Reflections on nonstandard satisfaction

Richard Kaye
School of Mathematics, University of Birmingham

11th June 2016

1 Introduction

This paper addresses a number of issues relating to definitions of truth or satisfaction over nonstandard models of arithmetic (PA).

The situation is familiar. We take one of the usual signatures for the language (such as $\mathscr{L} : +, \cdot, 0, 1, <$) and identify formulas with their Gödel number. Some usual Gödel-numbering is chosen, and we note that syntactic operations are well-behaved and absolute between a model of PA (in fact rather less is needed) and its standard initial segment.

Given a nonstandard $M \vDash \mathrm{PA}$ we wish to add a new relation or predicate $\mathrm{Tr}(\varphi)$ (for truth) for (possibly nonstandard) sentences φ or $\mathrm{Sat}(\varphi, a)$ (for satisfaction) for formulas φ and assignments of values for variables a.

Provided we do not require our truth/satisfaction predicate to satisfy too many axioms, such Tr (or Sat) can be defined for all models M. If we wish to make the inductive steps in Tarski's definition of truth hold, such Tr (or Sat) can be given for all *countable recursively saturated* M, a result due to Kotlarski, Krajewski and Lachlan [2]. That the non-trivial property of recursive saturation is necessary is a result due to Lachlan [3].

2 The relevance of nonstandard models

The objective for this first section is to justify the general methodology and framework to be used, that is, the target language for statements such as φ and the reason for using nonstandard models to analyse notions of truth.

I take it that our primary interest is in 'truth in the real world', and I shall take, as our model for the 'real world', the standard natural numbers $\mathbb{N}$, with arithmetical structure given by addition, multiplication and order. This seems to be a sensible minimum that has the expected problems of the actual (or mathematical) real world, in particular it has all the metamathematical difficulties presented by Gödel and others.

The main reason for considering nonstandard models and nonstandard sentences is to simplify the discussion and to decouple it from any additional considerations concerning the highly complex and, in some technical senses of the word, unknowable theory of the standard natural numbers $\mathbb{N}$. Our starting point is that there is some reasonable way of asserting statements such as φ; that is, we have some method or methods for determining those situations in which φ does or does not hold. I do not at the outset make any assumptions *ab initio* that these methods are Tarskian.

(Indeed, other possibilities for 'assertion' come to mind such as ones related to provability in a formal system. Or, alternatively, suppose expressions in the language were given Gödel numbers in some sophisticated and unusual way. Then the truth or falsity of a statement might be seen directly from some number theoretic properties of the Gödel number.)

To avoid circularity, methods for determining when φ does or does not hold ought not depend on complete knowledge of the real world, the structure $\mathbb{N}$, nor indeed on the full first order theory of $\mathbb{N}$, for we hope to be able to say something useful about when a statement φ is true without prior knowledge of *all* statements ψ in the same language and whether they are or are not true. Or, if it turns out that such a theory does necessarily depend on the full first order theory of $\mathbb{N}$, then that is a conclusion that we should like to draw, and not an assumption we wish to build into the investigation at the outset.

(For example, the Tarskian theory provides one such framework for truth in a model. It is based on induction, as indeed is all our knowledge of the real world $\mathbb{N}$. If the grounds for accepting that Tarskian theory were seen to be the same induction principles as those for knowledge of the real world $\mathbb{N}$, as is at least plausibly the case, then we would be left in the situation that our theory of truth of the real world is equivalent to full knowledge of the real world, a situation that would seem to present little progress. So we choose not to start with such a strong base. If this is of necessity the case, we would like to know about it, which makes a weak base necessary for our methodology so that we are able to attempt a proof.)

Thus it is essential none of our discussions about truth should be based on the assumption that we have full knowledge of $\mathbb{N}$ or even $\mathrm{Th}(\mathbb{N})$.

The simplest mathematical way to decouple our discussion of truth in $\mathbb{N}$ from $\mathbb{N}$ itself is to give a minimal set of axioms that we accept about $\mathbb{N}$ and investigate the class of (possibly nonstandard) models of that set of axioms. In the following I shall take PA as the base axioms, though with extra effort this could in most probability be replaced by some much weaker set if required.

There is the issue of the mathematical and set theoretical framework sufficient for the model theory required. In regards this I would comment that:

(a) It is indeed rather strong, and stronger than one would normally desire, but the methods of model theory has the major benefit of being clear: if some result depends on a model being an elementary extension of $\mathbb{N}$ (as opposed to just a model of PA) this will be obvious.

(b) The set theoretical basis is at least axiomatised and can be replaced if needed with an appropriate system for second order arithmetic (e.g. ACA_0, WKL_0).

(c) Proof theoretic methods are of course also possible. These seem best suited to a second and more refined approach, since they are often longer, more technical, and somewhat more error-prone, and lack the clarity mentioned in (a).

To summarise: a *theory of assertion* of statements such as φ is, we hope, substantially weaker than the full first order theory of the natural numbers. We want to examine various *theories of truth* that develop this theory of assertion. It is essential that full knowledge of 'the real world' is not built into our starting position, implicitly or explicitly. This can be seen most clearly using techniques from normal mathematical practice and model theory. Other approaches may later sharpen the results we obtain.

3 Reflections

In one sense at least the method works. Given a nonstandard model $M \models \mathrm{PA}$, by standard mathematical techniques we can say what it means for a standard formula φ to be satisfied under an assignment a, and we can define a predicate $\mathrm{Sat}(\varphi, a)$ on M capturing this exactly for such standard φ, and defined in some arbitrary way for nonstandard φ. Thus we get a satisfaction class satisfying the disquotational scheme

$$\text{(DS)} \qquad \mathrm{Sat}(\varphi, a) \leftrightarrow \varphi[a]$$

for all standard φ.

This seems to be the basis for the deflationist position on truth. However the main problem is that it does not say anything about truth for nonstandard formulas. Recall that the reason why nonstandard formulas are there in the first place is that we wish to decouple our theory of truth from complete knowledge of the real world. The theory we have at this point has the issue that it requires some additional knowledge of $\mathbb{N}$, and we are not in a position at the moment to say how much.

The usual approach is to add other axioms to our theory of satisfaction, if possible ones that have (DS) as a consequence. The most familiar set of axioms follow Tarskian lines.

$$\mathrm{Sat}(t = s, a) \leftrightarrow \mathrm{val}\,(t, a) = \mathrm{val}\,(s, a)$$
$$\mathrm{Sat}(t < s, a) \leftrightarrow \mathrm{val}\,(t) < \mathrm{val}\,(s)$$
$$\mathrm{Sat}(\neg\varphi, a) \leftrightarrow \neg\mathrm{Sat}(\varphi, a)$$
$$\mathrm{Sat}(\varphi \wedge \psi, a) \leftrightarrow \mathrm{Sat}(\varphi, a) \wedge \mathrm{Sat}(\psi, a)$$
$$\mathrm{Sat}(\forall \mathsf{v}\, \varphi, a) \leftrightarrow \forall x\ \mathrm{Sat}(\varphi, a[x/\mathsf{v}]).$$

The additional axioms are *reflections on truth*. Because we are so familiar with Tarski's definition we think of these as 'obvious', but we made no assumptions on how we (or some other agent) managed to 'assert' statements such as φ, and for some modes of assertion Tarski's axioms may not be at all obvious.

(For a natural language example, note that natural language has many ways of expressing conjunction. For example 'X is a yellow brick' can be construed to mean 'X is a brick and X is yellow'. Now consider, 'Y is a red panda'. This could reasonably mean 'Y is a panda and Y is red', but it could also mean that 'Y is a member of the species *Ailurus fulgens*'. The equivalence of 'Y is a panda and Y is red' and 'Y is a member of the species *Ailurus fulgens*' might well be a nontrivial true reflection on the world—involving for example, verification that no-one ever took a panda of species *Ailuropoda melanoleuca* and painted it red.)

Much mileage has been made of the fact that PA with the above axioms ($\mathrm{PA} + \mathrm{Sat}$, the theory of a *full satisfaction class* Sat) is conservative over PA.

Such arguments do indeed support the *safety* of the above reflection, but do not, as far as I see, support the idea that these axioms express nothing about 'truth' (or satisfaction) that was not already apparent from the assertability notion we started from.

For the development of mathematics, however, other reflections are essential. Key amongst these are the addition of an induction scheme in the language with Sat. The two most familiar of these are Δ_0–PA(Sat) (predicative induction, or induction for bounded formulas) and PA(Sat) (full induction). Neither of these is conservative over PA + Sat.

4 Non-conservativity and Lachlan's theorem

Indeed, I see no significant reason to think that conservativity for $\mathscr{L}$ is important for the deflationist view but conservativity for $\mathscr{L}_{\infty,\omega}$ is not. (To be precise, I am not arguing against deflationism here, nor am I arguing that conservativity for $\mathscr{L}_{\infty,\omega}$ is important, only that focusing on conservativity for $\mathscr{L}$ as being a major virtue is a strange position to take.) Indeed Lachlan's theorem (already mentioned) shows that PA + Sat is not conservative over PA for $\mathscr{L}_{\infty,\omega}$, since the principle 'If this model is nonstandard then is is recursively saturated' is expressed by the scheme

$$\forall a \left(\bigwedge_{n \in \mathbb{N}} \exists x \bigwedge_{i<n} \theta_i(x,a) \;\rightarrow\; \forall y \, \exists x \bigwedge_{i \in \mathbb{N}} (i < y \rightarrow \theta_i(x,a)) \right)$$

where $(\theta_i)_{i\in\mathbb{N}}$ runs over recursive sequences of $\mathscr{L}$ formulas.

Lachlan's theorem is rather subtle and of general interest and I want to explore it is a bit more detail here.

First, the condition that the satisfaction class is *full* says that it determines truth for every formula with Gödel number in M and every possible assignment of the variables. In fact this is not necessary for Lachlan's result: it turns out that it suffices that Sat decides truth for a reasonable class of formulas containing all formulas of size up to a given nonstandard value $\alpha > \mathbb{N}$ and closed under subformulas. (Having made this point I will simplify the discussion and restrict to full satisfaction classes.)

Less well known than Lachlan's theorem is a somewhat more general result that gives other information about nonstandard formulas. Many published proofs of Lachlan's theorem actually show the following result.

Theorem 4.1. Let $M \vDash \mathrm{PA}+\mathrm{Sat}$ be nonstandard and $(\theta_i)_{i\in\mathbb{N}}$ a coded sequence of formulas in a single free variable. Then the subsets

$$P_i = \{x \in M : \mathrm{Sat}(\theta_i, [x])\}$$

for $i \in \mathbb{N}$ do not define a partition of M indexed by $i \in \mathbb{N}$.

This generalises slightly.

Corollary 4.2. Let $M \vDash \mathrm{PA}+\mathrm{Sat}$ be nonstandard and $(\theta_i)_{i\in\mathbb{N}}$ a coded sequence of formulas in a single free variable. Suppose that the subsets

$$P_i = \{x \in M : \mathrm{Sat}(\theta_i, [x])\}$$

for $i \in \mathbb{N}$ are all disjoint. Then all but finitely many of the P_i are empty.

Proof. Given the statements $\theta_i(x)$ we define by a recursion new statements $\theta_{i,k}(x)$ (for $k \leqslant i$) as follows,

$$\theta_{0,0}(x) := \theta_0(x) \lor (x = 0 \land \forall v\ \neg\theta_0(v))$$
$$\theta_{i,0}(x) := \theta_i(x) \land (x = 0 \to \exists v\ \theta_0(v))$$

and, for $i > k > 0$,

$$\theta_{k,k}(x) := \theta_{k,k-1}(x) \lor (x = k \land \forall v\ \neg\theta_{k,k-1}(v))$$
$$\theta_{i,k}(x) := \theta_{i,k-1}(x) \land (x = k \to \exists v\ \theta_{k,k-1}(v))$$

Thus, putting $Q_{i,j} = \{x : \mathrm{Sat}(\theta_{i,j}, x)\}$ we have

$$\begin{array}{ll} Q_{0,0} = P_0 \text{ if } P_0 \neq \emptyset, & Q_{0,0} = \{0\} \text{ otherwise} \\ Q_{i,0} = P_i \text{ if } Q_{0,0} \neq \emptyset, & Q_{i,0} = P_i \setminus \{0\} \text{ otherwise} \\ Q_{1,1} = Q_{1,0} \text{ if } Q_{1,0} \neq \emptyset, & Q_{1,1} = \{1\} \text{ otherwise} \\ Q_{i,1} = Q_{i,0} \text{ if } Q_{1,0} \neq \emptyset, & Q_{i,1} = Q_{i,0} \setminus \{1\} \text{ otherwise} \end{array}$$

and so on. This defines a coded partition $\{Q_{i,i} : i \in \mathbb{N}\}$ of M. (More formally, assuming cofinally many P_i are nonempty, we can show by induction on k that

$$Q_{0,0} \cup \ldots \cup Q_{k,k} \subseteq P_0 \cup \ldots \cup P_k \cup \{0, 1, \ldots, k\}$$

and

$$Q_{0,0}, \ldots, Q_{k,k}, Q_{k+1,k}, Q_{k+2,k} \cdots$$

is a disjoint collection of sets whose union is M.) □

Incidentally, the last corollary also shows that full PA is not needed in Lachlan's theorem: provided the arithmetic theory of M is strong enough to handle syntax (a theory such as $\mathrm{I}\Delta_0 + \exp$ is certainly sufficient), if M has a satisfaction class then it will be recursively saturated. Full PA is usually convenient and used (for example in *Models of Peano Arithmetic* [1]) to simplify the reduction of Lachlan's Theorem to Theorem 4.1.

Lachlan's theorem also yields an interesting conclusion that is analogous to an induction principle for Sat in $\mathscr{L}_{\mathrm{Sat}}$ even though none was given in axiomatisation for PA + Sat. More precisely, the conclusion is an overspill principle, i.e. a consequence for the nonstandard numbers of M when the premise of an induction holds for the standard natural numbers.

Corollary 4.3. Let $M \vDash \mathrm{PA} + \mathrm{Sat}$ be nonstandard, and $\theta(x)$ a possibly nonstandard formula in a single free variable. Then if $\mathrm{Sat}(\theta, [k])$ holds for all $k \in \mathbb{N}$ there is some $x > \mathbb{N}$ such that $\mathrm{Sat}(\theta, [x])$.

Proof. If not, when applied to Sat the formulas '$\neg\theta(x)$' and '$x = i$' (for $i \in \mathbb{N}$) would give a coded partition of M. □

This raises an interesting question that I not currently able to resolve.

Question 4.4. Let $M \vDash \mathrm{PA}$ be nonstandard and $(\theta_i)_i < \nu$ is a coded sequence of sentences, where ν is nonstandard. Suppose that $\mathrm{Sat}(\theta_n, [a])$ holds for all $n \in \mathbb{N}$. Does it follow that $\mathrm{Sat}(\theta_x, [a])$ for some nonstandard $x < \nu$?

5 A language with full conservativity[1]

Taking $\mathscr{L}$ as the usual language for arithmetic, note that each formula of $\mathscr{L}$ has a natural *rank*:

(a) $\mathrm{rank}(\theta) = 0$ if θ is atomic or negated-atomic;

(b) $\mathrm{rank}(\exists x\ \theta) = \mathrm{rank}(\forall x\ \theta) = 1 + \mathrm{rank}(\theta)$; and

(c) $\mathrm{rank}((\theta \wedge \varphi)) = \mathrm{rank}((\theta \vee \varphi)) = 1 + \max(\mathrm{rank}(\theta), \mathrm{rank}(\varphi))$.

The *stratified sub-language* $\mathscr{L}_{\mathrm{strat}}$ is the sub-language that permits $(\theta \wedge \varphi)$ and $(\theta \vee \varphi)$ only when θ, ψ have the same rank. A formula is stratified if it is in this language.

Theorem 5.1. Let $M \vDash \mathrm{PA}$ be countable. Then there is a predicate Sat on M satisfying Tarski's axioms for all stratified formulas.

The proof goes by showing that in a nonstandard model M, template logic or FA-logic can simulate all instances of the M-rule. This is the original style of argument for the Kotlarski–Krajewski–Lachlan theorem [2], but they required full recursive saturation to handle $\mathrm{Th}(M)$. The objective is to reduce this requirement to Σ_n-recursive saturation for all standard n. This requires bounding the complexity of formulas that arise.

A template for a finite set of nonstandard *stratified* formulas can always be found that is Σ_n for some n: leave all standard-depth formulas alone and approximate all nonstandard formulas to some fixed (standard) depth. This works because stratified nonstandard formulas only relate (e.g. in the sense that one contains another as a sub-formula) in a small number of ways to standard depth, and we can ignore such relations at nonstandard depth. Thus the depth we need approximate our finite set $\Sigma, \theta(v)$ of nonstandard formulas is some fixed n depending on the standard-depth formulas in it.

It follows (from cut-elimination) that in template proofs of $\mathrm{Th}(M) \vdash \Sigma, \theta(a)$ for a in M, assumptions from $\mathrm{Th}(M)$ of complexity higher than n can be eliminated, so we only need assumptions from Σ_n–$\mathrm{Th}(M)$. So if $\mathrm{Th}(M) \vdash \Sigma, \theta(a)$ for all a in M, there are finitely many templates doing this, by Σ_n-recursive saturation. These finitely-many templates can be combined to a single proof of $\mathrm{Th}(M) \vdash \Sigma, \forall v\ \theta(v)$.

References

[1] Richard Kaye. *Models of Peano arithmetic*, volume 15 of *Oxford Logic Guides*. The Clarendon Press Oxford University Press, New York, 1991. Oxford Science Publications.

[2] H. Kotlarski, S. Krajewski, and A. H. Lachlan. Construction of satisfaction classes for nonstandard models. *Canad. Math. Bull.*, 24(3):283–293, 1981.

[3] A. H. Lachlan. Full satisfaction classes and recursive saturation. *Canad. Math. Bull.*, 24(3):295–297, 1981.

[1] The main result here is not in a detailed form suitable for general dissemination but a sketch is presented here. A write-up of the argument is planned for the near future.