On Interpretability in Peano Arithmetic

Robert M. Solovay

A letter to Petr Hájek, May 31, 1976

Annotation

This is a scan, created in June 2022, of a letter in Petr Hájek's personal archive. It contains an original result, supposedly the first proof of the fact that the set of all sentences φ such that (PA, φ) is interpretable in PA is Π_2-complete. The proof answered a question left open in [HH72] and it was never published. The paper [Lin79] contains an independent different proof of the same fact.

[Sol76a] is a self-citation to *this* file. [Sol76b] is another letter sent in the same year. It answers other question raised in [HH72], and it was also never published.

References

[HH72] M. Hájková and P. Hájek. On interpretability in theories containing arithmetic. *Fundamenta Mathematicae*, 76:131–137, 1972.

[Lin79] P. Lindström. Some results on interpretability. In V. Jensen, B. H. Mayoh, and K. K. Møller, eds., *Proc. 5th Scand. Logic Symp. 1979*, pp. 329–361, Aalborg, 1979. Aalborg University Press.

[Sol76a] R. M. Solovay. On interpretability in Peano arithmetic. Unpublished letter to P. Hájek, www.cs.cas.cz/~hajek/RSolovayIntpPA.pdf, May 31, 1976.

[Sol76b] R. M. Solovay. Interpretability in set theories. Unpublished letter to P. Hájek, www.cs.cas.cz/~hajek/RSolovayZFGB.pdf, Aug. 17, 1976.

File created by Zuzana Haniková, Dagmar Harmancová and Vítězslav Švejdar in June 2022. A BibTEX entry to cite this letter can be as follows. If both letters, this and the other one written in August, are needed in a list of references, then this one should probably come first. This is what `\noopsort` does:

```
@Preamble(
\providecommand{\href}[2]{#2}
\providecommand{\nolinkurl}[1]{\url{#1}}
\newcommand{\rurl}[1]{\href{http://#1}{\nolinkurl{#1}}}
)
@Unpublished(solo:cmpl76,
  author="Robert M. Solovay",
  title ="On Interpretability in {P}eano Arithmetic",
  note  ="Unpublished letter to P. H{\'a}jek,
          \rurl{www.cs.cas.cz/~hajek/RSolovayIntpPA.pdf}",
  year  ="{\noopsort{Auf}M}ay~31, 1976"
)
```

May 31, 1976

Dear Professor Hájek,

I can settle a question left open in your joint paper with Ms. Hájková "On interpretability in theories containing arithmetic".

Let T be either ZF or $\underline{P}$. (See my P.P.S. for the needed conditions on T.) Then the set $\{\varphi : (T, \varphi) \text{ is interpretable in } T\}$ is a complete Π^0_2 set. Hence it is not Σ^0_2. (Our proof in each case can be formalized in $\underline{P} + Con(T)$.)

I shall sketch the proof for $\underline{P}$; the proof for ZF is slightly simpler.

1. Let $T(e, x, y)$ be Kleene's T predicate. Our proof will construct a Π^0_3 formula $\Phi(x)$ so that $\underline{P} + \Phi(\underline{e})$ is interpretable in $\underline{P}$ iff $\forall n \exists y\, T(e, n, y)$.

Let $T_{e,n}$ be the following recursively axiomitized theory. $T_{e,n} = \{\varphi : \varphi$ is an axiom of $\underline{P}$, and if $\#(\varphi)$ is the Gödel no. of φ, $(\forall s \leq \#(\varphi)) \neg T(e, n, s)$.

<u>Lemma 1</u> $P \vdash Con(T_{e,n})$ iff $(\exists s)\, T(e,n,s)$.

<u>Proof</u>. if $(\exists s)\, T(e,n,s)$ then the theory $T_{e,n}$ is finitely axiomitizable subtheory of P, and P can prove this. So $P \vdash Con(T_{e,n})$.

Suppose ~~$T_{e,n}$~~ $\neg(\exists s)\, T(e,n,s)$. Then $T_{n,s} = P$, So if $P \vdash Con(T_{n,s})$, $T_{n,s} \vdash Con(T_{n,s})$. By Gödel's second incompleteness theorem (as presented in [2] of your paper), $T_{n,s}$ $(= P)$ is inconsistent, which is absurd. So P does not prove $Con(T_{e,n})$.

2. Our proof will use an auxillary sequence of finitely axiomitizable subtheories of P, $P_0, P_1, P_2, \ldots$. The essential properties we need are 1) P_0 is a large finitely axiomitizable theory; 2) P_{n+1} is "much larger" than P_n, n. 3) The definition and elementary properties of the P_i's can be formalized in P_0.

Here is a definition that will work.

Let $A(0,n) = 2^{n+1}$

$A(n+1,0) = A(n,n)$

$A(n+1,k+1) = A(n, A(n+1,k))$.

Let $P_n = \{\varphi : \varphi$ is an axiom of P and $\#(\varphi) \leq A(n+10,0)\}$.

($A(n,0)$ is a variant of Ackerman's function.)

In P_{n+1}, we can prove the following facts about P_n:

1) P_n is Σ^0_1-sound. I.e., if $P_n \vdash \varphi$, φ Σ^0_1, then φ is true.

2) $P_n +$ "$\neg P_n$ is Σ^0_1-sound" $\vdash \sigma$, σ Σ^0_1 sentence, then $P_n \vdash \sigma$. (Apply Gödel to $P_n + \neg\sigma$.)

3. We now define the sentence $\Phi(e)$: It says:

$(\forall n)[$ If P_n is Σ^0_1 sound, then $T_{e,n}$ is consistent$]$.

<u>Lemma 2</u> If $\forall n$, $P \vdash Con(P_n + \Phi(e))$, then $\forall n \exists y\, T(e,n,y)$.

Work in P.

Proof. Fix n. Now $P \vdash Con(P_{n+1} + \Phi(e))$. ~~Hence~~

$P_{n+1} + \Phi(e)$ is consistent. Also $P_{n+1} \vdash P_n$ is Σ^0_1-sound.

Whence $P_{n+1} + \Phi(e) \vdash Con(T_{e,n})$. Whence

$P_{n+1} + Con(T_{e,n})$ is consistent. So $T_{e,n}$ is consistent.

I.e. $P \vdash Con(T_{e,n})$. But then by Lemma 1,

$(\exists y)\, T(e,n,y)$.

Lemma 3. Suppose $(\forall n)(\exists y)\, T(e,n,y)$. Then

$(\forall n)\ P \vdash Con(P_n + \Phi(e))$.

Proof. Fix n. We show, in fact, P proves

$Con(P_n +$ "P_n is not Σ^0_1 sound" $+ \Phi(e))$.

Note that $P_n +$ "P_n is not Σ^0_1 sound" proves

P_i is Σ^0_1-sound iff $i < n$. So

$P_n +$ "P_n is not Σ^0_1 sound" proves

$\Phi(e) \longleftrightarrow (\forall i < n)\, Con(T_{e,i})$.

So it's enough to prove

$$P_n + \text{"}P_n \text{ is not } \Sigma^0_1 \text{ sound"} + (\forall i<n)[\mathrm{Con}(T_{e,i})].$$

consistent. By 2) of section 1, this is consistent iff

$$P_n + (\forall i<n)[\mathrm{Con}(T_{e,i})] \text{ is consistent iff}$$

$$(\forall i<n)\ \mathrm{Con}(T_{e,i}).$$

[The above argument can be formalized in P!)

But by Lemma 1, if $(\exists y)\ T(e,n,y)$,

$P \vdash \mathrm{Con}(T_{e,n})$. Apply this to each $m < n$ getting

$$P \vdash (\forall i<n)\ \mathrm{Con}(T_{e,i}).$$

This proves the lemma.

Has someone else noticed this? If not, I might write up a brief paper giving the proof.

Yours sincerely,

Bob Solovay

P.S. After July 1, my address will be

(over)

1) Until Sept. 1,
Robert Solovay
Dept. of Mathematics
University of California
Berkeley, CA. 94720

2) After Sept. 1 (Until July 1, 1977.)
Robert Solovay
Dept. of Mathematics
California Institute of Technology
Pasadena, California

P.P.S. The precise conditions on T are: 1) T is axiomitizible, 2) consistent, 3) P is intepreted in T, 4) T is strongly reflexive. The last condition means:

T proves $(\forall n \in \omega)\,[\Phi(n) \to \mathrm{Con}\,[\Phi(\underline{\underline{n}})]]$

(I.e. reflexinty for formulas with a free numerical variable.) Here I have formulated the condition for the case that P is a subtheory of T; the general case is essentially similar.)

The proof that if T satisfies 1) - 4), then the set of φ such that $T+\varphi$ is interpretable in T is Π^0_2 can be formalized in P.