Commentationes Mathematicae Universitatis Carolinae

Petr Hájek

On interpretability in theories containing arithmetic. II.

Commentationes Mathematicae Universitatis Carolinae, Vol. 22 (1981), No. 4, 667--688

Persistent URL: `http://dml.cz/dmlcz/106110`

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

22,4 (1981)

ON INTERPRETABILITY IN THEORIES CONTAINING ARITHMETIC II

Petr HÁJEK

Abstract: Investigated are Peano arithmetic PA and its conservative extension ACA_0 using classes. (Instead, one could speak on set theories ZF and GB.) I_{PA} (and I_{ACA_0}) denotes the class of all PA-sentences φ such that (PA + φ) is relatively interpretable in PA ((ACA_0 + φ) is relatively interpretable in ACA_0). Independent Σ_1^0 sentences φ are classified according to whether $\varphi \in I_{PA}$, $\varphi \in I_{ACA_0}$, $(\neg\varphi) \in$ $\in I_{ACA_0}$. (Note that $\neg\varphi$ can never be in I_{PA}.) This gives eight types of independent Σ_1^0 sentences; it is shown that each type is non-empty. This subsumes and completes most known results on the relation of I_{PA} and I_{ACA_0}. Main results are obtained by combining and generalizing methods of Solovay and Smoryński; a generalized fixed point calculation for a modal propositional calculus, which seems to be of independent interest, is presented and heavily used.

Key words: Relative interpretability, modal logic, arithmetic

Classification: 03F25, 03B45, 03F30

---

§ 1. Introduction

1.1. Let PA be Peano arithmetic and let ACA_0 denote the second-order theory with two sorts of variables (number variables x,y,... and class variables X,Y,...) having axioms PA minus the induction schema for number variables, a new predicate $\in$ such that $t \in Z$ is well formed iff t is a number term and X is a class term and two groups of second order axioms:

Arithmetical comprehension: for each formula φ in which no class variable is quantified and which does not contain the variable X, the following is an axiom:

$$(\exists X)(\forall x)(x \in X \equiv \varphi)$$

Induction axiom:

$$(0 \in X \,\&\, (\forall x)(x \in X \rightarrow S(x) \in X)) \rightarrow (\forall x)(x \in X).$$

It is well known that ACA_0 is a conservative extension of PA (each model of PA is expandable to a model of ACA_0) and that ACA_0 is finitely axiomatizable (imitate the proof of Metatheorem 1 in [2]). Thus we can claim that

$$PA:ACA_0 = ZF:GB$$

where ZF and GB is the Zermelo-Fraenkel and Gödel-Bernays set theory. And indeed, our results remain valid if we replace the pair (PA, ACA_0) by (ZF, GB) or another similarly related pair of theories containing PA. But since our investigation concerns PA-sentences we shall speak on PA and ACA_0.

1.2. For each theory T containing PA, let I_T denote the set of all PA-sentences φ such that $(T + \varphi)$ is relatively interpretable in T in the sense of Tarski, Mostowski and Robinson [17]. Let us survey the known facts on I_{PA} and I_{ACA_0}.

(1) $I_{PA} \neq I_{ACA_0}$; I_{PA} is Π^0_2-complete (Solovay [14]) but I_{ACA_0} is recursively enumerable.

(2) $I_{PA} - I_{ACA_0} \neq 0$. In [5], a Π^0_2 sentence φ is constructed such that $\varphi \in I_{PA} - I_{ACA_0}$ provided PA is ω-consistent; in [7] the assumption of ω-consistency is replaced by that of (mere) consistency. Solovay exhibited a Σ^0_1 sentence

$\varphi \in I_{PA} - I_{ACA_0}$ (cf. [10]). Lindström independently showed that for an appropriate binumeration α of PA, the Σ^0_1 sentence $\neg \mathrm{Con}_\alpha$ is in $I_{PA} - I_{ACA_0}$. Lindström also constructed a Π^0_2 sentence φ such that both φ and $\neg\varphi$ belong to $I_{PA} - I_{ACA_0}$ (see [8]).

(3) $I_{ACA_0} - I_{PA} \neq 0$. In [6] it is shown that if this difference is non-empty then it must contain a Π^0_1 sentence; Solovay constructed such a sentence [14]. His proof will be sketched and analyzed below.

(4) The following are equivalent: (i) $\varphi \in I_{PA}$; (ii) φ is Π^0_1 conservative (Π^0_1-con), i.e. for each Π^0_1 sentence π $(PA + \varphi) \vdash \pi$ implies $PA \vdash \pi$; (iii) for each n, $PA \vdash \mathrm{Con}_{(PA \upharpoonright n)+\overline{\varphi}}$ (where $PA \upharpoonright n$ denotes the set of all axioms of PA that (i.e. whose Gödel numbers) are less than n). See [3],[6]. Consequently, if π is a Π^0_1 sentence and $\varphi \in$ $\in I_{PA}$ then $PA \vdash \varphi$.

1.3. The above lead to the question what possibilities we have for independent Σ^0_1 sentences φ according to the questions whether $\varphi \in I_{PA}$, $\varphi \in I_{ACA_0}$, $(\neg\varphi) \in I_{ACA_0}$.(If φ is an independent Σ^0_1 sentence then necessarily $(\neg\varphi) \notin I_{PA}$, see the end of 1.2. Logically, we have eight types:

	$\varphi \in I_{PA}$	$\varphi \in I_{ACA_0}$	$(\neg\varphi) \in I_{ACA_0}$
1	no	yes	yes
2	no	yes	no
3	no	no	yes
4	no	no	no
5	yes	yes	yes
6	yes	yes	no
7	yes	no	yes
8	yes	no	no

We shall show that there are formulas of all these eight types.

1.4. Now let us make some preliminary observations. First it is easy to see that the formula $\neg \mathrm{Con}_{\alpha}$ (where α is the natural PR-binumeration of PA) is of type 6, since we have $PA \vdash (\mathrm{Con}_{\alpha} \equiv \mathrm{Con}_{ACA_0})$ (here Con_{ACA_0} is expressed using the finitely many axioms sufficient to axiomatize ACA_0); it is easy to show $(\neg \mathrm{Con}_{\alpha}) \in I_{PA}$, $(\neg \mathrm{Con}_{ACA_0}) \in I_{ACA_0}$, $\mathrm{Con}_{ACA_0} \notin I_{ACA_0}$ (cf. [1],[16]). But we shall show another sentence of type 6 below.

Second, observe that a formula φ of type 7 has the nice property that $\varphi \in I_{PA} - I_{ACA_0}$ and $(\neg\varphi) \in I_{ACA_0} - I_{PA}$; thus φ is a Σ^0_1 sentence showing that $I_{PA} - I_{ACA_0}$ is non-empty and $\neg\varphi$ is a Π^0_1 sentence showing that $I_{ACA_0} - I_{PA}$ is non-empty.

Third, we should make clear what means will be used in our proofs. Main tool for showing that something is in ACA_0 will be the Solovay's method described below. Main tool for showing that something is unprovable or is not in I_{ACA_0} will be a generalized Smoryński's fixed point calculation for fixed points defined by means of arithmetically interpreted modal logic. To show that something is or is not in I_{PA}, we shall show that the formula in question is or is not Π^0_1-con. And in one case, where these methods fail, we shall imitate a construction due to Lindström.

1.5. Most of our (non)interpretability results will follow rather quickly and easily from Solovay's construction and from our generalization of Smoryński's fixed point calculation. The contribution to arithmetical interpretations of modal logics presented in § 3 is hoped to be of independent interest. Note that § 3 does not depend on § 2.

§ 2. Solovay's construction analyzed

2.1. Solovay constructed a Π^0_1 sentence $\varphi \in I_{ACA_0} - I_{PA}$ (in fact, in $I_{GB} - I_{ZF}$) in 1976; a full proof is contained in a letter by Solovay to the present author. Since [14] has still not been finished, we shall give here a more or less detailed sketch of Solovay's proof in a form that enables us to obtain some general consequences concerning I_{ACA}. This is done with kind permission of Professor Solovay.

2.2. First, Solovay uses a rather specific provability predicate related to Herbrand's analysis. Let $(PA)_c$ be the conservative extension of PA having the following property: For each sentence $(\exists x)\psi(x)$ of $(PA)_c$ there is a witnessing constant $c_{(\exists x)\psi(x)}$ of $(PA)_c$ such that the following witnessing axiom is an axiom of $(PA)_c$:

$(\exists x)\psi(x) \longrightarrow c_{(\exists x)\psi(x)}$ is the minimal x such that $\psi(x)$.

Let $\Delta(PA)$ be the set of closed instances of axioms of $(PA)_c$, of equality and identity axioms and of the logical axioms $(\forall x)\psi(x) \longrightarrow \psi(t)$. Then we have the following lemma ([9] p. 49):

Let φ be a closed formula of (PA) . Then $(PA)_c \vdash \varphi$ iff φ is a tautological consequence of $\Delta(PA)$.

Following Solovay, call a <u>satisfactory sequence</u> on n each function s associating with each $(PA)_c$ sentence less than n <u>zero</u> or <u>one</u> such that s commutes with logical connectives and gives the value <u>one</u> to each element of $\Delta(PA)$. Then evidently we have the following:

Let φ be a closed formula of $(PA)_c$. Then $(PA)_c \vdash \varphi$ iff there is an n such that for each satisfactory sequence s on n we have $s(\varphi) = 1$.

Say that φ is <u>proved on level</u> n if each satisfactory s on n gives value 1 to φ . From now on, saying "φ is provable" for a $(PA)_c$-formula φ we shall always mean "there is an n such that φ is provable on level n".

2.3. Let us work in ACA_0 extended conservatively by adding witnessing constants from $(PA)_c$ and the corresponding witnessing axioms. Let us make the following definition: A class Z is a <u>satisfaction relation</u> on j (in symbols: $Tr(Z,j)$) if (roughly)Z is a function associating (1) with each pair (t,u) where t is a term of $(PA)_c$ whose Gödel number is less that j and u is a sufficiently long sequence of numbers a number and (2) with each pair (a,u) where a is a $(PA)_c$ formula whose Gödel number is less than n and u is a satisfactorily long sequence of numbers a truth value 0 or 1 <u>such that</u>

(a) $Z(x_j,u) = (u)_j$, $Z(t_1 + t_2,u) = Z(t_1,u) + Z(t_2,u)$ etc.,

(b) $Z(t_1 = t_2,u) = 1$ iff $Z(t_1,u) = Z(t_2,u)$ and

(c) the obvious Tarski's conditions for truth of composed formulas are valid.

Boring details of elaboration of this (evident) definition are left to the reader.

2.4. The following lemma is obvious:

<u>Lemma</u>. (1) $(\exists Z)Tr(Z,0)$

(2) $(\exists Z)Tr(Z,j) \rightarrow (\exists Z)Tr(Z,j + 1)$

(3) $Tr(Z_1,j_1)\ \&\ Tr(Z_2,j_2)\ \&\ j_1 \leq j_2 \rightarrow Z_1 \subseteq Z_2$.

<u>Caution</u>: But the statement $(\forall j)(\exists Z)Tr(Z,j)$ is unprovable in ACA_0 (pedantically: in $ACA_0)_c$) since it implies evidently Con_α where α is the natural binumeration of PA. This shows that the induction scheme

$(\psi(0)\ \&\ (\forall x)(\psi(x) \rightarrow \psi(S(x))) \rightarrow (\forall x)\psi(x)$

is unprovable in ACA_0 (which is well known).

2.5. In ACA_0, assume $Tr(Z,j)$. Then Z defines a <u>true</u> satisfactory sequence s on j - restriction of s to pairs $(a,\emptyset)$ where a is a $(PA)_c$-sentence, $a \leq j$. Thus: if φ is proved of level n and $Tr(Z,n)$ then φ is true, i.e. $Z(\varphi,\emptyset) = 1$.

2.6. "It's snowing" - it's snowing-metatheorem: Let φ be a $(PA)_c$-formula whose Gödel number is less than j. Then $ACA_0 \vdash Tr(Z,j) \rightarrow \varphi(x_0,\dots,x_n) \equiv Z(\overline{\varphi(x_0,\dots,x_n)},x_0,\dots,x_n) = 1.$

(Proof by induction on the length of φ.)

2.7. Let $Tr_n(x)$ be the Σ_n^0-predicate of PA which is a truth predicate for Σ_n^0-sentences constructed in the usual way; in particular, we have $PA \vdash \varphi \equiv Tr_n(\overline{\varphi})$ for each Σ_n^0 sentence φ. We have the following lemma:

Lemma (in ACA_0). Let $a \in \Sigma_n^0$ and let $Tr(Z,x)$ where x is the Gödel number of a. Then $Z(a,\emptyset) = 1$ iff $Tr_n(a)$.

(By induction on a.)

2.8 (cf. [18]). Say that n is occupable (Ocp(n)) if $(\exists Z)Tr(Z,n)$. By the above, Ocp(n) is not equivalent to any formula not containing bound class variables. The heart of Solovay's construction is the following theorem:

2.9. Theorem. Let φ be a PA-formula and let ACA_0' be an extension of ACA_0 such that ACA_0' proves "there is a satisfactory sequence s of non-occupable length such that $s(\varphi) = 1$". Then $(ACA_0 + \varphi)$ is interpretable in ACA_0'.

Sketch of the proof: First we define an interpretation of $((PA)_c + \varphi)$ in ACA_0' and then extend it to an interpretation of ACA_0. The first idea is: consider values s(a) for occupable a (pedantically: for a of occupable Gödel no.) - this gives something as a complete Henkin extension and one could try to use it for a definition of an interpretation of PA putting

$Number^*(x) \equiv x$ is a Henkin constant, $Ocp(x)$ and

$(\forall y < x)(y$ a Henkin constant $\rightarrow s(\ulcorner x = y \urcorner) = 0$;

$Number^*(x) \,\&\, Number^*(y) \,\&\, Number^*(z)$

$x +^* y = z$ iff $s(\ulcorner x + y = z \urcorner) = 1$

and analogously for successor and multiplication.
We would be obliged to prove an "it's snowing" - it's snowing theorem saying

$$(*) \quad \begin{aligned}&\mathrm{Number}^*(x_o) \,\&\, \ldots \to \\ &\to [\varphi^*(x_o, \ldots) \equiv s(\overline{\varphi}(x_o, \ldots)) = 1].\end{aligned}$$

But this requires closedness of Ocp to some operations; and we only know that Ocp is closed under successor. The alternative is not to use all Henkin constants of occupable Gödel no but to restrict oneself to x satisfying another non-arithmetical predicate $I(x)$ such that

(1) $I(x) \to \mathrm{Ocp}(x)$

(2) $I(0) \,\&\, (\forall x)(I(x) \to I(x+1))$

(3) I is satisfactorily closed

is provable in ACA_o'.

Solovay's analysis shows that (a) under an appropriate coding of formulas, it suffices to have in (3) $I(x) \to$ $\to I(x^{\log x})$ and (b) using $\mathrm{Ocp}(x)$, we can indeed define a predicate $I(x)$ such that (1) - (3) is provable. This concludes the construction of an interpretation of $(PA + \varphi)$ in ACA_o'.

Now this interpretation is extended to an interpretation of ACA_o as follows: Define
$\mathrm{Class}^*(x) \equiv x$ is a $(PA)_c$-formula with just one free variable v, $I(x)$ and $(\forall y < x)$(y is a $(PA)_c$ formula with just v free $\to s((\forall v)(x \equiv y)) = 0)$.
Then (in ACA_o') no x is both a number* and a class*; put
$\mathrm{Number}^*(x) \,\&\, \mathrm{Class}^*(y) \to (x \in^* y \equiv S(y(x)) = 1)$

where y(x) means formal substitution of the constant x into the formula y for the variable v).

Then the validity of the induction axiom for classes in the interpretation is clear (since s is satisfactory and, thanks to the sufficient closedness of I, if y is a formula as above and I(y) then for the sentence z expressing the least element principle for y we have also I(z)). To prove arithmetical comprehension in the interpretation it is useful to deal with "Gödel operations" as in [2] and to show closedness of classes under Gödel operations in the sense of the interpretation. Here again we profit from the satisfactory closedness of I: if a class Y is defined by a formula y such that I(y) then the formula defining the result of a Gödel operation applied to Y must also satisfy I. This concludes our proof-sketch.

The construction of a promised Π^0_1-sentence in $I_{ACA_0} - I_{PA}$ will be almost immediate from the preceding theorem and from the modal considerations of the next section.

§ 3. Some modal calculations

3.1. Arithmetical interpretations of some modal propositional calculi turned out to be a powerful tool for unifying some self-referential investigations and also for some negative results. See [15], [13], [11], [4]. We shall describe a modal system as close to that of Smoryński [11] as possible. We differ from Smoryński in two aspects: first, we want to prove a theorem applying to Rosser-like sentences as well as to Guaspari-like sentences, thus we have to generalize. On the other hand, in this paper we shall not need Sheperdson's generalization of Rosser sentences: in this aspect we are less general.

3.2. Language: Propositional variables p, q,...; propositional constants $\bot$, $\top$. Connectives $\&, \vee, \neg$, etc.; modalities $\Box$, Δ, ∇. Rosser witness comparisons $\preccurlyeq$, $\prec$; Guaspari witness comparison $\trianglelefteq$.

3.3. Formulas and S-formulas. Propositional variables and constants are formulas; formulas are closed under logical connectives and modalities. A formula is an S-formula if it begins with a modality (is of the form $\Box A$, ΔA, ∇A). If A, B are S-formulas then $A \preccurlyeq B$, $A \prec B$, $A \trianglelefteq B$ are formulas.

3.4. Arithmetical interpretation. For each propositional variable p, p^* is a sentence of PA. Modalities are interpreted by some Σ_1^0-formulas with one free variable and with just one unbounded existential quantifier. If a modality $\boxtimes$ is interpreted by $\alpha(x)$ and if $A^* = \psi$ then $(\boxtimes A)^* = \alpha(\overline{\psi})$. Necessity $\Box$ is in this paper always interpreted by the formula $(\exists y)$(x is proved (in $(PA)_c$) on level y), denoted by Pr(x). Δ and ∇ will be interpreted (1) either by the preceding provability formula or (2) by Intp(x) i.e. by $(\exists y)$(y is a witnessed interpretation of (ACA_0 + x) in ACA_0) (where a witnessed interpretation is a tuple consisting of formulas defining numbers, classes, basic arithmetical operations and membership in the sense of the interpretation and from an ACA_0-proof of the conjunction of interpretations of finitely many axioms axiomatizing ACA_0 plus of x); (3) or by (Pr(x) $\vee$ Intp(x)) (rewritten as a Σ^0_1-formula with one existential quantifier).

Note that Pr uses a fixed binumeration α of $(PA)_c$ (take the natural one); sometimes we shall write Pr_α instead of Pr. Similarly, Intp uses the natural binumeration β (by listing) of ACA_0; we write $Intp_\beta$ instead of Intp if necessary.

The arithmetical interpretation $*$ commutes with logical connectives.

If A, B are S-formulas, $A^* = (\exists y)\psi(y)$ $(= \Phi)$ and $B^* = (\exists x)\chi(z)$ $(= \Psi)$ then $A \preccurlyeq B$ and $A \prec B$ are interpreted as follows:

$(A \preccurlyeq B)^* = (\exists y)(\psi(y) \,\&\, (\forall z<y)\neg\chi(z))$

$(A \prec B)^* = (\exists y)(\psi(y) \,\&\, (\forall z\leq y)\neg\chi(z))$

In words, the former formula says that there is a witness y for Φ such that no $z<y$ is a witness for Ψ; similarly the latter.

The definition of $(A \trianglelefteq B)^*$ (for S-formulas A, B) is a bit more complicated.
Note that A^* can have one of the following three forms:

$Pr_\alpha(\bar{\varphi})$, $Intp_\beta(\bar{\varphi})$, $(Pr_\alpha \vee Intp_\beta)(\bar{\varphi})$

for some φ. Let $(\alpha + u)(x)$ be the formula $\alpha(x) \vee x = u$; similarly $(\beta + u)(x)$. Let Tr be the Σ^0_1-truth predicate for Σ^0_1-sentences. Then $(A \trianglelefteq B)^*$ says:

There is a witness y for $Pr_{\alpha+u}(\bar{\varphi})$ ($Intp_{\beta+u}(\bar{\varphi})$, $Pr_{\alpha+u}(\bar{\varphi}) \vee Intp_{\beta+u}(\bar{\varphi})$ respectively), where u is a true Σ^0_1-sentence, such that for no $z<y$, z is a witness for Ψ. (Recall that $\Psi = B^*$.)

For example, if $A^* = Intp_\beta(\overline{\varphi})$ then $(A \trianglelefteq B)^*$ says "There is a true Σ_1^0-sentence u such that there is a witnessed interpretation y of $(ACA_0 + \overline{\varphi})$ in $(ACA_0 + u)$ such that for no $z<y$, z is a witness for Ψ."

For provability, we may say that $\overline{\varphi}$ is proved on level y in $(PA)_c + u$ iff each satisfactory sequence s on y such that $s(u) = 1$ gives $s(\varphi) = 1$ (i.e., $u \dot{\rightarrow} \overline{\varphi}$ is proved on level y in $(PA)_c$).

In particular, if p^* is φ then $(\Box\neg p \preccurlyeq \Box p)^*$ is $(\exists y)(\neg\overline{\varphi}$ proved on level $y \,\&\, (\forall z<y)(\overline{\varphi}$ not proved on level z) and $(\Box\neg p \trianglelefteq \neg p)$ is $(\exists y)$ (for some true Σ_1^0-sentence u, $u \dot{\rightarrow} \neg\overline{\varphi}$ proved on level $y \,\&\, (\forall z<y)(\overline{\varphi}$ not proved on level z).

This completes the definition of an arithmetical interpretation $*$ of modal formulas.

Remark. The reader acquainted with [3] and/or [12] will now see why $\trianglelefteq$ is called Guaspari witness comparison: simply because witness comparison is combined with truth definition for Σ_1^0-formulas. (But apparentùy not all of our fixed points using $\trianglelefteq$ are particular cases of Smoryński's "Guaspari sentences of the first kind".)

3.5. Axioms for modal formulas. $\boxtimes$ varies over $\Box$, Δ, ∇; $\subseteq$ varies over $\preccurlyeq$, $\prec$, $\trianglelefteq$.

(A1) Propositional tautologies

(A2) Necessitations of tautologies

(A3) $A \rightarrow \Box A$, $A \subseteq B \rightarrow \Box(A \subseteq B)$ for all S-formulas A, B

(A4) $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$ for all A, B

(A5) $A \subseteq B \rightarrow A$; $A \preccurlyeq B \rightarrow A \trianglelefteq B$; $A \prec B \rightarrow A \preccurlyeq B$

$\boxed{S}(A \vee B \rightarrow (A \preceq B \equiv \neg (B \prec A)))$ for all S-formulas A, B

(A6) $(\Box (A \subseteq B) \& B) \rightarrow A$ for all S-formulas A, B

(A7) $\boxed{S} (\Box A \rightarrow \boxtimes A)$ for any A

(A8) $\Box (A \rightarrow B) \rightarrow (\boxtimes A \rightarrow \boxtimes B)$ for any A, B

(A9) $(\Box A \rightarrow \neg \Box \neg A)$ for any A

(A10) $\Box (A \rightarrow \neg \Box \neg A) \rightarrow \Box \neg A$ for any A

The only <u>deduction rule</u> is modus ponens. This concludes the definition of our (tentative) modal calculus.

<u>Remark</u>. (A1) - (A5) - (A7) - (A8) are like in Smoryński [11]. The axiom (A6) is important for $\subseteq$ being $\trianglelefteq$; for $\subseteq$ being $\preceq$ or $\prec$ it is easily derived from the remaining ones. (A9) is Smoryński's superconsistency; (A10) is Gödel's second incompleteness theorem.

3.6. We shall show that each arithmetical interpretation of each axiom is provable in $(PA + Con_{PA})$ (pedantically, in $(PA + Con_{\alpha})$; note that by our choice of α and β we have $PA \vdash Con_{\alpha} \equiv Con_{\beta}$).

Everything is clear except (1) $A \trianglelefteq B \rightarrow A$ and (2) $(\Box(A \trianglelefteq B) \& B) \rightarrow A$.

(1) First let A^* be $Pr_{\alpha}(\overline{\varphi})$. Reason in PA. Evidently, $(A \trianglelefteq B)^*$ implies $(\exists u)(u$ true Σ_1^0, $\overline{\varphi}$ is $((PA)_c + u)$-provable). But since each true Σ_1^0-sentence is $(PA)_c$-provable, we have $Pr_{\alpha}(\overline{\varphi})$. If A_0^* is $Intp_{\beta}(\overline{\varphi})$ then we reason in PA as follows: $(\exists u)(u$ true Σ_1^0, $(ACA_0 + \overline{\varphi})$ interpretable in $(ACA_0 + u))$. But since each true Σ_1^0-sentence is ACA_0-provable, there is an interpretation of $(ACA_0 + \overline{\varphi})$ in ACA_0. For $Pr_{\alpha} \vee Intp_{\beta}$ argue similarly.

(2) Let B^* be $(\exists z)\chi(z)$; first, let A^* be $Pr_\alpha(\overline{\varphi})$. Let b be a witness for $(\exists z)\chi(z)$. Then $\chi(b)$ and $Pr_\alpha((\exists y \leq b)(\exists u \text{ true } \Sigma_1^0)(\overline{\varphi}$ is $(PA_c + u)$-proved on level $y))$. Let $\sigma_1, \ldots, \sigma_n$ be all Σ_1^0-sentences such that $u \rightarrow \overline{\varphi}$ is $(PA)_c$-proved on a level $\leq b$; we have $Pr_\alpha(\bigvee_{i=1}^{n} \text{True}(\sigma_i))$, thus $Pr_\alpha(\bigvee \sigma_i)$, $Pr_\alpha(\bigvee \sigma_i \rightarrow \overline{\varphi})$ and hence $Pr_\alpha(\overline{\varphi})$.

For $A^* = Intp_\beta(\overline{\varphi})$ the proof is similar. (Note that if $i_1, \ldots, i_n$ are interpretations of $(ACA_0 + \overline{\varphi})$ in $(ACA_0 + \sigma_i)$ then they can be combined into a single interpretation i of $(ACA_0 + \varphi)$ in $(ACA_0 + \bigvee_{i=1}^{n} \sigma_i)$.)

Lemma 3.7 ([11]). $\Box(A \equiv B) \rightarrow (\boxtimes A \equiv \boxtimes B)$

3.8. <u>Main theorem.</u> Let $\subseteq$ be $\preccurlyeq$ or $\trianglelefteq$ and assume

$$\boxed{S}(p \equiv (\Delta \neg p \subseteq \nabla p)).$$

(1) From this assumption, the following is provable in our logic:

$$\neg p, \quad \neg \Box p, \quad \neg \Box \neg p, \quad \Delta \neg p \rightarrow \nabla p$$

(2) If, moreover, $\subseteq$ is $\preccurlyeq$ then the following is provable:

$$\neg \nabla p, \ \neg \Delta \neg p, \ \Box(p \rightarrow \neg(\nabla p \prec \Delta \neg p)), \ \neg \Box(\neg \nabla p \prec \Delta \neg p)$$

<u>Proof</u>. (1) Let A be $(\Delta \neg p \subseteq \nabla p)$.

(a) $p \vdash A \vdash \Box A$ (by (A3)) $\vdash \Box p$ (lemma) $\vdash \neg \Delta \neg p$ (A9)

$p \vdash A \vdash \Delta \neg p$ (by A5)

Thus $p \vdash$ contradiction, hence $\vdash \neg p$.

(b) $\Box p \vdash \Box A \,\&\, \nabla p$ (Lemma and A7) $\vdash \Delta \neg p$ (A6)

$\Box p \vdash \neg \Delta \neg p$ (A9)

(c) $\Box \neg p \vdash \Delta \neg p \vdash (\Delta \neg p \subseteq \nabla p) \vee (\nabla p \subseteq \Delta \neg p) \vdash$

$\vdash \nabla p \subseteq \Delta \neg p \vdash \nabla p$;

$\Box \neg p \vdash \neg \nabla \neg \neg p \vdash \neg \nabla p$

(d) $\Delta\neg p \vdash \nabla p$ as in (c).

(2)

(a) $\nabla p \vdash (\Delta\neg p \preceq \nabla p) \vee (\nabla p \prec \Delta\neg p) \vdash (\nabla p \prec \Delta\neg p) \vdash$
$\vdash \Box(\nabla p \prec \Delta\neg p) \vdash \Box(\neg(\Delta\neg p \preceq \nabla p)) \vdash \Box\neg p$, but
$\vdash \neg\Box\neg p$.

(b) $\vdash \Box(p \rightarrow (\Delta\neg p \preceq \nabla p))$, and $\vdash \Box((\Delta\neg p \preceq \nabla p) \rightarrow$
$\rightarrow \neg(\nabla p \prec \Delta\neg p))$, thus $\vdash \Box(p \rightarrow \neg(\nabla p \prec \Delta\neg p)$.

(c) $\Box(\neg(\nabla p \prec \Delta\neg p)) \vdash \Box(\nabla p \rightarrow (\Delta\neg p \preceq \nabla p))$

$\vdash \Box(\nabla p \rightarrow p)$

$\vdash \Box(\neg p \rightarrow \neg\nabla p)$

$\vdash \Box(\neg p \rightarrow \neg\Box p)$

$\vdash \Box(\neg\neg p)$

$\vdash \Box p$, a contradiction.

3.9. Corollary. Fix one of possible meanings of Δ, ∇ and $\subseteq$. Let φ be a fixed point such that the arithmetical interpretation of p by φ makes $\boxdot$ $(p \equiv (\Delta\neg p \subseteq \nabla p)$ PA-provable. Then

(1) φ is false, φ is unprovable, $\neg\varphi$ is unprovable; if $\neg\varphi$ is Δ then φ is ∇.

(2) If $\subseteq$ is $\preceq$ then φ is not ∇, $\neg\varphi$ is not Δ and φ is Π^0_1-nonconservative: the sentence interpreting $\neg(\nabla p \prec \Delta\neg p)$ shows it.

3.10. Remark. If $\subseteq$ is $\trianglelefteq$ and we succeed to show that φ is not ∇ (so that, consequently, $\neg\varphi$ is not Δ) then φ is Π^0_1-conservative: Let σ be a Σ^0_1-sentence such that $PA \vdash \varphi \rightarrow \neg\sigma$, i.e. $PA \vdash \sigma \rightarrow \neg\varphi$; then let d be a witness for $Pr_{\alpha+\overline{\sigma}}(\neg\overline{\varphi})$, $Intp_{\beta+\overline{\sigma}}(\neg\overline{\varphi})$, $(Pr_{\alpha+\overline{\sigma}} \vee Intp_{\beta+\overline{\sigma}})(\neg\overline{\varphi})$ respectively (choose according to the meaning of Δ). Argue in

(PA $+\neg\varphi$): If σ were true then beneath d there would be a witness for ∇p; since there is no such witness, σ must be false. We have proved $\neg\sigma$ in (PA $+\neg\varphi$).

§ 4. Interpretability in PA versus in ACA_0. Investigations of § 2 and § 3 yield almost immediately examples of seven types of independent formulas. Let us begin with Π^0_1-nonconservative φ, i.e. $\varphi \notin I_{PA}$:

4.1. $\varphi \equiv (\Box\neg\varphi \preccurlyeq \Box\varphi)$ (Solovay). Obviously, φ is independent. We show that ($ACA_0 + \neg\varphi$) is interpretable in ACA_0. It suffices to find an interpretation in ($ACA_0 + \varphi$). Argue in the last theory. There is a witness for $\Box\neg\varphi$; call least such witness n_0. Clearly, n_0 is not occupable (see 2.5 and 2.6). Consider $n_0 - 1$: it is not a witness for $\Box\varphi$, thus there exists a satisfactory sequence s on $n_0 - 1$ such that $s(\varphi) = 0$. Now 2.9 applies.

This is how Solovay constructed his example (except that he did not formulate explicitly 2.9). Observe, furthermore, that ($ACA_0 + \varphi$) is interpretable in ACA_0. Since ($ACA_0 + \neg Con_{ACA_0}$) is interpretable in ACA_0 (cf. e.g. [16]) it suffices to find an interpretation in ($ACA_0 + \neg Con_{ACA_0} + \neg\varphi$); but the last theory proves $\Box\varphi \prec \Box\neg\varphi$. Let n_0 be the least witness for $\Box\varphi$ and continue as above. Thus φ is of type (1)(from 1.3).

In the sequel, let Δ denote the modality of interpretability and let $\hat{\Box}$ denote disjunction of provability and interpretability.

4.2. $\varphi \equiv (\hat{\Box}\neg\varphi \precneq \Box\varphi)$.

Clearly, $\neg\varphi \notin I_{ACA_0}$; we show that $\varphi \in I_{ACA_0}$. Again it suffices to interpret $(ACA_0 + \varphi)$ in $(ACA_0 + \neg Con_{ACA_0} + \neg\varphi)$. The last theory proves $(\Box\varphi \prec \hat{\Box}\neg\varphi)$; let n_0 be the least witness for $\Box\varphi$. Then n_0 is not occupable and n_0 is neither a witness for $\Delta\neg\varphi$ nor a witness for $\Box\neg\varphi$. Thus there is a satisfactory s on n_0 such that $s(\varphi) = 1$. Apply 2.9. Thus φ is of type (2).

4.3. $\varphi \equiv (\Box\neg\varphi \precneq \hat{\Box}\varphi)$.

Clearly, $\varphi \notin I_{ACA_0}$. To prove $(\neg\varphi) \in I_{ACA_0}$ argue in $(ACA_0 + \varphi)$ as in (1). Thus φ is of type (3).

4.4. $\varphi \equiv (\Delta\neg\varphi \precneq \Delta\varphi)$ (Hájek [6]).

Clearly, φ, $(\neg\varphi) \notin I_{ACA_0}$. Thus φ is of type (4).

Now let us consider fixed points with $\trianglelefteq$; recall 3.10 telling that if we prove that φ is not ∇ then φ is Π^0_1-conservative, i.e. $\varphi \in I_{PA}$.

4.5. $\varphi \equiv (\Box\neg\varphi \trianglelefteq \Box\varphi)$.

Clearly, φ is independent. This already shows that φ is Π^0_1-conservative. We show that $(ACA_0 + \neg\varphi)$ is interpretable in $(ACA_0 + \varphi)$. Argue in the last theory. Let n_0 be the least number such that for some true Σ^0_1-sentence u, $u \rightarrow \neg\overline{\varphi}$ is proved on level n_0. If n_0 is occupable, $Tr(Z,n_0)$, then necessarily $Z(u,\emptyset) = 1$ (see 2.7) and $Z(\neg\varphi,\emptyset) = 0$ (see 2.6), thus for the true satisfactory sequence we have $s(u \rightarrow \neg\overline{\varphi}) = 0$, a contradiction. This shows that n_0 is not occupable and n_0 is not a witness for $\Box\varphi$; thus there is an s on $n_0 - 1$ such that $s(\neg\varphi) = 1$. Apply 2.9.

To prove that $(ACA_0 + \varphi)$ is interpretable in ACA_0, consider $(ACA_0 + \neg Con_{ACA_0} + \neg\varphi)$; the last theory proves $\Box\varphi \trianglelefteq \Box\neg\varphi$ (since $\neg$ Con implies $(\Box\neg\varphi \preccurlyeq \Box\varphi) \vee$ $\vee(\Box\varphi \prec \Box\neg\varphi)$ which implies $(\Box\neg\varphi \trianglelefteq \Box\varphi) \vee$ $\vee(\Box\varphi \trianglelefteq \Box\neg\varphi)$). Thus proceed analogously. We see that φ is of type (5).

4.6. $\varphi \equiv \hat{\Box}\neg\varphi \trianglelefteq \Box\varphi$.

Again, φ being not $\Box$, φ is Π^0_1-con. Consequently, $\neg\varphi$ is not $\hat{\Box}$ and hence not Δ, i.e. $(\neg\varphi) \notin I_{ACA_0}$. To prove $\varphi \in I_{ACA_0}$ consider $(ACA_0 + \neg Con_{ACA_0} + \neg\varphi)$ as above. Thus φ is of type (6).

4.7. $\varphi \equiv (\Box\neg\varphi \trianglelefteq \hat{\Box}\varphi)$.

We prove $\varphi \notin I_{ACA_0}$. Assume the contrary and let i be the least witness for $\Delta\varphi$. Work in $(ACA_0 + \varphi)$. Arguing as in the second half of 3.6 we show that $\neg\varphi$ is provable (in PA), which is a contradiction. Thus indeed $\varphi \notin I_{ACA_0}$. Consequently, φ is not $\hat{\Box}$ and therefore φ is Π^0_1-con. To show that $(\neg\varphi) \in I_{ACA_0}$, argue in $(ACA_0 + \varphi)$. Let n_0 be the least number such that $u \longrightarrow \neg\varphi$ is provable on level n_0, where u is true Σ^0_1-sentence. As in 4.5, show that n_0 is not occupable. Continue as usual; φ is of type (7).

4.8. Unfortunately, the author was unable to show that the fixpoint $\varphi \equiv (\Delta\neg\varphi \trianglelefteq \Delta\varphi)$ (or similar fixpoints with some Δ replaced by $\hat{\Box}$) is of type (8). This is definitely a fault of beauty; but this gives us an opportunity to present an entirely different method due to Lindström [8]. Our proof is a combination of his proofs of Theorem 2 and Theorem 5. (I was suggested by Švejdar to try to use Lindström's Theorem 2.)

We are going to construct a formula of type (8) as $\neg \mathrm{Con}_{\alpha'}$, where α' is an appropriate PR-binumeration of PA. Let α be the natural binumeration of PA and for each PA-sentence φ, let $\alpha[\varphi](x) \equiv (\alpha(x)\,\&$ beneath x, there is no Q-proof of $\overline{\varphi}$). (Q is the usual finite subsystem of PA.) Put

$f(\varphi) = \neg \mathrm{Con}_{\alpha[\varphi]}$, $Y_1 = \{\varphi ; f(\varphi) \in I_{ACA_0}\}$, $Y_2 = \{\varphi ;$
$\neg f(\neg\varphi) \in I_{ACA_0}\}$.

Claim. If $Q \vdash \neg\varphi$ then $\varphi \notin Y_1 \cup Y_2$, thus $Y_1 \cup Y_2$ is mono-consistent with Q in Lindström's terminology.

Proof of the claim. If $Q \vdash \neg\varphi$ then $PA \vdash \mathrm{Con}_{\alpha[\neg\varphi]}$, thus $ACA_0 \vdash \neg f(\varphi)$ and $f(\varphi) \notin I_{ACA_0}$. Furthermore, $PA \vdash \ulcorner Q \vdash \neg\varphi \urcorner$ thus $PA \vdash \ulcorner Q \nvdash \varphi \urcorner$ (since $PA \vdash \mathrm{Con}_Q$) and hence $ACA_0 \vdash$ $\vdash \alpha[\varphi] \equiv \alpha$, thus $ACA_0 \vdash \mathrm{Con}_{\alpha[\varphi]} \equiv \mathrm{Con}_\alpha$, which implies $\mathrm{Con}_{\alpha[\varphi]} \notin I_{ACA_0}$. The claim is proved.

By [8] Lemma 1, there is a φ such that neither φ nor $\neg\varphi$ is in $Th(Q) \cup Y_1 \cup Y_2$ (where Th(Q) is the set of all formulas provable in Q). We show that $f(\neg\varphi)$ is our formula of type (8). First, we have $f(\neg\varphi) = \neg \mathrm{Con}_{\alpha[\varphi]}$. Since $Q \nvdash \varphi$, $\alpha[\varphi]$ binumerates PA and hence $(\neg \mathrm{Con}_{\alpha[\varphi]}) \in I_{PA}$ (see [1]). Second, $(\neg\varphi) \notin Y_1$, thus $f(\neg\varphi) \notin I_{ACA_0}$; third, $\varphi \notin Y_2$, thus $\neg f(\neg\varphi) \notin I_{ACA_0}$. This concludes the proof.

4.1 - 4.8 prove the following

4.9. Main theorem II. Each type (from 1.3) is non-empty.

4.10. __Remark.__ After having read a preprint of this paper, Lindström gave simple alternative proofs of existence of sentences of types (2),(3),(4),(6),(7), assuming existence of sentences of type (1) and (5); his proofs use results of [8]. I present my original proofs since I believe that modal considerations of § 5, which make explicit the modal nature of proofs of existence of sentences of type (1) and (5), are of independent interest as a contribution to arithmetic interpretations of modal logic, and having our main theorem 3.8, proofs of existence of sentences of types (1) - (7) are reasonably simple.

References

[1] S. FEFERMAN: Arithmetization of metamathematics in a general setting, Fund. Math. 49(1960), 33-92.

[2] K. GÖDEL: The consistency of the axiom of choice etc., Princeton Univ. Press 1940.

[3] D. GUASPARI: Partially conservative extensions of arithmetic, Trans. Amer. Math. Soc. 254(1979), 47-68.

[4] D. GUASPARI, R. SOLOVAY: Rosser sentences, Annals of Math. Log. 16(1979), 81-99.

[5] P. HÁJEK: On interpretability in set theories, Comment. Math. Univ. Carolinae 12(1971), 73-79.

[6] P. HÁJEK: On interpretability in set theories II, Comment. Math. Univ. Carolinae 13(1972), 445-455.

[7] M. HÁJKOVÁ, P. HÁJEK: On interpretability in theories containing arithmetic, Fund. Math. 76(1972), 131-137.

[8] P. LINDSTRÖM: Some results on interpretability, Proc. 5th Scand. Log. Symp. Aalborg Univ. Press 1979.

[9] J.R. SHOENFIELD: Mathematical logic, Addison-Wesley

1967.

[10] C. SMORYŃSKI: Fifty years of self-reference in arithmetic, to appear.

[11] C. SMORYŃSKI: A ubiquitous fixed-point calculation, to appear.

[12] C. SMORYŃSKI: Calculating self-referential statements: Guaspari sentences of first kind, to appear.

[13] C. SMORYŃSKI: A short course in modal logic, handwritten notes.

[14] R. SOLOVAY: Interpretability in set theories, in preparation.

[15] R. SOLOVAY: Provability interpretations of modal logic, Israel J. of Math. 25(1976), 287-304.

[16] V. ŠVEJDAR: Degrees of interpretability, Comment. Math. Univ. Carolinae 19(1978), 789-813.

[17] A. TARSKI, A. MOSTOWSKI, R.M. ROBINSON: Undecidable theories, North-Holland Publ. Co. 1953.

[18] P. VOPĚNKA; P. HÁJEK: Existence of a generalized model of Gödel-Bernays set theory, Bull. Acad. Polon. Sci. 21(1973), 1079-1086.

Matematický ústav ČSAV, Žitná 25, 11000 Praha 1, Československo

(Oblatum 15.5. 1981)