CONSERVATION
by
Harvey M. Friedman
September 24, 1999

John Burgess has specifically asked about whether one give a finitistic model theoretic proof of certain conservative extension results discussed in [Si99]. Burgess asks this in connection with Hilbert's program. Specifically,

I. $WKL_0$ is a conservative extension of PRA for $\Pi$-0-2 sentences.
II. $ACA_0$ is a conservative extension of PA for arithmetic sentences.
III. $ATR_0$ is a conservative extension of IR for arithmetic sentences.
IV. $\Pi$-1-1-$CA_0$ is a conservative extension of ID(<omega) for arithmetic sentences.

Here IR is Feferman's IR, which can be taken to be the theory extending PA with new function symbols encoding the Kleene H-sets on each specific initial segment of the ordinal notation system Gamma0.

All of these conservative extension results are given model theoretic proofs in [Si99] except for III. The reader is referred to the exposition of my proof in Friedman, MacAloon, Simpson, "A finite combinatorial principle which is equivalent to the 1-consistency of predicative analysis", 1982, 197-230 in: G. Metakides (ed.), Patras Logic Symposion, Studies in Logic and the Foundations of mathematics, North-Holland, 1982.

We write PFA (polynomial function arithmetic) for the system in the language of 0,1,+,x,<, with the usual successor axioms and defining equations, together with induction for all bounded formulas. This is the same as what is called $I\Delta_0$ in the book {HP93] and elsewhere.

In [HP93], it is proved that PFA is fully capable of developing finite sequence coding and formalizing syntax. In fact, they devote all of Chapter V, section 3, to this topic, which is entitled "Exponentiation, Coding Sequences and Formalization of Syntax in $I\Delta_0$." In the Bibliographic Notes on page 406, they write "A formalization of syntax in $I\Delta_0$ is considered here for the first time, though the ideas on which it is based have been around for some time."

This makes PFA a good vehicle for taking a reverse mathematics point of view towards weak fragments of arithmetic.

From this point of view, it is natural to take EFA = exponential function arithmetic to be in the language of PFA, whose axioms are PFA plus the single axiom (forall n)($2^n$ exists). And we take EFA* to be PFA plus the axiom that asserts that for all n, there is a sequence of integers of length n, starting with 0, where each term is the base 2 exponentiation of the preceding term.

Here is what we will do in this note.

1. We isolate a crucial general fact about theories, which we call the Key Lemma. The Key Lemma has an easy model theoretic proof. But the Key Lemma also has a proof using the Criag interpolation theorem for predicate calculus with equality. The interpolation theorem has model theoretic proofs, but it also has a proof theoretic proof using Gentzen's cut elimination theorem for predicate calculus with equality. The usual proof of the cut elimination theorem with iterated exponential estimates (given by Gentzen) is readily formalized in EFA'. This yields a proof of the interpolation theorem and of the Key Lemma that is readily formalized in EFA'.

2. Using the Key Lemma, we give a purely model theoretic proof the conservation results in question. In fact, we adapt the usual model theoretic proofs of these conservation results to provide formal interpretations of the appropriate kind which establish the conservation results. The Key Lemma is crucial.

3. We round out the situation by claiming that EFA* is best possible in the following sense. That each of the conservation results are provably equivalent to EFA* over PFA.

**********

## 1. THE KEY LEMMA

THEOREM 1.1. Let T be a theory in predicate calculus whose language includes <,=,c, and which contains the axioms of linear order. Suppose that for each n ≥ 0, T is consistent with "there are at least n elements < c." Let I be a new monadic predicate symbol. Then T + "I is a cut below c" is consistent.

Proof: The model theoretic proof is straightforward. Consider T′ = T together with the axioms $c > c_1 > c_2 > \ldots$, where the $c_i$ are new constant symbols. Then T′ is consistent. Let M be a model of T′. There is obviously a cut below c in M. Set I to be a cut below c in M. Then (M,I) satisfies T + "I is a cut below c." QED

Let W be the following theory in the language {<,=,0,S}:

a) axioms of linear order;
b) 0 is the least element;
c) Sx is the successor of x;
d) every nonzero element is the successor of some element.

There is a well known elimination of quantifiers for W. From this elimination of quantifiers, one can also show that every formula □(x) in one free variable is provably equivalent over W to a Boolean combination of inequalities of the form x < S...S0.

KEY LEMMA. Let T be a finitely axiomatized theory in predicate calculus whose language includes <,=,0,c, and which contains the axioms of W. Suppose that for each n ≥ 0, T is consistent with "there are at least n elements < c." Let I be a new monadic predicate symbol. Then T + "I is a cut below c" is consistent.

Proof: This is an immediate consequence of Theorem 1.1. But now we wish to give an alternative proof using the Craig interpolation theorem for predicate calculus with equality, and quantifier elimination for W.

Fix T,I to be as in the hypothesis of the Key Lemma, and assume that T proves "I is not a cut below c." By the Craig interpolation theorem, let □ be a sentence in the common language of T and "I is not a cut below c" such that T proves □ and □ proves "I is not a cut below c."

Hence □ is in the language {<,=,c} and has the following properties:

i) for all $n \geq 0$, □ is consistent with W + "there are at least n elements < c."

ii) □ proves "I is not a cut below c."

By cut elimination for W, there exists $n \geq 0$ such that

□ if and only if $c \geq S^n(0)$

is provable in W.

Let (K,<) be the linear ordering omega + Z, I = omega, and c lies in Z. Then (K,<,I,c) satisfies W and therefore satisfies □. But (K,<,I,c) also satisfies "I is a cut below c." This contradicts that □ implies "I is not a cut below c." QED

THEOREM 1.2. The Key Lemma is provable in EFA'.

Proof: We indicate the essential points. First of all, we have used the interpolation theorem. One of the well known proofs goes as follows. Let □ arrows psi be provable. Then □ arrows psi has a cut free proof. And then an interpolant is constructed by recursion on the cut free proof. There are no blowups in size after one obtains the cut free proof.

The original proof by Gentzen of his cut elimination theorem coontained iterated exponential estimates, and so is readily formalized in EFA'.

The cut elimination theorem for W is readily formalizable in EFA, which is better than we need.

In the final steps, we considered the structure (K,<,I,c), where c lies in Z. We used that satisfaction implies consistency, which in general requires induction with respect to a truth predicate, and so could be a problem in a system like EFA' with its limited induction. However, because of the simplicity of the cut elimination procedure, the truth predicate for this structure can be built up by simple recursions. In fact, this can be carried out in EFA. Then the induction needed for the final step is easily available in EFA. QED

More delicate arguments will also establish more general versions of the Key Lemma in EFA’. In particular, with considerable effort, it can be proved in EFA’ with W replaced by the axioms for linear ordering (and with 0 removed), as in Theroem 1.1. But the Key Lemma is precisely what we need.

## 2. $WKL_0$ OVER PRA

We use the following version of the usual Ackerman hierarchy of functions from Z+ into Z+.

$A_1(n) = 2n$. $A_{i+1}$ is the indefinite iteration of $A_i$.

I.e., $A_{i+1}(n)$ is the result of applying $A_i$ n times starting at 1.

We need to consider a formalization of this hierarchy within PRA, or even within EFA. We have to take into account the limited language of PRA, and that no matter how this is formalized in PRA, we cannot prove that the A’s are total.

Let f(n) be the obvious algorithm for computing the n-th function in the usual Ackerman hierarchy of functions using stacks. Specifically, suppose the algorithm f(n) has been defined. The algorithm f(n+1) computes at m by starting with 1, and applying the algorithm f(n) m times.

Clearly f is a low level computable function. In PRA we cannot prove that every f(n) is total; i.e., halts at all arguments. However, we can obviously prove in PRA that

i) if f(n+1) is total and $m < n$ then f(n) is total;
ii) if f(n+1) is total then f(n+1) is the indefinite iteration of f(n) as in the usual Ackerman hierarchy of functions.

Here we have identified f(n) with the partial function that it computes, which is a harmless abuse of terminology. Thus we will write f(n)(m) for the output of the algorithm f(n) at the input m, which may not be defined.

We now let □ be a □-0-2 sentence that is consistent with PRA. We want to show that □ is consistent with $WKL_0$. Let □ = (□p)(□q)(R(p,q)). Fix p such that (□q)(R(p,q)).

We let PRA′ be PRA + □ + “f(c) is total.” Since PRA proves f(n) is total for each particular n, we see that the Key Lemma applies to PRA′. Hence the system PRA′′ = PRA + □ + “f(c) is total” + “I is a cut below c” is consistent.

We now build a sequence of closed intervals [xi,yi], 0 □ i □ c/4. For each such i, $y_i = f_{c-2i}$(xi). We start with $[x_0,y_0]$ = $[p,f_c(p)]$.

Suppose $[x_i,y_i]$ has been defined, $y_i = f_{c-4i}$(xi). Look at all Turing machines with index □ $f_{c-4i-2}(x_i)$ which produce an output at 0 that lies in $[x_i,y_i]$. Let E be the set of these at most $f_{c-4i-2}(x_i)$ values. By elementary combinatorial considerations, E must be disjoint from some subinterval $[x_{i+1},y_{i+1}]$, where $y_{i+1} = f_{c-2i-4}(x_{i+1})$, and $x_{i+1} > 2^{xi}$. Take this to be $[x_{i+1},y_{i+1}]$. The idea is that this choice of the next interval is to handle all appropriate instances of □-0-1 bounding whose front universal quantifier ranges over $[0,x_i]$, and whose parameters are subsets of $[0,y_{i+1}]$.

Note that the left endpoints are strictly increasing and lie at or above p. We now let J be the cut determined by looking at the left endpoints whose index i lies in the cut I.

We now define an interpretation of $WKL_0$ in PRA′′. The integers are taken to be the elements of J. And 0,<,+,x are are usual. From the perspective of PRA′′, take the sets to be the intersections of finite sets with the cut J. Since J cannot be a finite set, we have the usual overspill situation which allows us to prove the interpretation of weak Konig’s lemma and delta-0-1 comprehension in PRA′′. Also, the interpretation of □ is obviously provable in PRA′′. Every instance of □-0-1 induction becomes provable in PRA′′ under this interpretation; this has to be checked with set parameters. Hence $WKL_0$ + □ is consistent since PRA′′ is consistent.

THEROEM 2.1. EFA′ proves that $WKL_0$ is conservative over PRA for □-0-2 sentences.

3. $ACA_0$ OVER PA

Let □ be an arithmetic sentence that is consistent with PA. Let R be a binary relation symbol. Let T′ be the following theory in the language of PA together with c,R:

1) PA + □;
2) for all i <= c, R(i,n) codes the i-th Turing jump.
3) induction for all formulas in the language.

Obviously, for all n, T' + c > n is consistent. We can therefore apply the Key Lemma. Let T'' = T' + "I is a cut below c." Then T'' is consistent.

We now define a translation of $ACA_0$ into T''. We take the arithmetic part to be standard from the perspective of T''. We take the sets of integers to be the sets of integers coded by the unary predicates R(i,n) for i lying in I. We use 3) to verify that the interpretation of the induction axiom of $ACA_0$ is provable in T''. And since T'' proves that I is a cut below c, we see that the interpretation of each instance of arithmetic comprehension is provable in T''.

THEOREM 3.1. EFA' proves that $ACA_0$ is conservative over PA for arithmetic sentences.

## 4. $ATR_0$ OVER IR

Left to the reader. Of course, we are now so high up that there is absolutely no trouble formalizing the usual model theoretic proof of the relevant conservativity within ATR0.

## 5. □-1-1-$CA_0$ OVER ID(<□)

Left to the reader. Of course, we are now so high up that there is absolutely no trouble formalizing the usual model theoretic proof of the relevant conservativity within ID(<□).