Sonderdruck ISSN 0003-9268

Archiv für

MATHEMATISCHE LOGIK UND GRUNDLAGENFORSCHUNG

26/3–4

VERLAG W. KOHLHAMMER
STUTTGART · BERLIN · KÖLN · MAINZ

Arch. math. Logik 26 (1987), 169–176

A NOTE ON PROOFS OF FALSEHOOD *

Jan Krajíček

In these notes we shall sketch a few results motivated by Švejdar's question (see below). To make the paper self-contained we repeat, in Chapter 0, definitions of some notions and some results (without proofs) used in it. For more details the reader can consult e.g. [Pu].

Chapter 0

Let us begin with an informal recall of a few familiar notions. Q is *Robinson's arithmetic*, a *bounded formula* (or Δ_0) is a formula in the language of Q whose quantifiers are of the form $\exists x < t$ or $\forall x < t$ (t term not containing x), and *bounded arithmetic* $I\Delta_0$ is the extension of Q by instances of the induction scheme for all bounded formulas. *Exp* is the Π_2^0 formula in the language of Q saying "$\forall x \exists y : y = 2^x$".
A cut in a theory T is any formula $I(x)$, s.t. T proves the conjunction of the conditions: (i) $I(0)$, (ii) $I(x) \to I(s(x))$, (iii) $I(x) \& y < x \to I(y)$. We say that the cut is closed under the addition (resp. multiplication) iff T proves also: (iv) $I(x) \& I(y) \to I(x+y)$ (resp. $I(x) \& I(y) \to I(x \cdot y)$).
A theory T is *called sequential* iff it contains a reasonable fragment of a theory of finite sequences – for definitions and examples see [Pu]. Sequential theories and cuts in theories play an important role in the questions of interpretability (cf. [Pu]).
The last informal definition is: the *depth of a proof* is the maximal logical depth (not length) of a formula in it, where the logical depth is defined as usual.
We continue stating a few more formal definitions and recalling some results concerning the notions above. For proofs or details consult the cited papers.

Definition:

(1) a) $Con_y^{I(x)}(T) \leftrightarrows$ "there is no proof d of $0=1$ in T s.t. $I(d)$ and d has the depth $\leq y$"
b) $Con^I(T) \leftrightarrows \forall y;\ Con_y^I(T)$
c) $Con_y(T) \leftrightarrows Con_y^{x=x}(T)$
d) $Con(T) \leftrightarrows \forall y;\ Con_y(T)$

* Eingegangen am 16. 9. 1986, revidierte Fassung am 2. 2. 1987.

(2) a) $\mathrm{HCon}^I(T) \leftrightarrows$ "for any t, s.t. $I(t)$, t does *not* satisfy both (i) and (ii):
(i) t is a propositional tautology
(ii) t is a disjunction of Herbrand variants of a prenex normal form of a disjunction of negations of some axioms of T (cf. [Pu])"
b) $\mathrm{HCon}(T) \leftrightarrows \mathrm{HCon}^{x=x}(T)$
(3) $T \leqq S$ iff T is globally interpretable in S
(4) $T \overset{m}{\vdash} A$ iff formula A has a proof in T of depth $\leqq m$

Fact 0.0 (P. Pudlák): *For a reasonable sequential theory T and for any cut $I(x)$ in T:*
(i) $T \nvdash \mathrm{Con}^I(T)$
and even
(ii) *there exists $k<\omega$, $T \nvdash \mathrm{Con}^I_k(T)$.*

Fact 0.1 (P. Pudlák): *Let T be a finitely axiomatizable, sequential theory or $T = I\Delta_0$. Then there exists a cut $H(x)$ in T s.t.: $T \vdash \mathrm{HCon}^H(T)$.*

Fact 0.2 (P. Pudlák): *For T a finitely axiomatizable, sequential theory or for $T = I\Delta_0$: not $Q + \{\mathrm{Con}_k(T) \mid k<\omega\} \leqq T$.*

Fact 0.3 (A. Wilkie): *Let M be a countable model of $I\Delta_0 + \mathrm{Exp}$, $I(x)$ any formula of depth $\leqq k$ which is a cut in $I\Delta_0$ and have terms of depth $\leqq 1$. Let $a, b \in M$ be two nonstandard elements of M s.t.: $M \models 2^a_{(2\cdot k+3)} \leqq b$. Then there exists an initial substructure $M' \subseteq_e M$ s.t.:*
(i) $M' \models I\Delta_0$
(ii) $a \in M'$, $b \notin M'$
(iii) $M' \models I(a)$.
(2^y_x is the function defined: $2^y_0 = y$ and $2^y_{x+1} = 2^{(2^y_x)}$).

The results of this note are inspired by Švejdar's question: "When is it consistent for inconsistency-proofs to lie *between* cuts?" (the question is inspired by 0.0(i)). More precisely; for which T, $I(x)$ and $J(x)$ cuts in T is the theory "$T + \mathrm{Con}^J(T) + \neg \mathrm{Con}^I(T)$" consistent?
The arguments are sometimes only sketched and the paper should be considered as a preliminary report.
I am indebted for discussions to P. Pudlák, V. Švejdar, J. Paris, and A. Wilkie. I thank to a referee for important suggestions.

Chapter 1

For making results and arguments more readable let us extend the languages of the theories under consideration by constants $e, e_0, e_1, \ldots$ with the meaning:
(i) e is the least proof of $0=1$ in T or $e=0$ if $\mathrm{Con}(T)$
(ii) e_k is the least proof of $0=1$ in T of depth $\leqq k$ or $e_k = 0$ if $\mathrm{Con}_k(T)$.

Proposition 1.0: *For any cut $I(x)$ in $I\Delta_0$ there exists natural k s.t.*

$$I\Delta_0 \nvdash I(e) \rightarrow I(2_k^e).$$

Proof: Let $M \models I\Delta_0 + \mathrm{Exp} + \neg\mathrm{Con}(I\Delta_0)$ be countable (it is known that $I\Delta_0 + \mathrm{Exp} \nvdash \mathrm{Con}(I\Delta_0)$ – see [P-W]).
By 0.3 there exists $M' \subseteqq_e M$ s.t.:
(i) $e \in M'$ but $2_k^e \notin M'$,
(ii) $M' \models I(e)$,
(iii) $M' \models I\Delta_0$.
It is enough to choose $k = 2 \cdot$"the depth of cut I when written using terms of the depth $\leqq 1$"$+3$. We have done. q.e.d.

Proposition 1.1. *For any cut $I(x)$ in theory T, where T is a finitely axiomatizable, sequential Π_1^0 theory or $T = I\Delta_0$, there exist natural k, m s.t.:*

$$T \nvdash I(e_k) \rightarrow I(2_m^{(e_k)}).$$

Proof. According to 0.1 we can define a cut $H_0(x)$ in T closed under multiplication s.t.:

$$T \vdash \mathrm{HCon}^{H_0}(T)$$

Define a cut $H(x)$:

$$H(x) \equiv [(\mathrm{HCon}(T) \wedge x = x) \vee (\neg\,\mathrm{HCon}(T) \wedge H_0(x))]$$

Clearly:

$$T \vdash \mathrm{HCon}^H(T).$$

Assume that for all $k < \omega$ the theory "$T + I \subsetneqq H + \neg\,\mathrm{Con}_k^I(T)$" is inconsistent, i.e.

(1) $$T + I \subsetneqq H \vdash \mathrm{Con}_k^I(T), \quad k < \omega.$$

Since evidently:

(2) $$T + I \subsetneqq H \leqq T$$

(relativize to the cut $H(x)$; here $T \vdash H^H = H$ is needed, but this is easily verifiable) also:

(3) $$T + \{\mathrm{Con}_k^I(T) \mid k < \omega\} \leqq T.$$

Now:

(4) $$Q + \{\mathrm{Con}_k(T) \mid k < \omega\} \leqq T + \{\mathrm{Con}_k^I(T) \mid k < \omega\}$$

(relativize to $I(x)$ or to a suitable shortening of $I(x)$ if it is not closed under multiplication).
From (3) and (4):

$$Q + \{\mathrm{Con}_k(T) \mid k < \omega\} \leqq T,$$

which is a contradiction with 0.2.

Hence, since (1) is false there exists $k<\omega$ s.t.:

$$T+I\subseteq H+\neg\,\mathrm{Con}_k^I(T)$$

is consistent.
From Herbrand's theorem it is known how to transform a proof d of $0=1$ of depth $\leqq k$ into a Herbrand's disjunction of size $\leqq 2_m^d$, m polynomially depending on k, contradicting a given theory (in the sense of Definition (2)).
If it were true that:

$$T\vdash I(e_k)\rightarrow I(2_m^{(e_k)}),$$

we would also have:

(5) "$T+e_k\neq 0+H(2_m^{(e_k)})$" is consistent.

But this is in a contradiction with the choice of the cut $H(x)$. q.e.d.

Remark: When speaking about T as a Π_1^0-theory we implicitly assume that T is in the language of arithmetic. The assumption that T is Π_1^0 implies that in models of T the axioms of T are also satisfied in all initial segments closed under multiplication. An example: any theory of the form $(Q+A)$, A a true Π_1^0-sentence, is a finitely axiomatizable, sequential Π_1^0-theory.

Chapter 2

Now we shall use the preceeding chapter to obtain some results related to Švejdar's question.

Theorem 2.0: *For any cut $I(x)$ in $I\Delta_0$ there exists $k<\omega$ s.t. for any cut $J(x)$ in $I\Delta_0$ satisfying:*

$$I\Delta_0\vdash J(x)\rightarrow I(2_k^x)$$

the theory

$$I\Delta_0+\mathrm{Con}^J(I\Delta_0)+\neg\,\mathrm{Con}^I(I\Delta_0)$$

is consistent.

Proof: Let k be the natural number assigned to $I(x)$ by Proposition 1.0. Assume that "$I\Delta_0+\mathrm{Con}^J(I\Delta_0)+\neg\,\mathrm{Con}^I(I\Delta_0)$" is inconsistent, i.e.:

$$I\Delta_0\vdash I(e)\rightarrow J(e)$$

By the hypothesis of the theorem

$$I\Delta_0\vdash J(e)\rightarrow I(2_k^e)$$

and thus:

$$I\Delta_0 \vdash I(e) \rightarrow I(2_k^e)$$

This contradicts the choice of k. q.e.d.

Theorem 2.1: *For any cut $I(x)$ in T, where T is finitely axiomatizable, sequential Π_1^0-theory or $T = I\Delta_0$, there exist $k, m < \omega$ s.t. if a cut $J(x)$ in T satisfies:*

$$T \vdash J(x) \rightarrow I(2_m^x)$$

then the theory.

$$T + \mathrm{Con}_k^J(T) + \neg\, \mathrm{Con}_k^I(T)$$

is consistent.

Proof: Similar to the proof of 2.0 using 1.1 instead of 1.0. q.e.d.

The following observation complements the preceeding two results.

Proposition 2.2.: *For any cut $I(x)$ in a finitely axiomatizable, sequential Π_1^0-theory T or in $T = I\Delta_0$ there exists a cut $J(x)$ in T s.t.:*
(i) $T \vdash J(x) \rightarrow I(x)$
(ii) $T \nvdash I(x) \rightarrow J(x)$
but
(iii) $T + \mathrm{Con}^J(T) + \neg\, \mathrm{Con}^I(T)$ *is not consistent (and analogously when using* Con_k*'s instead of* Con*).*

Proof: With $I(x)$ given define:

$$J(x) \leftrightarrows (I(x) \,\&\, (\mathrm{Con}(T) \rightarrow (\forall y; I(y) \rightarrow I(y + 2^x))))$$

Evidently:
(i) $T \vdash$ "J is a cut"
(ii) $T \vdash J(x) \rightarrow I(x)$
(iii) $T \vdash \neg\, \mathrm{Con}^I(T) \rightarrow \neg\, \mathrm{Con}^J(T)$, since
$T \vdash \neg\, \mathrm{Con}^I(T) \rightarrow I = J$.
It remains to show (ii) of the proposition. Assume the contrary

(1) $$T \vdash I(x) \rightarrow J(x).$$

Then:

(2) $$T + \mathrm{Con}(T) \vdash I(x) \rightarrow I(2^x).$$

According to [P-D] there exists a model $M \models \mathrm{PA} + \mathrm{Con}\, T$ and an initial substructure $P \subsetneqq_e M$ s.t. in the structure $\langle M, P \rangle$ there is no definable a cut in P closed under 2^x. Choose such $\langle M, P \rangle$. Since $I\Delta_0 \subsetneqq \mathrm{PA}$ and $\mathrm{PA} \vdash \mathrm{Con}(T) \rightarrow T$ for T a Π_1^0 finite theory we have: $P \models T + \mathrm{Con}\, T$.
By (2) then: $P \models I(x) \rightarrow I(2^x)$. A contradiction. q.e.d.

Chapter 3

The aim of this chapter is to prove the following result.

Theorem 3.0: *For $T=I\Delta_0$ or T a finitely axiomatizable, sequential theory there exists an assignment of natural numbers to cuts in T, say $I\mapsto k_I$, s.t. the theory:*

$$T+\{\neg \mathrm{Con}^I_{k_I} \mid I \text{ cut in } T\}$$

is consistent.

Observe that since for all $k<\omega$ there is a cut I in T s.t. $T\vdash \mathrm{Con}^I_k(T)$ the theorem is equivalent with the proposition: "there is a model $M\models T$ s.t. e_i's are cofinal in all definable cuts (i.e. for each I a cut in T there exists e_i s.t. $M\models(I(e_i)\ \&\ e_i\neq 0)$ and $e\neq 0$ lies in all these cuts."
We shall need the following result proved in [F] (it can be also obtained from [Pu]).

Fact 3.1 (H. Friedman): *Let S, T be finitely axiomatizable, sequential theories. Then the following are equivalent:*
(i) $S\leqq T$
(ii) $I\Delta_0+\mathrm{Exp}\vdash \mathrm{HCon}(T)\rightarrow \mathrm{HCon}(S)$.
(This formulation is closer to [Pu]*).*

Another result we shall use is the "effective" version of 0.0(ii), namely:

Fact 3.2 (P. Pudlák): *Let T be a finitely axiomatizable, sequential theory. Then for any cut $I(x)$ in T there exists $k<\omega$ s.t.:*

$$I\Delta_0+\mathrm{Exp}\vdash \mathrm{HCon}(T)\rightarrow \mathrm{HCon}(T+\neg \mathrm{Con}^I_k(T))$$

(3.2 is obtained by an inspection of the proof of 0.0(ii)*).*

Finally we shall need:

Fact 3.3 (P. Pudlák): $I\Delta_0+\mathrm{Exp}$ *proves: "Let S, T be sequential theories, $S\leqq T$ and i be some interpretation of S in T. Then there exist cuts $I(x)$ in T and $J(x)$ in S s.t. there is a (definable) isomorphisms between $\langle I(x), +, \cdot\rangle$ and $\langle J(x)^i, +^i, \cdot^i\rangle$ (the structures definable in T)".*
(3.3 is proved in [Pu], *while its* $I\Delta_0+\mathrm{Exp}$*-provability is easily verifiable).*

Proof of the Theorem 3.0: Let $J_0, J_1,$ enumerate all cuts in T and define cuts I_0, $I_1,\ldots$:

$$I_n(X)\leftrightarrows J_0(x)\ \&\quad \&\ J_n(x)$$

So we have:
(i) $T \vdash I_0 \supseteq I_1 \supseteq \dots$ and
(ii) for any cut J in T there is $k < \omega$ s.t.

$$T \vdash J \supseteq I_k$$

T is from now on assumed to be finite.

Step 1: By 3.2 there exists $k_0 < \omega$ s.t.

$$I\Delta_0 + \mathrm{Exp} \vdash \mathrm{HCon}(T) \rightarrow \mathrm{HCon}(T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T)).$$

Hence by 3.1:

$$T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) \leqq T$$

(and this is provable in $I\Delta_0 + \mathrm{Exp}$ since it is true Σ_1^0-sentence) and let i_0 be the interpretation.

Step 2: Take a cut $I_1(x)$. We claim that there is $k_1 < \omega$ s.t.

$$I\Delta_0 + \mathrm{Exp} \vdash \mathrm{HCon}(T) \rightarrow \mathrm{HCon}(T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) + \neg\, \mathrm{Con}_{k_1}^{I_1}(T)).$$

Firstly argue informally: assume that we choose k_1 sufficiently big and that:

$$\neg\, \mathrm{HCon}(T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) + \neg\, \mathrm{Con}_{k_1}^{I_1}(T)).$$

Then there is $m < \omega$ s.t.:

$$T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) \mid\!\!\overset{m}{-}\, \mathrm{Con}_{k_1}^{I_1}(T).$$

Using the interpretation i_0 we can construct, by 3.3, a cut $I_1'(x)$ in T and m' s.t.:

$$T \mid\!\!\overset{m'}{-}\, \mathrm{Con}_{k_1}^{I_1'}(T).$$

(roughly speaking: I_1' is the image of I_1 in the interpretation i_0 intersected with the initial part of numbers common to the universe and to the universe of the interpretation), i.e.:

$$\neg\, \mathrm{Con}_{m''}(T + \neg\, \mathrm{Con}_{k_1}^{I_1'}(T)), \quad \text{for some } m'' \geqq m$$

From this it follows that:

$$\neg\, \mathrm{HCon}(T + \neg\, \mathrm{Con}_{k_1}^{I_1'}(T)).$$

If we choose k_1 sufficiently large w.r.t. $I_1'(x)$ we have, by 3.2.:

$$\neg\, \mathrm{HCon}(T).$$

Hence we proved:

$$\mathrm{HCon}(T) \rightarrow \mathrm{HCon}(T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) + \neg\, \mathrm{Con}_{k_1}^{I_1}(T)).$$

Now, the whole argument of Step 2 can be formalized and proved in $I\Delta_0 + \mathrm{Exp}$. For this we need only 3.2, 3.3 and the observations:

(1) for all m greater then the depth of A:

$$I\Delta_0 + \mathrm{Exp} \vdash \mathrm{HCon}(A) \equiv \mathrm{Con}_m(A).$$

(2) if $S \leqq T$ is true and i is the interpretation then

$$I\Delta_0 + \mathrm{Exp} \vdash \text{“}S \leqq T \text{ and } i \text{ is the interpretation”}.$$

So finally we have:

$$I\Delta_0 + \mathrm{Exp} \vdash \mathrm{HCon}(T) \to \mathrm{HCon}(T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) + \neg\, \mathrm{Con}_{k_1}^{I_1}(T))$$

and, by 3.1.:

$$T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) + \neg\, \mathrm{Con}_{k_1}^{I_1}(T) \leqq T.$$

In the same manner we prove, in the Step$(n+1)$, the formula:

$$T + \neg\, \mathrm{Con}_{k_0}^{I_0}(T) + \ldots + \neg\, \mathrm{Con}_{k_n}^{I_n}(T) \leqq T$$

(k_n's being constructed through the proof). By compactness we obtain consistency of the theory $T + \{\neg\, \mathrm{Con}_{k_j}^{I_j}(T) \mid j < \omega\}$ and we have done.
The result for $T = I\Delta_0$ is derived as follows. There is a finite, sequential $S \subseteqq I\Delta_0$ and a cut $J(x)$ in S s.t. $S \vdash (I\Delta_0)^J$ (see [Pu]). Let M be a model of the theory $S + \{\neg\, \mathrm{Con}_{k_j}^{I_j}(S) \mid j < \omega\}$ assured above. If we define the initial segment K of M:

$$K = \{m \in M \mid M \models J(m)\},$$

then clearly $K \models I\Delta_0$.
Moreover, if $I(x)$ is a cut in $I\Delta_0$ then for some $m < \omega$: $K \models \neg\, \mathrm{Con}_m^I(I\Delta_0)$. This is because for some $I_j(x)$ a cut in S, $S \vdash I_j \subseteqq I(x)^J$, i.e.

$$\{m \in M \mid M \models I_j(m)\} \subseteqq \{m \in K \mid K \models I(m)\}$$

and any proof in S (in particular of $0 = 1$ – we have $M \models \neg\, \mathrm{Con}_{k_j}^{I_j}(S)$) is a proof in $I\Delta_0$, too. We have done. q.e.d.

REFERENCES

[F] Friedman, H.: Translatability and relative consistency. II. Manuscript, 1980.
[P-D] Paris, J., Dimitracopoulos, C.: A note on the undefinability of cuts. J. Symb. Logic **48**, 564–569 (1983).
[P-W] Paris, J., Wilkie, A.: Δ_0-sets and induction. Ann. Pure Appl. Logic (to appear).
[Pu] Pudlák, P.: Cuts, consistency statements and interpretability. J. Symb. Logic **50**, 423–441 (1985).
[W] Wilkie, A.: On sentences interpretable in systems of arithmetic. In: Logic Colloquium '84, pp. 329–342. Paris, J.B., Wilkie, A.J., Wilmers, G.M. (eds.). Amsterdam: North-Holland 1986.

Jan Krajíček
Math. Inst. Acad. Sci.
Žitná 25,
CS-11567 Praha 1
Czechoslovakia