A Small Reflection Principle for Bounded Arithmetic
Author(s): Rineke Verbrugge and Albert Visser
Source: *The Journal of Symbolic Logic*, Vol. 59, No. 3 (Sep., 1994), pp. 785-812
Published by: Association for Symbolic Logic
Stable URL: http://www.jstor.org/stable/2275908
Accessed: 22/04/2010 12:10

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=asl.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Association for Symbolic Logic is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Symbolic Logic*.

http://www.jstor.org

THE JOURNAL OF SYMBOLIC LOGIC
Volume 59. Number 3. September 1994

A SMALL REFLECTION PRINCIPLE FOR BOUNDED ARITHMETIC

RINEKE VERBRUGGE AND ALBERT VISSER

Abstract. We investigate the theory $I\Delta_0 + \Omega_1$ and strengthen [Bu86. Theorem 8.6] to the following: if NP $\neq$ co-NP. then Σ-completeness for witness comparison foumulas is not provable in bounded arithmetic. i.e..

$$I\Delta_0 + \Omega_1 \nvdash \forall b \forall c (\exists a (\mathrm{Prf}(a. c) \wedge \forall z \leq a \neg \mathrm{Prf}(z. b)) \\ \rightarrow \mathrm{Prov}(\ulcorner \exists a (\mathrm{Prf}(a. \overline{c}) \wedge \forall z \leq a \neg \mathrm{Prf}(z. \overline{b}))\urcorner)).$$

Next we study a "small reflection principle" in bounded arithmetic. We prove that for all sentences φ

$$I\Delta_0 + \Omega_1 \vdash \forall x \, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x} (\mathrm{Prf}(y. \overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi)\urcorner).$$

The proof hinges on the use of definable cuts and partial satisfaction predicates akin to those introduced by Pudlák in [Pu86].

Finally. we give some applications of the small reflection principle. showing that the principle can sometimes be invoked in order to circumvent the use of provable Σ-completeness for witness comparison formulas.

§**1. Introduction.** A striking feature of Solovay's Theorem that *Löb's logic is complete for arithmetical interpretations* is its amazing stability. If one sticks to the unimodal propositional language and standard arithmetical interpretations, the result holds (modulo a trivial variation) for any decently axiomatized extension of $I\Delta_0$+ EXP. Such stability is in some sense a weakness: unimodal propositional logic combined with the standard interpretation cannot serve to classify or give information on specific theories in a broad range. Of course this weakness disappears when we extend the modal language, but that is not our subject here (however, see [Vi90], [Be91], [Be89]).

Is there life outside the broad range of arithmetical theories satisfying Solovay's Completeness Theorem? Clearly the question is only sensible if the theories under consideration verify Löb's logic, or perhaps some still interesting weakening of it.

Two directions of research come to mind. The first one is to weaken the logic of the arithmetical theory. Specifically one can study theories like Heyting Arithmetic (**HA**), the constructive version of Peano Arithmetic. It turns out that **HA** verifies the obvious constructive version of Löb's logic plus a wide variety of extra principles (see [Vi81], [Vi82], [Vi85]). The only definitive information that we have is a characterization of the closed fragment of **HA**. For all we know the provability logic corresponding to **HA** itself could be Π^0_2-complete. Moreover, extensions of

Received October 22. 1992.

The research of the first author was supported by the Netherlands Organization for Scientific Research (NWO).

©1994 Association for Symbolic Logic
0022-4812/94/5903-0003/$03.80

HA have quite different provability logics. Note by the way that provability logics need not be monotonic in their arithmetical theories.

The second direction of research is simply to look at classical arithmetical theories that are strictly weaker than, or even incompatible with, $\mathrm{I}\Delta_0+$ EXP. It turns out that there are two salient theories of this kind: Paris and Wilkie's $\mathrm{I}\Delta_0+\Omega_1$ and Buss' S^1_2, both of them satisfying Löb's logic (see [WP87], [Bu86]). Does Solovay's Theorem still hold for them? At present nobody knows—or to be precise, we haven't heard that anybody knows.

This paper is a first contribution to an understanding of the difficulties involved in proving or disproving Solovay's Theorem for theories like $\mathrm{I}\Delta_0+\Omega_1$ and S^1_2. Solovay's proof involves Rosser methods. The problem for us resides in the instances of Π^b_1-completeness that occur in the proof. Two points are important.

• We do not know whether the instances of Π^b_1-completeness used in Solovay's proof are provable in our target theories. Buss proved that provability of Π^b_1-completeness with parameters in S^1_2 implies NP = co-NP (see [Bu86]). In §3 we elaborate on this theme. To be specific, we prove that if NP $\neq$ co-NP, then Σ-completeness for witness comparison formulas is not provable in bounded arithmetic, i.e.,

$$\mathrm{I}\Delta_0+\Omega_1 \nvdash \forall b \forall c(\exists a(\mathrm{Prf}(a,c) \wedge \forall z \leq a \neg \mathrm{Prf}(z,b)) \\ \rightarrow \mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(a,\overline{c}) \wedge \forall z \leq a \neg \mathrm{Prf}(z,\overline{b}))\urcorner)).$$

• In many cases we can circumvent the use of instances of Π^b_1-completeness. Švejdar discovered the first alternative argument when he surprisingly provided a proof of Rosser's Theorem that genuinely differed from Rosser's own proof (see [Šv83]). To this end he introduced a principle which we have dubbed Švejdar's principle. In §4 we prove a "small reflection principle" in our target theories from which Švejdar's principle immediately follows. More precisely, we show that for all sentences φ,

$$\mathrm{I}\Delta_0+\Omega_1 \vdash \forall x\, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{Prf}(y,\overline{\ulcorner\varphi\urcorner}) \rightarrow \varphi)\urcorner).$$

Švejdar's principle is not sufficient to derive Solovay's Theorem. However, it has been fruitfully exploited in the dogged attempt to use Solovay-like methods to embed larger and larger classes of Kripke models for Löb's logic in our weak arithmetical theories. The state of this dogged art can be found in [BV93].

We end §4 with some other applications of the small reflection principle.

In §5, we use the small reflection principle in order to extend Krajíček and Pudlák's result on the injection of inconsistencies into models of $\mathrm{I}\Delta_0+$ EXP.

Theorem 3.7 and Theorem 4.20, the main results of §3 and §4, were published previously in the first author's technical report [Ve89], which in turn is based on her master's thesis [Ve88].

§**2. Preliminaries.** We assume that the reader is familiar with the standard references to the area of weak arithmetics (see [Bu86], [WP87], and Chapter V of [HP93]). However, for ease of reference, we quickly review those concepts that we need in the sequel.

The principal feature distinguishing various theories of Bounded Arithmetic from Peano Arithmetic is that in the former induction is restricted to bounded formulas.

2.1. $\mathbf{I\Delta_0 + \Omega_1}$.

DEFINITION 2.1. The language of $I\Delta_0 + \Omega_1$ as introduced in [WP87] contains $0, S, +, \cdot, =$, and $\leq$, and additionally the logical symbols $\neg, \rightarrow$, and $\forall$, and variables $v_1, v_2, \ldots$. With regard to logical axioms, we use a Hilbert-type system as in [WP87], but other choices are reasonable too. For example, a Gentzen style sequent calculus with cut rule or natural deduction would do. However, we do not use a logic in which only direct proofs (i.e., tableau proofs or cut-free proofs) are allowed.

As nonlogical axions we consider a set containg the following:

- a finite number of universal formulas defining the basic properties of the function and predicate symbols of the language:

(1) $0 \leq 0 \wedge \neg(S0 \leq 0)$;
(2) $\forall x(x + 0 = x \wedge x \cdot 0 = 0 \wedge x \cdot S0 = x)$;
(3) $\forall x \forall y(Sx = Sy \rightarrow x = y)$;
(4) $\forall x \forall y(x \leq Sy \leftrightarrow (x \leq y \vee x = Sy))$;
(5) $\forall x \forall y(x + Sy = S(x + y))$;
(6) $\forall x \forall y(x \cdot Sy = (x \cdot y) + x)$;

- a formula $\forall x \exists y \varphi(x, y)$, where φ is the Δ_0-formula defining the relation $y = \omega_1(x)(= x^{|x|})$;
- the scheme of induction for Δ_0-formulas.

2.2. Buss' systems of bounded arithmetic and the polynomial hierarchy.

DEFINITION 2.2. The language of Buss' bounded arithmetic consists of $0, S, +, \cdot$, $=, \leq, |x|(= \ulcorner \log_2(x + 1)\urcorner$, the length of the binary representation of x), $\llcorner \frac{1}{2}x \lrcorner$, and $x\#y(= 2^{|x|\cdot|y|}$, the smash function).

REMARK 2.3. Note that the smash function # allows us to express terms approximately equal to $2^{P(|x|)}$ for any polynomial P. More precisely, for every n, $x \geq 2$ the following holds:

$$2^{|x|^n} \leq \underbrace{x\#\cdots\#x}_{n \text{ times}} \leq 2^{2\cdot|x|^n-2},$$

as is easily proved by induction. This property of # is useful when we want to define polynomial time functions.

DEFINITION 2.4. The *hierarchy of bounded arithmetic formulas* is defined as follows:

(1) $\Sigma_0^b = \Pi_0^b = \Delta_0^b$ is the set of formulas with only sharply bounded quantifiers $\forall x \leq |t|$, $\exists x \leq |t|$ (wehre t is any term not involving x);
(2) Σ_{k+1}^b is defined inductively by
- $\Sigma_{k+1}^b \supseteq \Pi_k^b$, and is closed under $\wedge, \exists x \leq t$, and $\forall x \leq |t|$;
- if $B \in \Pi_{k+1}^b$, then $\neg B \in \Sigma_{k+1}^b$.

(3) Π_{k+1}^b is defined inductively by
- $\Pi_{k+1}^b \supseteq \Sigma_k^b$, and is closed under $\wedge, \forall x \leq |t|$, and $\exists x \leq |t|$;
- if $B \in \Sigma_{k+1}^b$, then $\neg B \in \Pi_{k+1}^b$.

(4) Σ^b_{k+1} and Π^b_{k+1} are the smallest sets which satisfy (2) and (3).

Definition 2.5. If R is a theory and A a formula, we say that A is Δ^b_{k+1} with respect to R iff there are formulas $B \in \Sigma^b_{k+1}$ and $C \in \Pi^b_{k+1}$ such that $R \vdash A \leftrightarrow B$ and $R \vdash A \leftrightarrow C$.

We never leave out the superscripts b from the levels of Σ^b_n and Π^b_n of Buss' bounded arithmetical hierarchy, so our use of Σ_n for Σ^0_n and Π_n for Π^0_n should not give rise to confusion.

The hierarchy of bounded arithmetic formulas is constructed in such a way that all levels Π^b_i and Σ^b_i except Σ^b_0 correspond to levels of the polynomial hierarchy, which is well known from structural complexity theory. Without defining all the basic notions of complexity theory, for which the reader may turn to [BDG87], we give one of the standard definitions.

Definition 2.6. The *polynomial hierarchy* is defined as follows.

(1) $\mathrm{P}=\Delta^p_1$ is the set of predicates on the natural numbers which are recognized by a deterministic polynomial time Turing machine;

(2) $\mathrm{NP} = \Sigma^p_1$ is the set of predicates on the natural numbers which are recognized by a nondeterministic polynomial time Turing machine;

(3) Σ^p_i is the set of predicates Q such that there is an $R \in \Delta^p_i$ and a polynomial P such that for all $\vec{x}$, $Q(\vec{x}) \iff \exists y \leq 2^{\mathrm{P}(|\vec{x}|)} R(\vec{x}, y)$;

(4) Π^p_i is the set of predicates Q such that there is an $R \in \Sigma^p_i$ so that for all $\vec{x}$, $Q(\vec{x}) \iff \neg R(\vec{x})$.

(5) Δ^p_{i+1} is the set of predicates which are recognized by a deterministic polynomial time Turing machine with some oracle from Σ^p_i.

As usual we use the name co-NP for Π^p_1. There are many open questions about the polynomial hierarchy. The most important one is: is there a k such that $\Sigma^p_k = \Sigma^p_{k+1}$, in which case the hierarchy collapses? More particularly, does NP $=$ co-NP? Or even P $=$ NP? It is also unknown whether for any k, $\Delta^p_k = \Sigma^p_k \cap \Pi^p_k$, and in particular, whether P $=$ NP $\cap$ co-NP.

Definition 2.7. A is *polynomially reducible* to B if there is a polynomial time computable function f such that $\forall x (x \in A \leftrightarrow f(x) \in B)$.

Note that polynomial reducibility is analogous to many-one reducibility from ordinary recursion theory.

Definition 2.8. B is *NP-complete* if all $A \in$ NP are polynomially reducible to B. Similarly, B is *co-NP-complete* if all $A \in$ co-NP are polynomially reducible to B.

Remark 2.9. It is easy to see that for every NP-complete set B, the following hold:

- if $B \in$ co-NP, then NP $=$ co-NP;
- if $B \in$ P, then P $=$ NP.

Remark 2.10. From results of Stockmeyer, Wrathall, and Kent and Hodgson [St76], [Wr76], [KH82] it follows that the bounded arithmetical hierarchy is related to the polynomial hierarchy in the following way: Σ^p_{k+1} is the class of predicates which are defined by formulas in Σ^b_{k+1}. In particular, NP is the class of predicates which are defined by Σ^b_1-formulas; similarly co-NP is the class of predicates defined by Π^b_1-formulas. We refer the reader to [Bu86, Chapter 1] for proofs of these correspondences.

DEFINITION 2.11. The theory S_2^i consists of BASIC, a finite list of axioms defining the basic properties of symbols in the language of bounded arithmetic, plus the following induction scheme PIND(Σ_i^b):

$$A(0) \wedge \forall x(A(\lfloor \tfrac{1}{2}x \rfloor) \to A(x)) \to \forall x A(x) \quad \text{for } A \in \Sigma_i^b.$$

DEFINITION 2.12. $S_2 := \bigcup_i S_2^i$.

One of the most important theorems about bounded arithmetic is Parikh's Theorem. It implies that every Δ_0-definable provably total function of S_2 can increase the length of its input only polynomially.

Parikh originally proved his theorem for $I\Delta_0$, for which the Δ_0-definable provably total functions are even more severely limited than for S_2: they can increase the length of the input only linearly.

We state a version of Parikh's Theorem for Buss' theories S_2^i.

THEOREM 2.13 (Parikh's theorem). *Let $i \geq 0$. Suppose that φ is a bounded formula and that $S_2^i \vdash \forall x \exists y \varphi(x,y)$. Then there is a term $t(x)$ such that $S_2^i \vdash \forall x \exists y \leq t(x) \varphi(x,y)$.*

PROOF. Buss gives a proof-theoretic proof (see [Bu86, Theorem 4.11]), but the theorem can also easily be proved in a model-theoretic way. □

2.3. Metamathematics for bounded arithmetic. In order to prove Gödel's Incompleteness Theorems for bounded arithmetic, Buss arithmetized the usual notions of metamathematics (see [Bu86, Chapter 7]). It turns out that most predicates needed can be Δ_1^b-defined (or sometimes $\exists\Delta_1^b$-defined) in S_2^1. Moreover, these definitions are intensionally correct in the sense of [Fe 60] which means that the usual connections between them can be proved in S_2^1.

Here follows a list of predicates used in the sequel.

- $\mathrm{Seq}(w)$ for "w encodes a sequence";
- $\mathrm{Len}(w) = a$ for "if w encodes a sequence, then the length of that sequence is a; otherwise $a = 0$";
- $\mathrm{Term}(v)$ for "v is the Gödel number of a term";
- $\mathrm{Fmla}(v)$ for "v is the Gödel number of a formula";
- $\mathrm{Prf}_\alpha(u,v)$ for $\mathrm{Fmla}(v)\wedge$ "u is the Gödel number of a proof of the formula with Gödel number v from the set of axioms given by formula $\alpha(x)$". When it is clear that the axioms of a theory T are given by the formula α, we sometimes write Prf_T instead of Prf_α; when α and T are clear from the context, we drop the subscript altogether.
- $\mathrm{Prov}_\alpha(v) := \exists u\, \mathrm{Prf}_\alpha(u,v)$; we sometimes abbreviate $\mathrm{Prov}(\ulcorner\varphi\urcorner)$ as $\Box\varphi$.

The predicates Seq, Len, Term, and Fmla are Δ_1^b-definable in S_2^1, and so is Prf_α where the formula α is Δ_1^b with respect to S_2^1. The condition on α is not a severe restriction. To any recursively enumerable set one can associate a polynomial time function having that set as its range, therefore one can suitably axiomatize any theory T which has a recursively enumerable set of axioms including BASIC.

Notation 2.14. Instead of the usual *numerals* S^k0 of Peano Arithmetic, we use canonical numerals $\overline{k}$ defined inductively by

- $\overline{0} = 0$;
- $\overline{2k+1} = \overline{2k} + (S0)$;
- $\overline{2k+2} = (SS0) \cdot (\overline{k+1})$.

Note that the length of the term $\overline{k}$ is linear in the length of the binary representation of k, a property that the $S^k 0$ obviously do not satisfy. The shortness of canonical terms plays a crucial rôle in many proofs, for example in Buss' proof that S^1_2 enjoys provable Σ^b_1-completeness.

S^1_2 can Σ^b_1-define a function $\mathrm{Num}(x)$ such that $\mathrm{Num}(x)$ stands for the Gödel number of the term $\overline{x}$. For ease of reading, we will however abuse notation; thus, if $A(x)$ is a formula with free variable x we write $\ulcorner A(\overline{a})\urcorner$ instead of $\mathrm{Sub}(\ulcorner A\urcorner, \ulcorner x\urcorner, \mathrm{Num}(a))$. Sometimes we are even more sloppy and leave out the numeral dashes altogether. In those cases the context should provide enough material for the reader to know what is meant.

THEOREM 2.15 (provable Σ^b_1-completeness, Buss). *Let A be any Σ^b_1-formula. Let $a_1, \dots, a_k$ be all the free variables of A. Then there is a term $t(a_1, \dots, a_k)$ such that*

$$S^1_2 \vdash \forall a_1, \dots, a_k (A(a_1, \dots, a_k) \rightarrow \exists w \leq t \, \mathrm{Prf}(w, \ulcorner A(\overline{a_1}, \dots, \overline{a_k})\urcorner)).$$

PROOF. See [Bu86, Theorem 7.4]. □

Using Theorem 2.15, we can easily see that Löb's logic is arithmetically sound with respect to S^1_2. In particular, this means that we can, in the standard way, prove Gödel's Second Incompleteness Theorem and its formalized version for S^1_2.

Sometimes, we will use the name $\mathrm{I}\Delta_0 + \Omega_1$ for Buss' theory S_2 (see Definition 2.12), in which induction for formulas from the hierarchy of bounded arithmetic formulas in a language containing # and | | is allowed. Because S_2 is a conservative extension of $\mathrm{I}\Delta_0 + \Omega_1$, the name change has no repercussions on results that do not hinge on the details of formalization.

2.4. Definable cuts. Because PA proves induction for all first-order formulas, no proper cuts of models of PA can be defined by formulas. In the context of weaker theories where induction is restricted to a proper subset of all formulas, on the contrary, definable cuts have proved to be highly useful tools.

DEFINITION 2.16. Let $T \supseteq Q$ be a Σ^b_1-axiomatized theory. A T- *cut* is a formula I such that

(1) $T \vdash I(0)$;
(2) $T \vdash \forall x \forall y (I(y) \wedge x \leq y \rightarrow I(x))$;
(3) $T \vdash \forall x (I(x) \rightarrow I(Sx))$.

DEFINITION 2.17. Let $T \supseteq Q$ be a Σ^b_1-axiomatized theory. A T-*initial segment* is a formula J such that

(1) $T \vdash J(0)$;
(2) $T \vdash \forall x \forall y (J(y) \wedge x \leq y \rightarrow J(x))$;
(3) $T \vdash \forall x \forall y (J(x) \wedge J(y) \rightarrow (J(Sx) \wedge J(x+y) \wedge J(x \cdot y)))$.

REMARK 2.18. For cuts I, we frequently write $x \in I$ instead of $I(x)$.

LEMMA 2.19. *Suppose that $T \supseteq \mathrm{I}\Delta_0$ and let I be a T-cut. Then there is a formula J such that*

(1) $T \vdash \forall x (J(x) \rightarrow I(x))$;
(2) *J is a T-cut*;
(3) $T \vdash \forall x \forall y (J(x) \wedge J(y) \rightarrow J(x+y))$, *i.e., J is closed under $+$.*

PROOF. Take

$$J(x) :\leftrightarrow I(x) \wedge \forall y (I(y) \rightarrow I(x+y)).$$

It is easy to see that $T \vdash \forall x(J(x) \rightarrow I(x))$ and that J is a T-cut.

For closure under $+$, reason in $\mathrm{I}\Delta_0$ and suppose that $x_1, x_2 \in J$ and that $y \in I$. Then by definition of J we have, first, $x_1 + x_2 \in I$. Also, $y + x_1 \in I$; thus, $y + (x_1 + x_2) = (y + x_1) + x_2 \in I$. We may conclude that $x_1 + x_2 \in J$. □

Lemma 2.20 (Solovay's shortening lemma [So76b]). *Suppose that $T \supseteq \mathrm{I}\Delta_0$, and let I be a T-cut. Then there is a formula K such that*

(1) $T \vdash \forall x(K(x) \rightarrow I(x))$;

(2) *K is a T-initial segment.*

Proof. First construct J from I as in Lemma 2.19. Next define

$$K(x) :\leftrightarrow J(x) \wedge \forall y(J(y) \rightarrow J(x \cdot y)).$$

We leave it to the reader to prove that K is indeed the desired T-initial segment. □

The following Lemma 2.21 is used in almost all applications of cuts. Note that it is essential that we use the efficient numerals $\overline{x}$ which are based on the binary expansion of x.

Lemma 2.21 (Pudlák). *Suppose J is a T-initial segment, where the set of axioms of T is given by the formula α. Then there is a polynomial P such that, for all n, $T \vdash J(\overline{n})$ by a proof of length $\leq P(|n|)$. Also we have $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, \mathrm{Prov}_\alpha(\ulcorner J(\overline{x})\urcorner)$.*

Proof. We give only a sketch, and leave the formal details to the reader. Essentially, in the proof of $J(\overline{x})$, we follow the $|x|$ steps it takes to build $\overline{x}$ from $\overline{0}$. At every step we instantiate either the proof of $\forall y(J(y) \rightarrow J(Sy))$ or the proof of $\forall y(J(y) \rightarrow J(SS0 \cdot y))$ with the appropriate efficient numeral. By using Modus Ponens a total of $|x|$ times, we finally derive $J(\overline{x})$. The length of the proof can evidently be bounded by a polynomial in $|x|$.

By inspection of the proof we see that it can be formalized to get $\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, \mathrm{Prov}_\alpha(\ulcorner J(\overline{x})\urcorner)$. Also, it is useful to remark that in the proofs of $J(\overline{x})$, only formulas of a fixed complexity depending only on J are used. □

§3. Σ-completeness and the NP = co-NP problem.

In this section, we will prove that, under the assumption that NP $\neq$ co-NP, the following holds:

$$\mathrm{I}\Delta_0 + \Omega_1 \nvdash \forall b \forall c\ (\exists a(\mathrm{Prf}(a,c) \wedge \forall z \leq a \neg \mathrm{Prf}(z,b)) \\ \rightarrow \mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(a,\overline{c}) \wedge \forall z \leq a \neg \mathrm{Prf}(z,\overline{b}))\urcorner)).$$

In the proofs of the lemmas leading up to this result, we will frequently, often without mention, make use of the following proposition and its corollary.

Proposition 3.1 ([Bu86]). *Suppose A is a closed, bounded formula in the the language of S^1_2, and let $\mathbf{R}$ be a consistent theory extending S^1_2. Then $\mathbf{R} \vdash A$ iff $\omega \models A$.*

Corollary 3.2 ([Bu86, Proposition 8.3]). *Suppose $A(\vec{a})$ is a bounded formula in the language of S^1_2, and let $\mathbf{R}$ be a consistent theory extending S^1_2. If $\mathbf{R} \vdash \forall \vec{x} A(\vec{x})$, then $\omega \models \forall \vec{x} A(\vec{x})$.*

In this section, we will use the name $\mathrm{I}\Delta_0 + \Omega_1$ for Buss' theory S_2 (see Definition 2.12) in which induction for formulas from the hierarchy of bounded arithmetic formulas in a language containing $|\ |, \lfloor \frac{1}{2}x \rfloor$, and # is allowed. Because S_2 is conservative over $\mathrm{I}\Delta_0 + \Omega_1$, the name change has no repercussions on the results of this section. (In the next section, where we need to construct formalized satisfaction predicates, we will be more careful.)

In order to prove the main theorem of this section, we need to prove a few seemingly far-fetched lemmas. Their proofs borrow heavily from the formalization carried out in [Bu86]. To make these lemmas understandable, we will give some details of the formalization of the predicate Prf. Buss uses a sequent calculus akin to Takeuti's (see [Ta75]). He considers a proof to be formalized as a tree, of which the root corresponds to the end sequent, and the leaves to the initial sequents of the proof. Every node of the proof tree is labeled by an ordered pair $\langle a, b\rangle$. The second member of this pair codes a sequent, and the first member codes the rule of inference by which this sequent has been derived from the sequents corresponding to the children of the node in question. For leaves, the first member of the corresponding ordered pair codes the axiom of which the initial sequent is an instantiation.

The only extra fact we need here is that logical axioms are all numbered 0; in particular, for all terms t, the tree containing just one node labeled $\langle 0, \ulcorner\rightarrow t = t\urcorner\rangle$ is a proof of $\rightarrow t = t$. Because of a peculiarity in the encoding of trees, by which 0 and 1 are reserved as codes for brackets, Buss encodes the proof just mentioned by $\langle 0, \ulcorner\rightarrow t = t\urcorner\rangle + 2$.

In the sequel, we will sometimes abuse Buss' conventions in order to keep the formulas legible. Thus, we will write

$$\langle 0, \ulcorner\rightarrow \overline{d} = \overline{d}\urcorner\rangle$$

for Buss' $\langle 0, (0 * \overline{\text{Arrow}}) * *(\ulcorner I_d\urcorner * \overline{\text{Equals}} * *\ulcorner I_d\urcorner)\rangle + 2$.

LEMMA 3.3. *Let $\psi(d, b)$ be the formula $\forall z \leq \langle 0, \ulcorner\rightarrow \overline{d} = \overline{d}\urcorner\rangle \neg \mathrm{Prf}(z, b)$. The predicate represented by ψ is co-NP-complete.*

PROOF. Straightforwardly, ψ is a Π^b_1-formula; hence, it represents a co-NP predicate. For the other side, viz. co-NP-hardness, begin by taking $A(a_1, \dots, a_k) \in$ co-NP. We will polynomially reduce A to ψ. (For definitions of the complexity theoretic concepts that we mention, see Definition 2.7 and Definition 2.8; and see Remark 2.10 or [Bu 86, Theorem 1.8]).

By provable Σ^b_1-completeness (see Theorem 2.15), there is a term $r(\vec{a})$ such that

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \neg A(\vec{a}) \rightarrow \exists z \leq r(\vec{a})\, \mathrm{Prf}(z, \ulcorner\neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner).$$

and thus,

$$\omega \models \neg A(\vec{a}) \rightarrow \exists z \leq r(\vec{a})\, \mathrm{Prf}(z, \ulcorner\neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner).$$

Because $r(\vec{a}) \leq \ulcorner\overline{r(\vec{a})}\urcorner \leq \langle 0, \ulcorner\rightarrow \overline{r(\vec{a})} = \overline{r(\vec{a})}\urcorner\rangle$ we also have

$$(1) \qquad \omega \models \neg A(\vec{a}) \rightarrow \exists z \leq \langle 0, \ulcorner\rightarrow \overline{r(\vec{a})} = \overline{r(\vec{a})}\urcorner\rangle\, \mathrm{Prf}(z, \ulcorner\neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner).$$

On the other hand, by Proposition 3.1 and the consistency of $\mathrm{I}\Delta_0 + \Omega_1$, we have

$$(2) \qquad \omega \models \exists z \leq \langle 0, \ulcorner\rightarrow \overline{r(\vec{a})} = \overline{r(\vec{a})}\urcorner\rangle\, \mathrm{Prf}(z, \ulcorner\neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner) \rightarrow \neg A(\vec{a}).$$

From (1) and (2), we conclude that

$$\omega \models A(\vec{a}) \leftrightarrow \forall z \leq \langle 0, \ulcorner\rightarrow \overline{r(\vec{a})} = \overline{r(\vec{a})}\urcorner\rangle \neg \mathrm{Prf}(z, \ulcorner\neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner).$$

This means by the definition of ψ that

$$\omega \models A(\vec{a}) \leftrightarrow \psi(r(\vec{a}), \ulcorner \neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner).$$

As both $\ulcorner \neg A(\overline{a_1}, \dots, \overline{a_k})\urcorner$ and $r(\vec{a})$ can be computed from $\vec{a}$ by polynomial time functions, we have reduced the co-NP predicate A to ψ. □

LEMMA 3.4. *Let $B(a_1, \dots, a_k)$ be a Π_1^b-formula representing a co-NP complete predicate. If NP $\neq$ co-NP, then*

$$\mathrm{I}\Delta_0 + \Omega_1 \nvdash \forall \vec{a}(B(\vec{a}) \to \mathrm{Prov}(\ulcorner B(\overline{a_1}, \dots . \overline{a_k})\urcorner)).$$

PROOF. An application of Parikh's Theorem for $\mathrm{I}\Delta_0 + \Omega_1$ (cf. Theorem 2.13). We leave the details, which are similar to part of the proof of [Bu86, Theorem 8.6], to the reader. □

LEMMA 3.5. *If NP $\neq$ co-NP, then*

$$\mathrm{I}\Delta_0 + \Omega_1 \nvdash \forall b \forall d\ (\forall z \leq \langle 0, \ulcorner \to \overline{d} = \overline{d}\urcorner\rangle \neg \mathrm{Prf}(z, \overline{b}) \\ \to \mathrm{Prov}(\ulcorner \forall z \leq \langle 0, \ulcorner \to \overline{d} = \overline{d}\urcorner\rangle \neg \mathrm{Prf}(z, \overline{b})\urcorner)).$$

PROOF. Directly from Lemma 3.3 and Lemma 3.4. □

LEMMA 3.6. $\mathrm{I}\Delta_0 + \Omega_1$ *proves the following*:

$$\forall b \forall d\ (\mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(a. \ulcorner \to \overline{d} = \overline{d}\urcorner) \wedge \forall z \leq a \neg \mathrm{Prf}(z, \overline{b}))\urcorner) \\ \to \mathrm{Prov}(\ulcorner \forall z \leq \langle 0, \ulcorner \to \overline{d} = \overline{d}\urcorner\rangle \neg \mathrm{Prf}(z, \overline{b})\urcorner)).$$

PROOF. It is not difficult to see that for Buss' formalization of Prf, we have the following:

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall d \forall a(\mathrm{Prf}(a, \ulcorner \to \overline{d} = \overline{d}\urcorner) \to a \geq \langle 0, \ulcorner \to \overline{d} = \overline{d}\urcorner\rangle),$$

and thus,

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall b \forall d\ (\exists a(\mathrm{Prf}(a, \ulcorner \to \overline{d} = \overline{d}\urcorner) \wedge \forall z \leq a \neg \mathrm{Prf}(z, b)) \\ \to \forall z \leq \langle 0, \ulcorner \to \overline{d} = \overline{d}\urcorner\rangle \neg \mathrm{Prf}(z, b)).$$

This in turn immediately implies our lemma. □

THEOREM 3.7. *IF NP $\neq$ co-NP, then*

$$\mathrm{I}\Delta_0 + \Omega_1 \nvdash \forall b \forall c\ (\exists a(\mathrm{Prf}(a, c) \wedge \forall z \leq a \neg \mathrm{Prf}(z, b)) \\ \to \mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(a, \overline{c}) \wedge \forall z \leq a \neg \mathrm{Prf}(z, \overline{b}))\urcorner)).$$

PROOF. Suppose that NP $\neq$ co-NP, and suppose, in order to derive a contradiction, that

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall b \forall c\ (\exists a(\mathrm{Prf}(a, c) \wedge \forall z \leq a \neg \mathrm{Prf}(z. b)) \\ \to \mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(a, \overline{c}) \wedge \forall z \leq a \neg \mathrm{Prf}(z, \overline{b}))\urcorner)).$$

Then, in particular,

$$\begin{aligned} \mathrm{I}\Delta_0+\Omega_1 \vdash \forall b \forall d\ (&\mathrm{Prf}(\langle 0, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner\rangle, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner) \\ &\wedge \forall z \leq \langle 0, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner\rangle \neg \mathrm{Prf}(z,b) \\ &\rightarrow \mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(z, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner) \wedge \forall z \leq a \neg \mathrm{Prf}(z,\overline{b}))\urcorner)). \end{aligned} \tag{3}$$

We know that

$$\mathrm{I}\Delta_0+\Omega_1 \vdash \forall d(\mathrm{Prf}(\langle 0, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner\rangle, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner)).$$

Combined with (3), this implies the following:

$$\begin{aligned} \mathrm{I}\Delta_0+\Omega_1 \vdash \forall b \forall d\ (&\forall z \leq \langle 0, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner\rangle \neg \mathrm{Prf}(z,b) \\ &\rightarrow \mathrm{Prov}(\ulcorner \exists a(\mathrm{Prf}(a, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner) \wedge \forall z \leq a \neg \mathrm{Prf}(z,\overline{b}))\urcorner)). \end{aligned}$$

Now we apply Lemma 3.6 to derive

$$\begin{aligned} \mathrm{I}\Delta_0+\Omega_1 \vdash \forall b \forall d\ (&\forall z \leq \langle 0, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner\rangle \neg \mathrm{Prf}(z,b) \\ &\rightarrow \mathrm{Prov}(\ulcorner \forall z \leq \langle 0, \ulcorner\rightarrow \overline{d}=\overline{d}\urcorner\rangle \neg \mathrm{Prf}(z,\overline{b})\urcorner)), \end{aligned}$$

in contradiction with Lemma 3.5. □

REMARK 3.8. We can prove that provable Σ_1^0-completeness fails already for a much simpler Π_1^b-formula $\chi(a,b,c)$ defined as follows:

$$\chi(a,b,c) := \forall x \leq c \forall y \leq c(a \cdot x^2 + b \cdot y \neq c).$$

The fact that Σ_1^0-completeness fails for χ follows immediately from Lemma 3.4 and the following lemma, to which A. Wilkie attracted our attention.

LEMMA 3.9 (Manders and Adleman, see [MA 78]). *The set of equations of the form* $(a \cdot x^2 + b \cdot y = c)$, *solvable over the natural numbers, with* a, b, c *positive natural numbers, is NP-complete.*

Note that Lemma 3.9 implies that the formula $\exists x \leq c \exists y \leq c(a \cdot x^2 + b \cdot y = c)$ represents an NP-complete predicate, and thus that χ as defined above represents a co-NP complete predicate.

§**4. The small reflection principle.** In this section, we will present a proof of the fact that $\mathrm{I}\Delta_0+\Omega_1$ proves the small reflection principle, i.e., for all φ:

$$\mathrm{I}\Delta_0+\Omega_1 \vdash \forall x \Box(\Box_x \varphi \rightarrow \varphi),$$

where $\Box\varphi$ is an abbreviation for $\mathrm{Prov}(\ulcorner\varphi\urcorner)$ and $\Box_x\varphi$ is a formalization of the fact that φ has a proof in $\mathrm{I}\Delta_0+\Omega_1$ of Gödel number $\leq x$. In fact, all arguments that we use can be carried out already in Buss' S_2^1, as the reader may check for him/herself.

In the proof, we will use the existence of partial truth- (or satisfaction-) predicates Sat_n for formulas of length $\leq n$. The intended meaning of $\mathrm{Sat}_n(x,w)$ will be "the formula of length $\leq n$ with Gödel number x is satisfied by the assignment sequence coded by w". Pudlák [Pu86] has constructed partial truth predicates much like the ones we need. (An analogous construction, where Sat_n is related to quantifier depth instead of length, can be found in [Pu87].)

However, our construction departs from Pudlák's in two ways. Firstly, whereas Pudlák presents his results for theories in relational languages, we allow function symbols.

Secondly and more importantly, $I\Delta_0 + \Omega_1$ is neither finitely nor sparsely axiomatized. Regrettably we cannot even apply to $I\Delta_0 + \Omega_1$ a trick of Pudlák's which turns some nonsparse theories like PA and ZF into sparse ones (see Theorem 5.5 of [Pu86]). Therefore, we introduce new satisfaction predicates $\mathrm{Sat}_{n.\Delta}(x, w)$ with as intended meaning: "the Δ_0-formula of length $\leq n$ with Gödel number x is satisfied by the assignment sequence coded by w". Using these satisfaction predicates, we will be able to prove by short proofs that the Δ_0-induction axioms are true.

In order to start the construction of short satisfaction predicates, we need a few more assumptions and definitions. First of all, when formalizing, we view $I\Delta_0 + \Omega_1$ in a restricted way more akin to Paris and Wilkie [WP87] than to Buss [Bu86]: see Definition 2.1.

For this system, we can define the appropriate Δ_1^b-predicates $\mathrm{Term}(v)$, $\mathrm{Fmla}(v)$, $\mathrm{Sent}(v)$, $\mathrm{Prf}(u, v)$ in S_2^1.

In Buss' formalization of sequences, $*$ stands for a function which adds a new element to the end of a sequence; $**$ stands for a function which concatenates two sequences; and $\beta(t, w)$ stands for the function giving the value of the tth place in the sequence coded by w.

In this paper, we denote concatenation of sequences sloppily by juxtaposition, and we leave our some outer parentheses; thus, for example, $y\ulcorner\rightarrow\urcorner z$ stands for Buss' $(0 * \overline{\mathrm{LParen}}) * *(y * \overline{\mathrm{Implies}}) * *(z * \overline{\mathrm{RParen}})$.

Definition 4.1. We formally define four concepts that we need in order to construct truth predicates.

- $w =_i w' := \mathrm{Len}(w) = \mathrm{Len}(w') \wedge \forall t(t \leq \mathrm{Len}(w) \wedge t \neq i \rightarrow \beta(t, w) = \beta(t, w'))$, i.e., the only possible difference between the sequences coded by w and w' is at the ith value,
- $\mathrm{Fmla}_n(v) := \mathrm{Fmla}(v) \wedge \mathrm{Len}(v) \leq n$, i.e., v is the Gödel number of a formula of length $\leq n$;
- $\mathrm{Fmla}_{n.\Delta}(v) := \mathrm{Fmla}_n(v)$ "and v codes a Δ_0-formula";
- $\mathrm{Evalseq}(x, w)$ will mean that the sequence coded by w is long enough to evaluate all variables appearing in x, i.e.,

$$\begin{aligned}\mathrm{Evalseq}(x, w) := {} & \mathrm{Seq}(w) \wedge (\mathrm{Fmla}(x) \vee \mathrm{Term}(x)) \\ & \wedge \forall i \text{ (``the variable } v_i \text{ occurs in the term or formula} \\ & \quad \text{with Gödel number } x\text{''} \rightarrow \mathrm{Len}(w) \geq i).\end{aligned}$$

Furthermore, we introduce the following two abbreviations:

- $\mathrm{Evalseq}_n(x, w) := \mathrm{Fmla}_n(x) \wedge \mathrm{Evalseq}(x, w)$;
- $\mathrm{Evalseq}_{n.\Delta}(x, w) := \mathrm{Fmla}_{n.\Delta}(x) \wedge \mathrm{Evalseq}(x, w)$.

Next we define, by Buss' method of p-inductive definitions, a function Val such that if $t(v_{i_1}, \dots v_{i_n})$ is a term of the (restricted) language of $I\Delta_0 + \Omega_1$ and w codes a sequence evaluating all variables $v_{i_1}, \dots v_{i_n}$ appearing in t, then $\mathrm{Val}(\ulcorner t\urcorner, w)$ gives the value of $t[\beta(i_1, w), \dots, \beta(i_n, w)]$.

DEFINITION 4.2. Let Val satisfy the following conditions:

- $\neg\,\mathrm{Term}(t) \vee \neg\,\mathrm{Evalseq}(t,w) \rightarrow \mathrm{Val}(t,w) = 0$;
- the p-inductive condition:

$$
\begin{aligned}
\mathrm{Term}(t) \wedge \mathrm{Evalseq}(t,w) \rightarrow (t = \ulcorner 0 \urcorner \wedge \mathrm{Val}(t,w) = 0) \\
\vee\, \exists i < t(t = \ulcorner v_i \urcorner \wedge \mathrm{Val}(t,w) = \beta(i,w)) \\
\vee\, \exists t_1, t_2 < t(\mathrm{Term}(t_1) \wedge \mathrm{Term}(t_2) \\
\wedge\, ((t = \ulcorner S \urcorner t_1 \wedge \mathrm{Val}(t,w) = S(\mathrm{Val}(t_1,w))) \\
\vee\, (t = t_1 \ulcorner + \urcorner t_2 \wedge \mathrm{Val}(t,w) = \mathrm{Val}(t_1,w) + \mathrm{Val}(t_2,w)) \\
\vee\, (t = t_1 \ulcorner \cdot \urcorner t_2 \wedge \mathrm{Val}(t,w) = \mathrm{Val}(t_1,w) \cdot \mathrm{Val}(t_2,w)))).
\end{aligned}
$$

By induction, we can show that $t\#w$ will be a bound for $\mathrm{Val}(t,w)$. Thus, by [Bu86, Theorem 7.3], Val is Δ_1^b-definable (thus, provably total) in S_2^1; furthermore, the definition of Val in S_2^1 is intensionally correct in that properties of Val can be proved in S_2^1 (and thus also in $I\Delta_0 + \Omega_1$) by the use of induction.

REMARK 4.3. Note that we cannot construct in $I\Delta_0 + \Omega_1$ a correct valuation function Val for a language that contains #. Indeed, to any a we can associate a formalized ferm $f(a)$ given informally as $1\#2\#\cdots\#2$, where the number of 2's is $|a|$. A correctly defined Val should give $\mathrm{Val}(f(a),w) = \exp(\exp(|a|+1) - 2) \geq \exp(a)$ (cf. [Ta88]). Therefore, by Parikh's Theorem (cf. Theorem 2.13), Val could not be Δ_0-definable and provably total in $I\Delta_0 + \Omega_1$.

In the sequel, we will freely make use of induction for $\Delta_0(\mathrm{Val})$-formulas in $I\Delta_0 + \Omega_1$, as is justified by the $I\Delta_0 + \Omega_1$-analogs of Buss' Theorem 2.2 and Corollary 2.3. We will especially need the following lemma.

LEMMA 4.4. *There exists a constant c such that for every term t with free variables among* $v_{i_1}, \ldots, v_{i_m}$ *and for every n with* $\mathrm{Len}(\ulcorner t \urcorner) \leq n$, *we can prove the following by proofs of length* $\leq c \cdot n$:

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}(\ulcorner t \urcorner, w) \rightarrow \mathrm{Val}(\ulcorner t \urcorner, w) = t[\beta(i_1, w), \ldots, \beta(i_m, w)].$$

PROOF. Straightforward by induction on the construction of t. □

For the definition of satisfaction predicates, we need one more definition.

DEFINITION 4.5. We formally define the following:

$$s(i,x,w) := (\mathrm{Subseq}(w,1,i) * x) ** \mathrm{Subseq}(w, i+1, \mathrm{Len}(w)+1).$$

Thus, if w is a sequence of length $\geq i$, $s(i,x,w)$ denotes the sequence which is identical to w, except that x appears in the ith place.

DEFINITION 4.6. We say that $\mathrm{Sat}_n(x, w)$ is a *partial definition of truth for formulas of length* $\leq n$ in $\mathrm{I}\Delta_0 + \Omega_1$ iff

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(x, w) \to \{\mathrm{Sat}_n(x, w) \leftrightarrow$$
$$[\exists t, t' < x(\mathrm{Term}(t) \wedge (\mathrm{Term}(t') \wedge x = t \ulcorner = \urcorner t' \wedge \mathrm{Val}(t, w) = \mathrm{Val}(t', w))$$
$$\vee \exists t, t' < x(\mathrm{Term}(t) \wedge \mathrm{Term}(t') \wedge x = t \ulcorner \leq \urcorner t' \wedge \mathrm{Val}(t, w) \leq \mathrm{Val}(t', w))$$
$$\vee \exists y < x(x = \ulcorner \neg \urcorner y \wedge \neg \mathrm{Sat}_n(y, w))$$
$$\vee \exists y, z < x(x = y \ulcorner \to \urcorner z \wedge (\mathrm{Sat}_n(y, w) \to \mathrm{Sat}_n(z, w)))$$
$$\vee \exists y, i < x(x = \ulcorner \forall v_i \urcorner y \wedge \forall w'(w =_i w' \to \mathrm{Sat}_n(y, w')))$$
$$\vee \exists y, i, t < x(\mathrm{Term}(t) \wedge x = \ulcorner (\forall v_i \leq \urcorner t \ulcorner) \urcorner y$$
$$\wedge \forall w' \leq s(i, \mathrm{Val}(t, w), w)(w =_i w' \wedge \beta(i, w) \leq \mathrm{Val}(t, w) \to \mathrm{Sat}_n(y, w')))]\}.$$

We denote the part between brackets [ ] on the right-hand side of the equivalence by $\Sigma(\mathrm{Sat}_n; x, w)$; note that these are just Tarski's conditions.

Similarly, we say that $\mathrm{Sat}_{n.\Delta}(x, w)$ is a *partial definition of truth for* Δ_0*-formulas of length* $\leq n$ in $\mathrm{I}\Delta_0 + \Omega_1$ iff

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_{n.\Delta}(x, w) \to \{\mathrm{Sat}_{n.\Delta}(x, w) \leftrightarrow$$
$$[\exists t, t' < x(\mathrm{Term}(t) \wedge \mathrm{Term}(t') \wedge x = t \ulcorner = \urcorner t' \wedge \mathrm{Val}(t, w) = \mathrm{Val}(t', w))$$
$$\vee \exists t, t' < x(\mathrm{Term}(t) \wedge \mathrm{Term}(t') \wedge x = t \ulcorner \leq \urcorner t' \wedge \mathrm{Val}(t, w) \leq \mathrm{Val}(t', w))$$
$$\vee \exists y < x(x = \ulcorner \neg \urcorner y \wedge \neg \mathrm{Sat}_{n.\Delta}(y, w))$$
$$\vee \exists y, z < x(x = y \ulcorner \to \urcorner z \wedge (\mathrm{Sat}_{n.\Delta}(y, w) \to \mathrm{Sat}_{n.\Delta}(z, w)))$$
$$\vee \exists y, i, t < x(\mathrm{Term}(t) \wedge x = \ulcorner (\forall v_i \leq \urcorner t \ulcorner) \urcorner y$$
$$\wedge \forall w' \leq s(i, \mathrm{Val}(t, w), w)(w =_i w' \wedge \beta(i, w) \leq \mathrm{Val}(t, w) \to \mathrm{Sat}_{n.\Delta}(y, w')))]\}.$$

We denote the part between brackets [ ] on the right-hand side of the equivalence by $\Sigma_\Delta(\mathrm{Sat}_{n.\Delta}; x, w)$. Note that the only difference between $\Sigma(\mathrm{Sat}_n; x, w)$ and $\Sigma_\Delta(\mathrm{Sat}_{n.\Delta}; x, w)$ is that in the latter, the disjunct for the unbounded quantifier $\forall$ is left out.

In the proof of the main theorem of this section, we will reason inside $\mathrm{I}\Delta_0 + \Omega_1$, and we will need the existence of Gödel numbers representing formulas Sat_n that provably satisfy the conditions of the preceding definition. Therefore, in the unformalized proofs below, we take care that the formulas Sat_n and the proofs that they have the right properties be bounded by suitable terms. The following lemmas provide us with such formulas. In [Pu86], [Pu87] Pudlák proves similar lemmas for a language without function symbols. Below, we sketch the adaptation of his method to our case. The parallel construction of a $\Delta_0(\mathrm{Val}, |\ |, \llcorner \frac{1}{2}x \lrcorner, \#)$-formula $\mathrm{Sat}_{n.\Delta}$ which works for Δ_0-formulas is particular to this paper. We use the formula $\mathrm{Sat}_{n.\Delta}$ only in our proof that Sat_n preserves the Δ_0-induction axioms, but there its use is essential.

LEMMA 4.7. *There exist formulas* $\text{Sat}_n(x, w)$ *for* $n = 0, 1, 2, \ldots$ *of length linear in* n, *and such that, by a proof of length linear in* n,

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \text{Evalseq}_{n+1}(x, w) \rightarrow (\text{Sat}_{n+1}(x, w) \leftrightarrow \Sigma(\text{Sat}_n; x, w)).$$

PROOF. Sat_n is constructed by recursion. We can define Sat_0 arbitrarily, as there are no formulas of length ≤ 0. If we have the formula Sat_k, we obtain Sat_{k+1} by substituting Sat_k for Sat_n in the formula $\Sigma(\text{Sat}_n; x, w)$ defined in Definition 4.6.

Remember that we have to ensure that the length of the formula Sat_n grows linearly in n. However, if we straightforwardly used $\Sigma(\text{Sat}_n; x, w)$ as defined above, the length of Sat_n would grow exponentially in n, because $\Sigma(\text{Sat}_n; x, w)$ contains more than one occurrence of Sat_n.

Ferrante and Rackoff (in [FR79, Chapter 7]) describe a general technique for writing short formulas, due to Fischer and Rabin. Using these techniques, one can replace $\Sigma(\text{Sat}_n; x, w)$ by a formula $\Sigma'(\text{Sat}_n; x, w)$ which contains only one occurrence of Sat_n and which is equivalent to $\Sigma(\text{Sat}_n; x, w)$ in a very weak theory—say predicate logic plus the axiom $S0 \neq 0$.

Ferrante and Rackoff use the inclusion of $\leftrightarrow$ in the language of the theory in an essential way. However, Solovay sent us a different construction of short formulas which circumvents the use of $\leftrightarrow$. With his kind permission, we present a sketch of his proof.

Solovay's basic idea is to shift attention from sets to characteristic functions. Without restriction of generality, we may assume that we work with unary predicates $\text{Sat}_n(x)$ instead of $\text{Sat}_n(x, w)$. Let

$$F_n(x, y) := (y = S0 \wedge \text{Sat}_n(x)) \vee (y = 0 \wedge \neg \text{Sat}_n(x)).$$

If we can find a formula H_n equivalent to F_n of length proportional to n, it will be easy to define using this formula our desired formula Sat_{n+1}.

Let L be the language of $\mathrm{I}\Delta_0 + \Omega_1$ enriched with a new binary predicate letter G. We can find a formula Φ of L in prenex normal form, having only the variables x and y free, such that if G is interpreted as F_n, then Φ is interpreted as F_{n+1}. We show how to find a formula Ψ which is equivalent to Φ and which has only one occurrence of G. Assume that Φ starts with the string of quantifiers $(Q_1 x_1) \ldots (Q_r x_r)$ and that there are k occurrences of G in the matrix of Φ, say $G(t_1, m_1), \ldots, G(t_k, m_k)$. The formula Ψ will have the form

$$(Q_1 x_1) \cdots (Q_r x_r)(\exists y_1) \cdots (\exists y_k)[M \wedge S].$$

Here $y_1, \ldots, y_k$ are fresh variables (for the moment—in the final definition we will be less liberal with variables). The formula M is obtained from the matrix of Φ by replacing each occurrence of $G(t_i, m_i)$ by $m_i = y_i$. The job of S is to ensure that the y_i's are chosen correctly. It is defined as follows.

$$S := \forall w_1 \exists w_2 \left[G(w_1, w_2) \wedge \bigwedge_{i=1}^{k} (w_1 = t_i \rightarrow w_2 = y_i) \right].$$

If we define H_{n+1} from H_n using Ψ, then we get a formula of length proportional to $n \log n$, because at every step we introduce fresh variables in order to avoid clashes. There are, however, tricks to get by with a finite set of variables, as the reader may enjoy figuring out (or look up in [FR79, Chapter 7]).

We will write $\Sigma'(\mathrm{Sat}_n; x, w)$ for the equivalent of $\Sigma(\mathrm{Sat}_n; x, w)$ resulting from an application of the techniques described above. The length of Sat_n thus constructed via iterated application of Σ' to Sat_0 is indeed linear in n. Moreover, for all n the *shape* of the proof of $\Sigma(\mathrm{Sat}_n; x, w) \leftrightarrow \Sigma'(\mathrm{Sat}_n; x, w)$ is the same. Thus, the proofs of $\Sigma(\mathrm{Sat}_n; x, w) \leftrightarrow \Sigma'(\mathrm{Sat}_n; x, w)$ grow linearly in n. Hence, as $\mathrm{Sat}_{n+1}(x, w) \equiv \Sigma'(\mathrm{Sat}_n; x, w)$, we have the following by proofs of length linear in n:

$$\text{(4)} \qquad \mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Sat}_{n+1}(x, w) \leftrightarrow \Sigma(\mathrm{Sat}_n; x, w) \qquad \square$$

Lemma 4.8. $\mathrm{I}\Delta_0 + \Omega_1$ *proves, by a proof of length of the order of* n^2, *that the formula* Sat_n *as constructed in Lemma* 4.7 *is a partial definition of truth for formulas of length* $\leq n$.

Proof. We want short proofs showing that Sat_n is a partial definition of truth for formulas of length $\leq n$ in $\mathrm{I}\Delta_0 + \Omega_1$, i.e.,

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(x, w) \rightarrow (\mathrm{Sat}_n(x, w) \leftrightarrow \Sigma(\mathrm{Sat}_n; x, w)).$$

By (4), it suffices to show that, by proofs of length of the order n^2,

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(x, w) \rightarrow (\mathrm{Sat}_n(x, w) \leftrightarrow \mathrm{Sat}_{n+1}(x, w)).$$

This can be proved by external induction on n. In fact, when we define

$$\Phi_n := \forall x \forall x (\mathrm{Evalseq}_n(x, w) \rightarrow (\mathrm{Sat}_n(x, w) \leftrightarrow \mathrm{Sat}_{n+1}(x, w))),$$

the proofs of $\Phi_n \rightarrow \Phi_{n+1}$ in $\mathrm{I}\Delta_0 + \Omega_1$ will have a shape which does not depend on n. (We refer those readers who seek elucidation by examples to [Pu86, Lemma 5.1].) We can observe that every proof in $\mathrm{I}\Delta_0 + \Omega_1$ of $\Phi_n \rightarrow \Phi_{n+1}$ is the instantiation of a single proof scheme. Thus, the length of the proofs of $\Phi_n \rightarrow \Phi_{n+1}$ increases only linearly in n, so that the length of the proof in $\mathrm{I}\Delta_0 + \Omega_1$ of

$$\forall x \forall w (\mathrm{Evalseq}_n(x, w) \rightarrow (\mathrm{Sat}_n(x, w) \leftrightarrow \mathrm{Sat}_{n+1}(x, w)))$$

is of the order n^2. $\square$

Lemma 4.9. *There exist formulas* $\mathrm{Sat}_{n.\Delta}(x, w)$ *for* $n = 0, 1, 2, \ldots$ *of lengths linear in* n, *and such that* $\mathrm{I}\Delta_0 + \Omega_1$ *proves, by proofs of length linear in* n, *that* $\mathrm{Sat}_{n+1.\Delta}(x, w) \leftrightarrow \Sigma_\Delta(\mathrm{Sat}_{n.\Delta}; x, w)$. *The resulting formulas* $\mathrm{Sat}_{n.\Delta}(x, w)$ *are* $\Delta_0(\mathrm{Val})$-*formulas.*

Proof. The proof is completely analogous to the proof of Lemma 4.7. Because $\Sigma_\Delta(\mathrm{Sat}_{n.\Delta}; x, w)$ contains only bounded quantifiers, and because all quantifiers introduced by the Solovay method can be bounded, the resulting formulas are indeed $\Delta_0(\mathrm{Val})$. $\square$

Lemma 4.10. $\mathrm{I}\Delta_0 + \Omega_1$ *proves by a proof of length of the order of* n^2 *that the formula* $\mathrm{Sat}_{n.\Delta}(x, w)$ *as constructed in Lemma* 4.9 *is a partial definition of truth for* Δ_0-*formulas of length* $\leq n$.

PROOF. We adapt the proof of Lemma 4.8, incorporating the fact that we are concerned with Δ_0-formulas only. Thus, instead of Φ_n, we define

$$\Phi_{n.\Delta} := \forall x \forall w (\text{Evalseq}_{n.\Delta}(x, w) \rightarrow (\text{Sat}_{n.\Delta}(x, w) \leftrightarrow \text{Sat}_{n+1.\Delta}(x, w))).$$

The proof of $\Phi_{n.\Delta} \rightarrow \Phi_{n+1.\Delta}$ runs along the same lines as the proof of $\Phi_n \rightarrow \Phi_{n+1}$, using the extra fact that if $x = y\ulcorner\rightarrow\urcorner z$ and $\text{Fmla}_{n+1.\Delta}(x)$, then $\text{Fmla}_{n.\Delta}(y)$ and $\text{Fmla}_{n.\Delta}(z)$, etc. □

We now show that the partial definitions of truth can, by proofs of quadratic length, be proven to satisfy Tarski's conditions, which justifies their name.

LEMMA 4.11 (cf.[Pu86], [Pu87]). *There exists a constant c such that for every formula φ with free variables among $v_{i_1}, \ldots, v_{i_m}$ and for every n with* $\text{Len}(\ulcorner\varphi\urcorner) \leq n$, *we can prove the following by proofs of length $\leq c \cdot n^2$:*

$$\begin{aligned} &\text{I}\Delta_0 + \Omega_1 \vdash \forall w(\text{Evalseq}(\ulcorner\varphi\urcorner, w) \\ &\qquad \rightarrow (\text{Sat}_n(\ulcorner\varphi\urcorner, w) \leftrightarrow \varphi[\beta(i_1, w), \ldots, \beta(i_m, w)])), \end{aligned} \tag{5}$$

and if φ is a Δ_0-formula, then we can also prove the following by proofs of length $\leq c \cdot n^2$:

$$\begin{aligned} &\text{I}\Delta_0 + \Omega_1 \vdash \forall w(\text{Evalseq}(\ulcorner\varphi\urcorner, w) \\ &\qquad \rightarrow (\text{Sat}_{n.\Delta}(\ulcorner\varphi\urcorner, w) \leftrightarrow \varphi[\beta(i_1, w), \ldots, \beta(i_m, w)])). \end{aligned} \tag{6}$$

PROOF. By cases. If φ is an atomic formula $t \leq t'$ of length $\leq n$ and with free variables among $v_{i_1}, \ldots, v_{i_m}$, Lemma 4.8 implies that we can prove the following by proofs of length linear in n:

$$\begin{aligned} &\text{I}\Delta_0 + \Omega_1 \vdash \forall w(\text{Evalseq}(\ulcorner t \leq t'\urcorner, w) \\ &\qquad \rightarrow (\text{Sat}_n(\ulcorner t \leq t'\urcorner, w) \leftrightarrow \text{Val}(\ulcorner t\urcorner, w) \leq \text{Val}(\ulcorner t'\urcorner, w))). \end{aligned}$$

By Lemma 4.4, we can then conclude that we can prove the following by proofs of length linear in n:

$$\begin{aligned} &\text{I}\Delta_0 + \Omega_1 \vdash \forall w(\text{Evalseq}(\ulcorner t \leq t'\urcorner, w) \\ &\qquad \rightarrow (\text{Sat}_n(\ulcorner t \leq t'\urcorner, w) \leftrightarrow (t \leq t')[\beta(i_1, w), \ldots, \beta(i_m, w)])). \end{aligned}$$

The case for $t = t'$ is analogous.

For the nonatomic cases, we define

$$\Psi_k(\psi) := \forall w(\text{Evalseq}(\ulcorner\psi\urcorner, w) \rightarrow (\text{Sat}_k(\ulcorner\psi\urcorner, w) \leftrightarrow \psi[\beta(i_1, w), \ldots, \beta(i_m, w)])).$$

Every formula φ of length $\leq n$ is constructed from atomic formulas in at most n steps. Therefore, we would like to prove the following in $\text{I}\Delta_0 + \Omega_1$ by proofs of length linear in k:

(1) $\Psi_{k-1}(\psi) \rightarrow \Psi_k(\neg\psi)$ for $\text{Len}(\ulcorner\neg\psi\urcorner) \leq k$
(2) $\Psi_{k-1}(\psi) \wedge \Psi_{k-1}(\chi) \rightarrow \Psi_k(\psi \rightarrow \chi)$ for $\text{Len}(\ulcorner\psi \rightarrow \chi\urcorner) \leq k$;
(3) $\Psi_{k-1}(\psi) \rightarrow \Psi_k(\forall v_i \psi)$ for $\text{Len}(\ulcorner\forall v_i \psi\urcorner) \leq k$;
(4) $\Psi_{k-1}(\psi) \rightarrow \Psi_k((\forall v_i \leq t)\psi)$ for $\text{Len}(\ulcorner(\forall v_i \leq t)\psi\urcorner) \leq k$.

If we can find these short proofs, then we have for every formula φ of length $\leq n$ a proof of $\Psi_n(\varphi)$ of length of the order of n^2, and we are done. We will leave the easy proofs of the four cases to the reader. □

LEMMA 4.12. $I\Delta_0 + \Omega_1$ *proves by a proof of length of the order of* n^2 *that* Sat_n *preserves the logical rules* (*Modus Ponens and Generalization*) *for formulas of length* $\leq n$, *i.e.*,

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(y\ulcorner\rightarrow\urcorner z, w) \wedge \mathrm{Sat}_n(y, w) \wedge \mathrm{Sat}_n(y\ulcorner\rightarrow\urcorner z. w) \rightarrow \mathrm{Sat}_n(z. w)$$

and

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(\ulcorner\forall v_i\urcorner y, w) \wedge \forall w'(w =_i w' \rightarrow \mathrm{Sat}_n(y, w')) \\ \rightarrow \mathrm{Sat}_n(\ulcorner\forall v_i\urcorner y. w).$$

PROOF. The lemma follows immediately from Lemma 4.8. □

LEMMA 4.13. $I\Delta_0 + \Omega_1$ *proves by a proof of length of the order of* n^2 *that* Sat_n *preserves the logical axioms and the equality axioms for formulas of length* $\leq n$, *e.g., axiom scheme* (1) *of* [WP87]:

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(y\ulcorner\rightarrow(\urcorner z\ulcorner\rightarrow\urcorner y\ulcorner)\urcorner, w) \\ \rightarrow \mathrm{Sat}_n(y\ulcorner\rightarrow(\urcorner z\ulcorner\rightarrow\urcorner y\ulcorner)\urcorner, w). \qquad (\mathbf{PW1})$$

Similarly, the other propositional schemes (2) *and* (3) *are preserved. Corresponding to axiom schemes* (4), (5), *and* (6) *we have the following*:

(**PW4**)(*corresponding to axiom* (4) *of* [WP87]).

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(\ulcorner\forall v_i\urcorner y \rightarrow \mathrm{Sub}(y, \ulcorner v_i\urcorner, t), w) \wedge \mathrm{SubOK}(y. \ulcorner v_i\urcorner, t) \\ \rightarrow \mathrm{Sat}_n(\ulcorner\forall v_i\urcorner y \rightarrow \mathrm{Sub}(y, \ulcorner v_i\urcorner. t), w),$$

where $\mathrm{SubOK}(y, \ulcorner v_i\urcorner, t)$ *is Buss' formalization of "the term with Gödel number t is free for the variable* v_i *in the* (*term or*) *formula with Gödel number y".*

(**PW5**) (*corresponding to axiom* (5) *of* [WP87]).

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(\ulcorner\forall v_i(\urcorner y\ulcorner\rightarrow\urcorner z\ulcorner) \rightarrow (\urcorner y\ulcorner\rightarrow \forall v_i\urcorner z\ulcorner)\urcorner, w) \\ \wedge \text{“}v_i \text{ does not appear free in the formula with Gödel number } y\text{”} \\ \rightarrow \mathrm{Sat}_n(\ulcorner\forall v_i(\urcorner y\ulcorner\rightarrow\urcorner z\ulcorner) \rightarrow (\urcorner y\ulcorner\rightarrow \forall v_i\urcorner z\ulcorner)\urcorner, w).$$

(**PW6**) (*corresponding to axiom* (6) *of* [WP87]).

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(v_1\ulcorner=\urcorner v_1, w) \rightarrow \mathrm{Sat}_n(v_1\ulcorner=\urcorner v_1, w)$$

and

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(v_i\ulcorner=\urcorner v_j\ulcorner\rightarrow(\urcorner y\ulcorner\rightarrow\urcorner z\ulcorner)\urcorner, w) \\ \wedge \mathrm{SubOK}(y, \ulcorner v_i\urcorner. \ulcorner v_j\urcorner) \wedge \mathrm{Somesub}(z. y, \ulcorner v_i\urcorner. \ulcorner v_j\urcorner) \\ \rightarrow \mathrm{Sat}_n(v_i\ulcorner=\urcorner v_j\ulcorner\rightarrow(\urcorner y\ulcorner\rightarrow\urcorner z\ulcorner)\urcorner. w),$$

where $\mathrm{Somesub}(z, y, \ulcorner v_i\urcorner, \ulcorner v_j\urcorner)$ *is the formalization of "the formula with Gödel* *number z is the result of substituting the term v_j for some of the occurrences of v_i in the formula with Gödel number y".*

PROOF. For the propositional axiom schemes (PW1), (PW2), and (PW3), the results follow almost immediately from Lemma 4.8. For (PW4), we need proofs in $I\Delta_0 + \Omega_1$ of length of the order of n^2 of the following "call by name = call by value" lemma:

$$\begin{aligned}&\mathrm{Evalseq}_n(\ulcorner\forall v_i\urcorner y \to \mathrm{Sub}(y, \ulcorner v_i\urcorner, t), w)\\ &\qquad \wedge \mathrm{SubOK}(y, \ulcorner v_i\urcorner, t) \to [\mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_i\urcorner, t), w)\\ &\qquad \leftrightarrow \mathrm{Sat}_n(y, s(i, \mathrm{Val}(t, w), w))].\end{aligned}$$

This can be proved by induction on n, in a way similar to the proof of Lemma 4.8. The rest of (PW4) then follows by Lemma 4.8 itself.

For (PW5), we need proofs in $I\Delta_0 + \Omega_1$ of length of the order n^2 of the following:

$$\begin{aligned}&\mathrm{Evalseq}_n(\ulcorner\forall v_i(\urcorner y\ulcorner\to\urcorner z\ulcorner) \to (\urcorner y\ulcorner\to \forall v_i\urcorner z\ulcorner)\urcorner, w)\\ &\qquad \wedge \text{“}v_i \text{ does not appear free in the formula with Gödel number } y\text{”}\\ &\qquad \wedge w =_i w' \to [\mathrm{Sat}_n(y, w) \leftrightarrow \mathrm{Sat}_n(y, w')].\end{aligned}$$

This can also be proved by induction on n; again, the rest of (PW5) follows by Lemma 4.8.

The first equality axiom of (PW6) is proved immediately by Lemma 4.8. The second one has a proof similar to that of (PW4). □

LEMMA 4.14. $I\Delta_0 + \Omega_1$ *proves by a proof of length of the order of* n^2 *that* Sat_n *preserves the basic nonlogical axioms for formulas of length* $\leq n$, *e.g.,*

$$I\Delta_0 + \Omega_1 \vdash \mathrm{Evalseq}_n(\ulcorner 0 \leq 0 \wedge \neg S0 \leq 0\urcorner, w) \to \mathrm{Sat}_n(\ulcorner 0 \leq 0 \wedge \neg S0 \leq 0\urcorner, w).$$

Similarly for the other five basic axioms relating the symbols $0, S, +, \cdot$, *and* $\leq$ *of the language.*

PROOF. The lemma follows immediately by Lemma 4.8 and Lemma 4.4. □

LEMMA 4.15. $I\Delta_0 + \Omega_1$ *proves by a proof of length of the order of* n^2 *that* $\mathrm{Sat}_{n.\Delta}$ *agrees with* Sat_n *on* Δ_0*-formulas of length* $\leq n$, *i.e.,*

$$\mathrm{Evalseq}_{n.\Delta}(x, w) \to [\mathrm{Sat}_{n.\Delta}(x, w) \leftrightarrow \mathrm{Sat}_n(x, w)].$$

PROOF. The proof is by induction on n as in the proof of Lemma 4.10. Here, we take

$$\Phi_n := \forall x \forall w(\mathrm{Evalseq}_{n.\Delta}(x, w) \to (\mathrm{Sat}_{n.\Delta}(x, w) \leftrightarrow \mathrm{Sat}_n(x, w))).$$

As in Lemma 4.10, we use the fact that if $x = y\ulcorner\to\urcorner z$ and $\mathrm{Fmla}_{n+1.\Delta}(x)$, then $\mathrm{Fmla}_{n.\Delta}(y)$ and $\mathrm{Fmla}_{n.\Delta}(z)$, etc. □

LEMMA 4.16. $I\Delta_0 + \Omega_1$ *proves, by a proof of length of the order of* n^2, *that* Sat_n *preserves the* Δ_0*-induction axioms of length* $\leq n$, *i.e.,*

$$\begin{aligned}&\mathrm{Fmla}_{n.\Delta}(y)\\ &\quad\wedge \mathrm{Evalseq}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, 0)^\ulcorner \wedge \forall v_1 (^\urcorner y^\ulcorner \rightarrow ^\urcorner \mathrm{Sub}(y, \ulcorner v_1 \urcorner, Sv_1)^\ulcorner) \rightarrow \forall v_1 {}^\urcorner y, w)\\ &\quad\rightarrow \mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, 0)^\ulcorner \wedge \forall v_1 (^\urcorner y^\ulcorner \rightarrow ^\urcorner \mathrm{Sub}(y, \ulcorner v_1 \urcorner, Sv_1)^\ulcorner) \rightarrow \forall v_1 {}^\urcorner y, w).\end{aligned}$$

PROOF. We work in $I\Delta_0 + \Omega_1$ and assume

$$\begin{aligned}&\mathrm{Fmla}_{n.\Delta}(y)\\ &\quad\wedge \mathrm{Evalseq}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, 0)^\ulcorner \wedge \forall v_1 (^\urcorner y^\ulcorner \rightarrow ^\urcorner \mathrm{Sub}(y, \ulcorner v_1 \urcorner, Sv_1)^\ulcorner) \rightarrow \forall v_1 {}^\urcorner y, w).\end{aligned}$$

Because Sat_n is a partial satisfaction predicate for formulas of length $\leq n$, we can, by a proof of length of the order of n^2, prove that the formula

$$\mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, 0)^\ulcorner \wedge \forall v_1 (^\urcorner y^\ulcorner \rightarrow ^\urcorner \mathrm{Sub}(y, \ulcorner v_1 \urcorner, Sv_1)^\ulcorner) \rightarrow \forall v_1 {}^\urcorner y, w)$$

is equivalent to the following formula:

$$\begin{aligned}&\mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, 0)w)\\ &\quad\wedge \forall w'(w' =_1 w \rightarrow (\mathrm{Sat}_n(y, w') \rightarrow \mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, Sv_1), w')))\\ &\quad\rightarrow \forall w'(w' =_1 w \rightarrow \mathrm{Sat}_n(y, w')).\end{aligned}$$

This formula in turn is equivalent to:

$$\begin{aligned}&\mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, 0), w)\\ &\quad\wedge \forall x(\mathrm{Sat}_n(y, s(1, x, w)) \rightarrow \mathrm{Sat}_n(\mathrm{Sub}(y, \ulcorner v_1 \urcorner, Sv_1), s(1, x, w)))\\ &\quad\rightarrow \forall x\, \mathrm{Sat}_n(y, s(1, x, w)),\end{aligned}$$

where $s(1, x, w)$ is as defined in Definition 4.5. This last formula is then, by a proof of length of the order of n^2 of a "call by name = call by value" lemma analogous to the one proved in Lemma 4.13, equivalent to the following formula:

$$\begin{aligned}&\mathrm{Sat}_n(y, s(1, 0, w)) \wedge \forall x(\mathrm{Sat}_n(y, s(1, x, w)) \rightarrow \mathrm{Sat}_n(y, s(1, Sx, w)))\\ &\quad\rightarrow \forall x\, \mathrm{Sat}_n(y, s(1, x, w)).\end{aligned}$$

This looks almost like an instance of induction. However, because Sat_n is not Δ_0, we replace it by its $\Delta_0(\mathrm{Val}, \#, |\ |, \llcorner \frac{1}{2} x \lrcorner)$-equivalent $\mathrm{Sat}_{n.\Delta}$, as is allowed by Lemma 4.15 and the assumption $\mathrm{Fmla}_{n.\Delta}(y)$, and we obtain the equivalent formula

$$\begin{aligned}&\mathrm{Sat}_{n.\Delta}(y, s(1, 0, w)) \wedge \forall x(\mathrm{Sat}_{n.\Delta}(y, s(1, x, w)) \rightarrow \mathrm{Sat}_{n.\Delta}(y, s(1, Sx, w)))\\ &\quad\rightarrow \forall x\, \mathrm{Sat}_{n.\Delta}(y, s(1, x, w)).\end{aligned}$$

As a true instance of $\Delta_0(\mathrm{Val}, \#, |\ |, \llcorner \frac{1}{2} x \lrcorner)$-induction, the above formula is at last provable from the assumptions. □

Now that we have the partial truth predicates in hand, we can proceed with the proof proper of the main theorem of this paper. We suppose that the reader is familiar with $I\Delta_0 + \Omega_1$-cuts and $I\Delta_0 + \Omega_1$-initial segments, and also with Solovay's method of shortening cuts (see Definition 2.16, Definition 2.17, and Lemma 2.20).

We have the following.

LEMMA 4.17. *If K is an $\mathrm{I}\Delta_0 + \Omega_1$-initial segment, then*

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x \, \mathrm{Prov}(\ulcorner K(\overline{x})\urcorner),$$

where $\overline{x}$ stands for the "efficient numeral" based on the binary expansion of x.

PROOF. See Lemma 2.21. It is not difficult to see that the proofs of $K(\overline{x})$ are of length of the order $|x|^2$.

However, in the formalized context in which we will use the result, the length of the formula K and the length of the proof $p_1(K)$ of $\forall y(K(y) \to K(Sy))$ and the proof $p_2(K)$ of $\forall y(K(y) \to K(SS0 \cdot y))$ also play a part in the computation of the length of the total proof, thereby making the length of the total proof of the order $|x|^2 \cdot |K| + |p_1(K)| + |p_2(K)|$.

In fact, if we analyze the proof, we find that

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall J \forall x (\Box(J \text{ "is an initial segment"}) \to \Box(J(\overline{x}))). \qquad \Box$$

DEFINITION 4.18. We formally define the following:

$$\mathrm{LPrf}_v(u.\ulcorner\chi\urcorner) := \text{ "u codes a proof of χ in $\mathrm{I}\Delta_0 + \Omega_1$ involving only formulas of length $\leq v$".}$$

LEMMA 4.19. *The following is provable in $\mathrm{I}\Delta_0 + \Omega_1$:*

$$\forall x \, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{Prf}(y, \overline{\ulcorner\varphi\urcorner}) \leftrightarrow \mathrm{LPrf}_{|x|}(y, \overline{\ulcorner\varphi\urcorner}))\urcorner).$$

PROOF. Formalize the following observation: if a formula v occurs in a proof y where $y \leq x$, then $\mathrm{Len}(v) \leq |v| \leq |y| \leq |x|$. $\Box$

THEOREM 4.20 (small reflection). *For all sentences φ the following holds*:

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x \, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{Prf}(y, \overline{\ulcorner\varphi\urcorner}) \to \varphi)\urcorner).$$

PROOF. By Lemma 4.19, it suffices to prove

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x \, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{Prf}_{|x|}(y, \overline{\ulcorner\varphi\urcorner}) \to \varphi)\urcorner).$$

We reason inside $\mathrm{I}\Delta_0 + \Omega_1$, and we take an x which we shall use to make a cut. The idea behind the proof is to find a Gödel number K_x standing for a formalized "Prov-initial segment" such that we have

$$\mathrm{Prov}(K_x(\overline{x})^{\ulcorner}\!\!\to \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner\varphi\urcorner}) \to \varphi)\urcorner).$$

(By abuse of notation we write $K_x(\overline{x})$ for the Gödel number that results by the appropriate application of the substitution function to K_x.) In the construction of the Prov-initial segment K_x, we will need the formalized versions of the lemmas which we proved above about the existence and the properties of partial satisfaction predicates for formulas of length smaller than some standard numeral n. In our formalized context, $|x|$ plays the rôle of "standard numeral", as will become clear when we define K_x. Again by abuse of notation, we let $\mathrm{Sat}_{|x|}(v, w)$ stand for a Gödel number instead of a formula; we will use the appropriate formalizations of lemmas we proved about the formulas $\mathrm{Sat}_n(v, w)$ to derive formalized facts about the Gödel number $\mathrm{Sat}_{|x|}(v, w)$.

Keeping these cautionary remarks in mind, we start the proof by defining the Gödel number J_x of a formalized "Prov-cut" (later to be shortened to the Prov-initial segment K_x that we need) as follows:

$$J_x(s) := \ulcorner \forall y, v \leq s(\mathrm{LPrf}_{|x|}(y, v) \rightarrow \forall w(\mathrm{Evalseq}(v, w) \rightarrow \urcorner \mathrm{Sat}_{|x|}(v, w)\ulcorner))\urcorner.$$

By the formalized version of Lemma 4.7, we may assume that this Gödel number exists, because the length of $\mathrm{Sat}_{|x|}(v, w)$ is linear in $|x|$. (Note that we are reasoning inside $\mathrm{I}\Delta_0 + \Omega_1$ all the time!) It is not difficult to prove directly from the definition of J_x (and from the fact that J_x is small enough) that the following holds:

$$\mathrm{Prov}(J_x(\overline{0})\ulcorner \wedge \forall y \forall z(\urcorner J_x(z)\ulcorner \wedge y \leq z \rightarrow \urcorner J_x(y)\ulcorner)\urcorner).$$

To prove that J_x is closed under successor, we remark that

$$\mathrm{Prov}(\ulcorner \mathrm{LPrf}_{|x|}(y, v) \rightarrow \mathrm{Len}(v) \leq |x|\urcorner).$$

Therefore, we can formalize Lemmas 4.12, 4.13, 4.14, and 4.16 to conclude by a proof of length of the order $|x|^2$ that $\mathrm{Sat}_{|x|}(v, w)$ is preserved by all logical and nonlogical axioms and rules for formulas of length $\leq |x|$, and thus, indeed,

$$\mathrm{Prov}(\ulcorner \forall y(\urcorner J_x(y)\ulcorner \rightarrow \urcorner J_x(Sy)\ulcorner)\urcorner),$$

proving J_x to be a Prov-cut.

By a formalization of the proof of Lemma 2.20, we can shorten the Prov-cut J_x to a Prov-initial segment K_x of length linear in $|x|$. The proof that K_x is a Prov-initial segment is of length polynomial in $|x|$.

Carefully analyzing the proof of Lemma 4.17 (see the remark at the end of that proof), we find, by proofs of length polynomial in $|x|$, that

$$\mathrm{Prov}(K_x(\overline{x})) \wedge \mathrm{Prov}(K_x(\overline{\ulcorner \varphi \urcorner})).$$

Thus, because we have $\mathrm{Prov}(\ulcorner \forall y(\urcorner K_x(y)\ulcorner \rightarrow \urcorner J_x(y)\ulcorner)\urcorner)$, we conclude that, by definition of J_x,

$$\mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner \varphi \urcorner}) \rightarrow \forall w(\mathrm{Evalseq}(\overline{\ulcorner \varphi \urcorner}, w) \rightarrow \urcorner \mathrm{Sat}_{|x|}(\overline{\ulcorner \varphi \urcorner}, w)\ulcorner))\urcorner).$$

Because we have $\mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner \varphi \urcorner}) \rightarrow \mathrm{Fmla}_{|x|}(\overline{\ulcorner \varphi \urcorner}))\urcorner)$, we can apply the formalized version of Lemma 4.11, taking note that φ is a sentence. Therefore,

$$\mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner \varphi \urcorner}) \rightarrow \forall w(\mathrm{Evalseq}(\overline{\ulcorner \varphi \urcorner}, w) \rightarrow \varphi))\urcorner).$$

This in turn is equivalent to the desired

$$\mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi)\urcorner).$$

Stepping out of $\mathrm{I}\Delta_0 + \Omega_1$ again, we conclude that indeed,

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall x\, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi)\urcorner). \qquad \square$$

REMARK 4.21. Looking carefully at the proof of Theorem 4.20, we notice that it is also possible to derive the following result, which is a little bit stronger:

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \forall v(\mathrm{Sent}(v) \rightarrow \forall x\, \mathrm{Prov}(\ulcorner \forall y \leq \overline{x}(\mathrm{LPrf}_{|x|}(y, \overline{\ulcorner v \urcorner}) \rightarrow \urcorner v\ulcorner)\urcorner).$$

Theorem 4.20 and its proof can also be adapted for the case that φ is a formula instead of a sentence (or, in the stronger result mentioned above, $\mathrm{Fmla}(v)$ instead of $\mathrm{Sent}(v)$).

COROLLARY 4.22 (Švejdar's principle is provable in $\mathrm{I}\Delta_0 + \Omega_1$). *For all sentences* φ, ψ, *we have the following*:

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \Box\varphi \to \Box(\Box\psi \leq \Box\varphi \to \psi),$$

i.e.,

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \exists x\, \mathrm{Prf}(x, \ulcorner\varphi\urcorner) \\ \to \mathrm{Prov}(\ulcorner \exists y(\mathrm{Prf}(y, \overline{\ulcorner\psi\urcorner}) \wedge \forall z \leq y \neg \mathrm{Prf}(z, \overline{\ulcorner\varphi\urcorner})) \to \psi\urcorner).$$

PROOF. We work inside $\mathrm{I}\Delta_0 + \Omega_1$ and suppose $\mathrm{Prf}(x, \ulcorner\varphi\urcorner)$. By provable Σ_1^b-completeness, this implies $\mathrm{Prov}(\ulcorner \mathrm{Prf}(\overline{x}, \overline{\ulcorner\varphi\urcorner})\urcorner)$. Hence, we have

$$\mathrm{Prov}(\ulcorner \exists y(\mathrm{Prf}(y, \overline{\ulcorner\psi\urcorner}) \wedge \forall z \leq y \neg \mathrm{Prf}(z, \overline{\ulcorner\varphi\urcorner})) \to \exists y \leq \overline{x}\, \mathrm{Prf}(y, \overline{\ulcorner\psi\urcorner})\urcorner).$$

Theorem 4.20 gives $\mathrm{Prov}(\ulcorner \exists y \leq \overline{x}\, \mathrm{Prf}(y, \overline{\ulcorner\psi\urcorner}) \to \psi\urcorner)$; therefore, we have the following:

$$\mathrm{Prov}(\ulcorner \exists y(\mathrm{Prf}(y, \overline{\ulcorner\psi\urcorner}) \wedge \forall z \leq y \neg \mathrm{Prf}(z, \overline{\ulcorner\varphi\urcorner})) \to \psi\urcorner).$$

Jumping outside $\mathrm{I}\Delta_0 + \Omega_1$ again, we conclude that

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \exists x\, \mathrm{Prf}(x, \ulcorner\varphi\urcorner) \\ \to \mathrm{Prov}(\ulcorner \exists y(\mathrm{Prf}(y, \overline{\ulcorner\psi\urcorner}) \wedge \forall z \leq y \neg \mathrm{Prf}(z, \overline{\ulcorner\varphi\urcorner})) \to \psi\urcorner). \qquad \Box$$

REMARK 4.23. Analogously to Remark 4.21, we may strengthen Švejdar's principle to the following:

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Sent}(u) \wedge \mathrm{Sent}(v) \wedge \mathrm{Prov}(u) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(v) \leq \mathrm{Prov}(u) \to \urcorner v).$$

Švejdar introduced a modal system in order to study generalized Rosser sentences, and he derived the formalized version of Rosser's Theorem in it [Šv83]. Because of Corollary 4.22, Švejdar's system is sound with respect to $\mathrm{I}\Delta_0 + \Omega_1$, and Rosser's Theorem holds in $\mathrm{I}\Delta_0 + \Omega_1$.

Below, we use an argument similar to Švejdar's to derive a more general theorem. For the case of PA, this theorem has been proved by Montagna and Bernardi (see [JM87]).

THEOREM 4.24 (Montagna-Bernardi in $\mathrm{I}\Delta_0 + \Omega_1$). *For every function h which is Σ_1^b-definable in $I\Delta_0 + \Omega_1$ and maps sentences to sentences, there is a sentence C such that*

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Prov}(\ulcorner C\urcorner) \leftrightarrow \mathrm{Prov}(h(\ulcorner C\urcorner)).$$

PROOF. Define C by diagonalization such that

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash C \leftrightarrow \mathrm{Prov}(h(\ulcorner C\urcorner)) \leq \mathrm{Prov}(\ulcorner C\urcorner).$$

Reason inside $\mathrm{I}\Delta_0 + \Omega_1$, and assume first that $\mathrm{Prov}(\ulcorner C\urcorner)$. Then, by definition,

$$\mathrm{Prov}(\ulcorner \mathrm{Prov}(h(\ulcorner C\urcorner)) \leq \mathrm{Prov}(\ulcorner C\urcorner)\urcorner).$$

Meanwhile Corollary 4.22 gives

$$\mathrm{Prov}(\ulcorner C\urcorner) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(h(\ulcorner C\urcorner)) \leq \mathrm{Prov}(\ulcorner C\urcorner) \to h(\ulcorner C\urcorner)\urcorner).$$

Combined, these two yield $\mathrm{Prov}(\ulcorner C\urcorner) \to \mathrm{Prov}(h(\ulcorner C\urcorner))$.

For the other side, we assume that $\mathrm{Prov}(h(\ulcorner C\urcorner))$. This implies

$$\mathrm{Prov}(\ulcorner \mathrm{Prov}(h(\ulcorner C\urcorner))\urcorner),$$

and thus,

$$\mathrm{Prov}(\ulcorner \mathrm{Prov}(h(\ulcorner C\urcorner)) \leq \mathrm{Prov}(\ulcorner C\urcorner) \vee \mathrm{Prov}(\ulcorner C\urcorner) \leq \mathrm{Prov}(h(\ulcorner C\urcorner))\urcorner).$$

By definition of C, we derive

$$\mathrm{Prov}(\ulcorner C \vee \mathrm{Prov}(\ulcorner C\urcorner) \leq \mathrm{Prov}(h(\ulcorner C\urcorner))\urcorner).$$

Now we apply Corollary 4.22 to conclude that because

$$\mathrm{Prov}(h(\ulcorner C\urcorner)) \to \mathrm{Prov}(\ulcorner \mathrm{Prov}(\ulcorner C\urcorner) \leq \mathrm{Prov}(h(\ulcorner C\urcorner)) \to C\urcorner),$$

indeed $\mathrm{Prov}(h(\ulcorner C\urcorner)) \to \mathrm{Prov}(\ulcorner C\urcorner)$. □

Note that the formalized version of Rosser's Theorem follows immediately from this construction. If we take R such that

$$\mathrm{I}\Delta_0 + \Omega_1 \vdash R \leftrightarrow \mathrm{Prov}(\ulcorner \neg R\urcorner) \leq \mathrm{Prov}(\ulcorner R\urcorner),$$

we derive $\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Prov}(\ulcorner R\urcorner) \leftrightarrow \mathrm{Prov}(\ulcorner \neg R\urcorner)$, and thus $\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Prov}(\ulcorner R\urcorner) \to \mathrm{Prov}(\ulcorner \bot\urcorner)$ and $\mathrm{I}\Delta_0 + \Omega_1 \vdash \mathrm{Prov}(\ulcorner \neg R\urcorner) \to \mathrm{Prov}(\ulcorner \bot\urcorner)$.

§**5. Injection of small (but not too small) inconsistency proofs.** Using the small reflection principle, we can strengthen Hájek's, Solovay's, and Krajíček and Pudlák's results on the injection of inconsistencies into models of $\mathrm{I}\Delta_0$+ EXP [Há83], [So89], and [KP89]. Instead of only injecting an inconsistency proof, we also take care to respect a fair number of consistency statements. Moreover, we do not need full exponentiation in our original model.

We cannot immediately apply the lemmas of [KP89], but the essential steps in our proof are the same as in that article. We first apply Pudlák's version of Gödel's Second Incompleteness Theorem (see [Pu86, Theorem 3.6]) to show that we can indeed inject an inconsistency proof; then we use the Omitting Types Theorem to prevent extra elements from creeping into the lower part of the new model that contains our injected inconsistency proof.

THEOREM 5.1. *Let* $\mathbf{T} \supseteq \mathrm{I}\Delta_0 + \Omega_1$ *be a* Σ_1^b*-axiomatized theory for which the small reflection principle (see Theorem* 4.20) *is provable in* $\mathrm{I}\Delta_0 + \Omega_1$. *Let* $\mathrm{Con}_T(x)$ *be a formalization of the consistency of* $\mathbf{T}$ *up to proofs of length* x. *Let* $\mathscr{M}$ *be a nonstandard countable model of* $\mathrm{I}\Delta_0 + \Omega_1$. *Let* a, c *be nonstandard elements of* $\mathscr{M}$ *such that the following conditions hold*:

- $\exp(a^c) \in \mathscr{M}$;
- $\mathscr{M} \models \mathrm{Con}_T(a^k)$ *for all* $k < \omega$.

Then there exists a countable model $\mathscr{K}$ *of* $\mathbf{T}$ *such that* $a \in \mathscr{K}$ *and*

(1) $\mathscr{M} \restriction a = \mathscr{K} \restriction a$,

(2) $\mathscr{M} \restriction \exp(a^k) \subseteq \mathscr{K}$ *for all* $k < \omega$,

(3) $\mathscr{K} \models \neg \mathrm{Con}_T(a^c)$,
(4) $\mathscr{K} \models \mathrm{Con}_T(a^k)$ *for all* $k < \omega$,
(5) $\mathscr{K} \models 2^{a^c} \downarrow$.

PROOF. Define $\mathscr{N} := \{x \in \mathscr{M} | x < \exp(a^k)$ for some $k < \omega\}$. Then $\exp(a^c) \in \mathscr{M} \backslash \mathscr{N}$; thus, $\mathscr{M}$ is a proper end-extension of $\mathscr{N}$. Therefore, by Theorem 1 of [WP89], $\mathscr{N} \models \mathrm{B}\Sigma_1$. (Remember that $\mathrm{B}\Sigma_1$ is $\mathrm{I}\Delta_0 +$ the scheme $\forall t(\forall x < t \exists y \varphi(x, y) \rightarrow \exists a \forall x < t \exists y < a \varphi(x, y))$ for $\varphi \in \Sigma_1^0$.) Also, it is easy to see that $\mathscr{N} \models \Omega_1$.

On the other hand, one of our assumptions is that $\mathscr{M} \models \mathrm{Con}_T(a^k)$ for all $k < \omega$. By Δ_0-overspill we conclude that there is a nonstandard $d < c$ in $\mathscr{M}$ such that $\mathscr{M} \models \mathrm{Con}_T(a^d)$. Thus, by Theorem 3.6 of [Pu86], there is a $k < \omega$ such that $\mathscr{M} \models \mathrm{Con}_{T+\neg \mathrm{Con}_T(a^d)}(a^{d/k})$, so certainly $\mathscr{M} \models \mathrm{Con}_{T+\neg \mathrm{Con}_T(a^c)}(a^{d/k})$. Indeed, because d/k is nonstandard, we even have $\mathscr{N} \models \mathrm{Con}(\mathbf{U})$, where $\mathbf{U} := \mathbf{T} + \neg \mathrm{Con}_T(a^c)$.

At this point we need some definitions analogous to the ones in [KP89]. Let $L(\mathscr{N})$ be the language of arithmetic expanded with domain constants for the elements of $\mathscr{N}$. We define a translation t from $L(\mathscr{N})$ to $\mathscr{N}$ by $t(A(a_1, \dots, a_k)) := \ulcorner A(\overline{a_1}, \dots, \overline{a_k}) \urcorner$, where $\overline{a}_i$ is the efficient numeral of a_i. We need one more definition:

$$\mathbf{U}^* := \{A(\vec{a}) \in L(\mathscr{N}) | \mathscr{N} \models \mathrm{Prov}_U(t(A(\vec{a})))\}.$$

It is easy to show that $\mathbf{U}^*$ is closed under the rules of predicate logic; that $\mathbf{U} \subseteq \mathbf{U}^*$; and that $\mathrm{Diag}(\mathscr{N}) \subseteq \mathbf{U}^*$. Also, because $\mathscr{N} \models \mathrm{Con}(\mathbf{U})$, we can conclude that $\mathbf{U}^*$ is consistent.

Moreover, by the small reflection principle for $\mathrm{I}\Delta_0 + \Omega_1$, we have

$$\mathscr{N} \models \forall x\, \mathrm{Prov}_U(\ulcorner \mathrm{Con}_T(|\overline{x}|) \urcorner);$$

thus, for all $k < \omega$, $\mathrm{Con}_T(a^k) \in \mathbf{U}^*$.

Finally, using Solovay's cuts, we can show that $\mathscr{N} \models \forall x\, \mathrm{Prov}(\ulcorner 2^x \downarrow \urcorner)$; thus, $2^{a^c} \downarrow \in \mathbf{U}^*$.

We construct the required model $\mathscr{K}$ by the Omitting Types Theorem in order to take care that $\mathscr{K}$ will contain no new elements below a. Let τ be the type in $L(\mathscr{N})$ defined by

$$\tau(x) := \{x \leq a\} \cup \{x \neq b | b \in \mathscr{M} \restriction a\}.$$

Claim 1. $\mathbf{U}^*$ *locally omits* τ.

Proof. Take any $A(x)$, and suppose that for all $b \leq a$ in $\mathscr{N}$ we have $\mathbf{U}^* \vdash \neg A(b)$ and that $\mathbf{U}^* \vdash A(x) \rightarrow x \leq a$. We want to show that $\mathbf{U}^* \vdash \neg \exists x A(x)$. By definition of $\mathbf{U}^*$, it is sufficient to prove the following:

$$\mathscr{N} \models \forall b \leq a\, \mathrm{Prov}_U(\ulcorner \neg A(\overline{b}) \urcorner) \rightarrow \mathrm{Prov}_U(\ulcorner \forall x \leq a \neg A(\overline{x}) \urcorner).$$

So suppose $\mathscr{N} \models \forall b \leq a\, \mathrm{Prov}_U(\ulcorner \neg A(\overline{b}) \urcorner)$. By $\mathrm{B}\Sigma_1$, there is a $q \in \mathscr{N}$ such that

$$\mathscr{N} \models \forall b \leq a \exists p < q\, \mathrm{Prf}_U(p, \ulcorner \neg A(\overline{b}) \urcorner).$$

Now we can use $\Delta_0(\omega_1)$-induction to show that we can combine these proofs for all $b \leq a$ into one proof p of $\forall x \leq a \neg A(x)$, where $|p| \leq a \cdot (|q| + k \cdot |a|) \leq a^m$ for some standard k, n, m; thus, $p \in \mathscr{N}$. We conclude that indeed

$$\mathscr{N} \models \mathrm{Prov}_U(\ulcorner \forall x \leq a \neg A(x) \urcorner). \qquad \square$$

At last we can construct a model $\mathscr{K}$ of $\mathbf{U}^*$ omitting τ . Using the facts that we proved about $\mathbf{U}^*$, we conclude that $\mathscr{K}$ satisfies all the properties that we want. $\square$

In Theorem 5.1, we require that $\mathbf{T} \supseteq \mathrm{I}\Delta_0 + \Omega_1$ is a Σ_1^b-axiomatized theory for which the small reflection principle is provable in $\mathrm{I}\Delta_0 + \Omega_1$. Examples of such theories are finite extensions of $\mathrm{I}\Delta_0 + \Omega_1$ itself, $\mathrm{I}\Delta_0+$ EXP, and PA. We hope to give an exact characterization of theories amenable to methods analogous to those of §4, [Pu86], and [Pu87] in a later paper.

Theorem 5.1 is only a slight extension of [KP89, Theorem 2.1]. We use the small reflection principle only to show that the length of injected inconsistency proofs can be bounded from below as well as from above.

A variation on the proof of Theorem 5.1 gives the following theorem. Its proof contains a more surprising use of the small reflection theorem than the proof of Theorem 5.1. In Theorem 5.3 we use it even in our application of the Omitting Types Theorem.

Recently, some papers (see [WP89], [Ad90], [Ad93]) appeared that partially answer the end extension problem which was formulated by Kirby and Paris in 1977 as follows: does every model of $\mathrm{I}\Delta_0 + \mathrm{B}\Sigma_1$ have a proper end extension to a model of $\mathrm{I}\Delta_0$? The theorem below gives a sufficient condition for a countable model of $\mathrm{I}\Delta_0 + \mathrm{B}\Sigma_1$ to have a proper end extension to a model of $\mathrm{I}\Delta_0$: if the model additionally satisfies $\Omega_1 + \mathrm{Con}(\mathrm{I}\Delta_0)$ and provable completeness for Π_2^b-formulas, then it does have such an end extension.

First, we need a definition.

DEFINITION 5.2. $\mathrm{C}\Pi_2^b(\mathbf{U})$ is the scheme

$$A(a_1, \ldots, a_k) \rightarrow \mathrm{Prov}_U(\ulcorner A(\overline{a_1}, \ldots, \overline{a_k}) \urcorner)$$

for $A(a_1, \ldots, a_k) \in \Pi_2^b$.

THEOREM 5.3. *Let* $\mathbf{U} \supseteq \mathbf{Q}$ *be a* Σ_1^b*-axiomatized theory, and suppose* $\mathscr{N}$ *is a countable model of* $\mathrm{B}\Sigma_1 + \Omega_1 + \mathrm{C}\Pi_2^b(\mathbf{U}) + \mathrm{Con}(\mathbf{U})$, *then there exists a countable model* $\mathscr{K}$ *of* $\mathbf{U}$ *such that* $\mathscr{K}$ *is an end extension of* $\mathscr{N}$.

PROOF. Define $\mathbf{U}^*$ from $\mathbf{U}$, $\mathscr{N}$ exactly as in the proof of Theorem 5.1. Again we construct the required model $\mathscr{K}$ of $\mathbf{U}^*$ using the Omitting Types Theorem. This time we define for all $a \in \mathscr{N}$ the type τ_a in $L(\mathscr{N})$ by

$$\tau_a(x) := \{x \leq a\} \cup \{x \neq b | b \in \mathscr{M} \restriction a\}.$$

Claim 2. $\mathbf{U}^*$ *locally omits* τ_a *for all* $a \in \mathscr{N}$.

Proof. Take any $a \in \mathscr{N}$ and any formula $A(x)$. As in the proof of Claim 1, it is sufficient to show the following:

$$\mathscr{N} \models \forall b \leq a\, \mathrm{Prov}_U(\ulcorner \neg A(\overline{b}) \urcorner) \rightarrow \mathrm{Prov}_U(\ulcorner \forall x \leq \overline{a} \neg A(\overline{x}) \urcorner).$$

So suppose

$$\mathscr{N} \models \forall b \leq a\, \mathrm{Prov}_U(\ulcorner \neg A(\overline{b}) \urcorner).$$

By $B\Sigma_1$, there is a $q \in \mathcal{N}$ such that

$$\mathcal{N} \models \forall b \leq a \exists p < q \, \mathrm{Prf}_U(p, \ulcorner \neg A(\overline{b}) \urcorner).$$

Now by $C\Pi_2^b(\mathbf{U})$, we derive

$$\mathcal{N} \models \exists q \, \mathrm{Prov}_U(\ulcorner \forall b \leq \overline{a} \exists p < \overline{q} \, \mathrm{Prf}_U(p, \ulcorner \neg A(\overline{b}) \urcorner) \urcorner).$$

Therefore, by the small reflection principle,

$$\mathcal{N} \models \mathrm{Prov}_U(\ulcorner \forall b \leq \overline{a} \neg A(\overline{b}) \urcorner). \qquad \square$$

We can now construct a countable model $\mathcal{K}$ of $\mathbf{U}^*$ omitting all τ_a for $a \in \mathcal{N}$. As before, it is easy to see that $\mathbf{U} \subseteq \mathbf{U}^*$, so $\mathcal{K} \models \mathbf{U}$.

By the way, note that by the small reflection principle for $I\Delta_0 + \Omega_1$, or simply by the isomorphism, we have $\mathrm{Con}_U(|\overline{x}|) \in \mathbf{U}^*$, and thus $\mathcal{K} \models \mathrm{Con}_U(|\overline{x}|)$ for all $x \in \mathcal{N}$. $\square$

§6. **Acknowledgments.** We would like to thank Dick de Jongh, Marianne Kalsbeek, Domenico Zambella, and Volodya Shavrukov for interesting discussions and for their helpful comments on previous versions of this paper.

REFERENCES

[Ad90] Z. Adamowicz, *End-extending models of* $I\Delta_0 + \mathrm{EXP} + B\Sigma_1$, ***Fundamenta Mathematicae***, vol. 136 (1990), pp. 133–145.

[Ad93] ———, *A contribution to the end-extension problem and the* Π_1 *conservativeness problem*, ***Annals of Pure and Applied Logic***, vol. 61 (1993), pp. 3–48.

[BDG87] J. L. Balcázar, J. Díaz, and J. Gabarró, ***Structural complexity I***, Springer-Verlag, Berlin, 1987.

[Be89] L. D. Beklemishev, *On the classification of propositional provability logics*, ***Mathematics of the USSR Izvestiya***, vol. 35 (1990), pp. 247–275.

[Be91] ———, ***On bimodal provability logics for*** Π_1***-axiomatized extensions of arithmetical theories***, ITLI prepublication series, X-91-09, University of Amsterdam, Amsterdam, 1991.

[BV93] A Berarducci and L. C. Verbrugge, *On the provability logic of bounded arithmetic*, ***Annals of Pure and Applied Logic***, vol. 61 (1993), pp. 75–93.

[Bu86] S. Buss, ***Bounded arithmetic***, Bibliopolis, Napoli, 1986.

[Fe60] S. Feferman, *Arithmetization of metamathematics in a general setting*, ***Fundamenta Mathematicae***, vol. 49 (1960), pp. 35–92.

[FR79] J. Ferrante and C. W. Rackoff, ***The computational complexity of logical theories***, Springer-Verlag, Berlin, 1979.

[GS79] D. Guaspari and R. M. Solovay, *Rosser sentences*, ***Annals of Mathematical Logic***, vol. 16 (1979), pp. 81–99.

[Há83] P. Hájek, *On a new notion of partial conservativity*, ***Logic colloquium '83*** (E. Börger et al, editors), vol. 2, Springer-Verlag, Berlin, 1983, pp. 217–232.

[HP93] P. Hájek and P. Pudlák, ***Metamathematics of first-order arithmetic***, Springer-Verlag, Berlin, 1993.

[Jo87] D. H. J. de Jongh, *A simplification of a completeness proof of Guaspari and Solovay*, ***Studia Logica***, vol. 46 (1987), pp. 187–192.

[JM87] D. H. J. DE JONGH and F. MONTAGNA, *Generic generalized Rosser fixed points*, ***Studia Logica***, vol. 46 (1987), pp. 193–203.

[JV88] D. H. J. DE JONGH and F. VELTMAN, ***Intensional logic***, lecture notes, Philosophy Department, University of Amsterdam, Amsterdam, 1988.

[JMM91] D. H. J. DE JONGH, M. JUMELET, and F. MONTAGNA, *On the proof of Solovay's theorem*, ***Studia Logica***, vol. 50 (1991), pp. 51–70.

[JM91] D. H. J. DE JONGH and F. MONTAGNA, *Rosser orderings and free variables*, ***Studia Logica***, vol. 50 (1991), pp. 71–80.

[KH82] C. F. KENT and B. R. HODGSON, *An arithmetical characterization of NP*, ***Theoretical Computer Science***, vol. 21 (1982), pp. 255–267.

[KP89] J. KRAJÍČEK and P. PUDLÁK, *On the structure of initial segments of models of arithmetic*, ***Archives of Mathematical Logic***, vol. 28 (1989), pp. 91–98.

[MA78] K. MANDERS and L. ADLEMAN, *NP-complete decision problems for binary quadratics*, ***Journal of Computer and System Sciences***, vol. 16 (1978), pp. 168–184.

[Ne86] E. NELSON, ***Predicative arithmetic***, Mathematical Notes 32, Princeton University Press, Princeton, New Jersey, 1986.

[Pa71] R. PARIKH, *Existence and feasibility in arithmetic*, this JOURNAL, vol. 36 (1971), pp. 494–508.

[Pu83] P. PUDLÁK, *A definition of exponentiation by a bounded arithmetical formula*, ***Commentationes Mathematicae Universitatis Carolinae***, vol. 24 (1983), pp. 667–671.

[Pu85] ———, *Cuts, consistency statements and interpretations*, this JOURNAL, vol. 50 (1985), pp. 423–441.

[Pu86] ———, *On the length of proofs of finitistic consistency statements in first order theories*, ***Logic colloquium '84*** (J. B. Paris et al., editors), North-Holland, Amsterdam, 1986, pp. 165–196.

[Pu87] ———, *Improved bounds to the length of proofs of finitistic consistency statements*, ***Logic and combinatorics*** (S. G. Simpson, editor): ***Contemporary Mathematics***, vol. 35, American Mathematical Society, Providence, Rhode Island, 1987, pp. 309–332.

[Sm85] C. SMORYŃSKI, ***Self-reference and modal logic***, Springer-Verlag, New York, 1985.

[So76] R. M. SOLOVAY, *Provability interpretations of modal logic*, ***Israel Journal of Mathematics***, vol. 25 (1976), pp. 287–304.

[So76b] ———, ***On interpretability in set theories***, manuscript, 1976.

[So89] ———, *Injecting inconsistencies into models of PA*, ***Annals of Pure and Applied Logic***, vol. 44 (1989), pp. 261–302.

[St76] L. J. STOCKMEYER, *The polynomial-time hierarchy*, ***Theoretical Computer Science***, vol. 3 (1976), pp. 1–22.

[Šv83] V. ŠVEJDAR, *Modal analysis of generalized Rosser sentences*, this JOURNAL, vol. 48 (1983), pp. 986–999.

[Ta75] G. TAKEUTI, ***Proof theory***, North-Holland, Amsterdam, 1975.

[Ta88] ———, *Bounded arithmetic and truth definition*, ***Annals of Pure and Applied Logic***, vol. 39 (1988), pp. 75–104.

[Ve88] L. C. VERBRUGGE, ***Does Solovay's completeness theorem extend to bounded arithmetic?***, Master's Thesis, University of Amsterdam, Amsterdam, 1988.

[Ve89] ———, ***Σ-completeness and bounded arithmetic***, ITLI prepublication series for mathematical logic and foundations, ML-89-05, University of Amsterdam, Amsterdam, 1989.

[Vi81] A. VISSER, ***Aspects of diagonalization & provability***, Ph.D. thesis, University of Utrecht, Utrecht, 1981.

[Vi82] ———, *On the completeness principle*, ***Annals of Mathematical Logic***, vol. 22 (1982), pp. 263–295.

[Vi85] ———, ***Evaluation, provably deductive equivalence in Heyting's Arithmetic of substitution instances of propositional formulas***, Logic Group Prepint Series, no. 4, University of Utrecht, Utrecht, 1985.

[Vi90] ———, *Interpretabilty logic*, ***Mathematical logic*** (P. P. Petkov, editor), Proceedings of the Heyting '88 summer school, Plenum Press, New York, 1990, pp. 175–209.

[Vi91] ———, *The formalization of interpretability*, ***Studia Logica***, vol. 50 (1991), pp. 81–105.

[WP87] A. J. WILKIE and J. B. PARIS, *On the scheme of induction for bounded arithmetic formulas*, ***Annals of Pure and Applied Logic***, vol. 35 (1987), pp. 261–302.

[WP89] ———, *On the existence of end extensions of models of bounded induction*, ***Logic, methodology and philosophy of science VIII*** (J. E. Fenstad et al, editors), North-Holland, Amsterdam, 1989, pp. 143–161.

[Wr76] C. WRATHALL, *Complete sets and the polynomial time hierarchy*, ***Theoretical Computer Science***, vol. 3 (1976), pp. 23–33.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF AMSTERDAM
1018 TV AMSTERDAM, THE NETHERLANDS

Current address: Department of Philosophy, University of Gothenburg, S-412 g8 Gothenburg, Sweden

E-mail: Rineke.Verbrugge@phil.gu.se

DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF GOTHENBURG
S-412 G8 GOTHENBURG, SWEDEN

E-mail: Albert.Visser@phil.RUU.nl