

Beth's Theorem and Deflationism

Timothy Bays

In 1999, Jeffrey Ketland published a paper which posed a series of technical problems for deflationary theories of truth. Ketland argued that deflationism is incompatible with standard mathematical formalizations of truth and that alternate deflationary formalizations are unable to explain some central uses of the truth predicate in mathematics. He also used Beth's definability theorem to argue that, contrary to deflationists' claims, the T-schema cannot provide an 'implicit definition' of truth. In this paper, I want to challenge this final argument. Whatever other faults deflationism may have, the T-schema *does* provide an implicit definition of the truth predicate. Or so, at any rate, I shall argue.¹

1 Notation and preliminaries

Let me start by setting out some context. Let $\mathcal{L} = \{0, 1, +, \times, <\}$ be the language of first-order arithmetic and let PA be the theory based on the first-order Peano axioms.² Our goal is to add a new truth predicate

¹The paper at issue here is Ketland 1999. For reasons of space, the present paper will focus fairly tightly on Ketland's discussion of implicit definition. I should note, however, that some of Ketland's other arguments have also sparked considerable discussion in the literature. The reader interested in this broader discussion is advised to start with the survey in Shapiro 2002. They should then turn to Field 1999, Tennant 2002, and Tennant 2005 for some responses to Ketland's arguments, and to Ketland 2005 for Ketland's replies.

²Some more-detailed remarks on notation are probably in order here. Throughout this paper, I will let $=, \neg, \&, \exists$ constitute the official first-order logical vocabulary, and I will treat $\vee, \rightarrow, \leftrightarrow,$ and \forall as abbreviations. (I will make free use of these abbreviations wherever they seem to improve readability.) Although the languages \mathcal{L} and \mathcal{L}^+ will officially contain only a small number of non-logical primitives, I will often abuse notation and write, e.g., ' $\phi \in \mathcal{L}$ ' to mean that ϕ is a formula built up from the primitives in \mathcal{L} (plus, of course, the logical connectives).

Throughout this paper, \mathbb{N} is the model which has the natural numbers as its domain and which interprets the symbols in \mathcal{L} in the ordinary manner. If \mathbb{M} is a model for \mathcal{L} and P is a new predicate, then $\langle \mathbb{M}; P \rangle$ is a model which results from expanding \mathbb{M} to the language $\mathcal{L} \cup \{P\}$ —that is, from choosing a new subset of \mathbb{M} to serve as the interpretation of P . Recall, here, that this kind of expansion does not add any new elements to our domain, and it does not change the satisfaction relation for the original language—so, if ϕ is a formula in \mathcal{L} , then $\mathbb{M} \models \phi \Leftrightarrow \langle \mathbb{M}; P \rangle \models \phi$. If m is an element of some model \mathbb{M} , I will write $\mathbb{M} \models \mathbb{T}[m]$ to mean that m lives in the subset of \mathbb{M} picked out by \mathbb{T} (similarly for $\mathbb{M} \models \phi[m]$ where $\phi(x)$ is a formula with only x free).

Finally, let ϕ be a formula in \mathcal{L} . Then $\ulcorner \phi \urcorner$ is the code of ϕ under some appropriate coding scheme, and $\dot{\ulcorner \phi \urcorner}$ is the term which represents this code in PA—that is, if the code of ϕ is n , then $\ulcorner \dot{\ulcorner \phi \urcorner} \urcorner = 0 + 1 + 1 + \dots + 1$ with n 1's. For convenience, I will assume that our coding scheme associates a sentence with every natural number. Readers who prefer alternate coding schemes should add the sentence $\forall n[\mathbb{T}(n) \rightarrow \mathbf{Sent}(n)]$ to each of the truth theories that we consider (where $\mathbf{Sent}(n)$ is a formula which numeralwise represents ' n is the code of a sentence' in PA).

to \mathcal{L} .³ In particular, we let $\mathcal{L}^+ = \mathcal{L} \cup \{\mathbb{T}\}$ where \mathbb{T} is a new unary predicate, and we then add some new axioms to PA which ensure that \mathbb{T} acts as a truth predicate for \mathcal{L} —that is, if $\langle \mathbb{N}; \mathbb{T} \rangle$ satisfies our new axioms, then the following biconditional holds for every sentence $\phi \in \mathcal{L}$:

$$\langle \mathbb{N}; \mathbb{T} \rangle \models \mathbb{T}(\ulcorner \phi \urcorner) \iff \mathbb{N} \models \phi.$$

Now, for our purposes, there two basic ways of going about this. First, we could simply add every instance of the \mathbb{T} -schema to PA. That is, for every sentence $\phi \in \mathcal{L}$, we could add the sentence

$$\mathbb{T}(\ulcorner \phi \urcorner) \leftrightarrow \phi$$

as a new axiom. Let us call the resulting theory PA_d . If we also expand PA's induction scheme to include formulas which contain our new predicate \mathbb{T} , then we will call the resulting theory PA_d^+ .⁴

Second, we could follow Tarski and give a full-fledged recursive definition of truth. This would involve adding some variant of the following axioms to PA:⁵

1. $\forall \phi [\mathbf{Atomic}(\phi) \rightarrow [\mathbb{T}(\phi) \leftrightarrow \mathbf{Tr}_0(\phi)]]$.
2. $\forall \phi [\mathbb{T}(\neg \phi) \leftrightarrow \neg \mathbb{T}(\phi)]$.
3. $\forall \phi \forall \psi [\mathbb{T}(\phi \& \psi) \leftrightarrow \mathbb{T}(\phi) \& \mathbb{T}(\psi)]$.
4. $\forall \phi \forall i [\mathbb{T}(\exists v_i \phi(v_i)) \leftrightarrow \exists n \mathbb{T}(\phi(\hat{n}))]$.

Let us call the resulting theory PA_T . If we also expand PA's induction scheme to include formulas which contain our new predicate \mathbb{T} , then we will call the resulting theory PA_T^+ .

³Ketland's original argument is formulated in somewhat more general terms, in that he works with a broader class of languages and considers base theories which extend PA. For our purposes, though, this level of generality is not necessary, so I will stick with PA itself. Moving to the general case would require only minor notational changes.

⁴I should note, here, that Ketland himself does not discuss the theory I'm calling PA_d^+ in Ketland 1999. That being said, several authors have discussed this theory in relation to Ketland's paper, and including it does not substantially complicate my argument.

⁵This way of formulating the axioms is perspicuous, but only because it suppresses a lot of the underlying coding apparatus. Technically, for instance, our language only lets us quantify over numbers, not formulas. So, each instance of ' $\forall \phi [\dots]$ ' in the above axioms really abbreviates a more-complicated expression of the form $\forall x [\mathbf{Sent}(x) \rightarrow [\dots]]$, where $\mathbf{Sent}(x)$ numeralwise represents 'x is the code of a sentence'. Similarly, when things are spelled out fully, we will have to use coding and numeralwise representability to capture the uses of negation, implication, quantification, and substitution which occur in axioms 2–4. So, modulo logical equivalence, axiom 2 really looks something like

$$\forall x \forall y [\mathbf{Sent}(x) \& \mathbf{Sent}(y) \& \mathbf{Neg}(y, x) \rightarrow [\mathbb{T}(y) \leftrightarrow \neg \mathbb{T}(x)]]$$

and axiom 4 looks like

$$\forall x \forall i \forall z [\mathbf{Formula}(x) \& \mathbf{Variable}(i) \& \mathbf{ExQuant}(z, i, x) \rightarrow [\mathbb{T}(z) \leftrightarrow \exists n \exists z' [\mathbf{Subst}(z', i, n, x) \& \mathbb{T}(z')]]]$$

where \mathbf{Neg} , $\mathbf{ExQuant}$, and \mathbf{Subst} numeralwise represent various negation, quantification, and substitution relations. (Note: in axiom 1, $\mathbf{Atomic}(x)$ represents 'x is an atomic sentence', and $\mathbf{Tr}_0(x)$ represents 'x is a true atomic sentence'. The latter can be already defined in PA—so, using a formula in \mathcal{L} —without appealing to the general notion of truth.)

Ketland’s main purpose in Ketland 1999 is to argue that Tarskian theories like PA_T^+ have real advantages over deflationary theories like PA_d and PA_d^+ . So, he highlights some positive features of PA_T^+ —it proves some general facts about truth, it helps to prove a semantic version of the incompleteness theorem, etc.—and some negative features of PA_d and PA_d^+ —they do not prove general facts about truth, they do not help to prove the incompleteness theorem, etc. It is in this later (negative) part of Ketland’s argument where his remarks on implicit definition occur, and so it is to this latter part of the argument that I turn next.

2 Ketland’s complaints

In section 4 of his paper, Ketland argues that PA_d and PA_d^+ have four negative features.⁶ First, he notes that these theories have non-standard models. In particular, let \mathbb{M} be a non-standard model of PA, and let m be any non-standard element of \mathbb{M} such that $\mathbb{M} \models \mathbf{Sent}[m]$. Then we can expand \mathbb{M} to a model for the language \mathcal{L}^+ such that $\langle \mathbb{M}; \mathbb{T} \rangle \models \text{PA}_d^+$ and $\langle \mathbb{M}; \mathbb{T} \rangle \models \mathbb{T}[m]$. This is true, even though there is no intuitive sense in which m codes up a true sentence of arithmetic. What is more, some of these non-standard models fail to satisfy even very basic principles concerning truth. So, for instance, we can build non-standard models of PA_d^+ which do not satisfy axiom 2 in the Tarskian truth definition—that is, $\langle \mathbb{M}; \mathbb{T} \rangle \not\models \forall \phi [\mathbb{T}(\neg\phi) \leftrightarrow \neg\mathbb{T}(\phi)]$.

Second, PA_d and PA_d^+ do not prove basic generalizations about truth. The example in the last paragraph shows that they do not prove the second of our Tarskian axioms. With a little work, we can extend this example to show that they do not prove *either* bivalence or non-contradiction—that is,⁷

- $\text{PA}_d^+ \not\models \forall \phi [\mathbb{T}(\phi) \vee \mathbb{T}(\neg\phi)]$.
- $\text{PA}_d^+ \not\models \forall \phi \neg[\mathbb{T}(\phi) \& \mathbb{T}(\neg\phi)]$.

Similar arguments show that PA_d and PA_d^+ do not prove basic facts about $\&$ and \exists .

Third, PA_d and PA_d^+ are ω -incomplete. Informally, we can note that although these theories do not imply axiom 2, they do imply all the relevant ‘instances’ of axiom 2—that is, for every $\phi \in \mathcal{L}$,

$$\text{PA}_d \vdash \mathbb{T}(\ulcorner \neg\phi \urcorner) \leftrightarrow \neg\mathbb{T}(\ulcorner \phi \urcorner).$$

More formally, although PA_d and PA_d^+ do not prove axiom 2, they do prove every sentence of the form

$$\forall y [\mathbf{Sent}(\dot{n}) \& \mathbf{Sent}(y) \& \mathbf{Neg}(\dot{n}, y) \rightarrow [\mathbb{T}(\dot{n}) \leftrightarrow \neg\mathbb{T}(y)]]$$

where n is a particular natural number.⁸ Hence, whether we think of ω -completeness in terms of *formulas* or in terms of *numbers*, PA_d and PA_d^+ are ω -incomplete.

⁶Although it is somewhat tangential to the main point of this paper, I should note that it is not clear just who these criticisms are really supposed to be directed against. Deflationists like Field tend to formulate their theories using substitutional quantifiers, and Ketland’s complaints do not obviously apply to such theories (see Field 1994, Field 1999, and Field 2001). In this sense, then, Ketland’s focus on PA_d and PA_d^+ may amount to attacking some straw men.

⁷As before, this way of formulating things suppresses a lot of the underlying coding apparatus. For convenience and perspicuity, I will continue to use this kind of notation/formulation throughout the remainder of this paper.

⁸Note that these sentences are simply instantiations of axiom 2 to particular natural numbers.

Finally, Ketland uses Beth's definability theorem to argue that PA_d and PA_d^+ do not provide an implicit definition of the truth predicate. To spell this argument out, I start by recalling the following definitions:

Definition 1: Let \mathbb{T} be a theory in some language \mathcal{L}^* , let P be a unary predicate not in \mathcal{L}^* , and let $\Phi(P)$ be a collection of sentences in $\mathcal{L}^* \cup \{P\}$. We say that $\mathbb{T} \cup \Phi(P)$ *implicitly defines* P , if for every \mathcal{L}^* -model \mathbb{M} such that $\mathbb{M} \models \mathbb{T}$, there is exactly *one* way to expand \mathbb{M} to $\mathcal{L}^* \cup \{P\}$ such that $\langle \mathbb{M}; P \rangle \models \mathbb{T} \cup \Phi(P)$.⁹

Definition 2: Let \mathbb{T} be a theory in some language \mathcal{L}^* , let P be a unary predicate not in \mathcal{L}^* , and let $\Phi(P)$ be a collection of sentences in $\mathcal{L}^* \cup \{P\}$. We say that $\mathbb{T} \cup \Phi(P)$ *explicitly defines* P if there is some $\psi(x) \in \mathcal{L}^*$ such that $\mathbb{T} \cup \Phi(P) \models \forall x [P(x) \leftrightarrow \psi(x)]$.

These definitions put us in a position to formulate Beth's definability theorem:

Theorem (Beth): Let \mathbb{T} be a theory in some language \mathcal{L}^* , let P be a predicate not in \mathcal{L}^* , and let $\Phi(P)$ be a collection of sentences in $\mathcal{L}^* \cup \{P\}$. Then $\mathbb{T} \cup \Phi(P)$ implicitly defines P if and only if $\mathbb{T} \cup \Phi(P)$ explicitly defines P .

Given all this, Ketland argues as follows. Suppose that PA_d or PA_d^+ implicitly defines \mathbb{T} . Then, by Beth's theorem, PA_d^+ must also *explicitly* define \mathbb{T} . Hence, there must be some $\psi(x) \in \mathcal{L}$ such that

$$\text{PA}_d^+ \vdash \forall x [\mathbb{T}(x) \leftrightarrow \psi(x)] \quad (1)$$

Further, since PA_d^+ contains the T-schema, we can use (1) to show that for every particular $\phi \in \mathcal{L}$,

$$\text{PA}_d^+ \vdash \psi(\ulcorner \phi \urcorner) \leftrightarrow \phi \quad (2)$$

Since we know that \mathbb{N} can be expanded to a model of PA_d^+ , (2) entails that

$$\mathbb{N} \models \psi(\ulcorner \phi \urcorner) \leftrightarrow \phi \quad (3)$$

for every $\phi \in \mathcal{L}$. This, however, makes ψ into a truth-predicate for \mathcal{L} , and that contradicts Tarski's theorem on the indefinability of truth.¹⁰ So, PA_d and PA_d^+ must not provide an implicit definition of \mathbb{T} after all.

⁹That is, there is exactly one subset of \mathbb{M} 's domain which, when chosen as the interpretation of P , will allow the model $\langle \mathbb{M}; P \rangle$ to satisfy $\mathbb{T} \cup \Phi(P)$. It is worth noting here that there is an alternate way of formulating the notion of implicit definition. Let P' be another new predicate, and let $\Phi(P')$ be the result of substituting P' for P throughout $\Phi(P)$. Then $\mathbb{T} \cup \Phi(P)$ implicitly defines P if

$$\mathbb{T} \cup \Phi(P) \cup \Phi(P') \models \forall x [P(x) \leftrightarrow P'(x)].$$

For some purposes, this alternate formulation is nicer than the one given above (though Ketland himself formulates the notion using expansions). I will say a bit more about the alternate definition when we get to section 3 (see especially fn. 13 on p. 5).

¹⁰Recall that Tarski's theorem says that there is no formula $\psi(x) \in \mathcal{L}$ such that for every $\phi \in \mathcal{L}$, $\mathbb{N} \models \psi(\ulcorner \phi \urcorner) \leftrightarrow \phi$. Note that this does not vitiate the overall project of using \mathbb{T} as a truth predicate, since $\mathbb{T}(x)$ does not live in \mathcal{L} . The trick in Ketland's proof comes in using the implicit definability of \mathbb{T} to generate a $\psi(x)$ which *does* live in \mathcal{L} and which perfectly 'matches' $\mathbb{T}(x)$. It is this ψ which generates a conflict with Tarski's theorem.

3 Tu quoque

Before turning to my main argument, I want to briefly highlight an initial oddity concerning the four arguments sketched in the last section. In particular, I will note that, when they are considered at a high enough level of abstraction, all four of Ketland’s claims about PA_d and PA_d^+ carry over nicely to PA_T and PA_T^+ . So, although Ketland may well have highlighted some negative features of PA_d and PA_d^+ , it is not clear that these features help to *distinguish* PA_d and PA_d^+ from PA_T and PA_T^+ .

First, a simple compactness argument shows that PA_T^+ has non-standard models and that, in any such model, $\langle \mathbb{M}; \mathbb{T} \rangle$, there will be some non-standard element m such that $\langle \mathbb{M}; \mathbb{T} \rangle \models \mathbb{T}[m]$.¹¹ Second, Gödelian considerations show that there are natural (and, indeed, true!) generalizations about truth which PA_T and PA_T^+ do not prove. So, for instance, let $\mathbf{Provable}^+(x)$ numeralwise represent the provability of \mathcal{L} -sentences in PA_T^+ .¹² Then, given the normal interpretation of \mathbb{T} on \mathbb{N} ,

$$\langle \mathbb{N}; \mathbb{T} \rangle \models \forall \phi [\mathbf{Provable}^+(\phi) \rightarrow \mathbb{T}(\phi)].$$

However, a simple application of Löb’s theorem shows that:

$$\text{PA}_T^+ \not\models \forall \phi [\mathbf{Provable}^+(\phi) \rightarrow \mathbb{T}(\phi)].$$

Third, PA_T and PA_T^+ are ω -incomplete. Let $\mathbf{Proof}^+(x, y)$ numeralwise represent the relation ‘ y is the code of a sentence in \mathcal{L} , and x is the code of a proof of that sentence in PA_T^+ .’ Then it is straightforward to show that, for every n ,

$$\text{PA}_T \vdash \neg \mathbf{Proof}^+(n, \ulcorner 1 \doteq 0 \urcorner).$$

But, by the second incompleteness theorem,

$$\text{PA}_T^+ \not\models \forall x \neg \mathbf{Proof}^+(x, \ulcorner 1 \doteq 0 \urcorner).$$

Finally, if we understand ‘implicit definition’ in the manner relevant to Beth’s theorem, then neither PA_T nor PA_T^+ provides an implicit definition of the truth predicate. This can be proved using *exactly* the same argument as we used to prove the corresponding claim for PA_d and PA_d^+ (just substitute PA_T^+ for PA_d^+ throughout the argument on page 4).¹³

Together, these points show that the simple development of Ketland’s arguments in section 4 of his paper—a development which simply runs through a series of awkward technical facts about PA_d and then

¹¹On this front, PA_d and PA_d^+ may seem to come out ahead of PA_T and PA_T^+ . If \mathbb{M} is a non-standard model of PA, then it is possible to expand \mathbb{M} to a model of PA_d^+ without making $\langle \mathbb{M}; \mathbb{T} \rangle \models \mathbb{T}[m]$ for any non-standard m . In contrast, *every* non-standard model of PA_T contains some non-standard m such that $\langle \mathbb{M}; \mathbb{T} \rangle \models \mathbb{T}[m]$. This latter claim follows from the fact that $\text{PA} \vdash \forall x \exists y [y > x \ \& \ \mathbf{Tr}_0(y)]$; hence, $\text{PA}_T \vdash \forall x \exists y [y > x \ \& \ \mathbb{T}(y)]$.

¹²So, for any natural number n , $\mathbb{N} \models \mathbf{Provable}^+(n)$ if and only if n is the code of some $\phi \in \mathcal{L}$ and $\text{PA}_T^+ \vdash \phi$. Note that the proofs at issue here may involve formulas in \mathcal{L}^+ , but the final sentence ϕ has to live in \mathcal{L} .

¹³In footnote 9, I gave an alternate formulation of implicit definition. Since it is equivalent to the definition given on page 4, it wo not help to evade the argument just given. Nonetheless it does suggest something interesting. Let \mathbb{T} and \mathbb{T}' be unary predicates, and let $\text{PA}(\mathbb{T})$ and $\text{PA}(\mathbb{T}')$ be full-fledged Tarskian truth theories—including induction—which are formulated in

leaves the matter at that—is not enough to distinguish PA_d and PA_d^+ from PA_T and PA_T^+ . I should note, here, that, although I think this point is technically interesting, it should not be overemphasized: there is room in principle for a more sophisticated argument which claims that (at least some of) Ketland’s points raise *deeper* problems for PA_d and PA_d^+ than for PA_T and PA_T^+ . Certainly the generalizations about truth which PA_d^+ fails to prove are more central to the theory of truth than those which PA_T^+ fails to prove, and the ω -incompleteness of PA_d^+ involves fundamental axioms of truth in a way in which the ω -incompleteness of PA_T^+ does not. (Indeed, one might even argue that Ketland’s second and third points about PA_d highlight a real weakness in that theory’s *account of truth*, while my analogous points about PA_T^+ simply reflect the underlying weakness of PA itself.) Finally, whereas deflationists sometimes claim that the T-schema serves to implicitly define the truth predicate, it is not clear that anyone has made a similar claim about the Tarskian axioms. Once again, therefore, there may be room for distinguishing the dialectical situation of the PA_d and PA_d^+ cases from that of the PA_T and PA_T^+ cases.¹⁴

4 Implicit definition

The above arguments show that neither PA_d^+ nor PA_T^+ provide an implicit definition of the truth predicate. How much should this fact bother us? In his paper, Ketland suggests that it should bother the deflationist quite a bit. After all, deflationists often claim that the T-schema *does* serve to implicitly define—or to ‘fix the extension of’—the truth predicate. On the surface, then, there seems to be a conflict between deflationists’ understanding of the T-schema and Ketland’s theorems on implicit definitions.

I think, however, that this particular conflict is mostly illusory and that it depends almost entirely on an equivocation concerning the phrase ‘implicit definition’. Beth’s theorem, after all, turns on a particularly strong reading of this phrase. For PA_d or PA_d^+ to ‘implicitly define’ \mathbb{T} in the sense of Beth’s theorem, they would have to fix the extension of \mathbb{T} *on all models of PA*, including all the non-standard models. Put

terms of these new predicates (so, $\text{PA}(\mathbb{T})$ is just a new notation for PA_d^+). Then Ketland’s argument shows that:

$$\text{PA}(\mathbb{T}) + \text{PA}(\mathbb{T}') \not\vdash \forall x [\mathbb{T}(x) \leftrightarrow \mathbb{T}'(x)].$$

However, if we expand our induction scheme so as to include formulas in the combined language, $\mathcal{L} \cup \{\mathbb{T}\} \cup \{\mathbb{T}'\}$, then we can use induction on the formula $\mathbb{T}(x) \leftrightarrow \mathbb{T}'(x)$ to show that:

$$\text{PA}(\mathbb{T}) + \text{PA}(\mathbb{T}') + \mathbb{I}(\mathbb{T} \cup \mathbb{T}') \vdash \forall x [\mathbb{T}(x) \leftrightarrow \mathbb{T}'(x)].$$

So, if we allow ourselves induction principles for a rich enough language, then we can prove that any two instances of the Tarskian truth definition are equivalent, and that is at least *similar* to the notion of implicit definition discussed in footnote 9. For more on this kind of argument, see Ketland’s remarks in Ketland 2003, pp. 9–10; see also McGee 1991, p. 73. For an interesting application of this type of implicit definition in the context of defining the logical particles, see Belnap 2006.

All that being said, this result still does not show that there is a unique way to expand an arbitrary model of PA to $\mathcal{L} \cup \{\mathbb{T}\} \cup \{\mathbb{T}'\}$ so that the resulting expansion satisfies $\text{PA}(\mathbb{T}) \cup \text{PA}(\mathbb{T}') \cup \mathbb{I}(\mathbb{T} \cup \mathbb{T}')$. It simply shows that any particular expansion will have to assign \mathbb{T} and \mathbb{T}' the *same* extension (although there may be—and, indeed, usually are—many ways of choosing that extension).

¹⁴That being said, people *have* claimed that the Tarskian axioms provide an *explicit* definition of truth, and this claim is also false on the conception of ‘explicit definition’ that is relevant to Beth’s Theorem.

otherwise, they would have to determine the application of \mathbb{T} , not just to every sentence in \mathcal{L} , but to every object that any model of PA *thinks* is a sentence in \mathcal{L} .¹⁵

It seems to me, however, that this is not what deflationists have in mind when they claim that the T-schema ‘fixes the extension’ of the truth predicate. As far as I can see, they are simply claiming that the T-schema fixes the application of \mathbb{T} to every *genuine* sentence in \mathcal{L} , and there is nothing in Ketland’s argument which tells against this more restricted claim. In fact, I think that this restricted claim is pretty clearly correct. Let me say two things about this.

First, if we limit ourselves to the standard model of PA, then PA_d does ‘fix the extension’ of \mathbb{T} . Let $\langle \mathbb{N}; \mathbb{T} \rangle \models \text{PA}_d$ and let n be the code of some sentence $\phi \in \mathcal{L}$. Then the fact that $\langle \mathbb{N}; \mathbb{T} \rangle$ satisfies the T-schema means that the number n will live *in* the extension of \mathbb{T} just in case $\mathbb{N} \models \phi$ (and will live *outside* the extension of \mathbb{T} just in case $\mathbb{N} \not\models \phi$). Since every sentence $\phi \in \mathcal{L}$ has a determinate truth value on \mathbb{N} , and since our coding scheme associates every natural number with a sentence, this will fix the extension of \mathbb{T} for every natural number.¹⁶ Hence, there is only one way to expand \mathbb{N} to the language \mathcal{L}^+ such that the resulting model satisfies PA_d . Further, this expansion ‘gives the right answer’ for all formulas of \mathcal{L} . Suppose, once again, that $\langle \mathbb{N}; \mathbb{T} \rangle \models \text{PA}_d$. Then for every $\phi \in \mathcal{L}$,

$$\langle \mathbb{N}; \mathbb{T} \rangle \models \mathbb{T}(\ulcorner \phi \urcorner) \iff \langle \mathbb{N}; \mathbb{T} \rangle \models \phi.$$

So, not only does PA_d fix the extension of \mathbb{T} on the standard model, but it does so correctly.

Second, if we turn our attention to non-standard models of PA, then we find that PA_d still uniquely defines \mathbb{T} *on the standard part* of those models. So, let \mathbb{N}' be a non-standard model of PA, and let \mathbb{N}'' and \mathbb{N}''' be expansions of \mathbb{N}' to \mathcal{L}^+ such that both expansions satisfy PA_d . Then for any natural number n ,

$$\mathbb{N}'' \models \mathbb{T}(n) \iff \mathbb{N}''' \models \mathbb{T}(n).$$

Further, these expansions continue to ‘give the right answer’ concerning \mathbb{T} ’s application to both ordinary natural numbers and to genuine sentences of \mathcal{L} . On the one hand, we know that for any number n ,

$$\mathbb{N}'' \models \mathbb{T}(n) \iff n \text{ is the code of an } \mathcal{L}\text{-sentence, } \phi, \text{ and } \mathbb{N}'' \models \phi.$$

¹⁵To further emphasize the strength of the notion of implicit definition that is in play in Beth’s theorem, we can examine that notion’s application to simple *recursive* definitions. Consider, for instance, the standard recursive definition of exponentiation:

$$\begin{aligned} x^0 &= 1 \\ x^{n+1} &= x^n \cdot x \end{aligned}$$

Although any mathematician would be happy with this definition—and, indeed, although this kind of recursive definition probably constitutes a paradigm case of ‘implicit definition’ in the mathematical context—it does not count as an ‘implicit definition’ for the purposes of Beth’s theorem (since it does not fix the interpretation of exponentiation on the non-standard parts of non-standard models of arithmetic). The fact that (even) this kind of recursive definition fails Beth’s test for being an ‘implicit definition’ highlights just how strong Beth’s conditions on implicit definition really are.

¹⁶If our coding scheme does not associate a sentence with every natural number, then the T-schema will only fix the extension of \mathbb{T} on those numbers which code sentences. But in this case, the axiom $\forall n[\mathbb{T}(n) \rightarrow \mathbf{Sent}(n)]$ (see fn. 2) will fix the extension of \mathbb{T} on those numbers which *do not* code sentences (i.e. by forcing those numbers to live *outside* the extension of \mathbb{T}).

On the other hand, we know that for any $\phi \in \mathcal{L}$,

$$\mathbb{N}'' \models \mathbb{T}(\ulcorner \phi \urcorner) \iff \mathbb{N}'' \models \phi.$$

So, even on non-standard models, PA_d will still fix the interpretation of \mathbb{T} on the standard parts of those models, and it will do so in exactly the way that we would want it to.

The upshot of these first two points is this: the *only* thing which keeps PA_d and PA_d^+ from providing implicit definitions of \mathbb{T} —in the sense of ‘implicit definition’ used in Beth’s theorem—is the fact that these theories do not fix the extension of \mathbb{T} on the non-standard parts of non-standard models of arithmetic. That is, if \mathbb{M} is a non-standard model of PA and m is a non-standard element of \mathbb{M} such that $\mathbb{M} \models \mathbf{Sent}[m]$, then PA_d and PA_d^+ do not fix the application of \mathbb{T} to m . But why should *this* bother the deflationist? After all, PA_d was supposed to fix the interpretation of \mathbb{T} on sentences of \mathcal{L} , and we already know that there is no intuitive sense in which non-standard elements like m code up \mathcal{L} -sentences. So, while the inability of PA_d and PA_d^+ to determine the value of \mathbb{T} on m may be a problem for Beth’s theorem, it is just not a problem for the deflationist’s conception of ‘implicit definition’.¹⁷

Let me make a purely textual comment about this matter. In claiming that deflationists regard the T-schema as an implicit definition of truth, Ketland cites remarks by Quine and Haack to the effect that the T-schema ‘fixes the extension’ of the truth predicate.¹⁸ A quick check, however, shows that both of these authors explain their remarks in ways that tell against Ketland’s invocation of Beth’s theorem. In particular, let $\text{TS}(\mathbb{T})$ and $\text{TS}(\mathbb{T}')$ be two versions of the T-schema, formulated using the predicates \mathbb{T} and \mathbb{T}' respectively. Then Quine and Haack both notice that, for every $\phi \in \mathcal{L}$:

$$\text{TS}(\mathbb{T}) \cup \text{TS}(\mathbb{T}') \vdash \mathbb{T}(\ulcorner \phi \urcorner) \leftrightarrow \mathbb{T}'(\ulcorner \phi \urcorner). \quad (4)$$

As we saw above, this claim is clearly true. Further, it underlies a perfectly natural conception of what it is for $\text{TS}(\mathbb{T})$ to ‘fix the extension’ of \mathbb{T} on \mathcal{L} . In this context, therefore, Ketland’s invocation of Beth’s theorem simply misses the point.

As far as I can tell, this problem generalizes pretty widely. Although I can find many deflationists who talk about the T-schema ‘implicitly defining’ or ‘fixing the extension of’ the truth predicate (the latter more

¹⁷From a technical perspective, this is really the key point. Ketland’s application of Beth’s theorem turns less on the details of our truth theory (see section 3) than on the fact that PA cannot pin down the appropriate extension for the $\mathbf{Sent}(x)$ predicate. This predicate *should* apply to all and only the codes of genuine \mathcal{L} -sentences, but in non-standard models it also applies to various non-standard elements (elements which do not, in any interesting sense, code sentences of \mathcal{L}). If there somehow *were* a way to correctly pin down the extension of $\mathbf{Sent}(x)$ across all models of PA, then theories like PA_d *would* implicitly define \mathbb{T} , and they would do so even by the standards of Beth’s Theorem.

An quick analogy might be helpful here. We can think of the T-schema as a recipe for determining the application of \mathbb{T} to particular sentences in \mathcal{L} (or, more formally, to the codes of those sentences). If a model insists that some odd, non-standard element in its domain is really a sentence code, then it is not the T-schema’s fault that it cannot determine the application of \mathbb{T} to that element. Similarly, a cake recipe tells you how to combine certain ingredients to make a chocolate cake. If you insist that talcum powder is really a special kind of sugar, then it is not your recipe’s fault when your cake turn out to be inedible.

¹⁸See Quine 1953, p. 136 and Haack 1978, p. 100. Ketland’s own remarks come on p. 84 of his paper.

often than the former), they all explain this usage by way of something like (4) above. To date, I cannot find *anyone* who actually argues that the T-schema implicitly defines truth in the strong sense of ‘implicit definition’ that is relevant to Beth’s theorem. Nor does Ketland’s own paper provide any evidence of such claims. Without such evidence, Ketland’s criticism of deflationism seems only to attack a straw man.

In the end, then, I just do not think that deflationists should be very troubled about the fact that PA_d and PA_d^+ do not implicitly define \mathbb{T} in the sense of ‘implicit definition’ that is used in Beth’s theorem. If you are only concerned with applying \mathbb{T} to standard codes of genuine formulas, then PA_d and PA_d^+ do a perfectly good job of fixing the extension of \mathbb{T} . The fact that they do not also define \mathbb{T} in some stronger sense—a sense which involves evaluating \mathbb{T} at objects which have no connection with the \mathcal{L} -sentences in which we are really interested—should not be too worrisome. Unless Ketland has an explicit argument connecting the deflationists’ modest use of ‘implicit definition’ with the strong sense of ‘implicit definition’ used in Beth’s theorem, his criticisms will simply wind up equivocating on two, very different, senses of this phrase.

This assessment reflects my overall reaction to the arguments in section 4 of Ketland’s paper. Even if we grant Ketland his mathematics, the philosophy in this section is unpersuasive. On the surface, all four of the criticisms that Ketland mounts against PA_d and PA_d^+ apply equally well to PA_T and PA_T^+ (with essentially identical arguments in the case of criticisms 1 and 4). So, it is not clear that these criticisms help to *distinguish* PA_d and PA_d^+ from PA_T and PA_T^+ . Further, although Ketland is certainly right that PA_d and PA_d^+ do not implicitly define \mathbb{T} in the sense of ‘implicit definition’ that is used in Beth’s theorem, this is not the sense of ‘implicit definition’ that is relevant to deflationism. On the deflationist’s conception of implicit definition—a conception which focuses solely on fixing the extension of \mathbb{T} on the class of genuine \mathcal{L} -sentences— PA_d and PA_d^+ do a perfectly good job of implicitly defining \mathbb{T} , and Beth’s theorem is not applicable. Or so, at any rate, I have argued.¹⁹

Department of Philosophy
University of Notre Dame
Notre Dame, IN 46556
USA
timothy.bays.5@nd.edu

TIMOTHY BAYS

References

Belnap, Nuel 2006: ‘Tonk, plonk and plink’. *Theoria*, 72, pp. 177–220.

Field, Hartry 1994: ‘Are our logical and mathematical concepts highly indeterminate’. *Midwest Studies in Philosophy*, 19, pp. 391–429.

¹⁹Let me return, here at the end, to a point that I made in footnote 1. For reasons of space, this paper has focused pretty tightly on section 4 of Ketland’s paper. I want to emphasize that this is only one of several sections in that paper, and that my views on the other sections are substantially more positive than my views on section 4 are. I strongly encourage the reader to give the rest of Ketland’s paper a good read-through.

- Field, Hartry 1999: 'Deflating the conservativeness argument'. *The Journal of Philosophy*, 96, pp. 533–40.
- Field, Hartry 2001: 'Postscript to "Deflationist views of meaning and content"', in his *Truth and the Absence of Fact*. Oxford: Clarendon Press 2001, pp. 141–56.
- Haack, Susan 1978: *Philosophy of Logics*. Cambridge: Cambridge University Press.
- Ketland, Jeffrey 1999: 'Deflationism and Tarski's Paradise'. *Mind*, 108, pp. 69–94.
- Ketland, Jeffrey 2003: 'On Wright's inductive definition of coherence truth for arithmetic'. *Analysis*, 63, pp. 6–15.
- Ketland, Jeffrey 2005: 'Deflationism and the Gödel phenomena: Reply to Tennant'. *Mind*, 114, pp. 75–88.
- McGee, Van 1991: *Truth, Vagueness, and Paradox*. Indianapolis: Hackett.
- Quine, Willard 1953: 'Notes on the theory of reference', in his *From a Logical Point of View*. Cambridge: Harvard University Press, pp. 130–8.
- Shapiro, Stewart 2002: 'Deflation and conservation', in Volker Halbach, ed., *Principles of Truth*. Frankfurt: Hansel-Hohenhausen, pp. 103–28.
- Tennant, Neil 2002: 'Deflationism and the Gödel phenomena'. *Mind*, 111, pp. 551–82.
- Tennant, Neil 2005: 'Deflationism and the Gödel phenomena: Reply to Ketland'. *Mind*, 114, pp. 89–96.