554
Notre Dame Journal of Formal Logic
Volume 32, Number 4, Fall 1991

On the Σ_1^0-Conservativity of Σ_1^0-Completeness

ALBERT VISSER

Abstract In this paper we show that $I\Delta_0 + \Omega_1$ verifies the sentential Σ_1^0-conservativity of schematical, sentential Σ_1^0-completeness. (This means that for any finite set of Σ_1^0-sentences S we can prove in $I\Delta_0 + \Omega_1$ that the statement expressing the completeness of S w.r.t. $I\Delta_0 + \Omega_1$ is conservative over $I\Delta_0 + \Omega_1$ w.r.t. Σ_1^0-sentences.) Some consequences are discussed. We formulate a system of provability logic based on the verifiable sentential Σ_1^0-conservativity of schematical, sentential Σ_1^0-completeness.

1 Introduction As is well known it is a difficult question whether $I\Delta_0 + \Omega_1$ proves Σ_1^0-completeness. From Buss [1], Chapter 8, we can extract the following point: let $A(x)$ be any coNP-complete Π_1^b formula. Suppose $I\Delta_0 + \Omega_1$ proves: $\forall x(A(x) \to \Box_{I\Delta_0+\Omega_1}A(x))$. Then by Parikh's Theorem for some polynomial $P(x)$, $I\Delta_0 + \Omega_1$ proves: $\forall x(A(x) \to \exists |y| < P(|x|)\ \mathrm{Proof}_{I\Delta_0+\Omega_1}(y, A(x)))$. Hence *in the standard model* we have: $\forall x(A(x) \leftrightarrow \exists |y| < P(|x|)\ \mathrm{Proof}_{I\Delta_0+\Omega_1}(y, A(x)))$. In other words, $A(x)$ is equivalent to a Σ_1^b-predicate. Ergo NP = coNP. On the other hand, if $I\Delta_0 + \Omega_1$ *proves* a suitable schematic version of NP = coNP, then—as is easily seen—$I\Delta_0 + \Omega_1$ proves Σ_1^0-completeness.

Verbrugge [7] shows that for $A(x)$ in the above argument we may also take a formula of the form: $\Box_{I\Delta_0+\Omega_1}B(x) < \Box_{I\Delta_0+\Omega_1}C(x)$. Such a formula is $\exists\Pi_1^b$. This means that if completeness for Rosser-ordered provabilities (with parameter) were provable in $I\Delta_0 + \Omega_1$, then again NP = coNP.

In Paris and Wilkie [4] it is shown that all principles of Löb's Logic are valid in $I\Delta_0 + \Omega_1$. Solovay's proof of the arithmetical completeness of Löb's Logic, however, uses essentially the verifiability of schematical, sentential Σ_1^0-completeness (in fact: completeness for Rosser-ordered provabilities) in the arithmetical theory (see [7]). As a consequence, the question of arithmetical completeness of Löb's Logic for interpretations in $I\Delta_0 + \Omega_1$ is still open.

In this paper we show that for any finite set S, $I\Delta_0 + \Omega_1$ verifies that the statement expressing the completeness of S w.r.t. $I\Delta_0 + \Omega_1$ is conservative over $I\Delta_0 + \Omega_1$ w.r.t. Σ_1^0-sentences. In other words: $I\Delta_0 + \Omega_1$ verifies the sentential Σ_1^0-conservativity of schematical, sentential Σ_1^0-completeness over $I\Delta_0 + \Omega_1$. This

Received October 2, 1990; revised February 27, 1991

fact gives rise to a rather natural system of provability logic. Let us add to the language of Löb's Logic propositional variables $s, s', \ldots$ for Σ_1^0-sentences. If we consider interpretations in a theory U extending $I\Delta_0 + \mathrm{EXP}$ (with Σ_1^0 provability predicate), the resulting arithmetically valid and arithmetically complete logic is Löb's Logic + $\{s \to \Box s \mid s \text{ a } \Sigma\text{-variable}\}$. (The proof is surprisingly easy: see Visser [8].) If we consider interpretations for the extended language in $I\Delta_0 + \Omega_1$, we can (by our present lights) only justify the system Löb's Logic + $\{\Box(\bigwedge\{s \to \Box s \mid s \text{ in } S\} \to s^*) \to \Box s^* \mid s^* \text{ is a } \Sigma\text{-variable}, S \text{ is a finite set of } \Sigma\text{-variables}\}$.

This logic is useful, for example, if one wants to formalize metamathematical reasoning involving the Rosser-ordering in $I\Delta_0 + \Omega_1$ (see the following article by Carbone on provable fixed points).

2 Prerequisites The reader should be acquainted with [1], [4], and Smorynski [5].

3 Programming cuts Let U be an arithmetical theory. A U-cut will be (in this paper) a formula $I(x)$, having only x free, such that U proves that $0 \in I$, that I is closed under successor, addition, multiplication, and ω_1, and that I is downwards closed w.r.t. $<$. If we speak simply about a cut, we mean: $I\Delta_0 + \Omega_1$-cut. We write A^I for the result of relativizing all quantifiers in A to I.

Let I and J be $I\Delta_0 + \Omega_1$-cuts. Define:

$I \leq J \quad :\Leftrightarrow I\Delta_0 + \Omega_1 \vdash \forall x(x \in I \to x \in J).$

$I = J \quad :\Leftrightarrow I \leq J$ and $J \leq I$.

$x \in \mathrm{ID} \quad :\Leftrightarrow x = x.$

$x \in I \circ J \quad :\leftrightarrow x \in J \wedge (x \in I)^J.$

$x \in I[A]J :\leftrightarrow (A \wedge x \in I) \vee (\neg A \wedge x \in J)$. (Here A is a sentence.)

We enumerate some elementary facts about cuts. The proofs are left to the diligent reader.

1. $I \circ J$ is a cut. The proof uses that $I\Delta_0 + \Omega_1 \vdash (I \text{ is a cut})^J$. Note that this would not work if we were considering $I\Delta_0 + \mathrm{EXP}$ and $I\Delta_0 + \mathrm{EXP}$-cuts instead of $I\Delta_0 + \Omega_1$ and $I\Delta_0 + \Omega_1$-cuts.
2. ID is a cut. Cuts are closed under union and intersection and $(\cdot)[A](\cdot)$.
3. $=$ is a congruence relation w.r.t. $\cap$, $\cup$, $\circ$ and $(\cdot)[A](\cdot)$ and $\leq$ is a po w.r.t. cuts modulo $=$.
4. ID is the identity w.r.t. $\circ$. Moreover ID is the maximum w.r.t. $\leq$.
5. $I \circ J \leq J$.
6. $I \leq I' \Rightarrow (I \circ J) \leq (I' \circ J)$.
7. $(I \cap I') \circ J = (I \circ J) \cap (I' \circ J)$.
8. $(I \cup I') \circ J = (I \circ J) \cup (I' \circ J)$.
9. For A a sentence: $I\Delta_0 + \Omega_1 \vdash A^{I \circ J} \leftrightarrow (A^I)^J$.
10. $\circ$ is associative.
11. $I\Delta_0 + \Omega_1 \vdash B^{I[A]J} \leftrightarrow ((A \wedge B^I) \vee (\neg A \wedge B^J))$
12. $I \circ (J[A]K) = (I \circ J)[A](I \circ K)$.
13. $(I[A]J) \circ K = (I \circ K)[A^K](J \circ K)$.

4 On schematical, sentential Σ^0_1-completeness Define $C(S) := \bigwedge\!\!\!\bigwedge \{s \to \Box s \mid s$ in $S\}$, where S is a finite set of Σ^0_1-sentences and where $\Box$ is provability in $I\Delta_0 + \Omega_1$. We have: for every S there is a cut $J(S)$ such that: $I\Delta_0 + \Omega_1 \vdash C(S)^{J(S)}$.

Proof: There exists a cut I such that for any Σ^0_1-sentence $s: I\Delta_0 + \Omega_1 \vdash s^I \to \Box s$. (Let me briefly sketch the proof that such a cut exists. The usual proof of Σ^0_1-completeness transforms a witness x of s into a witness p of $\Box s$. A crude estimate shows that $p \leq \exp(\exp(|s|^{|s|+k} \cdot |x|))$, where k is a fixed standard number, $|y| := \mathrm{entier}({}^2\log(y+1))$, $\exp(y) := 2^y$. Let $s := \exists x s_0(x)$, where $s_0 \in \Delta^0_1$. Using the estimate we can show in $I\Delta_0 + \Omega_1$:

$$\forall x, z((s_0(x) \wedge \exp(\exp(|s|^{|s|+k} \cdot |x|)) = z) \to \exists p \leq z \ \mathrm{Proof}(p, s).$$

Let I be any cut such that $I\Delta_0 + \Omega_1$ shows: $x \in I \to \exp(x)\downarrow$. Using the fact that s is standard and the closure properties of I we can easily show $I\Delta_0 + \Omega_1 \vdash s^I \to \Box s$.)

The proof of our theorem is by induction on the cardinality of S. Put $J(\varnothing) :=$ ID. Note that $C(\varnothing) = \top$. Suppose $S := S^* \cup \{s^*\}$, where $s^* \notin S^*$. Put $J(S) := (\mathrm{ID}[s^* \to \Box s^*]\,(J(S^*) \circ I)) \circ J(S^*)$. (Evidently our construction as it stands does not give a unique result. It can be made unique, e.g., by using some ordering of Σ^0_1-sentences.)

By the Induction Hypothesis $I\Delta_0 + \Omega_1 \vdash C(S^*)^{J(S^*)}$. Note that also: $I\Delta_0 + \Omega_1 \vdash C(S^*)^{J(S^*) \circ I \circ J(S^*)}$, because $I\Delta_0 + \Omega_1$ is again valid on $J(S^*) \circ I \circ J(S^*)$.

Reason in $I\Delta_0 + \Omega_1$ and reason 'inside' $J(S^*)$: we have $C(S^*)$ and $C(S^*)^{J(S^*) \circ I}$. In case $s^* \to \Box s^*$, clearly $C(S)$ and ipso facto $C(S)^{\mathrm{ID}}$. Otherwise it follows that $\neg \Box s^*$ and hence $(\neg s^*)^I$ (since $s^{*I} \to \Box s^*$). By the downward persistence of Π^0_1-sentences, also $(\neg s^*)^{J(S^*) \circ I}$ and thus $(s^* \to \Box s^*)^{J(S^*) \circ I}$. Combining this with $C(S^*)^{J(S^*) \circ I}$ we find: $C(S)^{J(S^*) \circ I}$. So we may conclude: $C(S)^{\mathrm{ID}[s^* \to \Box s^*]\,(J(S^*) \circ I)}$.

There is an alternative proof that is conceptually very simple: (in $I\Delta_0 + \Omega_1$) consider the set of true elements of S. Go inside I. Inside I the same elements of S are either true or less (because we can only lose witnesses). In the first case we are done: for any s in S we have: if s then s^I then $\Box s$. In case we have less, repeat the procedure inside I. This can go on no more than n times, because after each step S is left with strictly fewer truths and S contains only n elements. So in all cases we finish with $C(S)$! Below I give the alternative proof in a slightly more formal style.

Alternative proof: Let I be as before. Suppose the cardinality of S is n. Define $\mathrm{FIX}(S) := \bigwedge\!\!\!\bigwedge \{s \leftrightarrow s^I \mid s \in S\}$. Let $J_0(S) :=$ ID and $J_{k+1}(S) := \mathrm{ID}[\mathrm{FIX}(S)]\,(J_k(S) \circ I)$ and $J(S) := J_n(S)$. Reasoning in $I\Delta_0 + \Omega_1$ one easily sees that each time the right-hand side is chosen strictly fewer elements of S will be true. If this happens n times no elements will be left and $C(S)$ is trivially true. Otherwise at some stage k $\mathrm{FIX}(S)$ is true. Clearly $\mathrm{FIX}(S)$ implies $C(S)$.

Remark Let K be any $I\Delta_0 + \Omega_1$-cut. Define $K^0 :=$ ID, $K^{n+1} := K \circ K^n$. It is a nice exercise to show that for the $J_k(S)$ of the alternative proof we have: $J_k(S) = (\mathrm{ID}[\mathrm{FIX}(S)]\,I)^k$. (Hint: use 10 and 12 of Section 3.)

The sentential Σ_1^0-conservativity of schematical, sentential Σ_1^0-completeness:

$$\text{for all } S \text{ and } s: I\Delta_0 + \Omega_1 \vdash \Box(C(S) \rightarrow s) \rightarrow \Box s.$$

Proof: Reason in $I\Delta_0 + \Omega_1$: Suppose $\Box(C(S) \rightarrow s)$. Then $\Box s^{J(S)}$. Ergo: $\Box s$.

Remarks

i. It is an open question whether $I\Delta_0 + \Omega_1$ verifies the Σ_1^0-conservativity of full sentential Σ_1^0-completeness. As is easily seen it is sufficient to show: $I\Delta_0 + \Omega_1 \vdash \forall S, s\Box(C(S) \rightarrow s) \rightarrow \Box s$. I conjecture that this is the case. My reasons for believing this conjecture are given in a note.[1]

ii. Can we get e.g.: $I\Delta_0 + \Omega_1 \vdash \Box(\forall x C(S(x)) \rightarrow \forall x s(x)) \rightarrow \Box \forall x s(x)$, where $S(x)$ is a finite set of Σ_1^0-formulas having only x free and $s(x)$ is a Σ_1^0-formula having only x free? We can see that this is a difficult problem by the following argument due to Dick de Jongh: let $A(x)$ be a coNP-complete Π_1^b formula. Let $S(x) := \{A(x)\}$ and $s(x) := C(S(x))$. From the principle under consideration it would follow that $\Box\forall x(A(x) \rightarrow \Box A(x))$. The considerations in the introduction show that we cannot hope for an easy proof of this fact.

Corollary *Let L be Löb's Logic. Let I be an $I\Delta_0 + \Omega_1$-cut. An interpretation $(\cdot)^*$ of the modal language is an Ia-interpretation if $\Box A$ is interpreted as $\Box_{I\Delta_0+\Omega_1} A^{*I}$. $(\cdot)^*$ is an Ib-interpretation if $\Box A$ is interpreted as $\Box^I_{I\Delta_0+\Omega_1} A^*$. We have*:

(a) $L \vdash A \Leftrightarrow$ *for all $I\Delta_0 + \Omega_1$-cuts I and all Ia-interpretations* $(\cdot)^* : I\Delta_0 + \Omega_1 \vdash A^{*I}$.

(b) $L \vdash A \Leftrightarrow$ *for all $I\Delta_0 + \Omega_1$-cuts I and all Ib-interpretations* $(\cdot)^* : I\Delta_0 + \Omega_1 \vdash A^*$.

Sketch of the proof: We prove (a) and (b) simultaneously. The soundness side is trivial. Suppose $L \nvdash A$. Let K be a countermodel with extra node 0 added below. Say the domain of K is $\{0, \dots, n\}$. Define:

$$h(0) := 0,$$
$$h(n+1) := i \text{ if } h(n)Ri \text{ and } \mathrm{Proof}_{I\Delta_0+\Omega_1}(n, (L \neq \underline{i})^J),\ h(n+1) := h(n) \text{ otherwise},$$
$$L = i :\Leftrightarrow \exists x\, h(x) = i \wedge \forall y, z((h(y) = i \wedge z > y) \rightarrow h(z) = i),$$
$$J := J(\{\exists x\, h(x) = 1, \dots, \exists x\, h(x) = n\}).$$

It is easily seen that this definition can be made to work in $I\Delta_0 + \Omega_1$, using the Fixed Point Lemma to get both L and J. Note that L and J only occur as codes in the definition of h. Let me briefly indicate why h is provably total in $I\Delta_0 + \Omega_1$: first the function $\lambda A, J \cdot A^J$ can be formalized and proved total in $I\Delta_0 + \Omega_1$: the reason is that the recursion in its definition is over subformulas. (This fact is verified in detail in Kalsbeek [3].) Using this we can show that the function that assigns (a code for) $(L \neq \underline{d})^J$ to d, H, where H is a code for a formula defining h, is definable in $I\Delta_0 + \Omega_1$ and provably total. Define FCF(σ) (for: "σ codes a Finitely Changing Function") by:

$$\mathrm{FCF}(\sigma) :\Leftrightarrow ((\sigma)_0)_0 = 0 \wedge (\forall u < \mathrm{lth}(\sigma) \exists v, w < \sigma (\sigma)_u = \langle v, w \rangle) \wedge \forall u, v < \mathrm{lth}(\sigma)$$
$$(u < v \rightarrow ((\sigma)_u)_0 < ((\sigma)_v)_0).$$

Define further (for σ such that $\mathrm{FCF}(\sigma)$):

$$\sigma(x) = y :\Leftrightarrow \exists u < \mathrm{lth}(\sigma)\exists v < \sigma(v \leq x \wedge (\sigma)_u$$
$$= \langle v, y\rangle \wedge \forall w < \mathrm{lth}(\sigma)(u < w \rightarrow x < ((\sigma)_w)_0)).$$

It is easily seen that under this definition σ represents a function, when $\mathrm{FCF}(\sigma)$.

Let $B(x) := \langle\langle x,0\rangle,\langle x,1\rangle,\ldots,\langle x,n\rangle\rangle$; for a decent coding of sequences $B(x)$ is of order x^k for some standard k. We can write the equivalence proved by the Fixed Point Lemma as follows:

$$\begin{aligned}
&h(x) = y \leftrightarrow \exists\sigma < B(x)\\
&\quad(\mathrm{FCF}(\sigma) \wedge \sigma(x) = y \wedge \sigma(0) = 0 \wedge \forall z \leq x \exists u \leq z \exists d \leq n\\
&\quad\ (\sigma(u) = \sigma(z) = d \wedge\\
&\quad\ \ (u = 0 \vee \exists v < u \exists e \leq n\\
&\quad\ \ \ (u = v + 1 \wedge \mathrm{Proof}_{I\Delta_0+\Omega_1}(v, (L \neq \underline{d})^J) \wedge \sigma(v) = e \wedge eRd \wedge \forall w < z\\
&\quad\ \ \ \ (v < w \rightarrow \forall f \leq n\\
&\quad\ \ \ \ \ (\neg eRf \vee \neg\mathrm{Proof}_{I\Delta_0+\Omega_1}(w, (L \neq f)^J))))))).
\end{aligned}$$

The existence of L is trivial, the range of h being standardly finite.

To make the usual Solovay's argument work it is sufficient to provide sentences λi $(i = 0,\ldots,n)$, where we define $p^* := \bigvee\{\lambda i \mid i \Vdash p\}$ and where the λi satisfy:

(i) $\lambda 0$ is true,
(ii) $\vdash \bigvee\{\lambda i \mid i = 0,\ldots,n\}$,
(iii) $i \neq j \Rightarrow \vdash \neg(\lambda i \wedge \lambda j)$,
(iv) $iRj \Rightarrow \vdash \lambda i \rightarrow \Diamond\lambda j$,
(v) $i \neq 0 \Rightarrow \vdash \lambda i \rightarrow \Box\bigvee\{\lambda j \mid iRj\}$.

Here for the proof of (a): $\vdash A$ means $I\Delta_0 + \Omega_1 \vdash A^J$ and $\Box A$ means $\Box_{I\Delta_0+\Omega_1}A^J$. For the proof of (b): $\vdash A$ means $I\Delta_0 + \Omega_1 \vdash A$ and $\Box A$ means $\Box^J_{I\Delta_0+\Omega_1}A$.

Define for (a): $\lambda i := (L = i)$; for (b): $\lambda i := (L = i)^J$. Note that in this way the ultimate meanings of (i)–(v) are precisely the same for (a) and for (b). We leave it to the reader to verify from (i)–(v) the Embedding Lemma: for $i \neq 0$:

$$i \Vdash A \Rightarrow \vdash \lambda i \rightarrow A^*,$$
$$i \nVdash A \Rightarrow \vdash \lambda i \rightarrow \neg A^*.$$

We also leave to the reader the proof of Solovay's Theorem from the Embedding Lemma.

We turn to (i)–(v). The only interesting case to verify is (v). The crucial step is the verification of: for $i \neq 0$

$$I\Delta_0 + \Omega_1 \vdash (\exists x\, hx = i \rightarrow \Box_{i\Delta_0+\Omega_1}(\exists x\, hx = i)^J)^J.$$

Our construction of J gives us:

$$I\Delta_0 + \Omega_1 \vdash (\exists x\, hx = i \rightarrow \Box_{I\Delta_0+\Omega_1}\exists x\, hx = i)^J.$$

Moreover it is well known that:

$$I\Delta_0 + \Omega_1 \vdash (\Box_{I\Delta_0+\Omega_1}B \rightarrow \Box_{I\Delta_0+\Omega_1}B^J)^J.$$

Combining these two results we are done.

Corollary *$I\Delta_0 + \Omega_1 + \neg\mathrm{EXP} + \{s \to \Box s \mid s$ is a Σ_1^0-sentence$\}$ is locally interpretable in $I\Delta_0 + \Omega_1$.*

Proof: Let S be any finite set of Σ_1^0-sentences. Let us write $A \rhd B$ for: $I\Delta_0 + \Omega_1 + B$ is interpretable in $I\Delta_0 + \Omega_1 + A$. We use the principles for interpretability of the system ILW verified in Visser [9].

We have by our theorem $\top \rhd C(S)$ and (because the interpretation is a cut) $\Box(\top \rhd C(S))$; hence $\Box(\Diamond\top \to \Diamond C(S))$. By a result of Paris and Wilkie: $\mathrm{EXP} \rhd \Diamond\top$, so $\mathrm{EXP} \rhd \Diamond C(S)$. By the principles W and J5:

$$\mathrm{EXP} \rhd (\Diamond C(S) \wedge \Box\neg\mathrm{EXP}) \rhd \Diamond(C(S) \wedge \neg\mathrm{EXP}) \rhd (C(S) \wedge \neg\mathrm{EXP}).$$

Also: $(C(S) \wedge \neg\mathrm{EXP}) \rhd (C(S) \wedge \neg\mathrm{EXP})$. Hence by J3:

$$\top \rhd C(S) \rhd (\mathrm{EXP} \vee (C(S) \wedge \neg\mathrm{EXP})) \rhd (C(S) \wedge \neg\mathrm{EXP}).$$

Remark The fact that $\top \rhd \neg\mathrm{EXP}$ was first proved by Solovay in 1986. This was unknown to me when writing [9]. Solovay's proof is quite different from ours.

4 The s-system Let s-L be Löb's Logic in a language with two sorts of propositional variables: the usual $p, q, r, p', \ldots, p_1, p_2, \ldots$ and $s, s', \ldots, s_1, s_2, \ldots$. The s-variables stand for Σ_1^0-sentences. Let Σ be the smallest class of formulas in the enriched language such that formulas of the form $\bot, \top, \Box A, s$ are in Σ (for any formula A), and if B, C are in Σ, then so are $(B \vee C)$ and $(B \wedge C)$. s-L has the following additional rules:

s-Principle $\vdash \Box(C(S) \to s) \to \Box s$, for S a finite set of s-variables,
Substitution $\vdash A(p_1, \ldots, p_n, s_1, \ldots, s_n) \Rightarrow \vdash A(B_1, \ldots, B_n, \sigma_1, \ldots, \sigma_m)$, for any formulas $B_1, \ldots, B_n$ and for any $\sigma_1, \ldots, \sigma_m$ in Σ.

Equivalently we could take instead of s-Principle plus Substitution:

s^+-Principle $\vdash \Box(C(X) \to \sigma) \to \Box\sigma$, for X a finite subset of Σ and $\sigma \in \Sigma$.

An interpretation $(\cdot)^*$ of the language of s-L is a function from the elements of this language to sentences of the language of arithmetic, which satisfies the following conditions:

- $(s)^* \in \Sigma_1^0, (\bot)^* = \bot, (\top)^* = \top$,
- $(\cdot)^*$ commutes with the propositional connectives,
- $(\Box A)^* = \Box_{I\Delta_0 + \Omega_1} A^*$.

As is easily seen, s-L is arithmetically valid for interpretations in this sense, i.e.:

$$s\text{-}L \vdash A \Rightarrow \forall(\cdot)^* I\Delta_0 + \Omega_1 \vdash A^*.$$

Evidently the closure of s-L under the rule: $\vdash \Box A \Rightarrow \vdash A$, is also arithmetically valid. I conjecture that s-L is already closed under this rule.

We give some theorems in s-L:

S1 $\vdash \Box \bigvee S \to \Box \bigvee S^+$, *where* $S^+ := \{s \wedge \Box s \mid s \in S\}$
S2 $\vdash (\Box(\Box A \to \bigvee S) \wedge \Box(\bigvee S^+ \to A)) \to \Box A$
S3 $\vdash \Box(C(S) \to (A \to s)) \to \Box^+(\Box A \to \Box s)$, *where* $\Box^+ C := (C \wedge \Box C)$
S4 $\vdash \Box(C(S) \to (\Box s \to s)) \to \Box s$.

Proofs: S1 is trivial. For S2: suppose (in the *s*-System) $\Box(\Box A \to \Bbbk S)$, then $\Box(\Box\Box A \to \Box\Bbbk S)$. Hence by S1: $\Box(\Box\Box A \to \Box\Bbbk S^+)$. Suppose further $\Box(\Bbbk S^+ \to A)$, then $\Box(\Box\Bbbk S^+ \to \Box A)$. Combining: $\Box(\Box\Box A \to \Box A)$, and thus $\Box\Box A$. We may conclude $\Box\Bbbk S$, hence $\Box\Bbbk S^+$, hence $\Box A$.

Ad S3: From $\Box(C(S) \to (A \to s))$, we have $\Box(A \to (C(S) \to s))$. Hence $\Box^+(\Box A \to \Box(C(S) \to s))$. Hence $\Box^+(\Box A \to \Box s)$.

Ad S4: Suppose $\Box(C(S) \to (\Box s \to s))$. By S3: $\Box(\Box\Box s \to \Box s)$, hence $\Box\Box s$. Ergo $\Box(C(S) \to s)$ and thus $\Box s$.

Remark It is now easy to specify a reasonable system for Rosser logic valid in $I\Delta_0 + \Omega_1$. Take Svejdar's system Z (see Svejdar [6]). The validity of Z for interpretations in $I\Delta_0 + \Omega_1$ is verified in detail in Verbrugge [7]. Now add to it the Σ^*-substitution instances of the *s*-Principle, where Σ^* is the smallest class such that formulas of the form $\bot, \top, \Box A, \Box B < \Box C, \Box B \leq \Box C$ are in Σ^*, and if B, C are in Σ^*, then so are $(B \vee C)$ and $(B \wedge C)$. Call the resulting system *Z-s*.[2] Note that *Z-s* is not valid for the interpretations studied by Svejdar. *Z-s* is studied by Carbone and De Jongh. They show that the theorem by Montagna and De Jongh on provable fixed points is true for *Z-s*. See the following article (pp. 562–572) by Carbone: *Provable Fixed Points in* $I\Delta_0 + \Omega_1$. (For the original result by De Jongh & Montagna see: De Jongh & Montagna [2].)

Acknowledgment—The present result was found in direct interaction with the research by Allessandra Carbone and Dick de Jongh on provable Fixed Points. That research in its turn was inspired by ideas of Franco Montagna and by earlier work by De Jongh and Montagna. I thank the anonymous referee whose comments helped to improve the presentation of the paper.

NOTES

1. I sketch a Lakatosian Throught Experiment of which I hope it could be converted into a real proof.

   To formalize our argument in $I\Delta_0 + \Omega_1$ we should provide bounds for the cut $J(S)$ and for the $I\Delta_0 + \Omega_1$-proofs involved. If we follow the first proof of Section 4 it seems to me that $J(S)$ will grow too fast in S. So let us look at the alternative proof. By the remark following the proof we can use as a definition of $J(S)$: $(\mathrm{ID}[\mathrm{FIX}(S)]\,I)^n$. Let K be any cut and add a unary predicate variable X to the language. Let $(x \in K)^X$ be defined in the obvious way. By a method due to Ferrante and Rackoff we can rewrite $(x \in X \wedge (x \in K)^X)$ to a formula $P(x, X)$ with only one occurrence of X. (One needs a language with $\leftrightarrow$.) Let us define $K \circ X$ using $P(x, X)$ rather than the obvious formula. We can convert a proof of "K is a cut" into a proof of "X is a cut $\to K \circ X$ is a cut". Using these facts we can show that the length of K^n is linear in n. Since n is the number of elements of S, 2^n exists in $I\Delta_0 + \Omega_1$ and hence K^n will exist in $I\Delta_0 + \Omega_1$. Furthermore, one can show that the $I\Delta_0 + \Omega_1$-proof that K^n is a cut exists in $I\Delta_0 + \Omega_1$.

   Take $K := \mathrm{ID}[\mathrm{FIX}(S)]\,I$. Our induction hypothesis is: for k (with $0 \leq k \leq n$) we have an $I\Delta_0 + \Omega_1$-proof of: in K^k we have: $\mathrm{Fix}(S)$ or at least k elements of S are false. If we treat this naively then we explicate "at least k elements of S are false" by a big disjunction of conjunctions of negations of elements of S. It is easily seen that this big disjunction is so big that generally it won't exist in $I\Delta_0 + \Omega_1$. The alternative is to use a Σ_1^0-truthpredicate. The only problem is that such a truthpredicate is not available in $I\Delta_0 + \Omega_1$. However we can save ourselves by a trick: we can choose our cut I in such a way that there is (outside I) a Σ_1^0-truthpredicate for I; i.e., there is a predicate T such that $I\Delta_0 + \Omega_1$ proves: for all s there is an $I\Delta_0 + \Omega_1$-proof of $s^I \leftrightarrow T(s)$. Now we do our whole construction inside I using T to formulate "at least k elements of S are false". (Note that we have to convert T in truthpredicates for different cuts for the different k. This is easily done by extracting the witness for s from the witness for T and by demanding that the witness for s is in the desired cut.)

A different way to avoid the big disjunction is to say: there is a 0,1-sequence σ of length n such that 0 occurs at least k times in σ and $\bigwedge\!\!\!\bigwedge \{s_i \leftrightarrow (\sigma)_i = 1 \mid 0 \leq i \leq n\}$.

2. "s-" before a system signals the presence of special variables for Σ_1^0-sentences and that our system contains the s-Principle. "-s" behind a system means that we have substitution instances of the s-Principle for a suitable class of formulas (the 'Σ-formulas' of the system).

REFERENCES

[1] Buss, S., *Bounded Arithmetic*, Bibliopolis, Napoli, 1986.

[2] De Jongh, D. H. J. and F. Montagna, "Provable fixed points," *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, vol. 34 (1988), pp. 229–250.

[3] Kalsbeek, M. B., *An Orey Sentence for Predicative Arithmetic*, ITLI Prepublication Series X-8-01, University of Amsterdam, 1988.

[4] Paris, J. and A. Wilkie, "On the scheme of induction for bounded arithmetic formulas," *Annals for Pure and Applied Logic*, vol. 35 (1987), pp. 261–302.

[5] Smorynski, C., *Self-Reference and Modal Logic*, Springer Verlag, 1985.

[6] Svejdar, V., "Modal analysis of generalized Rosser sentences," *The Journal of Symbolic Logic*, vol. 48 (1983), pp. 986–999.

[7] Verbrugge, *Σ-completeness and Bounded Arithmetic*, ITLI Prepublication Series ML-89-05, University of Amsterdam, 1989.

[8] Visser, A., "A propositional logic with explicit fixed points," *Studia Logica*, vol. 40 (1981), pp. 155–175.

[9] Visser, A., "Interpretability logic," pp. 175–208 in *Mathematical Logic*, edited by P. Petkov, Plenum Press, New York, 1990.

Department of Philosophy
University of Utrecht
P.O. Box 80103
3508 TC, Utrecht
The Netherlands