

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Substantive Truth and Knowledge of Meaning

Permalink

<https://escholarship.org/uc/item/03f8n8t6>

Author

Khatchirian, Arpy

Publication Date

2014

Peer reviewed|Thesis/dissertation

Substantive Truth and Knowledge of Meaning

By

Arpy Khatchirian

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Philosophy

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Barry Stroud, Co-Chair
Professor John MacFarlane, Co-Chair
Professor Hannah Ginsborg
Professor Christopher Kutz

Spring 2014

Substantive Truth and Knowledge of Meaning

Copyright 2014

by

Arpy Khatchirian

Abstract

Substantive Truth and Knowledge of Meaning

by

Arpy Khatchirian

University of California, Berkeley

Professor Barry Stroud, Co-Chair
Professor John MacFarlane, Co-Chair

Deflationists have been hard at work convincing us that the concept of truth is far less interesting, and therefore far less mysterious, than long-lasting debates as to its true nature have made it seem. By denying that there is a substantive notion of truth expressed by our uses of the word ‘true,’ deflationism promises to dissolve a number of explanatory hurdles.

Donald Davidson rejected deflationism. He grew increasingly sceptical of attempts to assimilate our understanding of truth to any definition or schema, and came to see his own approach to meaning as crucially dependent on a substantive notion of truth.

Exactly how, and where, does Davidson’s approach to meaning depend on the availability of a substantive notion of truth? One of my goals is to answer this question. Another is to explain why it has proved so hard to answer it. A third goal is to bring to light what it would really be like to be a deflationist—in particular, what linguistic competence would look like if we took deflationism seriously.

In Chapter 1, I clarify Davidson’s proposal to use truth theories as meaning theories. I explain this proposal as rooted in a certain conception of a meaning theory as an account of a speaker’s understanding of her own sentences, and of this understanding as at least partly consisting in the speaker’s knowledge of the truth-conditions of her sentences. In Chapter 2, I argue that deflationism does not allow us to describe competent speakers as knowing the truth-conditions of their sentences. If this is right, deflationism is incompatible with Davidson’s conception of truth theories as theories of linguistic competence, but not for the reason usually given. That is, it is not because deflationism explains knowledge of truth-conditions as a trivial by-product of linguistic

competence rather than being constitutive of it, since deflationism does not even allow us to ascribe such knowledge to speakers.

In Chapter 3, I argue against the widespread assumption that Davidson's proposal concerning the form to be taken by a meaning theory is primarily an answer to the question: what knowledge would enable us to interpret any utterance in a given language? This assumption underlies recent attempts to exhibit Davidson's approach to meaning as hospitable to a deflationary account of truth. However, as I explain, it undermines the distinction Davidson insists on between meaning theories and translation manuals, and thus cannot be right. Chapter 4 further argues that the methodology of radical interpretation does not, by itself, explain the need for a substantive notion of truth, contrary to what some (including Davidson) have suggested. Finally, in Chapter 5, I highlight some consequences of these results for both our understanding of Davidson's philosophy of language, and our understanding of deflationism.

For Chaké

Contents

Chapter 1: A Truth Theory as a Meaning Theory?.....	1
1.1 Introduction.....	1
1.2 Meaning theory, linguistic competence, and knowledge of truth-conditions.....	3
1.3 A familiar story	7
1.4 Describing a speaker’s knowledge of a language	13
1.5 Dummettian argument and implausibility of the reductive reading.....	15
1.6 In what sense can competence in a language be said to partly consist in knowledge of truth-conditions?.....	21
1.7 Conclusion	23
Chapter 2: Deflationism and Knowledge of Truth-conditions.....	25
2.1 Introduction.....	25
2.2 A closer look at disquotationalism	28
2.3 Some puzzling remarks	29
2.4 Epistemic analyticity, belief, and the holding true of a sentence.....	31
2.5 Boghossian’s account of epistemic analyticity.....	33
2.6 Knowledge of truth-conditions and grasp of ‘is true’	36
2.7 A more charitable reading?.....	40
2.8 Conclusion	47
Chapter 3: Truth Theories as Meaning Theories: What’s Interpretation Got to Do with It?	51
3.1 Introduction.....	51
3.2 Infinite competence and knowledge of literal meaning	53
3.3 What could we know that would suffice for interpretation?.....	55

3.4	Instrumentalism	57
3.5	Lepore and Ludwig's proposal.....	59
3.6	Truth theories versus translation theories.....	61
3.7	Saying and showing	66
3.8	Textual evidence: speaker and interpreter.....	68
3.9	Conclusion	73
Chapter 4: Can a Deflationist Make Sense of Radical Interpretation?		75
4.1	Introduction.....	75
4.2	The need for a transcendent notion of truth.....	77
4.3	Radical interpretation from a deflationary perspective	81
4.4	Logic and truth.....	83
4.5	Charity, agreement, and deflationary truth.....	85
4.6	Observation sentences and the early stages of radical interpretation	86
4.7	Substantive truth and the reification of meaning.....	92
4.8	Conclusion	96
Chapter 5: Some Consequences.....		99
5.1	Truth, truth theories, and radical interpretation.....	99
5.2	Holding a sentence to be true	102
5.3	What is the point of the exercise?	104
5.4	The radicalness of deflationism	107
Bibliography.....		111

Acknowledgements

For a few years after my arrival in Berkeley, I had the pleasure, and privilege, of working for and with Donald Davidson. Notwithstanding the time since his passing, it is his mentorship, support and voice that have helped shape my vision of this project.

Subsequently, at various stages in the unfolding of my thesis, I have leaned on Hannah Ginsborg, Barry Stroud, and John MacFarlane. I am grateful to each of them, for in turn supporting, guiding, and encouraging me.

I am deeply indebted to Barry Stroud, for providing such a great model of how philosophy ought to be done, and how it ought to be taught. When I needed it most, he helped me sustain a clear view of what I really wanted to say.

My drive for the past couple of years is due in no small part to the encouragement from John MacFarlane, to make my writing “lean and mean.” He has been demanding and thorough, and has helped me be a better philosopher as a result.

I am grateful to Hannah Ginsborg for her support, her excellent feedback, and the many fruitful and thoroughly stimulating discussions we have had over the years. I look forward to many more of these in the future.

For their friendship throughout the years, for their critical and sympathetic thinking, and for keeping me up way past my bedtime (if not theirs), I thank Bethany Hoffman and Eliot Quataert. I thank Aparna Rao and Joshua Beattie for being “the friends I grew up with,” even though we met while in graduate school at Berkeley. For generous feedback that helped improve the presentation of Chapter 3, I thank Sudha Lakshmi.

I owe no one more than I owe my parents, Sona and Nerces, for showing me very early on how to be a free thinker and for being the free thinkers that they themselves are. I thank them, my sisters Karine and Sarah, and my husband Faraaz, for knowing me so well, for all the laughing and arguing we do together, and for their support during the more difficult times.

It is not easy living with a philosopher. Faraaz’s patience, humor, and soulfulness have made this journey possible. Finally, I am thankful to my daughter Sophie, for ensuring that there is never a dull moment in my life.

I dedicate this dissertation to my grandmother Chaké who, at 94, is still the wisest person I know.

Chapter 1

A Truth Theory as a Meaning Theory?

1.1 Introduction

For all its influence, no consensus has thus far been reached concerning either the significance or the plausibility of Davidson's proposal to use truth theories as meaning theories. Some now consider the Davidsonian project essentially defunct, citing a justificatory burden it has proved unable to meet¹. Others maintain that when appropriately understood and modified, the project turns out to be a plausible and fruitful one². I share the optimism of the latter group. However, I also believe that the significance of Davidson's proposal is too often misunderstood, leading to complex alleged improvements of it, which, despite their ingenuity, only take us further away from a clear view of Davidson's goals.

One of my goals in this dissertation is to ascertain whether, and if so, why, Davidson's approach to meaning is incompatible with a deflationary account of truth. Deflationists about truth believe that there is no explanatorily significant concept expressed by our uses of 'is true'. As they see it, 'is true' a purely syntactic device, whose role in our language we can adequately and exhaustively capture in terms of certain equivalences between sentences in which it occurs and sentences in which it does not. Thus, we have

¹ See especially Soames 2008.

² For different sorts of attempts to justify some form of the proposal but only by appeal to substantive departures from Davidson's ideas, see Larson and Segal 1995, Higginbotham 1992, and, more recently, Lepore and Ludwig 2005 and 2010. Lepore and Ludwig 2010 is a direct response to Soames 2008, which itself includes criticisms of Higginbotham 1992 and Larson and Segal 1995.

Far more promising than these, and, as I see it, more consonant with Davidson's own views, are Rumfitt 1995 and Heck 2007's defenses of a Davidsonian approach to meaning. Heck and Rumfitt have significantly influenced my own understanding of this approach.

understood all there is to understand about the meaning of ‘is true’ in our language once we have grasped the mechanism or schema governing these equivalences³, and appreciated the point of having such device in our language. Disquotationalists, for instance, think that it is an exhaustive explanation of the meaning of the predicate ‘is true’ in our language that for each sentence *s* of our language, the sentence in which ‘is true’ is applied to a quotation name for *s* is logically equivalent to *s*. The truth predicate allows us to replace talk about the world with equivalent talk about our sentences, thus enabling us to express infinite conjunctions or disjunctions (e.g., ‘Every sentence of the form <A iff A> is true’), or endorse sentences that we are not in a position to use (for example, ‘He is a good doctor, so his diagnosis must be true’ uttered in the absence of knowledge of the doctor’s diagnosis). For all its usefulness, however, it does not express a philosophically interesting concept, nor does it denote a substantive, explanatorily relevant property of our sentences.

Davidson, by contrast, held that “truth is a crucially important explanatory concept,” “one of the clearest and most basic concepts we have,”⁴ and that this primitive concept is essential to our understanding of meaning, of others, and of “why they act as they do.”⁵ He thought there was something essential to our concept of truth that any attempt to assimilate it to a schema would fail to capture, and came to see his own approach to meaning as crucially dependent on a primitive notion of truth. But where, exactly, does talk about truth play a role in Davidson’s approach that goes beyond the expressive functions recognized by deflationists?

The consensus until fairly recently had been that truth-conditional approaches to meaning do indeed place explanatory demands on truth that are incompatible with deflationism. Recently, however, philosophers have taken to questioning this assumption, and arguments in favor of the compatibility of deflationism with truth-conditional semantics, and with a Davidsonian approach in particular, have steadily been gaining steam.

³ There are disagreements among deflationists concerning what the relevant mechanism is. Disquotationalists such as Field and McGee take ‘is true’ to be a predicate of sentences governed by the schema < ‘s’ is true if and only if s>. Horwich prefers to speak of propositional truth, and takes the relevant schema to be <The proposition that *p* is true if and only if *p*>. Brandom and other prosententialists think that ‘is true’ is not a genuine predicate, but a prosentence forming operator which, when applied to an expression that designates a sentence *s* (or set of sentences) results in a sentence whose content is inherited from *s* (or from sentences in the designated set). See Field 2001, McGee 2005, Horwich 2005, Grover, Camp, and Belnap 1975, and Brandom 1994.

⁴ Davidson 2005a, pp. pp. 54, 55.

⁵ *Ibid.*, p. 73.

Disagreement concerning the compatibility of Davidson's own approach to meaning with a deflationary account of truth is traceable to a more basic disagreement concerning the significance of Davidson's proposals surrounding the notion of meaning. One of my goals is to explain why Davidson's proposal to use truth theories as meaning theories is to be understood in a much more straightforward sense than the many complex justificatory attempts in the literature make it out to be. Its motivation is the basic and intuitively compelling assumption that understanding one's own sentences involves knowledge of the conditions under which they are true. This crucial, yet often overlooked, assumption is, I shall argue, the very root of the incompatibility of Davidson's approach to meaning with a deflationary account of truth.

My goal in this chapter is to spell out the significance of this assumption and its importance in motivating Davidson's approach to meaning. In the next section, I argue that Davidson is driven by a conception of a meaning theory as, first and foremost, a theory of linguistic competence, and of a theory of linguistic competence as a constructive account of what a competent speaker knows in knowing the meaning of each sentence of her language. Section 3 sketches an influential account (the "T&M story" as I shall call it) of Davidson's goals that simply ignores this conception, and consequently ends up trivializing the role of the concept of truth in truth theories serving as meaning theories. Section 4 argues that by neglecting the conception outlined in section 2, the T&M story blinds itself to the following condition: a truth theory can serve as a meaning theory for a particular language only if it entails statements of what a speaker knows about each of its sentences that helps constitute her understanding of the language. Sections 5 and 6 further clarify the sense in which Davidson takes linguistic competence to partly consist in knowledge of truth-conditions. Section 5 begins with an apparent digression, by examining an influential and initially plausible argument against combining Davidson's approach to meaning with a deflationary account of truth. I argue that the argument only works on a certain reductive construal of the idea that linguistic competence partly consists in knowledge of truth-conditions. On the relevant construal, however, this idea is anyway implausible, whether or not it is combined with a deflationary account of truth. Section 6 then spells out a more plausible, non-reductive construal of the sense in which linguistic competence can be said to partly consist in knowledge of truth-conditions.

1.2 Meaning theory, linguistic competence, and knowledge of truth-conditions.

Let us examine the path that initially leads to Davidson's proposal that meaning theories take the form of recursive characterizations of truth. This proposal makes its first

appearance in “Truth and Meaning,”⁶ and in the passages preceding it, Davidson spells out his conception of a meaning theory as a “constructive account” of the meaning of sentences in a language. In “Theories of Meaning and Learnable Languages,” he connects this conception of the job of a meaning theory to a certain “learnability” requirement on natural languages:

I propose what seems to me clearly to be a necessary feature of a learnable language: it must be possible to give a constructive account of the meaning of the sentences in the language. Such an account I call a theory of meaning for the language, and I suggest that a theory of meaning that conflicts with this condition, whether put forward by a philosopher, linguist, or psychologist, cannot be a theory of a natural language; and if it ignores this condition, it fails to deal with something central to the concept of a language.⁷

Davidson goes on to reject different purported “theories of meaning” for failing to provide a constructive account of meaning, and thus for failing to exhibit the languages they deal with as learnable, or worse, for implying their unlearnability. But what exactly does Davidson mean by a “constructive account” of the meaning of the sentences of a language, and by a language being “learnable”?

A constructive account of the meaning of sentences in a language is, Davidson explains, one that specifies “in a way that depends effectively and solely on formal considerations, what every sentence means,” or again, one that reveals “the meaning of each sentence as a function of a finite number of features of the sentence.” Moreover, Davidson makes it clear, at the outset, that in seeking a constructive account of meaning for a language, his aim is to shed light on “the skill or ability of a person who has learned to speak [the] language.”⁸ In particular, to exhibit a language as “learnable,” or, make sense of how competence in the language might, in principle, be acquirable, involves “understand[ing] how an infinite aptitude can be encompassed by finite accomplishments.”⁹ Thus, in aiming to characterize the meaning of a potential infinity of sentences as a function of features of a finite number of repeatable expressions and their modes of combination, Davidson’s goal is to exhibit a speaker’s infinite competence as structured into finitely many abilities, thus making it intelligible it could be acquired. This, however, does not

⁶ [A] theory of meaning for a language L shows ‘how the meanings of sentences depend upon the meanings of words’ if it contains a (recursive) definition of truth-in-L. And so far, at least, we have no other idea how to turn the trick. (“Truth and Meaning,” Davidson 1984, p. 23)

⁷ “Theories of Meaning and Learnable Languages,” Davidson 1984, p. 4.

⁸ *Ibid.*, pp. 7-8.

⁹ *Ibid.*, p. 8.

involve explaining or speculating on the mechanisms by which the relevant capacities are, or may be, acquired.¹⁰

Thus, it is only because Davidson thinks of linguistic competence as competence in a particular language, and of a language as containing infinitely many sentences, that he thinks there is such a task as that of accounting for a language's learnability, or, making sense of linguistic competence as resting on finitely many abilities.¹¹ Indeed, once we agree with the assumption that linguistic competence involves an infinite capacity, it is hard to resist the demand for an intelligible description of such competence as resting on a finite basis, and as having a particular structure.¹²

Clearly, then, Davidson's conception of meaning theories as theories of linguistic competence should be made central in assessing both his demands on meaning theories and his proposals as to how they could be met. But what might be the role of a compositional truth theory in a theory of linguistic competence?

I do not think we can make any progress in answering this question unless we keep in mind the importance of a connection Davidson draws between competence in a language and the semantic features a meaning theory is to assign to its sentences. The connection can be spelled out in terms of the following condition on an adequate theory: for each sentence of the language, the theory is to entail a theorem such that knowing that the sentence has the features this theorem assigns to it is part of what is

¹⁰ In fact, Davidson not only emphasizes the distinction between a structure-revealing description of the capacities involved in knowing a language and an account of the mechanisms by which they can be acquired, he also disavows any direct preoccupation with the latter enterprise:

Often it is asserted or implied that purely a priori considerations suffice to determine features of the mechanisms, or stages, of language learning; such claims are suspect. (Ibid., p. 3)

It is not appropriate to expect logical considerations to dictate the route or mechanism of language acquisition, but we are entitled to consider in advance of empirical study what we shall count as knowing a language, how we shall describe the skill or ability of a person who has learned to speak a language. (Ibid., pp. 7-8)

¹¹ I think it is clear that by the "learning" of a language, Davidson does not mean to denote any specific mode of acquisition of linguistic abilities. As I understand it, there is nothing more to his "learnability" requirement than the idea that languages are, and should be exhibited as, graspable, in other words, that linguistic competence should be exhibited as acquirable (for a different interpretation, see Lepore and Ludwig 2005, pp. 25-37).

¹² I do not mean to suggest that a speaker's infinite competence is to be ascertained independently of the assumption that his capacities have a compositional basis. Rather, our conceptions of a language as compositional and of a speaker's competence as ranging over infinitely many (non-synonymous) sentences seem to go hand-in-hand.

involved in understanding the language. That Davidson imposes this condition on an adequate meaning theory is strongly suggested by his assimilating the question of what the sentences of a language mean to the question of what we know in knowing the meaning of each of these sentences, that is, what we know in understanding the language.¹³ Moreover, Davidson clearly thinks of competence in a language as involving the capacity for genuine thought – including particular beliefs and intentions¹⁴—rather than mere know-how. Finally, ascribing to competent speakers of a given language, knowledge (or at least, the capacity for knowledge) of sought-for theorems, enables us to see how the derivability of these theorems from structure-revealing axioms could shed light on the structure of their competence in the language.¹⁵

Unfortunately, this condition is too often overlooked in discussions of truth-conditional approaches to meaning, leading either to neglect of the conception of a truth theory as a theory of linguistic competence, or to puzzlement as to how it could be given substance. In the following section, I will sketch a familiar account of the motivation for Davidson’s proposal to use truth theories as meaning theories, in which this conception clearly seems to have fallen by the wayside.

¹³ This assimilation helps explain Davidson’s criticism of purported meaning theories for failing to account for facts about sentences of a language that we know, or are in a position to know, when we understand the language. For example, against the “hopeful thought” that a recursive syntax, together with a dictionary giving the meaning of each syntactic atom, could fulfill the goal of a meaning theory, Davidson simply notes that “knowledge of the structural characteristics that make for meaningfulness in a sentence, plus knowledge of the meanings of the ultimate parts, does not add up to knowledge of what a sentence means” (“Truth and Meaning,” Davidson 1984, p. 21).

¹⁴ Consider the following passage:

Someone who utters the sentence ‘The candle is out’ as a sentence of English must intend to utter words that are true if and only if an indicated candle is out at the time of utterance, and he must believe that by making the sounds he does he is uttering words that are true only under those circumstances. (“Thought and Talk,” Davidson 1984, p. 155)

¹⁵ I take this to be part of the point Wright is making on p. 109 of Wright 1981.

1.3 A familiar story ¹⁶

The story I am about to tell takes its departure from a famous passage in Davidson's "Truth and Meaning."¹⁷ I will thus refer to it as the "T&M Story," even though, as we will see, it neglects the very conception of a meaning theory as a theory of linguistic competence manifested earlier in the same paper.¹⁸

As the story goes, the main point of a meaning theory for L is to finitely generate theorems that in some sense "give the meaning" of sentences of L, and it is simply the technical difficulties encountered in trying to meet this demand that lead Davidson to his proposal to seek theorems of the form $\langle s \text{ is true in } L \text{ if and only if } p \rangle$. For instance, if we construe the target theorems as having the form $\langle s \text{ means in } L \text{ that } p \rangle$ ¹⁹, it is not clear through what form of axioms governing finitely many repeatable expressions and their modes of combination, and what sort of logic, such theorems could be derived.

Davidson's goal, then, is to find another form for the target theorems, one that would make the task of finitely generating them more tractable. And his idea here is to simply use, in place of the "means in L that" locution, words that would turn the target theorems into sentences whose logic is extensional. Thus, instead of $\langle s \text{ means in } L \text{ that } p \rangle$, Davidson proposes $\langle s \dots \text{ if and only if } p \rangle$ as the form of target theorems, where 'if and only if' is the material biconditional and the dots are appropriately filled-in. For an instance of this schema to be a well-formed sentence, we need a well-formed sentence on each side of its biconditional. This is something we can have, on the right hand-side, by simply substituting a sentence of the metalanguage (i.e. the language in which the theory is formulated) for 'p.' But what is to be substituted for 's' is not a sentence of L, but rather, a structural description, in the metalanguage, of a sentence of L. Replacing the dots with a one-place predicate applicable to sentences of L would thus

¹⁶ Different versions of this story can be found in Williams 1999, Kölbel 2001, Lepore and Ludwig 2010, and, tentatively, on pp. 7-8 of McDowell's "Truth-Conditions, Bivalence, and Verificationism" and pp. 32-33 of his "Meaning, Communication, and Knowledge" (both in McDowell 1998). However, some of McDowell's other writings on the subject, and in particular, Evans and McDowell 1999, cast considerable doubt on the idea that he accepts the T&M Story.

¹⁷ On pp. 22-24 of "Truth and Meaning" in Davidson 1984.

¹⁸ I do not mean to suggest that Davidson's remarks on pp. 22-24 of "Truth and Meaning" are to be interpreted as the story I am about to tell interprets them. Keep in mind, also, that Davidson came to express dissatisfaction with his way of putting things in "Truth and Meaning." See p. xvi of his introduction in Davidson 1984.

¹⁹ Where the expressions to be substituted for 's', 'L' are 'p' are, respectively, a structural description, in the metalanguage (the language in which the theory is formulated), of a sentence of the object language (the language the theory is about), a reference to the object-language, and a sentence of the metalanguage.

also give us a sentence on the left hand-side. Thus, the form of our target theorems becomes $\langle s \text{ is } T \text{ if and only if } p \rangle$. Following common usage, I will refer to sentences of this form as “T-sentences,” or “truth sentences,”²⁰ given the standard interpretation of the relevant predicate as a truth predicate, and of the compositional theory that entails them as a “T-theory” or “truth theory.”

Here is where Davidson perceives a connection with Tarski’s methods for “defining truth” for particular languages, since Tarski intended his truth definitions to entail sentences of precisely this form.²¹ Moreover, T-biconditionals entailed by a Tarskian truth definition are recursively generated from axioms governing finitely many expressions and finitely many ways of combining them. So if Davidson’s target theorems are given the form of T-biconditionals, Tarski’s methods can be exploited to generate these theorems from a finite and structure-revealing theory. To further reinforce the appearance of a connection between Tarskian truth definitions and Davidson’s goals, consider Tarski’s condition of material adequacy on a truth definition for a language, “Convention T”: to satisfy Convention T, the definition must entail, for each sentence s of the language, a T-biconditional in which the sentence used on the right-hand side is either s itself (when the metalanguage contains the object-language), or an acceptable translation of s in the metalanguage.

A T-theory for a language L meeting Convention T can thus be said to finitely generate, for each sentence s of L , a T-theorem in which s is “paired with” a sentence of the metalanguage that translates it. But the T-theory effects such pairings only in an indirect way. For each sentence s of L , the metalanguage sentence paired with s is the very sentence that is *used*—rather than mentioned—on the right-hand side of the relevant T-theorem for s . For example, the T-sentence ‘“La neige est blanche” is T if and only if

²⁰ There are two ambiguities in the notion of a “T-sentence,” as this term is commonly used. The first ambiguity concerns the form these sentences are to take. Since Tarski himself was interested in formal languages without context-sensitivity, his T-sentences were of the form $\langle s \text{ is true-in-}L \text{ if and only if } p \rangle$. But for languages like ours, the sought-for theorems have to involve relativization to various features of contexts of utterances, and are thus better thought of as universal generalizations (for example, and simplifying, $\langle \text{For any speaker } S \text{ and time } t, s \text{ for } S \text{ at } t \text{ is true if and only if } p \rangle$), rather than biconditionals. For ease of exposition, I will mostly ignore context-sensitivity in what follows, and treat T-sentences as biconditionals. The second ambiguity concerns the connection between T-sentences and Tarski’s Convention T. A truth theory for L satisfies Convention T if, for each sentence x of L , it entails a theorem of the form $\langle s \text{ is true in } L \text{ if and only if } p \rangle$, where the sentence substituted for ‘ s ’ denotes x , and the sentence substituted for ‘ p ’ is a translation of x in the metalanguage, i.e., the language of the theory (if the metalanguage includes L , then x itself is to be substituted for ‘ p ’). Let us call such theorems “translational.” At times, Davidson treats T-sentences as by definition, translational, though he sometimes uses ‘T-sentence’ to simply speak of sentences of the form $\langle s \text{ is true in } L \text{ if and only if } p \rangle$, whether or not they are translational. In what follows, I will be using ‘T-sentence,’ ‘T-theorem,’ and ‘T-biconditional’ in the second of these two senses.

²¹ See “Truth and Meaning,” in Davidson 1984, and Tarski 1944, 1956.

snow is white” ’ pairs the sentence ‘La neige est blanche’ in French with an acceptable translation, ‘Snow is white’, in English, by employing this very sentence on the right-hand side of its biconditional. If I know this about the T-sentence, and since I understand English (and therefore know that ‘Snow is white’ as used in this T-sentence means in English that snow is white), I know that I can describe utterances of ‘La neige est blanche’ in French as meaning that snow is white, or as cases of saying (among others things) that snow is white.

But how should we interpret the T-predicate in such a theory, if we are to make it at all plausible that the theory could do duty as a meaning theory? If our goal in constructing a T-theory is merely that of pairing object-language sentences with metalanguage sentences in use that translate them, then it does not seem to really matter how we interpret ‘is T.’ Nonetheless, since our so-called “theory” is supposed to be an empirical theory, rather than a system of uninterpreted sentences, we had better take the T-theorems to be saying something or other about object-language sentences, and thus to assign to the T-predicate some meaning, however arbitrary. As it turns out, we do have a pre-theoretical notion that would be a natural choice for what ‘is T’ can be taken to express, and that would makes the T-theorems come out true. This is our notion of truth, or at least, one pre-theoretical way of understanding truth for sentences. We can pin down this notion by drawing on the following seeming platitudes:

To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.²²

One who makes a statement or assertion makes a true statement or assertion if and only if things are as, in making that statement, he states them to be.²³

A true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.²⁴

The notion of truth that I am after is, I believe, the one Tarski is characterizing here, and it is also connected, even if not identical to, Strawson’s and Aristotle’s notions.²⁵ Each of these three characterizations is meant to capture, in an informal way, our

²² Aristotle, *Metaphysics* IV.7, 1011b25ff.

²³ Strawson 1971, p. 180.

²⁴ Tarski 1956, p. 155.

²⁵ The reason for this qualification is that we are after a notion of truth for sentences, whereas Strawson is treating truth as an attribute of statements or assertions, and it is not clear what Aristotle’s truth predicate applies to.

intuitions about truth. A common idea emerging from them is that it is only when the idea of “how things are stated to be” by a sentence (as uttered on a given occasion) is in place, that we can meaningfully speak of the sentence’s being true (as uttered on that occasion): it is true if things are as they are stated to be by the sentence as uttered on that occasion.²⁶

If a T-biconditional for a given sentence *s* satisfies Convention T (that is, if it employs, on its right-hand-side, a sentence that can be used to interpret utterances of *s* in *L*²⁷), then it will be true if ‘is T’ is taken to express the intuitive notion of truth I have just characterized.²⁸

If this is right, and if the intuitive notion of truth is not revealed to have any other features that would render it unsuitable for use in a T-theory, we might as well take it to be the notion employed in a T-theory serving as a meaning theory, rather than contemplate the possibility of other notions concocted for just this purpose. But for all

²⁶ McDowell seems to be recommending this very way of understanding truth in the following passage:

The basis for the truth-conditional conception of meaning, as I see it, is the following thought: to specify what would be asserted, in the assertoric utterance of a sentence apt for such use, is to specify a condition under which the sentence (as thus uttered) would be true. The truth-conditional conception of meaning embodies a conception of truth that makes that thought truistic. (I am inclined to think it is the only philosophically hygienic conception of truth there is). (“In Defense of Modesty”, McDowell 1998, p. 88)

It is with some hesitation that I interpret McDowell as proposing that ‘is T,’ as used in a truth theory serving as a meaning theory, be taken to express the intuitive notion of truth I have just characterized. My hesitation partly stems from some of the ways others have interpreted McDowell. Rattan 2005 describes McDowell as offering us the only coherent form of deflationism about truth, by espousing a deflationary approach to both truth and meaning. Kölbel 2001 and Williams 1999 take McDowell to be proposing that the sole function of ‘is T,’ as it occurs in our target theorems, is to allow us to pair sentences of a language with their interpretations in the metalanguage, thus underplaying the significance of the notion of truth in a truth-conditional meaning theory. I am not convinced that either of these claims gives us an accurate reading of McDowell’s views on the role of truth in our understanding of meaning, even though, to be fair, there is some textual evidence pointing in favor of Kölbel and Williams’s reading. For instance, in the passages mentioned in fn. 16 of this chapter, McDowell seems to be offering precisely the kind of instrumentalist account of the role of ‘is T’ to which the T&M story leads. For what it is worth, my own view is that the intuitive notion of truth that I am characterizing, and take McDowell to be characterizing in the above-quoted passage, is not deflationary, and that the role of ‘is true’ in a truth-conditional meaning theory goes beyond what Kölbel and Williams take it to be—namely, that of a mere syntactic device.

²⁷ That is, to say what these utterances (literally) mean, or how they state things to be.

²⁸ Though this claim is generally accepted, there are those who deny it. Ludwig, for example, argues that sentences containing vague terms are neither true nor false, and thus that truth-biconditionals for such sentences are simply not true. See Chapter 10 of Lepore and Ludwig 2005, and Ludwig and Ray 2002.

that the above story says about Davidson's goals in seeking T-theories, we have no compelling reason to reject other ways of interpreting 'is T'. Any notion that has the same extension, in L, as the pre-theoretical notion, and thus, any notion satisfying a materially adequate Tarskian definition of truth for L will be equally adequate. For example, if we interpret such a definition as giving content to an unstructured predicate 'true-in-L' (rather than as capturing features or a pre-existing notion), the predicate thus defined satisfies the relevant T-theory.

If, as the story goes, the function of T-theorems is merely that of pairing each sentence of L with an "interpretation" in the metalanguage, we need not understand 'is T' in one way rather than another. The pairing of object-language sentences with metalanguage sentences in use, and the derivability of our T-theorems through an extensional logic, is achieved regardless of how 'is T' is interpreted (and arguably, regardless of whether it is interpreted at all.)

Here, then, is the basic problem with the T&M Story. On the one hand, 'is T' is portrayed as a mere syntactic device, brought on to solve technical problem. On the other hand, attempts are made, by appeal to certain platitudes connecting truth and meaning, to convince us that 'is T' would be aptly, or even ideally, interpreted as expressing the intuitive notion of truth I have just characterized.²⁹

Another striking aspect of the T&M Story is its reluctance to characterize a truth theory as itself a meaning theory. As the story goes, and no matter how 'is T' is interpreted, a T-theory for L does not directly say what sentences of L mean. Rather, an adequate T-theory for L is supposed to be an indirect way of giving the meaning of sentences of L. As McDowell, puts it, an adequate truth theory "serves as" a meaning theory "without

²⁹ For example, Rattan 2005 argues that T-sentences do not state the content of speakers' knowledge of meaning, but can be *read* as specifying this knowledge, by "encoding more information than they explicitly state" (Rattan 2005, p. 8). In particular, the T-sentence "'La neige est blanche' is T iff snow is white" *can be read* as specifying French speakers' knowledge of meaning if we know that 'snow is white' as used in it translates 'La neige est blanche' in French. For we can then infer that 'La neige est blanche' means in French that snow is white, which is what speakers of French know about the sentence 'La neige est blanche' in knowing its meaning.

Rattan here seems to be offering nothing more than a version of the T&M Story. But he goes on to ask an important question: why, given the story that has just been told, ought we interpret 'is T' as expressing the concept or truth? To answer this question, he appeals to the schema $\langle s \text{ means that } p \rightarrow s \text{ is true iff } p \rangle$ governing our notions of truth and meaning, and argues that in light of this "fundamental platitude about truth and meaning," "the concept of truth is a particularly good, and natural, choice in specifying speakers' knowledge of meaning" (Rattan 2005, p. 15). However, as far as I can tell, Rattan does not give us any reason to demand that 'is T' be taken to express the intuitive notion of truth governed by the "fundamental platitude," since T-sentences can match mentioned sentences with what they mean regardless of how 'is T' is interpreted.

being one.”³⁰ Or consider Lepore and Ludwig’s instance that Davidson’s goal is not to use a truth theory as a meaning theory, but rather, “to exploit a truth theory to do the work of a meaning theory.”³¹ Davidson himself concedes, at some point, that “nothing strictly constitutes a theory of meaning,” and in particular, that “a theory of truth no matter how well selected, is not a theory of meaning.”³² As I hope will become clear in the course of my arguments, these sorts of remarks only serve to obscure Davidson’s considered view of the significance of truth theories, for this is a view on which a truth theory can, in a straightforward sense, *be* a meaning theory.

Let us go back to the question of how we are to interpret ‘is T.’ Do we have any good reason to *demand* that ‘is T’ be taken to express the intuitive notion of truth I have characterized?

I believe we do. In the midst of the search for a form of theorem that would permit the derivability of infinitely many theorems from a finite basis, we seem to have lost sight of the condition, on acceptable theorems, of describing what a competent speaker knows about each sentence of his language in understand it the way he does. Depending on how we interpret the T-predicate employed in our T-theorems, they may or may not express the content of such knowledge. And clearly, this gives us a compelling reason in favor of the intuitive notion I have just characterized. For speaking of truth in this sense, a competent speaker of a language can indeed be said to know the condition under which each sentence of the language would be true on particular occasions of utterance, at least insofar as he can be said to know how things would be stated to be by the relevant utterances of these sentences. It remains to be seen whether there are other conceptions of truth on which competence in a language can be said to involve knowledge of the truth-conditions of its sentences. In particular, does a deflationary account of truth allow us to think of linguistic competence as involving knowledge of truth-conditions?

The next chapter will argue that it does not. What we need, in the meantime, is a more precise characterization of the sought-for connection between the knowledge that a meaning theory is supposed to describe, and a speaker’s linguistic competence. The remainder of this chapter is devoted to this task.

³⁰ “On the Sense and Reference of a Proper Name,” McDowell 1998, p. 173.

³¹ Lepore and Ludwig 2011, p. 272.

³² “Reply to Foster,” Davidson 1984, pp. 178-179.

1.4 Describing a speaker's knowledge of a language

What connection should be taken to hold between the knowledge described by a theory and competence in a language, if the theory is to shed light on what it is to understand the language? It would surely be implausible to suggest that any theory that describes something a competent speaker knows about each sentence of his language is thereby a meaning theory for this language. For a theory to qualify as theory of linguistic competence, the knowledge that the theory describes has to be shown to play a central role in explaining the many things a speaker can do with sentences of the language. In other words, a speaker's competence in the language has to be plausibly taken to consist (or, partly consist) in the knowledge that the theory describes.

Davidson and others sometimes construe the point of a meaning theory for a given language L as that of spelling out the content of knowledge that would simply suffice for understanding L.³³ But this requirement is too weak, given our conception of a meaning theory as a theory of linguistic competence. A given condition can be sufficient for understanding L merely by virtue of presupposing this understanding, without being constitutive of it. The claim that knowledge of a certain kind of theory is sufficient for understanding L is also not sufficiently contentful, since we do not seem to have a definite enough independent conception of what this property of understanding a language is, for which knowledge of truth-conditions is being claimed to be sufficient. As I see it, our task here is precisely that of illuminating the notion of linguistic competence, by spelling out the sort of knowledge that it involves, and the role the attribution of such knowledge plays in explaining the variety of capacities associated with someone's understanding of a language.

My proposal, then, is that Davidson's conception of truth theories as meaning theories involves a commitment to the idea that competence in a language (partly) consists in knowledge of the truth-conditions of its sentences, where what is in question is a speaker's competence in, or understanding of, her own language (in the sense of idiolect.)

One might worry that my claim that a meaning theory is to describe what a speaker knows about each sentence of her language runs against Davidson's insistence that neither speaker nor interpreter has propositional knowledge of what is stated by a truth theory.³⁴ In characterizing a competent speaker as knowing the conditions under which each of her sentences would be true (on particular occasions of utterance), I do intend the knowledge to be construed as propositional or explicit, that is, as knowledge that

³³ In particular, see p. 125 of "Radical Interpretation" in Davidson 1984.

³⁴ Davidson 2005a, p. 52.

can be expressed by its possessor. This is fully compatible with Davidson's denial that the speaker has propositional knowledge of a truth theory, since Davidson is here only denying that the speaker has propositional knowledge of the axioms of such a theory.³⁵ For each sentence of a given language, a speaker may be able to state what she knows about the conditions under which the sentence is true, without being in a position to say how she knows it, or how her knowledge is to be explained in terms of her understanding of finitely many expressions of the language and of the significance of their modes of combination.³⁶

Here, then, we have a pretty definite account of the role of truth in Davidson's "truth-conditional" approach to meaning. By proposing that a truth theory can serve as a meaning theory, Davidson is claiming that a truth theory can serve as a structure-revealing description of the knowledge in which a speaker's understanding of her own language consists (or partly consists). But what exactly is the sense in which a speaker's understanding of a language is supposed to *consist in* her knowledge of truth-conditions?

A central goal of my dissertation is to explain why Davidson's approach to meaning cannot be successfully combined with disquotationalism, and by extension, other versions of deflationism about truth. In the rest of this chapter, I would like to clarify

³⁵ For pretty direct evidence that Davidson did think that competent speakers have explicit knowledge of the truth-conditions of each of their sentences, consider the following passage in Davidson 2005a:

[T]hrough the interpreter certainly does not need to have explicit knowledge of the theory, the theory does provide the only way to specify the infinity of things the interpreter knows about the speaker, namely, the conditions under which each of an indefinitely large number of sentences would be true if uttered. (Davidson 2005a, p. 52)

Clearly, if Davidson is prepared to ascribe explicit knowledge of the truth-conditions of a speaker's sentences to a successful interpreter of the speaker, it is only because he thinks that the speaker herself has explicit knowledge of these truth-conditions (at least when communication is successful).

Even the famous passage on pp. 22-24 of "Truth and Meaning," despite its erroneously identifying knowledge of the truth-conditions of sentences of a language L with knowledge of a Tarskian *definition* of 'truth-in-L,' concludes with the claim that "to know what it is for a sentence—any sentence—[of a given language] to be true" is "to understand the language" (Davidson 1984, p. 24). In other words, to know the conditions under which each sentence of a language is true is to understand the language.

³⁶ Knowledge of the axioms, by contrast, is to be characterized as tacit or non-propositional. But for the attribution of tacit knowledge of certain axioms to be understood as shedding light on the structure of an individual's linguistic competence, there must be more content to it than the claim that explicit knowledge of the axioms would suffice for explicit knowledge of the T-theorems, or that these axioms entail the theorems in question, or again, that the speaker behaves "as though" he had explicit knowledge of the axioms. For further discussion, see Wright 1981 and Evans 1981. For specific proposals about how the attribution of tacit knowledge is to be understood, see Evans 1981, Davies 1987, and Larson and Segal 1995.

the sense in which Davidson takes linguistic competence to partly consist in knowledge of truth-conditions, by first introducing an apparent digression. In the next section, I will examine an apparent argument against the possibility of successfully combining disquotationalism with Davidson's truth-conditional approach to meaning. I will show that the argument, though influential, turns on a misconstrual of the sense in which we should—and Davidson does—take linguistic competence to consist in knowledge of truth-conditions. I will then offer a more promising reading of the idea that linguistic competence partly consists in knowledge of truth-conditions.

1.5 Dummettian argument and implausibility of the reductive reading

If Davidson's approach turns on the idea that linguistic competence partly consists in knowledge of truth-conditions, any attempt to combine it with a deflationary account of truth would appear to be an obvious target of Dummett's seminal argument in "Truth".³⁷

It ... appears that if we accept the redundancy theory of 'true' and 'false' ... we must abandon the idea that we naturally have that the notions of truth and falsity play an essential role in any account either of the meaning of statements in general or of the meaning of a particular statement ...

[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 understand the sense of asserting P, which was precisely what was supposed to be being explained to him.³⁸

As I understand it, Dummett's argument concerns the possibility of successfully combining a truth-conditional approach to meaning with a disquotational account of truth (or, of 'is true'), rather than with the "redundancy theory" of 'is true', as this label

³⁷ I do not mean to suggest that Dummett himself had Davidson in mind in this passage. However, this passage is often the focal point of discussions surrounding the compatibility of deflationism with truth-conditional approaches to meaning. Thus we would do well to ask, as others have before me (see Rattan 2005 and Horisk 2008), whether Dummett's remarks have any force against combining Davidson's own truth-conditional approach to meaning with a deflationary account of truth.

³⁸ Dummett 1978, p. 5.

is usually understood.³⁹ I will offer a more precise characterization of what disquotationalism about truth comes to in the next chapter, but for now, the following remarks should suffice. For the disquotationalist, ‘is true’ is a predicate of sentences subject to the following two conditions: first, it can be meaningfully applied by a speaker only to sentences that he himself understands or is disposed to use a certain way—a speaker cannot meaningfully wonder whether a sentence in some language that he does not understand is or is not ‘true.’⁴⁰ Secondly, for each sentence *s* that a speaker understands, the application of ‘is true’ to a quotation-name for *s* is equivalent in meaning to *s*.⁴¹ Thus, for example, if I am using ‘is true’ in this sense, I cannot meaningfully apply it to a sentence of Japanese, since I have no understanding of any sentence of Japanese, that is, no dispositions to use any such sentence in any particular way. On the other hand, since ‘It is rarely foggy in San Francisco’ is a sentence of my language, I can meaningfully speak of this sentence as being true, and ‘“It is rarely foggy in San Francisco” is true’ and ‘It is rarely foggy in San Francisco’ are equivalent in meaning in my language.^{42, 43}

Let us return to Dummett’s argument, interpreted, as I am suggesting, as targeting the disquotational account of truth and its compatibility with a truth-conditional account of meaning.

We should remember Davidson and Dummett are both driven by a conception of a theory of meaning as a theory of understanding. Thus, when Dummett argues that a

³⁹ In support of this interpretation, note that the truth predicate that Dummett is targeting is supposed to be applicable to entities with meaning, or sense—thus, to sentences (or their utterances), rather than to propositions. Moreover, it is a predicate that satisfies the disquotationalist’s requirement that for any sentence *s*, one’s understanding of the sentence in which ‘is true’ is applied to a quotation name for *s* depends on one’s understanding of *s*. Thus, despite Dummett’s speaking of the “redundancy theory of ‘true,’” it is the disquotational account of ‘is true’ that Dummett seems to have in mind here. The “redundancy theory of truth,” as usually understood (and as Dummett himself seems to understand it in other contexts), is an account of the expression ‘it is true that...,’ thus at best an account of propositional truth, not of sentential (or utterance) truth.

⁴⁰ ‘True’ in the purely disquotational sense is simply not meaningful when applied by a speaker to a sentence she does not understand.

⁴¹ For some suitable notion of equivalence in meaning. Chapter 2 will take up the question of what the relevant notion of equivalence might be.

⁴² This means not only that an assertion of the one sentence is equivalent to an assertion of the other, but also, that these two sentences can be substituted for one another in the context of larger sentences without any relevant change in the meanings of these sentences.

⁴³ Note that both of these sentences are false. I can meaningfully, though falsely, characterize a sentence of my language as ‘true.’

deflationary account of ‘true’ cannot allow “truth and falsity to play an essential role in any account either of the meaning of statements in general or of the meaning of a particular statement,” he is, in effect, arguing that a deflationary account of truth cannot allow truth and falsity to play an essential role in an account either of what linguistic competence generally consists in, or of what competence in a particular language consists in.⁴⁴

However, Dummett’s focus on the question of how a speaker’s competence may be acquired is an unnecessary detour, since the main question Dummett is interested in is the nature of this competence, rather than the way it may be acquired.⁴⁵ To explicitly bring out both Dummett’s preoccupation with the nature of linguistic competence and the type of deflationism I take him to be targeting, I propose the following reconstruction of his argument:

1. Suppose truth is disquotational.
2. Since truth is disquotational, for any sentence P, a speaker’s understanding what it means to say of P that it is true depends on his understanding P (in other words, a speaker’s meaningful attribution of truth to a given sentence depends on his understanding of this sentence).

⁴⁴ This goes some way towards answering the following objection. According to Patterson, Dummett’s own reasoning fails “to support the view that deflationism is incompatible with the view that meaning consists at least in part in truth-conditions” (Patterson 2005, p. 274). This is because, as Patterson sees it,

there is no incompatibility between the views (a) that one might learn the meaning of a sentence by being told the condition under which it is true, that (b) a sentence and the claim of it that it is true are equivalent in meaning, and that (c) this meaning about which one can be told and which is shared by a sentence and the claim of it that it is true is in fact a truth-condition. (Ibid.)

Patterson’s objection turns on distinguishing between, on the one hand, the questions of how competence in a language may be acquired and of how the meaning of a sentence may be learned, and on the other hand, the question of what a sentence’s meaning what it does consists in. But Patterson here seems to be overlooking the conception, shared by Dummett and Davidson, of a theory of meaning as a theory of understanding. Given this conception, it is neither surprising nor problematic that Dummett’s argument focuses on a speaker’s understanding or knowledge of meaning.

⁴⁵ However, despite its potential to distract us from the main question Dummett is interested in, his focus on the question of how linguistic competence may be acquired is not illegitimate. For, even though the inference, from the assumption that linguistic competence may be acquired, in part, by acquiring knowledge of truth-conditions, to the conclusion that linguistic competence partly consists in knowledge of truth-conditions, is not valid, the converse inference is. If linguistic competence partly consists in knowledge of truth-conditions, it follows that it may be (indeed, it must be) acquired partly by acquiring knowledge of truth-conditions (thus, if linguistic competence cannot be acquired, in part, by acquiring knowledge of truth-conditions, it follows that it does not consist, even partly, in knowledge of truth-conditions).

3. Knowledge that P is true under certain conditions, or, knowledge of what is stated by a true T-biconditional for P, depends on understanding what it means to say of P that it is true.
4. Therefore, knowledge that P is true under certain conditions depends on understanding P. (from 2 and 3)
5. Therefore, a speaker's understanding of P cannot itself (partly) consist in his knowledge that P is true under certain conditions. (from 4)

Let us examine the inference from 4 to 5. It is clearly a legitimate move, on a certain default construal of "consists in" claims. On the relevant construal, the claim that some property or capacity A (partly) consists in property B belongs to an explanation of A in terms of more basic properties. And to say that the properties in question are more basic than A is to say that the attribution of any such property to an individual is intelligible independently of, or prior to, thinking of him as possessing A. For example, to explain competence in a language L as partly consisting in a certain complex of intentions, beliefs, and other psychological states, we would, on this construal, have to be able to intelligibly attribute each of the relevant psychological states to a speaker independently of assuming that he has competence in L. Or, consider the present proposal that competence in L partly consists in knowledge of the truth-conditions of sentences of L. The requirement would be that this knowledge be attributable to an individual without having to draw on the assumption that he has competence in L.

On this construal of "consists in" claims, the above argument is unobjectionable. For, if truth is purely disquotational, an individual's ability to ascribe truth to sentences in L (and thus, his knowledge of the truth-conditions of these sentences) depends on his competence in L, and thus cannot be attributed prior to attributing this competence to him. Disquotationalism simply does not allow us to explain competence in a language in terms of a prior knowledge of the truth-conditions of its sentences.

But disquotationalism aside, is it plausible at all to think that linguistic competence could be reductively explained in terms of knowledge of truth-conditions? And is it plausible at all to read Davidson himself as trying to do this?

It is clear that if we are not construing truth as purely disquotational, knowledge of the truth-conditions of sentences in L need not be taken to depend on understanding these sentences, or on competence in L. However, it does not follow from this that this knowledge can be plausibly attributed to an individual prior to assuming that he has any linguistic abilities, or, competence in any language. In fact, insofar as knowledge of what is stated by T-sentences involves grasp of the concepts expressed by these biconditionals, it depends on whatever linguistic abilities grasp of the relevant concepts require. In particular, in the case of a Davidsonian truth theory, each target T-theorem is supposed to employ (more or less) the same concepts on the right hand-side of its biconditional as the concepts expressible in L by uses of the sentence mentioned on the left hand-side (since it is supposed to state the truth-conditions of this sentence in terms

that can be used to describe the thoughts expressed by utterances of these sentences on particular occasions). Thus, knowledge of what is stated by all the target biconditionals for L depends on grasp of more or less the same concepts as are expressible in L—in addition, of course, to grasp of a notion of truth applicable to sentences of L.

How might we make sense of attributing to a speaker grasp of the concepts expressible in L prior to attributing to her competence in L? We have two options here. The first would be to construe the speaker's grasp of the concepts expressible in L as involving her competence in some another language (in which, it bears emphasizing, the concepts expressible in L must be expressible). The second would be to construe the speaker's grasp of the concepts expressible in L as independent of, or prior to, her competence in any language.

If L is any language as rich as ours, the latter option is clearly implausible, both in its own right, and as a reading of Davidson. With regard to many of the concepts expressible in our language, it is hard to see how grasp of these concepts in the absence of a language would be possible, or what it would amount to.⁴⁶ Moreover, any attempt to explain linguistic competence in terms of a pre-linguistic grasp of concepts would be antithetical to Davidson's insistence on the interdependence of language and thought.

Now consider the first option. What we would be doing, in explaining a speaker's competence in a language L in terms of her knowledge of the truth-conditions of sentences of L, is explaining her competence in L in terms of her competence in some other language. But this is exactly what we would be doing in explaining the speaker's competence in L in terms of her knowledge of how to translate L into another language L'. The only difference, as Dummett explains, is that while the explanation in terms of knowledge of a translation theory from L to L' presupposes the speaker's competence in a specific language (L'), the explanation in terms of knowledge of truth-conditions presupposes the speaker's mastery of "*some, though unspecified language.*"⁴⁷ Neither

⁴⁶ Of course, the potential difficulties involved in explaining linguistic competence in terms of a prior grasp of concepts go beyond the one I am highlighting here. Another problem is that it is not immediately clear what role the assumption of an individual's prior grasp of a concept could be playing in an explanation of how he comes to use a word as expressive of that concept. As Dummett explains,

[i]t is dubious whether there is any way to explain what it is to take a word as expressing a certain sense save by describing the use made of the word which constitutes its having that sense. This, however, will be an explanation which, while not denying a prior grasp of that sense or concept, does not presuppose it, either, and which therefore simply fails to exploit the assumption of an antecedent grasp of the concept. (Dummett 1991, p. 111)

⁴⁷ "What is a Theory of Meaning," in Dummett 1993, p. 103-104.

sort of account can shed any light on what an individual's competence in her first language might consist in.⁴⁸

Of course, a third possibility would be to accord knowledge of truth-conditions a more limited role in a reductive account of linguistic competence. For instance, consider the following two-stage strategy. First, we demarcate a class of basic concepts in a speaker's repertoire, and explain her grasp of these concepts (and thus her competence with sentences expressive of these concepts) without attributing to her explicit knowledge of truth-conditions. Secondly, we try to explain the speaker's competence with other sentences in terms of her knowledge of the truth-conditions of these sentences, where the truth-conditions are spelled out in terms of the basic concepts singled out in the first stage.

This picture might well capture a conception Dummett once held of what a satisfying truth-conditional account of linguistic competence would look like⁴⁹, but it is not hard to see that it has no place in Davidson's approach. For Davidson clearly insists on keeping the goals of a meaning theory distinct from those of conceptual analysis. As he

⁴⁸ Horwich, who interprets the idea that linguistic competence consists in knowledge of truth-conditions in the reductive sense I am opposing, makes a similar point in Horwich 2005. He asks, "How might we come to know that 'Tachyons can travel back in time' is true iff tachyons can travel back in time?" (Horwich 2005, p. 68). His answer is this:

[T]he picture that comes to mind is that we deliberately associate the sentence 'Tachyons can travel back in time' with a possible state of affairs, where the form of association is our decision to count the sentence 'Tachyons can travel back in time' true if and only if the state of affairs obtains. ...

It is surely impossible for an individual to conceive of such an explicit association unless he employs some sort of mental event—call it 'R'—to represent the possible state of affairs ... we must raise the question of what provides representation R with its meaning. In order to avoid an infinite regress it must be conceded that certain representational entities obtain their content by means other than having been explicitly associated with possible state of affairs. (Ibid., pp. 68-69)

In other words, any purported explanation of competence in a language L in terms of an independently ascribable knowledge of the truth-conditions of sentences of L really only amounts to an explanation of competence in L in terms of grasp of another system of "representational entities" (e.g., a public language or a language of thought). But the question then arise of what grasp of this system of representations consists in—or, if we think that it does not make sense to speak of an individual's "grasp" of his language of thought, the question is what it is in virtue of which sentences in this language have the meanings they do.

⁴⁹ Such as, for instance, in "What is a Theory of Meaning? (II)," in Dummett 1993. In particular, see pp. 44-45 of that essay.

explains⁵⁰, it is not part of the point of the sort of theory he is after to effect any kind of breakdown or explication of the concepts expressible in a speaker's language. To see this, note, once again, Davidson's requirement that acceptable T-theorems are to spell out the truth-conditions of sentences of the language in terms of more or less the same concepts as are expressible in these sentences. The goal is to describe the truth-conditions of sentences of L in terms that can be used to describe the linguistic acts that could be performed by the speaker's utterances of these sentences⁵¹, rather than in terms of any more basic concepts the speaker's grasp of which might help explain her capacity to perform these acts.

1.6 In what sense can competence in a language be said to partly consist in knowledge of truth-conditions?

Let us take stock. Competence in a language consists, at least in part, in knowledge of the conditions under which each of its sentences would be true on particular occasions of utterances. This idea, I have claimed, should be made central in understanding and assessing Davidson's proposal to use truth theories as meaning theories. But as the arguments of the previous section show, a speaker's competence in a language cannot, in general, or in the most basic cases, be explained in terms of her prior knowledge of the truth-conditions of its sentences. What, then, might be the sense in which linguistic competence can be said to partly consist in knowledge of truth-conditions?

Here is how I think we should understand this idea: simply put, a speaker's knowledge of the truth-conditions of sentences of a language is a central aspect of what is involved in her understanding the language. We can accept this without demanding that the knowledge in question be ascribable to the speaker prior to assuming that she understands the language.

Let me explain. There are many capacities associated with someone's understanding of a language, such as, for instance, the capacity to engage in complex or abstract reasoning, to exchange information, express beliefs and intentions, perform various speech acts, and so on. A speaker's competence in a language is part of what enables her to do the many things that someone who does not understand a language is unable to do. Or better, someone who understands and uses a given language L does these things in a distinctive way. Part of what we are after, in spelling out what competence in L amounts

⁵⁰ See "Truth and Meaning," Davidson 1984, pp. 31-33.

⁵¹ For example, to describe what the speaker would be saying in uttering a given sentence on a particular occasion.

to, is an illuminating description of what this way of doing these things comes to—or, what is distinctive of the way in which a speaker of L does, or can do, these things.

A speaker's understanding of her language is, then, part of what explains her abilities to do the many things she can do with the sentences of her language. Further, we think of her competence as ranging over infinitely many sentences, and thus seek a theory that exhibits her understanding of the language as having a particular structure, and a finite basis. Part of what is attractive about the truth-conditional approach is the prospect of being able to provide a compositional account of a speaker's understanding of a language—by exhibiting the speaker's knowledge of the truth-conditions of infinitely many sentences as derived from her understanding of finitely many parts and of the significance of their modes of combination.

Two points should be clear by now. First, in proposing that competence in L partly consists in knowledge of truth-conditions for sentences of L, we are not proposing that this knowledge is what enables the speaker's understanding of L. This would depend on the possibility of ascribing to someone knowledge of the truth-conditions of sentences of L prior to ascribing to her competence in a language. Rather, we are spelling out a central aspect of what is involved in understanding L, where this understanding is part of what enables the speaker to do various things by using sentences of L.

Secondly, the proposal that competence in L partly consists in knowledge of the truth-conditions of sentences of L should not be taken to entail that anyone who knows the conditions under which each sentence of L is true is a competent user of L. Rather, what distinguishes someone who is competent in L from someone who is not is her (compositionally derived) knowledge of the truth-conditions of sentences of L, together with her capacity to use this knowledge in certain ways, in thought and communication.

As Richard Heck⁵² has proposed, the key insight here is that knowledge of truth-conditions is only part of a full story of what linguistic competence involves—it is the central, compositionally derivable core, but equally important is the speaker's capacity to draw on this knowledge in ways that are characteristic of a competent speaker. A full account of linguistic competence would, then, have to explain the role this knowledge plays in reason-giving explanations of the speaker's behavior, or, explain how the speaker draws on her compositionally derived knowledge of truth-conditions in using language the way she does. Promising starting points can be found in Rumfitt 1995 and Heck 2007. As both of these authors suggest, an account of the role of the speaker's knowledge of truth-conditions in particular communicative contexts will invoke the speaker's belief, or expectation, that her audience also knows, or will come to know, the truth-conditions of her sentences. And as Rumfitt further suggests, the speaker's

⁵² See Rumfitt 1995 and Heck 2007.

relevant attitudes will likely also include the expectation that her audience's knowledge of the truth-conditions of her sentences will itself be compositionally derived.⁵³

1.7 Conclusion

A speaker's competence in a language L consists in her compositionally derived knowledge of the conditions under which each of the infinitely many sentences of L is true, and her ability to draw on this knowledge in certain distinctive ways (to be spelled out by a full account of linguistic competence). This chapter has been devoted to clarifying and defending this idea, and highlighting its importance in understanding Davidson's proposal to use truth theories as meaning theories.

In the course of my arguments, I have drawn attention to two contrasting yet common pitfalls in thinking about the nature of truth theories serving as meaning theories. The first is a tendency to construe the main goal of meaning theorists as that of giving a compositional account of knowledge that would simply *suffice for* understanding a language. The second lies in interpreting the claim that competence in a language consists in knowledge of truth-conditions as part of an attempt to explain linguistic competence in terms of more primitive capacities. As I have argued, this reductive reading is presupposed by Dummett's argument against the possibility of combining a truth-conditional account of linguistic competence with a deflationary account of truth. Therefore, if a truth-conditional account of linguistic competence is plausible in its own right, though incompatible with a deflationary account of truth, we need a different kind of argument to explain this incompatibility.

⁵³ See Rumfitt 1995, p. 855.

Chapter 2

Deflationism and Knowledge of Truth-conditions

2.1 Introduction

Can disquotationalism be successfully combined with a truth-conditional approach to meaning? Davidson himself clearly viewed his own explanatory goals as incompatible with disquotationalism and other forms of deflationism about truth. He described his own proposal to use of truth theories as meaning theories as taking “truth to be the central primitive concept”, and trying, “by detailing truth’s structure, to get at meaning.”¹ However, we have yet to find a convincing argument for the incompatibility of truth-conditional semantics with a disquotational account of truth.²

On the flipside, none of the compatibility arguments to date³ engage with the conception of linguistic competence that I have claimed underlies the truth-conditional approach. As I have argued in Chapter 1, the point of a truth theory for a language L is to say what a competent user of L knows about the conditions under which its sentences are true, and to exhibit this knowledge as derived from a finite compositional basis. A central commitment of a truth-conditional approach to meaning is, thus, the idea that competence in a language involves knowledge of the truth-conditions of its sentences. A question we need to ask, then, is whether this commitment is compatible with the disquotational account of truth. In this chapter, I argue that it is not.

¹ Davidson 1984, p. xvi. See also Davidson 2005a, p. 54.

² For criticisms of various incompatibility arguments, see Horisk 2008, Lance 1997, and Williams 1999.

³ For some compatibility arguments, see Kölbel 2001, Lance 1997, Williams 1999. Since none of these arguments addresses the question whether disquotationalism allows us to think of linguistic competence as involving knowledge of truth-conditions, they do not settle the case in favor of compatibility.

Suppose ‘is true’ as I use it is a pure disquotational truth predicate, in the sense introduced by Hartry Field. This is a predicate that I can only meaningfully apply to sentences that I understand a certain way. And it is an exhaustive account of the meaning of this predicate that for each such sentence *s*, the sentence in which ‘is true’ is applied to a quotation name for *s* is “cognitively equivalent,” in some suitable sense, to *s*.⁴ Disquotationalists hold that our primary notion of truth is the notion of pure disquotational truth.⁵

Disquotationalism has appeared to be just what we need to secure our knowledge of the truth-conditions of our own sentences and of the references of our terms in the face of various arguments purporting to establish the indeterminacy or ungroundedness of semantic facts.⁶ Vann McGee, for instance, proposes a disquotational conception of reference—an extension of the disquotational conception of truth—as a solution to the following “paradox” about reference:

It is a plain fact that, when I use the word “rabbit,” I am referring to rabbits, and to nothing else. But when we look at all the nonsemantic facts about a speaker and her usage that determine what she means by the words she uses, we find that the totality of relevant non-semantic facts ... don’t suffice to determine whether a speaker, whether it’s me or a speaker of Jungle, uses a word to refer to rabbits or to undetached rabbit parts. How can this be? There surely aren’t free floating semantic facts, facts about meaning and reference that aren’t somehow grounded in nonsemantic circumstances.⁷

On the disquotational conception of reference, “prefixing “the referent of” to the quotation name of a singular term undoes the effect of the quotation marks, as does prefixing “the referents of” to a quotation name of a general term.”⁸ Thus, since the claim ‘ “Rabbit” refers to rabbits’ is, on this account, equivalent to ‘Rabbits are rabbits,’

⁴ “Deflationism about Meaning and Content,” in Field 2001, pp. 105.

⁵ Though a disquotationalist need not deny the intelligibility of other notions of truth applicable to sentences we do not understand, he thinks that any such notion would have to be definable in terms of pure disquotational truth, together with, as Field puts it, “fairly limited additional resources” (“Disquotational Truth and Factually Defective Discourse,” Field 2001, p. 223). In particular, any notion of translation or synonymy employed in characterizing truth for other languages would have to be independent of any notion of truth applicable across languages.

⁶ Such arguments abound in the literature. Among the most influential ones are those of Quine 1960 and Kripke 1982.

⁷ McGee 2005, pp. 403-404.

⁸ *Ibid*, p. 410.

it is clearly not one that it would make sense to justify by appeal to facts about my usage of the word ‘rabbit.’

In the same vein, the disquotational reading of ‘is true’ is supposed to explain how I can hold onto claims about the truth-conditions of my own sentences even if I am convinced that all the relevant facts about any given speaker’s use of words do not suffice to determine any particular way of assigning truth-conditions to her sentences as the correct one.^{9, 10}

But can disquotationalism really help explain our knowledge of the truth-conditions of our own sentences, or can it only explain our entitlement to assert unproblematic instances of the T-schema (sentences of the form < ‘s’ is true if and only if s>)? I shall argue that despite initial appearances, disquotationalism cannot explain our knowledge of truth-conditions, though there is a sense in which it can make sense of our entitlement to assert unproblematic instances of the T-schema. If, as I suspect, it is the former task that the disquotationalist is really interested in, and needs to satisfactorily address the difficulties faced by inflationary approaches to truth, then, in showing that he is not able to carry it out, we would have undermined one central motivation for his position.

In the next section, I say a bit more about what disquotationalism, as I understand it, comes to. I then examine Field’s and McGee’s apparent attempts to explain facts about

⁹ Note that the difficulties to which disquotationalism is supposed to provide an answer do not depend on embracing Field and McGee’s commitment to the reducibility of the semantic to the non-semantic, but can stem from a commitment to the publicness of meaning—or, some version of the idea that there are no “hidden,” or “private” semantic facts. Just as with the reducibility requirement, disquotationalism can be invoked as a way to hold on to the requirement of publicness while acknowledging that all the publicly available facts about a speaker are compatible with different ways of assigning truth-conditions to his sentences and reference to his terms. For indeterminacy claims based on publicness rather than reducibility, see Davidson 1984, and in particular, “Radical Interpretation,” “Belief and the Basis of Meaning,” and “Reality without Reference.”

¹⁰ Similarly, Field describes the following difficulty for an “inflationist” about reference:

Consider ‘rabbit’: an inflationist presumably thinks that the set or property that my term ‘rabbit’ stands for is determined from the facts about this world’s conceptual role for me ... this raises the question of precisely how it is determined ... I don’t say that the inflationist can’t tell a reasonable story about this, only that there is a story to be told, and perhaps there is room for skepticism about the possibility of telling it adequately. If so, that provides a motivation for deflationism. For the deflationist view is that there is nothing to explain: it is simply part of the logic of ‘refers’ (or ‘is true of’) that ‘rabbit’ refers to (is true of) rabbits and to nothing else. (“Deflationist Views of Meaning and Content”, Field 2001, p. 116)

Paralleling this claim about the logic of ‘refers,’ Field would say that it is simply part of the logic of ‘is true’ that ‘Snow is white’ is true if and only if snow is white.

truth, or the truth of our T-sentences, in terms of the logic of ‘is true.’ Arguing that we cannot take these appearances at face value (section 3), I sketch a more plausible interpretation (section 4), on which what Field and McGee are really trying to do is explain our *knowledge* of facts about truth, or of the truth of our T-sentences, in terms of our grasp of the meaning of ‘is true.’ Section 4 then draws attention to a crucial distinction between knowing the truth of a given sentence and knowing the truth that the sentence expresses. As I argue, it is plausible that in explaining a speaker’s knowledge of the truth (in her language) of a given sentence, we have thereby explained her knowledge of the truth (or fact) the sentence expresses, only insofar as we assume that the speaker knows the truth-conditions of her own sentences. In section 5, I sketch Boghossian’s account of epistemic analyticity, that is, of how mere grasp of the meaning of a sentence can justify us in holding it to be true, and examine what would be involved in extending it into an account of how mere grasp of the meaning of a sentence can justify us in believing the proposition it expresses. I then show (in section 6) why our knowledge of the truth-conditions of our sentences is too fundamental to be amenable to any such account based on epistemic analyticity. Finally, in section 7, I offer an alternative account of McGee’s and Field’s aims, on which they can, in some sense, be said to be explaining our knowledge of the truth-conditions of our sentences in terms of our grasp of ‘is true.’ but only by adopting a deflated conception of belief and a non-factive conception of knowledge.

2.2 A closer look at disquotationalism

Consider the following four types of claims, made by, or on behalf of, disquotationalists about truth:

- Claims pertaining to the role of ‘is true’ in our language. For instance, the claim that the sentences ‘Snow is white’ and ‘“Snow is white” is true’ are, as Field puts it, “cognitively equivalent” in our language, and the claim that our acceptance of the T-sentence ‘“Snow is white” is true if and only if snow is white’ is “more or less indefeasible.”¹¹
- Claims about the content of our claims about truth, such as the claim that to call the sentence ‘Snow is white’ true is to call snow white.¹²

¹¹ Field 2001, p. 106.

¹² For example, consider Quine’s remark that “[by] calling the sentence [“Snow is white”] true, we call snow white” (Quine 1970, p.12). See also Field 2001, p. 121-122:

[F]or me, the claim that utterance *u* is true in the pure disquotational sense is cognitively equivalent to *u* itself as I understand it ... [This] feature of the pure disquotational notion of truth means that this notion is of a use-independent property: to call ‘Snow is white’

- Claims about the nature of truth as a property of sentences. For instance, the claim that truth is a use-independent property of sentences.¹³
- Claims about the conditions under which our sentences are true, and our knowledge of these conditions. For instance, the claim that given the logic of ‘is true,’ I cannot be wrong in thinking that ‘Snow is white’ as I use it is true if and only if snow is white.

In advocating the disquotational account of truth, Field and McGee seamlessly go back and forth between these four types of claims. It is not clear, however, that there is a successful position that encompasses them all. What is clear, I think, is that the disquotational account of truth is, in the first instance, an account of the central role of talk about truth, or, of the locution ‘is true’ in our language.¹⁴ The disquotationalist’s claims about the content of our attributions of truth, the nature of truth, and our knowledge of truth-conditions are meant to flow from his account of ‘is true’ as a pure disquotational truth predicate. But is there a legitimate route, from the disquotationalist’s account of the role of ‘is true’, to these various other kinds of claims? I want to suggest that there is no such route, unless the disquotationalist helps himself to a pre-existing notion of truth, thus undermining his commitment to the primacy of the pure disquotational notion.

2.3 Some puzzling remarks

According to Field, an “inflationist” about reference faces the task of explaining how the reference of our words is determined by facts about their use, and the difficulties encountered in trying to give such an account provide a motivation for deflationism. As he sees it,

disquotationally true is simply to call snow white; hence it is not to attribute it a property that it wouldn’t have had if I and other English speakers had used words differently.

¹³ Ibid.. See also McGee 2005, p. 410:

For the disquotationalist, what makes “Snow is white” true is the whiteness of snow; nothing more is required. For the correspondence theorist, what makes “Snow is white” true are the whiteness of snow together with the linguistic practices of the community of speakers in virtue of which the sentence means what it does.

¹⁴ And, by extension, an account of the central role of similar locutions in other languages—what, by analogy with our own language, we identify as the “truth predicates” of other languages.

the deflationist view is that there is nothing to explain: it is simply part of the logic of 'refers' (or 'is true of') that 'rabbit' refers to (is true of) rabbits and to nothing else.¹⁵

Paralleling this claim about the logic of 'refers', Field would say that it is simply part of the logic of 'is true' that 'Snow is white' is true if and only if snow is white.

Similarly McGee argues that for the disquotationalist,

[b]oth of the following biconditionals are made true by the meaning of the word "true":

"Harry is bald" is true (in the present context) if and only if Harry is bald.

"Harry is not bald" is true (in the present context) if and only if Harry is not bald.¹⁶

If these claims are interpreted literally, Field seems to be trying to explain the fact that 'rabbit' refers to rabbits purely in terms our meaning what we do by the sentence "'rabbit' refers to rabbits.' Similarly, McGee seems to be proposing that we explain the truth of our T-sentences purely in terms of our meaning what we do by these sentences. But is it really plausible to think that Field and McGee are explaining either the truth of a sentence, or the truth that the sentence expresses, in terms of our meaning what we do by the sentence? Consider any alleged example of "truth by virtue of meaning alone", such as the sentence 'If Mia is happy and Tom is bored, then Mia is happy.' Though we might be tempted to describe this sentence as true simply by virtue of its meaning what it does (namely, that if Mia is happy and Tom is bored, then Mia is happy), it is, on reflection, hard to deny that the truth of this sentence also has to do with the fact that if Mia is happy and Tom is bored, then Mia is happy. And it does not seem plausible to explain *this* necessary fact about the world in terms of contingent facts about our linguistic practices.

For these sorts of reasons, many philosophers, following Quine¹⁷, are now sceptical of so-called "metaphysical analyticity," or, the idea of truth by virtue of meaning alone. But even if we could make sense of this idea, it still would not be compatible with the disquotationalist's claims about truth. A sentence that is true solely by virtue of its meaning, if there is any such thing, is a sentence whose truth at least depends on its

¹⁵ "Deflationist Views of Meaning and Content," Field 2001, p. 116.

¹⁶ McGee 2005, p. 412

¹⁷ See Quine 1953.

meaning. But pure disquotational truth, Field and McGee claim, is not a use-dependent property of a sentence, that is, it is not a property of a sentence that depends on any facts about its use. For instance, it is supposed to follow from the disquotational account of ‘is true’ that the truth of ‘Snow is white’ only has to do with the whiteness of snow, not with facts about its meaning or use.

2.4 Epistemic analyticity, belief, and the holding true of a sentence

In light of both reasonable doubts about the intelligibility of the metaphysical notion of analyticity and of its incompatibility with the disquotationalist’s claims about truth, we should find another reading for the claims quoted at the beginning of section 3. Despite initial appearances, it would be more plausible, and more in line with their explanatory goals, to take Field and McGee to be employing the epistemic notion of analyticity that Paul Boghossian has usefully distinguished from the metaphysical one. Following Boghossian, let us characterize a sentence as analytic in the epistemic sense if mere grasp of its meaning justifies us in holding it true.¹⁸ As Boghossian has argued, once we distinguish the two notions of analyticity, it is far from obvious that the legitimacy of the epistemic notion depends on the legitimacy of the metaphysical one. I would venture to add that it is also not *completely* obvious that the claim that we can know a sentence to be true solely by virtue of grasping its meaning is incompatible with the disquotationalist’s depiction of truth as a use-independent property of sentences.

Let us, then, re-interpret Field and McGee’s claims as claims pertaining to epistemic, rather than metaphysical, analyticity. Field’s claim, thus reformulated, comes to this:

Mere grasp of the meaning of the sentence ‘“Rabbit” refers to rabbits and to nothing else’ justifies us in believing that ‘Rabbit’ refers to rabbits and to nothing else.

Similarly, McGee’s claim would be:

For the disquotationalist, mere grasp of the meaning of each of the following two sentences justifies us in holding them true:

“Harry is bald” is true (in the present context) if and only if Harry is bald.

“Harry is not bald” is true (in the present context) if and only if Harry is not bald.

¹⁸ Boghossian 1997, p. 337.

I would like to draw attention to a subtle but important difference between Field's and McGee's claims, as here interpreted. Field is claiming that mere grasp of the meaning of the sentence ' "Rabbit" refers to rabbits and to nothing else' justifies us in believing the proposition it expresses. McGee, on the other hand, is claiming only that mere grasp of the meaning of his two sentences justifies us in holding them true. He says nothing about whether it justifies us in believing the propositions these sentences express. Boghossian's account of epistemic analyticity, as spelled out in Boghossian 1997, is primarily an account of how mere grasp of the meaning of a sentence can justify us in holding it true. Assuming that this account is plausible, can it also help explain how mere grasp of the meaning of a sentence can justify us in believing the truth it expresses?

Boghossian does intend for his account of epistemic analyticity to shed light on how mere grasp of the meaning of a sentence can justify us in believing the proposition it expresses, or, explain our knowledge of the truth that it expresses. In an effort to work with a picture of belief that is "hospitable to Quine's basic outlook" (as opposed to the conception, that he himself favors, of belief as a relation to a proposition), Boghossian proposes that

for a person T to believe that p is for T to hold true a sentence S which means that p in T's idiolect ...

[F]or T to know that p is for T to justifiably hold S true, with a strength sufficient for knowledge, and for S to be true. And to say that T knows p a priori is to say that T's warrant for holding S true is independent of outer, sensory experience.¹⁹

The idea that my believing that p involves my holding true a sentence S that means that p is quite plausible, if we assume that I have a language (or idiolect) in which my thoughts are expressible. But what accounts for its plausibility? Why is it plausible at all that when S means that p in my idiolect, my holding S to be true amounts to my believing that p? Only because, I want to suggest, S's meaning that p in my idiolect involves my knowing, or at least believing, that S is true (in my idiolect) if and only if p.²⁰

Similarly, consider Boghossian's further proposal that when S in my idiolect means that p, my knowing that p involves S's being true, together with my justifiably holding S true

¹⁹ Boghossian 1997, p. 333.

²⁰ By contrast, S's meaning that p in the language of my community does not depend on my taking S to mean that p, or to be true if and only if p, in this language. For, I can surely be mistaken about what a given sentence means in my linguistic community.

(with a strength sufficient for knowledge). This claim derives its plausibility from the assumption that if S means that p in my idiolect, I know that S is true (in my idiolect) if and only if p. In other words, what I am suggesting here is that the assumption that a speaker knows the conditions under which her own sentences are true plays a crucial role in tying the speaker's knowledge of the truth of her own sentences to her knowledge of the facts they express.²¹ This idea will play a key role in my arguments.

2.5 Boghossian's account of epistemic analyticity

At the heart of Boghossian's account of epistemic analyticity is an account of the epistemic analyticity of logic—that is, of how we can explain our knowledge of the truth of sentences of logic, or of the validity of certain inferences, purely in terms of our grasp of the meaning of logical vocabulary. His account centers on the following thesis:

Implicit definition. It is by arbitrarily stipulating that certain sentences of logic are to be true, or that certain inferences are to be valid, that we attach a meaning to the logical constants. More specifically, a particular constant means that logical object, if any, which would make valid a specified set of sentences and/or inferences involving it.²²

Boghossian relies on Implicit Definition to justify the following argument form, where C is a logical constant, and S is an inference form involving C:

- (1) If logical constant C is to mean what it does, then S has to be valid, for C means whatever logical object in fact makes S valid.
- (2) C means what it does.

Therefore,

- (3) S is valid.²³

Since I am primarily interested in Boghossian's account as it bears on our knowledge of the truth of our sentences (rather than our knowledge of the validity of our inference

²¹ For, suppose S means that p in a language L that I do not understand. I can know that S is true in L without grasping the meaning of S in L. In this case, my knowledge of the truth of S does not amount to my knowledge that p, because I do not know that S is true in L if and only if p.

²² Boghossian 1997, p. 348.

²³ Ibid., p. 357.

forms), I will recast his proposed argument form in terms of sentences and truth rather than inference forms and validity. Thus, where C is a logical constant and S a sentence form that is constitutive of C's meaning what it does, the argument is:

- (1) If logical constant C is to mean what it does, then S has to be true, for C means whatever logical object in fact makes S true.
- (2) C means what it does.

Therefore,

- (3) S is true.

I shall refer this as “the Basic Argument.”²⁴ Suppose this argument successfully explains how mere grasp of the meaning of a sentence could justify me in holding it true. Can it also help explain how mere grasp of the meaning of a sentence can justify me in believing the truth it expresses?

Note, first, that I could know both that an expression C is meaningful in a given language L (or, that it means what it does in L), and that the meaning of C in L is fixed by stipulating that a given sentence S is to be true (in L), without knowing what either C or S means. In other words, I could know (1) and (2) without grasping the meaning of C or S. In this case, Boghossian's proposed justification for (3) would not depend on my grasp of the meanings of S and C. This of course does not mean that Boghossian is wrong to assume that grasp of the meanings of S and C is one way for me to know (1) and (2). What is striking about the case in which I know (1) and (2) without knowing what either S or C means is that even if my knowledge of (1) and (2) can explain my knowledge of the truth of S, it is irrelevant to explaining my knowledge of the truth that S expresses. The question I want to ask is, what is the relevant difference between this case and the case in which I know what S and C mean? How is it that my mere grasp of the meanings of S and C can explain, not just my knowledge of the *truth of S*, by also, my knowledge of the *truth that S expresses*?

²⁴ As Boghossian acknowledges, this form of argument can explain our knowledge of the truth of sentences of logic as derived solely from our knowledge of the meaning of logical constants only if we assume that knowledge of the meaning of a logical constant includes knowledge of how its meaning is fixed. As Laurence and Margolis 2005 further point out, this would amount to an account of the a-priority of logic only if we further assume that our knowledge of the meanings of our logical constants, or our knowledge of their meaningfulness, is itself a priori. Margolis and Laurence go on to question whether Boghossian gives us any good reason for this assumption. It is not part of my purposes here to examine whether, in explaining our knowledge of the truth of sentences of logic in terms of our knowledge of facts about meaning, Boghossian is really showing our knowledge of the truth of these sentences to be a priori. Nor do I wish to take a stand on whether Boghossian's explanation of our knowledge of the truth of sentences of logic in terms of our grasp of the meaning of logical vocabulary is successful.

Consider the following example. Suppose ‘if then’ (construed as the material biconditional) comes to have the meaning it does in my idiolect by my stipulating that certain argument forms containing it are valid, or that certain forms of sentences are true. Suppose that <If p and q, then p> is one such sentence form, and consider the following instance:

[A] If Mia is happy and Tom is bored, then Mia is happy.

Here is the Basic Argument, applied to this example:

- (1) If ‘if then’ is to mean what it does, then [A] has to be true, for ‘if then’ means whatever logical object in fact makes [A] (and other sentences of the form <If p and q, then p>) true.
- (2) ‘If then’ means what it does.

Therefore,

- (3) [A] is true.

If this argument does what it is supposed to do, what we here have is an explanation of how I know that [A] is true purely by virtue of grasping its meaning. But how might this also help explain how I know the truth that [A] expresses purely by virtue of grasping its meaning? Here is another way of understanding this question: what further premises, in addition to (1) and (2), might help us derive not just the conclusion that [A] is true, but also, the conclusion that if Mia is happy and Tom is bored, then Mia is happy? What we need, of course, is a premise that is justifiable purely in terms of my grasp of the meaning of [A]. Here is an obvious candidate:

(T_A) [A] is true if and only if, if Mia is happy and Tom is bored, then Mia is happy.

Adding this premise, we can argue as follows:

- (1) If ‘if then’ is to mean what it does, then [A] has to be true, for ‘if then’ means whatever logical object in fact makes [A] (and other sentences of the form <If p and q, then p>) true.
- (2) ‘If then’ means what it does.

(T_A) [A] is true if and only if, if Mia is happy and Tom is bored, then Mia is happy.

- (3) [A] is true. (from (1) and (2))

Therefore,

(4) If Mia is happy and Tom is bored, then Mia is happy. (from (T_A) and (3))

Let us call this “the Expanded Argument.” This argument explains my knowledge of (4) purely in terms of my grasp of the meaning of [A] only if we assume that I know (T_A) purely by virtue of grasping the meaning of [A]. Thus, assuming that mere grasp of the meaning of [A] justifies me in holding it true, it also justifies my belief that if Mia is happy and Tom is bored, then Mia is happy, as long as mere grasp of the meaning of [A] justifies my belief that [A] is true if and only if, if Mia is happy and Tom is bored, then Mia is happy.

One might, however, argue that there is no need for the further step involving in deriving (4) from (3), since the Basic Argument, by itself, shows that I am justified in believing (4), simply by showing that I am justified in believing (3). For, when S in my idiolect means that p, believing that p is, in my case, a matter of my holding S to be true, and justifiably believing that p is, in my case, a matter of my justifiably holding S to be true. If this is right, then once we have explained how I can justifiably believe (3), we have thereby explained how I can justifiably believe (4).

But why is it plausible to think when S in my idiolect means that p, my believing that p is a matter of my holding S to be true? As I explained in section 4, this idea derives its plausibility from the assumption that when S in my idiolect means that p, my grasp of the meaning of S involves my knowledge that S is true if and only if p. If this is right, then the Basic Argument explains how I know (4) purely by virtue of grasping the meaning of [A], only on the assumption that my grasp of the meaning of [A] involves my knowledge of (T_A) .

Thus, whether by appeal to the Basic Argument or to the Expanded Argument, we can explain how mere grasp of meaning of [A] justifies me in believing (4), only on the assumption that my grasp of the meaning of [A] involves my knowledge of (T_A) . More generally, suppose S means that p in a given speaker’s idiolect. Let us assume we have successfully explained how the speaker’s grasp of the meaning of S justifies her belief that S is true (without drawing on the assumption that the speaker’s grasp of the meaning of S justifies her belief that p). This does put us in a position to explain how the speaker’s grasp of the meaning of S justifies her belief that p, but only if we assume that the speaker’s grasp of the meaning of S involves her knowledge that S is true if and only if p.

2.6 Knowledge of truth-conditions and grasp of ‘is true’

In light of these results, let us see what might be involved in trying to explain my knowledge of the truth-conditions of my sentences in terms of my grasp of the meaning of a pure disquotational truth predicate. Drawing on Boghossian’s account of how the meaning of logical terms is fixed, we could say that ‘is true’ has the meaning that it does

in my idiolect by virtue of my stipulation that it is to have whatever meaning, if any, would make unproblematic instances of the T-schema come out true. One may reasonably wonder whether it would be legitimate to saddle disquotationalists with this sort of account of how the meanings of logical expressions (including pure disquotational truth locutions) are fixed. On the other hand, this is exactly the picture that McGee seems to be painting in the following passage:

A disquotational perspective provides the desired asymmetry between our situation, speaking our native tongue, and the situation of the field linguist trying to interpret [an alien language]. If our language does not yet contain any semantic terms, we are able to introduce the phrase “refers in the language I actually now speak” into the language by stipulating that the phrase is to be used in such a way that the (R) sentences are to be true. But stipulative definition is something that can only be done from the inside ... We alone, of all the creatures in the universe, are able to stipulate how a word is to be used in our language, but other creatures are able to stipulate how words are to be used in their own languages. (McGee 2005, pp. 414-415)

Similarly, we could, on this picture, introduce the phrase ‘is true in the language I actually now speak’ into our language by stipulating that this phrase is to be used in such a way that unproblematic instances of the T-schema are to be true.

Consider the following unproblematic instance:

[M] ‘Snow is white’ is true if and only if snow is white.

Let us invoke the thesis of Implicit Definition to explain how mere grasp of [M] justifies me in holding it to be true. Here is the Basic Argument, applied to this example:

- (1) If ‘is true’ is to mean what it does, then [M] has to be true, for ‘is true’ means whatever logical object in fact makes [M] (and other unproblematic instances of the T-schema) true.
- (2) ‘Is true’ means what it does.

Therefore,

- (3) [M] is true.

What we really want to explain, however, is how mere grasp of the meaning of ‘is true’ justifies me in believing that ‘Snow is white’ is true if and only if snow is white (i.e., the proposition expressed by [M]). If we could draw on the assumption that my believing this is a matter of my holding [M] to be true, then showing this belief to be justified would simply involve showing that I am justified in holding [M] to be true. But again, if

what I argued in section 4 is right, this assumption is only plausible if we assume that grasp of the meaning of [M] involves knowledge that it is true if and only if ('Snow is white' is true if and only if snow is white).²⁵

Alternatively, we can try to justify my belief that 'Snow is white' is true if and only if snow is white by appeal to the following Expanded Argument:

- (1) If 'is true' is to mean what it does, then [M] has to be true, for 'is true' means whatever logical object in fact makes [M] (and other unproblematic instances of the T-schema) true.
- (2) 'Is true' means what it does.

(T_M) [M] is true if and only if ('Snow is white' is true if and only if snow is white).

- (3) [M] is true. (from (1) and (2))

Therefore,

- (4) 'Snow is white' is true if and only if snow is white. (from (T_M) and (3))

As already explained with respect to the previous example, this argument would show my belief in (4) to be justified purely on the basis of my grasp of the meaning of [M] only if I could be said to know (T_M) simply by virtue of grasping the meaning of [M].

Thus, whether we appeal to a version of the Basic Argument or of the Expanded Argument, the outcome is the same. Explaining my knowledge of the truth-conditions of 'Snow is white' in terms of my grasp of the meaning of its T-sentence (namely, [M]) involves explaining my knowledge of the truth-conditions of 'Snow is white' in terms of my knowledge of the truth-conditions of its T-sentence. More generally, explaining my knowledge of the truth-conditions of my object-level sentences (that is, sentences that are, on the face of it, not about language) in terms of my grasp of the meaning of a pure disquotational truth predicate involves explaining my knowledge of the truth-conditions of these sentences in terms of my knowledge of the truth-conditions of my T-sentences. Therefore, the attempt to so explain my knowledge of truth-conditions fails, simply because it involves explaining my knowledge of the truth-conditions of some of my sentences by presupposing my knowledge of the truth-conditions of others.²⁶

²⁵ I am using parentheses to disambiguate the structure of this sentence.

²⁶ Even worse, the sentences knowledge of whose truth-conditions is being presupposed (namely, my T-sentences) contain the sentences knowledge of whose truth-conditions is purportedly being explained (namely, my object-level sentences). But it is hard to see how my knowledge of the truth-conditions of my

Here is another way to see why my knowledge of the truth-conditions of my sentences cannot be explained in terms of my grasp of the meaning of a pure disquotational truth predicate. The alleged account I just sketched would explain my knowledge of the truth-conditions of my object-level sentences in terms of my grasp of the meanings of my T-sentences only if my knowledge of the truth-conditions of my T-sentences could itself be explained in terms of my grasp of their meanings. But it is hard to see how my knowledge of the truth-conditions of my T-sentences could be so explained unless my grasp of the meanings of my T-sentences simply involved knowledge of their truth-conditions. And it seems arbitrary to hold that my grasp of the meaning of my T-sentences involves knowledge of their truth-conditions, while denying that my grasp of the meaning of my other sentences (in particular, my object-level sentences) involves knowledge of their truth-conditions. But if we do take grasp of the meanings of my sentences, in general, to involve knowledge of their truth-conditions, then we can explain my knowledge of the truth-conditions of each of my sentences simply in terms of my grasp of its meaning, thus dispensing with a Boghossian-style analytic explanation of this knowledge.²⁷

The upshot is this. Whether or not we think that our knowledge of certain facts can be explained in terms of our grasp of the meanings of the sentences we use to state these facts, our knowledge of the truth-conditions of our sentences is too basic to be amenable to this kind of explanation. For, as I explained, knowledge of the truth-conditions of our own sentences is invoked in any purported explanation of our knowledge of facts in terms of our grasp of the logic, conventions or stipulations governing our use of words.

T-sentences could be more fundamental than my knowledge of the truth-conditions of the object-level sentences that they contain.

²⁷ As a matter of fact, I think we can explain a speaker's knowledge of the *truth of* unproblematic instances of the T-schema simply in terms of her grasp of their meanings, even if we cannot explain the speaker's knowledge of the *truths expressed by* these instances in terms of her grasp of their meanings. In other words, I think these instances are analytic in Boghossian's epistemic sense. For instance, consider [M]. If we take grasp of the meaning of a sentence to involve knowledge of its truth-conditions, then my grasp of the meaning of [M] explains my knowledge of the fact that [M] is true if and only if ('Snow is white' is true if and only if snow is white). But since the sentence 'Snow is white' is used in [M] (on the right-hand side of its biconditional), my grasp of the meaning of [M] depends on my grasp of the meaning of 'Snow is white.' And since grasp of meaning involves knowledge of truth-conditions, my grasp of the meaning of 'Snow is white' explains my knowledge that 'Snow is white' is true if and only if snow is white. Thus, since my grasp of the meaning of [M] explains both my knowledge that [M] is true if and only if ('Snow is white' is true if and only if snow is white) and my knowledge that 'Snow is white' is true if and only if snow is white, it explains my knowledge that [M] is true.

2.7 A more charitable reading?

If this is right, Boghossian's strategy for explaining our knowledge of certain truths in terms of our grasp of the meanings of the sentences we use in stating these truths is simply not applicable, when our target explanandum is our knowledge of the truth-conditions of our sentences.

One may wonder, however, if there is any point to arguing for this conclusion here, since Boghossian's strategy does not even seem available to Field and McGee, if we take disquotationalism seriously.

Here is why. Helping ourselves to a pre-existing notion of truth, we could introduce a pure disquotational truth predicate 'is true_{pd}' into our language by stipulating that that 'is true_{pd}' is only meaningfully applied to sentences of our own language, and that all unproblematic instances of the T-schema '<s> is true_{pd} if and only if s' are true. But this cannot be what the disquotationalist is really doing, since his main claim is that our primary notion of truth just is the pure disquotational one, and that we have fully captured the content of this notion by specifying the schema governing the use of 'is true' in our language. If this is right, then the use of 'is true' in our language could not be spelled out by invoking an antecedently intelligible notion of truth. Thus, if, as McGee contemplates, 'is true' is introduced into our language by way of a stipulation concerning how it is to be used, the relevant stipulation cannot consist in the decision to use 'is true' in such a way that (unproblematic) instances of the T-schema come out true. For, to repeat, this would involve specifying the role of the pure disquotational truth predicate by appeal to a pre-existing notion of truth.

How else, though, is the disquotationalist meant to specify the role of 'is true'? Field seems more concerned than McGee with trying to avoid semantic vocabulary in characterizing the role of 'is true.' He claims, in favor of his account of this role, that it "provides a way to understand disquotational truth independent of any nondisquotational concept of truth or truth-conditions (and independent of any concept of proposition)."²⁸ This suggests that he does recognize the need for a conception of the meaning or role of 'is true' that does not invoke any antecedently intelligible notion of truth.

As we have seen, Field spells out the role of 'is true' in terms of the "cognitive equivalence" of sentences²⁹ in which it occurs and sentences in which it does not.³⁰

²⁸ Field 2001, p. 106.

²⁹ Is 'is true,' as Field understands it, a predicate of sentence-types (computationally rather than orthographically individuated, as Field insists) or of utterances? He sends us mixed signals: on the one hand, he spells out his account in terms of the cognitive equivalence of sentences in which the truth predicate is used and sentences in which it is not. On the other hand, when handling ambiguity and

Connected to the disquotationalist's notion of cognitive equivalence is some notion of analyticity or conceptual necessity. Given the cognitive equivalence of each sentence *s* that we understand with the sentence in which the pure disquotational 'is true' is applied to a quotation name for *s*, Field characterizes instances of the T-schema as "conceptually necessary,"³¹ "more or less indefeasible,"³² or even as "more or less 'analytic' or 'logically true' for [us],"³³ and seems to be using these labels interchangeably.

Just as with the notion of cognitive equivalence, we need to make sure that the idea of a sentence's being "conceptually necessary" or "logically true for us" does not smuggle in a pre-existing, unexplained notion of truth. The disquotationalist thus needs to give us a bit more elucidation on how he could be understanding these notions, if not in terms of truth. For instance, "conceptually necessary" here cannot be elucidated in terms of "necessarily true," and "logically true" or "analytic" cannot really mean, respectively, "true by virtue of logic alone" and "known to be true by virtue of knowledge of meaning."³⁴ How, then, is the disquotationalist understanding these notions, if not in

indexicals, he claims that "strictly speaking," 'is true' as he is thinking of it is a predicate of utterances rather than sentence types ("Deflationist Views of Meaning and Content," in Field 2001, pp. 134-137).

Despite the problems posed by context-sensitivity, however, Field's notion of cognitive equivalence seems to make better sense if its relata are construed as sentence types rather than utterances. For as we will see, Field spells out cognitive equivalence for sentences in terms of their intersubstitutability in the context of larger sentences. It is hard to make sense of this if 'sentence' is interpreted as 'utterance' (rather than 'sentence type'), since it is sentences, not their utterances, which occur, and can be substituted for one another, in the context of larger sentences.

³⁰ Similarly, McGee, who credits Quine with the disquotational conception of truth, characterizes the role of the pure disquotational 'is true' as follows:

Quine [...] thought of suffixing the words "is true" to the quotation name of an English sentence as undoing the effect of the quotation marks. (McGee 2005, p. 410)

We could put this by saying that for Quine, the "effect" of suffixing 'is true' to the quotation name of an English sentence *s* is a sentence that is "cognitively equivalent" to *s*.

³¹ Field 2001, p. 114.

³² Ibid., p. 173.

³³ Ibid., p. 222.

³⁴ This might be why Field encloses the latter two expressions in quotes, and speaks of the relevant sentences as being "'analytic' for us," or "'logically true' for us" (rather than simply as "analytic" or "logically true"), as if to avoid suggesting that a non-deflationary notion of analyticity (or logical truth) is doing any work here.

terms of truth? Field does not say much about what cognitive equivalence comes to, though we do find him sketching his “own preferred reading” of this notion in a footnote:

My own preferred reading, for what it’s worth, is that to call two sentences that a person understands ‘cognitively equivalent’ for that person is to say that the person’s inferential procedures license a fairly direct inference from any sentence containing an occurrence of one to the corresponding sentence with an occurrence of the other substituted for it; with the stipulation, of course, that the occurrence to be substituted for is not within the context of quotation marks or an intentional attitude construction ... I would also take the claim of cognitive equivalence to imply that the inferences are more or less indefeasible. (More specifically, that they are empirically indefeasible, and close to indefeasible on conceptual grounds as well, and that the person is not in possession of defeaters for them).³⁵

The relevant notion of cognitive equivalence for sentences is, then, some notion of sameness of conceptual or inferential role, where inferential role is to be understood independently of any notion of truth. Accordingly, in characterizing a sentence as “indefeasible,”³⁶ the disquotationalist is describing an aspect of the inferential role of the sentence for us. An indefeasible sentence is one that enjoys a certain privileged status in our inferential practices. Clearly, some notion of a speaker’s “acceptance” of a sentence is playing a crucial—albeit implicit—role here. Strictly speaking, it is our acceptance of a sentence, rather than the sentence itself, that can be said to be or not to be “indefeasible” or “unrevisable.” Or, if we could speak of truth, we would say that it is our belief that a sentence is true, or our “taking it to be true” that is or is not indefeasible.

Whether we can really make sense of the relevant notion of “acceptance” of a sentence without helping ourselves to a notion of truth, is an important and difficult question. The disquotationalist needs to be able to do this, since he needs to be able to spell out the role of the pure disquotational truth predicate without invoking any pre-existing notion of truth or of truth-conditions. But my goal here is not to assess the prospects of a successful deflationism about truth, but only its compatibility with the idea that competent speakers know the truth-conditions of their sentences. So I will grant, for the sake of argument, that the disquotationalist has a suitable notion of acceptance of a sentence that does not depend on a prior understanding of truth.

³⁵ Field 2001, p. 106, fn. 2.

³⁶ Field characterizes unproblematic instances of the disquotation schema as “more or less indefeasible,” that is, as “empirically indefeasible, and close to indefeasible on empirical grounds,” in light of the semantic paradoxes (see Field 2001, p. 106, fn. 2). To simplify exposition, I am ignoring this qualification.

In light of these restrictions, let us consider [M] again. Here is what the Basic Argument looks like, if we replace talk of truth with talk of indefeasible acceptability:

- (1) If 'is true' is to mean what it does, then my acceptance of [M] has to be indefeasible, for 'is true' has whatever conceptual role in fact makes my acceptance of [M] (along with my acceptance of other unproblematic instances of the T-schema) indefeasible.
- (2) 'Is true' means what it does.

Therefore,

- (3) My acceptance of [M] is indefeasible.

What this argument explains is how mere grasp of [M] justifies me in taking my acceptance of [M] to be indefeasible. But what we want to explain is how mere grasp of [M] justifies me in believing that 'Snow is white' is true if and only if snow is white. How could a disquotationalist explain this?

There is no obvious way of generating a suitable version of the Expanded Argument here, since there is no obvious way to get from (3) to

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

For, recall that the disquotationalist cannot characterize acceptance of a sentence as a matter of taking it to be true. But as with the previous examples, and assuming our Basic Argument does what it is supposed to do, appeal to an Expanded Argument might not necessary here. Let us assume, then, that our Basic Argument successfully explains how mere grasp of the meaning of [M] justifies me in taking my acceptance of [M] to be indefeasible. Could explaining this amount to explaining how mere grasp of the meaning of [M] justifies me in believing the proposition expressed by [M]—namely, that 'Snow is white' is true if and only if snow is white?

It would, if we could make sense of the idea that believing that p is, in my case, just a matter of accepting 'p.' For if believing that p is, in my case, a matter of accepting 'p,' justifiably believing that p is, in my case, a matter of justifiably accepting 'p.' If this is right, there is no more to explaining how mere grasp of [M] justifies my belief that 'Snow is white' is true if and only if snow is white, than explaining how mere grasp of [M] justifies my acceptance of [M]. One small fly in the ointment here is that what the Basic Argument directly explains is how grasp of the meaning of [M] justifies me in taking my acceptance of [M] to be indefeasible, rather than explaining how this grasp justifies me in accepting [M]. But this does not seem to be a serious difficulty, since it is plausible that if I am justified in taking my acceptance of a given sentence to be indefeasible, then I am justified in simply accepting this sentence. If this is right, the Basic argument can be said to explain how mere grasp of [M] justifies me in accepting it. And if my

believing that p is a matter of my accepting 'p,' what we would thereby have explained is how mere grasp of [M] justifies me in believing that 'Snow is white' is true if and only if snow is white.

But can the disquotationalist appeal to the assumption that believing that p is, in my case, a matter of accepting 'p'? Yes he can, simply because he is free to define belief in terms of a prior notion of acceptance of a sentence. And this is exactly what Field does, by putting forward his "linguistic view of meaning and content attributions."³⁷ On Field's linguistic view, in saying that S in X's language means that p, I am simply saying that S "has the same meaning characteristics as the meaning characteristics of my actual use of ['p']."³⁸ And in saying that S believes that p, I am saying that X "accepts a sentence (or mental representation) which has a role in her psychology like the role that the sentence 'p' (or a mental representation that [I] associate with it) plays in [mine]."³⁹ In particular, on this view, in saying that my own sentence 'p' means that p, what I am saying is that 'p' as I understand it has the meaning characteristics it actually has, or, more simply, that 'p' as I understand it means what it actually does.⁴⁰ And in saying that I believe that p, I am saying that I accept the sentence 'p.'

The disquotationalist is, then, using a primitive notion of acceptance of a sentence and defining belief in terms of it. Since he is doing this without relying on any pre-existing notion of truth, his assumption that believing that p is, in my case, a matter of accepting 'p' is perfectly in line with his commitment to the primacy of pure disquotational truth.

The upshot is this. If we accept disquotationalism along with Field's linguistic view of content attributions, my grasp of the meaning of 'is true' does, *in some sense*, justify my beliefs about the truth-conditions of my sentences, or explain my knowledge of these truth-conditions. For, on this view, there is nothing more to my having the beliefs I do about the conditions under which my sentences are true than my accepting unproblematic instances of the T-schema. For example, there is nothing more to my believing that 'Snow is white' is true if and only if snow is white than my accepting the sentence "'Snow is white" is true if and only if snow is white.' Therefore, there is nothing more to my justifiably believing what I do about the truth-conditions of my

³⁷ Field 2001, p. 158.

³⁸ Ibid., p. 159. Field goes on to "refine" his proposal, by suggesting that we replace the appeal to sameness of meaning characteristics with an appeal to a suitable notion of similarity or equivalence of meaning characteristics. These details do not matter for my purposes here.

³⁹ Ibid., p. 163.

⁴⁰ This means that regardless of how I am using or understanding 'p,' I am taking 'p' to mean that p, since regardless of how I am understanding 'p,' I am taking 'p' to mean what it actually does.

sentences than my justifiably accepting unproblematic instances of the T-schema. But my indefeasible, and therefore justified acceptance of these instances is constitutive of my grasp of the meaning of 'is true,' which, on this account, means that my justifiably believing what I do about the conditions under which my sentences are true is itself constitutive of my grasp of the meaning of 'is true.'⁴¹

Recall the question I posed in my introduction:

Can disquotationalism really explain our knowledge of the truth-conditions of our sentences, or can it only explain our entitlement to assert unproblematic instances of the T-schema?

The conclusion we have arrived at is that the disquotationalist can, in some sense, explain how mere grasp of 'is true' justifies our beliefs about the truth-conditions of our sentences. This is because on the disquotationalist's view of belief (or at least, *one* disquotationalist's view of belief), there is nothing more to our believing what we do about the truth-conditions of our sentences than our "accepting" unproblematic instances of the T-schema.

The disquotationalist's position on the above question, then, seems to be that it presents us with a false choice. For, on the line of response I have just considered, the disquotationalist takes himself to have explained our knowledge of the truth-conditions of our sentences simply by virtue of having explained our entitlement to assert unproblematic instances of the T-schema. His considered view, then, seems to be that there is nothing more to our so-called "knowledge" of the truth-conditions of our sentences than our entitlement to assert, or our indefeasible acceptance of, unproblematic instances of the T-schema.

If this is right, it means that the disquotationalist is ultimately deflating not just truth, but knowledge as well. In particular, his conception of belief does not seem to make room for a factive conception of knowledge. As knowledge attributions as usually understood, in describing X as knowing that p, I am claiming that p, though I am of

⁴¹ We may reasonably wonder if this deflationary conception of belief is really compatible with deflationists' tendency to depict our knowledge of the truth-conditions of our sentences as simply a trivial by-product of our linguistic competence, rather than as being constitutive of it (for instance, see Horwich 2005, pp. 69-70). If my indefeasible acceptance of (or my disposition to assert) unproblematic instances of the T-schema in application to sentences in my idiolect is constitutive of my grasp of 'is true,' then it is constitutive of my grasp of the meanings of these instances. So if my language contains the pure disquotational 'is true,' my indefeasible acceptance of unproblematic instances of the T-schema is, after all, constitutive of my linguistic competence, rather than being a by-product of it. But on Field's linguistic view of belief attributions, this means that my justifiably believing what I do about the truth-conditions of my sentences is, after all, constitutive of my linguistic competence, at least as long as my language contains a pure disquotational truth predicate.

course doing more than this: I am describing X as standing in some privileged relation to the fact that p—a relation that we can, for our purposes, think of as a matter of justifiably believing that p. Suppose we tried to hold on to both this feature of knowledge attributions and Field’s linguistic view of content attributions. Then our depiction of *ourselves* as knowing that p would involve a rather odd combination of two disparate claims: on the one hand, a claim about the world, or extra-linguistic reality (namely, that p), and on the other, a claim about our relation to a given sentence (e.g., that we are justified in accepting the sentence ‘p’), with nothing connecting the two. For, without construing acceptance of a sentence as a matter of holding it to be true, and invoking our knowledge of the fact that ‘p’ is true if and only if p, our acceptance of ‘p’ does not constitute our taking a stance on whether or not p.⁴²

In light of this, it would be more charitable to take the disquotationalist to be simply rejecting a factive conception of knowledge, rather than saddle him with the inchoate conception of knowledge I have just sketched. But this means that the disquotationalist is ultimately securing our knowledge of the truth-conditions of our sentences only by reducing such knowledge to nothing more than our indefeasible (or justified) “acceptance” of certain sentences. His point, thus understood, is that a speaker’s so-called “knowledge of truth-conditions” is just a matter of her being able to use each sentence s of her idiolect interchangeably with the sentence in which ‘is true’ is attached to a quotation name for s. In particular, a speaker who “knows the truth-conditions of her sentences” in this deflated sense can affix the predicate ‘is true’ to the quotation name of a sentence that she accepts and can conversely remove ‘is true’ and disquote sentences mentioned in assertions of the form <‘s’ is true>. Beyond this ability, there is nothing more to a speaker’s knowledge of truth-conditions, and thus nothing to explain, or justify, in terms of the disquotational account of ‘is true.’

If this is right, the disquotationalist’s answer to the difficulties facing inflationist approaches to truth is not the hopeless attempt I have made it out to be, to explain a speaker’s knowledge of truth-conditions—in a non-deflated, factive sense of knowledge—in terms of her competence with a pure disquotational truth predicate. Rather, it is a matter of simply embracing a deflated picture of linguistic competence as involving nothing more than syntactic manipulation. In saying this, I do not simply mean to draw attention to the fact that deflationists cannot make sense of linguistic competence as *essentially* involving knowledge of truth-conditions. This is an outcome that most deflationists gladly accept. My point is that, having committed to a non-truth-

⁴² The point I am making here depends on a factive conception of knowledge, but it does not depend on thinking of knowledge as anything more than justified true belief. My complaint is *not* that the disquotationalist’s conception of knowledge leaves unconnected the fact that p with the provenance or justification of our belief that p.

conditional conception of linguistic competence, the deflationist cannot then go on to explain our knowledge of such facts as that ‘Snow is white’ is true if and only if snow is white as a by-product of this competence.

2.8 Conclusion

To echo McGee’s words, how is it that, unmistakably, ‘snow is white’ as I understand it is true if and only if snow is white, even if my linguistic usage (or the linguistic usage of my community) fails to determine this as being the case? If what I have argued in this chapter is right, disquotationalism cannot adequately answer this sort of question. It cannot explain my knowledge of the truth-conditions of my sentences in terms of the logic of ‘is true’, though it can characterize my indefeasible acceptance of unproblematic instances of the T-schema as being part of the logic of this expression.

Does disquotationalism allow us to think of linguistic competence as involving knowledge of truth-conditions? One might argue that the above considerations are orthogonal to this question. For, from the result that disquotationalism cannot itself justify our assumptions about the truth-conditions of our sentences, it does not seem to follow that disquotationalism is incompatible with the idea that linguistic competence involves knowledge of truth-conditions. After all, it is not clear how, or whether, more inflationary approaches to truth can explain a competent speaker’s knowledge of truth-conditions.⁴³ Why, then, should the disquotationalist’s inability to explain our knowledge of the truth-conditions of our sentences be particularly troublesome?

Here is why. If we begin with a conception of meaning that encapsulates truth-conditions, and a conception of linguistic competence as involving (propositional) knowledge of meaning, then a speaker’s knowledge of truth-conditions is built into our depiction of him as linguistically competent.⁴⁴ From this perspective, any threat to our knowledge of the truth-conditions of our sentences constitutes a threat to our very

⁴³ As we have seen, one motivation for the disquotational account of truth is the inflationist’s perceived inability to explain the special status of our beliefs about what our own sentences mean and the conditions under which they are true—particularly in the face of arguments purporting to establish various indeterminacies of meaning, reference, or truth-conditions. Whether or not disquotationalism succeeds in meeting this explanatory demand is beside the point here—it is enough to point out that if there is any difficulty here, it is not peculiar to disquotationalism.

⁴⁴ Note, however, that not all inflationists about sentential truth believe that knowledge of truth-conditions is essential to linguistic competence. Soames, for instance, adopts a conception of meaning that does, in some sense, encapsulate, or at least determine, truth-conditions, but he denies that knowledge of truth-conditions has any part to play in characterizing linguistic competence (Soames 1989).

conception of ourselves as linguistically competent.⁴⁵ By contrast, because a disquotationalist has to begin with a conception of meaning that does not encapsulate truth-conditions, whatever ground he has for attributing to competent speakers knowledge of the truth-conditions of their sentences will have to be derivable from his account of the role of ‘is true’ as a pure disquotational truth predicate. As I have argued, however, the disquotational account of the role of ‘is true’ cannot help establish a competent speaker’s knowledge of the truth-conditions of her sentences.

We can conclude, then, that disquotationalism does not allow us to think of linguistic competence as involving knowledge of truth-conditions, and is thus incompatible with a truth-conditional approach to meaning. Another outcome of our reflections is that if the disquotationalist avoids the difficulties or explanatory demands faced by truth-conditional approaches to meaning, it is not by establishing our knowledge of the truth-conditions of our sentences as a by-product of our linguistic competence. Rather, the disquotationalist’s answer to the difficulties facing inflationary approaches to truth⁴⁶ seems to ultimately reduce to the rather drastic strategy of making unavailable any notions of truth and reference in terms of which the alleged difficulties can be spelled out.

Before closing I want to consider one final response on behalf of the disquotationalist. Here is one way of trying to avoid the choice I have presented between portraying the disquotationalist as embracing a sceptical view of meaning, or taking him to engage in the hopeless attempt to explain our knowledge of truth-conditions in terms of our grasp of ‘is true.’ My whole argument began as an attempt to make sense of the disquotationalist’s apparent goal of explaining our knowledge of facts about truth in terms of our grasp of the meaning of a pure disquotational truth predicate. In particular, I have focused on the disquotationalist’s claim that it is part of the logic of ‘is true’ that, for instance, ‘snow is white’ is true if and only if snow is white. One possibility I have not considered is that when Field and McGee say such things, they do not really mean to be explaining either facts about truth, or our knowledge of such facts, in terms of facts about our use of ‘is true.’⁴⁷ Rather, talk of the logic of ‘is true’ might simply be a

⁴⁵ While there might be some question as to whether, or how, we know what we think we know about the conditions under which our sentences are true, there is, from this perspective, no problem of explaining our knowledge of truth-conditions in terms of our linguistic competence.

⁴⁶ Or, perceived difficulties—whether or not the alleged difficulties are legitimate is beside the point here.

⁴⁷ In fact, in other places, Field himself expresses skepticism about the prospects of justifying any of our beliefs in terms of our grasp of the meanings of our words. See Field 2000 and 2005. This raises the question of how his claims in “Deflationist Views of Meaning and Content” are to be understood, if not as I am understanding them here. I should note, however, that “Deflationist Views of Meaning and Content” was first published in 1994, and thus, as far as I know, predates Field’s arguments against an analytic explanation of the a priori.

way of talking about truth itself. On this reading, when the disquotationalist says “It is part of the logic of ‘is true’ that ‘Snow is white’ is true if and only if snow is white,” what he is really saying is that our knowledge of the fact that ‘Snow is white’ is true if and only if snow is white counts as logical knowledge. In other words, what he is really saying is that our knowledge of this fact is akin to our knowledge of such facts as that if Mia is Happy and Tom is bored, then Mia is happy. Now this claim is plausible, if only because of the fundamental role that our knowledge of the truth-conditions of our sentences plays in our cognitive lives. But if what I have argued in this chapter is right, we cannot do justice to this role if we think of our grasp of truth as exhausted by our competent use of a pure disquotational truth predicate.

Chapter 3

Truth Theories as Meaning Theories: What's Interpretation Got to Do with It?

3.1 Introduction

Here is a crucial though seldom discussed question concerning the role of the notion of interpretation in Donald Davidson's philosophy of language. Does Davidson's account of radical interpretation presuppose an independent conception of meaning and of the point of a meaning theory? Or is the nature of a meaning theory, and the motivation for using truth theories as meaning theories, meant to be explained in terms of the ends and means of interpretation?

I suspect that most of those familiar with Davidson's writings would be inclined to accept Michael Williams's way of characterizing his approach to meaning:

If we (a) determine the theoretical form of the knowledge at which interpretation aims and (b) fix the methodology of interpretation, then we have said all there is to say about meaning. For Davidson, meaning just *is* whatever the practice of interpretation reveals.¹

This would not be surprising, since Davidson himself often presents his proposal to use truth theories as theories of meaning as primarily an answer to a certain question, or "problem," about interpretation. "Radical Interpretation," most notably, begins with the following queries:

Kurt utters the words 'Es regnet' and under the right conditions, we know that he has said that it is raining. Having identified his utterance as intentional and

¹ Williams 1999, p. 553.

linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it?²

He then goes on to generalize these questions, asking, “What knowledge would serve for interpretation?”³ The answer he goes on to give is that knowledge of an appropriately constrained Tarski-style truth theory would serve to interpret a speaker’s words. It would seem, then, that concern with the ends and means of interpretation is what explains and motivates Davidson’s proposal to use truth theories as meaning theories.

I shall argue against this widespread assumption. In the next section, I clarify certain aspects of the problem Davidson sets out to solve in “Radical Interpretation.” Section 3 spells out Davidson’s answer to this problem. Sections 4 and 5 examine two different accounts of the point of a meaning theory that this answer has generated. The first, “instrumentalism,” portrays a Davidsonian meaning theory as simply a recursive device for generating interpretations of a speaker’s utterances. The second, a joint proposal by Ernest Lepore and Kirk Ludwig, characterizes a meaning theory as an explicit statement of what we could know, about a compositional truth theory for a given language, that would put us in a position to interpret any arbitrary utterance in the language. I argue (in section 6) that Lepore and Ludwig’s account fares just as badly as the instrumentalist account: neither account adequately spells out the sense in which a truth theory can, while a translation theory cannot, be used to give a compositional account of our knowledge of a language. As I go on to explain, we can give substance to this contrast only by thinking of a meaning theory as a compositional account of a speaker’s understanding of her own language, rather than of the knowledge that would put us in a position to interpret the speaker’s utterances. But doing this involves forgoing the project of explaining the point of a meaning theory in terms of the ends and means of interpretation. In section 7, I show how thinking of a truth theory as a description of what a competent speaker knows about her own sentences puts us in a position to explain the otherwise elusive idea that a truth theory can be used to *show*, without itself *saying*, what sentences of a given language mean. Finally, in section 8, I provide further textual evidence in support my interpretation of Davidson, and explain away apparent evidence against it.

² Davidson 1984, p. 125.

³ *Ibid.*, p. 126.

3.2 Infinite competence and knowledge of literal meaning

When Davidson asks what we could know that would suffice for interpretation, he is not asking “What could we know that would suffice for understanding a speaker’s words as uttered on a particular occasion?” If this were the question Davidson really meant to ask about Kurt in the opening paragraph quoted earlier, a plausible answer would be: “That Kurt’s words, on the relevant occasion, meant that it is raining.” Clearly, this is something we could know even if we are not in a position to say what any other sentences in his language would mean on this or any other occasions of utterance.⁴ Indeed, Davidson makes it clear, by what he goes on to say, that what he is seeking is a “theory” knowledge of which would put us in a position to understand any of the potential infinity of sentences in the speaker’s language, as uttered on particular occasions. It is as an answer to this query that his proposal that knowledge of an appropriately constrained truth theory would “suffice for interpretation” is to be understood. Thus, the assumption that a speaker has, on any given occasion, linguistic abilities spanning infinitely many sentences is simply built into Davidson’s way of spelling out the “problem” of interpretation, rather than invoked to explain the possibility of interpreting an utterance on a particular occasion.⁵

But what does it take to interpret, or understand, a speaker’s utterance on a particular occasion? We think of ourselves as having understood a speaker’s utterance on a particular occasion if we understand what the speaker is doing in uttering the sentence on that occasion—in particular, what linguistic acts, and with what contents, the speaker is performing in uttering the sentence. Understanding an utterance, in this sense, involves understanding or interpreting the speaker, and thus depends on knowing a great deal about his beliefs and other psychological states. It is not plausible to expect that any systematic knowledge of a speaker’s “language” could put us in a position to interpret the speaker—to individuate his attitudes and determine the range of linguistic acts he would be performing on particular occasions of utterance.⁶

⁴ And without knowing what the speaker believes or what linguistic acts he is performing in making this utterance.

⁵ Pagin seems to have missed this crucial point in Pagin 1999, a paper devoted to the question whether compositionality is compatible with radical interpretability. In reply to Pagin, Davidson wrote that “radical interpretation,” as he understands it, simply cannot conflict with compositionality, since we should “view it as a given that any theory of meaning is compositional and then, and only then, ask how we can tell that a speaker is speaking in accordance with a specific compositional theory. It is only at this point that radical interpretation has a role to play” (Davidson 1999, p. 74). Thus, in asking whether any justification for compositionality can be derived from Davidson’s account of radical interpretation, Pagin seems to have misunderstood what this account is an account of.

⁶ The point I am making here is not the one Davidson makes in “A Nice Derangement of Epitaphs” (Davidson 2005b, pp. 89-107), when he argues that there is no systematic body of knowledge, acquired

Clearly, in asking what knowledge would enable us to interpret any utterance in a given language, we are assuming that there is a core component of the significance of an utterance that derives from a compositional account of the language as a whole, and on which the broader significance of the utterance depends. Of course, which aspect of the significance of an utterance is amenable to a compositional account, and how exactly it is to be described, are subjects of considerable disagreement amongst philosophers of language. Davidson himself speaks of a meaning theory as

describing the critical core of the speaker's potential and actual linguistic behavior—in effect, how the speaker intends his utterances to be interpreted. The sort of understanding is restricted to what we may call the literal meaning of the words, by which I mean, roughly, the meaning the speaker intends the interpreter to grasp, whatever further force or significance the speaker may want the interpreter to fathom.⁷

Thus, when Davidson asks, what could we know that would suffice for interpreting any arbitrary utterance in a language, he is asking what we could know that would put us in a position to determine the literal meaning of each utterance in the language⁸—for some suitable notion of literal meaning. While he often uses the term “interpretation” in the broader sense of interpreting a speaker, I will, in this chapter, be mostly concerned with his notion of “interpreting an utterance” in the sense of understanding, or determining, its literal meaning.⁹

prior to particular occasions of interpretation, that will tell us how a speaker is to be understood on any given occasion. To make that point, Davidson argues that a speaker's language can change from one occasion to the next without undermining successful communication. The problem I am getting at here is, rather, that there are too many things involved in understanding a speaker on a particular occasion, too many dimensions along which communication can be, or fail to, be successful. It would be absurd to expect a systematic account of all the linguistic acts a speaker would be (or could be) performing by uttering any of the infinitely many sentences in his language on particular occasions. It is not even clear that for each such sentence, and each possible occasion of utterance, there is a fact of the matter as to which linguistic acts a speaker would be performing in uttering the sentence on that occasion.

⁷ Davidson 2005a, p. 53.

⁸ On Davidson's idiolectal conception of a language, the “literal meaning” of a sentence as uttered on a particular occasion need not correspond to any conventionally determined meaning, or to the meaning it would have in a relevant linguistic community.

⁹ Davidson uses the verb ‘to interpret’ in at least the following three senses. He speaks of interpreting a speaker's words (on a particular occasion of utterance), in the sense of saying what they mean (or would mean, as uttered on that occasion), interpreting the speaker's utterance of these words, in the sense of saying what linguistic acts the making of that utterance involves, and finally, interpreting the speaker himself, in the sense of ascribing linguistic acts, beliefs and other propositional attitudes to him. Despite the occasional ambiguity, the context of his remarks usually sufficiently disambiguates his usage.

3.3 What could we know that would suffice for interpretation?

What is Davidson's answer to the problem of interpretation? What *could* we know that would suffice for grasping the literal meaning of any arbitrary utterance in the speaker's language?

Since Davidson's answer involves an appeal to truth theories, it is natural to suppose that this answer is, simply, that knowledge of an acceptable truth theory for a speaker's language (i.e., knowledge of the facts stated by the theory) would suffice for grasp of the literal meaning of the speaker's utterances. Indeed, this is precisely what Davidson seems to proposing in the following remarks:

[A truth theory] gives the substance of what a knowledgeable interpreter knows which enables him to grasp the meaning of the speaker's utterances.¹⁰

A theory of truth for a speaker is a theory of meaning in this sense, that explicit knowledge of the theory would suffice for understanding the utterances of that speaker.¹¹

Of course, Davidson cannot be claiming that any truth theory for a speaker gives the substance of what we know that enables us to interpret the speaker's utterances. It is knowledge of an *acceptable* truth theory that is in question here. What might be the relevant notion of acceptability?

We know what acceptability meant for Tarski. He famously imposed the following condition of adequacy, Convention T, on his truth definition for particular languages:

For any sentence of a language L, the recursive definition of 'is true-in-L' is to entail a theorem of the form <s is true-in-L if and only if p > (or T-sentence), where 's' is replaced by a structural description of the sentence, and 'p' is replaced by the sentence itself (if the metalanguage contains L), or by an appropriate translation of this sentence in the metalanguage.

Does Convention T help us understand Davidson's own requirements on acceptable truth theories? We know that Davidson and Tarski had different goals in seeking recursive characterizations of truth. Tarski's goal was to define logically well-behaved predicates capturing the extension of the concept of truth for particular languages. The point of Convention T, for him, was simply to ensure extensional adequacy—to ensure

¹⁰ Davidson 2005a, p. 52

¹¹ Ibid., p. 53.

that for each language L, his predicate ‘is-true-in-L’ picks out all and only the true sentences of L. Davidson, by contrast, wants to put recursive characterizations of truth to use as meaning theories. Thus, acceptability of a truth theory for a given language L is, for him, a matter of its being able to serve as—or serve the needs of—a meaning theory.¹² And in answering the question of how we determine whether a truth theory for a given language is acceptable, Davidson, unlike Tarski, wants to avoid reliance on any assumptions involving the notions of meaning and translation.

However, as evidenced by the following remarks, Davidson is not denying the relevance of Convention T to his own aims, but rather, trying to shed light on notion of translation on which it depends:

Since Tarski was interested in defining truth, and was working with artificial languages where stipulation can replace illumination, he could take the concept of translation for granted. But in radical interpretation, this is just what cannot be assumed. So I have proposed instead some empirical constraints on accepting a theory of truth that can be stated without appeal to such concepts as those of meaning, translation, or synonymy ... I have tried to show that if the constraints are met by a theory, then the T-sentences that flow from that theory will in fact have translations of s replacing ‘p’ [in theorems of the form <s is true-in-L if and only if p>].¹³

As Davidson further explains, “[t]o accept this change in perspective is not to give up Convention T but to read it in a new way.”¹⁴ Thus, when Davidson claims that explicit knowledge of an adequate truth theory would suffice for understanding the speaker’s utterances, we can take him to be claiming is that explicit knowledge of a translational truth theory¹⁵ would suffice for understanding the speaker’s utterances.

¹² The question I am asking in this chapter is whether “serving the needs of a meaning theory” for a given language can be explained as a matter of spelling out the content of knowledge that would enable us to interpret utterances in the language (in the sense of grasping their literal meaning).

¹³ Davidson 1984, pp. 172-173.

¹⁴ Ibid., p.173.

¹⁵ A truth theory for L is translational, in the sense relevant here, if it yields translational T-theorems, and a T-theorem is translational if the metalanguage sentence used on the right-hand side of it biconditional translates the sentence of L denoted on its left-hand side. Similarly, let us call a T-theorem “interpretive” if the metalanguage sentence used on its right-hand side can be used to interpret utterances of the sentence of L denoted on the left-hand side. Since our primary goal is to interpret utterances of sentences of L, rather than translate these sentences into sentences of our own, what we are really seeking are interpretive T-theorems. And given the presence of context-sensitivity in L, it would be wrong to say that, in general, interpretive T-theorems for sentences of L are translational ones, or that a truth theory for L can help us interpret utterances in L if it satisfies Convention T. However, to simplify exposition, I am

But as Davidson himself acknowledges in some of his more guarded moments, knowledge of what is stated by a translational truth theory for L is simply not sufficient for grasp of the literal meaning of any arbitrary utterance in L. This is because I can know what is stated by a translational truth theory without knowing that the truth theory is translational.¹⁶ Thus, even if we assume that Davidson's account of radical interpretation imposes constraints on a truth theory sufficient to guarantee that its T-theorems are translational, it does not follow that knowledge of a theory that satisfies these constraints is sufficient for interpreting utterances in L. There is, however, something we could know, in addition to knowledge of what is stated by a translational truth theory for L, that would put us in a position to grasp the literal meaning of any arbitrary utterance in L. This is the fact that this knowledge is entailed by such a theory. Davidson thus ends up accepting John Foster's appraisal of the situation: though knowledge of the facts entailed by a translational truth theory for L would not suffice for interpreting utterances in L, knowledge of the facts entailed by a translational truth theory for L, together with knowledge that these facts are entailed by such a theory, would suffice for interpreting any arbitrary utterance in L.¹⁷

3.4 Instrumentalism

If this is Davidson's answer to the question "What knowledge would suffice for interpretation?", it is hard to disagree with it. What is not clear, however, is what role the notion of a truth theory, or of truth itself, is really playing in an adequate account of what knowledge that would suffice for interpretation.

mostly ignoring context-sensitivity in what follows, and will thus think of Convention T as the criterion that a truth theory has to satisfy in order to serve for interpretation. In doing this, I am making the following methodological assumption: the main contrast between truth theories and translation manuals, as far as their ability to serve as meaning theories, is independent of context-sensitivity in the languages we are interested in. Thus, in trying to give substance to this contrast, we can simply ignore context-sensitivity.

¹⁶ As Davidson ends up conceding to John Foster, even when an interpreter "has a theory that satisfies Convention T, nothing in the theory itself tells him this" (Davidson 1984, p. 173).

¹⁷ Prompted by Foster's objections (Foster 1976) to his earlier statements of what knowledge would suffice for interpretation, Davidson ends up accepting Foster's proposal that "what we need to know, for the mastery of L, are both the facts which [a truth theory for L] states, and that those facts as are known by us, are T-theoretical"—that is, entailed by a truth theory that satisfies Convention T (Davidson 1984, p. 174). Of course, Davidson himself thinks that a statement of what someone needs to know about a truth theory would not qualify as a theory "in the formal sense." This is because such a statement will have to include claims about what the truth theory "states," which, on Davidson's paratactic account of indirect discourse, introduce "irreducible indexical element in the sentences that express it" (Davidson 1984, p. 179).

Indeed, a number of philosophers¹⁸ have argued that the work done by a truth theory in enabling us to interpret a speaker's utterances can be carried out whether or not we take its T-predicate to express a substantive notion of truth, and is thus compatible with a deflationary account of truth. Assuming that the point of a Davidsonian meaning theory is to put us in a position to interpret a speaker's utterances, they have plausibly argued that T-sentences can play a role in enabling us to do this without drawing on our understanding of a non-deflationary notion of truth.

For instance, Michael Williams argues that using a truth theory as a meaning theory for L does not involve explaining facts about what sentences of L mean as consisting in facts about their truth-conditions. Rather, as Williams sees it, a meaning theory is just a "recursive device for specifying the meaning of every sentence of a given language."¹⁹ Why, though, should such a theory center around the construction of a truth theory? Williams answers this question along familiar lines:

In specifying the meaning of sentences of another speaker's language, we associate his sentences with sentences of our own. In doing so, we make use of the truth predicate, which is what lends color to the idea that Davidson explains meaning in terms of truth-conditions. But the use of 'true' in a Davidsonian meaning specification for a particular speaker is expressive, not explanatory. He eschews 'means that' in favor of the material biconditional ... To replace 'means that' with a sentential connective, we need a sentence on the left side ... This is precisely what 'true' allows us to form.²⁰

On this picture, the function of a meaning theory for L is to recursively generate theorems in which sentences of L are paired (or "associated") with metalanguage sentences that "interpret them" (or, that can be used to interpret utterances of these sentences in L). A truth theory for L serves as a meaning theory for L insofar as it recursively generates such pairings. Thus, just as the truth theory is a recursive device for pairing sentences of L with metalanguage sentences in use, its T-predicate is a syntactic device that enables us to recursively generate such pairings.

It would not be incorrect to say that on this view, a translational truth theory for L serves a meaning theory for L not in virtue of what it says about sentences of L, nor in virtue of its saying anything at all about these sentences. This, as far as I can see, is the only explanation as to why Williams thinks that 'is true' can be taken to express a deflationary notion of truth compatibly with its doing the work it is supposed to do in a

¹⁸ In particular, see Williams 1999 and Kölbel 2001.

¹⁹ Williams 1999, p. 553.

²⁰ Ibid., p. 557.

truth theory. For, ‘is true’ can play its role as a recursive device regardless of what it is taken to mean, and arguably, regardless of whether it is taken to mean anything at all. The only crucial constraint on our interpretation of ‘is true,’ as far as allowing us to form metalanguage sentences in which sentences of L are paired with “interpretations” in the metalanguage, is that it be construed as of the right syntactic type—substituting ‘is true’ for ‘...’ should turn the following into a sentence:

‘La neige est blanche’ ... if and only if snow is white

A corollary to this result is that in order to play their role in a truth theory serving as a meaning theory for L, T-theorems need not make any true claims about sentences of L. And even when these theorems do express truths about the truth-conditions of sentences of L, knowledge of these truths does not, after all, play an essential role in enabling us to interpret utterances in L. For, suppose we do not know whether or not a given translational truth theory for L, which we know to be formulated in our own language, entails only truths about sentences of L.²¹ Let us assume that we have knowledge of a canonical proof procedure for this truth theory, know what its axioms are, and know that all of its canonically derived T-theorems are translational. Then for any given sentence *s* of L, we can derive a canonical T-theorem for *s*. And having done so, we are in a position to interpret any arbitrary utterance of *s* in L, since we know that the T-theorem derived for *s* is translational and understand the metalanguage sentence used on the right-hand side of its biconditional. Thus, despite not knowing either the truths stated by the T-theorems or whether these T-theorems do state truths, what we do know about the truth theory puts us in a position to interpret utterances in L.

3.5 Lepore and Ludwig’s proposal

Williams does not hesitate to describe a recursive truth theory *as* a meaning theory. For, after all, a meaning theory is for him just a recursive device for generating pairings of sentences of a language with metalanguage sentences in use that interpret them. But as we have seen, this is not quite how Davidson sees things in “Radical Interpretation.” A meaning theory as he conceives of it is not a mere recursive device, if it is that at all, but a statement of what we could *know* that would suffice for interpretation. How might we reconcile this conception of the point of a meaning theory with the conception of a truth theory as a recursive device for generating interpretations? This is precisely what Ernest Lepore and Kirk Ludwig seem to be trying to do, in proposing the following account of the role of a truth theory in a meaning theory.

²¹ For instance, suppose we are interpreting ‘is true’ in this theory as a truth predicate, but the arguments of Ludwig and Ray 2002 have convinced us that there are no true claims to be made about the conditions under which sentences containing vague expressions are true.

Lepore and Ludwig accept Davidson's idea that a meaning theory for a given language is to spell out the content of knowledge that suffices for interpreting utterances in the language. Only, they argue that Davidson and many of his commentators err in describing the knowledge in question as knowledge of a truth theory, or even, as including knowledge of the alleged facts entailed by a truth theory. Their proposal is to think of a meaning theory for L as an explicit statement of what we could know, about a compositional truth theory for L, that would put us in a position to interpret any arbitrary utterance in L. This includes knowledge of what the theory's axioms are, of what each axiom means or states, and knowledge of the fact that the theory satisfies a certain "interpretiveness" requirement, but it does not include knowledge of any alleged facts entailed by the truth theory.

Here is their suggested outline for "an explicit compositional meaning theory stated in terms of knowledge of an interpretive truth theory":

[1] Every instance of the following schema is true:

For all speakers S, times t, s for S at t in L means that p iff it is canonically provable on the basis of the axioms of an interpretive truth theory T for L that for all speakers S, times t, s for S at t is true in L iff p.

[2] τ is an interpretive truth theory for L whose axioms are...

[3] Axiom ... of τ means that...
Axiom ... of τ means that...
...

[4] A canonical proof in τ is ...²²

An interpretive truth theory, as Lepore and Ludwig are construing this notion, is a theory whose axioms are "interpretive"—where, roughly, an axiom is interpretive if it states the semantic contribution of an object-language expression using a metalanguage expression that translates it.²³ In [1], the notion of a canonical proof procedure is

²² Lepore and Ludwig 2005, p. 120.

²³ See Lepore and Ludwig 2005, p. 72 for the notion of an "interpretive axiom."

invoked to guarantee that each T-theorem whose proof only draws on the content of the axioms is itself interpretive.²⁴

The goal here is to finitely generate, for each sentence of the object-language, a uniquely identifiable metalanguage T-sentence, from which a corresponding M-sentence—i.e., a sentence of the form <For all speakers S, times t, s (for S at t in L) means that p>—can be derived. For, recall that what we ultimately want to do is interpret utterances of sentences of L, and we are in a position to do this if we know, for each sentence s in L, what any arbitrary utterance of s would literally mean. Assuming we can successfully construct truth theories that have the requisite properties, I think we have to agree with Lepore and Ludwig that knowledge of the sort of explicit “meaning theory” they describe (hereafter, an “LL-theory”) would indeed suffice for interpretation. But in what way does this vindicate the pursuit of LL-theories, or explain their significance?

As Lepore and Ludwig insist, what puts us in a position to understand L is not knowledge of any facts entailed by an interpretive truth theory for L, but rather, the sort of knowledge they describe, knowledge *about* an interpretive truth theory for L. But why should an account of what knowledge would put us in a position to interpret utterances in a given language center around the construction of a truth theory? Why should it not revolve around the construction of a compositional translation theory, in which each sentence of L is paired with a suitable translation in (what we know to be) our own language? An analogous question arises for Williams’s portrayal of a meaning theory as a recursive device for generating interpretations. If our goal is to be in a position to interpret any arbitrary utterance in a language, why should we do so on the basis of a recursive theory rather than of a translation theory? Depending on our purposes, either kind of theory could be construed as a recursive device for generating interpretations.

3.6 Truth theories versus translation theories

It is often stressed that knowing what a sentence or utterance means goes beyond knowing that it is equivalent in meaning to some other sentence or utterance—or, knowing how to translate it into a sentence of another language. Knowledge of what is stated by an acceptable translation manual from a language L onto a language L’ does not, by itself, put one in a position to understand either language. However, when L’ is our own language, knowledge of what is stated by an acceptable translation theory from L into L’, together with knowledge that L’ is our own language, would be sufficient for understanding any arbitrary utterance in L (or at least, for understanding those

²⁴ A T-theorem is interpretive if it satisfies Convention T, or a suitable analog of Convention T for context sensitive languages.

utterances in L that can be deemed sufficiently alike in meaning to potential utterances in our own language).

Thus, with respect to their ability to state the content of knowledge that would put us in a position to interpret a language L, an interpretive truth theory for L and a translation theory (from L into our own language) seem to be on a par. Neither sort of theory states facts knowledge of which would suffice for interpreting utterances in L. But knowledge of certain facts about an interpretive truth theory would suffice for interpreting L. So would knowledge of what is stated by a translation theory from L into L', along with knowledge that L' is our own language.

If the reference to one's own language is made explicit, a translation theory actually seems to be at an advantage here.²⁵ For, knowing that 'Snow is white' in my language translates 'La neige est blanche' in French puts me in a position to interpret utterances of 'La neige est blanche' in French—as meaning that snow is white. This is because I know what 'Snow is white' means in my language. By contrast, knowing that 'La neige est blanche' is true in French iff snow is white does not itself put me in a position to interpret utterances of 'La neige est blanche' in French. It's my knowledge of the fact that the sentence of my language that expresses this knowledge—the T-sentence '“La neige est blanche” is true in French iff snow is white'—uses, on the right hand side of its biconditional, a sentence that translates the French sentence named on the left hand side, that puts me in a position to interpret utterances of this sentence. But knowledge of this fact is knowledge that 'Snow is white' in my language translates 'La neige est blanche' in French, and this is a fact that a translation theory can explicitly entail.

Thus, if nothing more is said about the work done by a truth theory as—or in the context of—a meaning theory, we would have to agree with Soames's sceptical remarks:

The only role played by knowledge of that which is stated by a translational truth theory (with canonical theorems) is that of allowing the agent to identify a unique canonical claim in which S is paired with a certain content, which is recognized ... to be the content expressed by a translation of S, and hence by S itself. *Neither the truth of this canonical claim, nor the fact that it states the truth-conditions of S, plays any role in deriving the interpretation of S.* All it does is supply a translation, which could be supplied just as well in other ways—with or without the notion of truth. For example, one can get the same interpretive results by replacing the truth predicate in a translational truth theory ... with *any arbitrary predicate F whatsoever*. To be sure, the resulting F-theory might not be true. But that makes

²⁵ As I already explained in fn. 17, Davidson thinks that if we make the reference to our own language explicit, we no longer have a theory “in the formal sense.” But why should we insist on having such a thing, if our goal is simply that of spelling out what we could know that would suffice for interpretation?

no difference to its usefulness in coming to understand sentences of the language.²⁶

Lepore and Ludwig do in fact agree with Soames's claim that a truth theory need not be true, nor does its T-predicate have to be interpreted a truth predicate, for the theory to do its work in enabling us to understand sentences of a given language.²⁷ But despite this, they maintain that an interpretive truth theory does more than indirectly supply translations. What more does it do? More importantly, since a translation theory directly supplies translations, why bother with the indirectness involved in using truth theories?

Here is Lepore and Ludwig's answer:

Proofs of the canonical theorems *exhibit* how parts of sentences, in virtue of their meanings, contribute to fixing the truth-conditions of these sentences, by way of using terms the same in meaning. We see exhibited in the proof the semantic structure of the sentence and how it fixes truth-conditions. This is not what the proof says, but it can be culled from the proof. Someone in possession of such a theory and appropriate knowledge of it is in a position to understand the compositional structure of the language. That is more than being able to pair object language sentences with meta-language ones that translate them.²⁸

As Lepore and Ludwig see it, the point of a meaning theory for a language L is not just to characterize knowledge that would put us in a position to interpret utterances in L, but also, to shed light on the compositional structure of L. Knowledge of the facts entailed by an LL-theory for L not only enables us to interpret utterances in L, but also puts us in a position to understand how facts about the conditions under which different sentences of L are true depend on assumptions governing finitely many expressions and their modes of combination. By contrast, knowledge of the facts entailed by a translation theory that maps sentences of L onto sentences of L', together with knowledge that L' is our own language, does not give us an understanding of the compositional structure of L. This, as Lepore and Ludwig see it, is the main reason for

²⁶ Soames 2008, p. 11.

²⁷ See Lepore and Ludwig 2010, p. 273:

In response to Soames's claim that all the theory does is match object language sentences with meta-language sentences that translate them, we agree that it does that, but that is not all it does. Soames says that the theory need not be true to issue in theorems that match object language sentences with meta-language sentences that translate them. This too is correct, as we noted above. Any predicate could play the role.

²⁸ Lepore and Ludwig 2005, p. 273.

using a truth theory rather a translation theory as the “recursive device” around which to frame a meaning theory for L.

Thus, unlike a translation theory, a truth theory for L (or perhaps, the LL-theory framed around it) seems to be playing a dual role. First, it generates interpretations of utterances of L by indirectly supplying translations of sentences in L into sentences of the metalanguage. Secondly, it sheds light on the compositional structure of L by entailing statements of the conditions under which each sentence of L is true on the basis of assumptions governing its structure and the contributions of finitely many meaningful parts.

I have three points to make in response to this account of the main advantage of an LL-theory over a translation theory. My first point is that if we accept this account, it does, after all, seem to matter how we interpret ‘is true’ as it figures in our truth theory. This may not matter as far as our first goal is concerned, of generating interpretations of a speaker’s utterances. But unless we take ‘is T’ to express some notion of truth, the very same notion that Lepore and Ludwig are employing in speaking of the light that is shed on how the truth-conditions of whole sentences depend on the semantic features of their parts, it is not clear how the truth theory can shed any light on this. Moreover, the T-theorems have to express facts, or *true* claims, about the truth-conditions of sentences of L, in order for their derivability from the truth theory’s axioms to shed light on how the truth-conditions of sentences of L depend on their composition out of finitely many parts. Thus, Ludwig’s claim that a truth theory need not be true in order to play its assigned role in an LL-theory cannot be accepted.

Secondly, it is not clear how the two tasks Lepore and Ludwig assign a meaning theory are connected. They insist that a meaning theory for L should be construed as an account of what knowledge would suffice for interpreting utterances in L, rather than an account of what knowledge helps constitute a speaker’s competence in L.²⁹ Yet they also insist that an adequate meaning theory for L ought to give us an understanding of the compositional structure of L. What they do not explain is how, if at all, insight into the compositional structure of L serves our ability to interpret utterances in L. At some point, Davidson himself may have seemed to be offering an account of the connection between these two tasks, by proposing that knowledge of the compositional structure of the language (in the sense we have just specified) itself puts us in a position to interpret any arbitrary utterance in the language. Consider the following passage:

[W]e can interpret a particular sentence provided we know a correct theory of truth that deals with the language of the sentence. For then we know not only the T-sentence for the sentence to be interpreted, but we also ‘know’ the T-sentences for all other sentences; and, of course, all the proofs. Then we would

²⁹ See Lepore and Ludwig 2010, pp. 270-271.

see the place of the sentence in the language as a whole, we would know the role of each significant part of the sentence, and we would know about the logical connections between this sentence and others.³⁰

Clearly, however, seeing “the place of [a] sentence in a language” in the sense of knowing a true truth theory for the language of this sentence is neither necessary nor sufficient for being in a position to interpret utterances of this sentence. Not necessary, since I can know (for instance, by being told) that the sentence ‘La neige est blanche’, as uttered by Olivier on any given occasion, literally means that snow is white, without knowing anything about the rest of his language—and arguably, without even knowing what the semantically significant parts of this sentence are. Not sufficient, as Davidson himself acknowledges, but not just because a truth theory may be true without being translational, but because I may know a translational truth theory without knowing that it is translational.

This brings me to my third, and perhaps most important point. Since interpretation can get by without our being in a position to explain how the truth-conditions of the speaker’s sentences depend on their semantic structure, why might we care to explain this? There may of course be various goals involved that have little to do with illuminating the capacities involved in understanding a language.³¹ But as far as our goal has anything to do with shedding light on understanding, it is the speaker’s understanding of her own language, rather than our ability to interpret her utterances, that is the plausible subject of illumination here. It is only insofar as we are interested in capturing the structure of the speaker’s knowledge of her own language, rather than of our ability to interpret her utterances, that truth theories have any role to play over and beyond that of finitely generating interpretations of the speaker’s utterances. If this is right, the main point of a meaning theory for a given speaker’s language is not to explain how we are to understand the speaker’s sentences on the basis of understanding their parts³², but to explain how the speaker understands each of her sentences on the basis of her understanding of their parts. Thus, despite Lepore and Ludwig’s remarks to the contrary, Davidson is after all, explaining a speaker’s knowledge of her own language as partly consisting in her knowledge of the truth-conditions of her sentences, and it is only insofar as he is doing this that he is assigning a truth theory any role beyond that of a “recursive device for generating interpretations.”

³⁰ “Radical Interpretation,” Davidson 1984, pp. 138-139.

³¹ For instance, we have seen that Tarski’s goals in seeking recursive characterizations of truth for particular languages had little to do with shedding light on linguistic competence.

³² Lepore and Ludwig 2010, p. 265.

3.7 Saying and showing

To help us see how a truth theory could be “exploited to do the work of a meaning theory” even though it does not employ the concept of meaning, Lepore and Ludwig remark that “what the theory shows it does not say.”³³ Similarly, John McDowell has suggested that a Davidsonian truth theory is meant to *serve* as a theory of sense rather than *being* one³⁴, and that in the context of such a theory, an axiom that “does no more than state—in a suitable way—what the reference of an expression is may nevertheless give – or as good as give – that expression’s sense.”³⁵ These remarks are two instances of a familiar appeal to a distinction between “saying” and “showing” in an effort to make sense of the role of a truth theory in a meaning theory.

Despite this similarity, however, there is an important difference between McDowell’s and Lepore and Ludwig’s accounts of the way in which a truth theory is supposed to serve the needs of a meaning theory. As we have seen, an LL-theory entails M-sentences and thus can be said to explicitly *state* what sentences of the language mean, rather than merely *showing* it. This is meant to be an improvement over the sort of account McDowell gives, whereby no explicit statement of what sentences of the language mean is offered.³⁶

I would like to argue that, far from constituting an improvement over McDowell’s, Lepore and Ludwig’s account takes us further away from any plausible account of the special sense in which a truth theory can, while a translation theory cannot, be used to show what sentences of a language mean.

³³ Ibid., fn. 13 on p. 264-265.

³⁴ See McDowell 1998, p. 173:

[S]erving as a theory of sense is not the same as being one, on a certain strict view of what it is to be one. It was clear anyway that a truth-theory of the sort Davidson envisages does not, in saying what it does, state the sense of expressions. Why should we hanker after a theory that does that mysterious thing, if a theory that does some utterly unmysterious thing instead can be made to serve the purpose?

³⁵ McDowell 1998, p. 175.

³⁶ A different kind of attempt to exploit Davidsonian truth theories in the context of theories entailing M-sentences can be found in Kölbel 2001. Unlike an LL-theory, a Kölbel-style meaning theory for L itself includes a Davidsonian truth theory—in the sense of entailing everything that a truth theory entails—but in addition also entails an M-theorem for each sentence of L. To achieve this result, Kölbel proposes that we simply add to our logic an inference rule that permits us to derive an M-sentence from each canonically derived T-sentence. This proposal strikes me as even less plausible than Lepore and Ludwig’s, since Kölbel’s inference rule is simply not valid: it can be used to derive false M-sentences from axioms that are true but not interpretive.

But first, let us clarify the distinction between saying and showing. Richard Heck correctly points out that a sentence or theory, in and of itself, does not show anything, at least not in any sense of ‘showing’ that could be involved in the idea that the meaning of a sentence can be shown simply by a statement of its truth-conditions. Rather, it is the use we make of a truth theory that can show what we take sentences of a given language to mean.³⁷ Consider the following two true T-sentences:

- (a) ‘La neige est blanche’ is true in French if and only if snow is white.
- (b) ‘La neige est blanche’ is true in French if and only if grass is green.

In choosing (a) rather than (b) to state a condition under which the sentence ‘La neige est blanche’ is true, we can show the meaning we take this sentence to have, even if (a) itself does not state what the sentence means (or how we are to interpret its utterances). More generally, our choosing one true truth theory rather than another one can, in certain circumstances, show how we are prepared to interpret utterances in the language, even if the theory itself does not entail any claims as to what these utterances mean (or would mean).

This does seem plausible enough. But if it makes sense to say it of truth theories, could we not say the same thing of translation theories? A translation theory from a speaker’s language L into a language L’ says (or, entails theorems that say) how each sentence of L is to be translated into a sentence of L’. In appropriate circumstances, I can use such a theory to show, without saying it, what each sentence of L means. Of course, the theory can be used show this only to those who know what sentences in L’ mean. But similarly, a truth theory for L can be used to show what each sentence of L means only to those who know that its T-theorems are interpretive. What is the relevant difference between the two? What have we missed about the sense in which a truth theory can, while a translation theory cannot, be used to *show*, without saying it, what each sentence of a language means?

I think the difference can be explained only if we step outside of Davidson’s “Radical Interpretation.” For, as far as its usefulness in generating interpretations of a speaker’s utterances, a truth theory is no better than a translation theory from the speaker’s sentences into the interpreter’s own. And either kind of theory can help explain how we can come to understand any arbitrary utterance in a language on the basis of appropriate knowledge about finitely many expressions and ways of combining them. However, if we turn our attention to the task of describing a speaker’s understanding of his own language, an important contrast between truth theories and a translation theories clearly emerges: translational T-theorems for the language of a given speaker state something the speaker knows about each of his sentences in understanding them the way he does. A translation theory from the speaker’s language onto another language does not state

³⁷ See Heck 2007.

anything the speaker knows, or needs to know, in understanding his own language. For example, if the T-sentence ‘“La neige est blanche” is true in Olivier’s language if and only if snow is white’ is interpretive, what it says about the sentence ‘La neige est blanche’ is something Olivier knows about this sentence—something he knows about this sentence in using it or understanding it the way he does. By contrast, the fact that ‘Snow is white’ in my language translates ‘La neige est blanche’ in Olivier’s is not something Olivier needs to know to understand this sentence in the way he does.

How, then, can this help explain the sense in which a truth theory can be used to show, without saying it, what each sentence means in a given speaker’s language? The truth theory can be used to show what each sentence of the language means insofar as it entails theorems that state what the speaker knows about each of his sentences in understanding his language—what he knows about each of his sentences on the basis of his understanding of its syntactic parts. It bears emphasizing that these theorems *state*, rather than merely *showing*, what the speaker knows about each of his sentences in knowing what it means. But what no truth theory can itself entail is the claim that its T-theorems do have this property. This in the sense in which our using one true theory rather than another can show what we take sentences of the language to mean, even though neither theory entails any statements of what these sentences mean.³⁸

Thus, if we can use a truth theory to “show” what sentences in L mean, it is not, as Lepore and Ludwig claim, because knowledge of certain facts about the theory (the facts spelled out by an LL-theory) would suffice for interpreting utterances in L. Rather, it is because the truth theory states what a speaker of L knows about each of his sentences in understanding it the way he does.

3.8 Textual evidence: speaker and interpreter

As we have seen, in the early papers where Davidson’s proposal to use truth theories as meaning theories first emerges, it precedes any direct preoccupation with radical interpretation. I think that Davidson has not really changed his mind about the point of a meaning theory for a language, even when he goes on to characterize it primarily only

³⁸ In “On the Sense and Reference of a Proper Name” (McDowell 1998), McDowell claims that knowledge that ‘Hesperus’ in L stands for Hesperus is, in the context of further knowledge about other expressions of L (along with, for McDowell, knowledge of a theory of force), sufficient for being in a position to interpret utterances in L containing this name. This is not quite right, and we are now in a position to say why. If all I know about the role of ‘Hesperus’ in L (where L is the language of another speaker) is that ‘Hesperus’ in L refers to Hesperus, I do not thereby understand the meaning of ‘Hesperus’ in L, nor am I in a position to interpret utterances of sentences containing the name. This is because I can know that ‘Hesperus’ in L refers to Hesperus without knowing that this is something the speaker of L also knows, or that he draws on this knowledge in speaking the way he does.

in connection with a problem about the possibility of radical interpretation. For, despite his more explicit preoccupation with the question of what knowledge would suffice for interpretation, there is sufficient evidence for thinking that it is a speaker's knowledge of his own language (in the sense of idiolect) that Davidson all along seeks to capture by means of a compositional truth theory.

“Radical Interpretation” contains a brief discussion of the advantages of a truth theory over a translation theory, as far as each theory's ability to specify the content of knowledge that would suffice for interpretation is concerned. Davidson agrees that an interpreter can use a translation theory (a mapping of the speaker's sentences onto his own) to interpret another speaker's utterances, but argues that in doing this, the interpreter “brings to bear two things he knows and that the theory does not state: the fact that the subject language is his own, and his knowledge of how to interpret utterances in his own language.”³⁹

What I am suggesting is that this explanation as to why a truth theory would constitute a better solution to the “problem of interpretation” than a translation theory, only goes to show that what Davidson is really after, in asking what form should be taken by a meaning theory, does not primarily concern interpretation. Rather, Davidson's main concern is to give a theoretical description of a speaker's linguistic competence—to spell out, in structure-revealing terms, what a competent speaker knows about each of his sentences in understanding it the way he does. Davidson says that a translation theory leaves out “what we need to know that allows us to interpret our own language.”⁴⁰ But really, when it comes to our own language (that is, the language we actually now speak), our speaking it, or understanding it, does not involve “interpreting” our own utterances in it, in Davidson's sense of the word. Understanding why would help us shed further light on why a meaning theory should be construed as an account of a speaker's understanding of his own language, rather than of what knowledge would suffice for interpretation.

Though Davidson does, at times, appear to use “interpretation” and “understanding” interchangeably, he does, on other occasions, clearly suggest that an individual's understanding of his own language should not be described as a matter of interpretation. Consider the following passage from “Indeterminism and Antirealism”:

First person interpretations are necessarily tied to the homophonic translation manual (which is to say, translation, or interpretation, has no place here) ... It should not be concluded from the fact that a person is restricted to a unique way of interpreting himself (if this can be called interpretation: it would be

³⁹ “Radical Interpretation,” Davidson 1984, pp. 129-130.

⁴⁰ Ibid.

better to say that aside from pathological cases, our way of interpreting others has no application to ourselves) that therefore his words have unique reference.⁴¹

Why exactly does Davidson think that our way of interpreting others has no application to ourselves? We get some clarification in “First Person Authority,” where Davidson describes the presumption that a speaker knows what his own words mean as “essential to the nature of interpretation—the process by which we understand the utterances of a speaker.” Davidson continues:

This process cannot be the same for the utterer and for his hearers ... there can be no general guarantee that a hearer is correctly interpreting a speaker; however easily, automatically, unreflectively, and successfully a hearer understands a speaker, he is liable to serious error. In this special sense, he must always be regarded as interpreting a speaker. The speaker cannot, in the same way, interpret his own words.

The asymmetry rests on the fact that the interpreter must, while the speaker doesn't, rely on what, if it were made explicit, would be a difficult inference in interpreting the speaker.⁴²

Why, then, does Davidson think that a speaker cannot be said to interpret his own words “in the same way,” or in the same sense, as a hearer can be said to interpret the words of a speaker? This passage suggests that this has something to do with first person authority, or with the general presumption (essential to interpretation) that a speaker knows what he means by the words he uses. By contrast, Davidson wants to describe even the most mundane cases of one person's understanding of another utterance as being ultimately—even if not explicitly—“based on evidence and inference.”⁴³ Davidson thus seems to want to reserve “interpretation” to those cases that are based on evidence and inference, rather than to the understanding each speaker has of his own utterances.⁴⁴

⁴¹ Davidson 2001, p. 80.

⁴² Ibid., pp. 12-13.

⁴³ Ibid., p. 66.

⁴⁴ Does this mean that Davidson disagrees with Quine's remarks that radical translation begins at home, or with his own previous claim that “the problem of interpretation is domestic as well as foreign” (“Radical Interpretation”, Davidson 1984, p. 125)? No, I do not think that there is any conflict between these remarks and Davidson's denial that a speaker's understanding of her language involves interpretation. This is easy to see if we keep in mind the crucial distinction between an idiolect and a communal language. When Davidson claims that even ordinary communicative situations involve radical

Of course, the presumption that each speaker knows what his words mean is not indefeasible. The speaker might, on occasion, fail to be interpretable, and this is the sense in which he might fail to know what his words mean. However, what Davidson leaves no room for is the idea of a speaker's misunderstanding his own utterance, in the sense of its successfully meaning something while the speaker takes it to mean something else. In "What is Present to the Mind," he says:

I can do no better, in stating the truth-conditions for my utterance of the sentence 'The Koh-i-noor diamond is a crown jewel' than to say that it is true if and only if the Koh-i-noor diamond is a crown jewel. If I say this, I utter a tautology, but if you give the truth-conditions of my utterance using the same words, you are making an empirical claim, though probably a true one.⁴⁵

It is important to understand the sense in which a competent speaker's homophonic attributions of truth-conditions to her own sentences are supposed to be "tautologous." I do not think that Davidson means to be characterizing such utterances as logical or necessary truths. Rather, his claim is only that they not empirical, or, not based on "evidence and inference." When I say that 'Snow is white' as uttered by me now is true if and only if snow is white, the claim I am making is contingently true, but it is one that I know to be true simply by virtue of understanding it. Adding to this Davidson's assumption that when communication is successful, the speaker understands her own utterance (since the utterance means what she takes it to mean), we get the result that as long as a speaker is a successful communicator, and thus as long as she is linguistically competent, she knows the conditions under which her sentences are true. It is this basic knowledge that a meaning theory for a speaker's language should be describing. If this language contains infinitely many sentences, an illuminating theory will describe the speaker's knowledge of the conditions under which utterances of each of these sentences would be true as derived from a finite compositional basis.

Thus, instead of saying that a translation theory "leaves tacit and beyond the reach of theory what we need to know that allows us to interpret our own language,"⁴⁶ Davidson should have said that a translation theory leaves beyond the reach of theory what we know about the truth-conditions of our own sentences as we currently understand them (and use them to say how we are interpreting the utterances of others).

interpretation, he is claiming only that the account of radical interpretation can help shed light on the nature of successful communication between what we think of as speakers of "the same language." This does not mean that the correctness of a speaker's homophonic attributions of truth-conditions to her own sentences (as she currently understands them) can be explained by thinking of the speaker as being in the position of a radical interpreter with respect to her own utterances.

⁴⁵ Davidson 2001, p. 66.

⁴⁶ Davidson 1984, p. 130.

I have been insisting that a meaning theory for a given speaker's language should be construed primarily as an account of what the speaker knows in understanding her own sentences, rather than an account of what knowledge would suffice for interpreting the speaker's utterances. One might wonder, however, whether this contrast between the knowledge involved in speaking and interpreting a language really does come to anything. For, when communication is successful, both speaker and interpreter can be said to know the conditions under which the speaker's utterances are true. What, then, is the point of my insistence that a meaning theory ought to describe a speaker's, rather than an interpreter's, knowledge of a language? The following passage would indeed appear to suggest that Davidson himself directly opposes any substantive contrast between the two:

Because a speaker necessarily intends first meaning to be grasped by his audience, and it is grasped if communication succeeds, we lose nothing in the investigation of first meaning if we concentrate on the knowledge or ability a hearer must have if he is to interpret a speaker. What the speaker knows must correspond to something the interpreter knows if the speaker is to be understood, since if the speaker is understood he has been interpreted as he intends to be interpreted. The abilities of the speaker that go beyond what is required of an interpreter—invention and motor control—do not concern me here.⁴⁷

Davidson is here claiming that apart from matters of “invention and motor control,” there is no difference between the abilities involved in a speaker's understanding of his own language, and the abilities required of his interpreter. Does this not directly undermine my contrast between a speaker's and an interpreter's knowledge of a language? I do not believe it does, for in contrasting a speaker's knowledge of her own language with the knowledge on which an interpreter's understanding of the speaker might rest, I am not claiming that what a speaker knows about each of her sentences goes beyond what a successful interpreter knows about these sentences. I am, however, claiming that an account of what an interpreter could know that would enable him to understand a particular speaker need not, by itself, shed any light on what both speaker and interpreter know about each of the speaker's sentences when communication is successful.

What an interpreter could know that would put him in a position to understand a speaker are the facts entailed by an acceptable translation theory from the speaker's language onto his own. If and when the interpreter knows these facts, what he knows, in knowing them, does not correspond to anything the speaker needs to know in order to be understood. However, when the speaker is understood, she and her interpreter share knowledge of the conditions under which the speaker's utterances are true—even

⁴⁷ “A Nice Derangement of Epitaphs” in Davidson 2005b, pp. 92-93.

though the interpreter's knowledge of these truth-conditions might be based on something else he knows about the speaker's sentences that the speaker herself does not know.

3.9 Conclusion

Davidson clearly intended his account of radical interpretation to illuminate the notion of meaning, along with those of truth, rationality, and the propositional attitudes. He did not, however, seek to ground each and every one of his proposals about these notions in considerations about the ends and means of interpretation. A case in point is his proposal that truth theories can serve as meaning theories. As I argued in this chapter, and contrary to widespread assumptions, the point of a meaning theory, and the sense in which a truth theory can serve as one, cannot be explained in terms of its role in interpretation.

The prevailing assumption that it is to be so explained has had the unfortunate consequence of obfuscating a clear view of the role of truth in Davidson's approach to meaning. For, as we have seen, this assumption naturally leads to an instrumentalist view of the role of truth in truth theories serving as meaning theories. Thus, the very conception of the point of a meaning theory that underlies instrumentalism is one that undermines the contrast on which Davidson insists between meaning theories and translation manuals.

Chapter 4

Can a Deflationist Make Sense of Radical Interpretation?

4.1 Introduction

A central question of my dissertation is whether Davidson's approach to meaning is compatible with a deflationary account of truth. There seems to be increasing agreement among commentators that this question is to be answered by examining the role of truth in Davidson's account of the ends and means of interpretation.¹ This is not surprising, given the prominence of the notion of interpretation in Davidson's writings, and the widespread assumption that Davidson's account of the nature of meaning simply is his account of the ends and means of interpretation.² In particular, since Davidson's theory of meaning is supposed to include his account of the form to be taken by a meaning theory, it is assumed that the point of a Davidsonian meaning theory is itself to be explained in terms of its role in interpretation.

¹ See Kölbel 2001, Williams 1999, and Lepore and Ludwig 2003 (p. 54). For instance, Kölbel argues that

the method by which a Davidsonian theory of meaning is to be tested empirically should be the ultimate touchstone for the truth doctrine [or, the assumption that truth plays a crucial explanatory role in Davidson's account of the nature of meaning]: if the methodology of radical interpretation requires explanatory use of the notion of truth, the truth doctrine is justified. If it does not, then the truth doctrine can be thrown onto the scrapheap of unjustified dogmas. (Kölbel 2001, p. 628)

² Williams makes this assumption explicit in the following passage:

If we (a) determine the theoretical form of the knowledge at which interpretation aims and (b) fix the methodology of interpretation, we have said all there is to say about meaning. For Davidson, meaning just *is* whatever the practice of interpretation reveals. (Williams 1999, p. 553)

As I explained in Chapter 3, this assumption undermines the contrast between meaning theories and translation manuals³, and consequently obscures the crucial role of a substantive notion of truth in truth theories serving as meaning theories. We simply cannot explain the point of a Davidsonian truth theory, and the role played by its truth locution, in terms of a prior conception of the ends and means of interpretation.⁴

In this chapter I want to complete the work, started in Chapter 3, of showing that the inability to perceive the deep incompatibility of Davidson's approach to meaning with a deflationary account of truth stems from the mistaken assumption that for Davidson, "meaning just *is* whatever the practice of interpretation reveals."⁵ The question I want to ask is: does the methodology of radical interpretation give us any independent reasons to reject deflationism? In other words, without presupposing an independent account of the point of truth theories serving as meaning theories, I want to examine the prospects of being able to justify Davidson's opposition to deflationism simply in terms of his proposed methodology for radical interpretation.

In what follows, after introducing some terminology, I will examine some passages in which Davidson appears to be explaining the crucial role of a substantive notion of truth in interpretation. I will argue that his remarks do not really help explain why the notion of interpretation might be unavailable to a deflationist. I will then spell out (in section 3) a conception of the interpreter's task that is available to a deflationist. This conception makes it possible to ask, without begging the question, whether the deflationist can make sense of the methodology of radical interpretation. In sections 4 and 5, I will briefly explain why, despite initial appearances, two crucial aspects of the radical interpreter's methodology are available to a deflationist: first, the need to read a logic into a speaker's language, and secondly, the need to maximize truth or agreement in our attribution of beliefs to the speaker. Sections 6 and 7 will then focus on an apparently more promising attempt to explain why the methodology of radical interpretation might be unavailable to a deflationist: namely, the argument that

³ Since knowledge of either kind of theory could, in the context of further knowledge about it, put us in a position to interpret any arbitrary utterance in the speaker's language.

⁴ In particular, Davidson's proposal to use truth theories as meaning theories cannot be described in the way Williams does, as an account of "the theoretical form of the knowledge at which interpretation aims" (Williams, p. 553). We could, of course, simply define the "knowledge at which interpretation aims" as knowledge of an acceptable truth theory. But this does not involve explaining the point such a theory in terms of the notion of interpretation. Quite the contrary: it involves explaining the notion of interpretation—or, one notion of interpretation—in terms of an independent conception of the point of a truth theory. If, on the other hand, we characterize our goal as interpreters as that of interpreting any possible utterance in a speaker's language on a particular occasion (in the sense of determining its literal meaning), then there is no reason to think that the "knowledge at which interpretation aims" should be knowledge of a Tarskian truth theory, rather than that of suitably compositional translation theory.

⁵ Williams 1999, p. 553.

Davidson's distal approach to the interpretation of observation sentences makes explanatory demands on truth that a deflationary approach could not satisfy. I will argue that on closer examination, the advantages of Davidson's distal approach over Quine's proximal approach turn out not to depend on any commitment to a substantive notion of truth. If this is right, the methodology of radical interpretation cannot, by itself, explain the incompatibility of Davidson's approach to meaning with a deflationary account of truth.

4.2 The need for a transcendent notion of truth

Let us begin by reviewing, and in some cases introducing, some terminology. Following Field⁶, I would like to characterize the disagreement between deflationists and inflationists about truth in terms of the question whether or not our primary notion of truth is an immanent notion. In addition to helping distinguish my usage of the word 'deflationism' from some others in the literature⁷, this will be useful in locating some of Davidson's own worries about deflationism.

An immanent approach to truth, as I am using the term, is an approach that takes our primary notion of truth for sentences to be an immanent one. An immanent notion (of truth, reference, or meaning), in turn, is a notion that I can only meaningfully apply to expressions that I understand a certain way.⁸ For instance, Field's pure disquotational notion of truth is an immanent notion, since it does not make sense for me to wonder whether or not a given sentence in a language that I do not understand is true. As Field

⁶ Field 1992, 2003.

⁷ For example, Soames 2003 argues that "our ordinary truth predicate of propositions is deflationary," but "deflationary theories of truth for sentences don't tell the whole story about sentential truth" (*ibid.*, p. 370). Soames' approach to truth does not count as deflationary in my sense of the word, since it combines "deflationism" about propositional truth with a transcendent approach to sentential truth. Similarly for McGrath's "weak deflationism" in McGrath 1997.

⁸ Quine drew a distinction between "immanent" and "transcendent" linguistic notions in Quine 1970. He characterized a notion as "immanent when defined for a particular language; transcendent when directed to languages generally" (Quine 1970, p. 19). An "immanent" notion (of truth, reference, or meaning), as I am using the term, is a notion that a speaker can only meaningfully apply to expressions that she understands a certain way. Thus, my notion of immanence is not exactly the one Quine introduced. In particular, a notion could be immanent in Quine's sense without being immanent in my sense. For, a notion could be both immanent in Quine's sense and meaningfully applied by a speaker to expressions she does not understand, as long as these expressions belong to a language for which the notion is defined.

puts it, I understand ‘“s” is true’ only to the extent to which I understand ‘s.’ A transcendent notion of truth, by contrast, is a notion that is applicable to sentences across different speakers or languages. If ‘is true’ as I use it is a transcendent notion of truth, it does make sense for me to wonder whether or not a given sentence in a language that I do not understand is true: I may understand ‘“s” is true’ even if I do not understand ‘s.’⁹

An inflationist—that is, a non-deflationist—about truth, in the sense in which I am using these words, is someone who denies that our primary notion of truth is an immanent notion. Though inflationists may disagree on whether or how our concept of truth should be analyzed, and whether or how the nature of truth could be explained, what they share is a commitment to what we may call a ‘transcendent’ approach to truth: that is, to the idea that our primary notion of truth is a transcendent notion, not an immanent one. Or, we may say that what inflationists agree on is the claim that our primary notions of truth are transcendent notions—inflationists in my sense need not be committed to the existence of a single basic transcendent notion of truth.

Finally, I will be using the expression ‘substantive notion of truth,’ or ‘notion of substantive truth,’ as shorthand for ‘transcendent notion of truth that cannot be defined in terms of an immanent notion of truth.’ A deflationist can thus be characterized as someone who denies the intelligibility of a substantive notion of truth. For, even though the deflationist need not deny the intelligibility of transcendent notions of truth, he holds that any such notion is to be analyzed in terms of a more fundamental immanent notion of truth.¹⁰

With these definitions in place, let us examine some of Davidson’s apparent arguments for the crucial role of a substantive notion of truth in interpretation. Consider the following remarks in Davidson’s “Pursuit of the Concept of Truth”:

[O]nce the question is raised whether [a Tarskian truth theory], as stated in an interpreter’s language, is true of the language of a second person, the empirical and non-trivial character of the theory becomes obvious. What should count as

⁹ Disquotationalism is an immanent approach to truth since its primary notion of truth, pure disquotation, is an immanent notion.

¹⁰ My distinction between immanent and transcendent notions of truth seems to be in line with the one Davidson fleshes out in the following passage:

This is, I think, what Quine means when he says that truth is “immanent.” The point is not merely that the truth of a sentence is relative to a language; it is that there is no transcendent, single concept to be relativized. (Davidson 2005b, p. 67)

confirming that such a theory is true? To ask this question is to ask a question about the concept of truth that disquotation cannot answer.¹¹

Davidson seems to be arguing that insofar as we can use a Tarskian truth theory as an empirical theory of another speaker's language, there must be more to the concept of truth (that is, to the concept expressed by the truth locution employed in a Tarskian truth theory) than disquotation. In other words, the concept of truth employed in a theory whose aim it is to describe another speaker's language cannot be a concept whose content is exhausted by the disquotation schema. It is hard to disagree with this, since if I can meaningfully apply the pure disquotational truth predicate to a sentence 's,' then 's' and "'s' is true' are cognitively equivalent for me, and the question 'Is "'s'" true if and only if s?' cannot be an empirical one for me. But the fact that the concept of truth employed in a truth theory for another speaker's language cannot be that of pure disquotational truth does not, by itself, mean that using a truth theory as an empirical theory of another speaker's language incompatible with disquotationalism. For, as we have already noted, the disquotationalist need not deny the intelligibility of talk about truth in application to the sentences of others, or to sentences we do not understand. He just holds that any such notion of truth would have to be definable in terms of the more basic notion of pure disquotational truth. More generally, proponents of an immanent approach to truth need not deny the intelligibility of transcendent notions of truth—they just hold that any such notion would have to be characterized in terms of a more basic notion of immanent truth.¹²

¹¹ Davidson 2005b, p. 70. Here is a similar passage from Davidson 2005a:

Since all of us do understand some speakers of some languages, all of us must have adequate evidence for attributing truth-conditions to the utterances of some speakers; all of us have, therefore, a competent grasp of the concept of truth as applied to the speech behavior of others. (Davidson 2005a, p. 37)

¹² For instance, a disquotationalist can define 'is true' for the sentences of another speaker in terms of pure disquotational truth and some notion of translation or sameness of meaning. Of course, if we recognize the possibility of equally acceptable translation schemes that may result in different truth values being assigned to the same sentence (as uttered by the speaker on a given occasion), the T-sentences entailed by any given truth theory cannot strictly speaking be construed as empirical claims about the speaker's sentences. But there would still be an empirical question we can ask, about our truth theory and the translation scheme relative to which its truth predicate has been defined: namely, is this an acceptable translation scheme? It should be noted that some of Davidson's own responses to worries about indeterminacy seem to invite us to so construe the empirical import of truth theories:

Ian Hacking once put this puzzle to me: how can two theories of truth both be acceptable if one theory makes a certain utterance true and the other does not? Isn't this a contradiction? It is not a contradiction if the theories are relativized to a language, as all theories of truth are ... we admit that it is not entirely an empirical question what language a person speaks; the evidence allows us some choice in languages, even to the point of allowing us to assign conflicting truth-

However, Davidson's opposition to deflationism is more radical than the above-quoted remarks, and my response to them, suggest. His point is not simply to draw attention to the fact that we need a notion of truth applicable to sentences we do not already understand.¹³ What he rejects is the very idea that, as I have characterized it, is of the essence of deflationism: namely, that our *primary* notion of truth is an immanent notion. This comes out more clearly in the following passage:

I suggest that we omit the final step in Tarski's definitions, the step that turns his axiomatizations into explicit definitions. We can then in good conscience call the emasculated definition a theory, and accept the truth predicate as undefined. This undefined predicate expresses the *general*, intuitive, concept, applicable to any language, the concept against which we have always surreptitiously tested Tarski's definitions (as he invited us to do, of course) ... The empirical question is how to determine, by observation and induction, what the truth-conditions of empirical truth vehicles are. It bears emphasizing: absent this empirical connection, the concept of truth has no application to, or interest for, our mundane concerns, nor, so far as I can see, does it have any content at all.¹⁴

It is clear, then, that Davidson is not merely insisting that we need a notion of truth applicable to the sentences of other speakers (or to sentences that we do not already understand). Rather, his main point is that interpretation depends on our grasp of a single, primitive notion of truth, applicable to our own sentences as well as to the sentences of others. In other words, interpretation, as Davidson sees it, crucially depends on our grasp of a substantive notion of truth. However, what Davidson does not explain in these passages is why this should be so--that is, why interpretation should be taken to depend on our grasp of a substantive notion of truth. We can thus see why his claim that we all have "a competent grasp of the concept of truth as applied to the speech behavior of others"¹⁵ would be met with the following response:

But what does this grasp consist in? If we know how to interpret, we can extend our purely disquotational notion of truth to alien sentences, which would appear to give us just what Davidson demands. So the question remains: Why not give truth to the deflationists and let interpretation stand on its own feet? The only

conditions to the same sentence ... An empirical question remains, to be sure: is this language one that the evidence allows us to attribute to this speaker? (Davidson 1984, pp. 239-240)

¹³ As I see it, this is something that any plausible form of deflationism would have to accommodate.

¹⁴ "The Folly of Trying to Define Truth," in Davidson 2005b, pp. 35-36.

¹⁵ Davidson 2005a, p. 37.

objection to so doing would have to be that the canons of interpretation themselves make use of a rich notion of truth. I do not believe that they do.¹⁶

Again, what it does not make sense to disagree about is the need for a notion of truth applicable to the sentences of others. Without it, question of the correctness of a truth theory for the language of another speaker does not make sense. Williams is not disagreeing with this, but with the further claim that this notion has to be the sort of primitive, transcendent notion of truth on which Davidson insists.¹⁷ As we have seen in Chapter 3, Williams does not think that deflationism prevents us from using truth theories as meaning theories. This is why for him, the question of the compatibility of deflationism with Davidson's approach to meaning reduces to the question, "Do the canons of interpretation themselves make use of a rich notion of truth?" It is this question, understood the way Williams understands it, that is the topic of this chapter.

4.3 Radical interpretation from a deflationary perspective

What is the problem to which Davidson's account of radical interpretation is supposed to provide an answer? The problem that Davidson addresses in "Radical Interpretation," and on which I focused in Chapter 3, is that of spelling out the content of knowledge that could put us in a position to fix the literal meaning of any arbitrary utterance in a speaker's language. What we are after is a compositional theory knowledge of which could put *us* in a position to say what each of the speaker's sentences means, or would mean, on particular occasions of utterance.¹⁸ However, there is another, and in my view more fundamental notion of interpretation at work in Davidson's writings. One goal of interpretation, in this second sense, is to confirm or construct an acceptable truth theory for another speaker's language—where to be

¹⁶ Williams 1999, p. 559.

¹⁷ Some of Davidson's own remarks may be taken to suggest that he would indeed be willing to make room for an immanent approach to truth, as long as it accommodates talk of truth in application to the sentences of others. Consider:

I have been stressing the invasion of truth by considerations of meaning, an invasion we cannot ignore as soon as the question of the truth of sentences in languages other than our own arises. The invasion can be direct, if we ask when an alien sentence is true, or indirect if we characterize truth disquotationally for our own language first, and then translate the foreign tongue into our own. The difference, if there is one, is that the first strategy mingles issues of truth and meaning from the start, while the second approach allows for a division of labor. (Davidson 2005b, pp. 74-75)

¹⁸ Note that a theory knowledge of which would put us in a position to determine the literal meaning of the speaker's utterances need not constitute an account of the speaker's knowledge of her own language.

acceptable, a truth theory has to capture the content of the speaker's knowledge of her language, and thus, to generate statements of what the speaker knows about the truth-conditions of each of her sentences in speaking the way she does.

If what I have argued in previous chapters is right, Davidson's primary notion of interpretation is the latter notion. The task facing Davidson's radical interpreter is that of confirming or constructing an acceptable compositional account of the speaker's knowledge of her own language—what she knows about the truth-conditions of each of her sentences in understanding her language. Davidson's conception of radical interpretation is thus simply unavailable to the deflationist, since, as I have argued in Chapter 2, deflationism does not allow us to describe competent speakers as knowing the truth-conditions of their sentences. But if this is right, what could be the point of asking, as I do, whether the methodology of radical interpretation is compatible with deflationism, since the very problem that this methodology is brought in to solve cannot be articulated from a deflationary perspective? Moreover, since Davidson's conception of the ultimate evidence for radical interpretation—namely, the conditions under which the speaker *holds* various sentences *to be true*—ascribes to the speaker grasp of a substantive notion of truth, the methodology of radical interpretation is clearly unavailable to the deflationist. What work is left for us to do here?

To answer this question, note that even if Davidson's own conception of the radical interpreter's task is incompatible with deflationism, we might recognize the possibility of a less demanding conception that is available to the deflationist. More precisely, there seems to be a way of spelling out the radical interpreter's task, and the evidence for radical interpretation, that does not prejudge the question of the compatibility of Davidson's approach with deflationism. My question can be put thus: Once we do spell out the problem of interpretation in suitably deflationary terms, is there any reason to think that the methodology of radical interpretation is itself unavailable to the deflationist?

Here is how the "problem of interpretation" can be spelled out from a deflationary perspective on truth. The problem is not that of determining what the speaker knows about the truth-conditions of her sentences in understanding them the way she does, but rather, that of arriving at an acceptable way of matching each sentence of the speaker's language with a sentence of mine. An interpretive truth theory is, from this perspective, a mere "recursive tool" for generating "interpretations" of the speaker's sentences—or, a mapping of each these sentences onto an acceptable translation in my language. Similarly, the problem of "individuating" the speaker's beliefs and desires is also a problem of translation. What I am doing, in individuating a speaker's belief on the basis of his acceptance of a given sentence *s*, is putting forward a sentence of my language as an appropriate translation of *s*. For instance, all I am doing in describing the speaker as believing that the sun will rise tomorrow on the basis of her acceptance of the sentence 'Le soleil se lèvera demain,' is describing the speaker as accepting a sentence whose role in her language is suitably similar to the role that "The sun will rise

tomorrow' plays in mine. We should also remember that the deflationist's notion of acceptance, unlike Davidson's notion of the holding true of a sentence, is a primitive notion that is not to be explained in terms of a transcendent notion of truth. The deflationist would describe the evidence for radical interpretation as consisting of facts about the conditions under which the speaker accepts various sentences, where in describing the speaker as accepting a given sentence, we are not ascribing to her grasp of any notion of truth. Again, as I have already noted in Chapter 2, worries might be raised about the intelligibility of the deflationist's notion of acceptance. It is not clear that we have a firm enough grip on the notion of a speaker's acceptance of a sentence, as distinguished from, and prior to, the idea of a speaker's holding a sentence to be true. But as I also mentioned in that chapter, my main goal here is not to refute deflationism. It is to assess its compatibility with a truth-conditional approach to meaning, and in so doing, gain a clearer view of what it is really like to be a deflationist about truth.

Granting, then, the possibility of a suitably deflationary conception of the radical interpreter's task along these lines, is there any reason why the deflationist cannot go on to give a coherent account of the constraints governing radical interpretation?

4.4 Logic and truth

As Davidson sees it, one of the first tasks facing the radical interpreter is to identify the logical constants in the speaker's language, and assign logical form to his sentences. The evidence for this will consist in sentences the speaker accepts "come what may," along with patterns of inferences he is disposed to make. To read a logic into the speaker's sentences on the basis of such evidence, we need to assume overall logical consistency and sound inferential dispositions on his part, which is one aspect of the so-called "principle of charity." Is any part of this process unavailable to a deflationist?

Davidson thinks we have no choice but to read our own logic into the speaker's sentences, where this is presumed to be first order quantification theory with identity. Moreover, he construes this task as bound up with the need to satisfy the formal constraints imposed by seeking a Tarskian truth theory for the language:

First we look for the best way to fit our logic, to the extent required to get a theory satisfying Convention T, onto the new language ... The evidence here is classes of sentences always held true or always held false ... and patterns of inference. The first step identifies predicates, singular terms, quantifiers, connectives, and identity; in theory, it settles matters of logical form.¹⁹

¹⁹ Davidson 1984, p. 136.

There are constraints of a formal nature that flow from the demand that the theory be finitely axiomatized, and that it satisfy Convention T (as appropriately modified). If the metalanguage is taken to contain ordinary quantification theory, it is difficult, if not impossible, to discover anything other than standard quantificational strictures in the object language.²⁰

As far as our goal of identifying and translating the speaker's logical apparatus is concerned, what would we lose in abandoning the goal of a truth-theoretic semantics for his language? One thing we would lose, according to Davidson, is the promise of a unique logic ascribed to the speaker's language.²¹ But Davidson himself thinks that indeterminacy of interpretation is not generally "of genuine concern", since "it marks the fact that certain apparent distinctions are not significant."²² Given his standpoint on indeterminacy, it is not clear to what extent Davidson thinks indeterminacy concerning logical form would itself be problematic. On the other hand, his commitment to reading first order quantification theory into a speaker's language does seem to go hand in hand with the goal of constructing a truth-theoretic semantics for this language. But even if this is right, what forces this logic into the language are clearly the formal constraints imposed by the construction of a Tarskian truth theory, rather than any particular way of understanding its truth locution. These constraints are in force whether or not we take its 'is true' to express a substantive notion of truth.

Thus, as far as the need to impose our own logic into the language is concerned, there is no obvious reason why the deflationist cannot satisfy it in the very way Davidson proposes.

²⁰ Ibid., p. 150.

²¹ Consider:

[T]he semantic constraint in my method forces quantificational structure on the language to be interpreted, which probably does not leave room for indeterminacy of logical form. (Ibid., n. 16, p. 136)

The result of applying the formal constraints is, then, to fit the object language as a whole to the procrustean bed of quantification theory. Although this can no doubt be done in many ways if any, it is unlikely that the differences between acceptable theories will, in matters of logical form, be great. The identification of the semantic features of a sentence will then be essentially invariant: correct theories will agree on the whole about the quantificational structure to be assigned to a given sentence. (Ibid., p. 151)

See also p. 153.

²² Ibid., p. 154.

4.5 Charity, agreement, and deflationary truth

Can an explanatory role for truth be found in Davidson's much discussed appeal to the principle of charity? What might seem to make this question particularly difficult to answer is the absence of any consensus as to how Davidson's version of charity is to be understood. Is Davidson's radical interpreter supposed to maximize truth in the beliefs attributed to a speaker, or is he to maximize agreement between the speaker and himself? Does it make a difference which line we take? Or should we forgo the idea of maximization altogether, and think of our task as that of minimizing unexplained error (or unexplained disagreement between the speaker and ourselves), thus replacing an unqualified appeal to truth maximization with something more along the lines of Richard Grandy's principle of humanity²³? For our purposes here, I do not think it matters how exactly we answer these questions.

If we accept Davidson's holism about the attribution of attitudes, and in particular, his depiction of the individuation of a belief as a matter of assigning it "a location in a pattern of beliefs,"²⁴ it does seem plausible to expect that without attributing to a speaker a rich network of *true* beliefs, we lose our grip on what any his beliefs and other attitudes might be about. As Davidson puts it, "too much mistake simply blurs the focus."²⁵ For example, to intelligibly attribute to an individual the concept of a cat, or, the ability to think thoughts about cats, depends on attributing to him true general beliefs about cats, as well the ability to recognize particular cats as such. Alternatively, charity can be spelled out in terms of agreement rather than truth, as Davidson himself often does. Without finding the speaker's beliefs to largely agree with ours, we lose our grip on what our particular disagreements might be about.

Whether or not there is an explanatory appeal to truth here might seem to depend on which of truth or agreement we take to be the crucial notion. As Michael Williams sees it, "the key notion here is agreement, not truth," since charity is a matter of "using our own beliefs as the basis for interpreting the beliefs of others."²⁶ And as he further notes,

²³ Davidson qualifies his appeal to "truth maximization" in such passages as Davidson 1984, p. 152, p. 169. Many have come to think that once we spell out the role that agreement really plays in constraining interpretation, we will come to favor Grandy's "principle of humanity" over the appeal to charity, where the principle of humanity involves "the condition that the imputed patterns of relations among beliefs, desires and the world be as similar to our own as possible" (Grandy 1973).

²⁴ "Thought and Talk," Davidson 1984, p. 168.

²⁵ Ibid.

²⁶ Williams 1999, p. 561.

when charity is spelled out in terms of agreement, truth does not “get so much as a mention” and therefore cannot be playing any explanatory role.²⁷

While this last point is right, it makes no difference to our purposes whether we spell out charity in terms of truth or of agreement. Whichever way we spell it out, it is easy to see that no explanatory appeal to truth is involved here.²⁸ Spelled out in terms of truth maximization, the principle involves attributing, for the most part, true beliefs to the other person. The truth predicate is clearly playing only a generalizing role here. Attributing mostly true beliefs to a speaker is a matter of attributing to him, for the most part, the belief that some object is a chair only if it is chair, that cats can fly only if cats can fly, and so on, for any candidate belief. Spelled out in terms of agreement, the principle need not even mention truth: it is a matter of attributing, for the most part, beliefs that agree with ours. That is, it is matter of attributing to the speaker the belief that some object is a chair only if we believe that it is a chair, that cats can fly only if we believe that cats can fly, and so on.

4.6 Observation sentences and the early stages of radical interpretation

Whether we construe charity in terms of truth maximization, agreement maximization, or the minimizing of unexplained error, this kind of principle is useful only when sufficient progress has already been made in interpreting a speaker’s words and individuating her beliefs. Charity in any of these senses thus does not seem to be of much use in the early phases of radical interpretation, for generating, *ex nihilo*, hypotheses about the interpretation of particular sentences.²⁹ What we need here is a

²⁷ Ibid.

²⁸ In any case, the concept of truth invoked here would be a concept of truth for beliefs or propositions, not a concept of sentential truth; whereas the deflationary accounts of truth whose compatibility with radical interpretation is in question here are accounts of sentential truth.

²⁹ In other words, while such holistic principles may plausibly capture the features an acceptable overall theory will end up having, they do not tell us how to get there. To see this, suppose I am trying to translate the sentence ‘Il neige ici’ as used by my friend Olivier. How might the principle of truth maximization help me form a hypothesis about what this sentence means? What I need, first, is a statement of the evidence, that is, of the conditions under which Olivier holds ‘Il neige ici’ to be true. Put differently, I am to find a sentence in my language that is true (or, that I hold true) on just those occasions in which Olivier holds ‘Il neige ici’ to be true. Clearly, however, this does not get me very far, since there are bound to be too many such sentences. Of course, Davidson did at some point hope that the formal constraints imposed by the demand for a recursive theory, together with the need to match truths with truths throughout the language, would sufficiently narrow down the range of available choices as to the truth-conditions of each sentence (see “Radical Interpretation”, Davidson 1984, p. 134). But I do not

principle that enables us to break into the circle of meaning and belief “from without,” by connecting these beliefs and sentences to objects and events in our shared environment.³⁰

Davidson is, of course, aware of the need for some principle connecting the speaker’s sentences to objects and events in our shared environment. Unsurprisingly, he assigns a crucial role to a speaker’s so-called observational sentences, the holding true (or false) of which by the speaker seems to systematically vary with observable changes in the speaker’s environment. Without such sentences to connect the speaker’s utterances with observable features of the environment, it is not clear how we could generate a rich enough network of hypotheses about what the speaker believes³¹—hypotheses that can then collectively be evaluated for coherence and rational intelligibility.

Davidson and Ramberg both think that the crucial role of a primitive transcendent notion of truth is to be found here, in Davidson’s account of the principles guiding the early phases of radical interpretation. Contrasting his “distal” approach to the meaning of observation sentences with Quine’s “proximal” approach, Davidson assimilates the difference between the two to “the opposition between a theory of meaning that makes evidence primary, and a theory of meaning that makes truth primary.”³²

In what follows, I will argue that the crucial differences between Davidson’s distal approach and the proximal approach that Quine once favored need not be spelled out as differences in their approaches to truth, or in any explanatory burdens imposed on

think that this addresses our difficulty here, which is, how can we so much as generate initially plausible statements of the evidence for radical interpretation?

³⁰ In light of this, we can appreciate Ramberg’s efforts to spell out charity as a “precondition for interpretation” rather than a pragmatic constraint for choosing between interpretations (Ramberg 1989, p. 77). As Ramberg explains, if we construe charity as a pragmatic principle for choosing between interpretation theories, we will come to favor such principles as Grandy’s principle of humanity. But

speaking of the principle of charity in terms of agreement and matching beliefs is to think of it only as giving a criterion for evaluating rival theories of interpretation of a language, not as a principle of theory construction. Why? Because until we have such theories, it makes no sense to speak of the interpreter as trying to match her beliefs with those of the speakers of the language she is trying to understand. (Ibid., pp. 71-72)

³¹ Davidson writes:

The interpretation of common predicates and names depends heavily on indexical elements in speech, such as demonstratives and tense, since it is these which most directly allow predicates and singular terms to be connected to objects and events in the world. (Davidson 2005a, p. 64)

³² “Meaning, Truth, and Evidence,” Davidson 2005b, p. 58.

this notion. So if we do not presuppose that our goal, as interpreters, is to spell out the content of a speaker's knowledge of her language, there is no reason left to think that deflationists could not exploit Davidson's proposed methodology.

Here is how Davidson sums up the difference between the proximal and the distal approach:

On the proximal theory, ... [observation] sentences have the same meaning if they have the same stimulus meaning—if the same patterns of stimulation prompt assent and dissent. In such a case we may speak of stimulus synonymy. Stimulus synonymy is not much direct help in translating non-observation sentences, but it does about as well as can be done for observation sentences, and supplies the basis for all translation.

The distal theory, on the other hand, depends primarily on shared causes which are salient for speaker and interpreter, learner and teacher. Meanings are shared when identical events, objects or situations cause or would cause assent and dissent. As a radical interpreter I correlate verbal responses of a speaker with changes in the environment. Inferring a causal relation, I then translate those verbal responses with a sentence of my own that the same changes in the environment cause me to accept or reject. This is the distal theory at its simplest, subject to various fairly obvious caveats.³³

The first thing to note is that the notion of truth does not explicitly figure in this summary of the contrast between the two approaches. The main difference is that what matters to interpretation are, in the one case, the matching of sentences alike in stimulus meanings (where stimuli are construed as proximal), and in the other case, the shared external causes of our holding sentences true.

Might crucial differences in their conceptions of truth nonetheless emerge as the real source of this contrast between the proximal and the distal approach? It might seem so, since Davidson's main worry about Quine's approach to the meaning of observation sentences is that it leaves us with no satisfactory conception of what it would take for any such sentence to be true. As Davidson sees it, the proximal approach divorces truth and meaning, thereby opening the door to either skepticism about the external world, or to an unpalatable form of relativism about truth³⁴. For example, given Quine's assimilation of the meaning of such sentences as 'Lo, a rabbit!', 'Red!' to their stimulus meaning, and given the intuitive connection between meaning and of truth, it seems

³³ Ibid., p. 54.

³⁴ Davidson characterizes this as involving a further relativization of sentential truth to individuals, over and beyond the "familiar" relativization of truth to a language (Davidson 2005b, p. 57).

reasonable to expect these sentences are being construed as descriptions, or reports, of our sensations, and thus as answerable, for their truth, only to how things are with these sensations. As Davidson notes, however, Quine rejects this account of the subject matter of our sentences, or of what it takes for them to be true³⁵, and thus faces the following difficulty:

[L]et us imagine someone who, when a warthog trots by, has just the patterns of stimulation I have when there's a rabbit in view. Let us suppose the one-word sentence the warthog inspires him to assent to is 'Gavagai.' Going by stimulus meaning, I translate his 'Gavagai' by my 'Lo, a rabbit!', though I see only a warthog and no rabbit when he says and believes (according to the proximal theory) that there is a rabbit ... According to the proximal theory [the speaker] will be wrong to some degree about the world as conceived by a normal interpreter ... yet [he] has a theory that saves the structure of his sensations.³⁶

Tying meaning to proximal causes and truth to features of the environment can thus result in our ascribing systematic error to the beliefs of others. And this is problematic because, as Davidson explains, once each speaker "notices how globally mistaken others are, and why, it is hard to think why he would not wonder whether he had it right. Then he might wonder what it could mean to get it right."³⁷

Now, if we accept Quine's proximal approach to meaning, the connection that here went missing between meaning and truth can be restored if we adopt a proximal approach to truth as well. We could, in this vein, think of truth for our sentences (or at least, our observation sentences) as a matter, not of how things are with the world they appear to be about, but of how things are with our proximal stimulations. Thus, we could think of 'Lo, a rabbit!' as uttered by me as true in just those circumstances in which there occurs a certain pattern of stimulations of my nerve-endings. Similarly, my interlocutor's sentence 'Gavagai!' would be true (as uttered by him) only on those occasions in which his nerve endings exhibit the same, or a relevantly similar, pattern of stimulations. 'Gavagai!' and 'Lo, a rabbit!' thus have the same meaning, but different truth-conditions, since the truth of 'Lo, a rabbit!', as uttered by me, depends on the state of my nerve endings, while the truth of 'Gavagai!', as uttered by my interlocutor, depends on the state of his nerve endings.

³⁵ "I have forces from real external objects impinging on our nerve endings, and I have us acquiring sentences about real external objects" (Quine 1981, p. 181, quoted in Davidson 2005b, p. 56).

³⁶ Davidson 2005b, p. 56.

³⁷ Ibid.

Suppose, as in Davidson's example, that my interlocutor's 'Gavagai!' utterances and my 'Lo, a rabbit!' utterances are caused by different conditions in our shared environments: my utterances are caused by the presence of rabbits, while my interlocutor's are caused by the presence of warthogs. Since we are tying truth to proximal stimuli, this discrepancy does not matter at all, as far as the truth of our respective utterances are concerned. A proximal approach to both meaning and truth thus avoids the risk of attributing global error to speakers' utterances and beliefs. However, it does this at the cost of two unhappy outcomes: first, it deprives truth of any connection with the external world that we take ourselves to be talking and thinking about; secondly, it leads to a different kind of divorce between truth and meaning, by severing the connection between sameness of meaning and sameness of truth-conditions.

The proximal approach to meaning thus leaves us with no plausible conception of truth and of its connection with meaning. And this is what motivates Davidson's distal approach. But does a distal approach to meaning, in and of itself, depend on the availability of a substantive notion of truth, or is there an intelligible version of it available to a deflationist?

To help us answer this question, let us further examine how a distal approach to truth is supposed to help avoid the difficulties facing Quine's proximal approach. As Davidson explains, an interpreter has no way to identify the salient cause of a speaker's utterances but to single out features that are salient from her point of view, and gather evidence for the hypothesis that these features are also salient for the speaker. The evidence for this will consist in the speaker's exhibiting similar responses, on different occasions, to these very features. Thus:

What makes communication possible is the sharing, inherited and acquired, of similarity responses. The interpreter's verbal responses class together or identify the same objects and events that the speaker's verbal responses class together. If the interpreter also classes together the verbal responses of the speaker, he can correlate items from two of his own classes; verbal responses of the speaker he finds similar and distal objects and events that he finds similar. To the latter he has his own verbal responses; these provide his translation or interpretation of the speaker's words. Thus the common cause becomes the common subject matter of speaker and interpreter.³⁸

Note, again, that truth is not explicitly invoked in this story about how a radical interpreter is to fix the interpretation or translation of basic observation sentences. The translation of such sentences is explained in terms of the notion of mutually salient causes of utterances, which in turn relies on the notion of shared similarity responses or

³⁸ Ibid., p. 61.

shared interests. It is not clear that any of this depends on an explanatory appeal to truth.

From a perspective that takes translation to be the relevant task, our goal, in construing the relevant causes as distal, is to yield a notion of translation that allows for truth to be preserved by translation. But this does not mean that we need to appeal to a prior notion of truth and take truth preservation as a constraint on translation. The difficulties with a proximal approach to the interpretation of observation sentences are independent of whether we take our primary task to be that of translating the speaker's sentences into our own, or that of assigning truth-conditions to these sentences. Accordingly, the argument in favor of a distal approach is not, in itself, an argument for a primitive transcendent notion of truth, that is, an argument against defining truth for the sentences of others in terms of a prior notion of translation. Davidson's goal, in construing the relevant causes as distal, is simply to make sure that translation does not "lose track of truth."³⁹ But this goal can be met whether or not we think of truth as intelligible prior to translation, and thus whether or not we attribute to the interpreter grasp of primitive, transcendent notion of truth.

For instance, consider Field's way of making sense of truth in application to the sentences of others. On his approach, I am meant to make sense of translation first, without appealing to any transcendent notion of truth, and can then define truth for the sentences of others in terms of pure disquotation and translation. For example, to say that a given sentence *s* is true in Field's "extended disquotational" sense is to say that there is an acceptable translation of *s* into a sentence of my language that is true in the purely disquotational sense.⁴⁰ Given the way that truth for the sentences of others has been defined, translation does preserve truth-values, though this does not serve as a constraint on acceptable translation.⁴¹

³⁹ See p. 79 of "Pursuit of the Concept of Truth," in Davidson 2005b.

⁴⁰ It bears emphasizing that the notion of translation involved here does not appeal to sameness of truth-values, or any other constraints pertaining to the truth or falsity of the sentences of others.

⁴¹ In "Deflationism about Meaning and Content" (Field 2001, pp. 104-140), Field characterizes pure disquotational truth as follows:

As a rough heuristic, we could say that for a person to call an utterance true in this pure disquotational sense is to say that it is true-as-he-understands-it. ... As the heuristic suggests, a person can meaningfully apply 'true' in the pure disquotational sense only to utterances that he has some understanding of; and for such an utterance *u*, the claim that *u* is true (true-as-he-understands-it) is cognitively equivalent (for the person) to *u* itself (as he understands it). (p. 105)

Of course, if we take ourselves to grasp a substantive notion of truth⁴², then we will feel the need to impose such conditions on correct translation as the matching of truth-values. But if, as on Field's disquotational approach, we have no such notion to begin with, then we cannot and need not spell out any such constraints. In particular, we can adopt Davidson's own approach to the translation of observation sentences in terms of mutually salient causes of our utterances of such sentences. Once translation is under way, we can then secure the sought-for connection between meaning and truth by defining truth for the languages of others in such a way that it is preserved by translation.⁴³

4.7 Substantive truth and the reification of meaning

Ramberg would disagree with these remarks. As he sees it, "interpretation from one language into another works only because the interpreter possesses a pre-theoretical understanding of truth."⁴⁴ Moreover, this pre-theoretical grasp of truth required for interpretation is, according to Ramberg, grasp of transcendent notion—or, as Ramberg puts it, a "trans-linguistic" notion, applicable across speakers and languages:

It is this very intuition of truths-for-languages as somehow *the same* that drives interpretation ... The concept of truth that underlies a theory of interpretation is a concept of absolute truth. That is to say, in a true T-sentence, s and p are appropriate to the occasions of empirical observation in the same manner. It is

Field puts forward at least two different proposals about how we could define 'extended disquotational truth,' or, truth for sentences in other languages. On p. 128 of Field 2001, he defines 's is true' in the extended disquotational sense as "s is synonymous with a sentence of ours that is true in the purely disquotational sense." But soon afterwards he introduces a weaker construal of this notion, one that does not commit itself to a notion of interpersonal synonymy, construing 's is true' as "there is a good translation of s into a sentence of ours that is true in the purely disquotational sense" (p. 129), where standards of good translation can be taken to be interest-relative or context dependent. See also Field 1986.

⁴² As I have argued we should, if we are to think of ourselves as knowing the truth-conditions of our sentences.

⁴³ In other words, the connection between translation and truth can be secured whether we invoke a transcendent notion of truth to constrain translation, or appeal to an independently intelligible notion of translation to constrain our definition of truth for the languages of others.

⁴⁴ Ramberg 1989, p. 75.

by virtue of this trans-linguistic notion that [an interpreter] is able to formulate an empirical theory that ... actually interprets the language.⁴⁵

Ramberg further argues that when Davidson follows Quine in speaking of translation as a legitimate goal, he encourages “slippage into reification, thinking of meaning as something to be captured by, but given independently of, the sentences we use,”⁴⁶ and thus obscures the main advantages of his approach over Quine’s. Ramberg’s insistence on the need for a substantive notion of truth and his renouncing of translation as a legitimate goal in radical interpretation are, I think, connected. As I will explain, it is a certain perceived—though ultimately spurious—difficulty with translation as our goal that provides the only plausible rationale for Ramberg’s insistence on the crucial role of a substantive notion of truth in interpretation.

Why does Ramberg think that radical interpretation depends on the interpreter’s pre-theoretical grasp of a “trans-linguistic” notion of truth?⁴⁷ The above-quoted remarks suggest that this has to do with the alleged fact that in deciding to use a sentence *p* to assign truth-conditions to another’s sentence *s*, the interpreter relies on the assumption that “*s* and *p* are appropriate to the occasions of empirical observation in the same manner”—that is, that they are “true” in the very same sense, on those occasions. If the interpreter needs to rely on such an assumption, surely he must come equipped with grasp of a substantive notion of truth!

⁴⁵ Ibid., pp. 76-77.

⁴⁶ Ibid., p. 67.

⁴⁷ Consider Ramberg’s more modest claim that an interpreter’s pre-theoretical grasp of truth is a necessary condition for radical interpretation. This, according to Ramberg, means that the interpreter must be able to use the truth predicate in application to sentences of some language—namely, his own—in order to be in a position to interpret another speaker. But as Ramberg further explains, this merely “amounts to the requirement that the interpreter understands the language into which [or in which?] another language is to be interpreted” (p. 75). If this were all that Ramberg was after in demanding a pre-theoretical understanding of truth, there would not be much to disagree with.

However, from a deflationist point of view, a question does arise as to whether we might be able to understand a language without being in a position to speak of the truth or falsity of its sentences, that is, without having a purely disquotational truth predicate at our disposal. One might think that since we can meaningfully apply a purely disquotational truth predicate only to sentences we understand, competence in a language must be possible prior to grasp of a purely disquotational truth predicate. But I don’t find this convincing. Compare: one can meaningfully use logical operators only in the context of sentences that one understands (e. g., one can meaningfully use the sentence ‘It is not raining’ only if one understands the sentence ‘It is raining’), but it does not follow from this that it must be possible to speak a language devoid of logical operators.

But why does Ramberg think that interpretation depends on the interpreter's taking the speaker's sentences and his own to be true, in the very same sense, on the occasions of empirical observation? We can easily make sense of the interpreter's need to regard his own sentence *p* as true on these occasions. This simply amounts to the need to assume that, in using *p*, he has correctly characterized the conditions that cause the speaker to accept *s* (or to hold *s* true).⁴⁸ As for the assumption that the speaker's sentence *s* is true, it simply follows from the alleged need to directly infer a T-sentence for *s* from a statement of the conditions that cause the speaker to hold *s* true. For, without assuming that the speaker holds *s* to be true under those very conditions in which it is true (or would be true, as uttered by the speaker), I could not derive, from any assumptions as to the conditions that cause the speaker to hold *s* to be true, the conclusion that *s* as uttered by him would true under those very conditions.

Thus, Ramberg's claim that interpretation depends on holding truth "constant" between speaker and interpreter simply boils down to the claim that interpretation cannot get off the ground unless the interpreter assumes that, for the most part, the conditions that cause a speaker to hold a sentence true are the very conditions under which it is true. The interpreter could not rely on this assumption unless she has a pre-theoretical grasp of a notion of truth applicable to sentences she does not already understand. But clearly, the only reason for thinking that the interpreter needs to rely on this assumption is a construal of her task as that of directly deriving truth sentences from generalizations about the conditions that cause the speaker to hold various sentences to be true. In thinking of the early stages of radical interpretation as involving such inferences, Ramberg is already portraying the radical interpreter as employing a substantive notion of truth. His further observation, that such inferences would depend on the assumption that a speaker's sentence is true under the very conditions that cause him to hold it true, is beside the point.

If this is right, Ramberg's claims that we need to "hold truth constant" between the speaker and ourselves simply relies on, rather than helping explain, the assumption that we bring to our task a substantive notion of truth. But is there any reason, stemming from consideration of the ends and means of interpretation, for accepting this assumption? As far as I can tell, the only explanation of Ramberg's insistence on attributing to the interpreter grasp of a substantive notion of truth is his dissatisfaction with the perceived alternative—namely, Quine's portrayal of the interpreter's task as involving translation of the speaker's sentences into her own. Without attributing to the interpreter grasp of a primitive transcendent notion of truth, in terms of which she can formulate hypotheses about the relevant semantic features of a speaker's sentences, we

⁴⁸ Note, however, that this is not an assumption the interpreter invokes in inferring a T-sentence from a statement of these conditions—it is, rather, something to which he is already committed in putting forward a statement of these conditions. Thus, there are no further assumptions about truth involved here, beyond the necessity of taking one's own beliefs and judgments at face value.

would, fears Ramberg, be stuck with a model of her task as not essentially different from the radical translator's, despite certain differences in their methodologies.

But why would we want to avoid this outcome? As I have explained, the main advantages of Davidson's distal approach over Quine's proximal approach can be characterized whether or not we assimilate the radical interpreter's task to the radical translator's.

Let me spell this out in more detail. On Quine's approach, the translation of observation sentences is based on the matching of stimulus meanings. Having identified the proximal causes of the speaker's assent to a sentence, the radical translator then tries to find a sentence of her own that she is caused to assent to by the same (or, relevantly similar) proximal causes. Radical translation of the speaker's observation sentences is thus a two-step process; involving, first, identification of the proximal causes of the speaker's assent to a given sentence, and secondly, the choice of a sentence of our own to which we are caused to assent by relevantly similar proximal causes. In other words, we are to first assign a stimulus meaning to the speaker's sentence, and then find a sentence of our own that has the same, or a similar enough, stimulus meaning.

As we have seen, Davidson does at times describe his distal approach just along the lines of Quine's two-step process, but with the appeal to proximal causes replaced by an appeal to mutually salient distal causes.⁴⁹ We might thus say that for Davidson, as for Quine, what matters to the meaning or translation of observation sentences is their "stimulus meanings," but that they have divergent approaches to the kind of stimuli that matter to meaning: proximal for Quine, distal for Davidson. From a perspective that underplays the significance of the differences between "radical interpretation" and "radical translation," this seems to be a legitimate way of describing the differences between the two approaches.

As Ramberg sees it, however, this way of spelling out the difference between Davidson and Quine only helps distort the advantages of Davidson's truth-centered approach to meaning over Quine's. For, insofar as we describe Davidson's approach as involving the matching of stimulus meanings, we are thinking of it as committed to objective, meaning-determining features of sentences, thus undermining what Ramberg takes to be its essential promise: that of avoiding any explanatory appeal to the notion of meaning.

Instead, Ramberg proposes that when a radical interpreter is describing the conditions that cause a speaker to assent to a sentence (or to hold it to be true), she is "ipso facto producing the truth-conditions of the sentence, and that is all she has to do to get her

⁴⁹ For example, consider the longer passage I quoted towards the beginning of section 6, from Davidson 2005b, p. 54.

theory started.”⁵⁰ So if we approach radical interpretation in the way Ramberg thinks we should—by invoking a pre-theoretical notion of trans-linguistic truth—there needs to be only one step involved: that of describing the truth-conditions of a sentence by describing the conditions that cause the speaker to hold it true. There is no *further* task of pairing a sentence of ours with the speaker’s, or of finding a sentence of ours that is true, or that we hold true, under the same conditions. What makes it possible for the very sentence we use to describe the cause of the speaker’s holding true of *s* to qualify as a plausible candidate for the translation of *s* in our language is, of course, our conception of the relevant cause as distal. If we instead adopted Quine’s proximal approach, then the sentence used in describing the relevant cause of the speaker’s holding true of *s* could not plausibly be used to assign truth-conditions to *s*, nor, for that matter, constitute a plausible candidate for its translation into our language.

All of this seems right, and the way Davidson sometimes depicts the early phases of radical interpretation may obscure important advantages of his approach over Quine’s. When Davidson describes the process as a matter of first identifying the conditions that cause a speaker to hold a sentence true, and then finding a sentence of ours to “match” these conditions, he obscures the dispensability of this last step, given a distal approach to the relevant cause. But it is not clear that this advantage of Davidson’s distal approach to meaning over Quine’s depends on treating truth rather than translation to be the operative notion in interpretation. For, paralleling Ramberg’s suggestion that in describing the conditions that cause a speaker to hold a sentence *s* to be true, we have already “produced” the truth-conditions of *s*, we can insist that in describing the relevant causes, we have thereby “produced” a sentence that translates, interprets, or mean the same as *s*. Once we replace Quine’s appeal to proximal causes with an appeal to mutually salient features of the environment, this way of simplifying the methodology of radical interpretation is available, whether we take our operative notion to be that of truth or of translation.

4.8 Conclusion

Davidson’s distal approach to the interpretation of observation sentences has important advantages over Quine’s proximal approach. One advantage, as just discussed, is that a distal approach gives us a more plausible method of generating hypotheses about how a speaker’s observation sentences are to be understood. Another important advantage for

⁵⁰ Ramberg 1989, pp. 67-68.

the distal approach is that it helps avoid the imputation of systematic error to the speaker, or the possibility of systematic disagreement between speaker and interpreter.⁵¹

These two advantages for the distal approach are there, whether we think of the interpreter's primary task as that of assigning truth-conditions to the speaker's sentences or that of translating the speaker's sentences into her own. In other words, these two advantages can be spelled out and acknowledged from a deflationary perspective on truth. Moreover, as we have seen, Davidson's goal of ensuring that translatability, or sameness of meaning, guarantees sameness of truth-conditions, does not depend on treating truth-preservation as a basic constraint on translation.⁵² The sought for connection between meaning and truth can be secured in the context of an immanent approach to truth, by using an independently intelligible notion of meaning to constrain our notion (or notions) of truth for languages other than our own.

Does this answer the question that is the title of the present chapter? Can a deflationist, after all, make sense of radical interpretation? Well, in a sense, yes, and in another sense, no. If the goal of interpretation is to construct or confirm a truth theory for another speaker's language (construed as an account of the speaker's knowledge of her language), then the deflationist cannot make sense of radical interpretation, simply because the problem it is brought on to solve cannot be articulated from a deflationary perspective. If, on the other hand, interpretation is a process by which we arrive at an acceptable way of pairing sentences of another speaker's language with sentences of our own, then the deflationist can make sense of radical interpretation. For, as I have argued in this chapter, once we agree to suitably deflationary conceptions of the radical interpreter's task and starting point, no aspect of his methodology turns out to be unavailable to the deflationist. What I hope to have shown, however, is that radical interpretation takes on a radically different significance depending on whether we accept deflationism.

⁵¹ As we have seen, the danger of this lurks when we take sameness of proximal causes, rather than sameness of external circumstances, to be what matters in interpretation.

⁵² Nor does it depend on rejecting Quine's proximal approach to meaning.

Chapter 5

Some Consequences

I have argued that deflationism is incompatible with Davidson's conception of a truth theory as a compositional account of a speaker's knowledge of her own language, because it does not allow us to ascribe to competent speakers knowledge of the truth-conditions of their sentences. Reflection on the role of truth in Davidson's approach has led to me to paint a different picture of Davidson's project, and of the relationship between its two components—the proposal to use truth theories as meaning theories, and the account of radical interpretation—than the one commonly presupposed. In the first three sections of this final chapter, I would like to more explicitly bring out some consequences of the results of Chapters 1 through 4 for our understanding of Davidson's philosophy of language. Finally, in section 5, I will highlight some more general implications of the results of Chapter 2, beyond those that bear on our understanding Davidson's own approach to meaning.

5.1 Truth, truth theories, and radical interpretation

Does Davidson think that Tarskian truth theories shed any light on the concept of truth? Is his own proposal to use such theories as meaning theories meant to shed light on this concept? Consider the following passages:

[E]ven without an answer to the question of how we know when a definition of truth applies to a given language, Tarski has shown how the concept of truth can be used to give a clear description of a language.¹

[W]e have asked what the formal properties of the concept are when it is applied to relatively well-understood structures, namely languages ... It remains to

¹ Davidson 2005a, p. 30.

indicate how a theory of truth can be applied to particular speakers or groups of speakers.²

These passages suggest that Davidson assigns truth theories an important, albeit limited role to play in illuminating the concept of truth. As he sees it, such theories shed light on the formal properties of truth, that is, on the structure of truth as applied to particular languages. But Davidson thinks that there is more to the concept of truth than its formal properties. He takes truth to be a crucial explanatory concept, with constitutive ties to the concepts of meaning and of the propositional attitudes. It is thus natural to expect that any light that Davidson aims to shed on these connections, and on the empirical content of the concept of truth, is to be found in his account of radical interpretation. Indeed, this is precisely how Davidson seems to understand the significance of radical interpretation. He describes his goal, in radical interpretation, as that of illuminating the concept of truth by tracing “the connections between the concept of truth and the human attitudes and acts that give it body.”³ Further along in the same paper, he describes a theory of truth as a structure

we can find, with an allowable degree of fitting and fudging, in the behavior of more or less rational creatures gifted with speech. It is in the fitting and fudging that we give content to the undefined concepts of ... belief and desires ... and by way of theories like Tarski’s, to the undefined concept of truth.⁴

Here, then, is how Davidson seems to construe the division of labor between truth theories and the account of radical interpretation. He takes truth theories to illuminate the purely formal properties of truth, and the account of radical interpretation to shed light on the content of the concept, by tracing its connections to “the human attitudes and acts that give it body.”

If what I have argued in this dissertation is right, however, there are two points that Davidson’s remarks fail to bring out. First, even before embarking on the topic of radical interpretation, Davidson has already forged an important connection between truth and human attitudes, in proposing that a truth theory can be used as a meaning theory. In proposing that a Tarskian truth theory can be used to give a clear description of the language spoken by a speaker on a particular occasion, Davidson is proposing that such a theory can be used to give a compositional account of the speaker’s knowledge of her language. Further, I have argued that this involves attributing to the speaker knowledge of the truth-conditions of sentences of her language. If this is right,

² Ibid., p. 37.

³ “The Folly of Trying to Define Truth,” Davidson 2005b, p. 35.

⁴ Ibid., p. 37.

then in proposing that a truth theory can be used as a meaning theory, Davidson has already conveyed something very important about the content of the concept of truth, something that goes beyond an account of its “formal properties”: he has described a speaker’s competence as involving knowledge of the truth-conditions of its sentences. So while Davidson is not exactly wrong in saying that “it is in the fitting and fudging” that we give content to the concept of truth employed in a truth theory serving as a meaning theory, it is not *only* in the fitting and fudging that he is giving content to this concept.⁵

This brings me to my second point. Davidson says that his account of radical interpretation illuminates the concept of truth by answering the question “what makes a theory of truth correctly apply to a speaker?”⁶ But if our goal, in “applying” a truth theory to a particular speaker, is to use the theory as a meaning theory for the speaker, then the theory “correctly applies” to the speaker if its T-theorems capture what the speaker knows about each of her sentences in understanding it the way she does.

If this is right, then our primary question, in empirically confirming a truth theory as a meaning theory for a given speaker, is not “Is our truth theory true of the speaker’s language?”, or “Are its T-theorems true of the speaker’s sentences?”, but rather: “Do these theorems state the content of the speaker’s knowledge of her language?” or “Can knowledge of what is stated by these theorems be reasonably attributed to the speaker?”⁷ So the fitting and fudging that radical interpretation involves is not primarily a matter of confirming the truth of T-sentences, construed as generalizations about the speaker’s linguistic behavior. It is a matter of confirming the attribution to the speaker of knowledge of what is stated by these T-sentences. Thus, if radical interpretation does give empirical content to the concept of truth, it is by giving content to our attribution to a speaker of knowledge of the conditions under particular sentences are true.⁸

⁵ Of course, this is not to deny that the account of radical interpretation can shed further light on the concept of truth by spelling out the empirical constraints governing the construction of truth theories in particular cases.

⁶ *Ibid.*, p. 37

⁷ Of course, unless a T-theorem is true, it cannot be stating something someone knows. So our truth theory’s being true of the speaker’s sentences is a necessary condition on its being an empirically adequate meaning theory for her language.

⁸ In making these observations, I do not mean to suggest that the methodology of radical interpretation will necessarily be affected, depending on whether we construe our primary task as that of constructing a truth theory that is simply true of the speaker’s sentences, or that of constructing a truth theory that captures the speaker’s knowledge of her language.

5.2 Holding a sentence to be true

I would now like to highlight two features of Davidson's notion of the holding true of a sentence, and further examine its connections with the notions of meaning and belief.⁹

First, as I have already noted in previous chapters, Davidson's notion of holding true, unlike Quine's notion of assent, is not available to a deflationist. Let us spell out exactly why that is so.

Davidson tells us that a speaker's holding true of a sentence is not a use of language and not something a speaker does. It is an attitude that can be revealed in the speaker's linguistic behavior.¹⁰ By contrast, Quinean assent to a sentence is something a speaker does, an episode of behavior. This, however, does not explain why Quinean assent is, while the holding true of a sentence is not, available to a deflationist. As we have seen, the deflationist's analog of Davidson's notion of holding true is the notion of acceptance, not the notion of assent. Accepting a sentence, just like holding one to be true—and unlike assenting to or asserting a sentence—is not itself something a speaker does. However, just like the holding true of a sentence, a speaker's acceptance of a sentence is supposed to be revealed in her behavior—in what she does with her words, and how she responds to the utterances of others: for instance, in her asserting the sentence,¹¹ or her assenting to it.

Here is the reason why Davidson's notion of holding true is not available to a deflationist. In describing another speaker as holding a sentence of her language to be true, I am using a notion of truth applicable to the speaker's sentences and attributing to the speaker grasp of the very same notion. This notion of truth, grasp of which I am attributing to the speaker, cannot be an immanent notion of truth applicable by the speaker only to her own sentences, since this would be a notion that I myself cannot grasp. And it cannot be a transcendent notion of truth that I myself would define in terms of an immanent notion (applicable to my own sentences) since this would be a notion that the speaker does not grasp. Therefore, the notion of truth grasp of which I am attributing to the speaker must be a substantive notion of truth, in the sense defined

⁹ Davidson came to assign a crucial role, not only to hold-true attitudes, but to an agent's "preferring true" of one sentence over another. In fact, he came to think of the preferring true attitude as the most basic of the attitudes: "truth ... rests in the end on belief, and, even more ultimately, on the affective attitudes" (Davidson 2005a, p. 75). To simplify exposition, I am mostly ignoring affective attitudes in what follows, but whatever I say about Davidson's notion of the holding true of a sentence applies, with suitable modifications, to the notion of preferring true.

¹⁰ "Epistemology and Truth," Davidson 2005b, p. 190.

¹¹ By 'asserting a sentence,' I mean 'using a sentence assertorically.'

in the previous chapter—that is, a transcendent notion of truth that cannot be defined in terms of an immanent notion. And this is a notion that is not available to a deflationist.¹²

Secondly, the holding true of a sentence is not a psychological state that an individual could be in without having beliefs and other intentional states. It is not a more “basic” psychological state, in the sense of being attributable to an individual without having to assume that she has beliefs. To see this, let us reflect on the following questions: what are we doing when we describe a speaker as holding a sentence of her own language to be true? What assumptions are we making about the speaker and her propositional attitudes? Here is how Davidson describes one way of connecting these notions:

If a sentence *s* of mine means that *P*, and I believe that *P*, then I believe that *s* is true. What gives my belief its content, and my sentence its meaning, is my knowledge of what is required for the belief or the sentence to be true.¹³

In such cases, it is my belief that *P*, together with my knowledge that *s* (as I use it) is true if and only *P* that explain my holding *s* to be true. More generally, in the normal cases, a speaker holds (or believes) a sentence *s* of her language to be true because she knows the conditions under which it is true and believes these conditions to hold. And this is precisely why facts about the conditions under which a speaker holds various sentences true can reveal what the speaker means and what she believes. In other words, as Davidson sees it, a speaker’s holding true of sentences is a reliable guide to what she means and what she believes insofar as, and precisely because, it can be explained as a result of her beliefs about the conditions under her sentences are true, and her beliefs about the world.

The only sense in which Davidson treats “holding true” as basic is, therefore, an epistemic one. His crucial assumption is not that we could describe a speaker as holding sentences of her language to be true without presupposing that *she* knows the conditions under which they are true. It is, rather, that we could come to know that a speaker holds this or that sentence true without *ourselves* drawing on any particular assumptions about the meaning of her sentences or the content of her other attitudes. In particular,

¹² My understanding of Davidson’s notion of holding true stands in stark opposition to Dummett’s in the following passage:

In Davidson’s formulation, ‘holds *S* true’ is not to be construed as ‘holds that *S* is true’, that is, as appealing to an already understood notion of a statement’s being true. Rather, it is meant to express a relation between a speaker and a possible utterance, by him or another, which we can grasp before we attain the concept of a statement’s being true. (Dummett 1991, pp. 108-109)

¹³ “Epistemology and Truth,” Davidson 2005b, p. 189.

Davidson is assuming that we could detect a speaker's hold-true attitude towards a sentence without knowing what the speaker knows about the conditions under which this sentence is true (and therefore without knowing what she thereby believes). Without this assumption, radical interpretation, as Davidson conceives of it, could not get off the ground.

5.3 What is the point of the exercise?

These two points help explain why it would be a mistake to read into Davidson's account of radical interpretation a "full-blooded" theory of meaning in Dummett's sense.¹⁴ Here is how Rattan proposes to understand Dummett's full-bloodedness requirement on a theory of meaning:

A theory of meaning is full-blooded to a greater or lesser degree to the extent that it provides a description in an idiom lesser or greater intentional or semantic of the abilities to use the language that constitutes speakers' knowledge of a meaning theory.¹⁵

As Rattan characterizes it here, full-bloodedness is a matter of degree. A theory of meaning is full-blooded, at least to some degree, if it explains full-fledged intentional and semantic capacities as constituted by capacities that can somehow be characterized as intentional or semantic "to a lesser degree." What I have been urging is that even though radical interpretation depends on treating facts about a speaker's holding true and preferring true of sentences as basic—relative to facts about a speaker's beliefs, desires, and knowledge of truth-conditions—in a certain epistemic sense, Davidson should not be taken to be construing such facts as basic in any conceptual or metaphysical sense. If this is right, his account of radical interpretation cannot be described as full-blooded in even a weak sense.

In *Truth and Predication*,¹⁶ Davidson explains that his objective, in appealing to hold-true and preferring-true attitudes towards sentences,

was not to avoid intentional states; it was to avoid *individualized* intentional states, *intensional* states, states with (as one says) a propositional object. A preference for

¹⁴ See pp. 107-113 of Dummett's "What Is a Theory of Meaning" (Dummett 1993) and Chapter 5 of Dummett 1991 for different attempts at spelling out his full-bloodedness requirement.

¹⁵ Rattan 2004, p. 225.

¹⁶ Davidson 2005a.

the truth of one sentence over another is an extensional relation that relates an agent and two sentences (and a time). Because it can be detected without knowing what the sentences mean, a theory of interpretation based on it can hope to make the crucial step from the nonpropositional to the propositional.¹⁷

However, it is important not to misunderstand the nature of the “step” from the nonpropositional to the propositional that Davidson’s account of radical interpretation is meant to bridge. Davidson is assuming that we can “detect” a speaker’s preferences between, or holding true of, her own sentences, without knowing what these sentences mean and thus without knowing what the speaker believes and desires. But this does not mean that we can intelligibly take an individual to stand in these extensional relations to sentences without taking her to have beliefs, desires and other attitudes. So the step from the nonpropositional to the propositional that Davidson is aiming to bridge does not involve explaining facts of the one kind as determined, or constituted by, facts of the other kind.

The point of the account of radical interpretation to illuminate the notions of belief, desire, and meaning, by showing how we could “arrive at all of them at once,”¹⁸ that is, how we could, on the basis of minimal assumptions about an individual’s psychology, assign content to a person’s beliefs, desires, and linguistic acts, and meaning to her sentences. The role of hold-true and prefer-true attitudes towards sentences is not to provide a reductive base for an account of the propositional attitudes, for the point of the exercise is not to show how the propositional (or, individualized intentional states) can be reduced to, or constructed out of, the non-propositional (or, non-individualized intentional states). It is to shed light on our concepts of meaning, truth, and propositional attitudes, and the structure of ordinary psychological explanation, by locating the crucial role that each of these concepts plays in making rational sense of a speaker. The attribution to a speaker of hold-true and prefer-true attitudes towards sentences are minimal assumptions that Davidson thinks we can justifiably make without already knowing what any of the speaker’s sentences or utterances means (or the conditions under which it is true).

Dummett proposed to construe Davidson’s account of radical interpretation as a full-blooded, albeit holistic, theory of meaning. What I am suggesting is that mere reflection on Davidson’s concept of holding true is enough to convince us that this would amount to a serious misconstrual of his project. The account of radical interpretation could not involve a complicated form of reduction of the propositional to the non-propositional simply because, in describing a speaker as holding a sentence true, I am already thinking

¹⁷ Ibid., p. 67.

¹⁸ Davidson 2005a, p.74.

of her as having propositional attitudes—albeit ones of whose content I may be ignorant.

In light of this, let us pry apart some of what is right and wrong on different sides of the debate over the full-bloodedness of Davidson's approach. McDowell has famously defended "modesty" (that is, the rejection of full-bloodedness) in theories of meaning,¹⁹ and argued that Davidson's own account cannot but be a modest one. His argument highlights Davidson's commitment to the interconnectedness of meaning, propositional attitudes, and the holding true of sentences, as I myself have done here.²⁰ But as McDowell sees things, the interconnectedness of these notions also means that we cannot plausibly regard facts about holding true as evidence for a meaning theory.

Dummett, Peacocke and more recently, Rattan, see things differently. Rattan, in particular, argues that McDowell fails to appreciate the special status of the notion of holding true as an intentional, albeit non-individualized, attitude. As Rattan sees it, this special status allows Davidson to treat facts about holding true as not only ultimate evidence for radical interpretation, but as the basis for a "logical construction of semantic competence out of use."²¹ More specifically, Rattan describes Davidson's conception of, as he calls it, the "interaction of the metaphysics and epistemology of language" as centered on the idea that "meaning is made out of what the radical interpreter, idealized, needs to have access to in order to come to know meanings."²²

¹⁹ See "In Defense of Modesty" and "Another Plea for Modesty," in McDowell 1998.

²⁰ Here is a telling passage:

As I understand the holism that Davidson accepts, it is this thesis: attributions of content to sentences in a community's language, to their linguistic acts, and to their psychological states are systematically interlocked, in such a way that ... there is no explaining, "as from outside" the entire system, what it means to ascribe some specific content to an appropriate item ... Clearly this entails repudiating an aspiration to be anything but modest in theories of meaning. The notion of holding sentences true functions in Davidson's conception of radical interpretation (properly understood) as the key notion in certain judgments, already within the interlocking system, from which an interpreter would find it useful to begin in working his way into the whole. It is not, as in Dummett's picture, material for an account of content—what knits the interlocking system together—"as from outside." (McDowell 1998, p. 103)

McDowell further argues that describing facts about a speaker's holding true of sentences as "evidence" for a meaning theory would be "contrary to the thesis of holism about the interlocking system" (*ibid.*, p. 104).

²¹ Rattan 2004, p. 231.

²² *Ibid.*, p. 225.

If what I have said about Davidson's notion of holding true is right, we have to side with McDowell on the question of modesty. Davidson's account of radical interpretation does not amount to a logical construction of meaning facts out of facts about the holding true of sentences, since in describing an individual as holding sentences of her language to be true, we are presupposing that she knows the truth-conditions of her sentences and has beliefs about the world. This, however, does not mean that we cannot know what sentences a speaker holds true (and under what conditions) without drawing on assumptions about what it is that she knows about the truth-conditions of these sentences and what it is that she believes. We can thus acknowledge and explain the modesty of Davidson's approach without denying that facts about holding true can serve as evidence for a meaning theory.

But this also means that we cannot accept Rattan's formula for characterizing the interaction of Davidson's epistemology and metaphysics of language. For, again, while Davidson's account of radical interpretation is meant to illuminate the nature of linguistic competence, the illumination does not take the form of a logical construction of linguistic competence out the facts that constitute the evidence for radical interpretation.

5.4 The radicalness of deflationism

The argument of Chapter 2 has far-reaching implications for our understanding of the options available to us in the theory of meaning. By exposing the radicalness of deflationism, it shows that a certain familiar way of characterizing the space of available approaches cannot be sustained.

Field claims that we should be "methodological deflationists," that is,

we should start out assuming deflationism as a working hypothesis; we should adhere to it unless and until we find ourselves reconstructing what amounts to the inflationist's relation 'S has the truth-conditions p.' So methodological deflationism is simply a methodological policy, which if pursued could lead to the discovery that deflationism in the original sense ('metaphysical deflationism') is workable or could lead to the discovery that inflationism is inevitable.²³

Though Field himself anticipates that deflationism will ultimately prove workable, there are a couple of hurdles he does acknowledge as serious: the first is a worry about whether deflationism can accommodate the role of truth-conditions in explanations of behavior (and in particular, of success in behavior). The second is a worry about

²³ "Deflationism about Meaning and Content," Field 2001, p. 119.

whether it can make sense of vagueness and various forms of non-factuality in language. A good portion of Field 2003 is devoted to addressing this second worry.

However, what the argument of Chapter 2 suggests is that our discovery of the inevitability of inflationism comes in pretty early, before we have had a chance to raise questions concerning explanation, vagueness, or non-factuality. It is our commitment to the idea that, as a matter of fact, linguistically competent speakers know the truth-conditions of their sentences that forces us to abandon deflationism.

This is not how deflationists generally see things. Field describes the “main idea behind deflationism’ as requiring “that what plays a central role in meaning and content not include truth-conditions (or relations to propositions, where propositions are conceived as encapsulating truth-conditions).”²⁴ He seems to take it as definitional that deflationism involves “keeping truth-conditions (and hence ‘that’ clauses) out of the fundamental characterization of content.”²⁵ At the same time, as we have seen, Field thinks that deflationism explains our knowledge that ‘Snow is white’ is true if and only if snow is white as derived from our competence with a pure disquotational truth locution.

Similarly, Brandom argues that prosententialism, his own brand of deflationism, precludes the notion of truth “from playing an important role in global explanations of meaningfulness in general,” though it allows it to play a role in “local explanations of meaning,” that is, “explanations of the meanings of particular expressions.”²⁶ Thus, according to him, deflationism allows us to say what a sentence means by stating the conditions under which it is true, but it does not allow us to explain what it is for sentences to mean what they do in terms of the notion of truth. As he puts it:

[D]eflationists ought to acknowledge the possibility of *expressing* semantic content truth-conditionally, while denying the possibility of *explaining* semantic content in general truth-conditionally.²⁷

Finally, consider the following passages from Horwich:

A certain philosophical view of truth—known as deflationism—helps to dissolve a certain problem regarding aboutness—the notorious problem of

²⁴ Ibid., p. 108.

²⁵ Ibid., p. 112.

²⁶ Brandom 2005, p. 253.

²⁷ Ibid., p. 256.

intentionality—and thereby puts us in a good position to discern the nature of meaning.²⁸

[W]hile understanding a sentence does indeed usually *coincide* with an explicit knowledge of its truth-condition, understanding does not *consist* in such knowledge. It consists, rather, in appreciating the sentence’s syntactic structure and understanding its constituent words, which, in turn, consists in knowing the basic regularities in their use ... Once ‘Tachyons can travel back in time’ is understood in this way by someone with a conception of truth, then the minimalist account entails that he knows that ‘Tachyons can travel back in time’ is true iff tachyons can travel back in time.²⁹

One alleged advantage of Horwich’s approach to meaning and truth is to make it intelligible how our knowledge of the truth-conditions of our sentences could be manifested in our use of words. For, as Horwich sees it, a (non-truth-conditional) use-based account of meaning, combined with a deflationary account of truth, allows us to explain knowledge of truth-conditions “as the product of a knowledge of meaning ... and a grasp on the concept of truth.”³⁰

Here, then, is the picture that these prominent deflationists want to paint of the space of available options in the theory of meaning. The first option is to construe content truth-conditionally, but this opens up a host of alleged difficulties: for instance, that of explaining how knowledge of meaning or truth-conditions could be manifested in linguistic behavior, or of explaining how facts about our use of sentences determine their truth-conditions. The second option is to deflate our notion of truth, thus leaving it out of the “fundamental characterization” of meaning and content, and explain our knowledge of truth-conditions as a by-product of our linguistic competence together with our competent use of a deflationary truth locution (be it Field’s pure disquotational truth predicate, Brandom’s prosentence forming operator, or a sentential truth predicate definable in terms of Horwich’s minimal notion of propositional truth). Further, in addition to helping explain our knowledge of the truth-conditions of our sentences, our competence with a deflationary truth locution may also allow us to use the notion of truth in “local explanations” of the meanings of our sentences.

If what I have argued in Chapter 2 is right, this picture cannot be sustained. Far from enabling us to explain knowledge of truth-conditions as a by-product of linguistic competence rather than being constitutive of it, deflationism does not even allow us to

²⁸ Horwich 1999, p. 103.

²⁹ Horwich 2005, p. 69.

³⁰ *Ibid.*, p. 70.

describe competent speakers as knowing the truth-conditions of their sentences. Therefore, as far as the theory of meaning is concerned, here are our options: we can reinstate the fundamental role of truth in explaining what it is for our words to mean what they do, or we can deflate truth and forgo the possibility of describing competent speakers in any idiom “more intentional” than the ones deflationists allow in fundamental characterizations of meaningfulness.

Field has suggested that “the division between the inflationist and the deflationist is in some ways the most fundamental division within the theory of content and meaning.”³¹ My arguments reveal this division to be even more fundamental than deflationists are prepared to acknowledge. Or at least, they reveal the *difference* between being a deflationist and being an inflationist to run deeper than deflationists have assumed. To put it in Horwich’s terms, if deflationism does help “dissolve” any problem regarding aboutness, it is not by trivializing aboutness, but by depriving us of the tools needed to ascribe it to our use of words.

³¹ Field 2001, p. 107

Bibliography

- Aristotle (1984). *The Complete Works of Aristotle: The Revised Oxford Translation*. (Edited by Jonathan Barnes, 2 Volume Set). Princeton University Press.
- Boghossian, Paul (1997). "Analyticity." Bob Hale and Crispin Wright, eds., *A Companion to the Philosophy of Language*, Blackwell Publishers, pp. 331-368.
- Brandom, Robert (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.
- Brandom, Robert (2005). "Expressive versus Explanatory Deflationism about Truth." Bradley P. Armour-Garb and JC Beall, eds., *Deflationary Truth*, Open Court, pp. 237-257.
- Davidson, Donald (1984). *Inquiries Into Truth and Interpretation*. Oxford University Press.
- Davidson, Donald (1999). "Reply to Peter Pagin." *Zeglen* 1999: pp. 72-74.
- Davidson, Donald (2001). *Subjective, Intersubjective, Objective*. Oxford University Press.
- Davidson, Donald (2005a). *Truth and Predication*. Belknap Press: Harvard University Press.
- Davidson, Donald (2005b). *Truth, Language and History*. Clarendon Press, Oxford.
- Davies, Martin (1987). "Tacit Knowledge and Semantic Theory: Can a Five per cent Difference Matter?" *Mind*, Vol. 96, No. 384, pp. 441-462.
- Dummett, Michael (1978). "Truth". *Truth and Other Enigmas*. Harvard University Press.
- Dummett, Michael (1991). *The Logical Basis of Metaphysics*. Harvard University Press.
- Dummett, Michael (1993). *The Seas of Language*. Clarendon Press, Oxford.
- Evans, Gareth and John McDowell (1976). "Introduction." Gareth Evans and John McDowell, eds., *Truth and Meaning*, Oxford University Press.
- Evans, Gareth (1981). "Reply: Semantic Theory and Tacit Knowledge." Holtzman and Leich 1981, pp. 118-137.
- Field, Hartry (1986). "The Deflationary Conception of Truth." Graham McDonald and Crispin Wright, eds., *Fact, Science and Value*, Oxford: Blackwell, pp. 55-117.

- Field, Hartry (1992). "Critical Notice: Paul Horwich's *Truth*." *Philosophy of Science*, Vol. 59, No. 2, pp. 321-330.
- Field, Hartry (2000). "A Priority as an Evaluative Notion." Paul Boghossian and Christopher Peacocke, eds., *New Essays on the A Priori*, Oxford University Press.
- Field, Hartry (2001). *Truth and the Absence of Fact*. Oxford University Press.
- Field, Hartry (2005). "Recent Debates About the A Priori." Tamar Szabo and John Hawthorne, eds., *Oxford Studies in Epistemology: Volume 1*, Oxford University Press.
- Foster, David (1976). "Meaning and Truth Theory." Gareth Evans and John McDowell, eds., *Truth and Meaning*, Oxford University Press, pp. 1-32.
- Grandy, Richard (1973). "Reference, Meaning and Belief." *The Journal of Philosophy*, Vol. LXX, pp. 439-52.
- Grover, Dorothy, Joseph Camp and Nuel Belnap (1975). "A Prosentential Theory of Truth." *Philosophical Studies*, 27, pp. 73-125.
- Heck, Richard (2007). "Meaning and Truth-Conditions." Dirk Griemann and Geo Siegwart, eds., *Truth and Speech Acts: Studies in the Philosophy of Language*, New York: Routledge, pp. 349-76.
- Holtzman, Steven H. and Christopher M. Leich, (eds.) (1981). *Wittgenstein: To Follow a Rule*. Routledge.
- Horisk, Claire (2007). "The Expressive Role of Truth in Truth-Conditional Semantics." *Philosophical Quarterly* 57: 229, pp. 535-557.
- Horisk, Claire (2008). "Truth, Meaning, and Circularity." *Philosophical Studies*, 137, pp. 269-300.
- Horwich, Paul (1999). *Meaning*. Oxford University Press.
- Horwich, Paul (2005). *Truth*. Clarendon Press, Oxford. (First published 1990.)
- Kölbel, Max (2001). "Two Dogmas of Davidsonian Semantics." *The Journal of Philosophy*, Vol. XCIII, No. 12, pp. 613-635.
- Kripke, Saul (1982). *Wittgenstein on Rules and Private Language*. Harvard University Press, Cambridge, Massachusetts.
- Larson, Richard, and Gabriel Segal (1995). *Knowledge of Language*. (The MIT Press.)

- Lepore, Ernest and Kirk Ludwig (2003). "Truth and Meaning." Kirk Ludwig, ed., *Donald Davidson*, Cambridge University Press, pp. 35-63.
- Lepore, Ernest and Kirk Ludwig (2005). *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford University Press.
- Lepore, Ernest and Kirk Ludwig (2011). "Truth and Meaning Redux." *Philosophical Studies*, Vol. 154, Issue 2, pp. 251-277.
- Ludwig, Kirk and Georges Ray (2002). "Vagueness and the Sorites Paradox." In James Tomberlin, ed., *Philosophical Perspectives 16: Language and Mind*, Oxford: Blackwell., pp. 419-61.
- Margolis, Eric and Stephen Lawrence (2001). "Boghossian on Analyticity." *Analysis* 61.4, pp. 293-302.
- McDowell, John (1998). *Meaning, Knowledge, and Reality*. Cambridge: Harvard University Press.
- McGee, Vann (2005). "Inscrutability and its Discontents." *Nous* 39:3, pp. 397-425.
- McGrath, Matthew (1997). "Weak Deflationism." *Mind*, Vol. 106, 421, pp. 69-98.
- Pagin, Peter 1999. "Radical Interpretation and Compositional Structure." *Zeglen* 1999, pp. 59-72.
- Patterson, Douglas (2005). "Deflationism and the Truth-conditional Theory of Meaning." *Philosophical Studies*, 124, pp. 271-294.
- Peacocke, Christopher (1979). *Holistic Explanation: Action, Space, Interpretation*. Oxford University Press.
- Quine, W. V. O. (1953). "Two Dogmas of Empiricism." *From a Logical Point of View*. Harvard University Press, Cambridge, MA.
- Quine, W. V. O. (1960). *Word and Object*. The MIT Press, Cambridge, Massachusetts.
- Quine, W. V. O. (1970). *Philosophy of Logic*. Harvard University Press, Cambridge, Massachusetts.
- Quine, W. V. O. (1981). *Theories and Things*. Cambridge, MA: Harvard University Press.
- Ramberg, Bjorn (1989). *Donald Davidson's Philosophy of Language: an Introduction*. Basil Blackwell.

- Ramsey, F. P. (1927). "Facts and Propositions." *Proceedings of the Aristotelian Society*, sup. vol. 7: 153-170.
- Rattan, Gurpreet (2004). "The Theory of Truth in the Theory of Meaning." *European Journal of Philosophy* 12:2, pp. 214-243.
- Rattan, Gurpreet (2005). "Davidson, Semantic Deflationism, and Dummett's Dilemma." *Iyyun* 54: pp. 39-63.
- Rumfitt, Ian (1995). "Truth-conditions and Communication." *Mind*, Vol. 104, No. 416, pp. 827-862.)
- Soames, Scott (1992). "Truth, Meaning, and Understanding." *Philosophical Studies*, 65, pp. 17-35.
- Soames, Scott (2003). "Understanding Deflationism." John Hawthorne and Dean Zimmerman, eds., *Philosophical Perspectives*, 17, *Language and Philosophical Linguistics*, Wiley Blackwell, pp. 369-383.
- Soames, Scott (2008). "Truth and Meaning-In Perspective." *Midwest Studies in Philosophy*, XXXII, pp. 1-19.
- Strawson, P. F. (1971). "Meaning and Truth." In *Logico-Linguistic Papers*, (London: Methuen), pp. 170-189.
- Tarski, Alfred (1944). "The Semantic Conception of Truth." *Philosophy and Phenomenological Research*, 4, pp. 341-375.
- Tarski, Alfred (1956). "The Concept of Truth in Formalized Languages," in *Logic, Semantics, and Metamathematics*. Clarendon Press, Oxford
- Williams, Michael (1999). "Meaning and Deflationary Truth." *The Journal of Philosophy*, Vol. XCVI, No. 11, pp. 545-564.
- Wright, Crispin (1981) "Rule-Following, Objectivity, and the Theory of Meaning." Holtzman and Leich 1981, pp. 99-117.
- Zeglen, Urszula (ed.) (1999). *Donald Davidson: Truth, Meaning, and Knowledge*. Routledge.