

CHAPTER 1

A Critique of Deflationism

1.1. Introduction

The past century has witnessed two types of philosophical debates over the concept of truth. In the first, *substantive*, type of debate we find rival theories of truth put forward that seem to have, and whose proponents have taken them to have, significant metaphysical and epistemological implications. An early example of this type is the debate in the early 1900s between the British Idealists (F. H. Bradley and his followers) and the Logical Atomists (Bertrand Russell and his followers). The Idealists defended a coherence theory of truth, whereas the Atomists argued for a correspondence theory. This dispute over the theory of truth was not, and was not taken by the participants to be, a local disagreement. It was integral to the larger metaphysical debate between the two sides over monism and pluralism and over idealism and realism. A more recent example of the substantive type is the debate between the realist and the anti-realist found in Michael Dummett's writings. The crux of the debate here is what notion of truth is admissible. Dummett's anti-realist argues for a notion of truth that is constrained by evidence, while the realist defends the admissibility of a radically non-epistemic notion.

In the first type of debate, then, we find theses put forward and defended that have (or at least seem to have) substantial philosophical implications. Debates of this type presuppose that truth has a substantial role to play in philosophical inquiry. In the debates of the second, *metaphilosophical*, type the presupposition is called into question. An early example of this type is the debate over the claim, made by some Logical Positivists, that truth is a metaphysical concept and hence ought

to be banished from all rigorous and scientific thought. A decisive contribution to this debate was made by Alfred Tarski, who gave a definition of truth (for certain languages) that was adequate by the Positivists' own strictures. Tarski's definition used only terms that the Positivists found legitimate, and it defined a notion that was provably coextensive with truth. Tarski's work was widely viewed as establishing the legitimacy and the usefulness of truth in philosophical inquiry. One result of its influence was a shift away from a syntactical conception of language and towards a semantical one.¹

Ironically, Tarski's work, while refuting one sort of skepticism about the usefulness of truth, provided a basis for a different, more compelling, kind of skepticism. This new kind of skepticism, *deflationism*, maintains that truth is a simple and clear concept and has no substantial role to play in philosophy. Substantive debates over truth, according to deflationism, are in error not because they work with a notion that is metaphysically loaded (and hence corrupt), but because they work with a notion that is metaphysically lightweight. Deflationism has provoked a large debate among philosophers, a debate that provides a contemporary instance of the second, metaphilosophical, type of debate distinguished above.

A deflationary view typically consists of two parts: (i) a description of the meaning and function of 'true' and (ii) a derivation from that description of deflationary consequences concerning truth. As an example of (i), consider the following passage from Michael Williams; it contains a popular account of the meaning and function of 'true'. (In the next section I shall explain and discuss the account in detail.)

[W]hen we have pointed to certain formal features of the truth-predicate (notably its 'disquotational' feature) and explained why it is useful to have a predicate like this (e.g. as a device for asserting infinite conjunctions), we have said just about everything there is to be said about truth.²

Examples of (ii) can be found in §§1.3 and 1.4 below. The following extracts illustrate the sorts of deflationary consequences that are often drawn. The first extract is from Richard Rorty; the remaining two are from, respectively, Scott Soames and Paul Horwich.³

[T]ruth is not the sort of thing one should expect to have a philosophically interesting theory about.⁴

What does seem right about Tarski's approach is its deflationist character. . . . Truth is a useful notion, but it is not the key to what there is, or to how we represent the world to ourselves through language.⁵

[Truth is not] a deep and vital element of philosophical theory. . . . [T]he realism/anti-realism issue (together with various related questions in the philosophy of science) have nothing at all to do with truth.⁶

In short, deflationism holds that once we understand the meaning and function of 'true'—and this understanding, according to deflationism, is not hard to achieve—we shall see that truth has no substantial role to play in philosophy. Many contemporary philosophers find the deflationary account of 'true' attractive and plausible, and they have accepted (sometimes enthusiastically, sometimes regretfully) its negative verdict on the role of the concept of truth in philosophy.

I want to oppose deflationary attitudes in philosophy. The main problem with deflationism, in my view, lies in the descriptive account it gives of 'true'. The deflationary account makes (and, to sustain its conclusions, needs to make) some very strong claims about the meaning of 'true'—claims that on examination prove to be highly problematic. The account appears plausible, I think, only because we read its claims in a weaker way. But the weaker readings do not, I believe, yield any deflationary conclusions.

The argument I shall develop against deflationism, then, is this. The deflationary description of ‘true’, when it is taken in the strong and intended way, motivates the deflationary conclusions, but is highly problematic. On the other hand, when it is taken in the weaker way, the description is correct enough, but does not yield the deflationary conclusions. I shall substantiate this by considering deflationary arguments on two issues: the possibility of a physicalistic theory of truth (§1.3), and truth and meaning (§1.4). Deflationists take the concept of truth to be transparent, one capable of a complete and simple philosophical analysis. Towards the end of the chapter (§1.5) I shall point out some reasons to think that truth is a highly puzzling notion, one that defies all our attempts at its analysis.

1.2. The Disquotational Theory

Let us consider the disquotational account of the meaning of ‘true’, which we encountered briefly in the extract from Williams.⁷ Its original source is the following well-known passage from W. V. Quine’s *Philosophy of Logic*.⁸

By calling the sentence [‘snow is white’] true, we call snow white. The truth predicate is a device of disquotation. We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent. (p. 12)

Stephen Leeds provides, in the following extract, a useful elaboration of the disquotational account.

It is not surprising that we should have use for a predicate P with the property that “‘___’ is P ” and “___” are always interderivable. For we frequently find ourselves in a position to assert each sentence in a certain infinite set z (e.g., when all the members of z share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of z is warranted. A predicate P with the property described allows us to construct such a sentence: $(x)(x \in z \rightarrow P(x))$. Truth is thus a notion that we might reasonably want to have on hand, for expressing semantic ascent and descent, infinite conjunction and disjunction. And given that we want such a notion, it is not difficult to explain how it is that we have been able to invent one.⁹

The core thought here is that the function of the truth predicate is to serve certain expressive purposes, namely, that of expressing certain infinite conjunctions and disjunctions. The truth predicate serves these functions in virtue of its disquotational character; that is, in virtue of the fact that it undoes the effect of quotation marks.¹⁰ For example, the role of ‘true’ in

(1) ‘snow is white’ is true

is to cancel the quotation marks: (1) says no more nor less than the sentence

snow is white.

We shall get clearer on the disquotational theory if we consider a situation in which, as Quine puts it, “we want to affirm some infinite lot of sentences.” Suppose we wish to affirm all sentences of the form

___ & snow is white [= A , say].

That is, we want to affirm the conjunction of all sentences obtained by filling the blank in A with sentences of English:

(2) [Sky is blue & snow is white] & [Chicago is blue & snow is white] & . . .

We lack explicit and direct means of formulating the infinite conjunction, but the truth predicate, according to Quine and Leeds, provides us with an indirect means. Observe that we cannot generalize on the ‘___’ position in A using ordinary first-order variables. We cannot say, for example,

For all x : x & snow is white.

For the variable ‘ x ’ is pronominal and occupies name positions; it cannot meaningfully be put in sentence positions. The way the truth predicate helps here, according to the disquotational account, is this. The disquotational feature of truth makes (2) equivalent to

(3) [‘Sky is blue’ is true & snow is white] & [‘Chicago is blue’ is true & snow is white] & . . .

But the position ‘___’ in

___ is true & snow is white

is nominal and can be quantified using the pronominal variable ‘ x ’. We can say,

(4) For all sentences x : [x is true & snow is white].

But (4) is equivalent to (3) and, consequently, in virtue of disquotation, to (2). The truth predicate thus provides us with a means of expressing the infinite conjunction (2). Truth is, on the disquotational account, essentially a logical device. It enables us to generalize over sentence positions *while using pronominal variables such as ‘ x ’* and, thus, endows us with additional expressive power.

It will be useful to separate out four component ideas of the disquotational theory.

The Disquotation Thesis: The truth predicate is a device of disquotation.

The Infinite Conjunction Thesis: The truth predicate enables us to express certain infinite conjunctions and disjunctions; (4), for instance, expresses (2) and (3).¹¹

The Generalization Thesis: The truth predicate provides a means for generalizing over sentence positions even when the variables are pronominal.

The Connection Thesis: The truth predicate serves its expressive functions in virtue of its disquotation feature.¹²

The first two of these theses contain important ambiguities. Let us demarcate a little the sense in which the deflationists understand these theses (and need to understand them).

Let us call instances of the form

(T) ‘___’ is true if and only if ___

T-biconditionals.¹³ Then, the Disquotation Thesis is understood by the deflationists as saying not just that the T-biconditionals are true, nor just that they are necessarily true.¹⁴ The claim is rather that the T-biconditionals issue from our very understanding of ‘true’, that they explain (at least partially) the meaning of ‘true’.¹⁵ This way of reading the Disquotation Thesis is not always explicit in the writings of the deflationists. But, as we shall see, it is required by key deflationary arguments. Furthermore, some authors *are* explicit on the point. Horwich has stated that our understanding of ‘true’ *consists* in our “disposition to accept, without evidence, any instantiation of the schema [(T)].” And he speaks of the T-biconditionals as constituting a *definition* of ‘true’.¹⁶ Even philosophers opposed to deflationism have often been attracted to this reading of the Disquotation Thesis.¹⁷ Hartry Field’s influential paper “Tarski’s Theory of Truth” argues for a view as far removed from deflationism as any. Yet it contains a description of ‘true’ that would fit comfortably in any deflationary text:¹⁸

[L]et's note one obvious fact about how the word 'true' is standardly learned: we learn how to apply it to utterances of our own language first, and when we later learn to apply it to other languages it is by conceiving the utterances of another language more or less on the model of utterances of our own language. The obvious model of the first stage of this process is that we learn to accept all instances of the schema

(T) X is true if and only if p .

where ' X ' is replaced by a quotation-mark name of an English sentence S and ' p ' is replaced by S .¹⁹

In summary, we shall understand the Disquotation Thesis as stating that disquotation provides an analysis of 'true', that it explains (at least partially) what the word means and what our understanding of it consists in. The thesis should be sharply distinguished from weaker ideas such as that the T-biconditionals are necessarily true.

The Infinite Conjunction Thesis separates out for consideration the claim, often made by the deflationists (and sometimes by the nondeflationists also), that the truth predicate is a device for expressing certain infinite conjunctions and disjunctions. The thesis is ambiguous because of an ambiguity in 'express'. Is the thesis to be read so that it yields only that (4) and (2) are *materially* equivalent? Or that they are necessarily equivalent? Or that they have the same sense? Or something yet different? The deflationists have not been explicit on the point. We shall see, however, that the use they make of the Infinite Conjunction Thesis requires that 'express' be read in a strong way.

One argument of the deflationists—that for the Connection Thesis—requires much too strong a reading of the Infinite Conjunction Thesis. According to the Connection Thesis the truth predicate *needs* to be disquotational if it is to serve its

expressive functions—in particular, its function of expressing certain generalizations. The argument for the thesis was implicit in our exposition above: The function of (4) is to express (2). But this is possible only if (2) and (3) are equivalent. Here is where disquotation comes in; it is needed to ensure that the equivalence of (2) and (3) holds. Hence, disquotation is needed to ensure that truth plays its desired role in generalizations such as (4). The role of the Infinite Conjunction Thesis in the argument is to show that (2) and (3) need to be equivalent, if the truth predicate is to play its expressive role. But this motivates the need for a disquotational truth predicate only if the equivalence of (2) and (3) is required to be something like sameness of sense. Anything weaker will yield the need, not for disquotational truth, but for something weaker. If, for example, the role of truth in (4) requires only that (2) and (3) be necessarily equivalent, then the argument will yield only that the T-biconditionals must be necessarily true if ‘true’ is to serve its role. It will not yield the Disquotation Thesis.

In the strong sense needed for the Connection Thesis, the Infinite Conjunction Thesis is plainly false. A universal statement (e.g., (4)) does not have the same sense as the conjunction of its instances (e.g., (3)). The two typically do not even imply the same things; they are equivalent only in a much weaker sense.²⁰ I think that the proponents of the disquotational theory have gone astray because they have ignored the difference between wanting to affirm a generalization and wanting to affirm each of its instances. Quine writes in the passage quoted above, “if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use.” In the situation envisaged by Quine, where we can demarcate some infinite lot of sentences only by talking about them, what we typically want to do is affirm a generalization, and the truth predicate enables us to do this. We can, for example, generalize on the ‘___’ position in

___ & snow is white

with the aid of the truth predicate, as we saw above. But this is not the same as affirming “some infinite lot of sentences,” which requires an infinitary conjunction. It is because two distinct things (which, to repeat, are affirming the universal and affirming all the instances) are confused that the infinitary conjunction seems to be strongly equivalent to the generalization, and leads in turn to the Disquotation Thesis. (I think the same confusion is going on in Leeds’s claim in the passage quoted above that “we frequently find ourselves in a position to assert each sentence in a certain infinite set z (e.g., when all the members of z share a common form).”)

The Connection Thesis, then, rests on a confusion. This, I think, is a blemish on deflationism. It means that the deflationary accounts of the meaning and the function of ‘true’ are not connected in the neat way that the deflationists supposed. But this does not damage deflationism in a material way. For the arguments for the deflationary attitude towards the role of truth in philosophy rest not on the Connection Thesis but on the Disquotation and the Infinite Conjunction Theses. Let us now examine some of these arguments.

1.3. Physicalism About Truth

One question that philosophers have debated in recent years is whether truth is amenable to a physicalistic reduction—in other words, whether truth is a complex physical property. Two compelling philosophical pictures, when combined, suggest a positive answer: (i) the correspondence theory of truth and (ii) a physicalistic ontology. The former suggests that underlying truth there is a systematic relation between words and the world; the latter suggests that this relation can be understood in physical terms. The combination of the two pictures, in fact, makes each a little more attractive. Traditional correspondence theories are confronted with the embarrassment that they have had little to say (beyond such vacuous claims as ‘snow’ refers to snow) about the relation between words and the world. Physicalistic ontology is useful here: it provides a framework in which a substantial

account of the relation might be spelled out. Physicalistic ontology has faced, on the other hand, the problem of giving an account of psychological and semantic properties. A correspondence theory helps here: it provides a scheme for making sense of at least one semantic property. In short, the idea that truth is a complex physical property makes the two philosophical pictures a little more coherent and attractive.

The deflationist position on the question is, as one would expect, that truth is not amenable to a physicalistic reduction, that to suppose otherwise is to misunderstand the meaning and function of ‘true’. We shall examine the deflationary arguments for this claim after we have briefly reviewed the debate within which the arguments arose.

Hartry Field initiated the debate by arguing (in his paper, “Tarski’s Theory of Truth”) that truth is amenable to a physicalistic reduction. Field argued that just as the usefulness of the concept of valence in chemistry is a reason to expect a physicalistic reduction for it, so with truth: the usefulness of the notion of truth is a reason to think that it has a physicalistic reduction. Stephen Leeds pointed out a problem with this argument.²¹ What provides us with a reason to expect a reduction of “valence” is that it is a *causal-explanatory* notion, as is shown by its role in the *law* of valences. Mere usefulness does not establish the requisite analogy of truth with valence. What must be shown is that there are laws of truth analogous to the law of valences. (Leeds went on to suggest that the utility of truth can be explained by seeing it as a device for expressing infinite conjunctions and disjunctions.) Hilary Putnam took up Leeds’s challenge.²² He argued that, like valence, truth does play a causal-explanatory role. He offered several generalizations as examples of causal-explanatory laws involving truth, such as the following:

- (5) The laws of a mature science are typically approximately true.
- (6) True beliefs about how to attain our goals tend to facilitate success in achieving them.

(7) Beliefs formed as a result of our methods of inquiry tend to be true.

The first law, Putnam suggested, helps explain the success of the mature sciences; the last two help explain *our* success in attaining our goals.²³ The deflationists responded that Putnam's examples do not pose a difficulty for them; the examples, they argued, can be explained within their framework. Their arguments seem to have been widely accepted and have contributed to the prevalent skepticism of the possibility of a physicalistic reduction of semantic concepts.

Let us consider how the deflationary arguments go for one of Putnam's examples. (The others are treated in a parallel way.) Here is how Williams responds to (6).

I see no reason to think of [(6)] as a law. . . . If I want a cold drink and believe that the refrigerator, rather than the oven, is the place to look, I will increase the likelihood of getting what I want. This is because cold drinks are more likely to be found in the refrigerator than in the oven. To say that my having true beliefs makes it more likely that I will attain my goals is just a compact way of pointing to the indefinite number of mundane facts of this sort. It involves nothing so arcane as a physical correspondence theory of truth.²⁴

Williams argues here that (6) is not a law, since it is "just a compact way of pointing to the indefinite number of mundane facts" of the sort he cites. Let A_1, A_2, A_3, \dots be these mundane facts. Williams's argument rests on the idea that (6) expresses—in some sense of "expresses"—the infinite conjunction

(8) $A_1 \ \& \ A_2 \ \& \ A_3 \ \dots$

It is plain that Williams's argument does not work if "express" is understood in an extensional way; i.e., if we suppose only that (6) is materially equivalent to (8). Nor does the argument work if we take "express" in an intensional way; i.e., if we

suppose only that (6) is necessarily equivalent to (8). For, of two sentences that are necessarily equivalent, one can be a law and the other not. Here is an example:

(9) Cicero is Tully.

(10) No chemical reaction will produce caustic soda from saltpeter and sulfuric acid.

Both these statements are necessary truths and, hence, are necessarily equivalent.²⁵ The second states a law, but not the first. Only when the equivalence between two sentences is very strong can we infer the nomological character of one from the nomological character of the other. Williams's argument presupposes therefore a strong reading of the Infinite Conjunction Thesis.

Horwich responds to Putnam in a different way. He does not deny that (6) is a law. He argues instead that (6) is sufficiently explained by the T-biconditionals and, hence, that we do not need a substantial correspondence theory of truth to explain it. Horwich writes:

[I]t is clear, in general, how true beliefs contribute to practical success. Nothing beyond the minimal theory [which consists essentially of the T-biconditionals] is called for to explain this phenomenon.²⁶

The way the T-biconditionals explain (6), according to Horwich, is this. Suppose that

(11) Bill believes that he will get a beer if he nods

and that

(12) Bill wants a beer.

Sentences (11) and (12) explain Bill's nod. The truth of Bill's belief yields, in virtue of the T-biconditionals, that

If Bill nods, he will get a beer.

Hence, we obtain the conclusion that Bill will get a beer and, consequently, that his want will be fulfilled. Other examples of beliefs and desires may require a more complex explanation, but, as the above example illustrates, none will need a substantial theory of truth.²⁷

This argument needs the support of the Disquotation and the Infinite Conjunction Theses to work. The argument invites two challenges. First, it may be argued that even if the T-biconditionals explain (6), the possibility of a substantial theory of truth remains. It may be that a substantial theory of truth will provide a deeper explanation of the T-biconditionals and, consequently, of (6). Second, it may be argued that what Horwich proposes is an explanation only of the instances of (6), not of (6) itself.

The Disquotation Thesis provides a response to the first challenge. Since the T-biconditionals are definitional of truth, the response goes, they are not open to a deeper explanation;²⁸ the substantial theory of truth has no work to do. Observe that this response will not work on the weaker readings of the Disquotation Thesis. It will not work, for example, if all we have available is the thesis that the T-biconditionals are necessary truths. For, necessary truths *can* sometimes be given a deeper explanation. Sentence (10) expresses a necessary truth, yet chemistry provides a deep explanation of why it holds.

The Infinite Conjunction Thesis provides a response to the second challenge: Since the T-biconditionals explain all the instances of (6), they explain their infinite conjunction. But by the Infinite Conjunction Thesis, (6) expresses this infinite conjunction. Hence, the T-biconditionals explain (6) also. Again, observe that this response requires a strong reading of 'express'. The necessary equivalence of (6) with an infinite conjunction is insufficient ground for it. For to explain one of two sentences that are necessarily equivalent is not thereby to explain the other (see examples (9) and (10)).

We may conclude, then, that the deflationary arguments against a substantial theory of truth need the support of the Disquotation Thesis and the Infinite Conjunction Thesis read in a strong way; weaker versions of the theses are insufficient. This is a major weakness in the arguments, for on the strong reading the Infinite Conjunction Thesis is false. Williams's argument requires the generalization (6) to be equivalent to the infinite conjunction (8) in a sense strong enough to guarantee sameness of nomological character. But the two plainly are not equivalent in such a strong sense. The conjuncts of (8) are particular in character. So, (8) itself is particular in character. But this is not true of (6), which is general. Further, (6) gives us information about counterfactual situations that lie beyond the scope of the infinite conjunction (8).

Horwich's argument presupposes that a generalization is equivalent to the conjunction of its instances in a sense strong enough to guarantee that an explanation of one is an explanation of the other. But, as the following example shows, this is not true. We can explain each instance of the generalization "everyone on the boat died" by providing a separate explanation for the death of each person on the boat: Jack died of a heart attack; Mohini drowned; etc. But these separate explanations do not necessarily explain the generalization. The generalization may in fact have no explanation at all—it may be true accidentally. Or it may have an altogether different explanation, such as that the boat capsized. In any case, an explanation of the instances is not necessarily an explanation of the generalization.²⁹

We can accept the Infinite Conjunction Thesis when 'express' is understood as implying only material equivalence. We can even suppress several doubts and grant the thesis when 'express' is understood as implying necessary equivalence. But the thesis is false when 'express' is understood in the strong way needed in Williams's and Horwich's arguments.³⁰ I think the cause of error here is the same oversight that we found in Quine's passage in §1.2: a neglect of the distinction between "affirming the universal" and "affirming all the instances." Once the distinction is

neglected, it becomes easy to read the Infinite Conjunction Thesis in a strong way. Once the distinction is marked, the strong readings are seen to be plainly false.

In conclusion: The deflationary arguments against a substantial theory of truth presuppose an unacceptably strong reading of the Infinite Conjunction Thesis. I myself see nothing in the meaning and function of ‘true’ to rule out the possibility or the usefulness of a substantial theory of truth.³¹

1.4. Meaning and the Concept of Truth

The theory of meaning is another area in which the deflationists deny truth a substantial role. Two paradigms dominate philosophical studies of meaning. One paradigm seeks to understand meaning in terms of language-*world* relations. On this paradigm the concept of truth plays a central role in an account of meaning. Indeed, on many theories within this paradigm, meaning (of a sentence) is identified with truth conditions. The other paradigm seeks to understand meaning in term of language-*user* relations. On this paradigm language-world relations are not so central in an account of meaning. What is central is the *use* to which sentences are put. The debate between the two paradigms is large and of large significance. The deflationist contribution to the debate is the argument that the meaning and function of ‘true’ rule out a truth-conditional account of the meanings of sentences.

An early formulation of the argument occurs in Michael Dummett’s paper “Truth”:

[I]n order that someone should gain from the explanation that *P* is true in such-and-such circumstances an understanding of the sense of *P*, he must already know what it means to say of *P* that it is true. If when he enquires into this he is told that the only explanation is that to say that *P* is true is the same as to assert *P*, it will follow that in order to understand what is meant by saying that *P* is true, he must already know the sense

of asserting P , which was precisely what was supposed to be being explained to him. (p. 7)

Dummett goes on to write, in the concluding paragraph of his paper, that

[F]or most ordinary contexts the account of these words ['true' and 'false'] embodied in the laws 'It is true that p if and only if p ', and 'It is false that p if and only if not p ' is quite sufficient: but it means facing the consequences of admitting that this is the *whole* explanation of the sense of these words, and this involves dethroning truth and falsity from their central place in philosophy and in particular in the theory of meaning.³²

Dummett's argument brings out a tension between two ideas: the idea that the T-biconditionals explain the meaning of 'true' and the idea that meaning is to be explained in terms of truth conditions. If T-biconditionals are definitional of truth, if they explain what our understanding of 'true' consists in, then our understanding of 'true' presupposes a prior grasp of the meanings of the sentences of our language. Hence, truth cannot play a fundamental role in the theory of meaning; it cannot provide an explanation of our grasp of the meanings of sentences.

The tension is particularly vivid if one follows Donald Davidson and conceives of the theory of meaning for a language as a theory of truth for it.³³ The tension is now over two ways of reading the T-biconditionals: as elucidating the meanings of sentences and as elucidating 'true'. The two ways preclude each other. The former presupposes the concept of truth and uses the T-biconditionals to explain meaning; the latter presupposes meaning and uses the T-biconditionals to explain truth. By holding one element (truth or meaning) fixed, it appears, one can obtain the other. But one cannot use the T-biconditionals to extract both. As Horwich says, this is like having one equation and two unknowns.³⁴ Fixing one unknown we can solve for the other, but we cannot solve for both simultaneously.

Notice that these considerations provide an argument against truth-conditional semantics only when they are supplemented with the full force of the Disquotation Thesis. A weaker thesis such as that the T-biconditionals are necessarily true is insufficient. Not only is there no tension between this weaker thesis and truth-conditional semantics, the very formulation of truth-conditional semantics requires a sense of ‘true’ for which the weaker thesis holds. Consider an arbitrary sentence ‘ p ’ and an arbitrary possible situation w .³⁵ Truth-conditional semantics identifies the meaning of ‘ p ’ with its truth conditions, say, X . Now suppose w is in X . The very formulation of truth-conditional semantics requires that there be a sense of ‘true’ on which ‘ p ’ is true in w . Since w is in the truth conditions of ‘ p ’, the T-biconditional

‘ p ’ is true if and only if p

holds in w . By a parallel argument the biconditional holds also if w is not in X . Since w is arbitrary, the biconditional must be necessarily true. It follows that truth-conditional semantics requires a sense of ‘true’ on which the T-biconditionals are necessarily true.

The deflationary argument, if it is to work, requires the strong idea that the T-biconditionals explain the *meaning* or *sense* of ‘true’. This suggests the following picture of our acquisition of ‘true’: we first learn some first-order words (‘snow’, ‘white’, etc.) and then we arrive at ‘true’ definitionally through the T-biconditionals.³⁶ Given this picture, it follows immediately that we cannot explain our understanding of ‘snow is white’ in terms of our understanding of ‘true’, for our understanding of ‘true’, according to the picture, rests on our prior understanding of ‘snow is white’.

But now a basic difficulty with the argument comes into view. If anything like the above picture of the meaning of ‘true’ is correct, then an understanding of ‘true’ requires the possession of massive conceptual resources. For consider again the picture with which we are presented. We are told that we gain our understanding

of ‘true’ through the T-biconditionals, that we acquire ‘true’ by laying down the totality of T-biconditionals as definitional of ‘true’. But each biconditional plays an important role in the resulting definition: it defines what it means to apply truth to one particular sentence. If some of the biconditionals are omitted, the result is at best a partial definition of ‘true’. An individual who does not lay down some of the T-biconditionals as definitional of ‘true’ would have at best a partial notion of truth. To have a full notion of truth—to have a full understanding of the meaning of ‘true’—requires, on this picture, a grasp of all the T-biconditionals. But this is possible only if the individual possesses all the concepts expressed by the terms in the right-hand sides of the biconditionals. Hence, on the above picture of the meaning of ‘true’, a full understanding of ‘true’ is possible only for someone with massive conceptual resources.

An immediate response to this argument is to say that what defines ‘true’ is not the T-biconditionals, but something in their neighborhood—perhaps the form (T), or perhaps the general fact lying behind the T-biconditionals, or perhaps the rule of inference embodied in them. These suggestions are interesting but, as I shall argue in the next section, they do not provide a viable explanation of the meaning of ‘true’. Furthermore, they cannot play the role that the Disquotation Thesis plays in Dummett’s deflationary argument. Let us therefore set aside these suggestions for the moment and return to our original concerns: Should we think of the T-biconditionals as definitional of ‘true’? Do the T-biconditionals explain what our understanding of ‘true’ consists in? In short, is the Disquotation Thesis true? Let us approach these questions indirectly. Let us ask: What are we denying in denying the Disquotation Thesis?

(i) It is plain that we are not denying the T-biconditionals. Nor are we denying that the T-biconditionals are necessarily true. If the slogan “truth is a device of disquotation” is meant to say nothing more than this, then we are not denying the

slogan. But the slogan so read does not provide a foundation strong enough to support deflationism.

(ii) In denying the Disquotation Thesis, we need not even deny that there is a sense of ‘definition’ on which the T-biconditionals define ‘true’: we can accept the idea that the T-biconditionals fix the *extension*, and even the *intension*, of ‘true’.³⁷ What we deny is that the T-biconditionals fix the *sense* of ‘true’. When we evaluate a definition that aims to fix the extension or the intension of a predicate, we consider only how it carves the domain of its application into those objects that fall under the predicate and those that do not. The *ideology of* the definition, that is, the totality of the concepts employed in the definiens of the definition, is entirely irrelevant.³⁸ So, the fact that the ideology of the T-biconditionals is vast does not cast any doubt on the idea that the biconditionals fix the extension and the intension of ‘true’. But when we evaluate a definition that aims to capture the sense of a term, the ideology is of critical importance. For, the definition is now meant to capture what our understanding of the term consists in. If the definition is correct, a full understanding of the definiendum requires possession of the concepts in the definition’s ideology. Let the *ideology of* a term consist of those concepts that are necessary and sufficient for an understanding of the term (assuming that there is such a totality).³⁹ Then, it is an adequacy condition on a definition that aims to capture the sense of a term that the ideology of the definition coincide with the ideology of the term. It follows that a definition that aims to capture sense may be inadequate simply because of the ideology that it employs.⁴⁰ This explains why the T-biconditionals are not an adequate definition of the sense of ‘true’. If the T-biconditionals were adequate, then, given that their ideology is vast, it would follow that a full understanding of ‘true’ would require a massive repertoire of concepts. But, plainly, one can have a perfect understanding of ‘true’ even though one lacks, e.g., the concept of set or that of relativistic mass. The T-biconditionals fail

to define the sense of ‘true’ because they attribute much too large an ideology to ‘true’.

(iii) In denying the Disquotation Thesis, we are not denying the observation that lies at the foundation of deflationism: that in asserting “ ‘snow is white’ is true” one typically asserts nothing more nor less than “snow is white.” Deflationism goes on to explain this pragmatic fact in a certain way. And it is this explanation that we deny. According to deflationism, the pragmatic fact obtains because the sentences “ ‘snow is white’ is true” and ‘snow is white’ are synonymous, and the synonymy obtains because of the meaning of ‘true’. Deflationism thus explains the pragmatic fact solely on the basis of the meaning of ‘true’. But the deflationary explanation is not the only one possible, nor the most plausible. The pragmatic fact is sufficiently explained by the observation that in a typical situation the T-biconditional

(13) ‘snow is white’ is true if and only if snow is white,

is common knowledge, and indeed trivial common knowledge. Deflationism goes wrong because it reads the pragmatic fact into the very analysis of ‘true’.

(iv) This point is connected to the previous one. In denying the Disquotation Thesis we are not forced to deny that the T-biconditionals are trivial. Nor are we forced to deny that there is a sense of ‘analytic’ on which the T-biconditionals are analytic. We can grant, for example, that a person who knows the meanings of all the parts of (13) will thereby know that

‘snow is white’ is true if and only if snow is white.

We can grant therefore that there is a sense in which the T-biconditionals are “true solely in virtue of meaning.”⁴¹ But this is not to say that the T-biconditionals are “true solely in virtue of the meaning of ‘true’ ”; that someone who knew only the meaning of ‘true’ would thereby know the biconditionals. I think the point is important because the intuitive pull of the Disquotation Thesis comes from the seeming triviality and analyticity of the T-biconditionals. This makes us think that

the T-biconditionals explain the meaning of ‘true’ and that an adequate definition of ‘true’ must imply the biconditionals (Tarski’s Convention T). But in thinking thus we make an unwarranted leap, a leap from common sense to deflationism.

In denying the Disquotation Thesis, then, we are not denying any of our commonsensical ideas about truth. We are denying a very specific claim about the meaning of ‘true’, a claim that plays a central role in the deflationary argument from Dummett considered above. And we are denying a picture of how we arrive at our understanding of ‘true’, a picture that makes the deflationary attitude compelling. Once we shed the claim and the picture nothing remains, I believe, to make plausible the deflationary attitude in the theory of meaning.

1.5. The Meaning of ‘True’

The T-biconditionals make it tempting to think that the concept of truth is simple, that a complete analysis of the meaning of ‘true’ is easily given. Philosophers readily grant that analysis of meaning is, in general, a difficult task: even the meaning of such a simple word as ‘table’ is difficult to specify. But when it comes to ‘true’ the T-biconditionals make it tempting to suppose that a reductive analysis of its meaning is possible. Even if one accepts the point that an explanation of the meaning of ‘true’ should not employ the massive ideology of the T-biconditionals—and that therefore the T-biconditionals themselves do not explain the meaning of ‘true’—the thought persists that something in the neighborhood of the T-biconditionals does explain it. What matters to the meaning of ‘true’, one thinks, is not the details of the particular T-biconditionals, but the general idea captured by them. One is tempted to say that the meaning of ‘true’ is explained by the *form* (T),

(T) ‘___’ is true if and only if ___,

not by the particular biconditionals. And evidently the form does not carry with it a heavy ideology.

But how does a form explain the meaning of a predicate? This type of explanation of meaning is quite different from the usual sort. Form (T) does not explicitly state the application conditions of 'true' (otherwise it would not have overcome the ideology problem), but this is what we expect from an explanation of the meaning of a predicate.⁴² So how does (T) constitute an explanation of the meaning of 'true'? Several approaches suggest themselves as ways of answering this question. Let us examine a few. Our examination will cast doubt on the idea that a reductive analysis of 'true' is possible.

(A) The Generalization Approach. This approach tries to make sense of the idea that (T) explains the meaning of 'true' by appealing to the general truth corresponding to (T). What explains the meaning of 'true', on this approach, is not the totality of the T-biconditionals but the general fact that

(GT) All instances of the form (T) [i.e., all T-biconditionals] are true.

The suggestion has some attractive features. It explains the meaning of 'true' using a formula whose ideology is highly limited. Moreover, the formula is plainly analytic of the terms it involves. Anyone who understands the meaning of 'form (T)', 'true', etc., must grant the truth of (GT).

Unfortunately, however, the suggestion faces an obvious but overwhelming problem. It explains the meaning of 'true' using a formula that itself involves 'true'. The circularity is not intrinsically objectionable.⁴³ But the particular form it takes here violates material aspects of the meaning of 'true': the proposal fails to yield the T-biconditionals. Imagine we give (GT) as an explanation of 'true' to someone who does not yet understand the word. This person will be able to deduce from (GT) that

(14) 'Snow is white' is true if and only if snow is white' is true.

But how can he eliminate the last occurrence of 'true' and arrive at the T-biconditional

(15) 'Snow is white' is true if and only if snow is white?

To eliminate it, he needs to derive the T-biconditional

- (16) ‘Snow is white’ is true if and only if snow is white’ is true if and only if
[‘snow is white’ is true if and only if snow is white].

But, again, (GT) does not yield (16) but only that (16) is true. A parallel difficulty blocks attempts to eliminate this new unwanted occurrence of ‘true’. Our imaginary learner can derive of any T-biconditional that it is true, but he cannot derive the T-biconditional itself.

Note that if we *presuppose* the notion of truth, then the present strategy is a good way of spelling out the idea that a form explains the meaning of a predicate. The trouble is that the strategy works only if the meaning of ‘true’ is taken as given, not otherwise. The strategy cannot therefore be used to explain the meaning of ‘true’. Rather, the meaning of ‘true’ is needed to make sense of the strategy.

(B) The Syntactic Approach.⁴⁴ This approach reads (T) as expressing a syntactic rule, a rule to the effect that a declarative sentence can be transformed by the addition (and deletion) of the marks,

- (17) ‘ ’ is true,

without altering the sense of the original sentence. The approach thus views addition of the marks (17) as analogous to the passive transformation. Both transformations have a limited utility, but are insubstantial modifications of the original.

The syntactic approach gets around the ideology problem. And, it, unlike the previous approach, avoids problematic circularity in its explanation. However, it cannot be regarded as explaining the meaning of ‘true’. At best, the approach explains the meaning of (17) *when (17) is viewed as one syntactic unit*; it does not explain the meanings of the parts of (17). The approach does not even entitle us to treat ‘is true’ as a predicate. It therefore does not explain the role ‘is true’ plays when it occurs in combination with pronouns and general terms. If we follow

the syntactic approach, we face problems explaining such simple inferences as the following:

The sentence Bill uttered is 'snow is white'; 'snow is white' is true; hence, the sentence Bill uttered is true.

The approach does not entitle us to treat the occurrence of ' 'snow is white' ' in the second premise as a singular term. Consequently, we cannot explain the inference as an instance of Leibniz's principle of indiscernibility of identicals.

(C) The Inferential Approach. This approach uses the idea that the meaning of certain items in our language is specified by their inferential roles. The meaning of the truth predicate, it is suggested, is given by the rules of inference embodied in (T): to infer ' '___' is true' from ' ___' and, conversely, to infer ' ___' from ' '___' is true'.

This approach to the explanation of meaning, while attractive for some parts of our language, is distinctly less attractive when applied to the truth predicate. For, if truth is explained in terms of inference, how do we explain our understanding of inference? How do we explain inference without appeal to the notion of truth? The natural response to the query is as follows. Inference is to be explained in terms of its role in our conceptual practices—practices of assertion, denial, supposition, verification, etc. These practices, the suggestion goes, are governed by various norms, and an explanation of inference will specify the role that it plays in these norms. Thus, our understanding of inference consists in understanding such things as that if q can be inferred from p then the assertion of p commits one to q , that one cannot assert p and also deny q , that a verification of p counts as a verification of q , etc.

Obviously, the suggestion is viable only if one can explain "assertion," "denial," "commitment," etc., without appeal to truth. But can this be done? How is one to make sense of our conceptual practices without any appeal to the notion of

truth? The inferential approach remains a large promissory note until it provides a satisfactory answer to this question. I will not attempt to speculate on how the approach might be developed, but I would like to make two remarks about it.

First, the inferential approach to meaning does not need to forgo the notion of truth in order to stay true to its philosophical motive. *Use of the notion of truth in an explanation of our conceptual practices (and of meaning) does not immediately commit one to a referential picture of language.* It seems to me that the burden of explaining truth for a body of discourse in non-referential terms is lighter than that of giving a reductive explanation of the concept of truth. The inferential approach to meaning, it seems to me, takes a wrong turn when it denies itself the use of truth and takes on the burden of explaining the meaning of truth in inferential terms.

Second, even if truth does not play a substantive role in the explanation of our conceptual practices, it most likely does play an expressive role in their description. That is, truth is probably needed to describe basic facts about our conceptual practices—facts which are constitutive of them. For example, a description of our understanding of inference will, in all probability, need to mention our knowledge of the general fact of which the following is an instance.

If ‘snow is white’ can be inferred from ‘everything is white’,
then snow is white if everything is white.

But how else can one express the general fact than by using the truth predicate: If a sentence A can be inferred from a sentence B, then A is true if B is? If this thought is correct, then the prospects for an inferential approach to the meaning of ‘true’ are bleak indeed.

None of the above approaches, then, is likely to yield a viable account of the meaning of ‘true’. Let us observe also this. Even if these approaches were to lead to a viable account, it is an open question whether the resulting account would support any deflationary claims, and, if it did support some, it is an open question which

ones it would support. For example, suppose that the inferential approach overcomes the obstacles in its way and offers an acceptable account of the meaning of 'true'. This account will not, as far as I can tell, support a blanket deflationism in the theory of meaning. The account plainly could not play the role that the Disquotation Thesis played in the deflationary argument from Dummett considered in §IV. That argument rested on the idea that our understanding of '____' is true' presupposes an understanding of '____'. The inferential approach, far from supporting this idea, is designed to overcome the problem that the idea creates. Further, the account the inferential approach proposes will, presumably, explain our understanding of 'true' in terms of our understanding of a limited range of terms, conceptual practices, etc. It is intuitively plausible that 'true' could not be used to provide an explanation of our understanding of the terms, practices, etc., within this range. But this allows truth to play an important role in an explanation of those terms and conceptual practices that lie outside the range. The account therefore will not make plausible a blanket deflationism in the theory of meaning.

Let us return to the original, disquotational, account of truth and take stock. The account, to review briefly, goes as follows. "The usefulness of truth lies in the expressive power that it provides. The truth predicate, by providing us with an indirect means of quantifying over sentence positions, enables us to express certain infinite conjunctions and disjunctions. To perform this function truth must be a device of disquotation: applied to a quoted sentence it must undo the effect of the quotation marks. This function therefore requires the T-biconditionals to be definitional of 'true'." There are readings of this account on which it contains no errors, only insights. The key is how we understand "express," "device of disquotation," and "definition." Suppose we understand them extensionally. Then the account reads (in part): "The generalizations involving truth are materially equivalent to the corresponding infinite conjunctions/disjunctions. To ensure this equivalence,

truth needs to be a device of disquotation in the sense that the T-biconditionals need to be true. The T-biconditionals define ‘true’ in the sense that they fix the extension of ‘true’.” There is also an “intensional” reading of the disquotational account. This reading is parallel to the one just given, but it takes “express” to require necessary equivalence, “device of disquotation” to require necessary truth of the T-biconditionals, and “definition” to require the fixing of intension. The disquotational account, when it is read in either of these ways, is good, true, and insightful. The only point that I have insisted on is that on these readings the account is not strong enough to carry the burden of deflationism.

Deflationists read (and, to sustain their conclusions, need to read) the disquotational account in very strong ways. Here is one such reading:⁴⁵ “The generalizations involving truth are abbreviations for (and, hence, mean the same as) the corresponding infinite conjunctions/disjunctions. To ensure this equivalence, truth needs to be a device of disquotation in the sense that guarantees the synonymy of ‘‘___’ is true’ and ‘___’. Thus, if ‘true’ is to perform its function, the T-biconditionals must be definitional in the sense that they explain the meaning of ‘true’.” The reading goes on to add: “This makes truth a simple concept. What it means, what our understanding of it consists in, and how we acquire it—these all have a simple explanation. The meaning of ‘true’ is given by the T-biconditionals, our understanding of it consists in our acceptance of the T-biconditionals, and we acquire it by laying down the biconditionals as its definition.”

The objection from ideology, given in §1.4, puts in doubt each element of this reading. The T-biconditionals do not provide an adequate account of the meaning of ‘true’ because they impute to ‘true’ a massive ideology. The sentences ‘‘___’ is true’ and ‘___’ are not always synonymous, for the concepts needed to understand the latter are not necessarily needed to understand the former. The generalizations involving ‘true’ do not mean the same as the corresponding infinite

conjunctions/disjunctions, for again the two do not involve the same conceptual resources.

It is remarkable that not only do the deflationary claims fail, but that an explanation of the usefulness of 'true' lies in their failure. An example: One important reason why generalizations involving 'true' are useful is precisely that they do *not* mean the same as the corresponding infinite conjunctions/disjunctions. Consider the generalization 'all men are mortal' for comparison. One reason why this generalization is useful is that it enables us to express a fact about all men without requiring of us the ability to say of each man that he is mortal. The generalization expresses, in a weak sense, the conjunction of its instances *without* being synonymous with the conjunction. It thus enables us to express (weakly) a conjunction that we lack the resources to formulate explicitly—here lies its usefulness. The same holds for generalizations involving 'true'. One reason for their usefulness is that they are not synonymous with the corresponding infinite conjunctions/disjunctions. They allow us to express (weakly) these infinite conjunctions/disjunctions, even though our conceptual resources are meager.

More generally, one important source of the usefulness of 'true' is its remarkable double character: (i) that an understanding of 'true' requires only a limited range of conceptual resources, and yet (ii) 'true' enables us to talk and think about things that lie far beyond this range. (So, one reason 'true' is useful is precisely that the T-biconditionals do not define its sense.) This double character also constitutes, it seems to me, the most fundamental mystery of truth. The meaning of 'true', like that of many other words, is difficult to explain; it becomes mysterious when we consider what 'true' enables us to do. 'True' appears simple to the deflationists, I think, because they overlook its most puzzling feature.

I have focused in this essay on homophonic truth ("true in English") because it is here that the deflationist account appears most plausible. And I have tried to show that even here it fails. When we turn to heterophonic truth ("true in Inuit,"

“true in such and such an infinitary language”) the problems facing deflationism become more vivid, but in essence remain the same. Suppose we have somehow gained ‘true in English’, and suppose L is a language that can express things that are inexpressible in English (perhaps L is spoken by some alien creatures; perhaps L is an infinitary language that we find theoretically useful to talk about). How then can we gain ‘true in L ’ when all we have to work with is ‘true in English’? We cannot say that a sentence of L is true iff it has a true translation in English, for this will make all untranslatable sentences of L untrue. How then will the explanation go?⁴⁶ Deflationism needs to explain ‘true in L ’ without using the conceptual resources of L . The problem that must be solved is in essence the same as the fundamental problem we pointed to above. How to reconcile two features of ‘true in L ’: (i) that it enables us to talk about the inexpressible contents of L , but (ii) the explanation of its meaning does not appeal to those contents?

In conclusion: Deflationists think that truth is a simple concept, one that has a simple analysis. The analysis the deflationists offer *is* simple but, unfortunately, it makes truth far too complicated—it attributes to truth a vast ideology. We examined several attempts to get around this problem, but none resulted in a plausible account of the meaning of ‘true’. Now we are left with questions: *What does our understanding of ‘true’ consist in? How can one explain the meaning of ‘true’ using a limited ideology?* It is a fact that we understand truth attributions even when truth is attributed to a sentence (or thought or representation) that lies beyond our conceptual resources. *What do we understand by such attributions?* We seem to grasp something general about what it is for a sentence (or thought or representation) to be true. *But what is it that we understand?* Once we overcome the spell of deflationism we are no longer inclined to brush these questions aside with simple answers. We regain our original sense that there is something very mysterious about truth and that an exploration of this mystery may illuminate the nature of our thought and our language.

Postscript

I have argued above for two claims. First, deflationism can sustain its metaphilosophical theses about the role of truth only if it makes strong descriptive claims about the concept of truth. Second, the strong descriptive claims are false. Deflationism gains its plausibility, I suggested, through ambiguities in its descriptive claims. There are readings of the claims that are perfectly acceptable, but they are too weak to sustain the metaphilosophical theses. On the other hand, there are readings that sustain the metaphilosophical theses, but they are much too strong to be true.

Paul Horwich and Christopher Hill have responded to two of my objections to deflationism. The first objection is directed to the deflationary thesis that the T-biconditionals explain all the facts about truth. I argued (in §1.3) that there are generalizations about truth that the T-biconditionals fail to explain. The biconditionals explain at best only the instances of these generalizations, not the generalizations themselves. The cause of the deflationists' error here, I suggested, lies in the confusion of "affirming the universal" with "affirming its instances"—a confusion that is evident in the strong readings of the Infinite Conjunction Thesis. Horwich has responded to this objection and has offered a way of sustaining the deflationary thesis.

The second of my objections is directed to the thesis that the T-biconditionals fix the sense of 'true'. I argued that this thesis cannot be correct because it attributes to truth a massive ideology. It implies that to fully grasp the sense of 'true' one would need to have a grasp of all the concepts in this massive ideology—something that is plainly false.⁴⁷ Hill has responded to this objection.

In this "Postscript" I assess Horwich's and Hill's responses. Let me first get a preliminary point out of the way. My objections above were directed to a deflationary account of *sentential* truth. The responses of Horwich and Hill are directed

to defending a deflationary account of *propositional* truth. There is thus a mismatch. It is, however, of little consequence, since the two deflationary accounts run parallel to one another. For example, one account takes the instances of the schema

(T) '____' is true if, and only if, ____

to fix the sense of sentential truth, while the other takes the instances of the schema

(T*) that ____ is true if, and only if, ____

to fix the sense of propositional truth. For another example, one account takes the instances of (T) to explain all the facts about sentential truth, while the other takes the instances of (T*) to explain all the facts about propositional truth. There are, of course, some significant differences between the two accounts, but as far as our present concerns go, the differences between them can be ignored. The objections I offered carry over to the deflationary account of propositional truth,⁴⁸ and the responses of Hill and Horwich can be translated into a defense of the deflationary account of sentential truth.

Horwich on the generalization problem. Horwich argues that the instances of (T*) can explain generalizations about truth. He writes,

[I]t is plausible to suppose that there is a truth-preserving rule of inference that will take us from a set of premises attributing to each proposition some property, F, to the conclusion that all propositions have F We commit ourselves to [this rule], implicitly, in moving from the disposition to accept any proposition of the form '*x* is F' (where *x* is a proposition) to the conclusion 'All propositions are F'. So we can suppose that this rule is what sustains the explanations of the generalizations about truth with which we are concerned. Thus we can, after all, defend the

thesis that the basic theory of truth consists in some subset of the instances of the equivalence schema [(T*)].⁴⁹

The truth preserving rule Horwich invokes is

(R) That p_1 is F, that p_2 is F, that p_3 is F, . . . ; therefore, all propositions are F.

The premisses of this rule form an infinite totality—indeed a totality so large that its size cannot be measured by any of the alephs.⁵⁰ For each proposition p , the totality contains the premiss that p is F. From this infinite totality the rule allows us to derive the conclusion that all propositions are F. Horwich claims that (a) this rule makes possible an explanation of generalizations about truth, (b) we are implicitly committed to the rule, and finally (c) the rule sustains the thesis that some instances of (T*) constitute the basic theory of truth.

Horwich's move illustrates the deflationists' dilemma that I tried to highlight above: whether to argue for the strong or for the weak versions of their descriptive claims. The former can enable the deflationists to reach their goals, but they are difficult to establish. The latter, on the other hand, are easier to establish, but they do not help the deflationists reach their goals. The thesis that the instances of (T*) can explain, *in the presence of* (R), generalizations about truth is a weak version of Horwich's original explanatory claim. The weaker claim is easier to establish, but it will not sustain Horwich's deflationary ambitions.

Horwich holds that his Minimal Theory—which consists of (some of) the instances of (T*)—is the *basic* theory of truth: no reduction is possible of the Minimal Theory to a more fundamental theory.⁵¹ This thesis is important to Horwich because he wants to rule out the search for a more substantive theory of truth. Now, the thesis can perhaps be sustained if the original, strong version of Horwich's explanatory claim were in play. But it cannot be sustained on the basis of the weak version defended by Horwich. Plainly, a theory of truth that explains the phenomena *without appeal* to (R) is a better and more fundamental theory than Horwich's

Minimal Theory, which has to resort to (R) in its explanations. Moreover, such a theory remains more fundamental even if the scope of its explanations extends no further than that of the Minimal Theory and even if it has, like the Minimal Theory, infinitely many axioms and a massive ideology. For the kinds of explanation theories give of phenomena are relevant to assessing their relative merits and statuses. The explanations provided by the Minimal Theory are of a low grade; they leave open the possibility that a different theory would provide explanations of a higher grade. And this is no mere possibility: Tarskian theories of truth provide explanations of a range of generalizations about truth without invoking any infinitary rule. These explanations are plainly superior to those provided by the Minimal Theory with the aid of (R). So, even if we grant Horwich the weaker version of his explanatory claim, we can nevertheless reject his deflationary thesis that the search for a substantive theory of truth is futile.

Another idea that the weaker explanatory claim fails to sustain concerns the meaning of 'true'. Horwich maintains—and needs to maintain—that the instances of (T*) explain our use of 'true'. The question naturally arises: how can a plausible explanation of use appeal to an infinitary rule? Parallel question: how can our *acceptance* of arithmetical claims be explained on the basis of the ω -rule ("0 is F, 1 is F, ...; therefore, all natural numbers are F")? Plainly, something like Peano Arithmetic has a better chance of explaining our acceptance of arithmetical claims⁵² than a theory consisting of true numerical equations and the ω -rule. (The latter theory does not have the resources to explain our failure to accept certain true arithmetical generalizations.) So it is more plausible to say that Peano Arithmetic constitutes an implicit definition of natural number than to say that true numerical equations do so. Horwich's claim (b) is perhaps meant to address this difficulty. But I do not see that our practices contain, even implicitly, the sort of commitment to (R) that will sustain the idea that the instances of (T*) explain our use of 'true'.

Finally, Horwich's claim (c) reveals an obvious but insurmountable problem with his invocation of (R). Claim (c) is important to Horwich because he takes the theory of truth to contain only some of the instances of (T*), not all of them: Horwich excludes the instances for paradoxical propositions such as the Liar. The problem is that Horwich's theory of truth does not entail the requisite generalizations about truth even if we allow the use of (R). Consider, for example, the generalization

(18) If a proposition is true then the conjunction of the proposition with itself is also true.

To deduce (18) using the rule (R) we need each instance of the formula

(19) If x is true then $(x \& x)$ is true.

Suppose that A is a Liar-like proposition for which the instance of (T*) is excluded from the theory of truth. Now, although we shall be able to derive

If A then A and A ,

which is a law of logic, we shall be unable to derive

(20) If that A is true then (that A & that A) is true,

because we lack the relevant instances of (T*).⁵³ Thus, we shall not have the requisite premisses to sustain an application of (R) and, therefore, we shall be unable to derive (18).⁵⁴

Hill on the ideology problem. Hill suggests the following defense of the claim that the instances of (T*) define the concept of (propositional) truth. We admit that the correct definition of truth imputes to it a vast ideology, that each of the instances of (T*) is essential to the definition. But we reject the following principle, which Hill calls the *Intimacy Principle*:

If a class of propositions constitutes the correct definition of a concept, then, in order to possess the concept, an agent must

stand in an intimate cognitive relationship to each individual proposition in the class.⁵⁵

Hill thinks that the Intimacy Principle is plausible only when the propositions constituting the definition are formally disparate, but not when they all have the same form.⁵⁶ He maintains that so long as the agent is “cognitively linked” to the form that the propositions share, the agent counts as possessing the concept. Hill illustrates the point with the concept of classical negation. He writes of the inference pattern

It is not the case that it is not the case that p ; therefore, p

that

It is plausible that the concept of classical negation is partially defined by inferences that have this form. Moreover, it is plausible that every such inference counts as a component of the definition. . . . [I]n many cases, . . . constituent concepts [of these inferences] include concepts that lie altogether outside the ken of human agents. Yet it is clear that we are fully prepared to attribute the concept of classical negation to virtually the entire family of human agents.⁵⁷

Like Hill, I too want to reject the Intimacy Principle on some readings of the expression ‘definition of concept’. If the definition of a concept aims to fix only the extension or the intension of the concept, then the Intimacy Principle is false. If, however, the definition aims to articulate the concept in a stricter way—e.g., in a way that will let concepts serve as Fregean senses of terms—then the Intimacy Principle may well be true.⁵⁸ I argued above that a stricter articulation of concepts is needed to sustain various deflationary theses.

Hill says that an agent counts as possessing the concept of truth if he is “cognitively linked” to the form (T*). But when is an agent “cognitively linked” to the

form? Hill's answer here is complex. Roughly, Hill requires that an agent satisfy the following condition: the agent should be disposed to accept the instances of (T*) for all propositions that the agent can grasp, and should be disposed to do so "without empirical evidence and without supporting *a priori* argumentation."⁵⁹ This requirement seems to me to be simultaneously too strong and too weak. An agent should be forgiven for wanting supporting argumentation before accepting some of the instances of (T*)—for example, those for paradoxical propositions. But Hill's requirement would unforgivingly deem such an agent to lack the concept of truth. On the other hand, it would be forgiving of an agent who insists that only propositions he can grasp are true. Such an agent can meet Hill's requirements and thus count as possessing the concept of truth.

I do not wish to question Hill's suggestion that we can grasp a concept through "cognitive linkage" to a form, say F . I wish to question only his explanation of what the appropriate "cognitive linkage" amounts to. Plainly, the correct explanation should ensure that cognitive linkage to F commits the agent to all the instances of F , but it should not require the agent to grasp all those instances. The proper way to achieve this is simple: the cognitive linkage to form F is just the acceptance, perhaps *a priori*, of the claim that all propositions of the form F are *true*. To explain the grasp of the concept of classical negation through some inference patterns XYZ is to explain this grasp through the acceptance of the claim that all inferences that exemplify the patterns XYZ are valid, i.e., are necessarily truth preserving. We can explain what it is to grasp a concept via a cognitive linkage to a form *if* we invoke truth. Truth allows us to capture correctly the commitments undertaken by a concept user without burdening the user with a massive repertoire of concepts.

Once truth is invoked to explain concept possession, we have no reason to doubt the Intimacy Principle. If a concept can be grasped via a form F , then it can also be defined via the form F : the concept is defined by the principle that all

propositions of the form F are true.⁶⁰ The possession condition of the concept now matches its (implicit) definition. Of course, this sort of procedure is unavailable for the concept of truth. As I suggested in §1.5, the concept of truth is needed to make sense of the idea that a form defines a concept, not the other way around. We cannot explain the concept of truth via the idea of form.

Notes

¹See, for instance, Rudolf Carnap, *Introduction to Semantics*.

²Michael Williams, “Epistemological Realism and the Basis of Scepticism,” p. 424.

³I shall base my account of deflationism on the writings of a number of philosophers. I want to emphasize that while there are important similarities in the ideas of the philosophers I rely on, there are also important differences. No views, unless explicitly attributed to the individual authors, should be ascribed to them.

⁴Richard Rorty, *Consequences of Pragmatism*, p. xiii.

⁵Scott Soames, “What Is a Theory of Truth?,” p. 429.

⁶Paul Horwich, *Truth*, p. 54.

⁷Deflationists have offered several closely related descriptions of ‘true’. In this paper I choose to focus on just one description, that contained in the disquotational account. Nevertheless, the arguments developed below apply in a straightforward way to many other deflationary descriptions. One notable exception is the strand of deflationism that relies on the Prosentential Theory of Truth of Dorothy Grover, Joseph Camp, and Nuel Belnap. A development of this strand can be found in Dorothy Grover’s essays in *Prosentential Theory of Truth* and in Robert Brandom’s “Pragmatism, Phenomenalism, and Truth Talk.” My view is that the Prosentential Theory has important insights into the logical grammar of truth. But these insights

need to be supplemented with subsidiary theses before we can derive deflationary conclusions from them. I would want to argue that the subsidiary theses are problematic.

⁸Although Quine's writings have provided much inspiration to the deflationists, a reasonable case can be made that Quine himself is no deflationist. First, the concept of truth seems to play a substantial role in Quine's philosophy of logic. Second, Quine takes a skeptical attitude towards many of the notions used in the defense of deflationism.

⁹Stephen Leeds, "Theories of Reference and Truth," p. 121.

¹⁰The presence of ambiguity, context-sensitivity, self-reference, etc., in our language poses a challenge to the disquotational account. It forces us to recognize, for instance, that truth is not a simple predicate of sentences. I shall assume, for the sake of argument, that the deflationists can meet the challenge. I shall often write as if the problematic elements were not present in our language. Also, when the context allows it, I shall suppress relativity to language. I shall write 'true' in place of the longer 'true in English'.

¹¹I suppose I should call this thesis 'The Infinite Conjunction and Disjunction Thesis', but I want to save a few syllables.

¹²See Horwich, *Truth*, pp. 52 and 127. Recall also Quine's statement, "we *need* [a disquotational truth predicate] to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent" (emphasis added).

¹³Sometimes the notion "T-biconditional" is understood in a wider sense. This allows a nonquotational name of a sentence to appear in the left-hand side of the biconditional and a translation of the sentence to appear in the right hand side. Tarski constructed a definition (for certain languages) that implies the T-biconditionals in this wider sense. Since the definition implies the biconditionals, there could be no doubt that it was coextensive with truth. This refuted the skepticism of the

Positivists. At the same time it made it seem that truth was a clear and simple notion. This paved the way for modern-day deflationism. John Etchemendy's paper "Tarski on Truth and Logical Consequences" contains a good account of how Tarski's definition can be read in a deflationary way.

¹⁴Observe that the mere truth (or even the necessary truth) of the T-biconditionals will not yield that disquotation is a "formal feature" of the truth predicate (Williams); nor will it yield the interdeducibility of the two sides of the T-biconditionals (Leeds); nor Quine's claim that "[b]y calling the sentence ['snow is white'] true, we call snow white."

¹⁵I put in the qualification "at least partially" because a full explanation of 'true' may require not only the T-biconditionals but also some such claim as "only sentences are true." I shall sometimes take the qualification as read and will not state it explicitly.

¹⁶Horwich actually states this for a propositional notion of truth. But he wants to give a parallel account of the sentential notion. See Horwich, *Truth*, pp. 36–38, 52, 116, and 125.

¹⁷Perhaps this explains why opponents of deflationism have been on the defensive in recent years.

¹⁸Hilary Putnam accepted at one time an account of 'true' similar to the one sketched by Field. See his *Meaning and the Moral Sciences*, pp. 15–17. Putnam's arguments against deflationism can be found in, among other places, his "On Truth," and in "Does the Disquotational Theory Really Solve All Philosophical Problems."

Christopher Hill is another philosopher who accepts parts of the disquotational account; see his "Rudiments of a Theory of Reference."

¹⁹Hartry Field, "Tarski's Theory of Truth," p. 104. This paper no longer represents Field's present position. Field is now much more sympathetic to deflationism; see his "Deflationary Conception of Truth."

²⁰This causes a problem for any attempt to derive the strong reading of the Infinite Conjunction Thesis—and, more specifically, in deriving the synonymy of (2) and (4)—from the Disquotation Thesis. The Disquotation Thesis yields, we can grant, that (2) and (3) are synonymous. But to derive that (2) and (4)—are synonymous we need the synonymy of (3) and (4), which unfortunately does not hold.

²¹See his “Theories of Reference and Truth.”

²²See *Meaning and the Moral Sciences*.

²³Putnam rejects the idea that causal-explanatory laws are a reason to expect a physicalistic reduction. So, while Putnam thinks, *pace* Leeds, that truth is causal-explanatory, he rejects Field’s quest for a physicalistic reduction of truth. See *Meaning and the Moral Sciences*, lects. 3–5.

²⁴Michael Williams, “Do We (Epistemologists) Need a Theory of Truth?,” p. 232. The formulation of (6) that Williams is commenting on is this: “If we have true beliefs about how to attain our goals, we will generally attain them.” Williams gives one other argument for not regarding (6) as a law. Since this argument does not rely on the disquotational account of ‘true’, I will not consider it here.

²⁵I am assuming here the Kripke-Putnam theory of reference.

²⁶Horwich, *Truth*, p. 45.

²⁷See Horwich, *Truth*, pp. 23–24 and 44–47 for a fuller account.

²⁸Horwich, *Truth*, p. 52.

²⁹An analogy may make it clear that Horwich’s argument is unsuccessful. Consider the generalization:

In aircrafts with autoland systems, accurate instrumentation promotes safe landings.

If Horwich’s argument were successful then one could give a parallel deflationary explanation of this generalization and could argue that no further substantial explanation should be expected. But the deflationary explanation here is plainly unsatisfactory. An adequate explanation of why accurate instrumentation promotes safe

landings would show how the actions of the autoland system are linked with the readings of the instruments and how these actions affect the flight behavior of the aircraft.

³⁰It seems to me that the Infinite Conjunction Thesis is false also if ‘express’ is taken to mean “interdeducible” or “warranted on the same occasions” (see the extract from Leeds’s paper “Theories of Reference and Truth” given in §1.2). Let z be a set that contains sentences of a certain form. Then, the generalization ‘all members of z are true’ is not interdeducible with (nor is it warranted on the same occasions as) the infinite conjunction of the members of z . Neither the infinite conjunction nor the generalization carries information about what all the members of z are. But this is needed if we are to deduce one from the other.

³¹To avoid misunderstanding let me say explicitly that I am not here defending a correspondence, or a physicalistic, or any other particular theory of truth. What I am defending is the claim that the meaning of ‘true’ does not make the search for a substantial theory of truth futile.

³²Dummett, “Truth,” p. 19. Dummett later rejected this conclusion. See the preface to *Truth and Other Enigmas*, esp. xx–xxii.

³³For Davidson’s views on the theory of meaning see his essays in *Inquiries into Truth and Interpretation*. Davidson criticizes deflationism in “The Structure and Content of Truth.”

³⁴Horwich, *Truth*, p. 71.

³⁵Let us understand the variable p substitutionally in this argument.

³⁶Recall the extract from Field’s paper “Tarski’s Theory of Truth” given in §1.2.

³⁷I follow Carnap and think of “intension” as “extension-across-possible-situations.” For a theory of how the T-biconditionals fix the extension (more precisely, the signification) of ‘true’, even in the presence of vicious self-reference, see Nuel Belnap’s and my *Revision Theory*.

³⁸Quine uses ‘ideology’ in an analogous, though not identical, way.

³⁹I shall understand “ideology of a concept” in a parallel way.

⁴⁰Here is a simple illustration. Consider the definition,

x is a human iff x is an animal with such and such a DNA structure.

Supposing that the ‘such and such’ is properly filled out, the definition fixes correctly the intension of ‘human’—the ideology of the definition is irrelevant to this assessment. But when we consider whether the definition captures the sense of ‘human’, the ideology is highly relevant. The fact that the ideology of the definition includes the concept “DNA structure” makes the definition implausible as an explanation of the sense of ‘human’.

⁴¹In this sense of ‘analytic’, some analytic truths are open to substantial explanations. Contrast the analytic truths, “all bachelors are males” and “‘snow’ means snow.” It is unreasonable to expect a substantial answer to the question “why are all bachelors males?” but not to the question “why does ‘snow’ mean snow?”

⁴²I am setting aside circular predicates here.

⁴³As I argue in ch. 3 below.

⁴⁴Mark Wilson and Eric Dalton, independently, suggested this approach to me; they do not endorse it.

⁴⁵This is not the only possible strong reading, but it will highlight the points made earlier.

⁴⁶This kind of problem rules out the most obvious deflationist response to the ideology objection. According to the response, each of us learns “true” first, not as it applies to English, but as it applies to one’s own personal idiolect. That is, one first acquires the concept “true-in-my-present-idiolect” and then using it acquires the full-fledged “true.” The problem of explaining how one goes from “true-in-my-present-idiolect” to “true” seems to me to be much harder than that of explaining “true” using a limited ideology. The response reduces a very hard problem to a virtually impossible one.

⁴⁷See §1.4 and my “Minimalism.”

⁴⁸See my “Minimalism.”

⁴⁹“Minimalist Conception of Truth,” pp. 254-255. See also Horwich, “A Minimalist Critique of Tarski,” pp. 90-91 and p. 96, fn. 16.

⁵⁰Horwich errs in calling the totality of the premisses a *set*.

⁵¹“Minimalist Conception of Truth,” p. 255.

⁵²Of course, even here there are major problems.

⁵³The theory will not imply (20) because it leaves the behavior of truth completely indeterminate on the proposition that *A* and its self-conjunction.

⁵⁴Horwich offers a somewhat different response to the generalization problem in “A Defense of Deflationism.” I believe that the principal points made above hold *mutatis mutandis* for this response also.

⁵⁵*Thought and World*, p. 68. Hill’s own theory of truth is quite different from Horwich’s Minimal Theory. Hill makes the suggestion *en passant* in the course of developing his own view. I have heard several responses to the ideology problem. Hill’s is the best I have heard so far.

⁵⁶Hill can, and should, strengthen this claim, since not all the propositions in a minimalist definition of truth have the same form. Horwich accepts that the definition will contain the claim “only propositions are true.”

⁵⁷*Thought and World*, p. 69. Horwich also brought up the example of the logical constants in his response to the ideology objection in a letter to me.

⁵⁸Whether it is true will depend on how expressions such as ‘intimate cognitive relationship’ are spelled out.

⁵⁹See (PC) on p. 65 of *Thought and World*. Hill imposes two further requirements, but they do not affect the argument below.

⁶⁰In this discussion I have suppressed several doubts about implicit definitions.