

Axiomatic Theories of Truth

By VOLKER HALBACH

CAMBRIDGE UNIVERSITY PRESS, 2011. IX + 364 PP. £50.00

The Tarskian Turn: Deflationism and Axiomatic Truth

By LEON HORSTEN

MIT PRESS, 2011. XII + 165 PP. £24.95

The 'Oxford Readings' volume *Truth* (edited by Simon Blackburn and Keith Simmons in 1999) surely gets it right when it organizes most of the papers reprinted there under the two headings 'Early minimalist theories' and then 'Modern minimalism and doubts about it'. The idea that the notion of truth is a thin, broadly logical, notion and not a metaphysically substantive one, has now been at the very centre of debates for at least a quarter of a century.

Flicking through the pages of that collection, however, you can't help but be struck by the almost total absence of symbols, which betokens the absence of any detailed discussion of formal theories of truth. That's odd, given that a central topic is indeed the idea that truth is in some sense a logical one. Remarkably, the editors don't even reprint Kripke's game-changing 'Outline of a Theory of Truth'.

Much of the work on metaphysically light-weight theories of truth post-Kripke has continued to be done in a formal key, for a very good reason. As Timothy Williamson has nicely put it,

One clear lesson [of the formal investigations] is that claims about truth need to be formulated with extreme precision, not out of kneejerk pedantry but because in practice correct general claims about truth often turn out to differ so subtly from provably incorrect claims that arguing in impressionistic terms is a hopelessly unreliable method.

We need to explore the regimentation and formalization of variations on core minimalist or deflationist ideas about truth if we are to understand properly their relative strengths and weaknesses. We need too to understand how they relate to more semantically driven proposals such as Kripke's, if we are to do better than make hopeful armwaving gestures at crucial choice-points. But then, what we can offer our more advanced students to supplement the Blackburn/Simmons collection and bring them up to speed on some of the relevant recent formal work on truth, especially if we want to do that without scaring off all but the most technically apt?

The Tarskian Turn aims to fill the gap, at least partially. It is short, very clearly and attractively written (often with a pleasingly light touch), and it nicely explains the content of some key formal theories and presents the key results about them, sometimes sketching proofs, but often just pointing to proofs in the literature for enthusiasts to follow up. One could argue a bit about matters of presentation here or there. I would have been inclined to give somewhat more generous explanations of the background material in logic and formal arithmetic, perhaps in an appendix (remembering how limited is the logical knowledge of so many philosophy students now). I'm not sure quite how well some of the proof sketches in the book work: in particular, the proof of Theorem 30 (the theory *TC* proves global reflection for *PA*) is opaque. And, as I'll remark below, there's a surprising gap in the presentation of the key theory *PKF*. But that's nit-picking. Overall, I think Leon Horsten has succeeded brilliantly at the task of outlining and so making accessible a range of recent work on theories of truth.

The book, however, is more than a merely expository essay. It is philosophically opinionated, takes sides, and even is prepared to endorse one particular axiomatic theory of truth as philosophically in good shape. This makes *The Tarskian Turn* engagingly provocative.

Here's the headline story. After an initial chapter 'About This Book', Chapter 2 on 'Axiomatic Theories of Truth' explains the project and urges that we should prefer an axiomatic theory to a semantic theory such as Kripke's (or the Revision Theory of Gupta, Belnap and Herzberger). Compare, though, the project of giving a theory of necessary truth: *there* it seems natural to suppose that semantic ideas about the import of the modal operator should take centre stage and will be needed to underpin our choices among different modal logics presented as axiomatic systems. Why should it be different for theories of plain truth?

Horsten suggests a couple of reasons for aiming straight for a theory of truth presented either axiomatically or in the form of rules of inference, with the theory perhaps thus implicitly defining truth in something like the way that inferentialists think that the logical rules fix the content of other logical operators. First, as philosophers we want to characterize truth-in-our-own-language in our own language (in this project we have no metalanguage to ascend to): but semantic theories of truth can't do this. Second, formal semantic theories presuppose set-sized domains, but the domain of discourse of our language is too big to be a set. Some readers will find neither consideration particularly compelling (being suspicious about what Horsten calls 'universalist ambitions', and resisting reading too much significance into artifacts of set-theoretic regimentations). But be that as it may. Let's agree that it is worth 'giv[ing] the axiomatic approach a try' to see where it gets us. The testbed for this approach in the rest of the book will be the project of extending first-order Peano Arithmetic *PA* with axioms/rules governing a truth-predicate (the arithmetic in our base theory enabling us to do syntax via coding in a well-understood way).

Chapter 3, ‘On the Shoulders of Giants’, reminds us about PA , Gödel-coding, the Diagonalization Lemma, the incompleteness theorems, and proves Tarski’s Undefinability Theorem, thus showing that the naive theory of truth which adds to PA all T -biconditionals in the language of PA augmented with a truth-predicate T is inconsistent. This sets us up for Chapter 4, ‘The Disquotational Theory’ which discusses the theory you get by augmenting PA with a predicate T allowed now in induction axioms – call that PA^T – and then adding the T -biconditionals for sentences in the *original* unaugmented language of PA . It is shown that this axiomatic theory DT is consistent and has nice models.

Chapter 5 is a philosophical interlude on ‘Deflationism’, before we turn to more formal details in Chapter 6, ‘The Compositional Theory’. One snag with DT is it that doesn’t allow us to prove even the simplest general facts about how the truth-predicate interacts with the logical operators. For example, following Horsten’s helpfully sloppy notation, although DT trivially proves case by case that $T(\neg\varphi) \leftrightarrow \neg T(\varphi)$ for any PA -sentence φ , it can’t prove the generalization $\forall\varphi(T(\neg\varphi) \leftrightarrow \neg T(\varphi))$. This seems an evident shortcoming, so what to do? A natural proposal is to follow Tarski and go compositional: so let TC be the theory which adds to PA^T the T -biconditionals for atomic sentences of PA (i.e. for equations between closed terms) and then adds as axioms the clause $\forall\varphi(T(\neg\varphi) \leftrightarrow \neg T(\varphi))$ where the quantifier runs over sentences of PA , together with similar clauses for conjunction and the universal quantifier. Then TC can prove all the T -biconditionals DT could prove but also proves lots of desirable general claims about truth. However, while DT is an arithmetically conservative expansion of pure PA , TC enables us to prove new truths such as the arithmetized consistency sentence for PA .

That familiar result usually strikes students as a quite a surprise when they first encounter it. *Why* should the move from DT to TC , which looks to be just a matter of getting the truth-predicate to interact properly with *logical* operators, lead to new *arithmetical* results? It would have been nice if Horsten had paused to try to make this formal result seem less puzzling, but he moves on quickly to Chapter 7, ‘Conservativeness and Deflationism’, where he considers whether exhibiting such arithmetical non-conservativeness means that a theory is no longer really in the spirit of a minimalist or deflationist approach to truth. The discussion at this stage, though, is thin and inconclusive. As Horsten notes, though, we can precisely calibrate the arithmetical strength of TC : it proves the same arithmetical sentences as the predicative second-order theory ACA . He returns to consider the significance of this at the end of the book.

DT and TC are restricted theories in that they don’t say anything about iterated uses of the truth-predicate: they can’t even prove e.g. $T(T(0 = 0))$. To remove this limitation we can go in two directions: we can introduce a hierarchy of typed truth-predicates, each with their own axioms. Or, surely better if we are trying to regiment something akin to our pre-theoretic notion of truth, we can aim for an untyped theory which allows iteration. Chapter 8, ‘Maximizing Classical Compositionality’, considers the theory you get by (i) now allowing

the compositional truth-axioms of TC to run over *all* sentences of the language of PAT (including sentences involving the predicate T), and also (ii) adding a pair of rules-of-proof, that allow us to infer $T(\varphi)$ from a proof of φ and conversely allow us to infer φ from a proof of $T(\varphi)$. This theory FS is equivalent to a theory proposed by Friedman and Sheard, and Horsten goes on to show that it has some pleasing features. It also turns out (perhaps surprisingly) to have a close relation to the Revision Theory of truth. However, as Horsten also shows (following McGee) FS is omega-inconsistent – which in almost anyone’s book is Bad News. We are going to have to box more cleverly, then, if we are going to get an untyped compositional theory of truth allowing iteration of the truth-predicate.

So Chapter 9 moves on to consider ‘Kripke’s Theory Axiomatized’, first outlining Kripke’s semantic theory, and then Feferman’s theory KF which ‘can be viewed as an attempt to axiomatically describe the construction of Kripke’s fixed point models’. This theory too has nice features, but it also seems to rule itself out of serious contention as our finally preferred theory of truth since it proves $L \wedge \neg T(L)$, where L is the liar sentence. So KF overshoots as an attempt to axiomatize the truths about truth generated in Kripke’s theory: it closes too many truth-value-gaps. The trouble is that KF is formulated in a classical logic which forces gap-closing. A natural suggestion is to reformulate the theory in a gappy logic. So Horsten now adopts the strong Kleene version of partial predicate logic (in an odd expositional glitch, although he has previously told us about the Kleene evaluation scheme, Horsten doesn’t pause to present a corresponding axiomatic or natural deduction framework, even though this will be unfamiliar to most students reading the book). We are then introduced to a theory of truth dubbed PKF . This adds the rule-counterparts of the truth axioms of KF – these are basically the rules that allow us to move in both directions between $T(\neg\varphi)$ and $\neg T(\varphi)$ and similarly allow us to commute the T -predicate with other logical operators – and the theory also allows us to move in both directions between $T(\varphi)$ and $T(T(\varphi))$.

PKF ’s T -rules are all highly intuitive and the theory again has nice features. Isn’t working with a gappy logic rather horrid in practice? Well, everything remains classical for T -free sentences, and classical PA can proceed undisturbed. So it is only some sentences involving T for which PKF ’s non-classicality really matters and where e.g. excluded middle fails. And, as Horsten goes on to point out in the final Chapter 10 ‘Truth and Philosophy’, it isn’t that PKF rejects excluded middle in such cases, but rather it is silent. But is silence really what we want from a theory here? Horsten cheerfully says

The system PKF ... is not vulnerable to a strengthened liar attack because it makes no claim concerning the truth value of the liar sentence. PKF simply does not assert the liar sentence, nor its negation, nor that it is true, nor that it is not true.

Indeed. But now the theory is out there, on the table for all to see, can't we as philosophers stand back and reflect on it, and forcefully raise the question of the truth or otherwise of claims on which *PKF* fails to give a verdict? If we are philosophers with the universalist ambitions that Horsten hasn't eschewed, isn't that exactly what we'll want to do? And off we go again ...

Horsten himself is remarkably silent on this question. Instead he concludes the book by considering whether *PKF* is at least consonant with a broadly deflationist or minimalist stance. He thinks it is. For the theory treats truth as an insubstantial property without 'a fixed nature or essence' in the sense that there is no more to truth than is grasped in grasping some inference rules (though Horsten doesn't rule out there being inference rules beyond those codified in *PKF*). But what about the fact that, like *TC*, the compositional theory *PKF* is also non-conservative over arithmetic? Indeed *PKF* is arithmetically as strong as a transfinitely ramified system of predicative analysis that goes by the label ACA_{ω^ω} , whose first-order arithmetical consequences go far beyond those of *PA*.

Horsten is remarkably unworried. I think he is too swayed by a (quoted) claim of Feferman's that suggests that systems of predicative analysis only elaborate commitments that are already implicit in accepting *PA*, a suggestion that runs clean against Isaacson's well-known thesis that *PA* marks the natural boundary of those truths that can be reached by purely arithmetical reasoning. We can't examine who is right about that here. But the complaint is that neither does Horsten: he fails to acknowledge just how contentious it is to suppose that the progression through systems of predicative analysis stronger than the arithmetically conservative system ACA_0 can somehow be regarded as insubstantial rather as involving new infinitary ideas (and so it remains equally contentious to suppose that a theory of truth arithmetically equivalent to a strong system of analysis can still count as deflationary).

We might dissent at the end of the book, then, about Horsten's philosophical assessment of the merits of *PKF*. But we should still be very grateful for a beautifully structured guided tour, with thought-provoking commentary, making some recent formal work on truth accessible to a wide student audience interested in the truth about truth (and accessible as well as to non-expert colleagues who want to know what the logicians down the corridor have been up to).

Even if you want to run a higher level course for logicians concentrating on recent formal theories of truth, I'd strongly recommend shaping it by starting from Horsten's admirably clear overview of (some of) the lie of the land to provide an initial guide. Then to flesh out some details and examine in depth some of the proofs which Horsten gestures to, you can turn to Volker Halbach's *Axiomatic Theories of Truth*. This much longer, more expansive, more technically demanding, book is structured very similarly. Part I is a longer introduction advertising the merits of axiomatic theories and providing some formal background. Part II discusses typed theories like *DT* and *CT* and their hierarchical extensions. The much longer Part III treats untyped theories like *FS*, *KF*, etc. A final Part IV

offers some philosophical reflections. The theories which Horsten discusses (I didn't mention them all) therefore form a kind of backbone to Halbach's book, but there are more variations and a lot more detail. Many of the proofs presented here for the first time in book form are due to Halbach himself (either solo or in joint papers with Horsten): specialists will be very grateful to have his scattered work brought together and organized in this way.

I'm not so sure, however, whether this is as successful a book of its kind as Horsten's is of *its* kind. Halbach doesn't seem to write with a consistent idea of an intended audience in mind. Relatively straightforward philosophical discussion is mixed in with relatively high-level technical arguments. By p. 68, for example, we are launched into a fourteen page hard-core cut-elimination proof that a certain theory is arithmetically conservative (not one of the initial backbone theories either). Most readers of this journal interested in truth – aren't we all? – will want to pick and choose their route through the book *very* selectively, though that will be fairly easy to do if you have already read Horsten's shorter guide through the territory (or indeed, if you have read Halbach's own *Stanford Encyclopedia* essay on 'Axiomatic Theories of Truth').

Technically, Part III is entertaining enough as it explores some fifteen different theories of truth, if you like that sort of thing. But after struggling through the details it is difficult not to end up suspecting that, having given the type-free axiomatic approach with universalist ambitions a try, maybe this isn't after all what we need to soothe our philosophical troubles about truth. But what is Halbach's assessment? And in particular, how does he judge the theory *PKF* which Horsten proposes as the current best buy?

Well, Halbach clearly isn't happy with *PKF*. Partly this is because of a general resistance to the idea of departing from classical logic if that can be avoided. And partly this is because of technical observations about the mathematical limitations of *PKF* (it can't do much transfinite induction on open wffs involving the *T*-predicate). I don't see why Horsten need be moved by *these* considerations, however. Logical regimentation always involves trade-offs of costs and benefits: and Horsten will say that given pure arithmetic remains classical, the cost of officially having to go non-classical in the face of some cases that involve troublesome uses of the *T*-predicate and thereby avoiding paradox is a price worth paying. And if that leads us to restrict transfinite induction in cases of no ordinary mathematical interest, why care?

But Halbach's more general assessment of the import of the technical investigations he reports is less clear. Indeed, the final philosophical chapters strike this reader, at any rate, as disappointingly unsatisfying.

Peter Smith

University of Cambridge

Sidgwick Avenue, Cambridge CB3 9DA

ps218@cam.ac.uk

