

# 1 About This Book

## 1.1 What Is It About?

Horwich's book *Truth* first appeared in 1990. It established deflationism about truth as a central subject of philosophical investigation. In the first decade after the appearance of this book, deflationism was mostly associated with an axiomatic theory of truth that is known as the disquotational theory. This theory has its origins in Tarski's work: It is the earliest axiomatic theory of truth that meets minimal adequacy conditions.

In recent years, the attention of philosophers is gradually shifting to stronger axiomatic theories of truth. There is a growing consensus that truth theories stronger than the disquotational theory are needed for the uses to which the concept of truth is put in ordinary language, philosophy, and science.

A debate about the question of whether these stronger axiomatic theories are compatible with the main tenets of deflationism is currently in its formative stages. So far this discussion has been rather murky. There are several reasons for this. First, there is disagreement about what deflationism about truth is committed to. Second, it is less than clear which axiomatic theories of truth are philosophically attractive for the deflationist and which are not. Third, there is no unanimity about how strong truth theories need to be for performing the functions that the concept of truth needs to fulfill. This book aims to clarify these issues, and thereby to contribute to the question of whether the theories of truth that we need can be interpreted in a deflationary manner. A new philosophical theory of truth that provides a deflationary interpretation of our best axiomatic theory of truth is proposed and defended against possible objections.

My objective is to present the central concerns, questions, theories, and arguments perspicuously and concisely, and not to get carried too far into argumentative exchanges that somehow involve intangibles.

Here are some of the questions that are given only scant attention in this book:

1. Is truth a property?
2. Of what is truth a property: sentences, theories, utterances, assertions, judgments, propositions, etc.?

My answers are “yes,” and “sentences,” respectively. There is an extensive body of literature devoted to these questions. Some of the classical articles on these matters are reprinted in (Blackburn & Simmons 1999, Part III). A cursory reading of this literature reveals that my answers to the two questions agree with many contemporary philosophers’ answers to these questions.

On the first question, I have little to say. It lives in a sphere where the air is almost too thin to breathe. My reason for taking truth to be a property is simple. The truth predicate has an extension—the collection of all true sentences—and this collection does not, unlike the “extension” of the expression “exists,” consist of everything or even of all sentences. This does not entail that the property of truth is metaphysically deep or that there is a nature of truth to be discovered in reality. Truth may be a metaphysically uninteresting property; it is a property nonetheless.<sup>1</sup> Note that this reason for taking truth to be a property can be adapted to a view that takes propositions or utterances to be the truth-bearers.

On the second question, I have two remarks to offer. First, under certain conditions, one might be prepared to compromise for all that is said in this book. For instance, one *might* be willing to take truth to be a property of propositions, as long as propositions are structured in the way that sentences are. However, every such proposal must be carefully evaluated because it may introduce problems of its own. Some argue, for instance, that because the notion of proposition presupposes the notion of truth, it should not be used to explicate the notion of truth—on pain of circularity.

Second, it must be admitted that, strictly speaking, utterances are better candidates for being truth-bearers than sentences. This is because one sentence can be used to express a truth in one context and a falsehood in another. Take, for instance, the following short conversation between Bertie Wooster and Bingo Little [Wodehouse 1930, p. 15]:

*Bingo:* [...] if your aunt supposed I was a pal of yours, she would naturally sack me on the spot.

*Bertie:* Why?

*Bingo:* Why? Be reasonable, Bertie. If you were your aunt, and you knew the sort of chap you were, would you let a fellow you knew to be your best pal tutor your son?

1. Arguments against the thesis that truth is a property can be found in [Grover 2001].

A generous dose of pragmatic enrichment is needed to answer Bingo's retorical question truthfully. When uttered by Bingo when addressing Bertie, the answer is undoubtedly "no," as even Bertie is ready to admit. But when uttered by you, addressing your best friend, the answer may very well be "yes." Nonetheless, I take sentences to be the truth-bearers. Here is why. When abstraction is made of contextual reference-fixings, disambiguations, and so on, what are left are, for all intents and purposes, "eternal" sentences about the world. We are concerned with theories of truth for such sentences. If we want to extend such theories to theories of truth for utterances, more work has to be done. But even if truth must ultimately be seen as a property of utterances, it is reasonable to expect that in developing a theory of truth for eternal sentences, we are off to a good start.

This book is about the relation between formal theories of truth and deflationism about truth. Deflationism is a view that takes the notion of truth to be a light and insubstantial notion. This is a truly philosophical view that can be and has been made more precise in multiple ways. Crucial in making the deflationary intuition precise is the way in which this philosophical view is related to formal or logical aspects of the notion of truth.

In analogy with the division between the syntactic and semantic view in the philosophy of science, theories of truth and the paradoxes can be divided roughly into two classes. On the one hand, there are *semantical* theories of truth, which are primarily interested in describing one or more models for languages that contain a truth predicate. On the other hand, there are *axiomatic* (or syntactic) theories of truth, which are primarily interested in explicating basic logical principles governing the concept of truth. It is argued that if we want a theory of truth for a natural language such as English, we ultimately have to opt for an axiomatic theory of truth, although semantical theories of truth can be, and often are, of great heuristic value. For this reason, we are more concerned with axiomatic than with semantical theories of truth.

Because we prefer axiomatic truth theories over semantical ones, this book is about the relation between *axiomatic* truth theories and deflationism. Every deflationist theory presupposes at least a quasi-formal theory of truth. Not so long ago, virtually the only axiomatic theory of truth that was discussed consisted of a collection of sentences of the form

$\phi$  if and only if it is true that  $\phi$ ,

which are known as *Tarski-biconditionals*. Today, many discussions of deflationism still proceed against the background of this theory. But in recent decades, a number of other natural axiomatic theories of truth have come to the fore in the logical literature. Most of them are proof-theoretically stronger than

(natural) consistent collections of Tarski-biconditionals: They prove more sentences. So a philosophical question arises: Which of these axiomatic theories should a deflationist about truth adopt?

Despite its importance, this question has not received the attention it deserves in the literature. We see that the plausibility of a deflationist theory of truth depends in part on the plausibility of the axiomatic theory of truth on which it is built. For one thing, it is argued that the attractiveness of Horwich's minimalist theory of truth is diminished by the fact that he takes the meaning of the concept of truth to be exhausted by the Tarski-biconditionals. If a plausible version of deflationism is to be found, it must be based on one of the best axiomatic theories of truth that are available today. These are deductively significantly stronger than the traditional axiomatic truth theories that consist solely of Tarski-biconditionals.

The question of whether our best axiomatic theories of truth are compatible with deflationism at all cannot be prejudged at the outset. Indeed, we see that the most influential versions of deflationism about truth do not harmonize with the best axiomatic theories of truth that are available today. Some authors infer from this that deflationism is simply misguided. But this is a rash conclusion. Instead, I argue that the insubstantiality of truth has been misunderstood in the literature. In the later chapters of this book, I develop and argue for a new kind of deflationism, which I call *inferential deflationism*. According to inferential deflationism, truth is a concept without a nature or an essence. This is betrayed by the fact that there are no unrestricted logical laws that govern the concept of truth. Inferential deflationism concerning the concept of truth is a philosophical position, so it cannot be literally entailed by any formal truth theory. Nonetheless, this form of deflationism is seen to flow naturally from some of our best contemporary axiomatic theories of truth.

One of the goals of this book is to provide lessons for philosophers interested in deflationism and for logicians interested in the concept of truth. Therefore, these writings are semi-philosophical and semi-technical nature. Technicality cannot be altogether avoided in the subject under investigation. One of the first lessons of proof theory is that the details of a formal system are important. Changing one apparently minute detail in a formal system can transform it from a weak theory to a strong theory, or vice versa. Therefore, the details of the truth theories behind deflationist positions should be made explicit. Philosophers have tended to neglect to do this. But if one wants to address the question of whether a given truth theory can be interpreted in a deflationary way, one needs to know whether it is weak or strong, and in which ways it is weak or strong. So the details of the system do matter!

A certain degree of recognition of the importance of precision and detail in philosophical debates about truth has been forthcoming in recent years [Williamson 2006, p. 179]:

One clear lesson [of technical work by philosophical and mathematical logicians] is that claims about truth need to be formulated with extreme precision, not out of 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.

Williamson is surely right about this. But when the logicians have done their business, it is the philosopher's job to interpret the formal truth-theoretical results that have been reached. This book intends to help bridge the gap between the logical and philosophical literature on truth. The goal is to make the essentials of existing logical knowledge about axiomatic theories of truth as accessible as possible to philosophers, and to explicate how they are related to the philosophical discussion about deflationism.

## 1.2 What Is It Not About?

The present book is not intended as a comprehensive critical discussion of all contemporary varieties of deflationism of truth. There are simply too many such deflationist positions around to do them justice in the span of a monograph. The philosophical literature on deflationism about truth is too extensive to be comprehensively covered here. Also, relatively little is said about the relation between the theory of truth and the theory of meaning. This is a subject in its own right, and a vast one at that.

The liar paradox puts pressure on classical logic. Aside from logic, so little is needed to generate the liar paradox that one wonders whether the laws of classical logic are unrestrictedly valid after all. So it comes as no surprise that many theories of truth have been formulated in an environment of nonclassical logic.

Within such environments, a distinction between two classes of theories can be made. The first kind of nonclassicality stays close to classical logic. It shares with classical logic the assumption that there are no more than two mutually exclusive truth values: *true* and *false*. But contrary to classical logic, it maintains that some sentences fail to have a truth value. Perhaps the liar sentence is a good candidate for being a sentence without a truth value. The logical calculus of sentences that do have truth values are just like classical logic. But the classical setting has to be extended to accommodate sentences that lack a truth value. The second kind of nonclassicality denies that there are only two truth values and/or denies that truth values are mutually exclusive. Thus, some multivalued

logics will claim that some sentences are “half true,” and dialetheist logicians will claim that some sentences are both true and false at the same time.

In this book, we hold onto classical logic as much as we possibly can. Bending but not breaking is our device. In response to the semantical paradoxes, we are willing to entertain—albeit grudgingly—the possibility that some sentences lack a truth value. But we are not concerned with theories which claim that sentences can have more than one truth value at the same time or a sentence can have an intermediate truth value. This means that theories of truth based on paraconsistent logic or fuzzy logic or combinations thereof will not be dealt with here.

As adumbrated earlier, we also are not concerned with semantical theories of truth. It is argued that semantical theories of truth can never pass the ultimate adequacy test. They can at best be of heuristic value: as stepping stones to an axiomatic theory of truth. For this reason, certain prominent semantical theories of truth are not treated in depth in this book. Kripke’s semantical theory of truth is discussed because it has inspired promising axiomatic theories of truth. Indeed, the version of deflationism that is defended is inspired by an axiomatisation of Kripke’s theory of truth. The revision theory of truth is also discussed in this book because it is closely connected to certain influential axiomatic truth theories.

### 1.3 For Whom Is It Intended?

The present book is intended for anyone interested in the debate about the relation between philosophical and formal theories of truth. It can be used as a textbook on this subject for senior undergraduate and beginning graduate students in philosophy. It is intended to be relevant for all students and teachers of analytic philosophy regardless of which area of philosophy carries their preference (epistemology, philosophy of language, moral philosophy, metaphysics, etc.).

No specific prior philosophical knowledge is presupposed. That being said, it is useful for the reader to have general background knowledge in analytic philosophy—but that almost goes without saying. In particular, it is useful, albeit not required, for the reader to be familiar with the distinction between substantial and deflationist views of truth. A decent introduction to this debate is [Kirkham 1995], and a good collection of articles about deflationism is [Armour-Garb & Beall 2005].

It is beneficial if the reader has taken an intermediate logic course. Too many philosophy students are required to go through introductory and intermediate courses of logic but graduate without being convinced that, in particular, the more advanced logic course was of essential importance in their philosophical education. One of the objectives of the present monograph is to show what

intermediate and advanced logic courses in philosophy departments are really good for.

It would be helpful if the reader has gone through some standard detailed proof of Gödel's completeness and incompleteness theorems. In particular, knowledge of Gödel's proof of the first incompleteness theorem facilitates a deeper understanding of this book than can otherwise be obtained. Even though, historically, Tarski's ground-breaking work on truth was carried out independently from Gödel's work on the incompleteness theorems, Tarski's results are best explained as applications of Gödel's proof techniques. If the reader already possesses ample knowledge of the incompleteness and undefinability results, she can safely skip chapter 3. In chapter 3, the technical results used in the remainder of the book are stated and explained. But most of them are not proved in detail there. Good expositions of the proofs of the incompleteness and undefinability theorems can be found in [Goldstern & Judah 1998] and [Boolos & Jeffrey 1989].

Aside from this, some basic set theory is presupposed in the later chapters of this book. [Enderton 1977] is a good source for obtaining even more than the set theoretical background needed for reading them. Elementary knowledge of transfinite ordinals and cardinalities is especially helpful. In particular, on several occasions, use is made of *Cantor's theorem*, which states that no set, finite or infinite, stands in a one-to-one correspondence with its power set, which is the set of all its subsets.

#### 1.4 How Is It Structured?

There is a pattern of alteration between logical and philosophical chapters. We start with a chapter on the *Dämmerung* of "substantial" theories of truth. This is followed by a technical chapter in which the background of Tarski's results on truth is reviewed. Subsequently, Tarski's disquotational theory is discussed. The deflationist view of truth, which is often closely associated with the disquotational theory, is then critically scrutinized. It is argued that, even according to deflationism, a truth theory that is stronger than the disquotational theory is needed. A first such theory is found in the compositional theory of truth, which also traces back to Tarski. It turns out that the compositional theory allows us to prove mathematical facts that go beyond what the background mathematical theory can prove on its own. In that sense, the compositional theory is said to be "non conservative." The question of whether this is compatible with the tenets of deflationism is probed. In this context, attention is also given to the wider question of what role the concept of truth can legitimately play in the empirical sciences and in subdisciplines of philosophy.

Then we resume our ascent on the ladder toward stronger truth theories. Above the compositional theory, the path forks. When we try to construct strong theories of truth that explain how the truth predicate behaves with respect to sentences that contain occurrences of the truth predicate, tough choices have to be made. Either we hold onto classical logic and try to construct a classical truth theory that is as strong as possible. This leads us to the Friedman–Sheard theory. Or we abandon classical logic in favor of truth value gaps and try to strengthen the truth theory along this path. This leads to the Kripke–Feferman theory and variants of it. Both of these classical and partial theories of truth are theories of *reflexive truth*: They allow us to prove truth-iterations (“It is true that it is true that so-and-so”). To conclude, a new version of deflationism, called *inferential deflationism*, is articulated. It is argued that this way of viewing the concept of truth is naturally suggested by the correct axiomatization of Kripke’s theory of truth. In the final chapter, we also reflect on the minimal strength of truth theories that is needed to reconstruct philosophical arguments in which the concept of truth plays a role.

Not reading the chapters in the indicated order might prove awkward and difficult: Later chapters tend to presuppose and build on what is said in earlier ones. But there is a natural stopping point for the weary intellectual traveller. A reader who simply wants to acquire insight into the relation between most existing versions of deflationism and the axiomatic theories with which they ally themselves can stop after chapter 7, which deals with the subject of conservativeness. What comes afterward is a bit more advanced. However, it is my expectation that as the debate on deflationism moves forward over the coming years, the material in the later chapters will turn out to become more and more essential for understanding what is going on.

### 1.5 Note on Technicality and Notation

A sustained and determined attempt has been made to keep the technicality of this book down to the essential minimum. But I must admit that, in my opinion, the essential minimum contains a bit more than some of the experts on deflationism think it does.

In this book, I try to bridge the gap that divides philosophers and logicians who think about truth. This requires striking a happy balance between the need for logical precision, on the one hand, and the need not to go beyond the level of precision and technicality that is absolutely required for addressing the philosophical questions that lie at the heart of the investigation, on the other hand. I labor under no illusion that I have succeeded in this task to everyone’s satisfaction. There are bound to be philosophers who will find many discussions

in the book needlessly and distractingly technical, whereas many a logician will lament the sloppiness with which certain logical issues are treated—and in the back of their minds the question will form: “*Can* he work out the details?”

One of the things that impede communication between philosophers and philosophical logicians on the subject of truth is simply gödel coding. Doing the coding correctly is difficult. In addition to that, formulae involving gödel coding are hard to interpret: They do not wear their intended interpretation on their sleeves.

The course that is followed is the following. In chapter 3, in which the meta-mathematical background results are reviewed, we are careful to express all formulae correctly. But at the end of chapter 3, we drop gödel coding completely. This results in a whole array of ungrammatical formulae throughout the remainder of the book. But the ungrammatical formulae have clear advantages that more than compensate for their being ungrammatical. First, they *do* wear their intended interpretation on their sleeves—they are just not spelled correctly. Second, the logically educated reader can, if she so desires, correct any ungrammatical sentences so as to proudly produce a correct, virtually unreadable formula. If, as a philosophical logician, one tries to give a logical analysis of an axiomatic truth theory, one can hardly forsake the gödel coding. But if one looks as a philosopher at formal truth theories and their formal properties, the gödel coding is something one can do without.

Aside from some standard mathematical symbols, fairly widely used logical notation is used throughout this book. Standard symbols are used for the logical connectives:  $\wedge$  (and),  $\vee$  (or),  $\neg$  (not),  $\rightarrow$  (material implication),  $\leftrightarrow$  (material equivalence),  $\exists$  (existential quantifier),  $\forall$  (universal quantifier),  $\Box$  (necessity),  $\Diamond$  (possibility),  $K$  (knowledge). The symbol  $=:$  stands for equality by definition. Lowercase Greek letters are used for formulae and sentences, whereas capital Greek letters are used for sets of formulae and sentences. Capital roman letters (such as  $S$ ,  $S'$ ) and abbreviations (such as  $DT$  for the *Disquotational Theory*) are used for theories (i.e., computably axiomatisable sets of sentences).  $\vdash$  stands for the classical logical derivability relation; a subscript on the  $\vdash$ -relation (as in  $\vdash_S$ ) indicates derivability in a particular theory.  $\models$  stands for the classical semantic consequence relation; a subscript on the  $\models$ -relation (as in  $\models_S$ ) indicates a nonclassical semantical modeling relation.  $\mathbb{N}$  stands for the natural numbers structure. Calligraphic letters (such as  $\mathcal{L}_{PA}$ ) (the language of Peano arithmetic) are used for formal languages; Gothic-type letters (such as  $\mathfrak{M}$ ) are used for models. In the first chapters, where the gödel coding is still used, we use  $\ulcorner \dots \urcorner$  for referring to gödel codes, and we use overlining ( $\overline{n}$ ) to refer to standard numerals. Additional specific details about notation are explained as we go along. After the bibliography, the reader finds a glossary of logical notation

that is used throughout the book. The beginning of the subject index contains a list of abbreviations of formal theories, axioms, and rules.

The final sections of some of the chapters in this book are a bit more technical than the rest. A thorough understanding of them presupposes some familiarity with certain basic concepts of proof theory. More on the elements of proof theory that are needed for a thorough understanding of these sections is contained in [Franzen 2004]. The reader can, if she so chooses, skip these sections without missing too much.

I do not present the proofs of all the theorems that we encounter. Especially for the more difficult proofs, the reader is referred to the specialized literature.