On the Relationship Between ATR_0 and $\widehat{ID}_{<\omega}$ *

Jeremy Avigad

April 2, 1996

Abstract

We show that the theory ATR_0 is equivalent to a second-order generalization of the theory $\widehat{ID}_{<\omega}$. As a result, ATR_0 is conservative over $\widehat{ID}_{<\omega}$ for arithmetic sentences, though proofs in ATR_0 can be much shorter than their $\widehat{ID}_{<\omega}$ counterparts.

1 Introduction

Let Γ_0 denote the least impredicative ordinal, as defined in [9] or [12]. Work of Feferman and Schütte in the sixties demonstrated that this is the proof-theoretic ordinal corresponding to theories embodying "predicative mathematics." In more recent years a number of classical theories without direct predicative justification have been discovered, whose proof-theoretic strength is also Γ_0. The aim of this paper is to clarify the relationship between two such theories, namely $\widehat{ID}_{<\omega}$ and Friedman's ATR_0.

The $\widehat{ID}_n$ are theories in the language of Peano Arithmetic augmented by new constants representing fixed points of arithmetic formulas involving positive occurences of a unary predicate. Each theory $\widehat{ID}_n$ allows n iterations of this inductive definition schema, while the theory $\widehat{ID}_{<\omega}$ allows arbitrarily many. Feferman [2] shows that the proof-theoretic ordinal of each $\widehat{ID}_n$ is γ_n, where the γ_n form a canonical fundamental sequence for Γ_0. As a result, the proof-theoretic ordinal of $\widehat{ID}_{<\omega}$ is Γ_0.

On the other hand, there are two methods currently in the literature for showing that the proof-ordinal of ATR_0 is Γ_0: a model-theoretic argument appears in [3], and a proof-theoretic argument which involves embedding ATR_0 into a fragment of set theory and carrying out a series of cut-eliminations is described in [6].

In Section 3 we show that ATR_0 is in fact a "limit" of the theories $\widehat{ID}_n$, in the sense that its key axiom is equivalent to a second-order schema asserting the existence of fixed points of positive arithmetic formulas. In Section 4 we use a

*This paper comprises a part of the author's Ph.D. dissertation [1].

cut-elimination argument to show that this implies that ATR_0 is in fact a conservative extension of $\widehat{ID}_{<\omega}$ for arithmetic sentences. (A model-theoretic proof is also possible.) This result, combined with the analysis in [2], represents what is perhaps the most direct determination of ATR_0's proof-theoretic strength.

Because the conservation argument mentioned above involves a cut-elimination, it allows for the possibility that short ATR_0 proofs may translate to proofs requiring superexponentially many fixed points. In Section 5 we show that this possibility is unavoidable by showing that ATR_0 has short proofs of the consistency of $\widehat{ID}_n$ for superexponentially large n. (Perhaps this fact may explain some of the difficulties encountered in analyzing ATR_0.)

2 Preliminaries

In what follows we'll assume that some method of coding ordered pairs of natural numbers has been chosen, and we'll use $\langle x, y\rangle$ to represent the code of the pair consisting of x and y. If $z = \langle x, y\rangle$ we'll write $z_0 = x$ and $z_1 = y$.

The theory $\widehat{ID}_1$ is a first-order theory in the language of Peano Arithmetic (PA), with an additional predicate P_φ for each arithmetic formula $\varphi(x, X)$ in which the unary predicate X occurs only positively. We'll sometimes refer to such an arithmetic formula as a "positive arithmetic operator" since it defines the monotone function

$$\Gamma_\varphi : P(\omega) \to P(\omega)$$

given by

$$\Gamma_\varphi(X) = \{x|\varphi(x, X)\}.$$

(The monotonicity means that for any sets A and B, $A \supset B$ implies $\Gamma_\varphi(A) \supset \Gamma_\varphi(B)$.) The axioms of $\widehat{ID}_1$ consist of the axioms of PA with induction extended to formulas involving the new constants, together with the fixed point axioms

$$\forall x(P_\varphi(x) \leftrightarrow \varphi(x, P_\varphi)).$$

In other words, these axioms assert that P_φ represents a fixed point of the operator Γ_φ, though not necessarily the least one. Similarly each theory $\widehat{ID}_{n+1}$ adds new constants for positive arithmetic formulas in the language of $\widehat{ID}_n$, and the corresponding fixed point axioms. $\widehat{ID}_{<\omega}$ is the union of the theories $\widehat{ID}_n$. See [2] for more details. [1]

ATR_0 is the second-order theory consisting of the weak base theory RCA_0 together with a schema (ATR) allowing for definitions by arithmetic transfinite recursion along any well ordering. RCA_0 consists of the basic quantifier-free

[1] Note that the presentation in [2] adds only one new predicate at each stage, so that each $\widehat{ID}_n$ has only n new predicates. This difference is inessential, since in any proof in our version of $\widehat{ID}_n$ one can "collapse" all the fixed point predicates of each iterative depth to a single one.

axioms of PA; induction for Σ^0_1 formulas, possibly involving set parameters; and a recursive comprehension schema

$$(RCA)\quad \forall x(\varphi(x) \leftrightarrow \psi(x)) \rightarrow \exists X \forall x(x \in X \leftrightarrow \varphi(x)),$$

where φ and ψ are Σ^0_1 and Π^0_1 respectively.

To describe (ATR) we need some definitions. In order to code countable sequences of sets as a single set, we'll use the notation Y_b to denote

$$\{x | \langle b, x\rangle \in Y\}.$$

In other words, Y_b codes the b^{th} set in the sequence. Given a binary relationship $\prec$ we'll use Y^b to denote

$$\{\langle a, x\rangle \in Y | a \prec b\},$$

i.e. Y^b codes $\bigoplus_{a \prec b} Y_a$. Finally, we use the abbreviation $WO(\prec)$ to represent the Π^1_1 assertion that the set $\prec$ codes a well-ordering. With these definitions in place, we can write the (ATR) schema as

$$(ATR)\quad WO(\prec) \rightarrow \exists Y \forall b, x(x \in Y_b \leftrightarrow \varphi(x, Y^b))$$

where φ ranges over arithmetic formulas (again, possibly involving set parameters). In words, (ATR) asserts we can build a hierarchy of sets by iterating an arithmetic comprehension along any well-ordering.

The choice of RCA_0 as the base theory is not crucial. Since over RCA_0 (ATR) easily proves arithmetic comprehension, we could equivalently have taken ACA_0 (the stronger system based on arithmetic comprehension) as our base theory instead.

In Section 5 we'll use the fact that ATR_0 proves the Σ^1_1 choice schema,

$$(\Sigma^1_1\text{-}AC)\quad \forall x \exists Y \varphi(x, Y) \rightarrow \exists Y \forall x \varphi(x, Y_x),$$

where φ is Σ^1_1. For more information on ATR_0 and its capabilities, see [15, 16, 17, 3, 18].

We'll use the abbrevation (FP) to represent the second-order axiom schema asserting the existence of arbitrary fixed points of positive arithmetic operators,

$$(FP)\quad \exists Y \forall x(x \in Y \leftrightarrow \varphi(x, Y)),$$

where $\varphi(x, Y)$ is an arithmetic formula in which Y occurs only positively. Note that once again, φ can have number and set parameters. Note also that if Y does not appear in φ, (FP) simply reduces to arithmetic comprehension. Let FP_0 represent the system consisting of (FP), the basic quantifier-free axioms of PA, and Σ^0_1 induction.

In the system $\widehat{ID}_{<\omega}$ neither the predicates $P_\varphi(x)$ nor the formulas $\varphi(x, X)$ are allowed to have parameters other than the ones shown. However, if $\varphi(x, X, \vec{y})$ has parameters $\vec{y}$, we can define the formula

$$\varphi'(\langle x, \vec{y}\rangle, X) \equiv \varphi(x, \lambda x.X(\langle x, \vec{y}\rangle), \vec{y}),$$

where atomic formulas $(\lambda x.X(\langle x,\vec{y}\rangle))(t)$ are interpreted as $X(\langle t,\vec{y}\rangle)$. Then we have that

$$\begin{aligned} P_{\varphi'}(\langle x,\vec{y}\rangle) &\leftrightarrow \varphi'(\langle x,\vec{y}\rangle, P_{\varphi'}) \\ &\leftrightarrow \varphi(x, \lambda x.P_{\varphi'}(\langle x,\vec{y}\rangle), \vec{y}) \end{aligned}$$

In Section 4 it will be convenient to allow the fixed-point predicates of $\widehat{ID}_{<\omega}$ to have parameters, taking the axioms of $\widehat{ID}_{<\omega}$ to be of the form

$$P_\varphi(x,\vec{y}) \leftrightarrow \varphi(x, \lambda x.P_\varphi(x,\vec{y}), \vec{y}). \tag{1}$$

By the above considerations, these axioms and predicates can be replaced by the parameter-free ones via coding of pairs and sequences.

3 The Equivalence of (ATR) and (FP)

The goal of this section is to prove the following

Theorem 3.1 *ATR_0 and FP_0 are equivalent.*

Proof. First, reasoning in FP_0, we'll use (FP) to derive (ATR). Let $\prec$ be a well-ordering and let $\varphi(n,X)$ be an arithmetic formula. We need to show the existence of a set Y such that

$$\forall b, n(n \in Y_b \leftrightarrow \varphi(n, Y^b)).$$

Rather than show the existence of Y directly we'll show the existence of its characteristic function Z, defined by

$$\langle n, 0\rangle \in Z \leftrightarrow n \notin Y$$

and

$$\langle n, 1\rangle \in Z \leftrightarrow n \in Y.$$

Given our coding scheme for hierarchies note that this amounts to saying that we want Z to satisfy

$$\langle\langle b, m\rangle, 0\rangle \in Z \leftrightarrow m \notin Y_b$$

and

$$\langle\langle b, m\rangle, 1\rangle \in Z \leftrightarrow m \in Y_b$$

where the hierarchy coded by Y satisfies the conclusion of (ATR). The idea is to write down an arithmetic formula describing the inductive buildup of Z and then take a fixed point, although some care has to be taken to insure positivity.

Starting with the formula $\varphi(n,X)$ construct the formula $\widehat{\varphi}(n,Z,b)$ as follows. First put φ into negation-normal form, so that all negations appear in front of atomic formulas. Then replace formulas of the form $t \in X$ by

$$(t_0 \prec b \wedge \langle t, 1\rangle \in Z),$$

and formulas of the form $t \notin X$ by

$$(t_0 \not\prec b \vee \langle t, 0\rangle \in Z).$$

Note that Z occurs only positively in $\widehat{\varphi}$. The meaning behind the above substitutions is this: if Z represents the characteristic function of a set Y, we have that for all n and b

$$\widehat{\varphi}(n, Z, b) \leftrightarrow \varphi(n, Y^b).$$

Do the same to $\neg\varphi$ to obtain a formula $\widehat{\neg\varphi}$ in which Z occurs postively and so that whenever Z represents the characteristic function of Y we have

$$\widehat{\neg\varphi}(n, Z, b) \leftrightarrow \neg\varphi(n, Y^b).$$

Now consider the formula $\psi(n, X)$ given by

$$\begin{aligned} &\psi(\langle\langle b, m\rangle, k\rangle, X) \equiv \\ &\qquad (\forall c \prec b \forall l(\langle\langle c, l\rangle, 0\rangle \in X \vee \langle\langle c, l\rangle, 1\rangle \in X) \wedge \\ &\qquad ((k = 0 \wedge \widehat{\neg\varphi}(m, X, b)) \vee (k = 1 \wedge \widehat{\varphi}(m, X, b)))). \end{aligned}$$

Since X occurs only positively in ψ, by (FP) there is a set Z such that for any triple $\langle\langle b, n\rangle, k\rangle$ we have that

$$\langle\langle b, n\rangle, k\rangle \in Z \leftrightarrow \psi(\langle\langle b, n\rangle, k\rangle, Z).$$

We claim that Z represents the characteristic function of the set Y we seek. Under this interpretation, the inductive definition given by ψ can be paraphrased as follows: we can decide whether or not to put an element m into Y_b only after all the elements of Y^b have been decided; at that point, we put m into Y_b if $\varphi(n, Y^b)$ holds and keep it out otherwise.

Of course, there's no immediate guarantee that the set Z obtained from the fixed point definition even defines a characteristic function at all; that is, there is nothing per se to assure us that for any element n we have

$$\langle n, 0\rangle \in Z \leftrightarrow \langle n, 1\rangle \notin Z.$$

This is where the fact that $\prec$ is a well-ordering comes in. Assume that for some n the above fails. By arithmetic comprehension, we can consider the set of elements b such that for some m we have

$$\neg(\langle\langle b, m\rangle, 0\rangle \in Z \leftrightarrow \langle\langle b, m\rangle, 1\rangle \notin Z).$$

Since $\prec$ is a well-ordering we can find the $\prec$-least such b, i.e. the first place where "things go wrong." Then Z up until this point *does* represent the characteristic function of a set Y^b; but then, referring back to ψ, we see that for each m exactly one of $\langle\langle b, m\rangle, 0\rangle$ or $\langle\langle b, m\rangle, 1\rangle$ is in Z, which gives us a contradiction.

Letting Y be the set whose characteristic function is Z, the reader can verify that the definition of ψ again guarantees that

$$\forall m, b(m \in Y_b \leftrightarrow \varphi(m, Y^b))$$

as desired. Thus we've proven the left to right direction of our theorem.

To prove the converse direction, we use the method of "pseudohierarchies" described in [18]. Working in ATR_0, we want to show how to define a fixed point of any positive arithmetic formula $\varphi(n, Z)$. Consider the usual way of building such a fixed point: one starts with the empty set and then iterates the process

$$\emptyset, \Gamma_\varphi(\emptyset), \Gamma_\varphi(\Gamma_\varphi(\emptyset)), \ldots$$

through the ordinals, taking unions at limit stages. This hierarchy has the property that it is increasing; and any number that enters the union of the sets along this hierarchy enters at a successor stage.

Now, for any well-ordering $\prec$, ATR_0 proves the existence of the hierarchy defined by this process along $\prec$. The only problem is that it may not necessarily have a well-ordering long enough for the procedure to terminate. But we will show that ATR_0 *can* iterate the process "too" long, i.e. along a linear ordering that is *not* well-ordered. Dividing the hierarchy beneath a non-well-ordered part will give us our result.

The details are as follows. First, we need the following

Lemma 3.2 *For any Σ^1_1 formula $\psi(\prec)$, ACA_0 proves*

$$\neg\forall X(\psi(X) \leftrightarrow WO(X)).$$

Proof. This amounts to showing that ACA_0 can carry out the usual proof that WO is a complete Π^1_1 predicate, and hence not equivalent to any Σ^1_1 formula.

More specifically, if T is a tree on $\omega \times \{0,1\}$ and X is a set, let $X[m]$ denote the initial segment of the characteristic function of X of length m, let T^X be the tree on ω given by

$$\sigma \in T^X \leftrightarrow \langle \sigma, X[length(\sigma)] \rangle \in T,$$

and let $KB(T^X)$ be its Kleene-Brouwer ordering. By the usual reductions (see [8, 7]), for each Π^1_1 formula $\theta(X)$ ACA_0 proves the existence of a tree T on $\omega \times \{0,1\}$ so that for any set X,

$$\theta(X) \leftrightarrow \text{“}T^X \text{ is well-founded”} \leftrightarrow WO(KB(T^X)).$$

Given a Σ^1_1 formula $\psi(X)$, we diagonalize by letting T be the tree corresponding to the Π^1_1 formula $\theta(X)$ given by

$$\theta(X) \equiv \text{“}X \text{ is a tree on } \omega \times \{0,1\}\text{”} \wedge \neg\psi(KB(X^X)).$$

Then we have

$$WO(KB(T^T)) \leftrightarrow \theta(T) \leftrightarrow \neg\psi(KB(T^T)),$$

so $KB(T^T)$ is a set witnessing the conclusion of the lemma. □

Let $LO(\prec)$ be the arithmetic assertion that $\prec$ is a linear ordering. Define $\psi(\prec)$ as follows:

$$\begin{aligned}
\psi(\prec) \equiv LO(\prec) \wedge \exists Y & \\
& (Y_0 = \emptyset \wedge \\
& \forall\alpha(Y_{\alpha+1} = \{n|\varphi(n, Y_\alpha)\}) \wedge \\
& \forall\alpha(\lim(\alpha) \rightarrow Y_\alpha = \bigcup_{\beta\prec\alpha} Y_\beta) \wedge \\
& \forall\alpha, \beta(\alpha \prec \beta \rightarrow Y_\alpha \subseteq Y_\beta) \wedge \\
& \forall\alpha, n(n \in Y_\alpha \rightarrow \exists\beta(n \notin Y_\beta \wedge n \in Y_{\beta+1})).
\end{aligned}$$

In words, $\psi(\prec)$ says that $\prec$ is a linear order and there is a hierarchy along $\prec$ starting with the empty set, applying Γ_φ at successor stages, and taking unions at limit stages, with the following additional properties: the hierarchy is increasing (we'll call this condition *) and any number to enter the hierarchy enters at some successor stage (we'll call this property **).

As already remarked above, ATR_0 proves $WO(\prec) \rightarrow \psi(\prec)$. By Lemma 3.2 we can conclude, in ATR_0, that there is some set $\prec$ such that $\psi(\prec)$ but $\prec$ is *not* a well ordering. Then $\prec$ is a linear ordering, and there is pseudohierarchy Y satisfying the conditions set down by ψ. Let W be a set with no $\prec$-least element (without loss of generality we can assume W is closed upwards), and let $W' = \{c|\forall b \in W(c \prec b)\}$. So W is an ill-founded part of our linear ordering, and W' contains the elements beneath W. By arithmetic comprehension let $Z = \bigcap_{b\in W} Y_b$, and let $Z' = \bigcup_{c\in W'} Y_c$; that is, Z is the intersection of the top part, and Z' is the union of the bottom part. We claim that $Z = Z'$, and that this is the fixed point we're looking for.

The fact that $Z' \subset Z$ follows from property (*), since every set in the bottom part is contained in every set in the top part. Conversely, the fact that $Z \subset Z'$ follows from property (**). Suppose $n \in Y_b$ for some b in W. By (**) take d so that $n \notin Y_d$ but $n \in Y_{d+1}$. If d is in W' we have that n is in both Z and Z'; if d is in W then n is in neither.

But clearly $Z' \subset \Gamma_\varphi(Z')$: since each Y_c in the bottom part (i.e. for $c \in W'$) is a subset of Z', $Y_c \subset Y_{c+1} = \Gamma_\varphi(Y_c) \subset \Gamma_\varphi(Z')$, and so $\bigcup_{c\in W'} Y_c \subset \Gamma_\varphi(Z')$. Similar reasoning applies to show that $\Gamma_\varphi(Z) \subset Z$, so $Z = Z'$ is the desired fixed point.

This completes the proof of Theorem 3.1. □

4 The Conservation Result

The aim of this section is to prove the following

Theorem 4.1 *ATR_0 is conservative over $\widehat{ID}_{<\omega}$ for arithmetic formulas with no fixed-point predicates. In other words, if φ is a formula in the language of Peano Arithmetic such that ATR_0 proves φ, then $\widehat{ID}_{<\omega}$ proves φ as well.*

Proof. By the previous section, we can take FP_0 as our axiomatization of ATR_0. The model-theoretic proof is straightforward: given a model M of $\widehat{ID}_{<\omega}$ one can convert it to a second-order model M' of FP_0 by interpreting the second-order part of M' by the "projections" of the denotations of the fixed-point predicates of M (see axiom (1) at the end of Section 2). The proof-theoretic analog is not much more difficult. We present it below.

First we introduce the auxilliary system FP_0' with terms naming the fixed points guaranteed to exist by (FP). In other words, for every arithmetic formula $\varphi(x, X, \vec{y}, \vec{Y})$ in which X occurs positively we introduce a term $F_\varphi(\vec{y}, \vec{Y})$ with the free variables shown. FP_0' then contains axioms

$$(FP') \quad \forall x(x \in F_\varphi(\vec{y}, \vec{Y}) \leftrightarrow \varphi(x, F_\varphi(\vec{y}, \vec{Y}), \vec{y}, \vec{Y})),$$

as well as FP_0's induction and quantifier-free axioms. Note that (FP') easily implies (FP) in a standard axiomatization of two-sorted predicate logic.

We assume that the reader is familiar with cut-elimination arguments as they appear in [13, 9, 14]. To formalize second-order deductions we use a two-sorted Tait-style system with equality. If T is a set of axioms, we use T^c to denote the universal closure of these axioms, and $\neg T^c$ to denote their negations. The notation $[T]$ denotes some finite subset of T.

Suppose ATR_0 proves φ in a standard Hilbert-style proof system. Then by the deduction theorem there is a proof of $\bigwedge[FP_0'^c] \rightarrow \varphi$. By cut-elimination, there is a cut-free proof of $[\neg FP_0'^c], \varphi$ in a Tait-style system.

Note that second-order universal quantifiers in the closed axioms of $FP_0'^c$ become existential quantifiers in $[\neg FP_0'^c]$. To eliminate these we use a second-order version of Herbrand's theorem:

Lemma 4.2 *Suppose there is a cut-free proof of $\Gamma, \exists X \varphi(X)$ in which the formulas in $\Gamma, \varphi(X)$ are arithmetic. Then there are terms $T_i(\vec{y}_i)$ and a cut-free proof of*

$$\Gamma, \ldots, \exists \vec{y}_i \varphi(T_i(\vec{y}_i)), \ldots.$$

Proof. As in the proof of Herbrand's theorem, inductively replace inferences of the form

$$\frac{\Delta, \varphi(T(\vec{y}))}{\Delta, \exists X \varphi(X)}$$

by inferences

$$\frac{\Delta, \varphi(T(\vec{y}))}{\Delta, \exists \vec{y} \varphi(T(\vec{y}))}.$$

(Note that Δ may already contain the formula $\exists X \varphi(X)$, so that the elimination of the existential set quantifier may require several terms.) □

By (a suitable generalization of) the previous lemma, we can now translate the proof of

$$[\neg FP_0'^c], \varphi$$

to a proof of

$$[\neg FP_0^*], \varphi$$

where FP_0^* consists of first-order universal closures of substitution instances of axioms of FP_0'. If there are any free second-order variables in FP_0^*, by the substitution lemma (see [13]) we can replace them by arbitrary closed terms. As a result, we can assume that the formulas in the deduction are arithmetic and contain no second-order variables.

Now go through the proof and replace each formula φ by a formula $\hat{\varphi}$ in the language of $\widehat{ID}_{<\omega}$, by iteratively replacing atomic formulas

$$s \in F_{\psi(x,X,\vec{y},\vec{Y})}(\vec{t}(\vec{z}), \vec{T}(\vec{z}))$$

by $\widehat{ID}_{<\omega}$-terms

$$P_{\hat{\psi}(x,X,\vec{t}(\vec{z}),\vec{T}(\vec{z}))}(s, \vec{z}).$$

This has the net effect of replacing the fixed-point axioms of FP_0^* by fixed-point axioms (1) of $\widehat{ID}^c_{<\omega}$, and induction axioms of FP_0^* by induction axioms of $\widehat{ID}^c_{<\omega}$, while leaving the rules of inference sound. The result then is a proof of

$$[\neg \widehat{ID}^c_{<\omega}], \varphi$$

which can be converted back to a Hilbert-style proof if desired. □

5 The Speedup Result

Because the argument of the previous section involves a cut-elimination, it allows for a possibly superexponential increase in length when translating an ATR_0 proof to one in $\widehat{ID}_{<\omega}$. We aim to show that this increase is unavoidable, in that ATR_0 has short proofs of the sentences $Con(\widehat{ID}_{2^0_n})$, where 2^0_y is a formalization of the stack-of-twos function, i.e. the function $f(x) = 2^x$ iterated y times starting with 0. To that end we will use the following theorem, due to Solovay (see [11, 4]):

Theorem 5.1 *Let $I(x)$ be a cut in the theory T, i.e. a formula such that T proves $I(0)$ and $I(x) \to I(x+1)$. Then for every natural number n there is a cut J_n such that T proves*

$$\forall x (J_n(x) \to I(2^x_n)).$$

This has the following important corollary (see [10, 5]):

Corollary 5.2 *If I is a cut in T, then there is a polynomial p such that for each numeral $\bar{n}$, T proves $I(2^0_{\bar{n}})$ with a proof of length at most $p(n)$.*

In this section we'll demonstrate a cut I such that ATR_0 proves

$$\forall x(I(x) \to Con(\widehat{ID}_x)).$$

This, combined with the results just cited, will give us the following:

Theorem 5.3 *There is a polynomial p such that for each numeral $\bar{n}$, ATR_0 has a proof of $Con(\widehat{ID}_{2^0_{\bar{n}}})$ with length at most $p(n)$.*

Our cut $I(x)$ will say, roughly, that there exists an ω-model of $\widehat{ID}_x$. To that end, we need to make some definitions from within ATR_0. Assuming that the language and axioms of $\widehat{ID}_n$ have been formalized uniformly in ATR_0, let $Sent(\widehat{ID}_x)$ be the set of (Gödel codes) of sentences in the language of $\widehat{ID}_x$. We'll say that M is an ω-model for the language of $\widehat{ID}_x$ if it is a sequence of sets S_φ (coded as a single set), one for each set constant P_φ. A valuation for M is a map

$$f : Sent(\widehat{ID}_x) \to \{0, 1\}$$

such that

$$f(\ulcorner \bar{n} \in P_\varphi \urcorner) = 1 \leftrightarrow n \in S_\varphi,$$

f assigns 1 to true atomic formulas in the language of PA and 0 to false ones, and f obeys Tarski's truth conditions for the other logical connectives. Note that for any ω-model M ATR_0 easily proves there is a unique valuation for M, by iterating comprehension along a well-ordering of length ω. We'll say that $M \models \varphi$ if this unique valuation assigns φ the value 1, and we'll say M is an ω-model for $\widehat{ID}_x$ if M models each axiom of $\widehat{ID}_x$. (The reason we are calling M an ω-model is that we are implicitly assuming that its first-order universe is the same as that of ATR_0.)

Let $I(x)$ represent the statement

"There exists an ω-model of $\widehat{ID}_x$."

We claim $I(x)$ is the desired cut. It isn't difficult to show, by a standard soundness argument, that $\forall x(I(x) \to Con(\widehat{ID}_x))$. $\widehat{ID}_0$ is simply PA, and it is easy to show that the empty sequence is an ω-model of PA. (That is, there is a valuation that assigns a value of 1 to the axioms of PA; just use an iteration of length ω to define a truth-predicate for arithmetic sentences, as in [18].) This gives $I(0)$. And so we are reduced to showing from within ATR_0 that $I(x) \to I(x+1)$, i.e. that the existence of an ω-model for $\widehat{ID}_x$ implies the existence of an ω-model for $\widehat{ID}_{x+1}$.

One approach to this is as follows. Feferman [2] sketches Aczel's proof that each $\widehat{ID}_{n+1}$ can be interpreted in Σ^1_1-$AC(\widehat{ID}_n)$, where the latter represents the second-order theory obtained by adding $(\Sigma^1_1\text{-}AC)$ and the scheme of full induction to $\widehat{ID}_n$. Simpson [17, 18] shows that ATR_0 proves that any countable sequence of sets can be expanded to an ω-model of $(\Sigma^1_1\text{-}AC)$. So given an ω-model of $\widehat{ID}_x$, we can first expand it to a model of $(\Sigma^1_1\text{-}AC)$ and use that to determine the intepretations of the constants of $\widehat{ID}_{x+1}$. The method we present here is more direct. We reduce the proof to two lemmas.

Lemma 5.4 *ATR_0 proves the following: Suppose M is an ω-model of $\widehat{ID}_n$, and for every positive arithmetic formula $\varphi(z,Y)$ in the language of $\widehat{ID}_x$ there is a set S_φ so that when P_φ is interpreted as S_φ,*

$$M \cup \{S_\varphi\} \models \forall z(z \in P_\varphi \leftrightarrow \varphi(z, P_\varphi)).$$

Then there is an ω-model of $\widehat{ID}_{x+1}$.

In other words, if we can interpret each new P_φ individually, we can obtain a model of $\widehat{ID}_{x+1}$. The proof is straightforward: since ATR_0 proves $(\Sigma^1_1\text{-}AC)$, we can combine all the S_φ with M to obtain a new model M' (and also combine the valuations for each model $M \cup \{S_\varphi\}$ into a sequence of valuations $\langle f_\varphi \rangle$). Let f' be a valuation for M'; it isn't hard to show that f' has to agree with f_φ on the language involving just the one new constant P_φ, so f validates all the new fixed-point axioms of $\widehat{ID}_{x+1}$.

We've now reduced the proof of Theorem 5.3 to the following

Lemma 5.5 *ATR_0 proves the following: Let M be an ω-model of $\widehat{ID}_x$, and let $\varphi(z,Y)$ be a positive arithmetic formula in the language of $\widehat{ID}_x$. Then there is a set S_φ such that*

$$M \cup \{S_\varphi\} \models \forall z(z \in P_\varphi \leftrightarrow \varphi(z, P_\varphi))$$

when P_φ is interpreted as S_φ.

Proof. We'll use the axiom (FP) to prove this; but rather than define the set S_φ alone we'll define both it *and* a partial evaluation F for the language with the new constant at the same time. In other words, we'll present a formula $\psi(n,Y)$ for which a fixed point Y will represent an ordered pair $\langle F, S\rangle$, where S is the set S_φ and F is an evaluation for sentences of the new language in which the constant P_φ occurs positively.

It will be convenient to assume that all formulas are identified with their negation-normal-form equivalents, in which all negations have been pushed down to the atomic level. As such, a sentence in which P_φ occurs positively is one in which there are no occurences of subformulas of the form $\neg t \in P_\varphi$.

To code pairs of sets we'll write $Y = Y_0 \oplus Y_1$, where $Y_0 = \{n | \langle 0, n\rangle \in Y\}$ and $Y_1 = \{n | \langle 1, n\rangle \in Y\}$.

Without further ado, we define $\psi(n, Y)$ as follows.

$$\begin{aligned}\psi(n,Y) \equiv\ & n = \langle 0, \langle \ulcorner \bar{m} \in P_\varphi \urcorner, 1\rangle\rangle \wedge m \in Y_1 \\ & \vee\ n = \langle 0, \langle \ulcorner \theta \urcorner, 1\rangle\rangle \wedge \text{“}\theta \text{ is atomic and true in } M\text{”} \\ & \vee\ n = \langle 0, \langle \ulcorner \theta \wedge \nu \urcorner, 1\rangle\rangle \wedge (\langle \ulcorner \theta \urcorner, 1\rangle \in Y_0 \wedge \langle \ulcorner \nu \urcorner, 1\rangle \in Y_0) \\ & \vee\ \ldots \\ & \vee\ n = \langle 1, m\rangle \wedge \langle \ulcorner \varphi(\bar{m}, P_\varphi)\urcorner, 1\rangle \in Y_0.\end{aligned}$$

In other words, a sentence gets assigned a truth value of 1 by Y_0 (our putative valuation) if it is either of the form $\bar{m} \in P_\varphi$ and m is in Y_1, or if it is inductively true by the clauses of Tarski's truth definition. An element m makes it into Y_1 (our attempt at building S_φ) if and only if $\varphi(\bar{m}, P_\varphi)$ has been assigned a truth value of 1 by Y_0.

Since Y occurs only positively in the above formula, by (FP) there is a fixed point Z. Let $S_\varphi = Z_1$ and $F = Z_0$. Let $M' = M \cup \{S_\varphi\}$ and let f' be an evaluation for M' (in the language of $\widehat{ID}_x$ plus the new constant P_φ). We claim that for every sentence θ in which P_φ occurs only positively, $f'(\ulcorner\theta\urcorner) = 1$ if and only if $F(\ulcorner\theta\urcorner) = 1$; that is, the partial evaluation F is correct for these sentences. This is easy to verify by induction on the complexity of θ. (Recall that we only have to deal with negations at the atomic level, and no negated instances of $t \in P_\varphi$.) But then f' satisfies

$$f'(\ulcorner \bar{n} \in P_\varphi \urcorner) = 1 \leftrightarrow f'(\ulcorner \varphi(\bar{n}, P_\varphi)\urcorner) = 1,$$

since F does, and so,

$$f'(\ulcorner \forall x(x \in P_\varphi) \leftrightarrow \varphi(n, P_\varphi)\urcorner) = 1.$$

So M' is a model of $\widehat{ID}_x$ together with the new fixed point axiom, proving the lemma. This also completes the proof of Theorem 5.3. □

6 Comments and Acknowledgements

In Section 5 we used the fact that ATR_0 proves the Σ^1_1 axiom of choice in our proof of Lemma 5.4. Solomon Feferman has pointed out that one can avoid the use of (Σ^1_1-AC) and still obtain the speedup result, say, by defining the cut $I(x)$ to mean “there is an ω-model of any x fixed-point axioms of $\widehat{ID}_{<\omega}$.” I am grateful to him for this observation as well as a suggestion simplifying the proof of Theorem 3.1.

I'd also like to thank Stephen Simpson for sending me a preprint of his manuscript, and my advisor, Jack Silver, whose support has been invaluable to me.

References

[1] Avigad, Jeremy, *Proof-Theoretic Investigations of Subsystems of Second-Order Arithmetic*, Ph.D. Dissertation, U. C. Berkeley, 1995.

[2] Feferman, Solomon, "Iterated Inductive Fixed-Point Theories: Application to Hancock's Conjecture", in G. Metakides ed. *Patras Logic Symposium*, North-Holland Publishing Company, 1982.

[3] Friedman, Harvey, Kenneth McAloon, and Stephen Simpson, "A Finite Combinatorial Principle Which is Equivalent to the 1-Consistency of Predicative Analysis," in G. Metakides ed. *Patras Logic Symposium*, North-Holland Publishing Company, 1982.

[4] Hájek, Petr and Pavel Pudlak, *Metamathematics of First-Order Arithmetic*, Springer Verlag, 1991.

[5] Ignjatovic, Aleksandar, *Fragments of First and Second Order Arithmetic and Length of Proofs*, Ph.D. Dissertation, U. C. Berkeley, 1990.

[6] Jäger, Gerhard, *Theories for Admissible Sets: A Unifying Approach to Proof Theory*, Bibliopolis, 1986.

[7] Jech, Thomas, *Set Theory*, Academic Press, 1978.

[8] Martin, Donald, "Descriptive set theory: Projective sets," in Jon Barwise ed., *The Handbook of Mathematical Logic*, North-Holland, 1991.

[9] Pohlers, Wolfram, *Proof Theory: An Introduction*, Springer Verlag Lecture Notes in Mathematics 1407, 1989.

[10] Pudlák, Pavel, "On the length of proofs of finitistic consistency statements in first order theories," in Paris et al. editors, *Logic Colloqium '84*, North Holland, 1986.

[11] Pudlák, Pavel, "Cuts, Consistency Statements, and Interpretations," Journal of Symbolic Logic, 50:423-441, 1985.

[12] Schutte, Kurt, *Proof Theory*, Springer-Verlag, 1977.

[13] Schwichtenberg, Helmut, "Proof theory: Some applications of cut-elimination," in Jon Barwise ed., *The Handbook of Mathematical Logic*, North-Holland, 1991.

[14] Sieg, Wilfried, "Herbrand Analysis," Archive for Mathematical Logic, 30:409-441, 1991.

[15] Simpson, Stephen G., "Σ^1_1 and Π^1_1 Transfinite Induction," in D. van Dalen et al. eds. *Logic Colloqium '80*, North-Holland, 1982.

[16] Simpson, Stephen G., "Subsystems of Z_2 and Reverse Mathematics," appendix to Gaisi Takeuti, *Proof Theory* (second edition), North-Holland, 1987.

[17] Simpson, Stephen G., "On the strength of König's duality theorem of countable bipartite graphs," Journal of Symbolic Logic, 59:113-123, 1994.

[18] Simpson, Stephen G., *Subsystems of Second Order Arithmetic*, preprint.