mentioned. In [11] a modal study is made of (B1), (B2), (B3), (B5), and (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

$$\vdash \forall x \Box (\Box \ulcorner xA \rightarrow A)$$

from which it follows that

$$\vdash \forall x \Box \Diamond \ulcorner x \top.$$

Proof of B1:

$$\begin{aligned} \vdash A &\Rightarrow \text{for some } n \vdash \Box \ulcorner nA \\ &\Rightarrow \text{for some } n \vdash \Box \ulcorner nA \wedge \Diamond \ulcorner n \top \\ &\Rightarrow \vdash \triangle A. \end{aligned}$$

Proof of B4: We will prove the stronger principle

(32) Let S be Σ_1, then $\vdash S \rightarrow \triangle S$.

Let $\Box_Q$ stand for provability in Robinson's Arithmetic. For some q, PA $\vdash \Box_Q A \rightarrow \Box \ulcorner qA$. Let S be Σ_1. We have that

$$\begin{aligned} \vdash S &\rightarrow \Box_Q S \\ &\rightarrow \Box \ulcorner qS \\ &\rightarrow (\Box \ulcorner qS \wedge \Diamond \ulcorner q \top) \\ &\rightarrow \triangle S. \end{aligned}$$

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

Proof of B5: It is clearly sufficient to prove (6); i.e., $\vdash \Box A \rightarrow \triangle \Box A$. To do this we formalize the reasoning for (B1) as follows:

$$\begin{aligned} \vdash \Box A &\rightarrow \exists x \Box \Box \ulcorner xA \\ &\rightarrow \exists x \Box (\Box \ulcorner xA \wedge \Diamond \ulcorner x \top) \\ &\rightarrow \Box \triangle A. \end{aligned}$$

Just as for $\triangle^{\mathrm{mBM}}$, $\triangle^{\mathrm{F}}$ has precisely two nonequivalent *explicit* Henkin sentences. I will now show that $\triangle^{\mathrm{F}}$ has in fact infinitely many pairwise nonequiva-

lent Henkin sentences. 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 Rosser sentence. We have that

(33) If A is Σ-minded then $\vdash \triangle A \leftrightarrow (\Box A \wedge (\Box \bot \rightarrow A))$.

Proof: Suppose that A is Σ-minded. To prove left to right, it is clear that $\vdash \triangle A \rightarrow \Box A$. Moreover,

$$\begin{aligned} \vdash (\triangle A \wedge \Box \bot) \rightarrow (\neg A &\rightarrow (\neg A \wedge \Box \neg A) \\ &\rightarrow \triangle (\neg A \wedge \Box \neg A) && \text{(B1),(B2),(32)} \\ &\rightarrow \triangle \bot && \text{(B1),(B2)} \\ &\rightarrow \bot. && \text{(B3)} \end{aligned}$$

To prove right to left, we have that

$$\begin{aligned} \vdash (\Box A \wedge (\Box \bot \rightarrow A)) \rightarrow (\Box \bot &\rightarrow (A \wedge \Box A) \\ &\rightarrow \triangle (A \wedge \Box A) && \text{(B1),(B2),(32)} \\ &\rightarrow \triangle A). && \text{(B1),(B2)} \end{aligned}$$

Moreover,

$$\vdash (\Box A \wedge (\Box \bot \rightarrow A)) \rightarrow (\neg \Box \bot \rightarrow \triangle A). \qquad \text{(B6)}$$

Note that our proof uses only (B1), (B2), (B3), (B6), and (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 7.1.

To prove that there are infinitely many pairwise nonequivalent Henkin sentences I have to borrow some material and definitions from [20]. The reader not familiar with this paper can at least get the essential idea of the argument by considering the ordinary Σ_1 Rosser sentence 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 himself that R and S are Σ-minded and satisfy $\vdash R \leftrightarrow (\Box R \wedge (\Box \bot \rightarrow R))$, $\vdash S \leftrightarrow (\Box S \wedge (\Box \bot \rightarrow S))$.

Consider a tail model $\mathbf{K}$. 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 $\vdash ([\phi] \wedge \Box [\phi]) \leftrightarrow [\phi \wedge \Box\phi]$. 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 tail model shown in Figure 1, where 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 $\vdash [p_i] \leftrightarrow \triangle [p_i]$. 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 nonequivalent Henkin sentences of $\triangle$. Since n can be freely chosen it follows that there are infinitely many pairwise nonequivalent Henkin sentences of $\triangle$.

We state two open problems:

Open problem Are there Henkin sentences of $\triangle^F$ that are not provably equivalent to Σ_1 sentences?

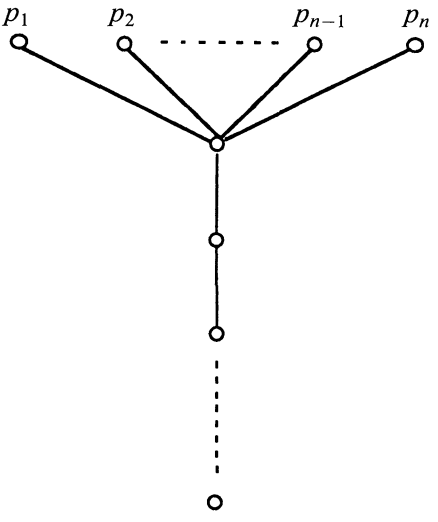


Figure 1.

Open problem What are the possible truth values of the *literal* Henkin sentences of $\triangle^F$?

We turn next to Gödel sentences. Let G satisfy (9), so as in the case of $\triangle^{mBM}$ G is nonexplicit. (We can, using the observations about tail models above, also see this "vom höheren Standpunkt," for consider the 'minimal' tail model, 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 $\triangle$ (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 this 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 nonuniqueness of the Gödel sentences of $\triangle^{mBM}$ 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 by Montagna [11].

Montagna's problem Is G unique?

The Gödel sentence of $\triangle^F$ is an Orey sentence. Before defining what an Orey sentence is, I want to note that the fact that G is such a sentence only depends upon (B1), (B2), (B3), (6), and (32). Let us call a $\triangle$ that satisfies these principles *precocious*.

An *Orey sentence* is a sentence A that has the property $A \lhd \top$ and $\neg A \lhd \top$. (Strictly speaking, my usage is at variance with the tradition: e.g., Gödel sentences and Rosser sentences 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*. Rosser sentences have the Rosser Property, but other sentences do as well. So the more correct usage would be: sentence with

the Orey Property.) Trivially, the negation of an Orey sentence is again an Orey sentence.

We now show that the Gödel sentence of any precocious $\triangle$ is an Orey sentence. Suppose that $\triangle$ is precocious. First we prove

(34) $\vdash A \lhd \triangledown A$.

Proof: Let $B = \neg A$; we have

$$\begin{aligned}
&\vdash \forall x \Box (\Box \restriction xB \to B) \Rightarrow \\
&\vdash \forall x \Box \triangle (\Box \restriction xB \to B) \Rightarrow && (6)\\
&\vdash \forall x \Box (\triangle \Box \restriction xB \to \triangle B) \Rightarrow && (B2)\\
&\vdash \forall x \Box (\Box \restriction xB \to \triangle B) \Rightarrow && (32)\\
&\vdash \forall x \Box (\triangledown A \to \Diamond \restriction xA) \Rightarrow \\
&\vdash A \lhd \triangledown A.
\end{aligned}$$

Secondly, one easily proves, using (I1), (I2), (I3)

(35) $\vdash A \lhd \neg A \to A \lhd \top$.

We have

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

hence by (35)

$$\vdash G \lhd \top$$

moreover

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

hence by (I1), (I2), and (35)

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

A curious fact is that the Gödel sentences of $\triangle^{\mathrm{F}}$ is *precisely* the Orey sentence discovered independently by Lindström and Švejdar (see [9] and [18]); by 3.4 *this* Orey sentence is $\triangle_2$. In Section 7 we will see that there are infinitely many nonequivalent Orey sentences.

Before leaving the subject of Orey sentences, I want to note that Orey sentences are Σ_1- and Π_1-flexible and that they are Kent sentences. 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 *Kent sentence* if $(A \wedge \Box A)$ is not provably equivalent to a Σ_1 sentence. I will show that an Orey sentence is a Kent sentence and leave the proof that Orey sentences are Σ_1- and Π_1-flexible to the reader. Suppose that A is an Orey sentence and suppose, for a reductio, that A is a Kent sentence. Then clearly $(\neg A \vee \Diamond \neg A)$ is provably equivalent to a Π_1 sentence and hence

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

mF (The modified Feferman System) The modified Feferman System is a modification of both the Feferman System and of the BM System. Define

$$\mathrm{mF}_0 = \varnothing$$

$$\mathrm{mF}_{n+1} = \begin{cases} \mathrm{mF}_n \cup \{A\}, & \text{if } \mathrm{Proof}(n,A) \text{ and } \mathrm{mF}_n \cup \{A\} \text{ is consistent} \\ \mathrm{mF}_n, & \text{otherwise.} \end{cases}$$

Let $\triangle_x A$ be the arithmetization of $A \in \mathrm{mF}_{x+1}$. Define further

$$\triangle A \Leftrightarrow \exists x \triangle_x A$$
$$\triangle^* A \Leftrightarrow \exists x \Box_{\triangle x} A.$$

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

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

This last observation brings out the Rosserlike character of $\triangle$.

We claim that $\triangle$ satisfies (B1)–(B7). The argument for the validity of (B1)–(B6) is similar to the one for the case of $\triangle^{\mathrm{F}}$. We treat of example (B5).

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

$$\begin{aligned}
\vdash \triangle A &\to \Box A\\
&\to \exists x \Box \Box_x A\\
&\to \Box \triangle A.
\end{aligned}$$

The argument for (B7) is similar to the one for the case of $\triangle^{\mathrm{mBM}}$. Just like $\triangle^{\mathrm{F}}, \triangle^{\mathrm{mF}}$ satisfies (32), i.e., $\triangle^{\mathrm{mF}}$ is provably Σ_1-complete. *Prima facie*, $\triangle$ is Σ_2. It is seen to be $\triangle_2$ by the following observation

$$\vdash \neg \triangle A \leftrightarrow (\neg \Box A \vee (\Box \bot \wedge \triangle \neg A)). \qquad \text{(B6),(B7)} \tag{36}$$

Concerning the Henkin sentences of $\triangle^{\mathrm{mF}}$ the same remarks can be made as for $\triangle^{\mathrm{F}}$. Just like $\triangle^{\mathrm{mBM}}$, $\triangle^{\mathrm{mF}}$ has at least two nonequivalent Gödel sentences. Clearly, $\triangle^{\mathrm{mF}}$ is precocious. It is now easy to see that the two nonequivalent Gödel sentences and their negations give us four pairwise nonequivalent Orey sentences. In Section 6 we will show that $\triangle^{\mathrm{mF}}$ has in fact infinitely many pairwise nonequivalent Gödel sentences; thus, there are infinitely many pairwise nonequivalent Orey sentences.

$\triangle^{\mathrm{mF}}$ is our final system and the main object of study of this paper. In Section 5 we will study the principles (B1)–(B7) from the modal point of view. In Section 6 we give a partial result on embedding Kripke models for our modal system into arithmetic. In Section 7 we will apply the result of Section 6 to relative interpretability.

5 *The system BMF*

5.1 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) $\vdash\phi \Rightarrow \vdash\Box\phi$
(L2) $\vdash\Box(\phi\to\psi)\to(\Box\phi\to\Box\psi)$
(L3) $\vdash\Box\phi\to\Box\Box\phi$
(L4) $\vdash\Box(\Box\phi\to\phi)\to\Box\phi$
(B1) $\vdash\phi \Rightarrow \vdash\triangle\phi$
(B2) $\vdash\triangle(\phi\to\psi)\to(\triangle\phi\to\triangle\psi)$
(B3) $\vdash\neg\triangle\bot$
(B4) $\vdash\Box\phi\to\triangle\Box\phi$
(B5) $\vdash\triangle\phi\to\Box\triangle\phi$
(B6) $\vdash\neg\Box\bot\to(\triangle\phi\leftrightarrow\Box\phi)$
(B7) $\vdash\Box\bot\to(\triangle\phi\vee\triangle\neg\phi)$.

This list is very long and rather redundant. A more economical list would consist of (B1), (B2), (B3), (B5), (B7), and the principles

(B8) $\vdash\Box\phi\leftrightarrow(\triangle\phi\vee\Box\bot)$
(B9) $\vdash\triangle(\Box\phi\to\phi)\to\triangle\phi$.

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

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

(a)	$\vdash\phi\to(\Box(\phi\wedge\Box\phi)\to(\phi\wedge\Box\phi))$	(L1)
(b)	$\vdash\triangle\phi\to\triangle(\Box(\phi\wedge\Box\phi)\to(\phi\wedge\Box\phi))$	(a),(B1),(B2)
(c)	$\vdash\triangle\phi\to\triangle(\phi\wedge\Box\phi)$	(b),(B9)
(d)	$\vdash\triangle\phi\to\triangle\Box\phi$	(c),(B1),(B2)
(e)	$\vdash\neg\Box\bot\to(\Box\phi\to\triangle\Box\phi)$	(d),(B6)
(f)	$\vdash\Box\bot\to(\triangle\Box\bot\vee\triangle\neg\Box\bot)$	(B7)
(g)	$\vdash\neg\triangle\neg\Box\bot$	(B3),(B9)
(h)	$\vdash\Box\bot\to\triangle\Box\bot$	(f),(g)
(i)	$\vdash\Box\bot\to\Box\phi$	(L1),(L2)
(j)	$\vdash\triangle\Box\bot\to\triangle\Box\phi$	(i),(B1),(B2)
(k)	$\vdash\Box\bot\to\triangle\Box\phi$	(h),(j)
(l)	$\vdash\Box\phi\to\triangle\Box\phi$.	(e),(h)

Finally, (L3) follows from (B4) and (B8).

We list a few further convenient consequences of BMF:

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

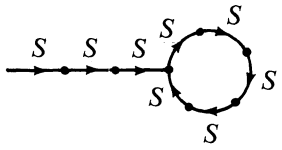
5.2 Nonuniqueness and nonexplicitness in BMF Clearly, the discussion in Section 4 on nonuniqueness and nonexplicitness of Henkin and Gödel sentences under mBM can be carried out in BMF; e.g., one can show

$$\vdash \Box(p \leftrightarrow \neg\triangle p) \rightarrow \Box(\triangledown p \leftrightarrow \neg\triangle\triangledown p)$$
$$\vdash \Box(p \leftrightarrow \neg\triangle p) \rightarrow (\Box(p \leftrightarrow \triangledown p) \rightarrow \Box\bot).$$

5.3 Kripke Semantics for BMF

5.3.1 Definitions

(a) Let K be a finite, nonempty set. Let S be a binary relation on K. The structure $\langle K, S\rangle$ is called *a lolly* if
  (i) for each k, k' in K, kS^Tk'. Here S^T is the transitive, symmetric, and reflexive closure of S
  (ii) for each k in K, there is *precisely one* k' in K such that kSk'. We will call k' with kSk' the *S-successor* of k
  (iii) there is *at most* one k in K such that, for no k' in K, $k'Sk$.
  It should be clear that a lolly looks like this:



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

(c) A structure $\langle K, R, S\rangle$ is called a *lolly-frame* if K is nonempty, R and S are binary relations in K, and
  (i) R is transitive
  (ii) R is upwards wellfounded
  Let $K_0 = \{k \text{ in } K \mid \text{for no } k' \text{ in } K,\ kRk'\}$
  $K_1 = K \setminus K_0$
  $S_0 = S \restriction K_0$
  S_0^T = the transitive, symmetric, and reflexive closure of S_0.
  (iii) $k \in K_1 \Rightarrow (kSk' \Leftrightarrow kRk')$
  (iv) Suppose that $k \in K_0$. Let $[k] = \{k' \mid k'S_0^Tk\}$. Then $\langle [k], S_0 \restriction [k]\rangle$ 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 $\langle K, R, S, \Vdash\rangle$, where $\langle K, R, S\rangle$ 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) $k \Vdash (\phi \wedge \psi) \Leftrightarrow (k \Vdash \phi \text{ and } k \Vdash \psi)$
  (iv) $k \Vdash (\phi \vee \psi) \Leftrightarrow (k \Vdash \phi \text{ or } k \Vdash \psi)$

(v) $k \Vdash (\phi \to \psi) \Leftrightarrow (k \Vdash \phi \Rightarrow k \Vdash \psi)$
(vi) $k \Vdash \neg\phi \Leftrightarrow k \nVdash \phi$
(vii) $k \Vdash \Box\phi \Leftrightarrow$ for all k' such that kRk', $k' \Vdash \phi$
(viii) $k \Vdash \triangle\phi \Leftrightarrow$ for all k' such that kSk', $k' \Vdash \phi$.

5.3.2 Remark It is easy to verify that lolly-frames also satisfy

$$kRk'Sk'' \Rightarrow kRk'', \text{ and } kSk'Rk'' \Rightarrow kRk''.$$

A lolly-frame is best visualized as a conventional frame for provability logic where the top nodes are blown up to lollies, as shown in Figure 2. Here, e.g., Figure 3 means Figure 4. Note that we don't draw the arrows to exhibit the transitivity of R. Also, since $R \subseteq S$, we don't write 'S' next to R-arrows.

5.3.3 Soundness Consider any lolly-model $\mathbf{K} = \langle K, R, S, \Vdash \rangle$. We write $\mathbf{K} \Vdash \phi$ for: for all $k \in K$, $k \Vdash \phi$. We then have that BMF $\vdash \phi \Rightarrow \mathbf{K} \Vdash \phi$.

Proof: The proof is entirely routine.

5.3.4 Completeness *Suppose that BMF* $\nvdash \phi$; *then there is a finite lolly-model* $\mathbf{K}$ *such that* $\mathbf{K} \nVdash \phi$.

We proceed with some preliminaries for the proof of 5.3.4.

5.3.5 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 $\phi \in \Delta$ such that $\phi \in \Gamma$.

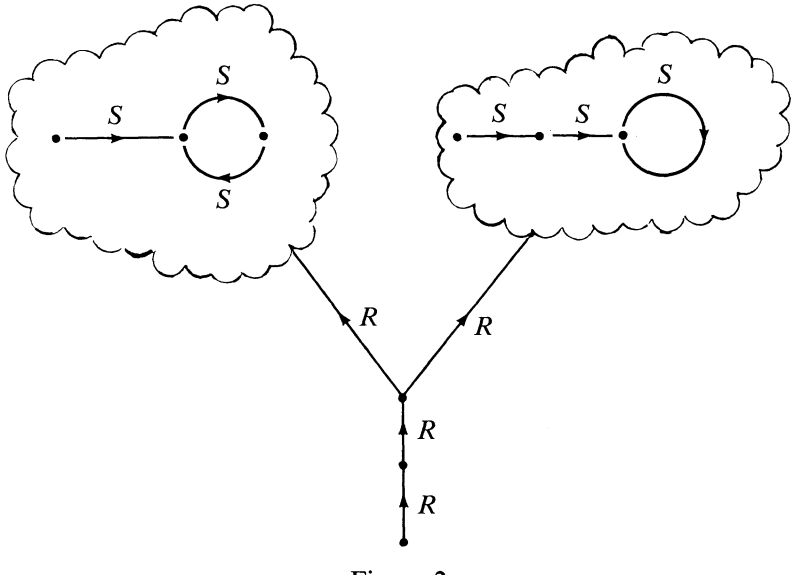


Figure 2.

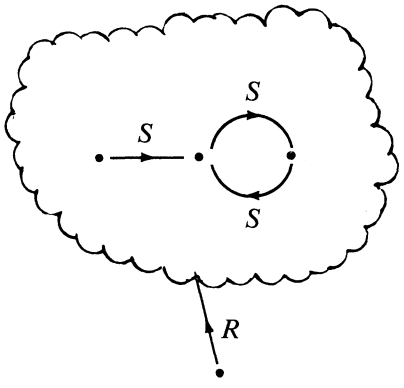


Figure 3.

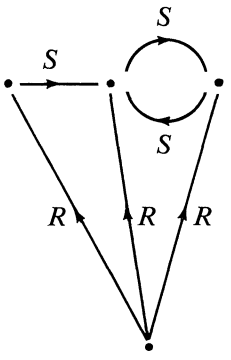


Figure 4.

5.3.6 Lemma *Suppose that* $\Gamma \nvdash \Delta$, *and 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: The proof is entirely routine.

5.3.7 The Henkin construction Let X be a finite set of formulas that is closed under subformulas, such that $\neg\Box\bot \in X$ and $(\triangle\phi \in X \Leftrightarrow \Box\phi \in X)$. We now 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 $\triangle\psi$ and both ψ and $\triangle\triangle\psi$ are in y.

Clearly, y consists of elements of X plus, for certain χ in $X \cap y$, $\triangle\chi, \triangle\triangle\chi, \triangle\triangle\triangle\chi, \ldots$. As is easily seen, K is finite and nonempty (by 5.3.6). For $x, y \in k$ we define:

$$xRy \Leftrightarrow (\Box\phi \in x \Rightarrow \phi, \triangle\phi, \triangle^2\phi, \ldots \in y)$$
$$\text{and (there is a } \Box\psi \in y \text{ with } \Box\psi \notin x)$$

$$\text{xSy} \Leftrightarrow ((\neg\Box\bot) \in \text{x and } xRy)$$
$$\text{or } (\Box\bot \in x, \Box\bot \in y \text{ and } (\triangle\phi \in x \Rightarrow \phi \in y))$$
$$\text{and } ((\triangle\phi \notin x \text{ and } \triangle\phi \in X) \Rightarrow \phi \notin y)).$$

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 $xRySz \Rightarrow xRz$.

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

Claim 3 *For* $\phi \in X$, $x \Vdash \phi \Leftrightarrow \phi \in x$.

Proof: We prove this claim by induction on ϕ in X. The only nontrivial cases are when ϕ is of the form $\Box\psi$ or $\triangle\psi$.

Case 1: Suppose that $\phi = \Box\psi$.
Subcase 1.1: Suppose that $\Box\psi \in x$ and xRy. Then $\psi \in y$, hence by IH $y \Vdash \psi$. Therefore, $x \Vdash \Box\psi$.
Subcase 1.2: Suppose that $\Box\psi \notin x$. Let $x^R = \{\chi, \Delta\chi, \Delta^2\chi, \ldots \mid \Box\chi \in x\}$. We claim that $x^R, \Box\psi \nVdash \psi$, for 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 5.3.6 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, xRy and $\psi \notin y$ and so by IH $y \nVdash \psi$. Hence $x \nVdash \Box\psi$.

Case 2: Suppose that $\phi = \Delta\psi$. In case $(\neg\Box\bot) \in x$, this reduces to the previous case. So we assume that $\Box\bot \in x$.
Subcase 2.1: Suppose that $\Delta\psi \in x$ and xSy. Then $\psi \in y$, hence by IH $y \Vdash \psi$. So we conclude that $x \Vdash \Delta\psi$.
Subcase 2.2: Suppose that $\Delta\psi \notin x$. 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 that $x^S \nVdash x_S$, for otherwise $x \vdash \Delta\mathbb{W}x_S$, ergo (by (B12), using the fact that $\Box\bot \in x$) $x \vdash \mathbb{W}\{\Delta\chi \mid \chi \in x_S\}$. Hence $x \vdash \Delta\chi$ for *some* χ in x_S, which is a contradiction. By 5.3.6 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 now show that $y \in K$. Suppose that $\nu \in y$ and $\nu \notin X$. Clearly $\nu \in x^S$, hence $\Delta\nu \in x$. Since $\Delta\nu \notin X$, ν and $\Delta\Delta\nu$ are in x, and since $\nu \in x$ and $\nu \notin X$, ν must be of the form $\Delta\rho$. We conclude that ρ and $\Delta\Delta\rho$ are in x^S. Next we show that xSy. 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 $\chi \notin y$. So we conclude that xSy.

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

Claim 4 *There is a y such that $xRy \Leftrightarrow (\neg\Box\bot) \in x$.*

Claim 5 *For every x there is a y such that xSy.*

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 needed. 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

(i) R' is empty
(ii) $yS'z \Leftrightarrow y = x_j$ and $z = x_{j+1}$ for some $j \in \{0, \ldots, n\}$
(iii) $y \Vdash' p_i \Leftrightarrow p_i \in y$.

Claim 6 *For $y \in \{x_0, \ldots, x_n\}$ and $\phi \in X$, $y \Vdash' \phi \Leftrightarrow y \Vdash \phi$.*

Proof: By induction on ϕ in X for all x_j simultaneously. The atomic case and the cases of $\wedge$, $\vee$, $\neg$, and $\to$ 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 that $\phi = \Delta\psi$. 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 $x_i \Vdash' \Delta\psi \Leftrightarrow x_{i+1} \Vdash' \psi \overset{\text{IH}}{\Leftrightarrow} x_{i+1} \Vdash \psi \Leftrightarrow x_i \Vdash \Delta\psi$.

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

$K^* = \{x \in K \mid \neg\Box\bot \in x\} \cup \{\langle y,x\rangle \mid \Box\bot \in x \in K$, where y 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 $L_x\}$
$K_1^* = \{x \in K \mid \neg\Box\bot \in x\}$
$uR^*v \Leftrightarrow (u,v$ are in K_1^* and $uRv)$ or
$(u$ is in K_1^*, v is in K_0^*, where $v = \langle y,x\rangle$ and $uRx)$
$uS^*v \Leftrightarrow uR^*v$ or $(u$ is in K_0^*, v is in K_0^*, $u = \langle y,x\rangle$, $v = \langle z,x\rangle$ and $yS'z$, 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 $p_i \in y)$
$\mathbf{K}^* = \langle K^*, R^*, S^*, \Vdash^*\rangle$.

We claim that
(A) $\mathbf{K}^*$ is a lolly-model
(B) If $u \in K_1^*$, then $(u \Vdash^* \phi \Leftrightarrow u \Vdash \phi)$ for $\phi \in X$. If $u \in K_0^*$ and $u = \langle y,x\rangle$, then $(u \Vdash^* \phi \Leftrightarrow y \Vdash \phi)$ for $\phi \in X$.

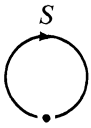
Proof of (A): One easily verifies that K^* is finite and that R^* is transitive, irreflexive, and hence upwards wellfounded. Moreover, $u \in K_1^* \Leftrightarrow$ there is a v in K^* such that uR^*v. The definition of S^* implies that $u \in K_1^* \Rightarrow (uS^*v \Leftrightarrow uR^*v)$. We leave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

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

- If $u \in K_0^*$ and $u = \langle y,x\rangle$ then $u \Vdash^* \phi \Leftrightarrow y \Vdash' \phi \Leftrightarrow y \Vdash \phi$.
  (The first equivalence is by a completely trivial induction.)
- Suppose that $u \in K_1^*$. The case where ϕ is atomic is trivial and so are the cases of $\wedge$, $\vee$, $\rightarrow$, and $\neg$.
  - Suppose that $\phi = \Box\psi$
    - Suppose that $u \Vdash \Box\psi$ and uR^*v. If v is in K_1^*, we have that uRv, hence $v \Vdash \psi$, so by IH $v \Vdash^* \psi$. If v is in K_0^*, say $v = \langle y,x\rangle$, we have that uRx. Using Claim 2, we can show that uRy. It follows that $y \Vdash \psi$. Hence by IH $v \Vdash^* \psi$, so we conclude that $u \Vdash^* \Box\psi$.
    - Suppose that $u \Vdash^* \Box\psi$ and uRy. If $(\neg\Box\bot) \in y$, we have that 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 $y \Vdash \psi$, so we conclude that $u \Vdash \Box\psi$.
  - The case where $\phi = \triangle\psi$ is similar.

Proof of 5.3.4: Suppose that BMF $\nVdash \phi$. Let X_0 be the smallest set that is closed under subformulas and contains ϕ and $\neg\Box\bot$, and let $X = X_0 \cup \{\triangle\psi \mid \Box\psi \in X_0\} \cup \{\Box\psi \mid \triangle\psi \in X_0\}$. Construct a finite lolly-model $\mathbf{K}^*$ as in 5.3.7 for X. By 5.3.6 there is an X-saturated $x_0 \subseteq X$ such that $x_0 \nVdash \phi$. x_0 will correspond to a node of $\mathbf{K}^*$, say u, and $u \nVdash^* \phi$.

5.3.8 Application In Section 4 under mBM we showed that neither in BMF nor in PA is there an explicit Gödel sentence for $\triangle$, where $\triangle$ is interpreted in PA as $\triangle^{mBM}$, or $\triangle^{F}$, or $\triangle^{mF}$. In the case of BMF this fact can be easily shown by considering the following Kripke Model:



6 Embedding circle-tail models in arithmetic We would like to generalize the result of Solovay [17] to the logic BMF, interpreting $\triangle$ as $\triangle^{MF}$. To do this we must embed lolly-models in arithmetic. This program, however, meets with a difficulty I could not 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 the powerful machinery to construct arithmetical sentences (see also Section 7). 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 section I follow the presentation of [20].)

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 Figure 5. We can arrange it so that the nodes of the finite model at the top are numbered $1, \ldots, N$, and the nodes of the down-going tail (except the bottom) $N + 1$, $N + 2, \ldots$, and the bottom is numbered by 0. The nodes numbered $N + 1$, $N + 2, \ldots, 0$ will be called *tail elements.* 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-tail models. Clearly, a circle-tail model *is* a circle-model.

An immediate consequence of our definition is

6.1 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: By 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-tail models satisfy the principle:

(C) $\vdash \Box(\Box\bot \to \triangle\phi) \to \Box(\Box\bot \to \phi)$.

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

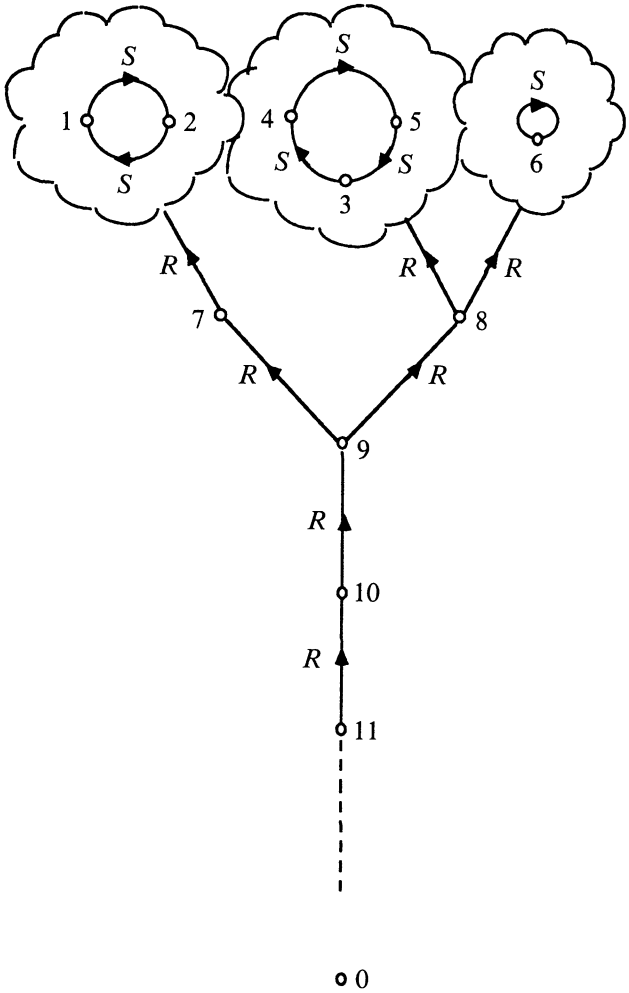


Figure 5

Open problem Is (C) arithmetically valid when $\triangle$ is interpreted as $\triangle^{mF}$?

For the rest of this section the following are to be kept in mind:

(i) We fix a circle-tail model **K**
(ii) We assume that **K** is suitably described in arithmetic. Specifically, we assume that R and S are given $\triangle_0$ definitions in such a way that all their simple properties are verifiable
(iii) We interpret $\triangle$ as $\triangle^{mF}$ 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 now need to 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. To this end, define by the Recursion Theorem

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

$$h0 = 0$$

$$h(k+1) = \begin{cases} a, & \text{if for some } a \text{ such that } hkRa\ \mathrm{Proof}(k, \neg\mathrm{COF}a) \\ a, & \text{if for some } a \text{ such that } hkSa\ \mathrm{Proof}(k, \neg\mathrm{COF}a) \\ & \text{and } (\neg\mathrm{COF}a) \in \mathrm{mF}_{k+1} \\ hk, & \text{otherwise.} \end{cases}$$

It is easy to see that the arithmetization of '$(\neg\mathrm{COF}a) \in \mathrm{mF}_{k+1}$' is $\triangle_2$ and hence that h is $\triangle_2$. An important difference with Solovay's original construction is that we use '$\mathrm{COF}a$' instead of '$l = a$'. Later we will see that $\vdash \mathrm{COF}a \leftrightarrow l = a$; but to show this we need h to be defined using 'COF' rather than 'l'.

We now prove a sequence of lemmas about h.

6.2 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 $x + z = y$.

6.3 Lemma
(i) $\vdash ((\forall z < x\ hz \in K_1) \wedge hx = y) \rightarrow \triangle hx = y$
(ii) $\vdash \forall x \forall y (hx = y \rightarrow \Box hx = y)$.

Proof of (i): The proof is conducted in PA as follows. The proof is by induction on x. The case where $x = \underline{0}$ is trivial. Suppose that $x = u + \underline{1}$, $\forall z < x\ hz \in K_1$, $hx = y$, and $hu = v$. There are three possibilities:

(a) hx was computed by the first clause of the definition of h. So we then have that vRy and $\mathrm{Proof}(u, \neg\mathrm{COF}y)$, hence, $\triangle vRy$ and $\triangle\mathrm{Proof}(u, \neg\mathrm{COF}y)$. By the induction hypothesis, $\triangle hu = v$, so we conclude that $\triangle hx = y$.
(b) hx was computed by the second clause of the definition of h. So we then have that vSy and $\mathrm{Proof}(u, \neg\mathrm{COF}y)$. Since $v \in K_1$, we also have vRy. So hx was also computed by the first clause.
(c) hx was computed by the third clause. Clearly, $hu = v = y = hx$ and $\forall z < u\ hz \in K_1$. By the induction hypothesis $\triangle hu = v$. Moreover, either for no w $\mathrm{Proof}(u, \neg\mathrm{COF}w)$ or for some w $\mathrm{Proof}(u, \neg\mathrm{COF}w)$ and not vRw (whence not vSw, since by hypothesis $u < x$ implies that $v \in K_1$).

Hence we have $\triangle \forall w < u \neg Proof(u, \neg\mathrm{COF}w)$ or $(\triangle\mathrm{Proof}(u, \neg\mathrm{COF}w)$ and $\triangle(\neg vRw \wedge \neg vSw))$, so we conclude that $\triangle hx = y$.

Proof of (ii): By the argument used in the proof that $\vdash \triangle A \to \Box \triangle A$ one shows that $\vdash \forall u \forall v (u \in \mathrm{mF}_v \to \Box (u \in \mathrm{mF}_v))$. The claim now follows by an easy induction in PA.

We define $\mathrm{LIM}a \Leftrightarrow \exists x\, hx = a \wedge \forall x \forall y ((hx = a \wedge x \leq y) \to hy = a)$.

6.4 Lemma $\vdash \forall a (\mathrm{COF}a \to \mathrm{LIM}a)$.

Proof: By 6.2 it is clear that $\vdash \forall a ((\mathrm{COF}a \wedge a \in K_1) \to \mathrm{LIM}a)$, so it is sufficient to show that $\vdash \forall a ((\mathrm{COF}a \wedge a \in K_0) \to \mathrm{LIM}a)$. We reason in PA as follows:

Suppose that $\mathrm{COF}a$ and $a \in K_0$. Assume that a is on the circle C with, say, $a = \underline{a}_1 S \underline{a}_2 S \ldots S \underline{a}_n S \underline{a}_{n+1} = \underline{a}_1$. 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 6.3(i) $\triangle h(x_0 + 1) = \underline{a}_j$, so we conclude that $\triangle \bigvee \{\mathrm{COF}\underline{a}_i \mid i = 1, \ldots, n\}$.

Now suppose for a reductio that $\neg \mathrm{LIM}a$. Clearly, by 6.2 $\mathrm{COF}\underline{a}_1$, $\mathrm{COF}\underline{a}_2$, $\ldots, \mathrm{COF}\underline{a}_n$. It follows from the definition of h that $\triangle \neg \mathrm{COF}\underline{a}_1, \triangle \neg \mathrm{COF}\underline{a}_2$, $\ldots, \triangle \neg \mathrm{COF}\underline{a}_n$ (or how else could h move on and on?). Hence, $\triangle \bigvee \{\mathrm{COF}\underline{a}_i \mid i = 1, \ldots, n\}$ and $\triangle \bigwedge \{\neg \mathrm{COF}\underline{a}_i \mid i = 1, \ldots n\}$, therefore $\triangle \bot$ and thus $\bot$. So we conclude that $\mathrm{LIM}a$.

6.5 Lemma $\vdash \exists a \mathrm{LIM}a$.

Proof: It is easily seen that $\vdash \exists a \mathrm{COF}a$.

6.6 Lemma $\vdash \exists x\, hx \in K_0 \leftrightarrow \Box \bot$.

Proof: Reason in PA as follows:

Right to left is trivial.

From left to right, suppose that $hx = \underline{a}_1$, where $\underline{a}_1$ is on the circle C, given by $\underline{a}_1 S \underline{a}_2 S \ldots S \underline{a}_n S \underline{a}_{n+1} = \underline{a}_1$. We have that $\Box hx = \underline{a}_1$ by 6.3(ii), hence $\Box \bigvee \{\mathrm{COF}\underline{a}_i \mid i = 1, \ldots, n\}$. h moved up to $\underline{a}_1$ by the first or by the second clause. In either case we have $\Box \neg \mathrm{COF}\underline{a}_1$.

We now show for $k = 0, \ldots, n-1$ that $\bigwedge \{\Box \triangle^k \neg \mathrm{COF}\underline{a}_j \mid j = 1, \ldots, k+1\}$, by (external) induction on k. The case where $k = 0$ is simply $\Box \neg \mathrm{COF}\underline{a}_1$. Suppose that $\bigwedge \{\Box \triangle^k \neg \mathrm{COF}\underline{a}_j \mid j = 1, \ldots, k+1\}$. By (B11) $\bigwedge \{\Box \triangle^{k+1} \neg \mathrm{COF}\underline{a}_j \mid j = 1, \ldots, k+1\}$. We now need to show $\Box \triangle^{k+1} \neg \mathrm{COF}\underline{a}_{k+2}$. Clearly

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

hence

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

We conclude using (L1), (L2), (B1), (B2), and (B11) that

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

Moreover, by (B11) we have from $\Box hx = \underline{a}_1$ that $\Box \triangle^k hx = \underline{a}_1$. So finally

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

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

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

6.7 Lemma

(i) $\vdash(\mathrm{COF}u \wedge uSv) \rightarrow \nabla\mathrm{COF}v$
(ii) $\vdash(y \in K_0 \wedge \mathrm{COF}y) \rightarrow \triangle\mathrm{COF}\sigma y$
(iii) $\vdash(y \in K_0 \wedge \Box\bot) \rightarrow (\triangle\mathrm{COF}\sigma y \rightarrow \mathrm{COF}y)$
(iv) $\vdash(y \in K_0 \wedge \Box\bot) \rightarrow (\mathrm{COF}y \leftrightarrow \triangle\mathrm{COF}\sigma y)$.

Proof:

(i) Reason in PA as follows. Suppose that $\mathrm{COF}u$, uSv, and $\triangle\neg\mathrm{COF}v$. By 6.4 $\mathrm{LIM}u$. Suppose that $hx = u$ and for all $y \geq x$ $hy = u$. For some z $(\neg\mathrm{COF}v) \in \mathrm{mF}_z$. Consider w such that $w > x$, $w > z$, and $\mathrm{Proof}(w, \neg\mathrm{COF}v)$. Clearly, $(\neg\mathrm{COF}v) \in \mathrm{mF}_{w+1}$. Hence h would move up to v at $w + 1$. *Quod non.* So we conclude that $\neg\triangle\neg\mathrm{COF}v$.

(ii) Immediate from (i) using $\vdash(y \in K_0 \wedge \mathrm{COF}y) \rightarrow \Box\bot$, which follows directly from 6.6, and $\vdash\Box\bot \rightarrow (\nabla A \leftrightarrow \triangle A)$.

(iii) Reason in PA as follows. Suppose that $y \in K_0$, $\Box\bot$, and $\triangle\mathrm{COF}\sigma y$. From $\Box\bot$ we have by 6.6 that for some $z \in K_0$ $\mathrm{COF}z$. By (ii) $\triangle\mathrm{COF}\sigma z$. Hence by 6.4, (B1), (B2), and $\triangle\mathrm{COF}\sigma y$, $\triangle y = z$ and thus $y = z$. (We have Π_1-Reflection for $\triangle$!)

(iv) By (ii) and (iii).

6.8 Definitions

(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 '$\Box$' shifts its meaning!)
- $(\triangle\phi)^f = \triangle(\phi)^f$

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

$$[\phi] = \mathbb{W}\{\mathrm{COF}\underline{i} \mid i \Vdash \phi\} \qquad \text{(we take } \mathbb{W}\varnothing = (\underline{0} = \underline{1})$$

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

$$[\phi] = \text{M}\{\neg\mathrm{COF}\underline{i} \mid i \nVdash \phi\} \qquad \text{(we take } \text{M}\varnothing = (\underline{0} = \underline{0})$$

Note that $[\phi]$ is simply an arithmetization of $\exists x \in [\![\phi]\!]\ \mathrm{COF}x$

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

6.9 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 $\triangle$. The cases of the propositional constants are trivial (using 6.4). We show that (i) $\vdash[\Box\psi] \leftrightarrow \Box[\psi]$ and (ii) $\vdash[\triangle\psi] \leftrightarrow \triangle[\psi]$.

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

From right to left, 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 that $\Box[\psi]$. Clearly, $\Box\neg\mathrm{COF}\underline{j}_k$. Suppose that $\mathrm{Proof}(p, \neg\mathrm{COF}\underline{j}_k)$ 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 $\mathrm{COF}x$, 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, $\mathrm{COF}x \to \neg xR\underline{j}_k$.
On the other hand, it is easily seen that if $x \nVdash \Box\psi$ then $xR\underline{j}_k$ for some k. Hence, $\mathrm{COF}x \to x \Vdash \Box\psi$, so we conclude that $[\Box\psi]$.

From left to right, suppose that $\mathrm{COF}\underline{i}$ for some $\underline{i}$ such that $\underline{i} \Vdash \Box\psi$. 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 that $\Box\neg\mathrm{COF}\underline{i}$. Suppose that $hx = \underline{i}$. By 6.3(ii), $\Box hx = \underline{i}$. If $\underline{i} \in K_0$ we have by 6.6 $\Box\bot$, and hence $\Box[\psi]$. If $\underline{i} \in K_1$ we see that $\Box\neg\mathrm{COF}\underline{i}$ and $\Box hx = \underline{i}$ imply $\Box\forall y(\mathrm{COF}y \to \underline{i}Ry)$, thus we conclude that $\Box[\psi]$.

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

To complete the argument we need to show that $\vdash\Box\bot \to ([\triangle\psi] \leftrightarrow \triangle[\psi])$.

$$
\begin{aligned}
\vdash\Box\bot \to ([\triangle\psi] &\leftrightarrow ([\triangle\psi] \wedge \Box\bot) \\
&\leftrightarrow [\triangle\psi \wedge \Box\bot] \\
&\leftrightarrow \mathbb{W}\{\mathrm{COF}\underline{j} \mid j \Vdash \triangle\psi \wedge \Box\bot\} \\
&\leftrightarrow \mathbb{W}\{\mathrm{COF}\pi\underline{i} \mid i \Vdash \psi \wedge \Box\bot\} \\
&\leftrightarrow \mathbb{W}\{\triangle\mathrm{COF}\underline{i} \mid i \Vdash \psi \wedge \Box\bot\} && \text{(6.7(iv))} \\
&\leftrightarrow \triangle\mathbb{W}\{\mathrm{COF}\underline{i} \mid i \Vdash \psi \wedge \Box\bot\} && \text{(B12)} \\
&\leftrightarrow \triangle[\psi \wedge \Box\bot] \\
&\leftrightarrow \triangle[\psi]). && \text{(B1),(B2),(B4)}
\end{aligned}
$$

6.10 Remark The reduction result proved as (33) in Section 4 clearly applies to $\triangle^{\mathrm{mF}}$. It implies that for the arithmetical embedding of *traditional* tail models we have that $\vdash\triangle[\phi] \leftrightarrow (\Box[\phi] \wedge (\Box\bot \to [\phi]))$. We can now understand this result in a new way: the arithmetical embedding of traditional tail models is similar to the arithmetical embedding of circle-tail models *which have just singleton circles*! (This point will become even clearer in the light of Lemma 7.3.)

6.11 Application There are infinitely many nonequivalent Gödel sentences for $\triangle^{\mathrm{mF}}$.

Proof: It is clearly sufficient to prove that for any n there are n nonequivalent Gödel sentences for $\triangle$. Consider the circle-tail model shown in Figure 6.

Let s be a sequence $c_1c_2\ldots c_n$ of 0's and 1's. Consider an atom p_s. Let

$$
\begin{aligned}
&a_{0i} \Vdash p_s \Leftrightarrow c_i = 0 \\
&a_{1i} \Vdash p_s \Leftrightarrow c_i = 1 \\
&b_j \Vdash p_s \text{ for all } j \\
&0 \Vdash p_s.
\end{aligned}
$$

Define $G_s = [p_s]$. It follows immediately from the Embedding Theorem that $\vdash G_s \leftrightarrow \neg\triangle G_s$. Moreover, if $s \neq s'$ then $\vdash\Box(G_s \leftrightarrow G_{s'}) \to \Box\bot$.

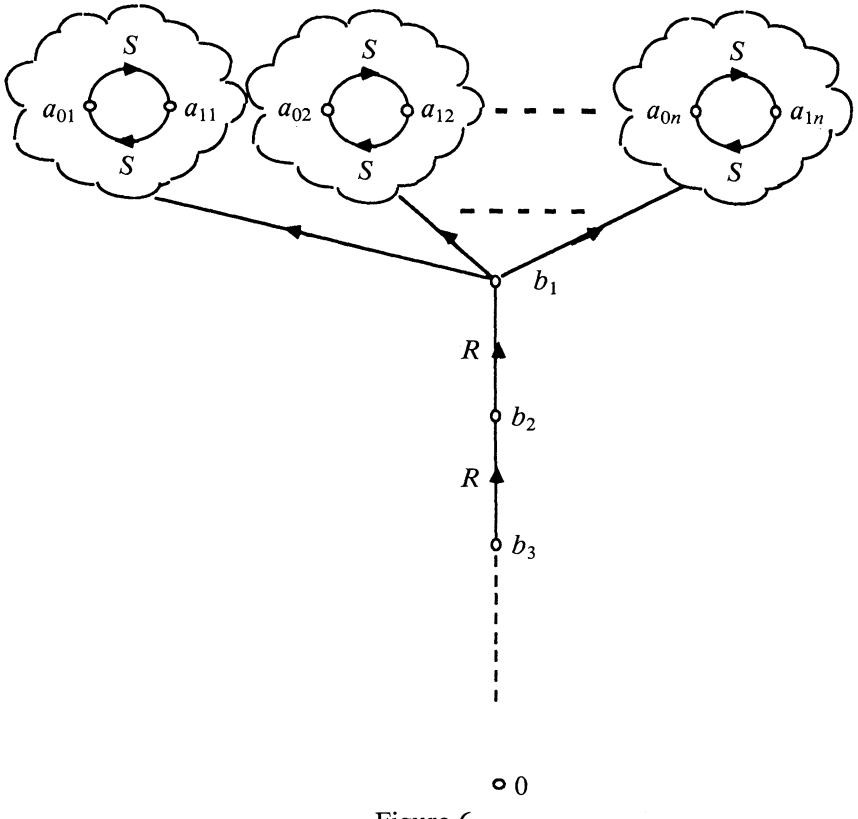


Figure 6.

Because Gödel sentences of $\triangle$ are Orey sentences it follows that there are infinitely many nonequivalent Orey sentences.

7 $\triangle^{\mathbf{mF}}$ ***meets relative interpretability*** In this section '$\triangle$' will stand for '$\triangle^{\mathrm{mF}}$' in arithmetical contexts, '$\vdash$' will stand for 'PA$\vdash$'. We fix a circle-tail model **K**.

For convenience we now repeat the derivability conditions we collected for relative interpretability in 3.5:

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

We add the (for our purposes) essential (34) of Section 4

(J3) $\vdash A \lhd \nabla A$.

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

We list some immediate consequences of our list:

(J4) $\vdash A \lhd \triangle A$. (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) $\vdash \bigwedge\{A_i \lhd B_i \mid i = 1, \ldots, n\}$
$\rightarrow \bigvee\{A_i \mid i = 1, \ldots, n\} \lhd \bigvee[B_i \mid i = 1, \ldots, n\}$ (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 now wish to take a closer look at the interaction between $\lhd$ and the sentences $[\phi]$ constructed in Section 6. The classes of sentences $[\phi]$, constructed for different circle-tail models, are too poor to refute all modal principles *not* valid in PA in a language with $\Box$ and $\lhd$. 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 to $[\![\phi]\!]$ all circles C such that $C \cap [\![\phi]\!] \neq \emptyset$.

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

To prove this we need a few lemmas.

7.2 Definition We define a *recursive* function h_0 as follows:

$$h_0 0 = 0$$

$$h_0(k+1) = \begin{cases} a, & \text{if for some } a \text{ such that } h_0 k R a, \text{Proof}(k, \neg \text{COF} a) \\ hk, & \text{otherwise.} \end{cases}$$

Here 'COF' is as in Section 6. Note that COF is based on h and not on h_0.

7.3 Lemma
(i) $\vdash (\forall z < x\ hz \in K_1) \rightarrow hx = h_0 x$
(ii) *Let S^* be the transitive, reflexive closure of S; then* $\vdash \forall x\ h_0 x S^* hx$.

Proof: The proof in both cases is by a simple induction on x in PA. These inductions are much like the proof of 6.3(i).

7.4 Corollary $[\phi]$ *is* $\triangle_2$.

Proof: It is clearly sufficient to show that sentences of the form $\mathrm{COF}\underline{i}$ are $\triangle_2$. In case $i \in K_0$ we have by 6.9 that $\vdash \mathrm{COF}\underline{i} \leftrightarrow \triangle\mathrm{COF}\sigma\underline{i}$, hence $\mathrm{COF}\underline{i}$ is in $\triangle_2$. In case $i \in K_1$ we have by 6.4 and 7.3 that

$$\vdash \mathrm{COF}\underline{i} \leftrightarrow (\exists x\, h_0 x = \underline{i} \wedge \forall x \forall y((h_0 x = \underline{i} \wedge x \leq y) \rightarrow h_0 = \underline{i})).$$

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

7.6 Lemma *Suppose that* $[\![\phi]\!]$ *is upwards persistent. Then* $[\phi]$ *is provably equivalent to a* Σ_1 *sentence.*

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

$$\vdash [\phi] \leftrightarrow \mathbb{W}\{\exists x\, h_0 x = \underline{i} \mid i \Vdash \phi\}.$$

We reason in PA as follows:

From right to left, suppose that $h_0 x = \underline{i}$ for $\underline{i} \in [\![\phi]\!]$. $\underline{i}S^* hx$ by 7.3(ii), hence by the upwards persistence of $[\![\phi]\!]$, $hx \in [\![\phi]\!]$. Thus $\forall z > x\, hz \in [\![\phi]\!]$, so we conclude that $[\phi]$.

From left to right, suppose that $\mathrm{COF}\underline{i}$ for $\underline{i} \in [\![\phi]\!]$. In case $\underline{i} \in K_1$ we have by 7.3(i) $\exists x\, h_0 x = \underline{i}$. Suppose that $\underline{i} \in K_0$, 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 7.3(i) $h_0 y = u$.

Then $[\![\phi]\!]$ is upwards persistent, $\underline{i}$ is in $[\![\phi]\!]$, $\underline{i}$ is on C, hence $C \subseteq [\![\phi]\!]$. We conclude that $u \in [\![\phi]\!]$, and so $\exists y\, h_0 y \in [\![\phi]\!]$.

7.7 Lemma *Suppose that* $\underline{i}$ *is on circle* C. *Then* $\vdash \mathrm{COF}\underline{i} \lhd \mathbb{W}\{\mathrm{COF}\underline{j} \mid j \in C\}$.

Proof: Reason in PA as follows: By 6.7(ii) we have that $\Box(\mathrm{COF}\pi\underline{i} \rightarrow \triangle\mathrm{COF}\underline{i})$; hence by (I1) $(\triangle\mathrm{COF}\underline{i}) \lhd \mathrm{COF}\pi\underline{i}$. By (J4) and (I2)

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

and similarly we have that

$$\begin{array}{c} \mathrm{COF}\pi\underline{i} \lhd \mathrm{COF}\pi^2\underline{i} \\ \vdots \\ \mathrm{COF}\pi^{n-2}\underline{i} \lhd \mathrm{COF}\pi^{n-1}\underline{i}. \end{array}$$

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

$$\begin{array}{l} \mathrm{COF}\underline{i} \lhd \mathrm{COF}\underline{i} \\ \mathrm{COF}\underline{i} \lhd \mathrm{COF}\ \pi\underline{i} \\ \quad\vdots \\ \mathrm{COF}\underline{i} \lhd \mathrm{COF}\pi^{n-1}\underline{i}; \end{array}$$

hence by (I3)

$$\mathrm{COF}\underline{i} \lhd \mathbb{W}\{\mathrm{COF}\pi^k\underline{i} \mid 0 \leq k < n\}.$$

In other words

$$\mathrm{COF}\underline{i} \lhd \mathbb{W}\{\mathrm{COF}\underline{j} \mid j \in C\}.$$

7.8 Lemma $\vdash [\phi] \equiv [\phi^*]$.

Proof: It is immediate that $\vdash [\phi^*] \lhd [\phi]$, so we need to show that $\vdash [\phi] \lhd [\phi^*]$. Reason in PA as follows: First, note that by 7.9 $\Box([\phi] \leftrightarrow ([\phi \wedge \Box\bot] \vee [\phi \wedge \neg\Box\bot]))$. 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 that $[\![\phi]\!] \subseteq K_0$. Clearly $[\![\phi^*]\!]$ consists precisely of those circles C such that $[\![\phi]\!] \cap C$ is not empty. We have by 7.7 and (J6)

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

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

7.9 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 7.6, and we are done.

Proof of 7.1: We have

$$\begin{aligned} \vdash [\phi] \lhd A &\leftrightarrow [\phi^*] \lhd A && (7.8),(J5)\\ &\leftrightarrow \Box(A \rightarrow ([\phi^*] \vee \Diamond[\phi^*])). && (7.9),(J8) \end{aligned}$$

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

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

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

7.11 On a question of Orey Orey asks: for which sets Γ of propositional formulas in the variables $p_1, \ldots, p_n$ are there arithmetical sentences $B_1, \ldots, B_n$ such that $\Gamma = \{\phi \mid \phi(B_1, \ldots, B_n) \lhd \top\}$? (I learned this formulation of Orey's problem from Per Lindström. Actually, the question is asked 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 us say that $\{\phi \mid \phi(B_1, \ldots, B_n) \lhd \top\}$ is *the interpretability class* of $B_1, \ldots, B_n$. A moment's reflection shows that interpretability classes Γ should satisfy

(i) $\top \in \Gamma$
(ii) $\bot \notin \Gamma$
(iii) $\phi \vdash_{\mathrm{Prop}} \psi$ and $\phi \in \Gamma \Rightarrow \psi \in \Gamma$.

We will show that, conversely, every set Γ of propositional formulas in $p_1, \ldots, p_n$ satisfying (i), (ii), and (iii) is an interpretability class.

Proof: Let Γ be a class of propositional formulas in $p_1, \ldots, p_n$ satisfying (i), (ii), and (iii). The plan of the proof is to construct a circle-tail model $\mathbf{K}$ and to take $B_i = [p_i]$. 7.10 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 then 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 represen-

tatives of the equivalence classes of Γ and let $\psi_1, \ldots, \psi_m$ be representatives of the equivalence classes of Γ^C. Define:

$$K_0 = \{\langle i,j\rangle | 1 \leq i \leq k,\ 1 \leq j \leq m\}$$
$$\langle i,j\rangle S\langle i',j'\rangle \Leftrightarrow j = j' \text{ and } ((1 \leq i < k \text{ and } i' = i + 1) \text{ or } (i = k \text{ and } i' = 1)).$$

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

Consider a node $\langle i,j\rangle$. Clearly $\phi_i \nvdash_{\text{Prop}} \psi_j$, so there is an assignment f of truth values 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 $(\Box\bot \to \phi_i^*)$ and $\Box(\Box\bot \to \phi_i^*)$ are forced everywhere in the model.

On the other hand, no node $\langle i,j\rangle$ on C_j forces ψ_j, hence $\langle i,j\rangle \Vdash \Box\bot \to \psi_j^*$. It follows that $(\Box(\Box\bot \to \psi_j^*) \to \Box\bot)$ 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 that $\chi(B_1, \ldots, B_n) = \langle\chi\rangle$. We have by 6.9 that

$$\begin{array}{ll}
\vdash\Box(\Box\bot \to [\phi_i^*]) \Rightarrow & (7.10)\\
\vdash[\phi_i] \lhd \top \Rightarrow & (7.9),(I1),(J5)\\
\vdash\langle\phi_i\rangle \lhd \top \Rightarrow & \\
\vdash\phi_i(B_1, \ldots, B_n) \lhd \top \Rightarrow & \text{(Reflection Principle)}\\
\phi_i(B_1, \ldots, B_n) \lhd \top. &
\end{array}$$

Moreover, by 6.9

$$\begin{array}{ll}
\vdash\Box(\Box\bot \to [\psi_j^*]) \to \Box\bot \Rightarrow & (7.10)\\
\vdash[\psi_j] \lhd \top \to \Box\bot \Rightarrow & (6.9),(I1),(J5)\\
\vdash\langle\psi_j\rangle \lhd \top \to \Box\bot \Rightarrow & \\
\vdash\psi_j(B_1, \ldots, B_n) \lhd \top \to \Box\bot \Rightarrow & \\
\neg(\psi_j(B_1, \ldots, B_n) \lhd \top). & \text{(Reflection Principle)}
\end{array}$$

Note that the uses of the Reflection Principle are eliminable here: we could just have proved the necessary lemmas externally, i.e., in nonformalized form. (In case the theory under consideration is not PA it may even be necessary to reason externally.)

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

7.12 Remark Note that in the proof of 7.11 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

[1] Auerbach, D. D., "Intentionality and the Gödel theorems," *Philosophical Studies*, vol. 48 (1985), pp. 337–351.

[2] Bernardi, C. and F. Montagna, "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, Italia.

[3] Bowie, G. L., "Lucas' number is finally up," *Journal of Philosophical Logic*, vol. 11 (1982), pp. 279–285.

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

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

[6] Hájek, P., "Experimental logics and Π_3-theories," *The Journal of Symbolic Logic*, vol. 42 (1977), pp. 515–522.

[7] Hilbert, D. and P. Bernays, *Grundlagen der Mathematik*, 2nd edition, Springer-Verlag, Berlin, 1970.

[8] Jeroslow, R. G., "Experimental logics and Δ_2-theories," *Journal of Philosophical Logic*, vol. 4 (1975), pp. 253–267.

[9] Lindström, P., "Some results on interpretability," pp. 329–361 in *Proceedings of the 5th Scandinavian Logic Symposium*, Aalborg, 1979.

[10] Lucas, J. R., "Minds, machines and Gödel," *Philosophy*, vol. 36 (1961), pp. 120–124.

[11] Montagna, F., "On the algebraization of Feferman's predicate," *Studia Logica*, vol. 37 (1978), pp. 221–263.

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

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

[14] Putnam, H., "Trial and error predicates and the solution to a problem of Mostowski," *The Journal of Symbolic Logic*, vol. 30 (1965), pp. 49–57.

[15] Smoryński, C., "Modal logic and self-reference," pp. 441–496 in *Handbook of Philosophical Logic*, edited by D. Gabbay and F. Guenthner, Reidel, Boston, 1984.

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

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

[18] Švejdar, "Degrees of interpretability," *Commentationes Mathematicae Universitatis Carolinae 19*, pp. 789–813.

[19] Švejdar, "Modal analysis of generalized Rosser sentences," *The Journal of Symbolic Logic*, vol. 48 (1983), pp. 986–999.

[20] Visser, A., "The provability logics of recursively enumerable theories extending Peano arithmetic at arbitrary theories extending Peano arithmetic," *Journal of Philosophical Logic*, vol. 13 (1984), pp. 79–113.

[21] Webb, J., *Mechanism, Mentalism and Metamathematics*, Reidel, Dordrecht, 1980.

Central Interfaculty Department of Philosophy
State University of Utrecht
3508 TC Utrecht
The Netherlands