

## A System of Complete and Consistent Truth

VOLKER HALBACH

**Abstract** To the axioms of Peano arithmetic formulated in a language with an additional unary predicate symbol  $T$  we add the rules of necessitation  $\varphi/T\bar{\varphi}$  and conecessitation  $T\bar{\varphi}/\varphi$  and axioms stating that  $T$  commutes with the logical connectives and quantifiers. By a result of McGee this theory is  $\omega$ -inconsistent, but it can be approximated by models obtained by a kind of rule-of-revision semantics. Furthermore we prove that FS is equivalent to a system already studied by Friedman and Sheard and give an analysis of its proof theory.

*1 Preliminaries* Let  $\mathcal{L}$  be the first-order language of arithmetic with symbols for all primitive recursive functions; that is, if  $[e]$  is a primitive recursive function with index  $e$ , a function symbol  $f_e$  for  $[e]$  is available in  $\mathcal{L}$ . We suppose that  $\mathcal{L}$  has  $=$ ,  $\neg$ ,  $\rightarrow$  and  $\exists$  as logical symbols. If we expand  $\mathcal{L}$  by adding the new predicate constant  $T$  we obtain the language  $\mathcal{L}_T$ . Throughout the whole paper we shall identify every expression of  $\mathcal{L}_T$  with its Gödel number (under a standard gödelnumbering). Because we also identify languages with the set of their formulas, a language will be a set of natural numbers. All theories we shall discuss are extensions of Peano arithmetic: PA is the theory containing all defining equations of the primitive recursive functions and all the induction axioms in the full language  $\mathcal{L}_T$ . The index  $e$  of a primitive recursive function  $[e]$  will provide the defining equation(s) for the symbol  $f_e$  associated with the index  $e$ .

If a primitive recursive function  $h$  is explicitly given by some equations, we have a natural index  $e$  for this function which is again associated with a function symbol  $f_e$  in the language  $\mathcal{L}$ . Usually we shall denote this function symbol for  $h$  by  $h$ . So  $h$  naturally represents  $h$  in PA in the language  $\mathcal{L}$ . It is useful to conceive of the logical connectives as functions of expressions (i.e., of natural numbers). So we have for negation a function symbol  $\neg$  representing the operation of prefixing a negation symbol to an expression (and similarly for material implication and the existential quantifier). Hence we can show for every formula  $\varphi \in \mathcal{L}_T$  that:

$$\text{PA} \vdash \neg\bar{\varphi} = \overline{\neg\varphi}$$

*Received April 24, 1994; revised October 11, 1994*

In general,  $\bar{n}$  will denote the numeral of  $n$ , for instance in the above line  $\bar{\varphi}$  is the numeral of the sentence  $\varphi$ .

Similar properties hold for the two-place function symbols  $\rightarrow$  and  $\exists$ . As usual we use a dot above variables in order to indicate that the variable can be bound from “outside” by the substitution function.

In order to state the axioms of our theories of truth we need some predicates in the language  $\mathcal{L}$  binumerating (strongly representing) certain properties of expressions in a natural way. Let  $\text{Sent}_{\mathcal{L}}(x)$  be a formula expressing in  $\mathcal{L}$  the property of being a sentence of  $\mathcal{L}$ . Similarly  $\text{At}(x)$  shall mean that  $x$  is an atomic sentence of  $\mathcal{L}$ ,  $\text{Ver}(x)$  that  $x$  is a true atomic sentence of  $\mathcal{L}$ , and  $\text{Var}(x)$  that  $x$  is a variable.

**2 Introduction** Whereas most axiomatic theories of type-free truth were guided by non-classical semantic constructions using partial or many-valued logic, our aim in this paper is to give a theory of truth which is thoroughly classical. As a starting point we consider the well-known theory  $\text{Tr}(\mathcal{L})$  of truth for the language  $\mathcal{L}$ , which is a formalization of the Tarskian definition of truth. It is equivalent to the statement that there is a satisfaction class plus full induction in the language  $\mathcal{L}_{\text{T}}$ .

**Definition 2.1**  $\text{Tr}(\mathcal{L})$  is given by the following axioms:

- (i) axioms of PA formulated in the language of  $\mathcal{L}_{\text{T}}$  (including full induction)
- (ii)  $\forall x[\text{At}(x) \rightarrow (\text{T}x \leftrightarrow \text{Ver}(x))]$
- (iii)  $\forall x[\text{Sent}_{\mathcal{L}}(x) \rightarrow (\text{T}\neg x \leftrightarrow \neg \text{T}x)]$
- (iv)  $\forall x\forall y[\text{Sent}_{\mathcal{L}}(x) \wedge \text{Sent}_{\mathcal{L}}(y) \rightarrow (\text{T}(x \rightarrow y) \leftrightarrow (\text{T}x \rightarrow \text{T}y))]$
- (v)  $\forall x\forall v[\text{Sent}_{\mathcal{L}}(x(\bar{0}/v)) \wedge \text{Var}(v) \rightarrow (\text{T}\exists vx \leftrightarrow \exists y\text{T}x(\dot{y}/v))]$

In the last axiom  $x(\bar{0}/v)$  designates the result of substituting the numeral 0 for the free variable  $v$  in the formula  $x$ . The substitution function is understood to be defined in such a way that  $x(\bar{0}/v)$  is a formula of  $\mathcal{L}_{\text{T}}$  only if  $v$  is a variable.  $x(\dot{y}/v)$  is written to indicate that the numeral of  $y$  is substituted for  $v$  in  $x$ .

For atomic formulas of  $\mathcal{L}$   $\text{Tr}(\mathcal{L})$  states in axiom (ii) simply that T coincides with the truth definition for atomic  $\mathcal{L}$ -sentences which can be given within the language  $\mathcal{L}$  itself, while the other axioms (iii)–(v) say that T commutes with all logical connectives of sentences of the language  $\mathcal{L}$  thus simulating Tarski’s definition of truth.

It is well-known that it is possible to show Tarski’s (uniform) biconditional for each formula  $\varphi(x_1, \dots, x_n)$  of  $\mathcal{L}$  with just  $x_1, \dots, x_n$  free by induction on the complexity of  $\varphi(x_1, \dots, x_n)$ :

$$\text{Tr}(\mathcal{L}) \vdash \overline{\text{T}\varphi(\dot{x}_1, \dots, \dot{x}_n)} \leftrightarrow \varphi(x_1, \dots, x_n).$$

In order to give axioms for a truth theory for the whole language  $\mathcal{L}_{\text{T}}$ , we can try to state the principles involved in the axiomatization of  $\text{Tr}(\mathcal{L})$ , not only as above, for sentences of  $\mathcal{L}$ , but also for sentences of the language  $\mathcal{L}_{\text{T}}$  including sentences with the truth predicate T. So we keep axioms (i) and (ii) and enlarge the range of the quantifiers in axioms (iii)–(v). For this purpose let  $\text{Sent}_{\mathcal{L}_{\text{T}}}(x)$  express the property of being a sentence of  $\mathcal{L}_{\text{T}}$ .

**Definition 2.2**  $\text{FS}_1$  is the theory consisting of:

- (i) and (ii) as above

- (iii)  $\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}\neg x \leftrightarrow \neg \text{T}x)]$
- (iv)  $\forall x\forall y[\text{Sent}_{\mathcal{L}_T}(x) \wedge \text{Sent}_{\mathcal{L}_T}(y) \rightarrow (\text{T}(x \rightarrow y) \leftrightarrow (\text{T}x \rightarrow \text{T}y))]$
- (v)  $\forall x\forall v[\text{Sent}_{\mathcal{L}_T}(x(\bar{0}/v)) \wedge \text{Var}(v) \rightarrow (\text{T}\exists vx \leftrightarrow \exists y\text{T}x(y/v))]$

With an easy model theoretic argument (see also Corollary 4.3 below) one can prove the consistency of  $\text{FS}_1$ . The axioms of  $\text{FS}_1$ , especially (iii), are incompatible with axioms usually used to characterize truth in non-classical models. Note that the left-to-right direction of (iii) is easily seen to be equivalent to the “consistency” axiom

$$\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow \neg(\text{T}x \wedge \text{T}\neg x)], \quad (\text{T-Cons})$$

whereas the right-to-left direction is equivalent to the completeness principle

$$\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow \text{T}x \vee \text{T}\neg x]. \quad (\text{T-Comp})$$

So axiom (iii) says that the extension of  $\text{T}$  is a complete and consistent set of sentences of  $\mathcal{L}_T$  ruling out interpretations of  $\text{T}$  as a partial (non-complete) predicate. In contrast to the completeness principle  $\text{T-Comp}$ , axiom systems for partial truth usually include the axiom of consistency  $\text{T-Cons}$ . For similar reasons the right-to-left direction fails in these systems, but  $\text{T}$  distributes over implication in such partial interpretations.

Now  $\text{FS}_1$  has a major drawback:  $\text{FS}_1$  does not contain any axiom concerning iterations of truth. For example it is impossible to deduce the sentence  $\overline{\text{T}\bar{0}} = \bar{0}$  within  $\text{FS}_1$ . A first idea to overcome this deficiency might consist in the addition of an axiom resembling the other axioms of  $\text{FS}_1$ :

$$\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}\bar{\text{T}}x \leftrightarrow \text{T}x)]. \quad (1)$$

Unfortunately (1), together with the other axioms of  $\text{FS}_1$ , yields an inconsistency. By Gödel’s diagonal lemma we choose a closed term  $t$  satisfying  $\text{PA} \vdash t = \bar{\gamma}$ , where  $\gamma$  is the sentence  $\neg \text{T}t$ . We can employ axiom (iii) to derive the following contradiction in  $\text{FS}_1$ :

$$\begin{aligned} \text{FS}_1 + (1) \vdash \gamma &\leftrightarrow \neg \text{T}t \\ &\leftrightarrow \neg \overline{\text{T}\bar{\text{T}}t} \\ &\leftrightarrow \overline{\text{T}\neg \text{T}t} \\ &\leftrightarrow \text{T}\bar{\gamma}. \end{aligned}$$

So we get  $\text{FS}_1 + (1) \vdash \gamma \leftrightarrow \neg\gamma$  by the fixed point property of  $\gamma$ . Instead of adding the full axiom (1) we could try to weaken (1) by discarding one direction of the biconditional:

$$\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}x \rightarrow \text{T}\bar{\text{T}}x)] \quad (2)$$

$$\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}\bar{\text{T}}x \rightarrow \text{T}x)]. \quad (3)$$

Both  $\text{FS}_1+(2)$  and  $\text{FS}_1+(3)$  are consistent as shown by Friedman and Sheard in [5]. There they constructed models for variants of the two theories in the sections B “It is true that everything is true” and C “It is true that everything is false.” It will follow from a theorem below that the construction of these models may be formalized

within the system ACA of arithmetical comprehension (see Theorem 5.9). As  $FS_1$  is already as strong as ACA we shall be able to conclude the following proof-theoretical equivalences:

$$FS_1 \equiv FS_1 + (2) \equiv FS_1 + (3) \equiv \text{Tr}(\mathcal{L}) \equiv \text{ACA}.$$

But neither (2) nor (3) is an attractive axiom for truth, at least when taken together with the other axioms of  $FS_1$  because one can show:

**Lemma 2.3** *There is a sentence  $\gamma$  such that*

$$FS_1 + (2) \vdash \overline{TT\bar{\gamma}} \wedge \overline{TT\neg\bar{\gamma}}.$$

*Proof:* We need the following result that will be shown later: for any sentence  $\varphi$  of  $\mathcal{L}_T$  we have:

$$\text{PA} \vdash \varphi \implies FS_1 \vdash T\bar{\varphi}.$$

Let  $\gamma$  be the liar sentence as above.

$$\begin{aligned} \text{PA} &\vdash \gamma \leftrightarrow \neg T\bar{\gamma} \\ FS_1 &\vdash \overline{T(\gamma \leftrightarrow \neg T\bar{\gamma})} \\ FS_1 &\vdash T\bar{\gamma} \leftrightarrow \neg \overline{TT\bar{\gamma}} \\ FS_1 + (2) &\vdash T\bar{\gamma} \rightarrow \overline{TT\bar{\gamma}} \\ FS_1 + (2) &\vdash \neg T\bar{\gamma} \\ FS_1 + (2) &\vdash \overline{TT\bar{\gamma}} \\ FS_1 + (2) &\vdash T\neg\bar{\gamma} \\ FS_1 + (2) &\vdash \overline{TT\neg\bar{\gamma}} \end{aligned}$$

Because of this result (2) may hardly be considered as a good principle of iteration of truth, while (3) is no principle of iteration at all. For example, it can be shown by an easy model-theoretic argument that

$$FS_1 + (3) \not\vdash \overline{TT0 = 0}.$$

If  $S$  is a theory in the language  $\mathcal{L}_T$ , call the set of all  $S$ -derivable sentences the external logic of  $S$  and the set of all sentences  $\varphi$  of  $\mathcal{L}_T$  such that

$$S \vdash T\bar{\varphi}$$

the internal logic of  $S$ . Using this terminology we can restate Lemma 2.3: The internal logic of the internal logic of  $FS_1 + (2)$  is inconsistent. For a system  $S$  of truth it is a desirable feature that the internal logic of  $S$  equals the external logic of  $S$ . If we have

$$\text{internal logic of } S = \text{external logic of } S$$

or, in other words, for all sentences  $\varphi \in \mathcal{L}_T$

$$S \vdash T\bar{\varphi} \iff S \vdash \varphi,$$

then  $\overline{TT0 = 0}$  is derivable in  $S$ , if  $\overline{T0 = 0}$  is. Now we are able to expand  $FS_1$  by the principle stating the equivalence of internal and external logic:

**Definition 2.4** FS is the system  $FS_1$  with the following two additional rules:

$$(NEC) \quad \frac{\varphi}{T\varphi} \qquad (CONEC) \quad \frac{T\varphi}{\varphi}$$

NEC reminds of the necessitation rule of modal logic, while CONEC is conecessitation (a term coined by van Fraassen).

If the internal logic of a system  $S$ , formulated in the language  $\mathcal{L}_T$ , is classical, that is, if  $S$  proves all classical tautologies, then the internal logic of  $S$  includes its external logic, that is, NEC is a sound rule, and the internal logic is also classical. For example, we get

$$S \vdash T\overline{\varphi \rightarrow \varphi}$$

for every sentence  $\varphi \in \mathcal{L}_T$ . So NEC and CONEC are suitable for a system designed to be a thoroughly classical theory of truth, though it may fail for systems characterizing a conception of partial truth.

**3 An Alternative Axiomatization** FS is contained in the list of theories studied by Friedman and Sheard in [5] (hence the designation FS), but they use different axioms. We shall call the axiom system considered in [5]  $\widetilde{FS}$ . Because the system  $\widetilde{FS}$  is obtained by combining some attractive principles for truth it is interesting in itself and gives further motivation to investigate FS. Let PRE be the theory formulated in  $\mathcal{L}_T$  consisting of all the equations defining the primitive recursive functions, i.e. PA without induction. We assume that PRE contains an axiom  $\forall x S(x) \neq \bar{0}$ , where  $S$  is the successor symbol, such that all atomic sentences of  $\mathcal{L}$  are decided by PRE.

**Definition 3.1** The theory  $\widetilde{FS}$  is given by the following axioms and rules:

**Axioms:**

Base <sub>T</sub>	All axioms of PA including full induction in the language $\mathcal{L}_T$ $\forall x \forall y [\text{Sent}_{\mathcal{L}_T}(x) \wedge \text{Sent}_{\mathcal{L}_T}(y) \rightarrow (T(x \rightarrow y) \rightarrow (Tx \rightarrow Ty))]$
PRE-Refl	$\forall x [\text{Sent}_{\mathcal{L}_T}(x) \wedge \text{Bew}_{\text{PRE}}(x) \rightarrow Tx]$
T-Cons	$\forall x [\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\neg(Tx \wedge T\neg x))]$
T-Comp	$\forall x [\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (Tx \vee T\neg x)]$
U-Inf	$\forall a \forall v [\text{Sent}_{\mathcal{L}_T}(x(\bar{0}/v)) \wedge \text{Var}(v) \rightarrow (\forall y Tx(\dot{y}/v) \rightarrow T\forall vx)]$
E-Inf	$\forall x [\text{Sent}_{\mathcal{L}_T}(x(\bar{0}/v)) \wedge \text{Var}(v) \rightarrow (T\exists vx \rightarrow \exists y Tx(\dot{y}/v))]$

**Rules:**

T-Intro	$\varphi/T\varphi$ (NEC)
T-Elim	$T\varphi/\varphi$ (CONEC)
$\neg$ -T-Intro	$\neg\varphi/\neg T\varphi$
$\neg$ -T-Elim	$\neg T\varphi/\neg\varphi$

In [4] Feferman pointed out that PRE-Refl reminds of van Fraassen's supervaluation because by PRE-Refl all  $\mathcal{L}_T$ -sentences provable in classical logic are contained in the extension of the truth predicate.

**Theorem 3.2** *FS and  $\widetilde{\text{FS}}$  are identical theories.*

*Proof:* First we prove that all axioms of FS are derivable in  $\widetilde{\text{FS}}$ . In order to get axiom (ii), note that PRE decides all atomic and negated atomic sentences of  $\mathcal{L}$ . This can be shown in PA, because PA proves the consistency of PRE:

$$\text{PA} \vdash \forall x[\text{At}(x) \rightarrow (\text{Ver}(x) \leftrightarrow \text{Bew}_{\text{PRE}}(x))].$$

Hence using PRE-Refl we have:

$$\begin{aligned} \widetilde{\text{FS}} \vdash \forall x[\text{Sent}_{\mathcal{L}_{\text{T}}}(x) \wedge \text{At}(x) &\rightarrow (\neg\text{Ver}(x) \rightarrow \text{Bew}_{\text{PRE}}(\neg x))] \\ &\rightarrow (\neg\text{Ver}(x) \rightarrow \text{T}\neg x) \\ &\rightarrow (\neg\text{Ver}(x) \rightarrow \neg\text{T}x). \end{aligned}$$

On the other hand we get, again using PRE-Refl:

$$\begin{aligned} \widetilde{\text{FS}} \vdash \forall x[\text{Sent}_{\mathcal{L}_{\text{T}}}(x) \wedge \text{At}(x) &\rightarrow (\text{Ver}(x) \rightarrow \text{Bew}_{\text{PRE}}(x))] \\ \widetilde{\text{FS}} \vdash \forall x[\text{Sent}_{\mathcal{L}_{\text{T}}}(x) \wedge \text{At}(x) &\rightarrow (\text{Ver}(x) \rightarrow \text{T}x)]. \end{aligned}$$

As already mentioned, axiom (iii) of FS is equivalent to T-Cons and T-Comp. As one direction of axiom (iv) is already an axiom of  $\widetilde{\text{FS}}$ , it remains to show the other direction:

$$\begin{aligned} \widetilde{\text{FS}} \vdash \forall x\forall y[\text{Sent}_{\mathcal{L}_{\text{T}}}(x) \wedge \text{Sent}_{\mathcal{L}_{\text{T}}}(y) &\rightarrow \text{Bew}_{\text{PRE}}(\neg x \rightarrow (x \rightarrow y))] \\ &\rightarrow \text{T}(\neg x \rightarrow (x \rightarrow y)) \\ &\rightarrow (\text{T}\neg x \rightarrow \text{T}(x \rightarrow y)) \\ &\rightarrow \text{Bew}_{\text{PRE}}(y \rightarrow (x \rightarrow y)) \\ &\rightarrow (\text{T}y \rightarrow \text{T}(x \rightarrow y)) \\ &\rightarrow (\text{T}\neg x \vee \text{T}y \rightarrow \text{T}(x \rightarrow y)) \\ &\rightarrow ((\text{T}x \rightarrow \text{T}y) \rightarrow \text{T}(x \rightarrow y)). \end{aligned}$$

In a similar way we can prove axiom (v):

$$\begin{aligned} \widetilde{\text{FS}} \vdash \forall x\forall v[\text{Sent}_{\mathcal{L}_{\text{T}}}(x(\bar{0}/v)) \wedge \text{Var}(v) &\rightarrow \forall y\text{Bew}_{\text{PRE}}(x(\dot{y}) \rightarrow \exists vx)] \\ &\rightarrow \forall y\text{T}(x(\dot{y}) \rightarrow \exists vx) \\ &\rightarrow (\exists y\text{T}x(\dot{y}) \rightarrow \text{T}\exists vx). \end{aligned}$$

Having shown  $\text{FS} \subseteq \widetilde{\text{FS}}$  we still have to derive the axioms of  $\widetilde{\text{FS}}$  in FS. By a formalized induction on the length of the proofs it is possible to prove PRE-Refl within FS. All other axioms of  $\widetilde{\text{FS}}$  are easily seen to be contained in FS, and the rules  $\neg\text{T}$ -Intro and  $\neg\text{T}$ -Elim may be obtained from (NEC), (CONEC), T-Cons and T-Comp.

**4 Semantics** A consistent set of sentences containing all axioms of  $\widetilde{\text{FS}}$  and closed under (NEC) and (CONEC) was constructed by Friedman and Sheard in [5]. Instead of repeating their proof we give a slightly different proof of the consistency of FS, which can be converted into an estimate of the upper proof-theoretical bound of FS.

We obtain subsystems of FS by restricting the number of applications of the rules NEC and CONEC. Put  $\text{FS}_0 = \text{PA}$ , where PA is formulated in the full language  $\mathcal{L}_{\text{T}}$ .  $\text{FS}_1$  was already defined in the introduction and for  $n > 1$  we define:

- A formula is derivable in  $FS_n^*$ , if and only if it is derivable in FS by a proof with at most  $n - 1$  applications of NEC and  $n - 2$  applications of CONEC.
- A formula is derivable in  $FS_n$ , if and only if it is derivable in FS by a proof with at most  $n - 1$  applications of NEC and  $n - 1$  of CONEC.

So a formula can be derived in  $FS_2$  if it can be obtained from the  $FS_1$  axioms and an application of NEC and CONEC. Obviously  $FS \vdash \varphi$  if and only if there is  $n$  such that  $FS_n \vdash \varphi$ .

We shall construct  $\omega$ -models for the systems  $FS_n$ . Let  $M$  be the class of all expansions of the standard-model of  $\mathcal{L}$  to the language  $\mathcal{L}_T$ . Any such expansion  $\mathfrak{M}$  of the standard-model of arithmetic is determined by the extension  $\mathfrak{M}(T)$  it gives to the T-predicate. So the following function  $\Phi$  from  $M$  into  $M$  is well-defined:

$$n \in (\Phi(\mathfrak{M}))(T) \iff n \text{ is a sentence } \varphi \in \mathcal{L}_T \text{ such that } \mathfrak{M} \models \varphi.$$

So  $T\bar{n}$  is true in  $\Phi(\mathfrak{M})$ , if and only if  $n$  is a sentence valid in  $\mathfrak{M}$ . The function  $\Phi$  is exactly the rule of revision of truth as studied by Gupta, Herzberger and others. If  $N \subseteq M$ , we take  $\Phi(N)$  to be the image  $\{\Phi(\mathfrak{M}) : \mathfrak{M} \in N\} \subseteq M$  of  $N$  under  $\Phi$ . For the result of applying  $\Phi$   $n$ -times to a class  $N \subseteq M$  of models we write  $\Phi^n(N)$ .

#### Lemma 4.1

- (i)  $\Phi : M \rightarrow M$  is one-one.
- (ii)  $\Phi^n(\mathfrak{M}) \neq \mathfrak{M}$  for all  $n \neq 0$  and  $\mathfrak{M} \in M$ .
- (iii) If  $m \leq n$  then  $\Phi^n(M) \subseteq \Phi^m(M)$ .
- (iv) There is no infinite sequence of models  $\mathfrak{A}_0, \mathfrak{A}_1, \mathfrak{A}_2 \dots$  such that  $\Phi(\mathfrak{A}_{n+1}) = \mathfrak{A}_n$  for all  $n$ .
- (v)  $\bigcap_{n \in \omega} \Phi^n(M)$  is empty.

By (ii) it even follows that  $\Phi^n(M) \neq \Phi^m(M)$  if  $n \neq m$ ; that is,  $\Phi$  applied to  $M$  has no loops, and, in fact, cannot have any loops at all. Note that in general  $\Phi(N) \subseteq N$  fails for arbitrary  $N \subseteq M$ .

*Proof:* (i) If  $\mathfrak{A}$  and  $\mathfrak{B}$  are different expansions of the standard model of  $\mathcal{L}$  to  $\mathcal{L}_T$ , there is a sentence  $\varphi$  such that  $\mathfrak{A} \models \varphi$  and  $\mathfrak{B} \models \neg\varphi$ . Hence  $\Phi(\mathfrak{A}) \models T\bar{\varphi}$  and  $\Phi(\mathfrak{B}) \models T\neg\bar{\varphi}$ , and  $\Phi(\mathfrak{A})$  and  $\Phi(\mathfrak{B})$  are different, too.

(ii) Here we use liar sentences  $\gamma_n$  satisfying  $PA \vdash \gamma_n \leftrightarrow \neg \underbrace{T \dots T}_n \bar{\gamma}$ . Because

$$\Phi^n(\mathfrak{M}) \models T \dots T \bar{\gamma}_n \iff \mathfrak{M} \models \gamma_n$$

it follows that  $\Phi^n(\mathfrak{M})$  and  $\mathfrak{M}$  have to be different models.

(iii) Because  $\Phi^0(M) = M$  is the set of all expansions of the standard-model,  $\Phi^1(M) \subseteq \Phi^0(M)$  is trivial. Now it is sufficient to prove  $\Phi^{n+2}(M) \subseteq \Phi^{n+1}(M)$ . Supposing  $\mathfrak{A} \in \Phi^{n+2}(M)$  we know that there is a model  $\mathfrak{B} \in \Phi^{n+1}(M)$  such that  $\Phi(\mathfrak{B}) = \mathfrak{A}$ . By induction hypothesis  $\mathfrak{B}$  is in  $\Phi^n(M)$ , too. Consequently,  $\mathfrak{A}$  is in  $\Phi^{n+1}(M)$ .

(iv) Assume that there is such an infinite chain of models. Define a primitive recursive function  $f$  satisfying for all  $n \in \mathbf{N}$  und  $\varphi \in \mathcal{L}_T$ :

$$f(n, \varphi) := \underbrace{T \dots T}_n \bar{\varphi}.$$

By Gödel's diagonal lemma there is a sentence  $\gamma \in \mathcal{L}_T$  such that

$$\mathfrak{A}_0 \models \gamma \leftrightarrow \exists x \neg T f(x, \bar{\gamma}).$$

We shall show that both  $\mathfrak{A}_0 \models \neg\gamma$  and  $\mathfrak{A}_0 \models \gamma$  lead to a contradiction so that there cannot be such a chain. In the first case we have:

$$\begin{aligned} \mathfrak{A}_0 \models \neg\gamma &\implies \forall k > 0 \mathfrak{A}_k \models \gamma \\ &\implies \forall k > 0 \mathfrak{A}_k \models \exists x \neg T f(x, \bar{\gamma}) \\ &\implies \forall k > 0 \exists i > k \mathfrak{A}_i \models \neg\gamma. \end{aligned}$$

The first and the last of the above lines are contradictory. In the other case we reason in a similar way:

$$\begin{aligned} \mathfrak{A}_0 \models \gamma &\implies \mathfrak{A}_0 \models \exists x \neg T f(x, \bar{\gamma}) \\ &\implies \exists k > 0 \mathfrak{A}_k \models \neg\gamma \\ &\implies \exists k > 0 \forall i > k \mathfrak{A}_i \models \gamma. \end{aligned}$$

So we arrive again at a contradiction.

(v) follows easily from (iv).

In the proof of part (iv) of the lemma we used the same fixed point as McGee did in [9]. As far as we know, Cantini was the first to observe that McGee's theorem in [9] can be applied directly to FS to establish the  $\omega$ -inconsistency of FS. We shall briefly outline the reasoning of McGee's theorem. Choose  $\gamma$  as in the proof of (iv) of the preceding lemma. Then  $\text{FS}_1 \vdash \neg\gamma \rightarrow T\bar{\gamma}$  is easily established, and by an application of NEC we also obtain  $\text{FS}_2 \vdash T\bar{\gamma} \rightarrow \gamma$ , so we have  $\text{FS}_2 \vdash \gamma$ . By iterated application of NEC we derive  $\text{FS}_3 \vdash T f(0, \bar{\gamma})$ ,  $\text{FS}_4 \vdash T f(1, \bar{\gamma})$ ,  $\text{FS}_5 \vdash T f(2, \bar{\gamma})$ , and so on. Together with  $\text{FS}_2 \vdash \exists x \neg T f(x, \bar{\gamma})$  this renders FS  $\omega$ -inconsistent.

Part (iv) of the preceding lemma exhibits the consequences of McGee's theorem on rule-of-revision semantics. In particular, it shows that there cannot be an infinite descending chain of (standard) models where each model is obtained from the preceding one by an application of the revision rule.

We have the following ‘‘adequacy’’ result for the models in  $\Phi^n(\mathbb{M})$ .

**Theorem 4.2** *For all  $\mathfrak{M} \in \mathbb{M}$ :  $\mathfrak{M} \in \Phi^n(\mathbb{M})$  if and only if  $\mathfrak{M} \models \text{FS}_n$ .*

*Proof:* For  $n = 0$  the claim is trivial, for  $\text{FS}_0$  is PA formulated in the full language  $\mathcal{L}_T$  and  $\mathbb{M}$  is the set of all expansions of the standard model of  $\mathcal{L}$  to this language.

First we prove the left-to-right direction by induction on  $n$ . If  $\mathfrak{M} \in \Phi(\mathbb{M})$  it is easy to check that  $\mathfrak{M} \models \text{FS}_1$ . So assume  $\mathfrak{M} \in \Phi^{n+1}(\mathbb{M})$ , that is, there is  $\mathfrak{A}$  such that  $\mathfrak{M} = \Phi(\mathfrak{A})$  and  $\mathfrak{A} \in \Phi^n(\mathbb{M})$ . By (iii) of the above lemma  $\mathfrak{M} \in \Phi(\mathbb{M})$  and therefore  $\mathfrak{M} \models \text{FS}_1$ . Hence it remains to check that  $\mathfrak{M}$  also satisfies all sentences which can be deduced in FS by  $n$  applications of NEC and CONEC, respectively. If  $\text{FS}_n \vdash \varphi$  for closed  $\varphi$  then by induction hypothesis  $\mathfrak{A} \models \varphi$  and consequently,  $\mathfrak{M} \models T\bar{\varphi}$ . So if  $\text{FS}_{n+1}^* \vdash T\bar{\varphi}$  by an application of NEC we conclude  $\mathfrak{M} \models T\bar{\varphi}$ . Supposing  $\text{FS}_{n+1}^* \vdash T\bar{\psi}$  for closed  $\psi$  we now know for all  $\mathfrak{M} \in \Phi^{n+1}(\mathbb{M})$  that  $\mathfrak{M} \models T\bar{\psi}$ . From the definition of  $\Phi$  follows that  $\mathfrak{M} \models \psi$  for all  $\mathfrak{M} \in \Phi^n(\mathbb{M})$ . By part (iii) of the above

lemma this holds also true for all  $\mathfrak{M} \in \Phi^{n+1}(M)$  and the left-to-right direction of the theorem is proved.

Hence, if  $\mathfrak{M} \in \Phi^n(M)$ , we must have  $\mathfrak{M} \models \text{FS}_n$ . Consequently all systems  $\text{FS}_n$  are consistent and each of them has an  $\omega$ -model. We now show the right-to-left direction of the theorem, again by induction on  $n$ .

Assume  $\mathfrak{M} \models \text{FS}_1$  and  $A = \{\varphi : \mathfrak{M} \models T\bar{\varphi}\}$ ; so  $A$  is a set consisting of sentences of the language  $\mathcal{L}_T$ .  $A$  is a consistent set of sentences containing all true sentences of  $\mathcal{L}$  closed under logic and the  $\omega$ -rule; so  $A$  determines a unique model  $\mathfrak{A} \in M$ . Obviously,  $\mathfrak{M} = \Phi(\mathfrak{A})$ , therefore  $\mathfrak{M} \in \Phi(M)$ , and we are done with case  $n = 1$ .

If  $\mathfrak{M} \models \text{FS}_{n+1}$ , we put again  $A := \{\varphi : \mathfrak{M} \models T\bar{\varphi}\}$ . As in the above case,  $A$  determines a model  $\mathfrak{A} \in M$  and we have to show that  $\Phi^n(M)$  contains  $\mathfrak{A}$ . From  $\text{FS}_n \vdash \neg\varphi$  we can conclude  $\text{FS}_{n+1} \vdash T\bar{\neg\varphi}$  and by the axiom of consistency  $\text{FS}_{n+1} \vdash \neg T\bar{\varphi}$  and by assumption  $\mathfrak{M} \models \neg T\bar{\varphi}$  and hence  $\varphi \notin A$ . So  $A \cup \text{FS}_n$  is consistent and for this reason  $\mathfrak{A} \models \text{FS}_n$ . By induction hypothesis  $\mathfrak{A}$  is member of  $\Phi^n(M)$  and we have  $\mathfrak{M} = \Phi(\mathfrak{A}) \in \Phi^{n+1}(M)$ .

As a direct consequence we get the following corollary.

**Corollary 4.3** *FS is consistent.*

Without using McGee's direct proof we can show that part (v) of Lemma 4.1 and the theorem above suffice to establish the  $\omega$ -inconsistency of FS.

**Corollary 4.4** *FS is  $\omega$ -inconsistent.*

*Proof:* Because FS includes PA, the  $\omega$ -models of  $\text{FS}_0$  are exactly the models in  $M$ . Hence  $\bigcap_{n \in \omega} \Phi^n(M)$  is the set of all  $\omega$ -models of FS and, since this set is empty, FS has no  $\omega$ -models. By a well-known model theoretic argument involving the omitting types theorem we conclude that FS is  $\omega$ -inconsistent.

$M$  is the set of all standard models of PA with arbitrary extensions of the truth predicate. If  $\Phi$  is applied to  $M$ , we get models interpreting  $T$  as a truth predicate for the language  $\mathcal{L}$  without the truth predicate. So all models in  $\Phi(\mathfrak{M})$  are sound as models for noniterated truth. By further applications of  $\Phi$  we get models which are sound with respect to finite iterations of  $T$ , because  $\Phi^n(\mathfrak{M})$  is sound with respect to  $n$ -times iterated truth, and applications of the rules correspond to a gradual improvement of the models. But according to Lemma 4.1 the chain  $\Phi(\mathfrak{M}), \Phi^2(\mathfrak{M}), \Phi^3(\mathfrak{M}), \dots$  does not have a natural limit in the sense of a union of all models of the chain (in [6], [1], and [7] reasonable limit models were constructed by Gupta and Herzberger but these limit models do not satisfy the axioms of  $\text{FS}_1$  and are therefore completely different from the models of the chain from which these limit models are built up). So if we try to characterize in a formal system this semantical process of revision which is given by the iterated application of  $\Phi$ , we should not expect to obtain a pleasing model for the whole system. So the  $\omega$ -inconsistency of FS directly corresponds to the fact that the intersection of all  $\Phi^n(M)$  is empty.

**5 Proof Theory** By the observations above FS may be considered as a theory of finitely iterated truth. In this section we shall show that FS is also proof-theoretically equivalent to a system  $\text{RT}_{<\omega}$  of ramified truth up to  $\omega$ .  $\text{RT}_{<\omega}$  is a system for Tarski's

hierarchy of languages. By a result of Feferman stated in [4],  $\text{RT}_{<\omega}$  is again equivalent to the system  $\text{RA}_{<\omega}$  of ramified analysis up to  $\omega$ , that is, the system of  $\omega$ -times iterated arithmetical comprehension. For formulating the system  $\text{RT}_{<\omega}$  we need countably many new truth predicates  $\text{T}_n$  ( $n \in \mathbf{N}$ ). Let  $L(k)$  be the language  $\mathcal{L}$  expanded by all symbols  $\text{T}_n$  such that  $n < k$ . Hence  $L(0)$  has no truth predicates at all and is identical to  $\mathcal{L}$ . Furthermore, we need formulas  $\text{Sent}_{L(k)}$  binumerating the set of  $L(k)$ -sentences. Now the axioms of  $\text{RT}_i$  are all axioms of PA formulated in the full language  $\mathcal{L}_\Gamma$  plus for any  $n < i$ :

- (i)  $\forall x[\text{Sent}_{L(n)}(x) \wedge \text{At}(x) \rightarrow (\text{T}_n x \leftrightarrow \text{Ver}(x))]$
- (ii)  $\forall x[\text{Sent}_{L(n)}(x) \rightarrow (\text{T}_n \neg x \leftrightarrow \neg \text{T}_n x)]$
- (iii)  $\forall x \forall y[\text{Sent}_{L(n)}(x) \wedge \text{Sent}_{L(n)}(y) \rightarrow (\text{T}_n(x \rightarrow y) \leftrightarrow (\text{T}_n x \rightarrow \text{T}_n y))]$
- (iv)  $\forall x \forall v[\text{Sent}_{L(n)}(x(\bar{0}/v)) \wedge \text{Var}(v) \rightarrow (\text{T}_n \exists v x \leftrightarrow \exists y \text{T}_n x(\dot{y}/v))]$
- (v)  $\forall k < \bar{n} \forall x[\text{T}_n \text{T}_k \dot{x} \leftrightarrow \text{T}_n x \wedge \text{Sent}_{L(k)}(x)].$

In the last axiom the quantification of the index  $k$  is possible, because  $\text{T}_k$  is in the scope of another predicate. Although here we could easily drop the quantifier and replace (v) by the conjunction of  $n$  sentences, because  $\forall k < \bar{n}$  ranges only over finitely many numbers, the quantification becomes essential if we would give axioms also for transfinite levels  $\text{RT}_\alpha$ .

$\text{RT}_{<\omega}$  is simply the union of all  $\text{RT}_n$ :

$$\text{RT}_{<\omega} := \bigcup_{n \in \omega} \text{RT}_n.$$

Note that Tarski's equivalences are derivable in  $\text{FS}_1$  for sentences without truth-predicate in a uniform way. This can be verified by an easy (meta-)induction on the complexity of  $\varphi(\vec{x})$ .

**Lemma 5.1** *For all  $\varphi(\vec{x}) \in \mathcal{L}$  we have:*

$$\text{FS}_1 \vdash \forall \vec{x}[\overline{\text{T}\varphi(\vec{x})} \leftrightarrow \varphi(\vec{x})].$$

In order to reduce  $\text{RT}_{<\omega}$  to FS we define sublanguages  $\mathcal{L}_n$  of  $\mathcal{L}_\Gamma$  which will simulate the languages  $L(n)$ . Simultaneously we shall define predicates  $\text{Sent}_{\mathcal{L}_n}(x)$  in the language  $\mathcal{L}$  expressing that  $x$  is a closed formula of the language  $\mathcal{L}_n$ .

- (i)  $\mathcal{L}_0 := \mathcal{L}$
- (ii)  $\mathcal{L}_{n+1}$  is the language  $\mathcal{L}_n$  expanded by all formulas of the following form ( $t$  is an arbitrary term):

$$\text{T}t \wedge \text{Sent}_{\mathcal{L}_n}(t).$$

$\mathcal{L}_{n+1}$  is closed under the usual rules for the formation of formulas. It is important that  $\text{T}t$  appears only with the restriction  $\text{Sent}_{\mathcal{L}_n}(t)$  in a  $\mathcal{L}_{n+1}$ -formula.

**Theorem 5.2**

$$\text{FS}_{n+2} \vdash \forall x[\text{Sent}_{\mathcal{L}_n}(x) \rightarrow (\text{T}\dot{x} \leftrightarrow \text{T}x)].$$

*Proof:* By induction on  $n$ .

$n = 0$ . First we verify the claim for atomic  $x$ :

$$\begin{aligned}
 \text{FS}_1 &\vdash \forall x[\text{At}(x) \rightarrow (\text{Tx} \leftrightarrow \text{Ver}(x))] \\
 \text{FS}_2 &\vdash \forall x[\text{At}(x) \rightarrow (\text{T}\dot{\text{T}}\dot{x} \leftrightarrow \text{TVer}(\dot{x}))] \quad (\text{using NEC}) \\
 \text{FS}_1 &\vdash \forall x[\text{At}(x) \rightarrow (\text{Ver}(x) \leftrightarrow \text{TVer}(\dot{x}))] \quad (\text{by Lemma 5.1}) \\
 \text{FS}_1 &\vdash \forall x[\text{At}(x) \rightarrow (\text{Tx} \leftrightarrow \text{TVer}(\dot{x}))] \\
 \text{FS}_2 &\vdash \forall x[\text{At}(x) \rightarrow (\text{T}\dot{\text{T}}\dot{x} \leftrightarrow \text{Tx})].
 \end{aligned}$$

The proof of the lemma in the case  $n = 0$  is completed by an induction on the complexity of  $x$  formalized within  $\text{FS}_2$ . Auxiliary lemmas of the following type are useful in proving the induction step:

$$\text{FS}_2 \vdash \forall x \forall y [\text{Sent}_{\mathcal{L}}(x) \wedge \text{Sent}_{\mathcal{L}}(y) \rightarrow (\text{T}\dot{\text{T}}(\dot{x} \rightarrow \dot{y}) \leftrightarrow (\text{T}\dot{\text{T}}\dot{x} \rightarrow \text{T}\dot{\text{T}}\dot{y}))]. \quad (4)$$

Auxiliary Lemma (4) may be obtained in the following way:

$$\begin{aligned}
 \text{FS}_1 &\vdash \forall x \forall y [\text{Sent}_{\mathcal{L}}(x) \wedge \text{Sent}_{\mathcal{L}}(y) \rightarrow ((\text{Tx} \rightarrow \text{Ty}) \leftrightarrow \text{T}(x \rightarrow y))] \\
 \text{FS}_2 &\vdash \forall x \forall y [\text{Sent}_{\mathcal{L}}(x) \wedge \text{Sent}_{\mathcal{L}}(y) \rightarrow (\text{T}(\text{T}\dot{x} \rightarrow \text{T}\dot{y}) \leftrightarrow \text{T}\dot{\text{T}}(\dot{x} \rightarrow \dot{y}))] \quad (\text{by NEC}) \\
 \text{FS}_2 &\vdash \forall x \forall y [\text{Sent}_{\mathcal{L}}(x) \wedge \text{Sent}_{\mathcal{L}}(y) \rightarrow (\text{T}\dot{\text{T}}(\dot{x} \rightarrow \dot{y}) \leftrightarrow \text{T}\dot{\text{T}}\dot{x} \rightarrow \text{T}\dot{\text{T}}\dot{y})]
 \end{aligned}$$

For the proof of the induction step in the case of  $\rightarrow$  we can conclude employing (4) and the axioms of  $\text{FS}_1$ :

$$\begin{aligned}
 \text{FS}_2 &\vdash \forall x \forall y [\text{Sent}_{\mathcal{L}}(x) \wedge \text{Sent}_{\mathcal{L}}(y) \rightarrow \\
 &\quad [(\text{T}\dot{\text{T}}\dot{x} \leftrightarrow \text{Tx}) \wedge (\text{T}\dot{\text{T}}\dot{y} \leftrightarrow \text{Ty}) \rightarrow (\text{T}\dot{\text{T}}(\dot{x} \rightarrow \dot{y}) \leftrightarrow \text{T}(x \rightarrow y))].
 \end{aligned}$$

The cases of  $\neg$  and  $\exists$  can be treated in a similar way.

$n \rightarrow n + 1$ . As in the preceding case, the claim is shown by a formalized induction on the complexity of  $x$  using the following as induction hypothesis:

$$\text{FS}_{n+2} \vdash \forall x [\text{Sent}_{\mathcal{L}_n}(x) \rightarrow (\text{T}\dot{\text{T}}\dot{x} \leftrightarrow \text{Tx})].$$

By NEC and the  $\text{FS}_1$ -axioms we obtain:

$$\text{FS}_{n+3} \vdash \forall x [\text{Sent}_{\mathcal{L}_n}(x) \rightarrow (\text{T}\dot{\text{T}}\dot{\text{T}}\dot{x} \leftrightarrow \text{T}\dot{\text{T}}\dot{x})].$$

The induction step may be carried out in the same way as above, again using auxiliary lemmata resembling (4). Hence we have:

$$\text{FS}_{n+3} \vdash \forall x [\text{Sent}_{\mathcal{L}_{n+1}}(x) \rightarrow (\text{T}\dot{\text{T}}\dot{x} \leftrightarrow \text{Tx})].$$

From the theorem we can derive a generalization of Lemma 5.1 by an induction on the complexity of  $\varphi(\vec{x})$ :

**Corollary 5.3** *For all  $\varphi(\vec{x})$  in  $\mathcal{L}_n$ ,  $\text{FS}_n \vdash \overline{\text{T}\varphi(\vec{x})} \leftrightarrow \varphi(\vec{x})$  holds.*

Now we can inductively define a sequence  $\langle h_n : n \in \mathbf{N} \rangle$  of functions where  $h_n$  translates all formulas of  $L(n)$  into formulas of the language  $\mathcal{L}_n$ .

- (i) If  $k < n$  then  $h_k \subset h_n$ , so  $h_n$  is an extension of  $h_k$ .
- (ii) If  $i \notin L(n)$  let  $h_n(i) := \perp$ , where  $\perp$  abbreviates of  $\bar{0} = \bar{1}$ .
- (iii) If  $\varphi \in \mathcal{L}$  let  $h_n(\varphi) := \varphi$ .

- (iv) Each of the  $h_n$  commutes with the logical connectives and the quantifier.
- (v)  $h_{n+1}(T_n t) := \text{Sent}_{L(n)}(t) \wedge Th_n(t)$  for any term  $t$ .

Every function  $h_n$  maps the formulas of  $L(n)$  to formulas of  $\mathcal{L}_T$ . This can be proved within PA:

**Lemma 5.4**  $PA \vdash \forall x[\text{Sent}_{L(n)}(x) \rightarrow \text{Sent}_{L_n}(h_n(x))]$ .

Using the  $h_n$ 's we can reduce the systems  $RT_n$  of finitely ramified truth to FS.

**Lemma 5.5**  $RT_n \vdash \varphi \Rightarrow FS_n \vdash h_n(\varphi)$ .

*Proof:* It has to be shown that for every axiom  $\varphi$  of  $RT_n$  its translation  $h_n(\varphi)$  is derivable in  $FS_n$ . If  $n = 0$  the claim is trivial, because  $RT_0 = FS_0 = PA$  and  $h_0$  replaces only subformulas  $Tt$  by  $\perp$ . So suppose  $k < n$ . Then the translations of all axioms are easily established within  $FS_n$ , except the following:

$$\forall k < \bar{n} \forall x [T_n \dot{T}_k \dot{x} \leftrightarrow T_n x \wedge \text{Sent}_{L(k)}(x)].$$

Now we reason as follows:

$$\begin{aligned} FS_{n+1} \vdash & \forall x [\text{Sent}_{L(k)}(x) \rightarrow h_k(x) = h_n(x)] \\ & \forall x [\text{Sent}_{L(k)}(x) \rightarrow (T \dot{T} h_k(\dot{x}) \leftrightarrow Th_n(x))] \quad (\text{by Th. 5.2, Lem. 5.4}) \\ & \forall x [TTh_k(x) \wedge \text{Sent}_{L(k)}(x) \leftrightarrow Th_n(x) \wedge \text{Sent}_{L(k)}(x)] \\ & \forall x [\text{Sent}_{L(n)}(T_k x) \wedge T \dot{T} h_k(\dot{x}) \wedge \text{Sent}_{L(k)}(x) \leftrightarrow Th_n(x) \wedge \text{Sent}_{L(k)}(x)] \\ & \forall x [\text{Sent}_{L(n)}(T_k x) \wedge Th_{k+1}(T_k \dot{x}) \leftrightarrow Th_n(x) \wedge \text{Sent}_{L(k)}(x)] \\ & \forall x [\text{Sent}_{L(n)}(T_k x) \wedge Th_n(T_k \dot{x}) \leftrightarrow Th_n(x) \wedge \text{Sent}_{L(k)}(x)] \\ & \forall x [\text{Sent}_{L(n)}(T_k x) \wedge Th_n(T_k \dot{x}) \leftrightarrow \text{Sent}_{L(n)}(x) \wedge Th_n(x) \wedge \text{Sent}_{L(k)}(x)]. \end{aligned}$$

As  $k$  was arbitrary we conclude therefrom:

$$\begin{aligned} FS_{n+1} \vdash & \forall k < \bar{n} \forall x [h_{n+1}(T_n \dot{T}_k x) \leftrightarrow h_{n+1}(T_n x) \wedge \text{Sent}_{L(k)}(x)] \\ & h_{n+1}(\forall k < \bar{n} \forall x [T_n \dot{T}_k x \leftrightarrow T_n x \wedge \text{Sent}_{L(k)}(x)]). \end{aligned}$$

Hence we have reduced  $RT_{<\omega}$  to FS. For, if  $RT_{<\omega} \vdash \varphi$  and  $\varphi \in \mathcal{L}$ , then there exists an  $n$  satisfying  $RT_n \vdash \varphi$ ; by the above lemma it follows that  $FS_n \vdash h_n(\varphi)$ . But because  $h_n(\varphi) = \varphi$  for  $\varphi$  in  $\mathcal{L}$  we also have  $FS \vdash \varphi$ .

In the whole proof we did not use CONEC. By Theorem 5.9 below it can be concluded that every FS-derivable arithmetical sentence can be proved without CONEC. This result may be expanded to all sentences in any language  $\mathcal{L}_n$ . But it is still an open problem whether this holds true for all sentences of  $\mathcal{L}_T$ , that is, whether CONEC is superfluous in the axiomatization of FS.

We now take up the task of reducing FS to  $RT_{<\omega}$ . By the previous section,  $\Phi^n(\mathfrak{M})$  is a model for  $FS_n$  if  $\mathfrak{M}$  is a model in  $M$ . For simplicity, we take  $\mathfrak{M}$  to be the model with  $\mathfrak{M}(T) = \emptyset$  declaring everything false. Because of a problem concerning the rule CONEC we show within  $RT_{<\omega}$  that  $\Phi^{2^n}(\mathfrak{M})$  is a model for  $FS_n$  instead of employing  $\Phi^n(\mathfrak{M})$ . The following virtually shows that the construction of the model  $\Phi^{2^n}(\mathfrak{M})$  can be carried out within  $RT_{2^n}$  thus reducing  $FS_n$  to  $RT_{2^n}$ .

Again we shall define a sequence of functions  $g_n$  where  $g_n$ , applied to a formula of  $\mathcal{L}_T$ , gives a formula of the typed language  $L(n)$ .

- (i) Applied to a formula  $\varphi \in \mathcal{L}_T$ ,  $g_0$  replaces each subformula  $Tt$  of  $\varphi$  by  $\perp$ .
- (ii) For  $g_n$  there is a function symbol  $g_n$  in  $\mathcal{L}$  strongly representing  $g_n$  in PA. Now  $g_{n+1}$  replaces, applied to a formula of  $\mathcal{L}_T$ , each subformula  $Tt$  of  $\varphi$  by  $T_n g_n(t)$ .
- (iii)  $g_n(k) := \perp$ , if  $k$  is not a formula of  $\mathcal{L}_T$ .

We need the following two obvious properties of  $g_n$  and their respective formalizations:

**Lemma 5.6**

- (i)  $\forall k \in \mathbf{N} \quad g_n(k) \in L(n)$
- (ii)  $\text{PA} \vdash \forall x \text{Form}_{L(n)}(g_n(x))$
- (iii)  $g_n(\varphi) = \varphi$  for  $\varphi \in \mathcal{L}$
- (iv)  $\text{PA} \vdash \forall x[\text{Form}_{\mathcal{L}}(x) \rightarrow g_n(x) = x]$

In the lemma  $\text{Form}_{L(n)}(x)$  represents the property of being a formula in the language  $L(n)$ . Now note that  $\text{RT}_n$  proves the ramified Tarskian equivalences.

**Lemma 5.7** *Assuming that  $\varphi(\vec{x})$  is a formula of  $L(n)$  and  $n < k$ , we have:*

$$\text{RT}_k \vdash \forall \vec{x} [\overline{T_n \varphi(\vec{x})} \leftrightarrow \varphi(\vec{x})].$$

From this we get by part (i) of Lemma 5.6:

**Lemma 5.8** *For all  $\varphi(\vec{x}) \in \mathcal{L}_T$  we have:  $\text{RT}_{n+1} \vdash \forall \vec{x} [T_n g_n(\varphi(\vec{x})) \leftrightarrow g_n(\varphi(\vec{x}))]$ .*

After this preliminary work we are ready to prove that  $\text{FS}_n \vdash \varphi$  implies  $\text{RT}_{2n} \vdash g_i(\varphi)$ .

**Theorem 5.9** *If  $i \leq n$  and  $\varphi$  is sentence of the language  $\mathcal{L}_T$ , the following implication holds:*

$$\text{FS}_i \vdash \varphi \implies \text{RT}_{2n} \vdash g_i(\varphi) \wedge g_{i+1}(\varphi) \wedge \dots \wedge g_{2n-i}(\varphi).$$

*Proof:* Let  $n$  be fixed; then the claim can be proved by induction on  $i$ .

*Case 1:*  $i = 0$ . At first we show for  $k \leq 2n$

$$\text{PA} \vdash \varphi \implies \text{RT}_{2n} \vdash g_k(\varphi).$$

PA does not contain any axiom involving T, except the induction axioms. So if  $\text{FS}_0 \vdash \varphi$  and in  $\varphi$  all subformulas  $Tt$  are replaced uniformly by an arbitrary formula resulting in a new formula  $\hat{\varphi}$ , we still have  $\text{PA} \vdash \hat{\varphi}$  and therefore  $\text{RT}_{2n} \vdash g_k(\varphi)$  if  $k \leq 2n$ , too.

*Case 2:*  $i = 1$ . For this case we must show:

$$\text{If } \varphi \text{ is an axiom of } \text{FS}_1, \text{ then } \text{RT}_{2n} \vdash g_1(\varphi) \wedge g_2(\varphi) \wedge \dots \wedge g_{2n-1}(\varphi). \quad (5)$$

For that purpose we show that all translations of  $\text{FS}_1$ -axioms are derivable within  $\text{RT}_{2n}$ . The first axiom can be established in the following way ( $k < 2n$ ):

$$\begin{aligned} \text{RT}_{2n} &\vdash \forall x[\text{At}(x) \rightarrow (T_k(x) \leftrightarrow \text{Ver}(x))] \\ \text{RT}_{2n} &\vdash \forall x[\text{At}(x) \rightarrow (T_k g_k(x) \leftrightarrow \text{Ver}(x))] \quad (\text{from Lemma 5.6, part (iv)}) \\ \text{RT}_{2n} &\vdash g_{k+1}(\forall x[\text{At}(x) \rightarrow (T(x) \leftrightarrow \text{Ver}(x))]). \end{aligned}$$

So (5) is shown for the first axiom of  $\text{FS}_1$ . Now for the second we reason again assuming  $k < 2n$ :

$$\begin{aligned} \text{RT}_{2n} &\vdash \forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}_k \neg g_k(x) \leftrightarrow \neg \text{T}_k g_k(x))] \\ \text{RT}_{2n} &\vdash \forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}_k g_k(\neg x) \leftrightarrow \neg \text{T}_k g_k(x))] \\ \text{RT}_{2n} &\vdash g_{k+1}(\forall x[\text{Sent}_{\mathcal{L}_T}(x) \rightarrow (\text{T}(\neg x) \leftrightarrow \neg \text{T}(x))]. \end{aligned}$$

The other axioms are treated in a similar way, and from (5) the claim follows for the case  $i = 0$ . For, if  $\text{RT}_{2n} \vdash g_{k+1}(\varphi)$  and  $\psi$  is a logical consequence of  $\varphi$ , we also can conclude that  $\text{RT}_{2n} \vdash g_{k+1}(\psi)$ .

$i \rightarrow i + 1 \leq n$ . First we turn to the rule NEC of necessitation. Suppose  $\text{FS}_i \vdash \varphi$ . Hence  $\text{FS}_{i+1}^* \vdash \text{T}\bar{\varphi}$  if  $\varphi$  is a sentence of  $\mathcal{L}_T$ . By induction hypothesis it follows that

$$\begin{aligned} \text{RT}_{2n} &\vdash g_i(\varphi) \wedge g_{i+1}(\varphi) \wedge \dots \wedge g_{2n-i}(\varphi) \\ \text{RT}_{2n} &\vdash \text{T}_i g_i(\bar{\varphi}) \wedge \text{T}_{i+1} g_{i+1}(\bar{\varphi}) \wedge \dots \wedge \text{T}_{2n-i} g_{2n-i}(\bar{\varphi}) \quad (\text{by Lemma 5.8}) \\ \text{RT}_{2n} &\vdash g_{i+1}(\text{T}\varphi) \wedge g_{i+2}(\text{T}\varphi) \wedge \dots \wedge g_{2n-i+1}(\text{T}\varphi). \end{aligned}$$

The claim being already established for all  $\text{FS}_1$ -axioms the following condition holds for all  $\varphi \in \mathcal{L}_T$ .

$$\text{FS}_{i+1}^* \vdash \varphi \quad \Rightarrow \quad \text{RT}_{2n} \vdash g_{i+1}(\varphi) \wedge \dots \wedge g_{2n-i}(\varphi).$$

Now suppose  $\text{FS}_{i+1}^* \vdash \text{T}\bar{\varphi}$ ,  $\varphi$  being a sentence of  $\mathcal{L}_T$ . Using the induction hypothesis we conclude:

$$\begin{aligned} \text{RT}_{2n} &\vdash g_{i+1}(\text{T}\bar{\varphi}) \wedge \dots \wedge g_{2n-i}(\text{T}\bar{\varphi}) \\ \text{RT}_{2n} &\vdash \text{T}_i g_i(\bar{\varphi}) \wedge \dots \wedge \text{T}_{2n-i-1} g_{2n-i-1}(\bar{\varphi}) \\ \text{RT}_{2n} &\vdash g_i(\bar{\varphi}) \wedge \dots \wedge g_{2n-(i+1)}(\bar{\varphi}) \quad (\text{by Lemma 5.8}). \end{aligned}$$

So we have for all sentences  $\varphi \in \mathcal{L}_T$ :

$$\text{FS}_{i+1} \vdash \varphi \quad \Rightarrow \quad \text{RT}_{2n} \vdash g_{i+1}(\varphi) \wedge \dots \wedge g_{2n-(i+1)}(\varphi).$$

If we put  $i = n$  we obtain the following result:

$$\text{FS}_n \vdash \varphi \quad \Rightarrow \quad \text{RT}_{2n} \vdash g_n(\varphi).$$

Since  $\varphi$  in  $\mathcal{L}$  satisfies again

$$g_n(\varphi) = \varphi$$

we know from Lemma 5.5 that  $\text{RT}_{<\omega}$  and  $\text{FS}$  have the same mathematical content, i.e. a formula of  $\mathcal{L}$  is provable in  $\text{RT}_{<\omega}$ , if and only if it is in  $\text{FS}$ . From Theorem 5.9 we can also conclude the following corollary which is worth noting because of the  $\omega$ -inconsistency of  $\text{FS}$ .

**Corollary 5.10** *FS is arithmetically sound.*

*Proof:*  $\text{RT}_{<\omega}$  is easily seen to prove only true arithmetical statements, because its standard interpretation is given by Tarski's hierarchy of languages. If  $\text{FS} \vdash \varphi$  for a sentence  $\varphi \in \mathcal{L}$  we know from Theorem 5.9 that  $\varphi$  is also deducible in  $\text{RT}_{<\omega}$ , and must therefore be true.

There is also an easy model-theoretic argument proving the arithmetical soundness of FS. Assuming that  $\varphi \in \mathcal{L}$  is a sentence such that  $\text{FS} \vdash \varphi$  and  $\mathfrak{M} \in \mathbf{M}$ , there must be an  $n$  such that  $\text{FS}_n \vdash \varphi$  and therefore  $\Phi^n(\mathfrak{M}) \models \varphi$ . Since  $\Phi^n(\mathfrak{M})$  is an  $\omega$ -model and yields all true arithmetical statements,  $\varphi$  is true. So FS is sound with respect to its arithmetical content.

From a tiny variant of the theorem we can also obtain information about the system FS with arithmetical induction only. For the following corollary let  $\text{FS}_n \upharpoonright$ ,  $\text{FS} \upharpoonright$ ,  $\text{RT}_n \upharpoonright$  and  $\text{RT} \upharpoonright$  be the respective systems with the induction scheme restricted to the language  $\mathcal{L}$ .

**Corollary 5.11** *FS $\upharpoonright$  is conservative over PA.*

*Proof:* Note that  $\text{FS}_n \upharpoonright$  can be interpreted in  $\text{RT}_{2n} \upharpoonright$  in the same way as it was done for both systems with full induction in the theorem. We can now establish the claim by showing inductively each  $\text{RT}_{n+1} \upharpoonright$  to be conservative over  $\text{RT}_n \upharpoonright$  and hence over  $\text{RT}_0 \upharpoonright = \text{PA}$  with respect to all formulas of the language  $\mathcal{L}$ . If  $\text{RT}_n \upharpoonright \not\vdash \varphi$  there is a model  $\mathfrak{M}$  of  $\text{RT}_n \upharpoonright$  such that  $\mathfrak{M} \not\models \varphi$ . By the downward Löwenheim-Skolem theorem choose an elementarily equivalent countable submodel  $\mathfrak{M}_1$  of  $\mathfrak{M}$  and let  $\mathfrak{M}_2$  be a recursively saturated elementary extension of  $\mathfrak{M}_1$ . It follows from an argument similar to that of Kotlarski, Krajewski and Lachlan in [8] that  $\mathfrak{M}_2$  has a satisfaction class and hence can be expanded to a model of  $\text{RT}_{n+1} \upharpoonright$  satisfying exactly the same sentences of  $L(n)$  as  $\text{RT}_n \upharpoonright$ , so  $\text{RT}_{n+1} \upharpoonright \not\vdash \varphi$ .

Although we already know that  $\text{RT}_{<\omega}$  and FS prove the same arithmetical statements and that they are in this sense equivalent, we do not know whether both systems are equivalent if other notions of reducibility are considered. Our partial reductions via the functions  $h_i$  and  $g_i$  are problematic, because in contrast to many proof theoretical interpretations of theories we have not exhibited a *single* function commuting with the connectives and the quantifier when reducing FS to  $\text{RT}_{<\omega}$  and vice versa, and by the  $\omega$ -inconsistency of FS there cannot be such a function. But without much trouble we can prove that they are equivalent in the sense of Feferman [3].

**Theorem 5.12** *There is a partial recursive function  $f$  satisfying the following condition: If  $\varphi \in \mathcal{L}$  and  $B$  is a proof of  $\varphi$  within FS, then  $f(B)$  is defined and  $f(B)$  is a proof of  $\varphi$  in  $\text{RT}_{<\omega}$ . Moreover this can be shown within  $\text{RT}_{<\omega}$ .*

*Proof:* We roughly outline how to construct  $f$ . Assume that  $B$  is a proof of a formula  $\varphi \in \mathcal{L}$  in FS and let  $n$  be the number of applications of NEC in  $B$  and  $m$  that of CONEC. Hence if  $k := \max(m, n) + 1$  we know that  $\text{FS}_k \vdash \varphi$ . According to the construction in the proof of Theorem 5.9 let  $f$  replace  $B$  by a proof for  $g_k(\varphi) = \varphi$  in  $\text{RT}_{<\omega}$  making use of the function  $g_0, \dots, g_{2k}$  (It has to be checked that this can be done effectively.). As usual it is also left to the reader that this property of  $f$  may be established within  $\text{RT}_{<\omega}$ .

Of course,  $\text{RT}_{<\omega}$  is also reducible to FS in this sense as can be shown in a similar way as in the sketch of the proof of the above theorem.

Because FS is  $\omega$ -inconsistent it may seem difficult to obtain a global interpretation, that is, a single function interpreting FS into  $\text{RT}_{<\omega}$  and vice versa. But we can apply Orey's compactness theorem of [10] to get such an interpretation.

**Theorem 5.13** *There is a syntactical interpretation of FS in  $\text{RT}_{<\omega}$  and vice versa.*

*Proof:* Although Orey's paper deals only with theories in the language  $\mathcal{L}$ , his argument can be carried out as in [10] if we can show that both FS and  $\text{RT}_{<\omega}$  are reflexive theories, that is, they prove the consistency of all their finite subtheories, respectively. For the latter this is obvious. So let  $A$  be a finite set of theorems of FS. Hence there must be an  $n$  such that  $\text{FS}_n \vdash \varphi$  for all  $\varphi \in A$  and so  $\text{RT}_{2n} \vdash g_n(\varphi)$ . Since  $\text{RT}_{2n+1} \vdash \text{Con}_{\text{RT}_{2n}}$  and therefore by formalization  $\text{RT}_{2n+1} \vdash \text{Con}_{\text{FS}_n}$  and  $\text{Con}_{\text{FS}_n}$  is an arithmetical sentence,  $\text{Con}_{\text{FS}_n}$  is also derivable in  $\text{FS}_{2n+1}$  and therefore  $\text{FS} \vdash \text{Con}_A$ .

It should be noted that a global interpretation, like the above obtained by Orey's theorem, cannot map every arithmetical statement onto itself, in contrast to the local interpretations  $h_n$  and  $g_n$ , because the translation of an unrestricted quantifier in a global interpretation has to be a restricted quantifier. The reason for this is the  $\omega$ -inconsistency of FS.

**Acknowledgments** I am heavily indebted to Andrea Cantini for his advice which I enjoyed during a half year stay in Florence in 1993, and for his bringing the system FS to my attention. I owe thanks also to Wallace Fryer, Karl-Georg Niebergall, and Matthias Varga von Kibéd for their comments on earlier drafts of this paper.

## REFERENCES

- [1] Belnap, N. and A. Gupta, *The revision theory of truth*, MIT Press, Cambridge, 1993. [Zbl 0858.03010](#) [MR 95f:03003](#) 4
- [2] Feferman, S., "Systems of predicative analysis," *Journal of Symbolic Logic*, vol. 29 (1964), pp. 1–30. [Zbl 0134.01101](#) [MR 33:1228](#)
- [3] Feferman, S., "Hilbert's program relativized: proof-theoretical and foundational reductions," *Journal of Symbolic Logic*, vol. 53 (1988), pp. 364–384. [Zbl 0656.03002](#) [MR 89k:03074](#) 5
- [4] Feferman, S., "Reflecting on incompleteness", *Journal of Symbolic Logic*, vol. 56 (1991), pp. 1–49. [Zbl 0746.03046](#) [MR 93b:03097](#) 3, 5
- [5] Friedman, H. and M. Sheard, "An axiomatic approach to self-referential truth," *Annals of Pure and Applied Logic*, vol. 33 (1987), pp. 1–21. [Zbl 0634.03058](#) [MR 88d:03110](#) 2, 3, 3, 4
- [6] Gupta, A., "Truth and paradox," *Journal of Philosophical Logic*, vol. 11 (1982), pp. 1–60. [Zbl 0512.03002](#) [MR 84i:03020a](#) 4
- [7] Herzberger, H., "Notes on naive semantics," *Journal of Philosophical Logic*, vol. 11 (1982), pp. 61–102. [Zbl 0513.03003](#) [MR 83h:03008a](#) 4
- [8] Kotlarski, H., S. Krajewski and A. Lachlan, "Construction of satisfaction classes for nonstandard models," *Canadian Mathematical Bulletin*, vol. 24 (1981), pp. 283–293. [Zbl 0471.03054](#) [MR 84c:03066a](#) 5
- [9] McGee, V., "How truthlike can a predicate be? A negative result," *Journal of Philosophical Logic*, vol. 14 (1985), pp. 399–410. [Zbl 0583.03002](#) [MR 87g:03009](#) 4, 4
- [10] Orey, S., "Relative interpretations," *Journal of Symbolic Logic*, vol. 7 (1961), pp. 146–153. [Zbl 0121.25501](#) [MR 26:3608](#) 5, 5

*Ochsenfelder Straße 12  
85072 Eichstätt  
Germany  
email: [halbach@cis.uni-muenchen.de](mailto:halbach@cis.uni-muenchen.de)*