Notes on ω-inconsistent Theories of Truth in Second-Order Languages

Eduardo Barrio & Lavinia Picollo

Draft! Not final version!

Abstract

It is widely accepted that a theory of truth for arithmetic should be consistent, but ω-consistency is less frequently required. This paper argues that ω-consistency is a highly desirable feature for such theories. The point has already been made for first-order languages, though the evidence is not entirely conclusive. We show that in the second-order case the consequence of adopting ω-inconsistent truth theories for arithmetic is unsatisfiability. In order to bring out this point, well known ω-inconsistent theories of truth are considered: the revision theory of nearly stable truth $\mathbf{T}^{\#}$ and the classical theory of symmetric truth **FS**. Briefly, we present some conceptual problems with ω-inconsistent theories, and demonstrate some technical results that support our criticisms of such theories.

Keywords: theories of truth, second-order languages, ω-inconsistency, unsatisfiability

In this paper we argue that any consistent but ω-inconsistent theory of truth for arithmetic will fail to be adequate.[1] The point has already been made for first-order systems and is widely known and generally accepted.[2] The consequences of adopting second-order ω-inconsistent theories of truth have not been explored yet. Naturally, the negative results obtained for the first-order case continue to hold in the second-order case. However, we will show that in the second-order case things get considerably worse. We provide some new results and proofs for the well known ω-inconsistent revision theory of truth $\mathbf{T}^{\#}$ and classical theory of symmetric truth **FS** over second-order arithmetic.[3]

[1] The expressions 'truth theory' and 'truth system' will not be used in a technical sense, but will be taken to be interchangeable and to refer to any semantic or axiomatic formal approach to truth.

[2] See Leitgeb [13], Barrio [1] and Halbach [10, p. 134].

[3] We will focus on theories of truth over arithmetic, but applications to other more comprehensive base theories are presumably intended for these theories. Arithmetic is a convenient and relatively simple setting, since by fixing some Gödel coding it can express its own syntax. Limiting ourselves to arithmetic does not harm the generality of our claims: if a theory fails to provide a satisfactory account of truth for some (crucial) base system, then it is not an attractive theory of truth overall, for the general aim is lost.

The article is organized as follows. In section 1 we give some technical preliminaries. Section 2 is devoted to the introduction of **FS** and $\mathbf{T}^{\#}$ over second-order arithmetic. In 3 we provide unsatisfiability results for both systems and consistency and soundness results for the latter. We argue that consistency and soundness are not sufficient for an adequate theory of truth. In section 4 we draw conclusions.

1 Technical preliminaries

Let $\mathcal{L}^2$ be the usual second-order language of arithmetic. It contains $\neg$, $\wedge$ and $\forall$ as logical symbols ($\vee$, $\rightarrow$, $\leftrightarrow$ and $\exists$ are defined), denumerably many individual variables $x_1, x_2, ...$, denumerably many n-ary set variables $X_1^n, X_2^n, ...$, 0 as its only individual constant, the monadic function symbol s and finitely many additional symbols for primitive recursive functions that will be needed. $\mathcal{L}_T^2$ obtains from $\mathcal{L}^2$ by adding a new monadic predicate letter T for truth.

Let $\mathcal{N}^2$ be the intended model of $\mathcal{L}^2$, and let ω be its first-order domain. $\mathcal{N}^2$ interprets the constant 0 with the number 0, s with the successor function and other primitive recursive function symbols in the intended way. By 'standard' or 'intended model of $\mathcal{L}_T^2$' we mean any second-order interpretation of this language whose restriction to $\mathcal{L}^2$ is isomorphic to $\mathcal{N}^2$.

Shapiro has recently argued that unintended interpretations of first-order arithmetic demonstrate that first-order languages are inadequate for axiomatizing arithmetic.[4] He maintains that arithmetic should be formulated in a language whose resources transcend first-order logic. He proposes that second-order languages provide a suitable framework. Second-order languages contain not just first-order quantifiers that range over elements of the domain, but also second-order quantifiers that range over subsets of the domain. In *full* second-order logic, it is crucial that these second-order quantifiers range over *all* subsets of the domain.

Thus, by 'second-order interpretation' or 'second-order model' we understand any classical interpretation of a second-order language as $\mathcal{L}^2$ where the second-order domain is the power set of the first-order domain. If we did not limit ourselves to standard semantics we would not be really working with second-order languages but just with multi-sorted first-order ones.

Let PA^2 be the usual recursive axiomatization of second-order arithmetic. It contains the usual axioms of first-order arithmetic plus the second-order formulation of induction instead of the first-order induction schema. If the principles of arithmetic are formulated in a second-order language, then Dedekind's argument goes through and we have a categorical theory. PA_T^2 is PA^2 formulated in $\mathcal{L}_T^2$, with full comprehension over $\mathcal{L}_T^2$-formulae. We assume $\mathcal{N}^2 \vDash PA^2$.

$\mathcal{L}_T^2$ contains a term $\bar{n}$—the numeral of n—for each $n \in \omega$ given by n occurrences of s

[4]For a detailed account see Shapiro [15, pp. 70-76].

followed by the constant symbol 0. Given any piece of vocabulary A of $\mathcal{L}_T^2$, $\ulcorner A \urcorner$ denotes the numeral of the Gödel number of A, given some fixed coding.

The sets of atomic and true atomic sentences of $\mathcal{L}^2$ are recursive, as well as the set of $\mathcal{L}_T^2$-sentences, the set of first-order variables, the set of second-order variables and the set of predicate letters. They will be represented in PA^2—and thus expressed in $\mathcal{N}^2$—by $At(x)$, $Ver(x)$, $Sent(x)$, $var(x)$, $Var(x)$ and $Pred(x)$, correspondingly.

The function symbols $\dot{\neg}$, $\dot{\wedge}$ and $\dot{\forall}$ represent recursive functions such that for any formulae A and B, individual variable v and second-order variable V: $\dot{\neg}\ulcorner A \urcorner = \ulcorner \neg A \urcorner$, $\ulcorner A \urcorner \dot{\wedge} \ulcorner B \urcorner = \ulcorner A \wedge B \urcorner$, $\dot{\forall}\ulcorner v \urcorner \ulcorner A \urcorner = \ulcorner \forall v A \urcorner$ and $\dot{\forall}\ulcorner V \urcorner \ulcorner A \urcorner = \ulcorner \forall V A \urcorner$. $\dot{x}$ represents the function that maps any number n to the code of its numeral $\bar{n}$. Finally, $x(y/z)$ represents the substitution function, which applied to the code x of a formula A and the codes y and z of terms t_1 and t_2, gives the code of the formula that obtains by replacing t_2 in A with t_1; while when applied to the codes y and z of relation symbols or second-order variables of the same arity R_1 and R_2 it gives the code of the formula that obtains by substituting R_2 in A with R_1. As usual, we write $\ulcorner A(\dot{v}) \urcorner$ as short for $\ulcorner A(v) \urcorner (\dot{v}/\ulcorner v \urcorner)$ to bind the individual variable v from outside corner quotes.

2 Two ω-inconsistent theories of truth

2.1 The complete and consistent theory of truth **FS**

FS is an axiomatic theory of truth introduced by Friedman and Sheard [6], and studied in depth by Halbach [9], [10]. Given a base theory formulated in a classical language containing a predicate T for truth, **FS** adds axioms and rules governing T to the base theory that are consistent with it and intended to turn T into a truth predicate for the resulting system. **FS** provides compositionality principles and two rules of introduction and elimination for T that turn it into a symmetric notion: every provable formula of the theory will also be provably true and vice versa.

Let $\mathcal{L}_T^2$ be our classical language and PA_T^2 our base theory. We will also refer to the latter as 'FS_0^2'. FS_1^2 is FS_0^2 plus the following axioms:

(AT) $\forall x(At(x) \rightarrow (Tx \leftrightarrow Ver(x)))$

$(T\neg)$ $\forall x(Sent(x) \rightarrow (T\dot{\neg}x \leftrightarrow \neg Tx))$

$(T\wedge)$ $\forall x \forall y(Sent(x) \wedge Sent(y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow Tx \wedge Ty))$

$(T\forall v)$ $\forall x \forall v(Sent(x(0/v)) \wedge var(v) \rightarrow (T\dot{\forall}vx \leftrightarrow \forall y Tx(\dot{y}/v)))$

$(T\forall V)$ $\forall x \forall p \forall v(Var(v) \wedge Pred(p) \wedge Sent(x(p/v)) \rightarrow (T\dot{\forall}vx \rightarrow Tx(p/v)))$

$(T\neg)$, $(T\wedge)$ and $(T\forall v)$ collaborate to establish the compositional character of truth. One might expect $(T\forall V)$ to do its part, *i.e.* to state that T commutes with the second-order universal quantifier too. Such a principle, for instance, would require $\forall X A$ to

be true whenever each formula that results from replacing all free occurrences of X in A with a monadic predicate symbol is true. But there cannot be enough predicate symbols in $\mathcal{L}_T^2$ for each $X \subseteq \omega$, for $\mathcal{L}_T^2$ is a denumerable language; it may happen that all instances of $A(X)$ are true, while its universal closure is not.[5] $(T\forall V)$ states only the other direction, which is sound: if the universal closure of $A(X)$ is true, then the result of replacing X with a predicate letter of the same arity is true too.

Finally, let FS^2 be FS_1^2 plus the following inference rules:

$$(\textsc{Nec}) \quad \frac{\vdash A}{\vdash T^{\ulcorner}A^{\urcorner}} \qquad\qquad (\textsc{Conec}) \quad \frac{\vdash T^{\ulcorner}A^{\urcorner}}{\vdash A}$$

FS^2 is an ω-inconsistent system, that is, for some formula $A(x)$ with only one free individual variable x, $FS^2 \vdash A(\bar{n})$ for each $n \in \omega$ and, at the same time, $FS^2 \vdash \neg\forall x A(x)$. This is an immediate consequence of a theorem of McGee [14, p. 399].

2.2 The revision theory of truth $\mathbf{T}^{\#}$

The revision theory of truth is an attempt to show how a classical language may contain its own truth predicate by describing and explaining the behavior of this predicate in ordinary as well as in problematic cases such as those posed by semantic paradoxes. It was originally introduced by Gupta [7] and Herzberger [12] and fully developed by Gupta and Belnap [8], the *locus classicus* on the topic. The revision theory is said to be a semantic truth theory. But despite working with a class of models for the language whose truth predicate is to be explained, it does not provide a class of models for the language but instead provides a class of sentences intended to be the ones that are categorically assertible.

Let $\mathcal{L}_T^2$ be the language under investigation. The revision theory works as follows. A partial interpretation of $\mathcal{L}_T^2$ that just leaves T uninterpreted must be fixed to begin with: the *base model*. Let $\mathcal{N}^2$ be our base model. Then a hypothetical extension $S_0 \subseteq \omega$ for T must be chosen and $\mathcal{N}^2$ expanded to a full second-order model $(\mathcal{N}^2, S_0)$ of $\mathcal{L}_T^2$ with the same first-order domain,[6] that interprets T with S and every other non-logical symbol of $\mathcal{L}_T^2$ with the same objects as $\mathcal{N}^2$.

Now a revision process of the chosen hypothesis begins. $(\mathcal{N}^2, S_0)$ is transformed into another model $(\mathcal{N}^2, S_1)$, and this one into another, etc. obtaining a sequence of interpretations indexed by ordinal numbers of length On.[7] We use greek letters α, β, λ

[5] Adopting a satisfaction predicate instead of a truth predicate could solve the problem, but we would be forced to move to a third-order language allowing predicates to apply, not only to objects in the first-order domain of a model, but also to sets in the second-order domain. This would make things rather more complicated and it does not bear on the aim of our paper.

[6] And thus the second-order domain remains intact too.

[7] The class of all ordinal numbers.

to denote ordinals. Let A be any sentence of $\mathcal{L}_T^2$. At each successor step $\alpha + 1$ the extension of the truth predicate is given by the following rule:

$$S_{\alpha+1} = \{A : (\mathcal{N}^2, S_\alpha) \vDash A\}^8 \tag{1}$$

Only sentences that are true in the previous level fall inside the extension of the truth predicate. At limit levels things get more complicated. Clearly, results obtained in previous stages must be collected: those sentences stabilized inside the extension of T must remain there, and the same for sentences stabilized outside the extension of the truth predicate. What about unstable sentences? Gupta and Belnap [8] consider all possible ways of adding some unstable sentences to the extension of T at each limit stage λ.[9] Let A be any sentence of $\mathcal{L}_T^2$ and Γ_λ any subset of ω:

$$S_\lambda = \{A : \exists\alpha\forall\beta(\alpha \leq \beta < \lambda \Rightarrow A \in S_\beta)\} \cup \Gamma_\lambda - \{A : \exists\alpha\forall\beta(\alpha \leq \beta < \lambda \Rightarrow A \notin S_\beta)\}$$

Different choices of the initial hypothesis S_0 and the extension of Γ_λ at each limit stage λ give rise to different revision sequences. To prevent arbitrary choices from slanting the process, every possible sequence with $\mathcal{N}^2$ as base model must be considered. Sentences that stabilize *in some way* inside the extension of T in *every* revision sequence will be categorically true and assertible, while the ones that stabilize outside the extension will be categorically false and their negations will be assertible.

In $\mathbf{T}^\#$ categorical statements must be *nearly stable*.[10] An $\mathcal{L}$-sentence A is nearly-stably true in a sequence generated by $(\mathcal{N}^2, S_0)$ if and only if for every stage β after some stage α there is a natural number n such that for all natural numbers $m \geq n$, $A \in S_{\beta+m}$: and similarly for nearly-stably false sentences. Nearly stable sentences are allowed to fluctuate all along sequences, but those fluctuations must be confined to finite regions immediately after limit ordinals.

An $\mathcal{L}_T^2$-sentence A is valid in $\mathbf{T}^\#$ in $\mathcal{N}^2$—$T^\#_{\mathcal{N}^2}$ for short—if and only if it is nearly stably true in every sequence based on $\mathcal{N}^2$.

Gupta and Belnap [8, p. 225] prove that $\mathbf{T}^\#$ in $\mathcal{N}$—the standard model of first-order arithmetic—is ω-inconsistent, for it satisfies the hypothesis of McGee's [14] theorem. Since $T^\#_{\mathcal{N}^2}$ is an extension of that system, it is ω-inconsistent too. In fact, FS^2 is nearly-stably true and, thus, a subtheory of $T^\#_{\mathcal{N}^2}$.[11] So $T^\#_{\mathcal{N}^2}$ is a system of both compositional

[8] Actually, codes of sentences rather than sentences themselves belong to each S_α. However, for readability purposes we will frequently identify expressions with their codes.

[9] Previously, some alternatives have been explored in the literature by Belnap [2], Gupta [7] and Herzberger [12], for which Gupta and Belnap's [8] notion seems to be an improvement. Later, Yaqūb [16] and Chapuis [3] worked on several refinements.

[10] Gupta and Belnap present three diverse systems built around different ways a sentence may stabilize in a sequence, one of which is $\mathbf{T}^\#$. For a detailed exposition see Gupta and Belnap [8, chapter 6].

[11] Gupta and Belnap [8, p. 222] prove that first-order versions of $(T\neg)$, $(T\wedge)$ and $(T\forall v)$ are valid in $\mathbf{T}^\#$ in $\mathcal{N}$. The proof for (AT), $(T\neg)$, $(T\wedge)$, $(T\forall v)$ and $(T\forall V)$ in $T^\#_{\mathcal{N}^2}$ is analogous. Of course, since the latter entails every true-in-$\mathcal{N}^2$ $\mathcal{L}^2$-sentence, it is not axiomatizable. FS^2 can only be seen as a partial axiomatization.

and symmetric truth too.

3 Some relevant results

Dedekind's categoricity result states that any structures satisfying the axioms of second-order arithmetic are isomorphic. Philosophers of mathematics—*e.g.* Shapiro, Isaacson—have repeatedly claimed that this result has significant implications with respect to the determinacy of our understanding of the natural numbers. In the second-order case ω-inconsistency entails unsatisfiability. Both $T^{\#}_{\mathcal{N}^2}$ and FS^2 lack models.

Theorem 3.1 *FS^2 has no (full) models.*

Proof *Suppose for reductio $\mathcal{M} \vDash FS^2$. As FS^2 is ω-inconsistent, there is a formula $A(x)$ with exactly one free individual variable x such that $\mathcal{M} \vDash A(\bar{n})$ for each $n \in \omega$ and also $\mathcal{M} \vDash \neg\forall x A(x)$. Since FS^2 extends PA^2, $\mathcal{M} \vDash PA^2$. Then, by categoricity,[12] $\mathcal{M}$ must be an ω-model. Thus, $\mathcal{M} \vDash \forall x A(x)$ too, which is impossible.* □

Corollary 3.2 *The set of $\mathcal{L}^2_T$-sentences that are valid in $T^{\#}_{\mathcal{N}^2}$ has no (full) models.*

Proof *By theorem 3.1, since FS^2 is a subsystem of $T^{\#}_{\mathcal{N}^2}$.* □

These are definitively negative results. The lack of models for FS^2 implies, in the first place, that the non-logical vocabulary of FS^2 cannot be interpreted in any way. Thus, FS^2 'talks' about nothing, neither true statements nor natural numbers. In a sense, this turns it into a useless theory. In the second place, the lack of models shows that this formal system semantically entails everything, it is semantically trivial. One might feel inclined to believe that as a result of being unsatisfiable FS^2 is also inconsistent. However, as is well known, Theorem 3.1 does not entail that FS^2 is inconsistent, for there is no complete second-order calculus and what happens at the semantic level may carry no proof-theoretical consequences. In fact, FS^2 is consistent and even arithmetically sound, *i.e.*, it proves only true $\mathcal{L}^2$-statements.

Let FS^2_n be FS^2_1 plus at most $n-1$ applications of (Nec) and $n-1$ applications of (Conec). We will show that, for each $n \in \omega$, FS^2_n has an ω-model.[13] Since every theorem of FS^2 must be provable in some FS^2_n, FS^2 must be consistent and arithmetically sound. First we will prove an auxiliary lemma.

Lemma 3.3 *Let $n \in \omega$ and A be any formula of $\mathcal{L}^2_T$. If A is true in the n-th step of every revision sequence based on $\mathcal{N}^2$ then it is also true at stage $n+1$.*

[12] Dedekind [4] proved a categoricity result for PA^2: any model of PA^2 is isomorphic to $\mathcal{N}^2$ and, thus, an ω-model.

[13] A similar proof for the first-order case can be found in Halbach [10, chap. 14, sec. 1].

Proof Consider a revision sequence generated by S_0. Then, $(\mathcal{N}^2, S_n) \vDash A$. Consider now the sequence generated by S_1 instead. The n-th step of the latter sequence is $(\mathcal{N}^2, S_{n+1})$. Thus, $(\mathcal{N}^2, S_{n+1}) \vDash A$. □

Theorem 3.4 *If* $(\mathcal{N}^2, S_0)$ *is an extension of* $\mathcal{N}^2$ *to* $\mathcal{L}_T^2$, $(\mathcal{N}^2, S_n) \vDash FS_n^2$.

Proof Since FS_0^2 is PA_T^2 and the latter contains just logical axioms and rules for T, any extension of $\mathcal{N}^2$ to $\mathcal{L}_T^2$ satisfies it. Thus, $(\mathcal{N}^2, S_0) \vDash FS_0^2$.

We will prove the remaining cases by induction on n with $n = 1$ as our base step. By the last paragraph, $(\mathcal{N}^2, S_1) \vDash FS_0^2$. We need to show that (AT), $(T\neg)$, $(T\wedge)$, $(T\forall v)$ and $(T\forall V)$ are also true in $(\mathcal{N}^2, S_1)$. We will explicitly prove that $(T\neg)$ and $(T\forall V)$ are satisfied. Other cases are treated in a similar way.

Suppose for *reductio* that $(\mathcal{N}^2, S_1) \nvDash (T\neg)$. Then $(\mathcal{N}^2, S_1) \vDash Sent(\bar{n}) \wedge T\dot{\neg}\bar{n} \wedge T\bar{n}$ or $(\mathcal{N}^2, S_1) \vDash Sent(\bar{n}) \wedge \neg T\dot{\neg}\bar{n} \wedge \neg T\bar{n}$ for some $n \in \omega$. Thus, there is a sentence A of $\mathcal{L}_T^2$ such that $\neg A \in S_1$ and $A \in S_1$, or $\neg A \notin S_1$ and $A \notin S_1$. By (1), we have that either $(\mathcal{N}^2, S_0) \vDash \neg A \wedge A$ or $(\mathcal{N}^2, S_0) \nvDash \neg A \vee A$, which is absurd.

Similarly, assume that $(\mathcal{N}^2, S_1) \nvDash (T\forall V)$. Thus, $(\mathcal{N}^2, S_1) \vDash Var(\bar{k}) \wedge Pred(\bar{m}) \wedge Sent(\bar{n}(\bar{m}/\bar{k})) \wedge T\dot{\forall}\bar{k}\bar{n} \wedge \neg T\bar{n}(\bar{m}/\bar{k})$ for some $k, m, n \in \omega$. Then, there is a second-order variable V, predicate symbol P of the same arity as V and a formula A with possibly V as its only free variable such that $\forall V A \in S_1$ and $A(P) \notin S_1$. By (1), $(\mathcal{N}^2, S_0) \vDash \forall V A \wedge \neg A(P)$, which is impossible.

Now suppose that $(\mathcal{N}^2, S_n) \vDash FS_n^2$, *i.e.*, that FS_n^2 is true in the n-th step of every revision sequence based on $\mathcal{N}^2$. By (3.3), $(\mathcal{N}^2, S_{n+1}) \vDash FS_n^2$ too. So $(\mathcal{N}^2, S_{n+1})$ validates $n-1$ applications of (Nec) and (Conec). Finally we need to prove that $(\mathcal{N}^2, S_{n+1})$ validates one more application of each rule. Let A be an $\mathcal{L}_T^2$-sentence such that $FS_n^2 \vdash A$. By the inductive hypothesis, $(\mathcal{N}^2, S_n) \vDash A$. By (1), $A \in S_{n+1}$, that is, $(\mathcal{N}^2, S_{n+1}) \vDash T\ulcorner A\urcorner$. Now let $FS_n^2 \vdash T\ulcorner A\urcorner$. Then, $(\mathcal{N}^2, S_{n+1}) \vDash T\ulcorner A\urcorner$, *i.e.*, $A \in S_{n+1}$. By (1), $(\mathcal{N}^2, S_n) \vDash A$ and, by (3.3), $(\mathcal{N}^2, S_{n+1}) \vDash A$. Therefore, $(\mathcal{N}^2, S_n) \vDash FS_{n+1}^2$. □

So each FS_n^2 is true at stage n of every revision sequence for $\mathcal{L}_T^2$ based on $\mathcal{N}^2$, no matter what initial hypothesis we have chosen.

Corollary 3.5 *FS^2 is consistent.*

Corollary 3.6 *FS^2 is arithmetically sound.*

While theorem 3.1 shows that FS^2 is trivial from a semantic standpoint, corollary 3.5 keeps it safe from proof-theoretic trivialization and corollary 3.6 from arithmetical falsity. The failure of completeness for second-order systems allows these differences between the semantics and the calculus. Nonetheless, consistency and soundness are not enough; ω-consistency is necessary for a theory of truth.

Adding a truth predicate to some base theory should not interfere with the ontology of that theory. First, it does not make much sense to talk about the truth or falsity of uninterpreted formulae. We want sentences such as '$\neg\exists x(s(x) = 0)$' to come out true in our truth theories *because* they say something true about natural numbers, because they are true of the standard interpretation of the language we are working with. Although FS^2 entails many of these sentences and also their truth predication, it cannot be seen as expressing the truth of a sentence that concerns natural numbers (or anything at all), since FS^2 can be no longer seen as saying something true about ω.

Second, one might expect a truth theory to provide a better understanding of the standard interpretation of the base language. But despite being arithmetically sound and proving more true-in-$\mathcal{N}^2$ formulae than PA^2, including the Gödel sentence and the consistency statement for PA^2, FS^2 does not provide a better characterization of $\mathcal{N}^2$ since it is not true in it.

Then, as a consequence of theorem 3.1—which is itself a consequence of the ω-inconsistency of FS^2—the truth predicate of FS^2 is not a legitimate truth predicate for arithmetic, not even a partial one.

A good theory of truth over arithmetic should not only be arithmetically sound and entail as many intuitive truth principles as possible, but it should also not imply counterintuitive statements involving truth. Consider the following $\mathcal{L}_T^2$-sentence:

$(RFL_{FS^2})\ \forall x(Bew_{FS^2}(x) \rightarrow Tx)$

where Bew_{FS^2} is the provability predicate for FS^2, weakly representable in this system. RFL_{FS^2} is a global reflection principle for FS^2: it states that all FS^2-theorems are true (as long as T is capable of expressing truth, at least partially). This principle seems desirable to anyone embracing FS^2, for it establishes its soundness. Moreover, it appears to be true according to FS^2 itself since, by (Nec), T applies to every theorem of FS^2.

However, FS^2 proves the negation and falsity of RFL_{FS^2}.[14] $\neg RFL_{FS^2}$ is a highly counterintuitive principle but—worst of all—it is strictly false. Although FS^2 does not prove any arithmetically false statement, it entails incorrect truth-theoretical principles. FS^2 is arithmetically but not truth-theoretically sound.

As a consequence, supporters of FS^2 must regard their own theory as unsound, for they fall into the following dilemma: they commit themselves either to the falsity or to the truth of RFL_{FS^2}. The first alternative seems reasonable, for $\neg RFL_{FS}$ is entailed both by FS^2 and $T^{\#}_{\mathcal{N}^2}$. But this formula states the unsoundness of FS^2. The second choice also seems attractive, since RFL_{FS^2} states the soundness of FS^2. However, since FS^2 implies the negation of that principle, supporters are forced to admit that their theories entail falsities and, hence, are unsound.

[14] See Halbach and Horsten [11] for a proof for the first-order case. The second-order case is immediate, for FS^2 includes its first-order counterpart and the negation of the reflection principle for the latter system entails $\neg RFL_{FS^2}$.

Naturally, whatever is provable in FS^2 must be provable by a finite number of applications of (Nec) and (Conec). As a result, every FS^2-theorem is a theorem of an ω-consistent fragment of FS^2. ω-inconsistency is not the reason why FS^2 proves the negation of its own reflection principle, but rather why this negation is false according to the theory itself. For while a finiteness argument goes through and, thus, $\neg RFL_{FS^2}$ is provable in some ω-consistent fragment of FS^2, RFL_{FS^2} only becomes true—and its negation false—when applications of (Nec) are unrestrained. In fact, FS^2 entails $T\ulcorner A\urcorner$ whenever it entails $Bew_{FS^2}(\ulcorner A\urcorner)$ for each $\mathcal{L}_T^2$-sentence A and, at the same time, $\neg\forall x(Bew_{FS^2}(x) \rightarrow Tx)$ is a theorem of FS^2.

Regarding $T^{\#}_{\mathcal{N}^2}$, things get murkier. This truth theory provides a class of $\mathcal{L}_T^2$-sentences that are supposed to be correctly assertible. Gupta and Belnap [8, p. 219] show that this set is closed under classical logical consequence, for logical truths are true and logical rules are sound in every model, including all extensions of $\mathcal{N}^2$. Since the set of $T^{\#}_{\mathcal{N}^2}$-valid sentences lacks models we have that $T^{\#}_{\mathcal{N}^2}$ entails every $\mathcal{L}_T^2$-statement. In theorem 3.2 we show that $\mathbf{T}^{\#}$ is fully incapable of dealing with truth for second-order arithmetic. According to $\mathbf{T}^{\#}$ *every* sentence of $\mathcal{L}_T^2$ is correctly assertible, and so we get absolute triviality.

4 Conclusions

As is widely accepted, ω-inconsistent theories of truth for first-order arithmetic are undesirable, for they do not succeed in expressing genuine truth. Results for the second-order case are worse and completely decisive. Higher-order resources with standard semantics ban the existence of non-standard models. Thus, ω-inconsistency entails unsatisfiable theories of truth, *i.e.*, semantically trivial systems. While FS^2, by an incompleteness result, avoids trivialization at the proof-theoretical level, $T^{\#}_{\mathcal{N}^2}$ has the further flaw of entailing every sentence, rendering it completely useless as a semantic theory of truth.

In sum, ω-consistency is a highly desirable feature for a theory intended to provide a truth predicate for first-order arithmetic; but it becomes indispensable if the aim is to give a truth predicate for second-order arithmetic.

Acknowledgements

We owe thanks to Roy Cook, Volker Halbach, Hannes Leitgeb and Øystein Linnebo for very helpful comments on previous drafts of this paper. Earlier versions of this material were presented at conferences in MCMP-Munich, University of Oxford, the Logic Research Group of the University of Buenos Aires and Sadaf. We are grateful to the members of these audiences for their valuable feedback.

References

[1] Barrio, E. A. Theories of truth without standard models and Yablo's sequences. *Studia Logica 96* (2010), 375–391.

[2] Belnap, N. D. Gupta's rule of revision theory of truth. *Journal of Philosphical Logic 11* (1982), 110–116.

[3] Chapuis, A. Alternative revision theories of truth. *Journal of Philosphical Logic 25* (1996), 399–423.

[4] Dedekind, R. Was sind und was sollen die zahlen? In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, W. B. Ewald, Ed. Oxford University Press, 1996, pp. 787–832.

[5] Ewald, W. B., Ed. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Oxford University Press, 1996.

[6] Friedman, H., and Sheard, M. An axiomatic approach to self–referential truth. *Annals of Pure and Applied Logic 33* (1987), 1–21.

[7] Gupta, A. Truth and paradox. *Journal of Philosphical Logic 11* (1982), 1–60.

[8] Gupta, A., and Belnap, N. D. *The Revision Theory of Truth*. MIT Press, Cambridge, 1993.

[9] Halbach, V. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic 35* (1994), 311–327.

[10] Halbach, V. *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge, 2011.

[11] Halbach, V., and Horsten, L. The deflationist's axioms for truth. In *Deflationism and Paradox*, B. Armour-Garb and J. C. Beall, Eds. Oxford University Press, 2005.

[12] Herzberger, H. Notes on naive semantics. *Journal of Philosphical Logic 11* (1982), 61–102.

[13] Leitgeb, H. What theories of truth should be like (but cannot be). *Philosophy Compass 2*, 2 (2007), 276–290.

[14] McGee, V. How truth-like can a predicate be? A negative result. *Journal of Philosphical Logic 14* (1985), 399–410.

[15] Shapiro, S. *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford University Press, New York, 1991.

[16] Yaqūb, A. *The Liar Speaks the Truth. A Defense of the Revision Theory of Truth*. Oxford University Press, 1993.