

Volker Halbach, *Axiomatic Theories of Truth*, Cambridge: Cambridge University Press, 2011, ix+365pp., \$85.00, ISBN 978-0-521-11581-0.

The work under review provides the first book-length study wholly devoted to its topic, and will be indispensable both to those already at work on axiomatic theories of truth and to students seeking to enter the field. Let me begin by locating that field within the wider area of truth studies.

Halbach begins by distinguishing *definitional* theories, which attempt to fill in the blank in 'Something is true if and only if _____' without using the word 'true,' and *axiomatic* theories, which attempt to characterize truth by listing various principles about it. But an even more important distinction is between what may be called *soft* and *hard* approaches. Examples of soft definitional theories are the traditional coherence, pragmatist, and correspondence theories. What goes into the blank for them (e.g. 'it coheres with other beliefs into a significant whole') is generally something at least as much in need of clarification as the notion truth itself. The first hard definitional theory was that of Alfred Tarski, who attempted to fill in the blank with something mathematically rigorous, since his goal was to make the notion of truth palatable to mathematicians; the most influential subsequent hard definitional theory has been Saul Kripke's, which stimulated

the development of others. Soft theories tend to seek complete generality, and to be unconcerned with paradoxes (and to take propositions to be the primary truth-bearers). Hard theories tend to restrict themselves to fragments of language, and to arise out of attempts to deal with the liar paradox (and to take sentences to be the primary truth-bearers). The soft/hard distinction can be made not only among definitional theories, but also among axiomatic theories. Crispin Wright's pluralism, for instance, which attempts to characterize the role of truth through informal 'platitudes,' could in principle be called a soft axiomatic theory, even if in practice the label 'axiomatic theory' tends to be limited to the hard kind. Halbach's topic in this book, one to which he has made many contributions in the journal literature, is such hard axiomatic theories.

In practice, the bulk of his attention is given to a single test situation, the one that has been most studied. Here one is concerned with truths about natural numbers, and exploits Gödel's coding of sentences by numbers. One starts with the first-order language L_0 of arithmetic with variables for natural numbers and symbols for operations like addition and multiplication, and adds a predicate T for 'is the code number of a true sentence,' thus obtaining an expanded language L_1 . One takes the theory in L_0 known as Peano arithmetic PA, and adds axioms for T to obtain some truth theory X . The

main technical problems addressed are, in jargon, the existence and characterization of ω -models for various X , and the determination for various X of its proof-theoretic strength. (An ω -model for X would be a model in which the variables range over the genuine natural numbers and the addition and multiplication symbols denote the genuine addition and multiplication operators on them. X counts as proof-theoretically stronger than Y if X can prove the consistency of Y , as Y itself generally cannot by Gödel's second incompleteness theorem.) A feature of this situation is that one can always construct a *liar* sentence A in L_1 , for which one can prove $A \leftrightarrow \sim T\langle A \rangle$, where $\langle A \rangle$ is the numeral for the code number of A . Thus no consistent X can prove the so-called Tarski biconditionals $A \leftrightarrow T\langle A \rangle$ for *all* sentences of L_1 . Halbach divides the theories he considers into two kinds: *typed*, in which T applies only to sentences of L_0 , and *untyped*, in which T applies also to sentences of L_1 involving T itself. The book is in four chapters, devoted to miscellaneous preliminaries, typed theories, untyped theories, and concluding mainly philosophical reflections, and running to about 30, 80, 150, and 50 pages, respectively.

The wealth of material presented cannot be adequately summarized in the space available here. I can at most note some highlights. Tarski took the ability to prove the Tarski biconditional for every sentence of L_0 as his

criterion of 'material adequacy' for a truth definition, but did not think that one could rest content with the Tarski biconditionals alone, since they do not allow us to prove such *composition principles* as 'A conjunction is true if and only if both conjuncts are true.' The first substantial result discussed, Theorem 8.12, is one that has been stated without full proof by several authors and has been proved in different ways by several other logicians, namely, that simply adding the composition principles for formulas of L_0 as one's only truth axioms, produces a *conservative* extension of PA: In other words, any sentence of L_0 provable in the indicated truth theory was provable already in PA. Since the Tarski biconditionals are implied by the composition principles, it follows that adding just them would also produce a conservative extension. Unfortunately, the proof in the book is flawed (according to a private communication from the author to the present reviewer). The conservativeness result is an indication that even typed truth theories can have points of interest.

Still, almost twice as much space is given to untyped theories, which have drawn the most attention in the literature. The study of these begins with one that turns out (applying a criterion due to Vann McGee) not to have any ω -models, the Friedman-Sheard theory FS. Its axioms include the composition principles (for the whole of L_1 , not just L_0 , since we are in the

untyped realm now), as well as the rules of T-introduction and T-elimination, permitting inference from A to $T\langle\langle A \rangle\rangle$ and from $T\langle\langle A \rangle\rangle$ to A . Though one has these rules, one has neither $A \rightarrow T\langle\langle A \rangle\rangle$ nor $T\langle\langle A \rangle\rangle \rightarrow A$, and adding either one would result in inconsistency. The proof that FS itself is consistent involves a clever application of aspects of the Gupta-Herzberger revision approach to truth.

The heart of the book, however, lies in its treatment of theories for which the well-known Kripke construction provides models. There is some coverage of the work of Andrea Cantini on versions of the Kripke construction involving the van Fraassen supervaluation scheme, but the bulk of the coverage is given to theories based on the better-known version involving the Kleene strong trivalent scheme. The so-called Kripke-Feferman theory KF describes the common properties of all fixed points "from the outside" (using classical logic). A variant PKF describes them "from the inside" (using classical logic). Theorem 16.31, representing joint work of Halbach and Leon Horsten, gives an exact determination of the proof-theoretic strength of PKF, showing it to be significantly weaker than KF, whose proof-theoretic strength had earlier been determined by Feferman. Another variant (suggested by the reviewer), describing "from the

outside" the *minimal* fixed point specifically, is also discussed; it can be shown to be, unsurprisingly, of much higher proof-theoretic strength.

John P. Burgess

Department of Philosophy

Princeton University, Princeton, NJ 08544-1006, USA

jburgess@princeton.edu