Petr Hájek
On interpretability in set theories. II.

Commentationes Mathematicae Universitatis Carolinae, Vol. 13 (1972), No. 3, 445--455

Persistent URL: http://dml.cz/dmlcz/105433

Terms of use:
© Charles University in Prague, Faculty of Mathematics and Physics, 1972

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these Terms of use.

This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project DML-CZ: The Czech Digital Mathematics Library http://project.dml.cz
ON INTERPRETABILITY IN SET THEORIES II

Petr HÁJEK, Praha

This paper is a continuation of [2] and [3] and uses techniques developed in [1]. ZF denotes the Zermelo-Fraenkel set theory and GB the Gödel-Bernays set theory. We adopt conventions made in [3] § 1 (Preliminaries). GB is a conservative extension of ZF; so we have \( \text{Con}(ZF, \varphi) \iff \text{Con}(GB, \varphi) \) for each ZF-formula \( \varphi \). Denote by \( \mathcal{J}_{ZF}(J_{GB}) \) the set of all ZF-formulas \( \varphi \) such that \( \langle ZF, \varphi \rangle \) is interpretable in ZF ((GB, \varphi) is interpretable in GB). We know the following: (1) \( \varphi \in \mathcal{J}_{ZF} \cup \mathcal{J}_{GB} \Rightarrow \text{Con}(ZF, \varphi) \), (2) \( \mathcal{J}_{ZF} \cap \mathcal{J}_{GB} \neq \emptyset \), (3) \( \mathcal{J}_{ZF} \in \Pi_2^0 - \Sigma_1^0 \) and \( \mathcal{J}_{GB} \in \Sigma_1^0 \).

(We assume \( \text{Con}(ZF) \).) There remain the following questions:

(1) What is the exact position of \( \mathcal{J}_{ZF} \) in the arithmetical hierarchy? In particular, is \( \mathcal{J}_{ZF} \) a complete \( \Pi_2^0 \) set?

(2) What is the relation between \( \text{Con}(ZF, \varphi) \), \( \varphi \in \mathcal{J}_{ZF} \), \( \varphi \in \mathcal{J}_{GB} \)? In particular, is \( \mathcal{J}_{GB} - \mathcal{J}_{ZF} \) non-empty?

AMS, Primary: 02F35, 02K15 Ref. Z. 2.641.3, 2.653.1

- 445 -
Unfortunately, I have not succeeded to answer these questions exhaustively; but I hope that the results of this paper give some new information on both questions. We prove the following:

**Theorem 1.** If ZF is consistent then \( \mathcal{J}_{ZF} \not\equiv \Pi^0_1 \).

The question if \( \mathcal{J}_{ZF} \) is not a \( \Sigma^0_2 \)-set, in particular, if it is a complete \( \Pi^0_2 \)-set, remains open. According to question (2), if we had a (closed) formula \( \varphi \in \mathcal{J}_{GB} - \mathcal{J}_{ZF} \), it would satisfy the following:

\[ \text{Con}(ZF, \varphi), \text{Con}(ZF, \neg \varphi) \] (i.e. \( \varphi \) would be independent from ZF), \( \varphi \not\equiv \mathcal{J}_{ZF} \). I offer to the reader a formula with the following properties:

**Theorem 2.** If ZF is consistent then there is a closed ZF-formula \( \varphi \) such that (1) \( \varphi \) is independent from ZF, (2) neither \((ZF, \varphi)\) nor \((ZF, \neg \varphi)\) is interpretable in ZF and (3) neither \((GB, \varphi)\) nor \((GB, \neg \varphi)\) is interpretable in GB.

In Discussion, we mention possible generalizations of these results (in the spirit of [3]) for theories containing arithmetic and having some additional properties; we further show that if \( \mathcal{J}_{GB} - \mathcal{J}_{ZF} \) is non-empty then there is a very simple formula in this set. We conclude with some remarks.

It seems reasonable to use the following hierarchy of P-formulas (P is the Peano arithmetic): a P-formula is \( \Pi^m_n \) (or \( \Sigma^m_n \)) if it has a prefix containing \( m \) alternating quantifiers, the first one being universal (existential), followed by a PR-formula (see [1] for PR-formulas). There will
be no misunderstanding with the arithmetical hierarchy of sets of natural numbers (here we use $\Sigma_m^0$ and $\Pi_m^0$).

If $\Gamma$ is a class of formulas and $T$ is a theory then we say that a $T$-formula $\varphi$ is a $\Gamma$-formula in $T$ if there is a $\Gamma$-formula $\psi$ such that $T \vdash \varphi \equiv \psi$. Note that for each $T$ containing $\Sigma_1^1$-formulas in $T$ coincide with (Feferman's) RE-formulas in $T$.

**Lemma 1.** If $\kappa$ is an interpretation of ZF in ZF then there is a formula $\varphi$ with two free variables such that the following is ZF-provable ($x, y, ...$ are variables for natural numbers and $x^*, y^*, ...$ are variables for natural numbers in the sense of the interpretation):

1. $$(\forall x)(\exists! x^*) \varphi(x, x^*)$$
2. $\varphi(\overline{0}, \overline{x^*})$
3. $$(\varphi(x, x^*) \land \varphi(x + \overline{1}, y^*)) \rightarrow y^* = x^* + \overline{1}$$

**Proof.** Let $\text{Seq}(a)$ mean that $a$ is a finite sequence, let $\ell h(a)$ be the length of the sequence and let $(a)_i$ be the $i$-th member of $a$. We put

$$\varphi(x, x^*) \equiv (\exists a)(\text{Seq}(a) \land \ell h(a) = x + \overline{1}) \land (a)_{\overline{0}} = \overline{0} \land (\forall y < \ell h(a) = x + \overline{1})(x < (a)_{i+\overline{1}} = (a)_{i} + \overline{1})$$

One proves the above formulas by induction inside ZF.

**Lemma 2.** If $\kappa$ is an interpretation of ZF in ZF and if $\varphi$ is as in Lemma 1 then for each $\Sigma_1^1$-formula $\varphi(x, ...)$ we have:

$$(\ast) \ ZF \vdash (\varphi(x, x^*) \land ...) \rightarrow (\varphi(x, ...) \rightarrow \varphi^*(x^*, ...))$$

- 447 -
Proof. By [1] 3.9, it suffices to prove the present lemma for Feferman's BPF. First one proves by (metamathematical) induction

\[ ZF \vdash (\varphi(x,x^*) \& \ldots) \rightarrow (\psi(x,\ldots) \equiv \psi^*(x^*,\ldots)) \]

for each \( \psi \in EF \) using induction inside ZF; then one proves \( (*) \) for BPF (derive the following formulas from (1) - (3) in ZF):

(4) \( (\varphi(x,u^*) \& \varphi(\psi,u^*)) \rightarrow x = \psi \),

(5) \( (\varphi(x,u^*) \& u^* \leq *u^*) \rightarrow (\exists \psi < x) \varphi(\psi,u^*) \).

Corollary 1. If \( \varphi \) is a PR-formula then
\[ (\varphi(x,x^*) \& \ldots) \rightarrow (\varphi(x,\ldots) \equiv \varphi^*(x^*,\ldots)) \]
(since both \( \varphi \) and \( \neg \varphi \) are \( \Sigma_1 \)-formulas in ZF).

Corollary 2. If \( \varphi \) is a \( \Pi_1 \)-formula then
\[ (\varphi(x,x^*) \& \ldots) \rightarrow (\varphi^*(x^*,\ldots) \rightarrow \varphi(x,\ldots)) \]

Corollary 3. If \( \varphi \) is a closed \( \Pi_1 \)-formula and \( \varphi \in \mathcal{J}_{ZF} \) then \( ZF \vdash \varphi \).

It is of some interest that we can give an alternative proof of the last corollary using the Orey's result (cf. [3] Lemma 2):

Let \( \mathcal{A} \) be such that all the axioms of the arithmetic \( \mathcal{Q} \) are provable in \( ZF \vdash \mathcal{A} \). Then

\[ ZF \vdash \neg \varphi \rightarrow \mathcal{P}^*_{\mathcal{Q}_{(\neg \varphi)}} \rightarrow \mathcal{P}^*_{\mathcal{Z}_{\mathcal{Q}_{\mathcal{A}}}(\neg \varphi)} \rightarrow \neg \mathcal{C}_{\mathcal{Z}_{\mathcal{Q}_{\mathcal{A}}},(\varphi)} \]

i.e. \( ZF \vdash \mathcal{C}_{\mathcal{Z}_{\mathcal{Q}_{\mathcal{A}}},(\varphi)} \rightarrow \varphi \), which together with
Orey's result gives the corollary. (For the first implication see [1] 5.5.)

Lemma 3 (Feferman [1] 6.6 and 8.9). If \( \xi \) is a \( \text{PR} \)-bi-numeration of \( \text{ZF} \) then \( (\neg \text{Con}_\xi) \in \mathcal{J}_{\text{ZF}} \).

Proof of Theorem 1. Suppose that \( \mathcal{J}_{\text{ZF}} \) is \( \Pi^0_1 \), i.e. the complement of \( \mathcal{J}_{\text{ZF}} \) is recursively enumerable. Let \( \xi \) be a \( \text{PR} \)-bi-numeration of \( \text{ZF} \) in \( \text{ZF} \); then 

\( (\text{ZF}, \neg \text{Con}_\xi) \) is consistent and, by [3] Lemma 1, there is a "nice" numeration of \( \neg \mathcal{J}_{\text{ZF}} \) in \( (\text{ZF}, \neg \text{Con}_\xi) \), i.e. there is a \( \text{P} \)-formula \( \gamma \) such that

\[
\varphi \notin \mathcal{J}_\xi \iff (\text{ZF}, \neg \text{Con}_\xi) \vdash (\exists \gamma)(\gamma(\vec{\varphi}, \gamma) \iff \\
\iff (\exists \xi)(\text{ZF}, \neg \text{Con}_\xi) \vdash \gamma(\vec{\varphi}, \xi)).
\]

Note that \( \text{ZF} \vdash \gamma(\vec{\varphi}, \gamma) \equiv \alpha_0(\xi(\vec{\varphi}), \gamma) \) where \( \alpha_0 \) is a \( \Pi^0_1 \)-formula in \( \text{ZF} \) defined in [3] and \( \xi \) is a recursive function; hence \( \gamma(\vec{\varphi}, \gamma) \) is a \( \Pi^0_1 \)-formula in \( \text{ZF} \).

If \( \varphi \) is a formula and \( \varphi \notin \mathcal{J}_{\text{ZF}} \) then (i) \( \text{ZF} \vdash \varphi \), (ii) for some \( \kappa \) we have 

\( \text{ZF}, \neg \text{Con}_\xi \vdash \gamma(\vec{\varphi}, \kappa) \) and by (i) we have 

(iii) \( \text{ZF} \vdash (\forall \gamma < \kappa) \neg \Box_\xi(\vec{\varphi}, \gamma) \).

By the diagonal lemma [1] 5.1, find a closed \( \text{P} \)-formula such that

\( \text{ZF} \vdash \varphi \equiv (\exists \gamma)(\gamma(\vec{\varphi}, \gamma) \& (\forall x < \gamma) \neg \Box_\xi(\vec{\varphi}, x)) \).

Suppose \( \varphi \notin \mathcal{J}_{\text{ZF}} \); then, by (ii) and (iii) above, we have \( \text{ZF}, \neg \text{Con}_\xi \vdash \varphi \). Since \( \neg \text{Con}_\xi \in \mathcal{J}_{\text{ZF}} \) (by Lemma 3), we have \( \varphi \in \mathcal{J}_{\text{ZF}} \).
Suppose $\varphi \in I_{ZF}$, then

$\text{Con}(ZF, \neg \varphi, (\forall y) \neg \varphi(\vec{f}, y))$ by the properties of $\varphi$. Denote the interpretation of $(ZF, \varphi)$ in ZF by $\ast$ and the theory $(ZF, \neg \varphi)$ in ZF by $ZF_1$.

Then we have

1. $ZF_1 \vdash (\exists x) \mathcal{P}_x \mathcal{L}_x (\vec{f}, x)$ (from $\neg \varphi$)
2. $ZF_1 \vdash \neg (\exists y) \varphi(\vec{f}, y)$
3. $ZF_1 \vdash (\exists y^*) (\varphi^*(\vec{f}^*, y^*) \& (\forall x^* \neg \varphi^*) \neg \mathcal{P}_x \mathcal{L}_x (\vec{f}^*, x^*))$.

We proceed informally in $ZF_1$. Let $\varphi$ be as in Lemma 1. For $y^*$ from (3), there is no $y$ such that $\varphi(y, y^*)$ (say, $y^*$ is non-standard); otherwise we had $\varphi(\vec{f}, y)$ by Corollary 2. But if $x$ is as in (1) and if $\varphi(x, x^*)$ then $\mathcal{P}_x \mathcal{L}_x (\vec{f}^*, x^*)$ and necessarily $x^* < * y^*$ (cf. (5) in the proof of Lemma 21).

This contradicts (3). So we derived a contradiction in ZF. Hence we proved $\varphi \notin I_{ZF}$.

We see that the assumption $I_{ZF} \in \Pi_1^0$ leads to a contradiction; hence $I_{ZF}$ is not $\Pi_1^0$.

**Lemma 4** (Vopěnka [4]). $(GB, \neg \varphi)$ is interpretable in GB, i.e. $(\neg \varphi)_{GB} \in I_{GB}$.

**Proof of Theorem 2.** Let $\text{Int}_p (x, y)$ be a PR-formula saying "$y$ is an interpretation of $(GB, x)$ in $[GB]$" (cf. [2] or [3]) and find a $\varphi$ such that
$\text{ZF} \vdash \varphi \equiv (\forall x) (\text{Int}_p (\overline{\varphi}, x) \rightarrow (\exists y < x) \text{Int}_p (\overline{\varphi}, y))$

(by the way, $\varphi$ is the Rosser's formula with interpretability instead of provability).

(1) Let $d$ be the least interpretation of $(\text{GB}, \varphi)$ in GB; denote it by $\star$. Then $\text{GB} \vdash \varphi^*$, 
\[
\text{GB} \vdash (\text{Int}_p (\overline{\varphi}, \overline{d}))^*, \text{ i.e. }
\]
\[
\text{GB} \vdash [(\exists y < \overline{d}) \text{Int}_p (\overline{\varphi}, y)]^*. 
\]
The formula $(\exists y < \overline{d}) \text{Int}_p (\overline{\varphi}, y)$ is PR in GB, hence, by Corollary 1, $\text{GB} \vdash (\exists y < \overline{d}) \text{Int}_p (\overline{\varphi}, \overline{d})$ and hence there is a $d_1 < d$ which is an interpretation of $(\text{GB}, \neg \varphi)$ in GB. Denote it by $\square$. We have $\text{GB} \vdash \neg \varphi^\square$, $\text{GB} \vdash [\text{Int}_p (\overline{\varphi}, d_1^\square)]^\square$ and hence $\text{GB} \vdash [(\exists x < d_1^\square) \text{Int}_p (\overline{\varphi}, x)]^\square$ and

$\text{GB} \vdash (\exists x < \overline{d_1}) \text{Int}_p (\overline{\varphi}, x)$, so that there is a $d_2 < d_1 < d$ which is an interpretation of $(\text{GB}, \varphi)$ in GB. This is a contradiction, so that $\varphi \notin \text{J}_\text{GB}$.

(2) If $(\neg \varphi) \in \text{J}_\text{GB}$ then there is the least $d_1$ which is an interpretation of $(\text{GB}, \neg \varphi)$ in GB. By (1), then there is a $d_2$ which is an interpretation of $(\text{GB}, \varphi)$ in GB, which is a contradiction. Hence $(\neg \varphi) \notin \text{J}_\text{GB}$ and $\varphi$ is independent from GB (and from ZF).

(3) $\varphi \notin \text{J}_\text{ZF}$ since $\varphi$ is a $\Pi_1$-formula in ZF (cf. Corollary 3).
(4) To prove \((\neg \varphi) \notin \mathcal{J}_{ZF}\) we need the following

Lemma 5. If \(\xi\) is a PR-bi-numeration of ZF such that \(ZF \vdash \text{Con}_{\varepsilon} \Leftrightarrow \text{Con}_{\xi}\) and if \(\varphi\) is as above then \((ZF, \neg \text{Con}_{\xi}) \vdash \neg \varphi\).

Otherwise we had the following interpretations:

\[
\text{GB}, \neg \varphi \implies \text{GB}, \neg \text{Con}_{\xi} \implies \text{GB}, \neg \text{Con}_{\varepsilon \text{GB}} \implies \text{GB}.
\]

(Double arrows are identities; for the last arrow see Lemma 4.) By composition of interpretations, we would have an interpretation of \((\text{GB}, \neg \varphi)\) in GB, which is a contradiction. (Note that the "natural" bi-numeration of ZF has the desired property.)

We continue the proof of \((\neg \varphi) \notin \mathcal{J}_{ZF}\). Suppose the contrary. Then we have the following interpretations:

\[
ZF, \neg \varphi \implies ZF \implies ZF, \neg \text{Con}_{\xi}, \varphi.
\]

We consider the composed interpretation \(\ast\) of

\((ZF, \neg \varphi)\) in \((ZF, \neg \text{Con}_{\xi}, \varphi)\) and proceed in the last theory. Since \(\neg \text{Con}_{\xi}\) we have \(\neg \text{Con}_{\varepsilon \text{GB}}\) and hence there are \(\eta, \zeta\) such that \(\text{Int}_\mu(\overline{\varphi}, \eta)\) and \(\text{Int}_\mu(\overline{\varphi}, \zeta)\). Suppose that \(\eta\) and \(\zeta\) are least with the corresponding properties. Then, by \(\varphi\), \(\zeta\) is smaller than \(\eta\). On the other hand, we have \((\neg \varphi)^\ast\), which says

\[
(\exists \mu^\ast)(\text{Int}_\mu^\ast(\overline{\varphi}^\ast, \mu^\ast) \& (\forall \nu^\ast \leq \ast \mu^\ast) \neg \text{Int}_\mu^\ast(\overline{\varphi}^\ast, \nu^\ast)).
\]

If \(\varphi(x, x^\ast)\) then we have \(\text{Int}_\mu^\ast(\overline{\varphi}^\ast, x^\ast)\) and hence \(\mu^\ast \leq \ast x^\ast\); then there is a \(\mu\) such that

- 452 -
\( \varphi(\mu, \mu^*) \), \( \mu < \kappa \) and \( \text{Init} \left( \theta \varphi, \mu \right) \) which is a contradiction. Since \( (\text{ZF}, \neg \text{Con}_\varphi) \) is consistent by Lemma 5, there is no interpretation of \( (\text{ZF}, \neg \varphi) \) in \( \text{ZF} \), q.e.d.

Discussion. (1) Let us first discuss the possibility of generalizing Theorems 1 and 2 for theories containing arithmetic. Inspection of the proof of Theorem 1 shows that its assertion holds for any primitive recursively axiomatized theory \( T \) containing \( P \), which is consistent, essentially reflexive (so that [3] Lemma 1 applies) and satisfies Lemmas 1 and 2. The proofs of these lemmas apply to each theory \( T \) in which, in addition to the assumptions just made, the induction schema is provable for all \( T \)-formulas and in which sequences of arbitrary objects are definable. (Note in passing that in GB sequences of arbitrary classes are easily definable, but the induction schema is not provable for all formulas.) Concerning Theorem 2, let \( P \subseteq T \subseteq S \), where \( T \) is as above and \( S \) is a conservative finitely axiomatized extension of \( T \). We need two additional assumptions in \( S \): (i) There is a FR-bi-numeration \( \alpha \) of \( T \) in \( T \) such that \( T \vdash \text{Con}_{S_{[S]}} \equiv \text{Con}_\alpha \).

(This is the case e.g. if the formal statement saying " \([ S \) is a conservative extension of \( \alpha \) " is provable in \( T \).)

(ii) \( (S, \neg \text{Con}_{S_{[S]}}) \) is interpretable in \( S \).

This is an important assumption; it is not clear how to
modify Vopěnka's original proof of \( \neg \forall \langle \forall \rangle \in G_B \) \( \forall \langle \forall \rangle \notin G_P \) e.g. for a proof of \( \neg \forall \langle \forall \rangle \in G_P \) where \( G_P \) is the finitely axiomatized conservative extension of the Peano's arithmetic with classes (say, Gödel-Peano). Let us stress the fact that one cannot use Feferman's \([1]\) 8.9 for \( S \) since \( S \) is finitely axiomatizable and therefore not reflexive.

(2) Suppose that we would find a ZF-formula \( \varphi \) such that \( \varphi \in J_{GB} - J_{ZF} \). Then, by Orey's result, there is a natural number \( k \) such that \( \forall \langle \forall \rangle \in J_{ZF}, \langle \forall \rangle \notin G \) is not provable in ZF. Denote the last formula by \( \varphi_0 \). It is a \( \Pi_1 \)-formula and, moreover, a \( \Pi_1 \)-formula. Since \( ZF \vdash \varphi_0 \) we have \( \varphi_0 \notin J_{ZF} \) by Corollary 3. On the other hand, if \( * \) is an interpretation of \( (GB, \varphi) \) in \( GB \) then \( GB \vdash \varphi^* \), \( GB \vdash (\varphi \rightarrow \varphi_0)^* \) by essential reflexivity of ZF and by \( ZF \leq GB \); hence we have \( GB \vdash \varphi_0^* \) and \( \varphi_0 \in J_{GB} \). So we have proved the following

**Fact.** If \( J_{GB} - J_{ZF} \neq \emptyset \) then there is a \( \Pi_1 \)-formula in \( J_{GB} - J_{ZF} \).

This contrasts with Corollary 3; by this corollary, no \( \Pi_1 \)-formula is in \( J_{ZF} - J_{GB} \) (Examples of formulas in \( J_{ZF} - J_{GB} \) constructed in [2] and [3] are \( \Pi_2 \)-formulas.)

(3) It follows by Orey's result that \( \varphi \in J_{ZF} \) iff there is a recursive function \( f \) such that, for each \( \langle \forall \rangle \),
\( f \langle \alpha \rangle \) is a proof of \( \text{Con}_{\text{ZF} \cup \varphi} \) in ZF. Define \( \varphi \in J_{ZF}^{\text{prim}} \) iff there is a primitive recursive function \( f \) such that, for each \( \alpha \), \( f \langle \alpha \rangle \) is a proof of \( \text{Con}_{\text{ZF} \cup \varphi} \) in ZF. Then \( J_{ZF}^{\text{prim}} \) is \( \Sigma^0_2 \) (by the existence of a recursive function universal for primitive recursive functions). Inspection of the proof in [2] shows that \( J_{ZF}^{\text{prim}} - J_{GB} \) is non-empty (assuming that ZF is \( \omega \)-consistent).

Is \( J_{ZF}^{\text{prim}} \neq \emptyset \)? Can we weaken the assumption of \( \omega \)-consistency to \( \text{Con}(ZF) \) in the proof of \( J_{ZF}^{\text{prim}} - J_{GB} \neq \emptyset \) using methods of [3] or other methods?

References


Matematický ústav ČSAV

Žitná 25, Praha 1
Československo

(Oblatum 17.7.1972)