ARITHMETIC ANALOGUES OF MCALOON'S UNIQUE ROSSER SENTENCES

C. Smoryński

Department of Philosophy

University of Utrecht

Logic Group
Preprint Series
No. 36

Department of Philosophy

University of Utrecht

ARITHMETIC ANALOGUES OF MCALOON'S UNIQUE ROSSER SENTENCES

C. Smoryński

Department of Philosophy
University of Utrecht

April 1988

Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

Arithmetic Analogues of McAloon's Unique Rosser Sentences

C. Smoryński

It is always annoying to read what someone else has to say about one's papers. The writer-- usually a reviewer-- inevitably picks out some small point of tangential interest and expands on it. Such is what I intend to do to *McAloon 1975* here: McAloon prefaces his paper with an abstract which does not even mention the result on which I, perversely enough, wish to focus. This result, as is so subtly hinted in the title of the present note, is the uniqueness of a certain kind of Rosser sentence for **ZF**.

Rosser's original sentence is easily described. Let *Prov(x,y)* express "*x* proves *y* " (or, more precisely: "the derivation coded by *x* proves the formula coded by *y* "). The Rosser sentence is then any sentence φ provably satisfying

$$\varphi \leftrightarrow \forall x(Prov(x,\ulcorner\varphi\urcorner) \rightarrow \exists y < x\, Prov(y, \ulcorner\neg\varphi\urcorner)) . \tag{1}$$

A variant of this using the weak inequality in place of the strict one,

$$\varphi \leftrightarrow \forall x(Prov(x,\ulcorner\varphi\urcorner) \rightarrow \exists y \leq x Prov(y, \ulcorner\neg\varphi\urcorner)) , \tag{2}$$

is equivalent for the usual encodings because any derivation proves only one formula.

McAloon obtains his Rosseresque sentences for set theory by stepping temporarily into an infinitary language, or, if one prefers, into a hierarchy of such languages. Specifically, for any admissible ordinal α, let $\mathbf{ZF}_\alpha$ be the formulation of **ZF** in the admissible language of the set L_α with additional axioms,

$$\forall x(x \in \overline{a} \leftrightarrow \mathbb{W}_{b \in a}\, x = \overline{b}), \qquad a \in L_\alpha.$$

There is a finitary formula $Prov^{\infty}(x,y)$ asserting "*x* is an admissible ordinal and $\mathbf{ZF}_x$ proves *y* ". For this formula, McAloon considers sentences φ satisfying,

$$\mathbf{ZF} \vdash \varphi \leftrightarrow \forall\alpha(Prov^{\infty}(\alpha, \ulcorner\varphi\urcorner) \rightarrow Prov^{\infty}(\alpha, \ulcorner\neg\varphi\urcorner)). \tag{3}$$

Observing that sentences φ satisfy (3) iff they satisfy

$$\mathbf{ZF} \vdash \varphi \leftrightarrow \forall\alpha(Prov^{\infty}(\alpha, \ulcorner\varphi\urcorner) \rightarrow \exists\beta \leq \alpha\, Prov^{\infty}(\beta, \ulcorner\neg\varphi\urcorner)), \tag{4}$$

we see that such sentences φ are indeed analogues to Rosser sentences of the form (2). Using the well-ordering of the ordinals, McAloon proved the uniqueness up to **ZF**-provability of sentences satisfying (3). This result can also be proven by appeal to Löb's Theorem-- itself a well-foundedness result of sorts-- using the method of section 1, below.

The main goal of the present note is not to give a new proof of McAloon's result, but to attempt to mirror this result in arithmetic. By "arithmetic" I shall initially mean primitive recursive arithmetic, **PRA**, formulated in the language of ordinary arithmetic with Σ_1-induction. Eventually, I shall mean Peano arithmetic, **PA**. In place of **PRA** and **PA**, one could take any pair $\mathbf{T} \subseteq \mathbf{T'}$ of r.e. extensions of **PRA** of sufficient difference in strength. For the sake of definiteness, however, I shall stick to **PRA** and **PA**.

The "arithmetisation" of McAloon's construction is immediately suggested by rewriting $Prov^{\infty}(\alpha, \ulcorner\varphi\urcorner)$ as $Pr_{ZF_{\alpha}}(\ulcorner\varphi\urcorner)$. Formula (4) becomes

$$\mathbf{ZF} \vdash \varphi \leftrightarrow \forall\alpha[Pr_{ZF_{\alpha}}(\ulcorner\varphi\urcorner) \rightarrow \exists\, \beta \leq \alpha\, Pr_{ZF_{\beta}}(\ulcorner\neg\varphi\urcorner)]. \qquad (5)$$

To obtain arithmetical McAloon-Rosser sentences, I simply replace the hierarchy of admissible set theories,

$$\mathbf{ZF} = \mathbf{ZF}_{\omega} \subseteq \mathbf{ZF}_{\omega_1^{CK}} \subseteq \ldots \subseteq \cup_{\alpha} \mathbf{ZF}_{\alpha},$$

by a recursively enumerable "hierarchy" of arithmetic theories,

$$\mathbf{PRA} \subseteq \mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots \subseteq \cup_{n \in \omega} \mathbf{T_n}.$$

Thus, we get

$$\mathbf{PRA} \vdash \varphi \leftrightarrow \forall x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y \leq x\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)] \qquad (6)$$

as an analogue to (5), whence to (4) and, eventually, (3). Recalling the strict inequality of the original Rosser sentence (1), we have a second analogue,

$$\mathbf{PRA} \vdash \varphi \leftrightarrow \forall x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y < x\;\; Pr_{T_y}(\ulcorner\neg\varphi\urcorner)], \qquad (7)$$

to an unstated set theoretic companion to (5). Under some minimal restraints on the sequence $\{\mathbf{T}_n\}_{n \in \omega}$, both (6) and (7) have fixed points unique up to **PRA**-provable equivalence. I shall prove this in section 1, below.

Sections 2 and 3 are devoted to a more general question: If we let $\mathbf{T} = \cup_{n \in \omega} \mathbf{T_n}$, then formulae satisfying (6) and (7) are presumably Rosseresque sentences for **T**, not for **PRA**. If we relax (6) and (7) to

$$\mathbf{T} \vdash \varphi \leftrightarrow \forall x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y \leq x\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)] \qquad (8)$$

and $$\mathbf{T} \vdash \varphi \leftrightarrow \forall x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \to \exists y < x\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)],\tag{9}$$

respectively, do we have uniqueness up to equivalence provable in **T** of each of the two fixed points? I will prove in section 2 that, if the sequence $\{\mathbf{T_n}\}_{n\in\omega}$ grows sufficiently rapidly in strength, then the answer is yes. In particular, both types of fixed points are **PA**-unique for

$$\mathbf{T_n} = \mathbf{PRA} + \Sigma_{n+1}\text{-Induction.}$$

In section 3, I give a rather feeble counterexample if no growth requirement is made. In section 4, I prove uniqueness (and explicit definability) under a strong non-growth requirement.

The uniqueness proofs for (6) and (7) in section 1 are nearly identical, and the proofs for (8) and (9) in section 2 are still quite similar. Unlike the situation regarding (1) and (2), sentences satisfying (6) and (7) need not be equivalent and both cases must be checked. Indeed, for the generality in which I have described (6) and (7), a divergence of behaviour is readily demonstrated. This is done in the latter part of section 4, where I contrast "Henkin sentences" for the strong non-growth case. The non-uniqueness of such sentences under minimal growth is also observed.

Finally, in section 5, I take a look at the main results of McAloon's paper and prove analogues of them. These analogues demonstrate more readily the possible arithmetic interest of the McAloon-Rosser sentences, an interest obscured by sections 1 - 4 with their almost paedagogical emphasis on illustrating the non-uniqueness of the notion of uniqueness of fixed points.

Before getting down to business, let me introduce two abbreviations that will be useful in the sequel:

$$MPr(z):\ \exists x\, [Pr_{T_x}(z) \wedge \forall y \le x\ \neg Pr_{T_y}(neg(z))]$$

$$MPr'(z):\ \exists x\, [Pr_{T_x}(z) \wedge \forall y < x\ \neg Pr_{T_y}(neg(z))],$$

where $neg(\cdot)$ is the usual function satisfying,

$$neg(\ulcorner\varphi\urcorner) = \ulcorner\neg\varphi\urcorner,$$

for all formulae φ. Using these abbreviations, the McAloon-Rosser sentences of (6) and (8) and of (7) and (9) can be written simply as,

$$\varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)\tag{10}$$

and $\varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner)$, (11)

respectively.

1. Preliminary Uniqueness Results

Currently, the most general tool for proving the uniqueness of self-referential sentences is the modal uniqueness theorem for my system **SR⁻**:

1.1. Definition. **SR⁻** is the system of bimodal logic with language, axioms, and rules of inference as follows:

Language.

Propositional variables: $p, q, r, \ldots$

Truth values: $\top, \bot$

Propositional connectives: $\neg, \wedge, \vee, \rightarrow$

Modal Operators: $\Box, \nabla$.

Axioms.

A1. All boolean tautologies

A2. $\Box A \wedge \Box(A \rightarrow B) \rightarrow \Box B$

A3. $\Box A \rightarrow \Box\Box A$

A4. $\Box(\Box A \rightarrow A) \rightarrow \Box A$

A5. $\Box(A \leftrightarrow B) \rightarrow .\nabla A \leftrightarrow \nabla B$.

Rules.

R1. $A, A \rightarrow B \,/\, B$

R2. $A \,/\, \Box A$.

To state the necessary uniqueness result, let $\boxed{s}A$ abbreviate $A \wedge \Box A$ for modal formulae A.

The following result was proven as Theorem 4.1.8 in *Smoryński 1985* for a slightly stronger theory **SR**. The additional axiom schema of **SR** is, however, not used in the proof.

1.2. Modal Uniqueness Theorem.

$$\mathbf{SR}^{-} \vdash \boxed{s}(p \leftrightarrow \nabla p) \wedge \boxed{s}(q \leftrightarrow \nabla q) \rightarrow (p \leftrightarrow q).$$

The application of this theorem in a specific self-referential context is given first by choosing an r.e. theory **T** containing **PRA** and then interpreting $\Box$ by $Pr_T(\cdot)$. This will guarantee the validity of axiom schemata A2 - A4 and closure under R2, the truth of A1 and closure under R1 coming for free. If one now interprets ∇ by a formula $\rho(x)$ which satisfies,

$$\mathbf{T}\vdash Pr_T(\ulcorner\varphi \leftrightarrow \psi\urcorner) \rightarrow .\rho(\ulcorner\varphi\urcorner) \leftrightarrow \rho(\ulcorner\psi\urcorner), \qquad (*)$$

for all sentences φ, ψ, then schema A5 will also be valid. A formula $\rho(x)$ for which (*) holds will be called **T**-*substitutable* .

1.3. Arithmetic Uniqueness Theorem. Let **T** be an r.e. theory containing **PRA**, and let $\rho(x)$ be **T**-substitutable. If φ, ψ are sentences satisfying,

$$\mathbf{T}\vdash \varphi \leftrightarrow \rho(\ulcorner\varphi\urcorner) \text{ and } \mathbf{T}\vdash \psi \leftrightarrow \rho(\ulcorner\psi\urcorner),$$

then $\mathbf{T}\vdash \varphi \leftrightarrow \psi$.

The proof is very simple: The hypotheses and derivability conditions on $Pr_T(\cdot)$ yield,

$$\mathbf{T}\vdash (\varphi \leftrightarrow \rho(\ulcorner\varphi\urcorner))) \wedge Pr_T(\ulcorner\varphi \leftrightarrow \rho(\ulcorner\varphi\urcorner))\urcorner)$$

$$\mathbf{T}\vdash (\psi \leftrightarrow \rho(\ulcorner\psi\urcorner)) \wedge Pr_T(\ulcorner\psi \leftrightarrow \rho(\ulcorner\psi\urcorner)\urcorner).$$

Interpreting Theorem 1.2 in **T**, we have

$$\mathbf{T}\vdash \text{these things} \rightarrow (\varphi \leftrightarrow \psi),$$

whence $\mathbf{T}\vdash \varphi \leftrightarrow \psi$.

Theorem 1.3 is and is not the most general result one can state. If $\rho(x)$ is **T**-substitutable, then $\neg\rho(x)$, $\rho(\ulcorner\rho(\dot{x})\urcorner)$, etc. have unique fixed points as well, and Theorem 1.3 doesn't state this. However, if $\rho(x)$ is **T**-substitutable, then so are $\neg\rho(x)$, $\rho(\ulcorner\rho(\dot{x})\urcorner)$, etc., whence Theorem 1.3 yields this uniqueness. I refer the reader to Theorem 4.1.8 of *Smoryński 1985* for a discussion of the generality of the result; in the present note I wish only to consider a few specific **T**-substitutable formulae $\rho(x)$.

In fact, the formulae $\rho(x)$ I wish to consider are $\neg MPr(x)$ and $\neg MPr'(x)$, the fixed points of which are the arithmetic versions of McAloon's Rosser sentences. The uniqueness proof applies equally well to McAloon's original set theoretic sentences, but I shall only prove the uniqueness of the arithmetic analogues. In fact, since the proofs for the two types of sentences are virtually identical, I shall only give the details in the one case.

Perhaps the most interesting thing about the result is how little has to be assumed about the sequence $\{\mathbf{T_n}\}_{n\in\omega}$.

1.4. Theorem. Let $\mathbf{T_0}, \mathbf{T_1}, \ldots$ be an r.e. sequence of theories containing **PRA**-- provably so in **PRA**:

$$\mathbf{PRA} \vdash \forall x\, [Pr_{PRA}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\chi\urcorner)], \qquad (*)$$

for all sentences χ. Then: **PRA**-provable fixed points of $\neg MPr(x)$ and $\neg MPr'(x)$ are unique, i.e.

i. if φ, ψ are sentences such that

$$\mathbf{PRA} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner) \text{ and } \mathbf{PRA} \vdash \psi \leftrightarrow \neg MPr(\ulcorner\psi\urcorner),$$

then $\mathbf{PRA} \vdash \varphi \leftrightarrow \psi$;

and ii. if φ, ψ are sentences such that

$$\mathbf{PRA} \vdash \varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner) \text{ and } \mathbf{PRA} \vdash \psi \leftrightarrow \neg MPr'(\ulcorner\psi\urcorner),$$

then $\mathbf{PRA} \vdash \varphi \leftrightarrow \psi$.

Proof: I handle the case of $\neg MPr(x)$. It suffices, by Theorem 1.3, to prove the **PRA**-substitutability of $\neg MPr(x)$. Let θ, χ be any two sentences and observe:

$$\mathbf{PRA} \vdash Pr_{PRA}(\ulcorner\theta \leftrightarrow \chi\urcorner) \rightarrow \forall x Pr_{T_x}(\ulcorner\theta \leftrightarrow \chi\urcorner), \text{ by } (*)$$

$$\vdash Pr_{PRA}(\ulcorner\theta \leftrightarrow \chi\urcorner) \rightarrow \forall x\, [Pr_{T_x}(\ulcorner\theta\urcorner)) \leftrightarrow Pr_{T_x}(\ulcorner\chi\urcorner))],$$

by the derivability conditions, whence pure logic yields

$$\mathbf{PRA} \vdash Pr_{PRA}(\ulcorner\theta \leftrightarrow \chi\urcorner) \rightarrow \forall x\, [Pr_{T_x}(\ulcorner\theta\urcorner) \wedge \forall y \le x\, \neg Pr_{T_y}(\ulcorner\theta\urcorner) \leftrightarrow$$
$$\leftrightarrow Pr_{T_x}(\ulcorner\chi\urcorner)) \wedge \forall y \le x\, \neg Pr_{T_y}(\ulcorner\chi\urcorner))]$$

$$\vdash Pr_{PRA}(\ulcorner\theta \leftrightarrow \chi\urcorner) \rightarrow [MPr(\ulcorner\theta\urcorner) \leftrightarrow MPr(\ulcorner\chi\urcorner))]$$

$$\vdash Pr_{PRA}(\ulcorner\theta \leftrightarrow \chi\urcorner) \rightarrow . \neg MPr(\ulcorner\theta\urcorner) \leftrightarrow \neg MPr(\ulcorner\chi\urcorner)). \text{ QED}$$

2. A Second Look at Uniqueness

As remarked in the introduction, if the sequence $\{\mathbf{T_n}\}_{n\in\omega}$ forms a chain,

$$\mathbf{PRA} \subseteq \mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots \subseteq \cup_{n\in\omega} \mathbf{T_n} = \mathbf{T},$$

then the McAloon-Rosser sentences are sentences about **T** and it is the **T**-provable uniqueness of such sentences that would be nice to have. Stated in such generality, such uniqueness is not always possible. However, under some simple conditions on the sequence $\{\mathbf{T_n}\}_{n\in\omega}$, the stronger uniqueness result obtains. First, there is the condition that the $\mathbf{T_n}$'s provably contain enough arithmetic:

$$\mathbf{PRA} \vdash \forall x\, [Pr_{PRA}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\chi\urcorner)\,], \tag{1}$$

for all sentences χ. Second, there is the condition that the $\mathbf{T_n}$'s provably form a chain:

$$\mathbf{PRA} \vdash \forall x\, y\, [x < y \rightarrow (Pr_{T_x}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_y}(\ulcorner\chi\urcorner))\,], \tag{2}$$

for all sentences χ. Finally, there is a condition asserting that the $\mathbf{T_n}$'s grow in strength:

$$\forall k\, \exists n_k\, \forall n \geq n_k\, (\mathbf{T_{n+1}} \vdash Rfn_{\Sigma_k \cup \Pi_k}(\mathbf{T_n})), \tag{3}$$

where $Rfn_{\Gamma}(\mathbf{T_n})$ is the restriction of the local reflexion schema for $\mathbf{T_n}$ to sentences $\chi \in \Gamma$:

$$Pr_{T_n}(\ulcorner\chi\urcorner) \rightarrow \chi.$$

Note that these conditions do not include the formalisation of (3) in **PRA** or the provability within **PRA** that **T** is the union of the sequence. Such formalisations are only necessary if one wishes to prove the uniqueness results within **PRA**.

Before proving the uniqueness theorems, let me quickly note that these conditions are satisfied by the sequence

$$\mathbf{T_n} = \mathbf{PRA} + \Sigma_{n+1}\text{-Induction},$$

and even by the extremely short sequence,

$$\mathbf{T_0} = \mathbf{PRA}, \quad \mathbf{T_1} = \mathbf{PA},$$

(where we take $n_k = 0$ -- provided we agree to allow finite sequences at all, which will be done in the next section).

2.1. Theorem. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent theories containing **PRA** and satisfying (1) - (3), and let $\mathbf{T} = \bigcup_{n \in \omega} \mathbf{T_n}$. Then:

i. if φ, ψ are sentences such that

$$\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner) \quad \text{and} \quad \mathbf{T} \vdash \psi \leftrightarrow \neg MPr(\ulcorner\psi\urcorner),$$

then $\mathbf{T} \vdash \varphi \leftrightarrow \psi$;

and ii. if φ, ψ are sentences such that

$$\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner) \quad \text{and} \quad \mathbf{T} \vdash \psi \leftrightarrow \neg MPr'(\ulcorner\psi\urcorner),$$

then $\mathbf{T} \vdash \varphi \leftrightarrow \psi$.

Proof: i. First, let $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$.

Let n be large enough so that $\mathbf{T_n}$ proves this equivalence, and also assume $n > n_k$ where $\varphi \in \Sigma_k$. Observe

$$\mathbf{T_n} \vdash \varphi \leftrightarrow \forall x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y \leq x Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]$$

$$\vdash \varphi \leftrightarrow \forall x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\neg\varphi\urcorner)], \text{ by (2).} \tag{4}$$

The universally quantified assertion in (4) splits into two conjuncts,

$$\bigwedge_{k<n} [Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_k}(\ulcorner\neg\varphi\urcorner)] \tag{α}$$

$$\rho_n(\ulcorner\varphi\urcorner):\ \forall x \geq \overline{n}\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\neg\varphi\urcorner)]. \tag{β}$$

I claim that (α) is derivable in $\mathbf{T_n}$. For $k < n$,

$$\mathbf{T_n} \vdash Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow \varphi, \text{ by reflexion}$$

$$\vdash Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow [Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_k}(\ulcorner\neg\varphi\urcorner)], \text{ since } \varphi \rightarrow (\alpha)$$

$$\vdash Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_k}(\ulcorner\neg\varphi\urcorner)$$

$$\vdash (\alpha).$$

It follows that,

$$\mathbf{T_n} \vdash \varphi \leftrightarrow \neg\rho_n(\ulcorner\varphi\urcorner),$$

for $\rho_n(x)$ defined as in (β).

Suppose now that $\mathbf{T} \vdash \psi \leftrightarrow \neg MPr(\ulcorner\psi\urcorner)$. By the same reasoning, $\mathbf{T_n} \vdash \psi \leftrightarrow \neg\rho_n(\ulcorner\psi\urcorner)$ for all sufficiently large n. In particular, φ and ψ are $\mathbf{T_n}$-provably fixed points of $\rho_n(x)$ for some n. But $\rho_n(x)$ is clearly $\mathbf{T_n}$-substitutable, whence $\mathbf{T_n} \vdash \varphi \leftrightarrow \psi$.

ii. This proof follows the same lines, but is a bit more complicated. If $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner)$, then for sufficiently large n,

$$\mathbf{T_n} \vdash \varphi \leftrightarrow \forall x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y < x\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]. \tag{5}$$

The quantified expression in (5) is equivalent to the conjunction of four sentences:

$$\neg Pr_{T_0}(\ulcorner\varphi\urcorner) \tag{α}$$

$$\bigwedge_{0<k<n} [Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow \exists y < \overline{k}\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)] \tag{β}$$

$$Pr_{T_n}(\ulcorner\varphi\urcorner) \rightarrow \exists y < \overline{n}\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \tag{γ}$$

$$\forall x > \overline{n}\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y < x\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]. \tag{δ}$$

This time the claim is that (α) and (β) are provable in $\mathbf{T_n}$ and that (γ) and (δ) can be simplified.

Ad (α): Observe,

$$\mathbf{T_n} \vdash Pr_{T_0}(\ulcorner\varphi\urcorner) \rightarrow \varphi$$

$$\vdash Pr_{T_0}(\ulcorner\varphi\urcorner) \rightarrow \neg Pr_{T_0}(\ulcorner\varphi\urcorner), \text{ since } \varphi \rightarrow (\alpha)$$

$$\vdash \neg Pr_{T_0}(\ulcorner\varphi\urcorner).$$

Ad (β): Start again with reflexion for $0 < k < n$:

$$\mathbf{T_n} \vdash Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow \varphi$$

$$\vdash Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow [Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow \exists y < \overline{k}\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)], \text{ since } \varphi \rightarrow (\beta)$$

$$\vdash Pr_{T_k}(\ulcorner\varphi\urcorner) \rightarrow \exists y < \overline{k}\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)$$

$$\vdash (\beta).$$

Ad (γ): Using reflexion one more time, we have

$$\mathbf{T_n} \vdash \exists y < \overline{n}\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \rightarrow \neg\varphi$$

$$\vdash (\gamma) \rightarrow [Pr_{T_n}(\ulcorner\varphi\urcorner) \rightarrow \neg\varphi]$$

$$\vdash \varphi \rightarrow [Pr_{T_n}(\ulcorner\varphi\urcorner) \rightarrow \neg\varphi] \wedge (\delta), \text{ since } \varphi \rightarrow (\gamma) \wedge (\delta)$$

$$\vdash \varphi \rightarrow \neg Pr_{T_n}(\ulcorner\varphi\urcorner) \wedge (\delta).$$

Conversely,

$$\mathbf{T_n} \vdash \neg Pr_{T_n}(\ulcorner\varphi\urcorner) \wedge (\delta) \rightarrow [Pr_{T_n}(\ulcorner\varphi\urcorner) \rightarrow \exists y < \overline{n}\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)] \wedge (\delta)$$

$$\vdash \neg Pr_{T_n}(\ulcorner\varphi\urcorner) \wedge (\delta) \rightarrow (\gamma) \wedge (\delta)$$

$$\vdash \neg Pr_{T_n}(\ulcorner\varphi\urcorner) \wedge (\delta) \rightarrow \varphi, \text{ since } (\gamma) \wedge (\delta) \rightarrow \varphi.$$

Thus,

$$\mathbf{T_n} \vdash \varphi \leftrightarrow \neg Pr_{T_n}(\ulcorner\varphi\urcorner) \wedge (\delta). \tag{6}$$

Ad (δ): By (2),

$$\mathbf{T_n} \vdash \forall x > \overline{n}\, [\exists y < x Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \rightarrow \exists y < x(\overline{n} \leq y \wedge Pr_{T_y}(\ulcorner\neg\varphi\urcorner)].$$

Thus,

$$\mathbf{T_n} \vdash (\delta) \leftrightarrow \forall x > \overline{n}\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y < x(\overline{n} \leq y \wedge Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]. \tag{7}$$

Using (6) and (7), we see

$$\mathbf{T_n} \vdash \varphi \leftrightarrow \neg\rho_n'(\ulcorner\varphi\urcorner),$$

where

$$\rho_n'(\ulcorner\varphi\urcorner): \neg Pr_{T_n}(\ulcorner\varphi\urcorner) \wedge \forall x > \overline{n}\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow \exists y < x(\overline{n} \leq y \wedge Pr_{T_y}(\ulcorner\neg\varphi\urcorner)].$$

Now $\rho_n'(x)$ is again clearly $\mathbf{T_n}$-substitutable and the uniqueness of φ is readily established.

QED

2.2. Remark. Since $\neg MPr(\ulcorner\varphi\urcorner)$ and $\neg MPr'(\ulcorner\varphi\urcorner)$ are Π_2 , and since reflexion is only applied to φ and $\neg\varphi$ in the proof of Theorem 2.1, it is tempting to weaken (3) to

$$\exists n_0 \forall n \geq n_0 (\mathbf{T_{n+1}} \vdash Rfn_{\Sigma_2 \cup \Pi_2}(\mathbf{T_n})).$$

However, the proof that φ is Π_2 may not be available in the early theories $\mathbf{T_n}$ to which reflexion is applied. If we make this weakening, the proof given will, thus, only prove the uniqueness of fixed points in $\Sigma_2 \cup \Pi_2$.

3. Non-uniqueness; A Counterexample

A positive result is no good unless it is set off by a counterexample showing it to be best possible. Alas, I can only show that *some* growth condition like (3) of the previous section is necessary for the validity of Theorem 2.1. My counterexample may be viewed as a rather artificial construction of a sequence $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ which stops growing, or as a good example of a finite sequence $\mathbf{T_0} \subseteq \mathbf{T_1}$ with a minimal, but insufficient, growth throughout its short length.

3.1. Counterexample. Let $\mathbf{T_0} = \mathbf{PRA}$ (or any Σ_1-sound r.e. extension thereof) and $\mathbf{T_1} = \mathbf{T_0} + Con_{T_0}$. There are sentences φ,ψ such that

i. $\mathbf{T_1} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$ and $\mathbf{T_1} \vdash \psi \leftrightarrow \neg MPr(\ulcorner\psi\urcorner)$

ii. $\mathbf{T_1} \vdash \varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner)$ and $\mathbf{T_1} \vdash \psi \leftrightarrow \neg MPr'(\ulcorner\psi\urcorner)$,

and yet

iii. $\mathbf{T_1} \nvdash \varphi \leftrightarrow \psi$.

The proof is a simple application of Solovay's Second Completeness Theorem. In applying this Theorem, I follow my exposition in *Smoryński 1985* , Chapter III, section 2, in matters of notation. One tiny exception is this: I abbreviate $\Box(\neg\Box\bot \rightarrow A)$ (i.e. $Pr_{T_1}(\ulcorner A\urcorner)$) by ∇A . In any Kripke model, one will have

$$\alpha \Vdash \nabla A \text{ iff } \forall\beta > \alpha\, (\beta \text{ not terminal} \Rightarrow \beta \Vdash A).$$

The modal counterpart to $\varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$ is the formula,

$$p \leftrightarrow (\Box p \rightarrow \Box\neg p) \wedge (\nabla p \rightarrow \nabla\neg p).$$

The assertion of its provability in $\mathbf{T_1}$ reads,

$$\nabla[p \leftrightarrow (\Box p \rightarrow \Box\neg p) \wedge (\nabla p \rightarrow \nabla\neg p)].$$

The modal counterpart to $\varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner)$ and the assertion of its provability in $\mathbf{T_1}$ read,

$$p \leftrightarrow \neg\Box p \wedge (\nabla p \rightarrow \Box\neg p)$$

and $$\nabla[p \leftrightarrow \neg\Box p \wedge (\nabla p \rightarrow \Box\neg p)],$$

respectively.

By Solovay's Second Completeness Theorem, we can establish Theorem 3.1 by constructing a Kripke model $\underline{K} = (K,<,\alpha_0,\Vdash)$ of the provability logic **PrL** satisfying:

fixed point assertions

$$\alpha_0 \Vdash \nabla[p \leftrightarrow (\Box p \rightarrow \Box\neg p) \wedge (\nabla p \rightarrow \nabla\neg p)] \quad (1)$$

$$\alpha_0 \Vdash \nabla[q \leftrightarrow (\Box q \rightarrow \Box\neg q) \wedge (\nabla q \rightarrow \nabla\neg q)] \quad (2)$$

$$\alpha_0 \Vdash \nabla[p \leftrightarrow \neg\Box p \wedge (\nabla p \rightarrow \Box\neg p)] \quad (3)$$

$$\alpha_0 \Vdash \nabla[q \leftrightarrow \neg\Box q \wedge (\nabla q \rightarrow \Box\neg q)] \quad (4)$$

unprovability of the equivalence

$$\alpha_0 \Vdash \neg\nabla(p \leftrightarrow q) \quad (5)$$

instances of reflexion

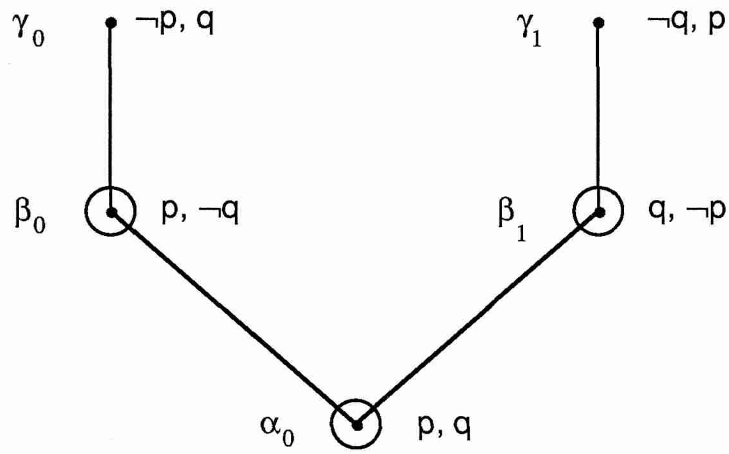
$$\alpha_0 \Vdash \nabla[\text{fixed point assertions}] \rightarrow (\neg\Box\bot \rightarrow \text{fixed point assertions}) \quad (6)$$

$$\alpha_0 \Vdash \Box p \rightarrow p,\ \alpha_0 \Vdash \Box q \rightarrow q,\ \alpha_0 \Vdash \Box\neg p \rightarrow \neg p,\ \alpha_0 \Vdash \Box\neg q \rightarrow \neg q \quad (7)$$

$$\alpha_0 \Vdash \nabla p \rightarrow (\neg\Box\bot \rightarrow p),\ \alpha_0 \Vdash \nabla q \rightarrow (\neg\Box\bot \rightarrow q) \quad (8)$$

$$\alpha_0 \Vdash \nabla(p \leftrightarrow q) \rightarrow (\neg\Box\bot \rightarrow (p \leftrightarrow q)). \quad (9)$$

The following model does all of this:



For convenience I have circled the nodes at which $\neg\Box\bot$ is forced.

To verify (1), observe that $\beta_i \Vdash \nabla A$ for any A. In particular, $\beta_i \Vdash \nabla p \to \nabla\neg p$ and $\beta_i \Vdash \nabla q \to \nabla\neg q$. Moreover,

$$\beta_0 \Vdash p \quad \text{and} \quad \beta_0 \Vdash \Box p \to \Box\neg p \quad (\text{since } \beta_0 \Vdash \Box\neg p\ ),$$

whence

$$\beta_0 \Vdash p \leftrightarrow (\Box p \to \Box\neg p) \wedge (\nabla p \to \nabla\neg p)\,.$$

On the other hand,

$$\beta_1 \Vdash \neg p \quad \text{and} \quad \beta_1 \Vdash \neg(\Box p \to \Box\neg p\,) \ (\text{since } \beta_1 \Vdash \Box p \wedge \neg\Box\neg p\,),$$

whence

$$\beta_1 \Vdash p \leftrightarrow (\Box p \to \Box\neg p) \wedge (\nabla p \to \nabla\neg p)\,.$$

Hence (1) holds.

Assertion (2) holds by a symmetric argument, and (3), (4) hold by similar arguments.

Skipping ahead, note that (7) and (8) hold since

$$\alpha_0 \not\Vdash \Box p\,, \Box\neg p\,, \Box q\,, \Box\neg q\,, \nabla p\,, \nabla q\,.$$

For precisely this reason, we also have

$$\alpha_0 \Vdash (\Box p \to \Box\neg p) \wedge (\nabla p \to \nabla\neg p)$$

$$\alpha_0 \Vdash (\Box q \to \Box\neg q) \wedge (\nabla q \to \nabla\neg q)$$

etc.

But, as $\alpha_0 \Vdash p,q$, we have $\alpha_0 \Vdash p \leftrightarrow (\Box p \to \Box\neg p) \wedge (\nabla p \to \nabla\neg p)$, etc., whence (6) also holds.

Finally, (5) holds since $\beta_i \not\Vdash p \leftrightarrow q$, and (9) holds since $\alpha_0 \Vdash p \leftrightarrow q$. This completes the proof of Counterexample 3.1.

The construction given readily extends to any finite iteration of consistency statements. The real question is the following.

3.2. Open Problem. Define the sequence $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ by

$$\mathbf{T_0} = \mathbf{PRA}$$
$$\mathbf{T_{n+1}} = \mathbf{T_n} + Con_{T_n}.$$

Let **T** be the union of this sequence. Are the **T**-provably McAloon-Rosser sentences for this sequence **T**-provably unique?

4. The Uniqueness Question for Sequences of Constrained Growth

There is another case besides that of strong growth given in section 2 in which uniqueness can be established. This is the case in which the sequence $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ provably does not grow in proof theoretic strength. That is, in addition to some normalising conditions,

$$\mathbf{PRA} \vdash \forall x\, [Pr_{PRA}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\chi\urcorner)], \tag{1}$$

$$\mathbf{PRA} \vdash \forall x\, y\, [x < y \rightarrow (Pr_{T_x}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_y}(\ulcorner\chi\urcorner))], \tag{2}$$

and

$$\mathbf{PRA} \vdash Pr_T(\ulcorner\chi\urcorner) \leftrightarrow \exists x Pr_{T_x}(\ulcorner\chi\urcorner), \tag{3}$$

for all sentences χ, we assume

$$\mathbf{PRA} \vdash \forall x\, (Con_{T_x} \rightarrow Con_{T_{x+1}}). \tag{4}$$

Assertion (3) is a new normality condition asserting **T** to be the union of the $\mathbf{T_n}$'s. Using (2) and (3), (4) readily yields

$$\mathbf{PRA} \vdash \forall x\, (Con_{T_x} \leftrightarrow Con_T). \tag{4'}$$

A trivial example of a sequence satisfying these conditions is the constant sequence,

$$\mathbf{T_0} = \mathbf{T_1} = \ldots = \cup_{n \in \omega} \mathbf{T_n} = \mathbf{PA}.$$

A less trivial example is given by

$$\mathbf{T_0} = \mathbf{PRA}$$
$$\mathbf{T_{n+1}} = \mathbf{T_n} + \text{Rosser}(\mathbf{T_n}),$$

where, by "Rosser($\mathbf{T_n}$)", I mean a genuine Rosser sentence for $\mathbf{T_n}$ as given by formula (1) or (2) of the introduction, above.

4.1. Theorem. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent theories containing **PRA**, let $\mathbf{T} = \cup_{n \in \omega} \mathbf{T_n}$, and assume (1) - (4) are satisfied. Then: For any sentence φ,

i. if $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$, then $\mathbf{T} \vdash \varphi \leftrightarrow Con_T \to Con_{T+Con_T}$

and ii. if $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner)$, then $\mathbf{T} \vdash \varphi \leftrightarrow Con_T$.

This theorem and a second one follow readily from the following lemma.

4.2. Lemma. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent theories containing **PRA**, let $\mathbf{T} = \cup_{n \in \omega} \mathbf{T_n}$, and assume (1) - (4) are satisfied. Then: For any sentence φ,

i. $\mathbf{T} \vdash MPr(\ulcorner\varphi\urcorner) \leftrightarrow Pr_T(\ulcorner\varphi\urcorner) \wedge Con_T$

ii. $\mathbf{T} \vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow Pr_T(\ulcorner\varphi\urcorner)$.

Proof: i. Observe,

$$\mathbf{T} \vdash MPr(\ulcorner\varphi\urcorner) \leftrightarrow \exists x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \neg Pr_{T_x}(\ulcorner\neg\varphi\urcorner)]$$

$$\vdash MPr(\ulcorner\varphi\urcorner) \leftrightarrow \exists x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge Con_{T_x}]$$

$$\vdash MPr(\ulcorner\varphi\urcorner) \leftrightarrow \exists x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge Con_T], \text{ by (4')}$$

$$\vdash MPr(\ulcorner\varphi\urcorner) \leftrightarrow Pr_T(\ulcorner\varphi\urcorner) \wedge Con_T .$$

ii. Observe,

$$\mathbf{T} \vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow \exists x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y < x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]$$

$$\vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow Pr_{T_0}(\ulcorner\varphi\urcorner) \vee \exists x > \overline{0}\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y < x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]. \quad (5)$$

But

$$\mathbf{T} \vdash Con_T \to [Pr_{T_x}(\ulcorner\varphi\urcorner) \to \forall y < x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)] \quad (6)$$

and

$$\mathbf{T} \vdash x > \overline{0} \wedge \forall y < x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \to \forall y < x\, Con_{T_y}$$

$$\vdash x > \overline{0} \wedge \forall y < x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \to Con_T , \quad (7)$$

by (4'). (5), (6) and (7) yield:

$$\mathbf{T} \vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow Pr_{T_0}(\ulcorner\varphi\urcorner) \vee \exists x > \overline{0}\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge Con_T]$$

$$\vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow Pr_{T_0}(\ulcorner\varphi\urcorner) \vee Pr_T(\ulcorner\varphi\urcorner) \wedge Con_T \quad (8)$$

$$\vdash MPr'(\ulcorner\varphi\urcorner) \to Pr_{T_0}(\ulcorner\varphi\urcorner) \vee Pr_T(\ulcorner\varphi\urcorner)$$

$$\vdash\ MPr'(\ulcorner\varphi\urcorner) \to Pr_T(\ulcorner\varphi\urcorner), \tag{9}$$

which is half of what we want.

To obtain the converse of (9), observe that

$$\begin{aligned} \mathbf{T} + Con_T \vdash\ & Pr_T(\ulcorner\varphi\urcorner) \to Pr_T(\ulcorner\varphi\urcorner) \wedge Con_T \\ \vdash\ & Pr_T(\ulcorner\varphi\urcorner) \to Pr_{T_0}(\ulcorner\varphi\urcorner) \vee Pr_T(\ulcorner\varphi\urcorner) \wedge Con_T \\ \vdash\ & Pr_T(\ulcorner\varphi\urcorner) \to MPr'(\ulcorner\varphi\urcorner), \end{aligned} \tag{10}$$

by (8). Also observe,

$$\begin{aligned} \mathbf{T} + \neg Con_T \vdash\ & \neg Con_{T_0}, \text{ by (4')} \\ \vdash\ & Pr_{T_0}(\ulcorner\varphi\urcorner) \\ \vdash\ & MPr'(\ulcorner\varphi\urcorner), \text{ by (8)} \\ \vdash\ & Pr_T(\ulcorner\varphi\urcorner) \to MPr'(\ulcorner\varphi\urcorner). \end{aligned}$$

Together with (10), this yields

$$\mathbf{T} \vdash Pr_T(\ulcorner\varphi\urcorner) \to MPr'(\ulcorner\varphi\urcorner),$$

which with (9) yields the desired conclusion. QED

Via Lemma 4.2, the proof of Theorem 4.1 is a simple matter of calculating the fixed points,

$$\mathbf{T} \vdash\ \varphi \leftrightarrow Pr_T(\ulcorner\varphi\urcorner) \to \neg Con_T\ ,$$

and

$$\mathbf{T} \vdash\ \varphi \leftrightarrow \neg Pr_T(\ulcorner\varphi\urcorner)\ ,$$

respectively, by the known algorithms (e.g. 2.3.15 of *Smoryński 1985*). The same holds for the calculation of the "Henkin" sentences:

<u>4.3. Theorem</u>. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent theories containing **PRA**, let $\mathbf{T} = \cup_{n \in \omega} \mathbf{T_n}$, and assume (1) - (4) are satisfied. Then: For any sentence φ,

i. $\mathbf{T} \vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner)$ iff $\mathbf{T} \vdash \neg\varphi$

ii. $\mathbf{T} \vdash \varphi \leftrightarrow MPr'(\ulcorner\varphi\urcorner)$ iff $\mathbf{T} \vdash \varphi$.

We can paraphrase 4.3 as saying that $\bot$ is the unique Henkin sentence for $MPr(x)$, while $\top$ is the unique one for $MPr'(x)$. Theorem 4.3 is not unusual for the obvious reason that we expect Henkin sentences to be provable: As Kreisel first observed, the Henkin sentences,

$$\mathbf{T} \vdash \varphi \leftrightarrow RPr(\ulcorner\varphi\urcorner),$$

for the "Rosser provability predicate"

$$RPr(z) : \exists x\, [Prov_T(x,z) \wedge \forall y \leq x\ \neg Prov_T(x, neg(z))], \tag{11}$$

include both $\top$ and $\bot$ among their number. The oddity of Theorem 4.3 is that the analogy with Rosser sentences only half holds, with different halves holding for $MPr(x)$ and $MPr'(x)$. The behaviour observed by Kreisel and expected by the cognoscenti returns as soon as a minimal increase in proof theoretic strength is assumed of the sequence. Moreover, as proven by Albert Visser, a bit more occurs.

4.4. Theorem. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent theories containing **PRA** and satisfying (1), and let $\mathbf{T} = \cup_{n \in \omega} \mathbf{T_n}$. Suppose further that $\mathbf{T} \vdash Con_{T_0}$. Then:

i. if $\mathbf{T_0} \vdash \varphi$, then $\mathbf{T} \vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner)$ and $\mathbf{T} \vdash \varphi \leftrightarrow MPr'(\ulcorner\varphi\urcorner)$

ii. if $\mathbf{T_0} \vdash \neg\varphi$, then $\mathbf{T} \vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner)$ and $\mathbf{T} \vdash \varphi \leftrightarrow MPr'(\ulcorner\varphi\urcorner)$

iii. if φ is the Σ_1-form of an ordinary Rosser sentence, i.e. if

$$\mathbf{T} \vdash \varphi \leftrightarrow RPr(\ulcorner\neg\varphi\urcorner),$$

with $RPr(z)$ as in (11), then $\mathbf{T} \vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner)$ and $\mathbf{T} \vdash \varphi \leftrightarrow MPr'(\ulcorner\varphi\urcorner)$,

and iv. there are infinitely many pairwise $\mathbf{T}$-inequivalent Henkin sentences for $MPr(x)$ and $MPr'(x)$.

Proof: The proofs of iii and iv can be obtained by translating the proofs in *Visser A* of the corresponding result for the Henkin sentences for the Feferman predicate into the present context. The proofs of i and ii are both trivial and repetitive, but I shall present them anyway in order to illustrate where the strict assumption that φ be $\mathbf{T_0}$-provable or $\mathbf{T_0}$-refutable (as opposed to $\mathbf{T}$-provable or $\mathbf{T}$-refutable) is used.

i. Assume $\mathbf{T_0} \vdash \varphi$ and observe,

$$\mathbf{T} \vdash Pr_{T_0}(\ulcorner\varphi\urcorner) \wedge \forall y \leq \overline{0}\ \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner), \text{ since } \mathbf{T} \vdash Con_{T_0}$$

$$\vdash MPr(\ulcorner\varphi\urcorner)$$

$$\vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner).$$

Also,

$$\mathbf{T} \vdash Pr_{T_0}(\ulcorner\varphi\urcorner) \wedge \forall y < \overline{0}\ \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)$$

$$\vdash MPr'(\ulcorner\varphi\urcorner)$$

$$\vdash \varphi \leftrightarrow MPr'(\ulcorner\varphi\urcorner).$$

(Observe that this latter proof makes no use of the assumption that $\mathbf{T}\vdash Con_{T_0}$ and affords us a simple proof of the right-to-left implication of 4.3.ii in the case $\mathbf{T_0}\vdash\varphi$.)

ii. Assume $\mathbf{T_0}\vdash \neg\varphi$ and observe,

$$\mathbf{T}\vdash MPr(\ulcorner\varphi\urcorner) \leftrightarrow \exists x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y \le x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]$$

$$\vdash \neg MPr(\ulcorner\varphi\urcorner),$$

since $\mathbf{T}\vdash Pr_{T_0}(\ulcorner\neg\varphi\urcorner) \rightarrow \forall y\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)$. Thus

$$\mathbf{T}\vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner).$$

(Again, we have not made use of the assumption that $\mathbf{T}\vdash Con_{T_0}$)

Next, observe

$$\mathbf{T}\vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow \exists x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y < x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]$$

$$\vdash MPr'(\ulcorner\varphi\urcorner) \leftrightarrow Pr_{T_0}(\ulcorner\varphi\urcorner), \qquad (12)$$

since $\mathbf{T}\vdash Pr_{T_0}(\ulcorner\neg\varphi\urcorner) \rightarrow \forall x > \overline{0}\; \forall y < x\, Pr_{T_y}(\ulcorner\neg\varphi\urcorner)$. But

$$\mathbf{T}\vdash Pr_{T_0}(\ulcorner\neg\varphi\urcorner) \wedge Con_{T_0} \rightarrow \neg Pr_{T_0}(\ulcorner\varphi\urcorner)$$

$$\vdash \neg Pr_{T_0}(\ulcorner\varphi\urcorner), \text{ since } \mathbf{T}\vdash Pr_{T_0}(\ulcorner\neg\varphi\urcorner) \wedge Con_{T_0}$$

$$\vdash \neg MPr'(\ulcorner\varphi\urcorner), \text{ by (12)}$$

$$\vdash \neg\varphi \leftrightarrow \neg MPr'(\ulcorner\varphi\urcorner)$$

$$\vdash \varphi \leftrightarrow MPr'(\ulcorner\varphi\urcorner).$$

QED

The proof made essential use of the fact that the provability or refutability of φ was in the theory $\mathbf{T_0}$ whose consistency is provable in $\mathbf{T}$. Thus, e.g., to conclude

$$\mathbf{T_n}\vdash\varphi \Rightarrow \varphi \text{ is a McAloon-Rosser-Henkin sentence,}$$

for $n > 0$ would require in the above proof the assumption that $\mathbf{T}\vdash Con_{T_n}$. That this is not a feature of the proof, but a genuine restriction is readily demonstrated.

4.5. Example. Consider the sequence

$$\mathbf{T_0} = \mathbf{PRA}$$

$$\mathbf{T_1} = \mathbf{T_0} + Con_{T_0}$$

$$\mathbf{T_{n+2}} = \mathbf{T_{n+1}} + \text{Rosser}(\mathbf{T_{n+1}}).$$

For this sequence, $\mathbf{T} \vdash Con_{T_0}$, but $\mathbf{T} \nvdash Con_{T_0} \leftrightarrow MPr(\ulcorner Con_{T_0} \urcorner)$.

Proof: Let φ abbreviate Con_{T_0}, and observe that the assumption $\mathbf{T} \vdash \varphi \leftrightarrow MPr(\ulcorner\varphi\urcorner)$

yields successively,

$$\mathbf{T} \vdash \varphi \leftrightarrow \exists x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y \le x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)]$$

$$\vdash \exists x\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y \le x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)], \text{ since } \mathbf{T} \vdash \varphi$$

$$\vdash Pr_{T_0}(\ulcorner\varphi\urcorner) \wedge Con_{T_0} \vee \exists x > \overline{0}\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge Con_{T_x}]. \qquad (13)$$

But $\mathbf{T} \vdash \neg Pr_{T_0}(\ulcorner\varphi\urcorner)$ by Gödel's Second Incompleteness Theorem, whence (13) yields

$$\mathbf{T} \vdash \exists x > \overline{0}\,[Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge Con_{T_x}]$$

$$\vdash \exists x > \overline{0}\; Con_{T_x}$$

$$\vdash Con_T,$$

contrary to the Second Incompleteness Theorem. QED

I leave it to the reader to generalise this Example to show the more general necessity of assuming $\mathbf{T} \vdash Con_{T_n}$ in establishing the Henkinness of all theorems of $\mathbf{T_n}$.

5. McAloon's Paper Revisited

In the present section, we assume given an ascending r.e. sequence $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots \subseteq \bigcup_{n \in \omega} \mathbf{T_n} = \mathbf{T}$ of consistent extensions of **PRA**. For the sake of brevity, we will only consider McAloon-Rosser sentences based on $MPr(x)$.

McAloon's simplest result-- one I have not yet explicitly cited-- is the independence of the McAloon-Rosser sentences.

5.1. Lemma. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent theories containing **PRA**, and let $\mathbf{T} = \bigcup_{n \in \omega} \mathbf{T_n}$. Assume $\mathbf{T}$ is Σ_1-sound and $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$. Then: φ is independent of $\mathbf{T}$.

Proof: First, observe

$$\mathbf{T} \vdash \varphi \Rightarrow \mathbf{T_n} \vdash \varphi, \text{ for some } n$$

$$\Rightarrow \mathbf{T_n} \vdash \varphi \wedge Pr_{T_n}(\ulcorner\varphi\urcorner)$$

$$\Rightarrow \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\neg\varphi\urcorner), \text{ by definition of } MPr(\ulcorner\varphi\urcorner)$$

$$\Rightarrow \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\bot\urcorner) \qquad (1)$$

$$\Rightarrow Pr_{T_n}(\ulcorner\bot\urcorner) \text{ is true, by } \Sigma_1\text{-soundness}$$

$\Rightarrow \mathbf{T_n} \vdash \perp$, a contradiction.

Next, observe

$$\mathbf{T} \vdash \neg\varphi \Rightarrow \mathbf{T_n} \vdash \neg\varphi, \text{ for some } n$$
$$\Rightarrow \mathbf{T_n} \vdash \exists x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \neg Pr_{T_x}(\ulcorner\neg\varphi\urcorner)]$$
$$\Rightarrow \mathbf{T_n} \vdash \exists x < \overline{n}\, Pr_{T_x}(\ulcorner\varphi\urcorner) \text{ , since } \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\neg\varphi\urcorner)$$
$$\Rightarrow \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\varphi\urcorner)$$
$$\Rightarrow \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\perp\urcorner) \text{, since } \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\neg\varphi\urcorner) \qquad (2)$$
$$\Rightarrow \mathbf{T_n} \vdash \perp,$$

and again we have a contradiction. QED

5.2. Remarks. i. In the example of Theorem 4.1, we have

$$\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner) \Rightarrow \mathbf{T} \vdash \varphi \leftrightarrow Con_T \rightarrow Con_{T+Con_T}\,.$$

Choosing such a sequence for which $\mathbf{T} \vdash \neg Con_T$, we have $\mathbf{T} \vdash \varphi$, whence the condition of Σ_1-soundness in Lemma 5.1 cannot be replaced by simple consistency.

ii. Assuming a weak ultimate growth condition,

$$\forall n\ \mathbf{T} \vdash Con_{T_n}, \qquad (3)$$

we can replace Σ_1-soundness by consistency in Lemma 5.1. For, one can use this growth to get contradictions from (1) and (2) as follows:

$$\mathbf{T} \vdash \varphi \text{ or } \mathbf{T} \vdash \neg\varphi \Rightarrow \mathbf{T_n} \vdash Pr_{T_n}(\ulcorner\perp\urcorner)$$
$$\Rightarrow \mathbf{T_n} \vdash \neg Con_{T_n}$$
$$\Rightarrow \mathbf{T} \vdash \neg Con_{T_n},$$

making $\mathbf{T}$ inconsistent.

McAloon's purpose in introducing his set theoretic Rosser sentences was to construct end extensions of models of set theory. The arithmetic analogue of his initial result is the following.

5.3. Theorem. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ be an r.e. sequence of consistent extensions of **PRA** in the langauge of arithmetic, let $\mathbf{T} = \cup_{n \in \omega} \mathbf{T_n}$, and assume condition (3) above. Assume further that $\mathbf{T}$ contains Π_2-induction. Let $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$. Then: Any model $\mathcal{M} \models \mathbf{T} + \varphi$ has an end extension $\mathcal{N} \models \mathbf{T} + \neg\varphi$.

Proof: Observe, for each n , that

$$\mathbf{T} + \varphi \vdash Pr_{T_n}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_n}(\ulcorner\neg\varphi\urcorner)$$

$$\vdash Pr_{T_n}(\ulcorner\varphi\urcorner) \rightarrow \neg Con_{T_n}$$

$$\vdash Con_{T_n} \rightarrow \neg Pr_{T_n}(\ulcorner\varphi\urcorner)$$

$$\vdash \neg Pr_{T_n}(\ulcorner\varphi\urcorner), \text{ by (3)}$$

$$\vdash Con_{T_n} + \neg\varphi. \qquad (4)$$

Applying the Arithmetised Completeness Theorem yields the desired conclusion. QED

5.4. Remark. As shown by McAloon in another paper (*McAloon 1978*), the assumption that **T** include Π_2 -induction is necessary to conclude the existence of an end extension via the construction in the proof of the Arithmetised Completeness Theorem. In the absence of Π_2 -induction, one still has (4) which is enough to conclude that $\neg\varphi$ is Π_1 -conservative over **T**:

$$\mathbf{T} + \neg\varphi \vdash \pi \Rightarrow \mathbf{T} \vdash \pi, \text{ for any } \Pi_1\text{-sentence } \pi.$$

For,

$$\mathbf{T} + \neg\varphi \vdash \pi \Rightarrow \mathbf{PRA} + Con_{T_n} + \neg\varphi \vdash \pi$$

$$\Rightarrow \mathbf{T} + \varphi \vdash \pi, \text{ by (4)}$$

$$\Rightarrow \mathbf{T} + \varphi \vee \neg\varphi \vdash \pi$$

$$\Rightarrow \mathbf{T} \vdash \pi.$$

Before we can continue presenting arithmetic analogues of McAloon's other results, we must take a closer look at McAloon's original Rosser sentences,

$$\mathbf{ZF} \vdash \varphi \leftrightarrow \forall\alpha(Prov^{\infty}(\alpha, \ulcorner\varphi\urcorner) \rightarrow Prov^{\infty}(\alpha, \ulcorner\neg\varphi\urcorner)), \qquad (5)$$

where, as said in the introduction,

$$Prov^{\infty}(x,y): \text{ "}x\text{ is an admissible ordinal and } \mathbf{ZF}_x \text{ proves } y\text{ ".} \qquad (6)$$

As McAloon noted, formula (6) can be modified by imposing an extra condition on the admissible ordinal. For any weak set theory **T**, he considered

$$Prov_t^{\infty}(x,y): \text{ "}x\text{ is an admissible ordinal and } L_x \models \mathbf{T} \text{ and } \mathbf{ZF}_x \text{ proves } y\text{ ".}$$

Each **T** has its own Rosser sentence φ_t analogous to φ in (5):

$$\mathbf{ZF} \vdash \varphi_t \leftrightarrow \forall\alpha(Prov_t^{\infty}(\alpha, \ulcorner\varphi_t\urcorner) \rightarrow Prov_t^{\infty}(\alpha, \ulcorner\neg\varphi_t\urcorner)).$$

McAloon then considered the question of the relation between φ_t and φ_u for different weak set theories **T** and **U**. He showed that, if **U** is somewhat stronger than **T** in that,

$$\mathbf{U} >> \mathbf{T}:\ \mathbf{U} \vdash \forall\alpha\, \exists\beta > \alpha\, (L_\alpha \models \mathbf{T}), \tag{7}$$

then

$$\mathbf{ZF} \vdash \varphi_t \vee \varphi_u . \tag{8}$$

The arithmetic analogue to varying the weak theories **T** and **U** is the variation of the hierarchies $\{\mathbf{T_n}\}_{n\in\omega}$. Thus, we consider two hierarchies $\{\mathbf{T_n}\}_{n\in\omega}$ and $\{\mathbf{U_n}\}_{n\in\omega}$ for the same theory **T**:

$$\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots \subseteq \bigcup_{n\in\omega} \mathbf{T_n} = \mathbf{T}$$

$$\mathbf{U_0} \subseteq \mathbf{U_1} \subseteq \ldots \subseteq \bigcup_{n\in\omega} \mathbf{U_n} = \mathbf{T}.$$

We will say that a hierarchy $\{\mathbf{U_n}\}_{n\in\omega}$ is *somewhat stronger than* $\{\mathbf{T_n}\}_{n\in\omega}$, if

$$\mathbf{PRA} \vdash \forall xy\, [Pr_{T_x}(y) \rightarrow Pr_{U_x}(y)\,] \tag{9}$$

and

$$\mathbf{PRA} \vdash \forall x\, Pr_{U_x}(\ulcorner Con_{T_{\dot{x}}} \urcorner) . \tag{10}$$

We also say that $\{\mathbf{U_n}\}_{n\in\omega}$ is *not too much stronger than* $\{\mathbf{T_n}\}_{n\in\omega}$, if

$$\mathbf{PRA} \vdash \forall xy\, [Pr_{U_x}(y) \rightarrow Pr_{T_{x+1}}(y)\,] \tag{11}$$

and

$$\mathbf{PRA} \vdash \forall x\, Pr_{T_{x+1}}(\ulcorner Con_{U_{\dot{x}}} \urcorner) . \tag{12}$$

We also write $MPr_t(x)$ and $MPr_u(x)$ for the McAloon proof predicates based on $\{\mathbf{T_n}\}_{n\in\omega}$ and $\{\mathbf{U_n}\}_{n\in\omega}$, respectively.

It is not hard to guess that conditions (9) and (10) are intended as the arithmetic analogues to (7). It turns out that one needs (11) and (12) as well: If, for example, $\{\mathbf{T_n}\}_{n\in\omega}$ satisfies the strong growth condition of section 2, above, and the sequence $\{\mathbf{U_n}\}_{n\in\omega}$ is defined by

$$\mathbf{U_n} = \mathbf{T_{n+1}},$$

then φ_t and φ_u are virtually identical and $\mathbf{T} \vdash \varphi_t \leftrightarrow \varphi_u$.

As for the normality conditions, first note that (9) and (11) yield the usual monotonicity conditions,

$$\mathbf{PRA} \vdash \forall x\, y\, [x < y \rightarrow (Pr_{T_x}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_y}(\ulcorner\chi\urcorner))\,]$$

$$\mathbf{PRA} \vdash \forall x\, y\, [x < y \rightarrow (Pr_{U_x}(\ulcorner\chi\urcorner) \rightarrow Pr_{U_y}(\ulcorner\chi\urcorner))\,],$$

for all sentences χ. The other necessary condition is

$$\mathbf{PRA} \vdash \forall x\, [Pr_{PRA}(\ulcorner\chi\urcorner) \to Pr_{T_x}(\ulcorner\chi\urcorner)\ ], \text{ for all sentences } \chi, \tag{13}$$

which, with (9), yields the corresponding

$$\mathbf{PRA} \vdash \forall x\, [Pr_{PRA}(\ulcorner\chi\urcorner) \to Pr_{U_x}(\ulcorner\chi\urcorner)\ ], \text{ for all sentences } \chi.$$

We won't need to assume the provability within **PRA** that **T** is the union of each of the sequences.

5.5. Theorem. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ and $\mathbf{U_0} \subseteq \mathbf{U_1} \subseteq \ldots$be r.e. sequences of consistent extensions of **PRA** satisfying (9) - (13), and let $\mathbf{T} = \bigcup_{n \in \omega} \mathbf{T_n}$. If $\mathbf{PRA} \vdash \varphi \leftrightarrow \neg MPr_t(\ulcorner\varphi\urcorner)$ and $\mathbf{PRA} \vdash \psi \leftrightarrow \neg MPr_u(\ulcorner\psi\urcorner)$, then $\mathbf{PRA} \vdash \varphi \vee \psi$.

Proof: First, observe

$$\begin{aligned} \mathbf{PRA} \vdash\ & Pr_{T_x}(\ulcorner\varphi\urcorner) \to Pr_{T_x}(\ulcorner Pr_{T_{\dot x}}(\ulcorner\varphi\urcorner)\urcorner) \\ \vdash\ & Pr_{T_x}(\ulcorner\varphi\urcorner) \to Pr_{U_x}(\ulcorner Pr_{T_{\dot x}}(\ulcorner\varphi\urcorner)\urcorner), \end{aligned} \tag{14}$$

by (9). But (9) also yields

$$\mathbf{PRA} \vdash Pr_{T_x}(\ulcorner\varphi\urcorner) \to Pr_{U_x}(\ulcorner\varphi\urcorner),$$

which, with (14) and the definition of $\neg MPr_t(\ulcorner\varphi\urcorner)$, yields

$$\begin{aligned} \mathbf{PRA} \vdash\ & Pr_{T_x}(\ulcorner\varphi\urcorner) \to Pr_{U_x}(\ulcorner Pr_{T_{\dot x}}(\ulcorner\varphi\urcorner) \wedge Pr_{T_{\dot x}}(\ulcorner\neg\varphi\urcorner)\urcorner) \\ \vdash\ & Pr_{T_x}(\ulcorner\varphi\urcorner) \to Pr_{U_x}(\ulcorner\neg Con_{T_{\dot x}}\urcorner) \\ \vdash\ & Pr_{T_x}(\ulcorner\varphi\urcorner) \to Pr_{U_x}(\ulcorner\bot\urcorner), \end{aligned} \tag{15}$$

by (10).

Similarly,

$$\mathbf{PRA} \vdash Pr_{U_x}(\ulcorner\psi\urcorner) \to Pr_{T_{w+1}}(\ulcorner\bot\urcorner). \tag{16}$$

Let θ abbreviate

$$Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y \le x\, \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \wedge Pr_{U_w}(\ulcorner\psi\urcorner) \wedge \forall y \le w\, \neg Pr_{U_y}(\ulcorner\neg\psi\urcorner),$$

so that $\neg\varphi \wedge \neg\psi \leftrightarrow \exists x\, \exists y\, \theta$. Observe,

$$\mathbf{PRA} \vdash Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y \le w\, \neg Pr_{U_y}(\ulcorner\neg\psi\urcorner) \to Pr_{U_x}(\ulcorner\bot\urcorner) \wedge \forall y \le w\, \neg Pr_{U_y}(\ulcorner\neg\psi\urcorner)$$

by (15), whence

$$\mathbf{PRA} \vdash Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \forall y \le w\, \neg Pr_{U_y}(\ulcorner\neg\psi\urcorner) \to w < x. \tag{17}$$

Similarly,

$$\mathbf{PRA} \vdash Pr_{U_w}(\ulcorner\psi\urcorner) \wedge \forall y \leq x \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \rightarrow Pr_{T_{w+1}}(\ulcorner\bot\urcorner) \wedge \forall y \leq x \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner)$$
$$\vdash Pr_{U_w}(\ulcorner\psi\urcorner) \wedge \forall y \leq x \neg Pr_{T_y}(\ulcorner\neg\varphi\urcorner) \rightarrow x < w+1 .$$

With (17) this yields,

$$\mathbf{PRA} \vdash \theta \rightarrow w < x \wedge x < w+1$$
$$\vdash \neg\theta$$
$$\vdash \neg \exists x \exists y\, \theta$$
$$\vdash \neg(\neg\varphi \wedge \neg\psi)$$
$$\vdash \varphi \vee \psi.$$

QED

5.6. Remark. If we also assume the strong growth requirement of section 2, then we can conclude the more general

$$\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr_t(\ulcorner\varphi\urcorner) \ \&\ \mathbf{T} \vdash \psi \leftrightarrow \neg MPr_u(\ulcorner\psi\urcorner) \Rightarrow \mathbf{T} \vdash \varphi \vee \psi.$$

This can be seen either by analysing the proof or invoking Theorem 2.1:

$$\mathbf{T} \vdash \varphi \leftrightarrow \varphi_0 \quad \text{and} \quad \mathbf{T} \vdash \psi \leftrightarrow \psi_0 ,$$

where $\mathbf{PRA} \vdash \varphi_0 \leftrightarrow \neg MPr_t(\ulcorner\varphi_0\urcorner)$ and $\mathbf{PRA} \vdash \psi_0 \leftrightarrow \neg MPr_u(\ulcorner\psi_0\urcorner)$. Thus, from the fact that $\mathbf{PRA} \vdash \varphi_0 \vee \psi_0$, we can conclude $\mathbf{T} \vdash \varphi \vee \psi$.

5.7. Corollary. Let $\mathbf{T_0} \subseteq \mathbf{T_1} \subseteq \ldots$ and $\mathbf{U_0} \subseteq \mathbf{U_1} \subseteq \ldots$ be r.e. sequences of consistent theories in the language of arithmetic containing **PRA** and satisfying (9) - (13), and let $\mathbf{T} = \bigcup_{n \in \omega} \mathbf{T_n}$. Assume further that **T** contains Π_2 -induction. Let $\mathbf{T} \vdash \varphi \leftrightarrow \neg MPr_t(\ulcorner\varphi\urcorner)$. Then: Any model $\mathcal{M} \models \mathbf{T} + \neg\varphi$ has an end extension $\mathcal{N} \models \mathbf{T} + \varphi$.

Proof: Let $\mathbf{PRA} \vdash \psi \leftrightarrow \neg MPr_u(\ulcorner\psi\urcorner)$ and observe:

$$\mathcal{M} \models \mathbf{T} + \neg\varphi \Rightarrow \mathcal{M} \models \psi, \text{ since } \mathbf{PRA} \vdash \varphi \vee \psi$$
$$\Rightarrow \exists \mathcal{N}(\mathcal{M} \subseteq_e \mathcal{N} \models \mathbf{T} + \neg\psi), \text{ by 5.3}$$
$$\Rightarrow \exists \mathcal{N}(\mathcal{M} \subseteq_e \mathcal{N} \models \mathbf{T} + \varphi), \text{ since } \mathbf{PRA} \vdash \varphi \vee \psi.$$

QED

5.8. Remarks. i. The end extension obtained in the proof of 5.7 is proper. The end extension promised in 5.3 can also be made proper-- under the presently assumed conditions-- by the simple expedient of applying 5.3, 5.7, and 5.3 in succession.

ii. If the strong growth condition of Theorem 2.1 is assumed, then Theorem 5.7 holds for all **T**-provably McAloon-Rosser sentences $\varphi \leftrightarrow \neg MPr(\ulcorner\varphi\urcorner)$.

iii. Moreover, if the strong growth condition and Π_2 -induction are assumed, the Corollary can be proven directly without appeal to Theorem 5.6: If $\mathcal{M} \models \mathbf{T} + \neg\varphi$, then $\mathcal{M} \models \exists x\ [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \neg Pr_{T_x}(\ulcorner\neg\varphi\urcorner)]$. Let a in the domain of $\mathcal{M}$ witness this formula. Thus, $\mathcal{M}$ believes that $\mathbf{T_a}$ proves φ. By reflexion, if a were finite, we would have $\mathcal{M} \models Pr_{T_a}(\ulcorner\varphi\urcorner) \rightarrow \varphi$, whence $\mathcal{M} \models \varphi$, a contradiction. Thus, a is infinite and $\mathcal{M} \models \neg Pr_{T_n}(\ulcorner\neg\varphi\urcorner)$, for all finite n, i.e. $\mathcal{M} \models Con_{T_n + \varphi}$ for all finite n, and the Arithmetised Completeness Theorem yields the result.

5.9. Remarks. i. Again, if we drop the requirement that $\mathbf{T}$ include Π_2 -induction, we can still conclude that φ is Π_1 -conservative over $\mathbf{T}$.

ii. If $\mathbf{T}$ is also Σ_1 -sound, then φ is also Σ_1 -conservative over $\mathbf{T}$: Let $\sigma \in \Sigma_1$ and suppose $\mathbf{T} + \varphi \vdash \sigma$. If $\mathbf{T} \nvdash \sigma$, then $\mathbf{T} + \neg\sigma$ is consistent and Σ_1 -sound (since $\neg\sigma \in \Pi_1$).

But
$$\begin{aligned} \mathbf{T} + \neg\sigma &\vdash \neg\varphi \\ &\vdash \exists x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \wedge \neg Pr_{T_x}(\ulcorner\neg\varphi\urcorner)] \\ &\vdash Pr_T(\ulcorner\varphi\urcorner). \end{aligned}$$

The Σ_1 -soundness of $\mathbf{T} + \neg\sigma$ would then tell us that $\mathbf{T} \vdash \varphi$, contrary to Lemma 5.1. Hence $\mathbf{T} \vdash \sigma$.

iii. Alternate proof of ii: Observe
$$\begin{aligned} \mathbf{T} + Con_T &\vdash \neg Pr_T(\ulcorner\varphi\urcorner), \text{ by Remark 5.2} \\ &\vdash \forall x\, [Pr_{T_x}(\ulcorner\varphi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\neg\varphi\urcorner)] \\ &\vdash \varphi, \end{aligned}$$

and Con_T is Σ_1 -conservative over $\mathbf{T}$ provided $\mathbf{T}$ is Σ_1 -sound. Thus φ, being a consequence of a Σ_1 -conservative sentence, is itself Σ_1 -conservative.

iv. Again assuming the Σ_1 -soundness of $\mathbf{T}$, $\neg\varphi$ is not Σ_1 -conservative over $\mathbf{T}$: As we saw in ii,
$$\mathbf{T} + \neg\varphi \vdash Pr_T(\ulcorner\varphi\urcorner)$$
and Σ_1 -conservation would yield
$$\mathbf{T} \vdash Pr_T(\ulcorner\varphi\urcorner),$$
whence Σ_1 -soundness would yield $\mathbf{T} \vdash \varphi$, contrary to Lemma 5.1.

It is an easy matter to produce examples of sequences $\{\mathbf{T_n}\}_{n\in\omega}$ and $\{\mathbf{U_n}\}_{n\in\omega}$ which satisfy the conditions of Theorem 5.5. One starts wilth a sequence $\{\mathbf{T_n}\}_{n\in\omega}$ like

$$\mathbf{T_0} = \mathbf{PRA}$$
$$\mathbf{T_{n+1}} = \mathbf{T_n} + Con_{T_n},$$

or

$$\mathbf{T_n} = \mathbf{PRA} + \Sigma_{n+1}\text{-Induction}, \tag{18}$$

or, indeed, any sequence satisfying

$$\mathbf{PRA} \vdash \forall x\, [Pr_{PRA}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_x}(\ulcorner\chi\urcorner)], \text{ for all sentences } \chi$$
$$\mathbf{PRA} \vdash \forall x\, y\, [x < y \rightarrow (Pr_{T_x}(\ulcorner\chi\urcorner) \rightarrow Pr_{T_y}(\ulcorner\chi\urcorner))], \text{ for all sentences } \chi$$

and $$\mathbf{PRA} \vdash \forall x\, Pr_{T_{x+1}}(\ulcorner Con_{T_{\dot{x}}} \urcorner).$$

From such a sequence one can define two new sequences,

$$\mathbf{T_n}' = \mathbf{T_{2n}}, \quad \mathbf{U_n} = \mathbf{T_{2n+1}},$$

and observe that $\{\mathbf{U_n}\}_{n\in\omega}$ is somewhat stronger but not too much stronger than $\{\mathbf{T_n}\}_{n\in\omega}$, i.e. Theorem 5.5 applies to them.

Also, if $\{\mathbf{U_n}\}_{n\in\omega}$ is somewhat stronger but not too much stronger than $\{\mathbf{T_n}\}_{n\in\omega}$, one can define

$$\mathbf{T_n}' = \mathbf{T_{n+1}}, \quad \mathbf{U_n}' = \mathbf{U_n},$$

and observe that $\{\mathbf{T_n}'\}_{n\in\omega}$ is somewhat stronger but not too much stronger than $\{\mathbf{U_n}'\}_{n\in\omega}$, thus reversing the roles of the given sequences.

And, of course, for the sequence (18), there is enough room between succesive elements of the sequence to interpolate a second sequence,

$$\mathbf{U_n} = \mathbf{T_n} + Con_{T_n}.$$

A bit more interesting than the construction of such examples is the construction of a strong counterexample, one which brings us full circle by returning us to the uniqueness question.

5.10. Counterexample. Consider the sequences,

$$\mathbf{T_n} = \mathbf{PRA} + \Sigma_{n+1}\text{-Induction}, \quad \mathbf{U_n} = \mathbf{PRA} + \Sigma_{n+2}\text{-Boundedness}.$$

Then: If $\mathbf{PA} \vdash \varphi \leftrightarrow \neg MPr_t(\ulcorner\varphi\urcorner)$ and $\mathbf{PA} \vdash \psi \leftrightarrow \neg MPr_u(\ulcorner\psi\urcorner)$, then $\mathbf{PA} \vdash \varphi \leftrightarrow \psi$.

The point to this example is that, although $\mathbf{T_n}$ and $\mathbf{U_n}$ are unequal, they have the same Π_{n+3}-consequences (as shown independently by Friedman and Paris, cf. *Paris 1981*). Hence, if we define $\rho_{t,n}$ and $\rho_{u,n}$ as in the proof of Theorem 2.1.i, we have

$$\mathbf{T_n} \vdash \rho_{t,n}(\ulcorner\chi\urcorner) \leftrightarrow \rho_{u,n}(\ulcorner\chi\urcorner), \tag{19}$$

for χ of low complexity. Thus, for sufficiently large n ,

$$\mathbf{T_n} \vdash \varphi \leftrightarrow \neg\rho_{t,n}(\ulcorner\varphi\urcorner), \text{ as in the proof of 2.1.i}$$

$$\vdash \varphi \leftrightarrow \neg\rho_{u,n}(\ulcorner\varphi\urcorner), \tag{20}$$

by (19). But we also have

$$\mathbf{T_n} \vdash \psi \leftrightarrow \neg\rho_{u,n}(\ulcorner\psi\urcorner), \tag{21}$$

and $\rho_{u,n}$ is $\mathbf{T_n}$-substitutable. Thus, (20) and (21) yield $\mathbf{T_n} \vdash \varphi \leftrightarrow \psi$.

References

K. McAloon

1975 Formules de Rosser pour ZF, C.R. Acad. Sc. Paris 281, ser. A, 669-672.

1978 Completeness theorems, incompleteness theorems and models of arithmetic, Trans. AMS 239, 253-277.

C. Smoryński

1985 *Self-Reference and Modal Logic* , Springer-Verlag, New York.

J. Paris

1981 Some conservation results for fragments of arithmetic, in: K. McAloon et al, eds., *Model Theory and Arithmetic* , Springer-Verlag, Heidelberg.

A. Visser

A Peano's smart children, to appear.

Logic Group Preprint Series
Department of Philosophy
University of Utrecht
Heidelberglaan 2
3584 CS Utrecht
The Netherlands

nr. 1 C.P.J. Koymans, J.L.M. Vrancken, *Extending Process Algebra with the empty process*, September 1985.
nr. 2 J.A. Bergstra, *A process creation mechanism in Process Algebra*, September 1985.
nr. 3 J.A. Bergstra, *Put and get, primitives for synchronous unreliable message passing*, October 1985.
nr. 4 A.Visser, *Evaluation, provably deductive equivalence in Heyting's arithmetic of substitution instances of propositional formulas*, November 1985.
nr. 5 G.R. Renardel de Lavalette, *Interpolation in a fragment of intuitionistic propositional logic*, January 1986.
nr. 6 C.P.J. Koymans, J.C. Mulder, *A modular approach to protocol verification using Process Algebra* , April 1986.
nr. 7 D. van Dalen, F.J. de Vries, *Intuitionistic free abelian groups*, April 1986.
nr.8 F. Voorbraak, *A simplification of the completeness proofs for Guaspari and Solovay's R*, May 1986.
nr.9 H.B.M. Jonkers, C.P.J. Koymans & G.R. Renardel de Lavalette, *A semantic framework for the COLD-family of languages*, May 1986.
nr.10 G.R. Renardel de Lavalette, *Strictheidsanalyse*, May 1986.
nr.11 A. Visser, *Kunnen wij elke machine verslaan? Beschouwingen rondom Lucas' argument*, July 1986.
nr.12 E.C.W. Krabbe, *Naess's dichotomy of tenability and relevance*, June 1986.
nr.13 Hans van Ditmarsch, *Abstractie in wiskunde, expertsystemen en argumentatie*, Augustus 1986
nr.14 A.Visser, *Peano's Smart Children, a provability logical study of systems with built-in consistency* , October 1986.
nr.15 G.R.Renardel de Lavalette, *Interpolation in natural fragments of intuitionistic propositional logic*, October 1986.
nr.16 J.A. Bergstra, *Module Algebra for relational specifications*, November 1986.
nr.17 F.P.J.M. Voorbraak, *Tensed Intuitionistic Logic*, January 1987.
nr.18 J.A. Bergstra, J. Tiuryn, *Process Algebra semantics for queues*, January 1987.
nr.19 F.J. de Vries, *A functional program for the fast Fourier transform*, March 1987.
nr.20 A. Visser, *A course in bimodal provability logic*, May 1987.
nr.21 F.P.J.M. Voorbraak, *The logic of actual obligation, an alternative approach to deontic logic*, May 1987.
nr.22 E.C.W. Krabbe, *Creative reasoning in formal discussion*, June 1987.
nr.23 F.J. de Vries, *A functional program for Gaussian elimination*, September 1987.
nr.24 G.R. Renardel de Lavalette, *Interpolation in fragments of intuitionistic propositional logic*, October 1987.(revised version of no. 15)
nr.25 F.J. de Vries, *Applications of constructive logic to sheaf constructions in toposes*, October 1987.
nr.26 F.P.J.M. Voorbraak, *Redeneren met onzekerheid in expertsystemen*, November 1987.
nr.27 P.H. Rodenburg, D.J. Hoekzema, *Specification of the fast Fourier transform algorithm as a term rewriting system*, December 1987.

nr.28 D. van Dalen, *The war of the frogs and the mice, or the crisis of the Mathematische Annalen,* December 1987.

nr.29 A. Visser, *Preliminary Notes on Interpretability Logic,* January 1988.

nr.30 D.J. Hoekzema, P.H. Rodenburg, *Gauß elimination as a term rewriting system*, January 1988.

nr. 31 C. Smorynski, *Hilbert's Programme,* January 1988.

nr. 32 G.R. Renardel de Lavalette, *Modularisation, Parameterisation, Interpolation,* January 1988.

nr. 33 G.R. Renardel de Lavalette, *Strictness analysis for POLYREC, a language with polymorphic and recursive types*, March 1988.

nr. 34 A. Visser, *A Descending Hierarchy of Reflection Principles,* April 1988.

nr. 35 F.P.J.M. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory,* April 1988.

nr. 36 C. Smorynski, *Arithmetic Analogues of McAloon's Unique Rosser Sentences,* April 1988.