Book Review | *The Tarskian Turn: Deflationism and Axiomatic Truth*

Matthew McKeon

Published online: 1 August 2012
© Matthew McKeon 2012

*The Tarskian Turn: Deflationism and Axiomatic Truth;* Leon Horsten; Cambridge MA: The MIT Press, 2011; 165 pages; $35 hardcover; 978-0262-01586-5.

“*The Tarskian turn*” refers to Tarski’s articulation of a theory of truth that is free from traditional philosophy. Rather than giving a philosophical account of the essence of truth, Tarski in his famous paper, “The Concept of Truth in Formalized Languages,” offers a formal theory of truth that highlights the disquotational function of the truth predicate (e.g., ‘snow is white’ is true if and only if (iff) snow is white). According to Tarski, a theory of truth for a language L is adequate iff (i) it is consistent and (ii) it entails for every truth apt sentence α of L its T-sentence: ‘α’ is true iff α (here and in what follows I play loose with use/mention conventions). Although Tarski favored the correspondence theory of truth, he did not want to burden a theory of truth for a language that satisfies (i) and (ii) with a defense and explication of a philosophical account of truth. Tarski’s “turn” away from traditional philosophical explications of truth sets the rhythm of Horsten’s book.

Tarski’s approach to defining truth has been immensely influential and has spawned a lot of technical research by logicians and mathematicians on truth. Although truth may be most naturally investigated in formal settings, it takes some work to see how the contours of a philosophical conception of truth can emerge from such an investigation. This book serves as a counterbalance to Tarski’s turn: it discusses some of the main axiomatic theories of truth with the aim of establishing their relevance to a philosophical understanding of the concept of truth. In particular, as I shall outline below, Horsten presents a series of axiomatic theories of increasing strength that informs his development of inferential deflationism, his favored version of deflationism.

Corresponding Author: Matthew McKeon
Michigan State University
email – mckeonm@msu.edu
The book is well-written. The presentations of the axiomatic theories of truth are clear and concise, as is the connected discussions of deflationism. The book is relevant to contemporary thinking about truth: both the axiomatic approach to characterizing truth and deflationism are in vogue nowadays. The book is useful for non-specialists with intermediate logic under their belts who want a refresher or quick tutorial on main axiomatic theories of truth and their relation to deflationism. Since the technical details discussed by Horsten are not new, I skip most of them and gloss over those I that mention. In what follows, I outline each Chapter and then conclude.

Chapter 1
This Chapter outlines the aims of the book and its structure.

Chapter 2
Horsten reviews considerations that favor an axiomatic approach to characterizing truth over a model-theoretic or substantial approach. Substantial theories of truth (e.g., correspondence, coherence, pragmatic) treat truth as a substantial notion and attempt to uncover its essence. The central tenet of each is generated by filling in the blank of,

\[ \text{a truth-bearer } \alpha \text{ is true if and only if } \_\_\_\_\_\_ \] ,

with a description of the condition that \( \alpha \) must satisfy in order to be true. Historically, the problems with the proffered descriptions are that they lack precision, are circular, or that they fail to reflect our actual uses of the concept of truth.

A model-theoretic characterization of truth for a given formal language \( L \) defines truth in the meta-language for \( L \) in two steps. First, a model for \( L \) is defined. Second, an intended model \( M \) for \( L \) is described in the meta-language for \( L \). An \( L \)-sentence is true iff it is true in \( M \). Of course, the truth of meta-language sentences requires a formalization of the meta-language and a distinct characterization of truth for the meta-language in the meta-meta-language.

According to Horsten, the model-theoretic approach to characterizing truth is inadequate because any such characterization is essentially relative to a formal language. It is desirable, however, that a theory of truth characterizes truth for ordinary languages such as English. Satisfying this desideratum seems to require that a truth theory for a language include the
language in which the truth theory is expressed. Since no model theory of truth is formulated in the object language, the model-theoretic approach is not promising as a means of satisfying this desideratum. Furthermore, the universe of discourse of English is not a set and so there can be no intended model for English since the domains of models are sets.

In sum, the axiomatic approach is preferable to the model-theoretic and substantial approaches to characterizing truth. An axiomatic theory of truth of sufficient strength promises precision, reflects the actual uses of the truth predicate, and can characterize (at least partially) the truth predicate of the language in which it is expressed. Horsten will identify what he takes to be such a theory and argue that it reflects inferential deflationism, which he regards as a plausible deflationist view of truth.

Chapter 3

Horsten presents Peano Arithmetic, which serves as the background theory for the axiomatic theories of truth that he offers later. These theories of truth are comprised of the Peano axioms and individuated by various sets of axioms or inference rules for the truth predicate. Horsten sketches proofs of several meta-logical results that bear on later discussions of axiomatic truth theories. For example, from Gödel’s theorems it follows that no consistent axiomatic truth theory can prove its own consistency. Using this result allows us to compare the comparative strength of truth theories. For truth theories S and S’, if S proves that S’ is consistent, then S is stronger than S’ (i.e., there is a theorem of S that isn’t a theorem of S’). Tarski’s theorem on the undefinability of truth entails that the property of being an arithmetical truth is undefinable in the language of Peano arithmetic (L_{PA}). If we extend L_{PA} by adding the truth predicate (generating the language L_T) and let PA^T be Peano arithmetic formulated in L_T (the truth predicate can occur in instances of the induction scheme), then it can be established employing Tarski’s theorem that no consistent extension of PA^T proves all the T-sentences for L_T. So, no axiomatic theory of truth that extends PA^T proves all T-sentences for L_T, and thus none fully captures the meaning of the truth-predicate for L_T.

Chapter 4

Consider the toy arithmetical language L which consists of just the two sentences ‘(0=0)’ and ‘(1=0).’ The axioms of the Disquotational theory of truth (DT) for L express all the T-sentences for L: ‘(0=0)’ is true iff (0=0) and ‘(1=0)’ is true iff (1=0). If we extend L by adding a truth predicate ‘___ is true,’ then we add ‘(0=0)’ is true’ is true iff ‘(0=0)’ is true,
and ‘(1=0)’ is true’ is true iff ‘(1=0)’ is true. Actually, given that the truth predicate is iterable, its addition to L requires infinitely many truth axioms.

Chapter 4 describes DT for $L_{PA}$. By the familiar technique of Gödel numbering, sketched in Chapter 3, this language contains a sentence S that can be read as saying that it is false. S says ‘S is false.’ Given classical logic, ‘$S$ is false’ is true iff $S$ is false entails a contradiction. The axioms of DT, which are formulated in $L_T$, are the axioms of $P A^T$ and the T-sentences for $L_{PA}$. So, following Tarski’s strategy for avoiding inconsistency, the T-sentence axioms of DT are restricted to those for the object language: if ‘$\alpha$’ is true iff $\alpha$ is an axiom of DT, then ‘$\alpha$’ is an $L_{PA}$ sentence and does not contain an occurrence of the truth predicate. There are two negative consequences of taking the restricted T-sentences as axioms of a theory of truth for $L_{PA}$. First, DT is unable to prove all truths about the truth predicate. For example, from ‘(0=0)’ and its T-sentence we may derive in DT that ‘(0=0)’ is true. Intuitively, from this it follows that ‘(0=0)’ is true, but we can’t derive this in DT. This reflects the fact that DT is not a theory of truth for the language it is formulated in and so does not do justice to the self-reflexive nature of truth. Furthermore, DT does not do justice to the compositional nature of truth. For example, we cannot prove that for all $L_{PA}$-sentences $\alpha$, $\beta$: $(\alpha$ is true and $\beta$ is true) iff $(\alpha$ and $\beta$) is true, although each instance of this generalization can be proven in DT. This second defect of DT motivates the stronger theory of truth presented in Chapter 6.

Chapter 5

The primary aim of much of the philosophical investigation of truth in western philosophy from Plato onwards has been to answer the question: what do all truth bearers that are true share that is lacked by all ones that are untrue? The deflationist demurs from answering such a question, because the deflationist thinks that there is nothing to say about truth in general. At minimum, a deflationary view of truth denies that there is a nature of truth in general. With an eye on DT and in anticipation of the axiomatic theories later presented, Horsten sketches his account of the aim of a philosophical investigation of truth that is in sync with deflationism.

Horsten asserts that a deflationist theory of truth consists of three related parts: (i) an account of the meaning of the concept of truth (pp. 61-2)), (ii) a description of the role that the concept of truth can validly perform (pp.63-4), and (iii) a description of the kind of concept truth really is (pp.65-6). Horsten’s view is that (ii) at least partially yields (i). This is discussed in Chapter 10. (iii) figures later in Chapters 7 and 10. Here I briefly elaborate on (ii) and (iii).
(ii): the truth predicate extends the expressive resources of language (e.g., blind truth descriptions such as ‘What Obama said this morning about Romney is true,’ truth-generalizations such as ‘every instance of $P$ or $\sim P$ is true’).

(iii): truth is a logico-linguistic notion. The truth predicate allows us to formulate logical principles like the above one. It is also a linguistic notion; the bearers of truth are linguistic entities (i.e., interpreted syntactic objects). Given the arithmetization of syntax, by virtue of being partly a linguistic notion truth is also partly a mathematical notion.

**Chapter 6**

Chapter 6 introduces the compositional theory of truth (TC), which like DT is an axiomatic theory of truth for $L_{PA}$ formulated in $L_T$. The axioms of TC are the axioms of $PA^T$ plus principles of the compositionality of truth restricted to the sentences of $L_{PA}$ (e.g., for all such sentences $\alpha,\beta$ (True ($\sim \alpha$) iff $\sim$True($\alpha$), True($\alpha \& \beta$) iff (True ($\alpha$) & True ($\beta$))). In order to express these principles, the axioms of TC, unlike those of DT, universally quantify over terms and formulae from $L_{PA}$.

The theorems of TC include all those of DT plus those whose proofs call upon principles of the compositionality of truth. Like DT, TC does not do justice to the self-reflexivity of truth (e.g., we cannot derive in TC ‘True(True(0=0))’ from ‘True(0=0)’). This is a consequence of following the Tarskian strategy for avoiding paradox: a truth theory does not characterize truth for the language that it is formulated in.

As reviewed in Chapter 3, Gödel’s second incompleteness results show that Peano Arithmetic (PA) cannot prove its own consistency (on the assumption that PA is consistent). So, *it is false that* $(0=1)$ *is a theorem of PA* cannot be derived from the PA axioms (plus first-order logic). Horsten sketches the reasoning that establishes that the consistency of PA can be proved in TC. So, from the principles of the compositionality of truth plus the axioms of PA, it follows that *it is false that* $(0=1)$ *is a theorem of PA*. This is summarized by saying that TC is not conservative over PA. DT, however, is conservative over PA. The worry here is that the mere addition of principles concerning the compositionality of truth to PA increases the mathematical strength of PA. Whether a sentence from $L_{PA}$ is a theorem of PA or not is an arithmetical fact. From PA alone it cannot be proven that $(0=1)$ is not a theorem of PA, but this is provable in TC.

It seems like a deflationist won’t be happy with TC. After all, according to the deflationist truth is an insubstantial notion. Following Blackburn and Simmons (*Truth*. Oxford: OUP,
1999 p. 4), one way of spelling this out is as follows. Consider the two truths (i) \( \neg(0=1) \) and (ii) \( \neg(0=1) \) is true. The deflationist is committed to believing that (ii) says nothing more than (i). This doesn’t mean that the truth predicate does no expressive work and is eliminable; it allows us to say things that can’t be said without it (see above). That (i) and (ii) say the same thing suggests that we can’t learn anything about (i) that essentially depends on a theory about the functioning of the truth-predicate in (ii). Since TC makes possible the proof that (i) is a theorem of PA, TC seems to be a theory of truth that a deflationist should reject, opting for a weaker one such as DT that is conservative over arithmetic. Horsten resists this, and aims to configure deflationism so that by its lights TC is philosophically sound even if philosophically incomplete (recall that TC doesn’t do justice to the self-reflexivity of truth). This raises the question: what is essential to deflationism? Horsten confronts this question in Chapter 10.

Chapter 7

The two central claims of Chapter 7 are: (i) deflationism is compatible with truth not being conservative over arithmetic (p. 93) and (ii) deflationism should not commit itself to the claim that truth plays no explanatory role in specific philosophical disciplines (p.92).

Horsten’s arguments for (i) and (ii) could be more clearly presented. As best as I can make out they are as follows. Note well: I am drawing on more than is explicit in Horsten’s book.

(i) What is essential to deflationism is the thesis that the characterization of the functions of the truth predicate is conceptually and explanatorily basic. This characterization does not follow from and is not explained by a definition of truth that relates it to more “basic” concepts in terms of which ‘true’ can be defined (here I borrow from Armour-Garb and Beall, *Deflationary Truth*. Chicago: Open Court, 2005 p.3ff). Truth is an inferential tool. This function of truth is compatible with truth not being conservative over arithmetic. As sketched in Chapter 6, “the notion of truth allows us to infer “new” mathematical propositions: consistency statements, for example” (p. 93). As long as the inferential properties of truth are taken as basic, Deflationism can live with truth not being conservative over arithmetic. Acknowledging the claim made in Chapter 5 that truth is partially a mathematical notion, lessens the surprise that it isn’t conservative over arithmetic.

(ii) Deflationism should not commit itself to the claim that truth plays no explanatory role in specific philosophical disciplines such as epistemology and semantics, because it is an open question whether truth does play such a role.
The concern arises that (i) and (ii) overly inflate deflationary truth so that it is no longer really deflationary. If truth may not be conservative over a background theory such as arithmetic and if it may play an explanatory role in philosophy, then deflationary truth starts looking substantial. Horsten tells the reader (p. 93) that he will add new content to the deflationist view of truth in Chapter 10. The challenge is to allay the concern that the added content will result in a concept of truth that is not meaningfully deflationary.

**Chapters 8 and 9**

Chapters 8 and 9 investigate type-free axiomatic truth theories that extend TC in the sense that they do justice to the compositional nature of truth and to the self-reflexivity of truth. Theories of truth are untyped if they prove sentences of the form ‘True(True(α))’ as opposed to merely proving ‘Truex(Truey(α))’ where ‘Truex’ and ‘Truey’ are different types of truth. Horsten sketches formal details for the theories that he discusses, some of which show that they do not also prove True(L) iff L, where L is the liar sentence (i.e., a sentence roughly equivalent with “This sentence is false”).

Chapter 9 ends with Horsten’s presentation of PKF (Partial Kripke-Feferman truth theory), a type-free axiomatic theory that is a formalization in partial logic of a close version of Kripke’s theory of truth. PKF is defined for L∀ and so it is a theory of truth for the language it is expressed in. Actually, PKF is a collection of inference rules (e.g., from True(True(α)) infer True(α) and from True~(α) infer ~True (α)), and so not an axiomatic theory of truth even though it is a proof-theoretic one. Besides the inference rules, PKF also includes PA T minus the induction axiom. The induction axiom is reformulated as an inference rule which reflects the partial logic that PKF uses. Kripke treats the liar sentence as neither true nor false, but since he formulates his theory of truth in classical logic it is vulnerable to the paradox that arises from the strengthened liar sentence, ‘This sentence is not true.’ Kripke’s theory seems to evaluate this sentence as both true and not true. PKF is formulated in the strong Kleene version of partial logic. Accordingly, if α is truth-value less, then so too is ~α, ~α v α, and ~(α & ~α). The strengthened liar is neither true nor not true in PKF, which Horsten takes to be an advance on Kripke’s theory. Like some other truth theories, PKF makes the truth predicate partial: it does not apply universally to all meaningful sentences (e.g., the liar sentence). To say that a predicate P doesn’t apply in this sense to an object δ is to say at least that δ is neither in the extension of P nor in its anti-extension (the collection of things that are non-P). In PKF, the liar and strengthened liar L∀-sentences are neither true nor not-true.
Chapter 10

The central claim of Chapter 10 is that PKF harmonizes well with a deflationist stance (p.143). Drawing on parts of earlier Chapters, Horsten argues in support of this claim for a new version of deflationism which he calls “inferential deflationism.” Much of the content of Chapter 10 devoted to defending this claim is in Horsten’s 2009 paper “Levity” *Mind* vol. 118 pp. 555-81.

According to Horsten, to explicate the meaning of the truth predicate, one must explain how it functions. According to inferential deflationism, to explain this one must account for the inferential behavior of the truth predicate, which PKF accomplishes in terms of its list of inference rules. I take it that inferential deflationism holds (i) that the inference rules are conceptually basic and (ii) that they are explanatorily basic. In a nutshell, (i) means that the rules are analytic and, therefore, they are necessarily truth-preserving and that this is *a priori* knowable. The import of (ii) is that the inference rules are not derived from concepts in terms of which truth can be defined. Also, the inference rules adequately explain all the functions of truth that we know and love, possibly in conjunction with claims about things other than truth which don’t presuppose anything about truth not captured by the inference rules. I now conclude.

*The Tarskian Turn* achieves its primary aim: it makes some main axiomatic theories of truth accessible to philosophers and shows how they can inform philosophical discussions about deflationism (p.5). It is worth emphasizing that one can’t simply read a philosophical theory of truth (deflationary or otherwise) off of an axiomatic theory of truth. For example, PKF is compatible with a correspondence theory of truth. Unlike the deflationist, the correspondence theorist would not take the inference rules to be conceptually and explanatorily basic. PKF doesn’t decide between deflationary and correspondence approaches to truth. What Horsten is doing is developing notions of deflationary truth that he takes to be in sync with the various axiomatic theories of truth he discusses.

Given the book’s limited aim, it cannot be faulted for not fully developing inferential deflationism and giving a substantive defense of it. However, by not doing these things the book does not allay the concern that arises in Chapter 7. That is, after Chapter 10 one wonders whether inferential-deflationist truth really is “deflationary.” Inferential deflationism cashes out our understanding of truth in terms of rule-following: understanding truth amounts to correctly deploying its inferential rules. This raises two related questions: Does the explanation of a concept of a rule call upon a concept of truth? Can a person be correctly described as following a rule without ascribing to that person possession of the concept of truth? As Horsten knows (pp. 147-48), an affirmative response to the first and a negative response to the second question suggests that inferential deflationism is not
deflationist enough. Since the nature of following a rule deserves its own book, Horsten’s responses to these questions are understandably light.

Horsten says in Chapter 10 that *prima facie* one could well imagine someone correctly applying a rule of inference without even possessing the concept of truth. Perhaps, on some notion of “correctly applying a rule of inference” this is imaginable. However, the inference rules that account for the functioning of truth are allegedly valid. How are we supposed to explain the notion of a valid inference rule without appealing to a notion of truth? On my view, an inference rule is valid only if it is necessarily truth preserving, i.e., only if it never sanctions inferring a non-truth from a truth. But then the concept of a valid inference rule calls upon a prior concept of truth. If so, inferential deflationist truth really isn’t deflationary. Of course, there are other explanations of the validity of an inference rule that do not directly appeal to truth. For example, one might say that a rule that sanctions the inference of $\alpha$ from $\beta$ is valid iff the assertion (or verification, or…) of $\beta$ commits one to $\alpha$. But this works for the inferential deflationist only if the epistemic practice(s) used to account for the validity of inference such as assertion, verification, commitment, etc., can be understood without appealing to truth. As previously mentioned, Horsten admits in Chapter 7 that it is an open question whether truth plays an explanatory role in epistemology and semantics. In the absence of further argument and analysis, it seems to be an open question whether inferential deflationism offers a truly deflationary notion of truth.