$C = C(R')$ and F is the field of quotiens of C. The algebra $R' \otimes_C F$ is nil, whence by assumption $(R' \otimes_C F)[x]$ is also nil. Since, however, $R'[x]$ is contained in $(R' \otimes_C F)[x]$, $R'[x]$ is nil, which is impossible.

Finally, let us assume that $P_5 F$ has a positive solution for any field F. Then obviously $P_6 F$ has a positive solution, whence — as we have just proved above — P_6 has a positive solution. Now, applying Proposition 1, we obtain a positive solution of P_5.

The converse implications are obvious.

References

[1] S. Amitsur, *Algebras over infinite fields*, Proc. Amer. Math. Soc. 7 (1956), pp. 35–48.
[2] — *Radicals of polynomical rings*, Canad. J. Math. 8 (1956), pp. 355–361.
[3] N. Divinsky, *Rings and radicals*, University of Toronto Press 1965.
[4] Е. С. Голод, *О ниль-алгебрах и финитно-аппроксимируемых p-группах*, ИАН СССР, сер. матем. 28 (1964), pp. 273–276.
[5] Е. С. Голод, И. Р. Шафаревич, *О башне полей классов*, ibidem 28 (1964), pp. 261–272.
[6] I. N. Herstein, *Theory of rings*, University of Chicago Mathematics Lecture Notes 1961.
[7] N. Jacobson, *Structure of rings*, A.M.S. Coll. Publ. XXXVII 1964.
[8] G. Koethe, *Die Structure der Ringe deren Restenblassenring dem Radikal vollständig ist*, Math. Zeit. 32 (1930), pp. 161–186.

INSTITUTE OF MATHEMATICS OF THE POLISH ACADEMY OF SCIENCES

Reçu par la Rédaction le 27. 4. 1971

On interpretability in theories containing arithmetic

by

Marie Hájková and **Petr Hájek** (Prague)

0. Introduction. In [3], a ZF-formula φ was constructed such that (ZF, φ) is relatively interpretable in ZF but (GB, φ) is not relatively interpretable in GB, provided ZF is ω-consistent. (ZF denotes the Zermelo–Fraenkel set theory, and GB the Gödel–Bernays set theory.) This result is generalized in the present paper in two ways: first, we replace the assumption of ω-consistency by the assumption of (usual) consistency and, secondly, we replace ZF and GB by an arbitrary couple of theories related similarly as ZF and GB and containing arithmetic. Similarly as in [3], our result is an immediate consequence of a general theorem (Theorem 1) concerning reflexive theories containing arithmetic. A technical lemma (Lemma 1) concerning "nice" numerations of recursively enumerable sets, which is the key device of removing the assumption of ω-consistency, is — in a certain sense — a generalization of the result of [1] and might be useful also in other connections. Some other consequences of Theorem 1 are listed at the end of the paper. The knowledge of [3] is not necessary to understand this paper, but the reader is supposed to be familiar with [2] and with some topics of the recursion theory.

We thank Professor A. Mostowski for his interest and valuable discussions during the second autbor's visit to Warsaw in December 1970.

1. Preliminaries. Theories are assumed to be formalized in the predicate calculus with equality, denumerably many predicates and functions of each finite arity, denumerably many constants and denumerably many sorts of variables (there are denumerably many variables of each sort). A *theory* is a pair consisting of a language and of a set of formulas of that language (special axioms), a language being a list of predicates, functions, constants and sorts of variables. A sort s is *subordinated* to a sort t in a theory T if $T \vdash (\forall x^s)(\exists y^t)(x^s = y^t)$ (where x^s is a variable of the sort s, etc.). A sort s is *universal* in T if each sort of the language of T is subordinated to s in T. We restrict ourselves to theories having a universal sort. *Definitions* of new predicates, functions, constants and sorts in a theory T are additional axioms δ of a particular form enriching the language of T but extending T *conservatively*; i.e. if (T, δ) is an extension of T by a definition and if φ is a T-formula provable in (T, δ), then φ is provable in T. An *interpretation* of T in S is a mapping $*$ which associates with each predicate P, function F, constant c and sort s of the language of T a predicate P^*, function F^*, constant c^* and sort s^* of the language of T such that the following holds true: (1) The arity of P^* equals the arity of P and the arity of F^* equals the arity of F. (2) If, for any T-formula φ, φ^* is the formula resulting from φ by replacing each predicate P, function F, constant c and variable x^s by P^*, F^*, c^*, x^{s*} respectively, then $\mathrm{S} \vdash \varphi^*$ for each axiom φ of T. A theory T is *interpretable* in S if there is an extension $\hat{\mathrm{S}}$ of S by definitions and an interpretation of T in $\hat{\mathrm{S}}$.

We hope that the reader can imagine an appropriate detailed construction of the formalism ontlined above; one can consult [10], Chapt. I, Sect. 2 for more details. In particular, interpretations (called direct syntactic models in [10]) are particular syntactic models and hence guarantee relative consistency. On the other hand, our notion of interpretation and interpretability is very closely related to the notion of relative interpretations and relative interpretability due to Tarski [9]; the present notions are obviously equivalent reformulations of the original ones for the logic with various sorts of variables.

Peano's arithmetic (as a theory with one sort of variables, no predicate except $=$, functions $'$, $+$, $\cdot$ and the constant $\bar{0}$) is denoted by P (see e.g. [2] for the axioms of P). A theory T is said to *contain* arithmetic if P is interpretable in T. As far as interpretability in T and of T is concerned, if T contains P one can assume without loss of generality that P is simply a subtheory of T, i.e. that the identical mapping is an interpretation of P in T. (One has only to replace T by a conservative extension and rename some symbols in the language.) For example, both ZF and GB contain arithmetic.

If P is a subtheory of T, then the universal sort of P is a sort of T; variables of this sort will be called *number variables* and denoted by x, y, z (with or without subscripts) throughout the paper. One has also the constant $\bar{0}$ and the terms $\bar{n} = \bar{0}\underbrace{'\cdots'}_{n\,\text{times}}$ in T and the notions of a numeration and bi-numeration of a set of natural numbers make sense. Similarly for other notions defined in [2]; for example, for every bi-numeration $\alpha(\mathrm{x})$ of T in T one has the formula Con_α, expressing consistency. T is called *reflexive* if, for each natural number n, we have $\mathrm{T} \vdash \mathrm{Con}_{[\mathrm{T} \restriction n]}$. T is called essentially reflexive if each extension of T with the same language is reflexive. Evidently, T is essentially reflexive if and only if, for every closed T-formula φ, $\mathrm{T} \vdash \varphi \rightarrow \mathrm{Con}_{[\{\varphi\}]}$. Note that both P and ZF are essentially reflexive. (See [2] for references.)

2. "Nice" numerations of recursively enumerable sets. Let W be are a recursively enumerable set of natural numbers. There are (primitive) recursive relations A such that $W = \{m; (\exists n) A(m, n)\}$. Let T be a recursively axiomatized theory containing arithmetic, let A be as above and let $\alpha(\mathrm{x}, \mathrm{y})$ be a bi-numeration of A in T. Then the formula $(\exists \mathrm{y})\, \alpha(\mathrm{x}, \mathrm{y})$ is a numeration of a recursively enumerable set W' in T, but in general W and W' need not be the same; the only thing we can say is that $W \subseteq W'$. If we assumed T to be ω-consistent, we could conclude that $W = W'$ but we want to avoid this assumption. So let us ask if, given W, there is a relation A and a bi-numeration $\alpha(\mathrm{x}, \mathrm{y})$ of A in T such that $W = \{m; (\exists n) A(m, n)\}$ and $(\exists \mathrm{y})\, \alpha(\mathrm{x}, \mathrm{y})$ numerates W in T.

LEMMA 1. *Let* T *be a theory containing arithmetic and let* T *be consistent, reflexive and recursively axiomatized. Then, for each recursively enumerable set* W *of natural numbers, there is a recursive relation* A *and a bi-numeration* $\alpha(\mathrm{x}, \mathrm{y})$ *of* A *in* T *such that* $W = \{m; (\exists n) A(m, n)\}$ *and the formula* $(\exists \mathrm{y})\, \alpha(\mathrm{x}, \mathrm{y})$ *numerates* W *in* T.

Proof. Similarly as in [1], it suffices to find one recursively enumerable set which is creative and for which our assertion is true. For, let W_0 be creative, let $W_0 = \{m; (\exists n) A_0(m, n)\}$, let α_0 bi-numerate A_0 in T and let $(\exists \mathrm{y})\, \alpha_0(\mathrm{x}, \mathrm{y})$ numerate W_0 in T. If W is an arbitrary recursively enumerable set, then there is a (total) recursive function f such that, for all n, $n \in W$ iff $f(n) \in W_0$. (See [6].) Since f is recursive, we have a formula χ such that

$$(1) \qquad \mathrm{T} \vdash \chi(\overline{m}, \mathrm{z}) \equiv \mathrm{z} = \overline{f(m)} \quad \text{for all } m\,.$$

Now,

$$W = \{m; (\exists n)(\exists q)\big(f(m)\big) = n \,\&\, A_0(n, q)\}$$
$$= \{m; (\exists n)\big(f(m)\big) = (n)_0 \,\&\, A_0\big((n)_0, (n)_1\big)\}\,.$$

Put

$$A(m, n) \equiv f(m) = (n)_0 \,\&\, A_0\big((n)_0, (n)_1\big)$$

and

$$\alpha(\mathrm{x}, \mathrm{y}) \equiv \chi\big(\mathrm{x}(\mathrm{y})_{\cdot\bar{0}}\big) \,\&\, \alpha_0\big((\mathrm{y})_{\cdot\bar{0}}, (\mathrm{y})_{\cdot\bar{1}}\big)\,.$$

Evidently, α bi-numerates A in T and it is easy to show (using (1) and the properties of α_0) that $(\exists \mathrm{y})\, \alpha(\mathrm{x}, \mathrm{y})$ numerates W in T.

Hence let us define an appropriate creative set. In fact, we take for W_0 the set D' defined in [1], p. 41. In more detail, if

$$A = \{n; (\exists p)\big[T_1\big((n)_0, n, p\big) \,\&\, (\forall q \leqslant p) \neg T_1\big((n)_1, n, q\big)\big]\}\,,$$

$$B = \{n; (\exists q)\big[T_1\big((n)_1, n, q\big) \,\&\, (\forall p \leqslant q) \neg T_1\big((n)_0, n, p\big)\big]\}$$

are effectively inseparable sets defined by Kleene [5] and if $\tau_1(x, y, z)$ is a PR-formula bi-numerating T_1 in P, we put

$$\Psi(x) = (\exists y)\left[\tau_1\left((x)_{\bar{0}}, x, y\right) \mathbin{\&} (\forall z \leqslant y) \neg\tau_1\left((x)_{\bar{1}}, x, z\right)\right]$$

and

$$W_0 = \{m;\ T \vdash \neg\Psi(\overline{m})\} = \left\{m;\ (\exists n) Prf_T\left(\neg\Psi(\overline{m}), n\right)\right\}.$$

(So $A_0(m, n) \equiv Prf_T(\neg\Psi(\overline{m}), n)$.) Obviously W_0 is creative (cf. [1]). By [2] 8.6, there is a bi-numeration $\beta^*(x)$ of T in T such that $T \vdash \mathrm{Con}_{\beta^*}$. Put $\alpha_0(x, y) = \mathrm{Prf}_{\beta^*}\left(\mathrm{Sb}^{\mathrm{vr}\bar{0}}_{\mathrm{nm}_x}(\overline{\neg\Psi}), y\right)$ and observe that $T \vdash \alpha_0(\overline{m}, y) \equiv \mathrm{Prf}_{\beta^*}\left(\overline{\neg\Psi(\overline{m})}, y\right)$. (Note that $\alpha_0(x, y) = \mathrm{Prf}_{\beta^*}\left(\overline{\neg\Psi(x)}, y\right)$ in the notation of [4], 1.4.) Evidently, α_0 bi-numerates A_0 in T. (See [2], 4.4 and p. 57.) It remains to show that $(\exists y)\,\alpha_0(x, y)$ numerates W_0 in T. If $m \in W_0$, then obviously $T \vdash (\exists y)\,\alpha_0(\overline{m}, y)$. To prove the converse it suffices to show $T \vdash \mathrm{Pr}_{\beta^*}\left(\overline{\neg\Psi(\overline{m})}\right) \to \neg\Psi(\overline{m})$ for each m. $\Psi(\overline{m})$ is equivalent to an RE-formula in T, whence $T \vdash \Psi(\overline{m}) \to \mathrm{Pr}_{\beta^*}\left(\overline{\Psi(\overline{m})}\right)$ by [2], 5.4. (Cf. also [4], 1.7 if necessary.) Consequently

$$T \vdash \left(\mathrm{Pr}_{\beta^*}\left(\overline{\neg\Psi(\overline{m})}\right) \mathbin{\&} \Psi(\overline{m})\right) \to \neg\mathrm{Con}_{\beta^*},$$

which implies

$$T \vdash \mathrm{Pr}_{\beta^*}\left(\overline{\neg\Psi(\overline{m})}\right) \to \neg\Psi(\overline{m}),$$

q.e.d.

3. Interpretability in essentially reflexive theories.

LEMMA 2. *Let* T *be a theory containing arithmetic and suppose* T *to be recursively axiomatized and reflexive. Then, for any* T*-formula* φ, (T, φ) *is interpretable in* T *if and only if* $T \vdash \mathrm{Con}_{[(T,\varphi)\upharpoonright n]}$ *for each* n.

Proof. The implication $\Leftarrow$ follows by [2], 6.2 and 6.9. Conversely, if (T, φ) is interpretable in T, then for each n there is an m such that $(T, \varphi) \upharpoonright n$ is interpretable in $T \upharpoonright m$; hence the implication $\Rightarrow$ follows by [2], 6.4 and by reflexivity.

THEOREM 1. *Let* T *be a theory containing arithmetic and suppose* T *to be recursively axiomatized and essentially reflexive. Let* W *be a recursively enumerable set of* T*-formulas such that* $\varphi \in W$ *implies* $Con(T, \varphi)$. *Then there is a* T*-formula* φ *such that* (T, φ) *is interpretable in* T *and* $\varphi \notin W$. (*Equivalently, denote by* $\mathfrak{J}_T$ *the set of all* T*-formulas such that* (T, φ) *is interpretable in* T *and by* $\mathcal{F}_T$ *the set of all* T*-formulas refutable in* T. *If* Con(T), *then there is no r.e. set* W *containing* $\mathfrak{J}_T$ *and disjoint from* $\mathcal{F}_T$.)

Proof. The theorem is trivial if T is inconsistent, so suppose that T is consistent. Let W be given and let A, α be as in Lemma 1. Let β be an arbitrary bi-numeration of T in T. Using the diagonal Lemma 5.1 of [2], we can construct a formula φ (containing only number variables) such that

$$T \vdash \varphi \equiv (\forall y)\left(\alpha(\overline{\varphi}, y) \to \neg\mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright y}\right)$$

(where $\beta(x) \cup \{z\} \upharpoonright y$ is the formula $\left(\beta(x) \vee x = z\right) \mathbin{\&} x \leqslant y$).

(a) $Con(T, \varphi)$. Otherwise we have $T \vdash (\exists y)\,\alpha(\overline{\varphi}, y)$ and therefore $\varphi \in W$, which implies $Con(T, \varphi)$.

(b) $\varphi \notin W$. Otherwise we have $A(\varphi, n)$ for some n; then $T \vdash \alpha(\overline{\varphi}, \overline{n})$ and $(T, \varphi) \vdash \neg\mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright \overline{n}}$. But since (T, φ) is consistent and reflexive (by the essential reflexivity of T), we have $(T, \varphi) \vdash \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright \overline{n}}$, which implies that (T, φ) is inconsistent, contradicting (a).

(c) (T, φ) is interpretable in T. By Lemma 2 it suffices to show $T \vdash \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright \overline{n}}$ for each n. Since $(T, \varphi) \vdash \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright \overline{n}}$ by the essential reflexivity of T, it suffices to show $(T, \neg\varphi) \vdash \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright \overline{n}}$. But $\neg\varphi$ is equivalent in T to $(\exists y)\left(\alpha(\overline{\varphi}, y) \mathbin{\&} \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright y}\right)$. Since $\varphi \notin W$ by (b) and since α bi-numerates A in T, we have $T \vdash \neg\alpha(\overline{\varphi}, \overline{n})$ for each n. Hence we have $(T, \neg\varphi) \vdash (\exists y)(y > \overline{n} \mathbin{\&} \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright y})$ for each n, which implies $(T, \neg\varphi) \vdash \mathrm{Con}_{\beta \cup \{\overline{\varphi}\} \upharpoonright \overline{n}}$. This completes the proof.

(An infinitistic remark.) If all the T-provable P-formulas are true (in the natural model of P) — e.g. if T is P — then the formula φ constructed above is evidently true since $\varphi \notin W$ and therefore $\neg(\exists y)\,\alpha(\overline{\varphi}, y)$ is true.

4. Interpretability and finitely axiomatized extensions of theories containing arithmetic.

LEMMA 3. *If* T *is finitely axiomatized, then the set of all* T*-formulas* φ *such that* (T, φ) *is interpretable in* T *is recursively enumerable.*

The lemma is more or less evident; to prove it in detail one must show the recursive enumerability of the following predicates: (1) The predicate $Def(\delta, T)$, saying that the formula δ is a definition of a new symbol (a predicate, function, constant, sort respectively) in T. ($Def(\delta, T)$ says that δ has a certain syntactic structure and that something is provable in T). (2) The predicate $Ext(T', T)$, saying that T′ is an extension of T by definitions such that the sequence of defined symbols forms a language L′ of the same type as the language L of T. (3) The predicate $Intp(T', T, \varphi)$, saying that $Ext(T', T)$, φ is a T-formula and all the axioms of (T, φ) understood in the sense of the language L′ are provable in T. Then our set is $\{\varphi;\ (\exists T') Intp(T', T, \varphi)\}$, T being fixed; this set is obviously recursively enumerable.

THEOREM 2. *If* T *is a consistent recursively axiomatized essentially reflexive theory containing arithmetic and* S *is a consistent finitely axiomatizable extension of* T, *then there is a* T*-formula* φ *such that* (T, φ) *is interpretable in* T *but* (S, φ) *is not interpretable in* S.

Proof. Put $W = \{\varphi; \varphi$ T-formula & (S, φ) interpretable in S$\}$. By Lemma 3, W is recursively enumerable (W is the intersection of a primitive recursive set and a recursively enumerable set). Since S is an extension of T, $Con(\mathrm{S}, \varphi)$ implies $Con(\mathrm{T}, \varphi)$ for each T-formula φ; and since S is consistent, the interpretability of (S, φ) in S implies the consistency of (S, φ). Hence, by Theorem 1, there is a T-formula φ such that (T, φ) is interpretable in T but $\varphi \notin W$, i.e. (S, φ) is not interpretable in S.

Remarks. (1) Let T be as above and let S be a conservative finitely axiomatizable extension of T. Then, for each T-formula φ, (i) $\mathrm{T} \vdash \varphi$ iff $\mathrm{S} \vdash \varphi$; consequently, (ii) $Con(\mathrm{T}, \varphi)$ iff $Con(\mathrm{S}, \varphi)$. But, by the preceding theorem, the demonstrability of $Con(\mathrm{T}, \varphi)$ by means of an interpretation of (T, φ) in T is not the same as the demonstrability of $Con(\mathrm{S}, \varphi)$ by means of an interpretation of (S, φ) in S. In particular, one can take ZF for T and GB for S (assuming that ZF is consistent); see [8] for a finitary proof of the fact that GB is a conservative extension of ZF. Following the remark in [7], p. 90 and using [8], one can construct for various "well axiomatized" theories their conservative finitely axiomatized extensions; one can describe the situation generally, but one must replace the rough description of what a permissible schema is by an inductive definition of a certain class of primitive recursive functions mapping formulas into formulas, such that each of these functions defines a permissible schema. Then "well axiomatized" means "axiomatized by a finite number of single axioms and finitely many permissible schemas". On the other hand, it is not necessary to restrict ourselves to theories with one sort of variables.

(2) Let T, S be as in the preceding remark and denote by $\mathfrak{I}_{\mathrm{T}}$ the set of all T-formulas φ such that (T, φ) is interpretable in T and by $\mathfrak{I}_{\mathrm{S}}$ the set of all T-formulas such that (S, φ) is interpretable in S. By Lemmas 2 and 3, $\mathfrak{I}_{\mathrm{S}}$ is a Σ_1^0-set, $\mathfrak{I}_{\mathrm{T}}$ is a Π_2^0-set and, by Theorem 1, $\mathfrak{I}_{\mathrm{T}}$ is not a Σ_1^0-set. We do not know whether $\mathfrak{I}_{\mathrm{T}}$ is a Σ_2^0-set or even a Π_1^0-set. By Theorem 2, $\mathfrak{I}_{\mathrm{T}} - \mathfrak{I}_{\mathrm{S}} \neq \emptyset$; is $\mathfrak{I}_{\mathrm{S}} - \mathrm{T}_{\mathrm{T}} \neq \emptyset$?

(3) The formula φ constructed in [3] has the form $(\forall x)(\exists y)\psi$, where ψ is a PR-formula, as one can easily see. Even if our present φ is more complicated, one can verify that it still has the same form. What is the simplest form of formulas such that there is a φ of that form such that $\varphi \in \mathfrak{I}_{\mathrm{ZF}} - \mathfrak{I}_{\mathrm{GB}}$?

(4) Some particular consequences of Theorem 1. (i) Let $\mathfrak{I}_{\mathrm{P}}$ be the set of all P-formulas φ such that (P, φ) is interpretable in P and let $W = \{\varphi; \mathrm{P} \vdash \varphi\}$; then $\mathfrak{I}_{\mathrm{P}} - W \neq 0$ (cf. [2], 6.6). (ii) Let (ax) be a ZF-formula and suppose that (ZF, ax) is consistent (e.g. (ax) is an axiom of large cardinals). Let $W = \{\varphi; \varphi$ P-formula & $(\mathrm{ZF}, \mathrm{ax}) \vdash \varphi\}$. Then $\mathfrak{I}_{\mathrm{P}} - W \neq \emptyset$. In other words, there is a formula φ such that the consistency of φ with P can be proved via interpretations, φ is true in the natural model, but φ cannot be proved in (ZF, ax). (iii) Let (ax) be as above, and let $W = \{\varphi; \varphi$ P-formula & (GB, φ) interpretable in $(\mathrm{GB}, \mathrm{ax})\}$. Then $\mathfrak{I}_{\mathrm{P}} - W \neq 0$.

References

[1] A. Ehrenfeucht and S. Feferman, *Representability of recursively enumerable sets in formal theories*, Arch. Math. Logik Grundlagenforsch. 5 (1960), pp. 37–41.

[2] S. Feferman, *Arithmetization of metamathematics in a general setting*, Fund. Math. 49 (1960), pp. 35–92.

[3] P. Hájek, *On interpretability in set theories*, Comment. Math. Univ. Carolinae 12 (1971), pp. 73–79.

[4] M. Hájková, *The lattice of bi-numerations of arithmetics* I, ibidem, pp. 81–104.

[5] S. C. Kleene, *Introduction to metamathematics*, Amsterdam 1952.

[6] J. Myhill, *Creative sets*, Z. Math. Logik Grundlagen Math. 1 (1955), pp. 97–108.

[7] I. L. Novak, *A construction of models for consistent systems*, Fund. Math. 37 (1950), pp. 87–110.

[8] J. R. Shoenfield, *A relative consistency proof*, J. Symbolic Logic 19 (1954), pp. 21–28.

[9] A. Tarski, A. Mostowski and R. M. Robinson, *Undecidable theories*, Amsterdam 1953.

[10] P. Vopěnka and P. Hájek, *The theory of semisets*, Amsterdam and Prague 1972.

Reçu par la Rédaction le 5. 5. 1971