NOÛS 36:4 (2002) 525–549

Learning from Others

GARY EBBS
University of Illinois at Urbana-Champaign

I. Introduction

I once believed that there are several naturally occurring isotopes of gold. When asked, "Are there several naturally occurring isotopes of gold?" either I would simply say, "Yes," or, if I was in the mood to be more explicit, I would say, "There are several naturally occurring isotopes of gold." But four years ago, while browsing through the Encyclopedia Britannica article titled "Gold", I read the sentence "The element's only naturally occurring isotope is gold-197." I took the author's words at face value—I took him to have *asserted* that the element's only naturally occurring isotope is gold-197—and I *accepted* that the element's only naturally occurring isotope is gold-197 solely because I trusted him. I thereby took myself to have *learned* from him that the element's only naturally occurring isotope is gold-197. It ordinarily goes without saying that we can learn from others in this way—by taking their words at face value and accepting what they write or say solely because we trust them.

C. A. J. Coady has recently argued[1] that we can *justify* our trust in what others write or say by appealing to Donald Davidson's principle of charity, which exhorts us to interpret another speaker's words in such a way that under the assigned interpretation, what the speaker asserts is true by our own lights. Against this, I will argue that if Davidson's principle of charity is a constraint on correctly interpreting what others write or say, then our ordinary impression that we can learn from others by taking their words at face value and accepting what they write or say is an illusion created by habitual but unjustified misinterpretations of their utterances.[2] The reason is that in a large number of cases in which we take ourselves to be learning from others by accepting what they write or say, what we *take* them to write or say, given (what Davidson sees as) our tacit interpretations of what we read or hear, is *not* true by

our own lights. In such cases, when we take ourselves to be learning from others, we violate Donald Davidson's principle of charity.

We must therefore either reject Davidson's principle of charity, or conclude that what others write or say cannot conflict with what we believe. I propose that we reject Davidson's principle of charity. For reasons I'll explain, this requires that we reject his conception of the problem of interpretation as well. I will sketch a systematic alternative that builds our practice of taking each other's words at face value into our understanding of truth and denotation, and thereby allows for and illuminates the familiar phenomenon of learning from others.

II. A case in which one person learns from another

My discussion of learning from others will focus on the following case[3]: A competent English speaker named Al accepts a number of English sentences that contain the word 'arthritis', including such sentences as "Arthritis can cause pain in one's joints," "Several members of my family have arthritis," and "I have arthritis in my left knee." Al has a pain in his thigh that resembles the arthritic pain in his left knee. He tells his friend Joe, "I have arthritis in my thigh." Joe replies, "You don't have arthritis in your thigh—arthritis afflicts the joints only." "I bet you're wrong, Joe," Al says, "but I'll ask my doctor—I have an appointment with her tomorrow." The next day, Al asks his doctor "Do I have arthritis in my thigh?" She replies "You don't have arthritis in your thigh—arthritis afflicts the joints only." Later that day, Al tells Joe, "I was wrong—the pain in my thigh is not arthritis."

To begin with, let me draw your attention to several obvious but important aspects of this case. First, since Al accepts the sentences "Arthritis can cause pain in one's joints," "Several members of my family have arthritis," and "I have arthritis in my left knee," it is natural to say that Al *believes* that arthritis can cause pain in one's joints, that several members of his family have arthritis, and that he has arthritis in his left knee. Second, it is natural to suppose that since Al, Joe, and the doctor are all at least minimally competent English speakers, they are at least minimally competent in the use of the words 'thigh' and 'joint', so that if they were asked, "Is a thigh a joint?" they would each say, "No, a thigh is not a joint," thereby expressing their belief that a thigh is not a joint. Third, Al takes Joe's and the doctor's utterances of the words "You don't have arthritis in your thigh—arthritis afflicts the joints only" at face value. Speakers of the same natural language typically take each other's words at face value in this way. In doing so, they attribute beliefs to each other and identify points on which they agree and disagree. In this case, Al takes Joe and the doctor to disagree with his belief that he has arthritis in his thigh.

Against this background, two aspects of the case stand out. First, Al does not think he can learn anything from Joe about arthritis, and so when Joe says "You don't have arthritis in your thigh—arthritis afflicts the joints only", Al

continues to believe that he has arthritis in his thigh. In contrast, Al assumes that his doctor knows more than he does about arthritis, and so when the doctor says "You don't have arthritis in your thigh—arthritis afflicts the joints only", Al takes himself to have *learned* from her that arthritis afflicts the joints only. The key observation is that

(1) Al takes his doctor's words at face value and accepts what she says.

Al thereby takes himself to have learned from her that arthritis afflicts the joints only. Since Al believes that a thigh is not a joint, he takes himself to have learned from his doctor that he does not have arthritis in his thigh.

Second, after his conversation with the doctor, Al tells Joe, "I was wrong—the pain in my thigh is not arthritis." Al and Joe remember that the previous day, Al told Joe, "I have arthritis in my thigh." When we hear this story, we naturally assume that after Al takes himself to have learned from his doctor that arthritis afflicts the joints only, he *also* takes at face value his previous utterance of "I have arthritis in my thigh," but rejects it. Hence Al takes for granted that for several days, up until he spoke with his doctor, he believed that the pain in his thigh was arthritis, and that he learned from his doctor that that belief was false. The key observation here is that

(2) After Al talks with his doctor, he takes his previous utterance of "I have arthritis in my thigh" at face value, but rejects it.

Al thereby takes himself to have learned from his doctor that his belief that the pain in his thigh was arthritis was false.

It seems a small step from these observations to the conclusion that by taking his doctor's words at face value and accepting them, Al learns from his doctor. But there is more to learning from another than simply taking her words at face value and accepting what she says. What she says must also be true. From ⌜A learns that p⌝, one can infer ⌜p⌝.[4] In this paper I will largely ignore this further condition and focus on cases in which we take for granted that what a given subject is told is true. I will highlight and explore the consequences of aspects (1) and (2) of Al's situation, which is typical of situations in which we take ourselves to have learned from others that some of our previous beliefs were false. In such cases we take what we are told at face value and accept it, we continue to take our previous utterances at face value, and we reject some of our previous beliefs because they conflict with what we take ourselves to have learned.

These aspects of ordinary cases in which we take ourselves to be learning from others, encapsulated for Al's case by (1) and (2), seem almost too obvious to state. We ordinarily take such observations for granted without any sense that they are in tension with each other, or that they violate any commonsense

constraints on understanding others. It is therefore natural to assume that the conjunction of (1) and (2) is compatible with Davidson's principle of charity.

In fact, however, the conjunction of (1) and (2) *violates* Davidson's principle of charity. To see why, one must first see that (1) and (2) are closely linked to judgments about the denotations of Al's and the doctor's words. In the next three sections I will explain this connection, and sketch my strategy for showing that the conjunction of (1) and (2) violates Davidson's principle of charity.

III. Practical judgments of sameness of denotation

Consider the following pattern for specifying the denotations of one's own words disquotationally:

(D) '_____' denotes *x* if and only if *x* is _____.[5]

Anyone who understands (D) can apply it to his own words. For instance, I accept the results of writing my word 'arthritis' in the blanks of (D), so I accept that my word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis.

Now suppose that I take another English speaker's word 'arthritis' at face value, in the sense that I take her word 'arthritis' to be my word 'arthritis'. Since I accept the results of writing my word 'arthritis' in the blanks of (D), when I take her word 'arthritis' at face value, I in effect take for granted that her word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis. This is *like* a judgment in that I might revise it in light of new information, but *unlike* a judgment in that it is unreflective and may never come up for review. I call it a *practical judgment of sameness of denotation*.

If I take another speaker's word 'arthritis' at face value *while* I am talking with her, I thereby make what I call a practical judgment of sameness of denotation *at a given time*. Often, however, the words I take at face value were written or spoken at some time before my practical judgments of sameness of denotation. At such times, I make what I call practical judgments of sameness of denotation *across time*. A special case of this kind of practical judgment of sameness of denotation occurs when I take my own past uses of a given word at face value. For instance, when I take my own past uses of the word 'arthritis' at face value, I thereby make *practical judgments of sameness of denotation across time* for my word 'arthritis'.

More generally, any speaker who accepts applications of (D) to his own words, and who takes a speaker A's word *w* at face value while he is talking with A, thereby makes a practical judgment of sameness of denotation for A's word *w*. And any speaker who accepts applications of (D) to his own words, and who takes at face value a given word *w* that he used at some previous time *t*, thereby makes a practical judgment of sameness of denotation across time for the word *w* that he used at *t*.

IV. Two conditionals

Let's assume that Al accepts applications of (D) to his own words, so that, in particular, Al accepts the following sentence:

(a) My word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis.

Then we can establish two conditionals, the first with (1) as antecedent, and the second with (2) as antecedent.

To establish the first conditional, suppose that

(1) Al takes his doctor's words at face value and accepts what she says.

This trivially implies that Al takes his doctor's words at face value. Given that Al accepts (a), when he takes his doctor's words at face value he accepts the following sentence

(b) The doctor's word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis.

and thereby makes a practical judgment of sameness of denotation *at a given time* for the doctor's word 'arthritis'. Now, given that Al accepts (a) and (b), we can infer that

(3) Al accepts that the denotation of his word 'arthritis' is the same as the denotation of his doctor's word 'arthritis'.

These reflections show that whether or not we accept (1), if Al accepts applications of (D) to his own words, then

(C1) If (1) then (3).

This is the first conditional.

To establish the second conditional, suppose that

(2) After Al talks with his doctor, he takes his previous utterance of "I have arthritis in my thigh" at face value, but rejects it.

This implies that Al takes his past use of the word 'arthritis' at face value. Given that Al accepts (a), when he takes his past use of the word 'arthritis' at face value, he accepts the following sentence

(c) The word 'arthritis' that I used before talking to the doctor denoted *x* if and only if *x* was (an instance of) arthritis.

and thereby makes a practical judgment of sameness of denotation *across time* for the word 'arthritis' that he used before talking to the doctor. Now, given that Al accepts (a) and (c), we can infer that

(4) Al accepts that the denotation of the word 'arthritis' that he used before talking to the doctor is the same as the denotation of the word 'arthritis' that he uses after talking with her.

These reflections show that whether or not we accept (2), if Al accepts applications of (D) to his own words, then

(C2) If (2) then (4).

This is the second conditional.

V. My strategy

The conjunction of (1) and (2), together with (C1) and (C2), truth functionally implies the conjunction of (3) and (4). I will use this implication to argue that if we accept Davidson's principle of charity, then we must reject the conjunction of (1) and (2). The heart of my argument, which I will present in the next several sections, is that if we accept Davidson's principle of charity, then we must reject the conjunction of (3) and (4). Since the conjunction of (1) and (2), together with (C1) and (C2), truth functionally implies the conjunction of (3) and (4), we may infer that if we accept Davidson's principle of charity, then we must reject the conjunction of (1), (2), (C1), and (C2). That is, we must reject either (1) or (2) or (C1) or (C2). But, as we shall see, (C1) and (C2) are unproblematic consequences of Davidson's methodology of interpretation. Hence if we accept Davidson's principle of charity, we must reject the conjunction of (1) and (2).

VI. What is Davidson's principle of charity?

To understand Davidson's principle of charity, one must see how it fits with his project of constructing empirically testable theories of interpretation for natural languages.

Davidson's project is to try to bridge an assumed gap between data about linguistic behavior, on the one hand, and semantical interpretations of that data, on the other. In practice speakers of the same natural language typically ignore this gap—they typically take each other's words at face value without reflect-

ing about whether they are justified in doing so. But Davidson thinks that if we want to interpret another speaker's words fairly and accurately, it is almost always wrong to take her words at face value. Instead, he thinks, we should suspend our unreflective trust in our practice of taking each other's words at face value, acknowledge the epistemological gap between data about linguistic behavior, on the one hand, and semantical interpretations of that data, on the other, and interpret speakers in accordance with a principle of charity.

To explain how Davidson arrives at this conclusion, I will answer three questions about his approach: (i) What are the philosophical roots of Davidson's project of constructing a theory of interpretation? (ii) What is Davidson's conception of the task and test of a theory of interpretation? (iii) How is Davidson's conception of the task and test of a theory of interpretation related to his principle of charity? The answer to question (i) puts constraints on the answer to question (ii), which in turn puts constraints on the answer to question (iii). I'll address them in order, starting with question (i).

Like many others, Davidson is convinced by W. V. Quine that traditional philosophical attempts to make sense of the idea of meaning are hopelessly obscure. Quine argues that all such attempts make use of notions such as *synonymy* (sameness of meaning), *analyticity* (truth "in virtue of" meaning), and *semantical rule* that are themselves no clearer than the idea of meaning. This would not be a problem if the idea of meaning were clear. But Quine argues that unlike the sentences used in the mature sciences, including logic, psychology, and physics, sentences that contain the term 'meaning' have no explanatory import. For Quine, explanatory import is the criterion of clarity, and so he concludes that the terms 'meaning', 'synonymy', 'analyticity', and 'semantical rule' are hopelessly obscure.

Quine recommends that we replace the obscure idea of meaning with the more tractable idea of translation, understood as a mapping from sentences of one language to sentences of another. In Chapter 2 of *Word and Object*, Quine sketches an empirical theory of translation that makes no essential appeal to the notion of meaning. In Quine's view, a translation of one language into another is acceptable if it preserves the relevant speakers' dispositions to assent to and dissent from sentences under various stimulus conditions.[6]

Like Quine, Davidson proposes that we replace the obscure idea of meaning with an idea that has clearer empirical consequences. But Davidson seeks a theory of interpretation that states the meanings of expressions of a natural language. A translation of one language A into another B comprises a syntactical correlation of the expressions of A with the expressions of B, and so it does not actually state the meanings of the expressions of A.[7] Davidson agrees with Quine that the notions of truth and denotation are clearer than the traditional idea of meaning,[8] he thinks that "... to give truth conditions is a way of giving the meaning of a sentence",[9] and he endorses Quine's suggestion that "... in point of *meaning*... a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined".[10] Combining these

considerations, Davidson claims that the "meaning" of a sentence S of language L may be specified by an empirically tested theory of truth for L.

This brings us to question (ii)—"What is Davidson's conception of the task and test of a theory of interpretation?" Davidson is impressed by Noam Chomsky's approach to constructing an empirically testable theory of the *syntax* of a speaker's language:

> While there is agreement that is it the central task of semantics to give the semantic interpretation (the meaning) of every sentence in the language, nowhere in the linguistic literature will one find, so far as I know, a straightforward account of how a theory performs this task, or how to tell when it has been accomplished. *The contrast with syntax is striking.* The main job of a modest syntax is to characterize meaningfulness (or sentencehood). We may have as much confidence in the correctness of such a characterization as we have in the representativeness of our sample and our ability to say when particular expressions are meaningful (sentences). *What clear and analogous task and test exist for semantics*?[11]

Davidson's answer is that the *task* of a theory of meaning for an infinitary natural language L is to specify the meanings of all the sentences of L on the basis of the meanings of a finite number of simple expressions of L without using the obscure notion of meaning. His conception of the *test* of such a theory is modelled on Chomsky's approach to syntax. Such theories are tested by checking evidence about whether or not speakers regard particular sentences as grammatical. We can clarify the obscure notion of meaning, Davidson argues, if we can propose a conception of meaning and evidence that is structurally similar to the linguists' conception of grammatical theories and the evidence to which such theories must be faithful.

Davidson's bold and ingenious proposal begins with the idea that "to give truth conditions is a way of giving the meaning of a sentence." He proposes that the task of a theory of meaning for a natural language L is to construct a Tarski-style truth theory for L that has a finite number of clauses which together specify the denotations of all the simple expressions of L. Davidson stipulates that "... a theory of meaning for a language L shows 'how the meanings of sentences depend on the meanings of words' if it contains a (recursive) definition of truth-in-L."[12]

This proposed clarification of the task of a theory of meaning suggests a simple test of such a theory. According to Davidson, a Tarski-style truth theory for a natural language L, relativized to persons, times, and circumstances, entails biconditionals of the following general form:

(T) Sentence *s* is true-in-L (speaker *p*'s language at *t*) if and only if conditions *c* obtain at *t*.

In order to test a Tarski-style truth theory for a natural language L, "... all that is needed is the ability to recognize when the required biconditionals are true."

This conception of the test for empirical theories of meaning puts them "on as firm a footing empirically as syntax."[13]

To answer question (iii)—"How is Davidson's conception of the task and test of a theory of interpretation related to his principle of charity?" —we must look more closely at Davidson's account of how an empirical theory of truth is tested. A theory of truth for a German speaker, stated in English, might entail the following biconditional:

(t) 'Es schneit' is true-in-L (speaker *p*'s language at *t*) if and only if it is snowing in the vicinity of *p* at *t*.

But how can we test (t) if we do not (yet) know what the sentence 'Es schneit' means? According to Davidson,

> A good place to begin is with the attitude of holding a sentence true, of accepting it as true. This is, of course, a belief, but it is a single attitude applicable to all sentences, and so does not ask us to be able to make finely discriminated distinctions among beliefs. It is an attitude an interpreter may plausibly be taken to be able to identify before he can interpret, since he may know that a person intends to express a truth in uttering a sentence without having any idea *what* truth.[14]

Suppose that statements of our evidence about which sentences a given speaker holds true have the following general form:

(E) Sentence *s* is held-true-in-L (by speaker *p* at *t*) if and only if conditions *c* obtain at *t*.

Then one bit of evidence might be:

(e) 'Es schneit' is held-true-in-L (by speaker *p* at *t*) if and only if it is snowing in the vicinity of *p* at *t*.

To test a proposed theory of truth for a language L, we must determine whether the biconditionals it entails are true. This is where Davidson's principle of charity comes in: *to determine whether* (t) *is true, we must in effect treat the phrase 'true-in-L' as roughly equivalent to 'held-true-in-L'*. As Davidson puts it:

> I propose that we take the fact that speakers of a language hold a sentence to be true (under observed circumstances) as prima facie evidence that the sentence is true under those circumstances.[15]

According to this proposal, for instance, (e) is evidence for (t).

The key point is that to test a theory of truth for a natural language, we must link our understanding of when the phrase 'true-in-L' applies to a given sentence to our understanding of when the phrase 'held-true-in-L' applies to

that sentence. The link we need is the principle of charity. In Davidson's view, this principle is not optional: the gap that Davidson describes between data about linguistic behavior, on the one hand, and semantical interpretations of that data, on the other, can only be bridged "by assigning truth conditions to alien sentences that make native speakers right when plausibly possible, according, of course, to our own view of what is right".[16] Davidson sometimes recommends that we "maximize" the number of alien sentences that come out true by our own lights, and sometimes that we "optimize" that number. No matter how the principle of charity is formulated, however, its role is to give content to Davidson's leading idea that the consequences of a particular theory of truth for a natural language L can be *tested* against evidence available to an interpreter who does not already know what the sentences of L mean.

VII. Davidson's framework for evaluating (3) and (4)

If we view Al as a Davidsonian interpreter, then (3) and (4) (introduced in section IV) imply that Al accepts particular interpretations of the doctor's words at the time he is learning from her, and of his own words on the previous day, respectively, so (3) and (4) are acceptable or not according as they are permitted by Davidson's principle of charity. This can be seen in three steps.

First, suppose that Al has the resources to construct a Davidsonian theory for another speaker's language. Then he has the resources to construct a disquotational truth theory for his own language, including clauses that specify the denotations of his own words disquotationally. Hence we may assume (as before) that Al accepts applications of (D) to his own words, and, in particular, that Al accepts the sentence

(a) My word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis.

We would ordinarily take Al's word 'arthritis' at face value, and assume that Al's assertion of (a) expresses Al's belief that his word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis. But Davidson thinks that if we are trying to understand Al, we should at first *suspend* our inclination to take Al's word 'arthritis' at face value, and interpret Al in that way only if it is justified by the principle of charity.[17] Even if we suspend our inclination to take Al's word 'arthritis' at face value, however, we know that when Al takes his doctor's words at face value, he accepts

(b) The doctor's word 'arthritis' denotes *x* if and only if *x* is (an instance of) arthritis.

and thereby makes a practical judgment of sameness of denotation at a given time for the doctor's word 'arthritis'. Similarly, we know that when Al takes his previous utterance of "I have arthritis in my thigh" at face value, he accepts

(c) The word ‘arthritis’ that I used before talking to the doctor denoted *x* if and only if *x* was (an instance of) arthritis.

and thereby makes a practical judgment of sameness of denotation across time for his own word ‘arthritis’. From here the simple reasoning presented earlier will take us the rest of the way to (C1) and (C2). In this sense, (C1) and (C2) are unproblematic consequences of Davidson’s approach to interpretation.

Second, suppose Al is constructing a Davidsonian theory of truth for the doctor’s language (idiolect) at the time he is learning from her, or for his own language (idiolect) at some previous time. Then his practical judgments of sameness of denotation for the doctor’s word ‘arthritis’ at the time he is learning from her, and for his own word ‘arthritis’ at some previous time, amount to “base clauses” that assign denotations to, and thereby in effect *interpret*, the doctor’s word ‘arthritis’ at the time he is learning from her, and his own word ‘arthritis’ at that previous time. These are the same practical judgments of sameness of denotation that underlie the conditionals (C1) and (C2), which, together with (1) and (2), imply (3) and (4).

Finally, Davidson thinks his principle of charity governs all interpretation. From Davidson’s perspective, then, (3) and (4) are acceptable or not according as the practical judgments of sameness of denotation that they presuppose are permitted by his principle of charity.

VIII. Why the conjunction of (3) and (4) violates Davidson’s principle of charity

We are now in a position to see that the conjunction of (3) and (4) is *not* permitted by Davidson’s principle of charity. I’ll show (first) that if (4) is permitted, then (3) is not, and (second) that if (3) is permitted, then (4) is not. These two conditionals are truth-functionally equivalent, but it is nevertheless instructive to see the symmetrical arguments that support them.

Suppose that Al is trying to construct a Davidsonian theory of truth for his doctor’s language at the time when she says, “You don’t have arthritis in your thigh—arthritis afflicts the joints only” and that Davidson’s principle of charity does not rule out (4), according to which Al accepts that the denotation of the word ‘arthritis’ that he used before talking to the doctor is the same as the denotation of the word ‘arthritis’ that he uses after talking with her. It’s part of the story that before Al talks with his doctor, Al takes his word ‘arthritis’ to denote the ailment in his thigh. And, as I stressed above, Davidson thinks we can solve the problem of interpretation only “by assigning truth conditions to alien sentences that make native speakers right when plausibly possible, according, of course, to our own view of what is right”.[18] Hence, since Davidson thinks that radical interpretation begins at home, his principle of charity implies that Al must judge denotation and truth by his own lights. Given (4), then, according to Davidson’s principle of charity, Al must take his own

word ‘arthritis’ to denote the ailment in his thigh even after he talks with his doctor.

Now suppose in addition that

(3) Al accepts that the denotation of his word ‘arthritis’ is the same as the denotation of his doctor’s word ‘arthritis’.

Then Al accepts that the doctor’s word ‘arthritis’ denotes the ailment in Al’s thigh. Hence, in effect, Al interprets the doctor’s word ‘arthritis’ in such a way that it denotes the ailment in Al’s thigh. But Al can see from the doctor’s linguistic behavior that she believes that her word ‘arthritis’ does *not* denote the ailment in Al’s thigh. Thus Al interprets the doctor’s words in such a way that under the interpretation, the doctor’s utterance expresses the belief that Al doesn’t have (what Al calls) arthritis in his thigh. This interpretation of the doctor’s utterance is unacceptable, according to Davidson, because it attributes what Al takes to be a false belief to the doctor, and thereby violates the principle of charity.

Davidson sometimes says that the principle of charity permits us to attribute false beliefs to other speakers, as long as the error is, by Davidson’s standards, explicable.[19] In this case, however, the error that Al would be attributing to the doctor if he were to take the doctor’s words at face value is not, by Davidson’s standards, explicable. To see why not, consider the following passage, in which Davidson describes a problem of interpretation that is similar to Al’s:

> If you see a ketch sailing by and your companion says, ‘Look at that handsome yawl’, you may be faced with a problem of interpretation. One natural possibility is that your friend has mistaken a ketch for a yawl, and has formed a false belief. But *if his vision is good and his line of sight favorable it is even more plausible that he does not use the word ‘yawl’ quite as you do, and has made no mistake at all about the position of the jigger on the passing yacht*.[20]

In other words, Davidson thinks that under the circumstances described, if your friend’s vision is good and his line of sight is favorable, you should not take his word ‘yawl’ to denote *x* if and only if *x* is a yawl, since that would be to attribute an inexplicable error to him—the error of believing that the ketch sailing by is a yawl.

You might object to Davidson’s treatment of this example on the grounds that it is natural to assume that your friend intends to speak “correctly”—to apply the word ‘yawl’ in the same way that other competent English speakers apply it. In Davidson’s view, however, your friend’s intention to speak “correctly” in this sense does not settle how his word ‘yawl’ should be interpreted. Davidson would agree that we should interpret speakers in the way that they intend to be interpreted.[21] But according to Davidson, our grasp on how the

speaker intends to be interpreted is ultimately rooted in our evidence of which sentences he holds true. By this criterion, together with the principle of charity, we must conclude that your friend intends to be interpreted in such a way that the ketch sailing by is what he calls a 'yawl'. He may also intend to apply the word 'yawl' in the same way that other competent English speakers apply it, but according to Davidson this metalinguistic intention is less fundamental to our understanding of what he means. As Davidson puts it, "A failed intention to speak 'correctly', unless it foils the intention to be interpreted in a certain way, does not matter to what the speaker means."[22]

Davidson's treatment of his ketch-yawl example reflects his methodological assumption that we can't explain a speaker's error simply by attributing other false beliefs to him, since our attribution of *any* false beliefs to the speaker is precisely what needs explaining. Instead we must appeal to factors that can be checked independently of the interpretation, such as whether or not a speaker's vision is good and his line of sight is favorable. In the ketch-yawl case, by hypothesis, the speaker's vision is good and his line of sight is favorable, so, Davidson suggests, if you were to attribute to your friend the mistaken belief that the passing yacht is a yawl, you would be attributing an inexplicable error to him. The clear implication of the passage is that you should interpret your friend's word 'yawl' in such a way that, unlike your word 'yawl', it denotes the ketch that is sailing by.

By hypothesis, both Al's and the doctor's vision is good and their lines of sight are favorable—they can both see Al's thigh clearly. Hence, by Davidson's standards, Al could not cite these facts to "explain" what (according to Davidson) Al *should* regard as the doctor's error, *if* he interprets her word 'arthritis' to have the same denotation as his word 'arthritis'. Moreover, I don't see how there could be *any* facts about Al's and the doctor's situation that by Davidson's standards Al could cite to "explain" this. I conclude that just as according to Davidson you should interpret your friend's word 'yawl' in such a way that, unlike your word 'yawl', it denotes the ketch that is sailing by, so according to Davidson Al should interpret the doctor's word 'arthritis' in such a way that, unlike Al's word 'arthritis', it does *not* denote the ailment in Al's thigh.[23]

These considerations show that if (4) is permitted by Davidson's principle of charity, then (3) is not. One might think, however, that when Al takes the doctor's words at face value and accepts the sentence "One can't have arthritis in one's thigh", the denotation of Al's word 'arthritis' *changes*, so that *before* Al spoke with his doctor Al's word 'arthritis' denoted the ailment in his thigh, but *after* he spoke with his doctor and accepted the sentence "One can't have arthritis in one's thigh", Al's word 'arthritis' no longer denotes the ailment in his thigh. For instance, one might think that when Al accepts the doctor's assertion, he unwittingly *exchanges* his old word 'arthritis' for a new word that is spelled the same way but has a different denotation. If we accept this description of what happens when Al accepts the doctor's sentence "One can't have

arthritis in one's thigh", then (3) may be permitted by Davidson's principle of charity.

Let's turn now to the second way of formulating the conflict between the conjunction of (3) and (4) and Davidson's principle of charity: if (3) is permitted by Davidson's principle of charity, then (4) is not.

Suppose that Al is trying to construct a Davidsonian theory of truth for the language he used just a few minutes before he spoke with his doctor, and that Davidson's principle of charity does not rule out (3), according to which Al accepts that the denotation of his word 'arthritis' is the same as the denotation of his doctor's word 'arthritis'. Al can see from the doctor's linguistic behavior that she believes that her word 'arthritis' does *not* denote the ailment in Al's thigh. Hence (3) is compatible with Davidson's principle of charity only if Al believes that his own word 'arthritis' does not denote the ailment in his thigh.

Now suppose, in addition, that

(4) Al accepts that the denotation of the word 'arthritis' that he used before talking to the doctor is the same as the denotation of the word 'arthritis' that he uses after talking with her.

Then Al accepts that before he talked to his doctor, his word 'arthritis' did not denote the ailment in his thigh. But Al remembers that he accepted the sentence, "I have arthritis in my thigh", so he can see that he believed that his word 'arthritis' *did* denote the ailment in his thigh. Thus Al interprets his own words in such a way that under the interpretation, his previous utterance of "I have arthritis in my thigh" expressed the belief that he has (what Al now calls) arthritis in his thigh. But when Al said "I have arthritis in my thigh," he did not make any mistakes that we could discern simply by observing the pattern of sentences he held true and using this data to construct a Davidsonian truth theory for his language (idiolect) at that time. By Davidson's standards, then, the mistake that Al attributes to himself—the mistake of believing that he had arthritis in his thigh—is inexplicable. In short, Al's interpretation of his own previous utterance violates the principle of charity. Hence if (3) is permitted by Davidson's principle of charity, then (4) is not.

I conclude that if (4) is permitted by Davidson's principle of charity, then (3) is not, and if (3) is permitted by Davidson's principle of charity, then (4) is not. In other words, the conjunction of (3) and (4) violates Davidson's principle of charity.

IX. My conclusion drawn, generalized, and explained

We can now see that if we accept Davidson's principle of charity, then we must reject the conjunction of (1) and (2). Recall that the conjunction of (1) and (2), together with (C1) and (C2), truth functionally implies the conjunc-

tion of (3) and (4). I have just shown the conjunction of (3) and (4) violates Davidson's principle of charity. We may infer that the conjunction of (1), (2), (C1), and (C2) violates Davidson's principle of charity. But, as we saw in section VII, (C1) and (C2) are unproblematic consequences of Davidson's methodology of interpretation. Hence we must conclude that the conjunction of (1) and (2) violates Davidson's principle of charity.

There is no way of avoiding this conclusion while still embracing Davidson's conception of the task and test of a theory of interpretation. Recall that according to that conception, the *task* of such a theory is to interpret another speaker's words by constructing a truth theory for her language without relying on any evidence about what her words mean, and the *test* is to determine whether the consequences of the theory—the biconditionals it entails—are true. The very idea of such a test builds in Davidson's principle of charity, which links evidence plausibly available to an interpreter—evidence about when speakers hold their sentences true—with particular biconditionals entailed by the theory. When we apply this principle we cannot take for granted that we understand what others are telling us—our entire grip on truth is exhausted by what we currently believe. This implies that we cannot learn from others by taking what they say at face value and accepting it solely because we trust them.

Why have so many philosophers missed or ignored this consequence of Davidson's approach to interpretation? The main reason, I think, is that our practice of taking each other's words at face value is so deeply entrenched that it is largely unresponsive to Davidson's *a priori* criticisms. Even those who have studied and absorbed Davidson's approach to interpretation apparently overlook the extent to which their practice of taking other speakers' words at face value shapes their "interpretations" of other speakers' utterances. Most interpreters who think they are following Davidson's recommendations in fact unwittingly rely on their practice of taking other's words at face value even when it conflicts with Davidson's principle of charity. They take comfort in Davidson's claim that his principle of charity makes room for error, but they do not realize that the errors that they are inclined to attribute to themselves or to others are not in fact compatible with his principle of charity.

There is also a tendency to think that it is *charitable* to take another speaker's words at face value and *trust* what she says. But, as we have seen, if I take another speaker's words at face value and trust what she says, I may violate Davidson's principle of charity. More generally, there is a difference between trust and charity. Unlike trust, charity is something we think of ourselves as exercising only if we take ourselves to be in a position superior in some respects to the position of the person to whom we aim to be charitable. Given Davidson's conception of what a theory of interpretation is and how such a theory can be tested, *we have no choice but to regard ourselves as ultimate authorities on truth*, and to interpret others in such a way that what they say or write is compatible with what we already believe.[24] In many cases, this

is incompatible with taking their words at face value and accepting what they say solely because we trust them.

Davidson points out that in his view "The methodology of interpretation is ... nothing but epistemology seen in the mirror of meaning".[25] We can now see that his methodology of interpretation, and hence his epistemology, is individualistic: it precludes learning from others by taking their words at face value and accepting what they say solely because we trust them.[26]

X. Is the principle of charity optional?

Is this a *criticism* of Davidson's principle of charity? The answer to this question depends on whether or not the principle of charity is optional.

Davidson insists that we have no choice but to accept a principle of charity in interpretation, since otherwise we would not be able to solve the problem of interpretation. He argues that even though we are inclined to trust our practice of taking each other's words at face value, careful reflection about linguistic interpretation shows we should not trust this practice. If that undermines our assumption that we learn from others by taking their words at face value and accepting what they write or say, he reasons, then so much the worse for this assumption, which must be rejected along with other tempting but unfounded assumptions, such as the assumption that synonymy (sameness of meaning) is an objective relation, or that some sentences are analytic (true "in virtue of" their meaning).

I agree that we cannot accept Davidson's conception of the task and the test of a theory of interpretation without also accepting his principle of charity. If Davidson's conception of the task and the test of a theory of interpretation were not optional, then his principle of charity would not be optional either.[27] The question, therefore, is whether Davidson's conception of the task and the test of a theory of interpretation is optional.

Recall that Davidson was driven to this conception by his disenchantment with traditional approaches to meaning, which presuppose such obscure notions as synonymy and analyticity. I agree with Davidson that an adequate philosophical description of our practice of interpreting each other should not make use of such notions. My question, therefore, is whether we can find a way of understanding linguistic interpretation that accommodates our practice of taking each other's words at face value, and thereby makes sense of the familiar phenomenon of learning from others, without relying on such obscure notions as synonymy and analyticity. My answer to this question is "yes".

XI. My proposal

I propose that we accept our practice of taking each other's words at face value as fundamental to our identification of words of our shared language. To accept this proposal is to see our practice of taking each other's words at face value

as part of the data that a description of interpretation must accommodate. Our description of how interpretation proceeds should be in effect a description of what we actually do when we interpret each other. We take each other's words at face value unless we have some reason in a context for not doing so. If we think of ourselves as applying a disquotational truth predicate to other speaker's words, then this amounts to trusting our practical judgments of sameness of denotation unless we have some reason in a given context for not trusting them. As I see it, our shared practice of taking each other's words at face value does not need justification. Instead, in my view, it is local *divergences* from that practice that need justification. But the sort of justification that is needed cannot be derived from an abstract principle of interpretation; it must be based in a description of how we actually proceed when we are trying to understand utterances that puzzle us.

The first thing we typically do is *ask our interlocutor what he or she is saying*. This is only possible if we take for granted some shared vocabulary with which we may discuss the question. Davidson describes the problem of interpretation in such a way that no common vocabulary can be assumed, and so every identification of one person's word with another's must be independently justified by some abstract principle that links linguistic behavior to theory. The key to my alternative is to build some of what Davidson regards as "interpretations" into the data with which we begin when we try to understand others.[28]

To see in more detail how my approach differs from Davidson's, recall first that when we take another speaker's words at face value, we are each in a position to make what I call practical judgments of sameness of denotation. We can each apply the disquotational pattern (D) to our own words. By combining applications of (D) to our own words with our practice of taking other speakers' words at face value, we in effect make practical judgments of sameness of denotation.

This simple observation can help us to describe the arthritis case. If Al accepts the results of writing his word 'arthritis' in the blanks of (D), he can see that when he takes the doctor's word 'arthritis' at face value, he in effect takes for granted that the doctor's word 'arthritis' denotes *x* if and only if *x* is arthritis, and Al thereby makes a practical judgment of sameness of denotation. Similarly, if the doctor accepts the results of writing her word 'arthritis' in the blanks of (D), she can see that when she takes Al's word 'arthritis' at face value, she in effect takes for granted that Al's word 'arthritis' denotes *x* if and only if *x* is arthritis, and the doctor thereby makes a practical judgment of sameness of denotation.

My proposal is that we accept such practical judgments of sameness of denotation unless we have some concrete reason in a given context for revising them. As I see it, we do not and should not begin with the assumption that we can't take another speaker's words at face value unless we can provide some independent justification for doing so. Since we need not begin with this

assumption, our practice of interpreting each other does not commit us to Davidson's principle of charity. If there are reasons in a given context for suspending or rejecting a practical judgment of sameness of denotation, they are not abstract reasons derived from a principle of charity, but concrete reasons rooted in our ordinary practice of interpreting each other.

But what counts as a "concrete reason" for suspending or rejecting a practical judgment of sameness of denotation? I do not believe that there is a correct general account of the reasons that should lead us to suspend or reject our practical judgments of sameness of denotation, so I can only offer examples. Here is one. Suppose John invites his friend Sally over for dinner, and asks her if she has any dietary restrictions. Sally says, "I don't eat meat; anything else would be fine." John serves chicken. Sally, who is fastidious about her diet, objects, saying, "John, I thought I told you that I don't eat meat." John replies, "Chicken is not meat, Sally—it's poultry." Assume that both are completely sincere and initially took each other's words at face value. They consult the Shorter Oxford English Dictionary (SOED), and discover that according to one of the SOED entries for "meat"—"The flesh of animals used as food, now esp. excluding fish and poultry …", John is right. Sally regards this dictionary entry for "meat" as the definition of a word that is different from, even if related to, her word "meat". Both Sally's and John's uses of the word "meat" are established among competent English users; there is good reason to think that the word is ambiguous between a looser use that encompasses chicken, and a more restricted use that does not. In this case, when Sally says, "John, I thought I told you that I don't eat meat," and regards the dictionary definition of "meat" as the definition of a word that is different from her word "meat", it becomes clear to John that he was wrong to take Sally's word "meat" at face value.[29]

In this and other cases in which we suspend or revise a previous practical judgment of sameness of denotation, we presuppose other practical judgments of sameness of denotation. For instance, Sally and John take each other's words "chicken" and "poultry" for granted in their discussion of whether or not poultry is meat. This illustrates the more general phenomenon that speakers of the same natural language typically take each other's words at face value and suspend this practice only locally, when good reasons emerge in discussion. I propose that we make sense of this phenomenon by accepting our practical judgments of sameness of denotation unless we have some concrete reason in a given context for revising them.[30]

If we adopt this proposal, we are free to take the linguistic interaction between Al and his doctor at face value. If we also think Al is entitled to *accept* what his doctor tells him, we can describe Al as having learned that arthritis applies to the joints only, and as being relieved that what he previously believed about his thigh is false. We can also describe the doctor as having corrected Al's false belief about arthritis. We therefore can describe this as a clear case of what I call *learning from others*, which presupposes practical judgments of

sameness of denotation—Al's practical judgment that the doctor's word 'arthritis' has the same denotation as Al's word 'arthritis', and the doctor's practical judgment that Al's word 'arthritis' has the same denotation as her word 'arthritis'.

Davidson's argument against describing the case in this way presupposes his conception of the problem of interpretation, as we have seen. If this conception is not mandatory, as I have suggested, then we can resist Davidson's argument by adopting a different description of our practice of interpreting each other. As I see it, our understanding of truth and denotation is rooted in our practical judgments of sameness of denotation. Divergences from this practice are local, as are the reasons for those divergences. No single abstract principle of interpretation is needed, since particular interpretations always presuppose some background of momentarily unquestioned assumptions, and are therefore never generated completely from scratch.[31]

XII. Radical interpretation and metaphysics

Davidson's conception of the problem of interpretation goes hand in hand with his assumption that we need a fully general philosophical theory of how we interpret others. Such a theory must cover all cases, including cases in which we know nothing about the language we are attempting to interpret. These are cases in which we must engage in what Davidson calls radical interpretation. Since he assumes that all interpretation must proceed by the same principles that we use in radical interpretation, he thinks that all interpretations are ultimately based on evidence that is available to an interpreter who knows nothing about the language he is interpreting. Even when we interpret our fellow English speakers, Davidson believes, we employ the same abstract principles of interpretation that govern radical interpretation.

I propose that we turn this reasoning on its head. Instead of assuming that radical interpretation is the model for all interpretation, I propose that we assume that ordinary interpretation, in which we take for granted a shared vocabulary for raising questions about what other speakers mean, is the model for all interpretation.

How then will I handle the abstract possibility of attempting to understand a language we know nothing about? I propose that we accept a crude principle of charity, not as a guide to interpretation, but as a heuristic for getting into a position in which we can then *start* interpreting speakers in something like the ordinary way: by talking with them and sometimes asking them what they mean by various words and sentences. In my view before we can even get started interpreting a given speaker, we must accept, at least for the moment, some ways of taking her words, so we can ask her what some of her words mean by using some of her other words. This way of understanding interpretation is incompatible with Davidson's, since once we begin talking with the native speakers, we are likely to find counterexamples to Davidson's prin-

ciple of charity—we are likely to find cases like Al's in which individuals express false beliefs but members of their community take their words at face value and evaluate their utterances accordingly.

Davidson assumes that we could in principle interpret a group of speakers without interacting with them at all, simply by observing when they hold their sentences true. In contrast, my account of the epistemology of radical interpretation implies that to understand or interpret other speakers, we must participate in their practice of making assertions, agreeing, disagreeing, adjudicating disputes, and clarifying confusions. As I see it, our practice of interpreting others always presupposes our practical identification of some shared vocabulary or other that we can use to express our agreement or disagreement with other speakers, and to ask them to say what they mean.

Many philosophers assume that the denotations of a person's words supervene on non-semantic facts about how she uses them. This metaphysical assumption goes hand in hand with the assumption that we could in principle interpret a group of speakers without interacting with them at all, simply by observing their linguistic behavior and the non-semantic relationships they bear to their social and physical environments. As I argued in detail elsewhere,[32] however, the standard metaphysical assumption that the denotations of a person's words supervene on independently specifiable facts about how that word is used conflicts with our confidence that we are able to make discoveries. Such confidence is based in our practical judgments of sameness of denotation across time, and our trust in these judgments, even for central and entrenched examples, such as the discovery that gold is the element with atomic number 79, conflicts with the metaphysical assumption. Many philosophers would reject any description of our practice of interpreting each other that conflicts with the metaphysical assumption. But in my view this assumption is not as compelling as my description of how we interpret each other. I therefore recommend that we accept our practice of taking each other's words at face value, trust our practical judgments of sameness of denotation, and reject any metaphysical assumptions that conflict with these aspects of our inquiries.

XIII. Methods and conclusions

I started this paper by describing a typical case in which one speaker learns from another by taking her words at face value and trusting her. I then explained why Davidson's principle of charity is incompatible with our ordinary observations about this typical case. If Davidson is right, then we can't learn from others by taking her words at face value and trusting them. The problem is that according to Davidson's understanding of interpretation, each individual is the ultimate authority on whether or not one of his words denotes a contextually salient object, such as a boat sailing by, or a painful condition in one's thigh.[33] But if we learn from others, then we are not the ultimate

authorities on whether or not one of our words denotes a contextually salient object. There is no way to avoid this consequence without rejecting Davidson's conception of the problem of interpretation.

I argued that this conception is not mandatory—there is an alternative. In particular, I proposed that we accept our ordinary practice of taking each other's words at face value as part of the data that any account of interpretation must accommodate. I proposed that we trust our practical judgments of sameness of denotation unless we find concrete reasons in a particular context for revising them. I then briefly sketched the consequences of this approach for our understanding of learning from others. I argued that if we take this alternative approach to describing our linguistic practices, we can accept our ordinary observations about typical cases, such as Al's, in which we take ourselves to have learned from others.

One might wonder what sort of argument this is. Am I saying that no reasonable person can accept Davidson's conception of the problem of interpretation, given the consequences I have detailed above? Or am I perhaps saying that my alternative conception of interpretation is correct and Davidson's is incorrect? To address these natural questions, I must say something about my philosophical methodology.

As Quine emphasized in §53 of *Word and Object*,[34] mathematicians sometimes replace old vocabulary that they find useful in some ways but problematic or unclear in other ways, by clearer vocabulary that is explicitly designed to capture the uses of the old vocabulary that mattered to them. Set theorists justify their definitions of ordered pair, for instance, by this method. My proposal that we trust our practical judgments of sameness of denotation can be seen as similar to this aspect of mathematical practice. I claim that to see ourselves as engaged in rational inquiry is to accept that we learn from others. To save and clarify this aspect of our epistemological practices, I propose that we build our practice of taking each other's words at face value—our practical identifications of when two words are the same—into our description of the denotations of our words.

I recommend this way of describing our practice because it enables us to make sense of those aspects of learning from others that matter to us when we see ourselves as engaged in rational inquiry. Thus my argument has ultimately a conditional form: if we wish to make sense of this aspect of our linguistic practices, then we should reject Davidson's conception of the problem of interpretation, and adopt a conception of interpretation that does not make learning from others impossible. I have explained how my proposed alternative makes sense of these aspects of our linguistic practices, and to that extent I have supported it.

One might object that Davidson's conception of the problem of interpretation is either correct or incorrect. If it is correct, then any alternative must be incorrect, so the fact that I can *describe* an alternative is irrelevant. But what could it be for a particular conception of the problem of interpretation to be

correct? I don't see how to answer this question, and so I find the objection obscure and unpersuasive.

Whatever the merits of the objection, however, Davidson and his followers are in no position to press it, since they make use of the mathematical methodology I described three paragraphs above. They insist that any aspect of the idea of meaning that isn't captured by their proposed theory is too obscure or unscientific to matter. But this is just to say that all that matters, and all that can genuinely be salvaged from the traditional ways of thinking about meaning, is captured by an empirically testable theory of truth.

I endorse the mathematical methodology, so I do not fault Davidsonians for defending their position in this way. Instead I propose an alternative that builds our practice of taking each other's words at face value into our understanding of truth and denotation, and thereby enables us to make sense of and clarify our ordinary assumption that we learn from others.[35]

Notes

[1] See C. A. J. Coady, *Testimony* (Oxford: Clarendon Press, 1992), chapter 9. For a similar argument, see Tyler Burge, "Content Preservation," *Philosophical Review* 102 (1993) 457–488. Burge argues that our trust in testimony is justified by an a priori principle that is "clearly similar to what is widely called a 'Principle of Charity' for translating or interpreting others." (487) Burge rejects Davidson's assumption that speakers of the same natural language should use the methods of radical interpretation to interpret each other. But Burge suggests that something like a Principle of Charity provides an a priori entitlement to accept what others say as true.

[2] In "Telling and Trusting: Reductionism and Anti-reductionism in the Epistemology of Testimony," *Mind* 104 (1995): 393–411, Elizabeth Fricker also raises doubts about Coady's attempt to use Davidson's principle of charity to justify our reliance on testimony, but not by raising doubts about Coady's assumption that Davidson's principle of charity is a constraint on correctly interpreting what others write or say. Fricker's point is that testimony may be unreliable even if many of our beliefs are true (409–410).

[3] This is an elaboration on the arthritis case that Tyler Burge first presented in his classic paper "Individualism and the Mental", in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, *Midwest Studies in Philosophy*, volume IV (Minneapolis: University of Minnesota Press, 1979), 73–122. I chose to elaborate on Burge's arthritis case, and not to construct a completely new case of my own, for two main reasons. First, many philosophers now accept our initial, commonsense description of Burge's case, even though this description conflicts with some theories of meaning that were once widely accepted. This consensus about how to describe Burge's case aids my argument. Second, even though Burge's arthritis case has been discussed extensively, to my knowledge no one has yet clearly articulated the points I will highlight, or fully appreciated their consequences.

[4] There is also an ordinary sense of "learn" in which a sentence of the form ⌜A learns that p⌝ can be true while the corresponding sentence ⌜p⌝ is false. This sense of "learn" goes with the ordinary sense in which a person may be said to "teach" another to accept a false sentence. For example, a person who accepts Darwin's theory of evolution may say "Zack is teaching Mary (and Mary is learning) that Darwin's theory of evolution is false." I will not be concerned with this sense of "learn" in this paper.

[5] To avoid Grelling's paradox, (D) must be restricted. And I have not said anything about what would motivate us to accept applications of (D). Following W. V. Quine, I think the best

reason for accepting such applications is that they enable us to specify a truth predicate that we can use to state such logical generalizations as 'every sentence of the form 'p ∨ ¬p' is true', 'every sentence of the form '$\forall x(Fx \rightarrow Fx)$' is true', and 'every sentence of the form '$\exists x \forall y Gxy \rightarrow \forall y \exists x Gxy$' is true'. Also like Quine, I favor using a Tarski-style truth predicate defined for regimented sentences of one's own language in terms of restricted applications of a disquotational pattern similar to (D). For more detail on this motivation for accepting a suitably restricted version of (D), see my paper "Truth and Trans-Theoretical Terms," in James Conant and Urszula Zeglen, eds., *Hilary Putnam: on Pragmatism and Realism* (London: Routledge, 2002). I agree with Quine on many technical points, but my view of truth differs fundamentally from his. I reject his behavioristic account of translation, and, for reasons I will partly explain below, I propose that we incorporate our practice of taking each other's words at face value into a disquotational account of truth and denotation.

[6] For a more thorough and accurate account of Quine's views on meaning and translation, see chapter 2 of my book, *Rule-Following and Realism* (Cambridge, Mass.: Harvard University Press, 1997).

[7] Donald Davidson, *Inquiries into Truth and Interpretation* (Oxford: Oxford University Press, 1984), 129.

[8] But note that unlike Davidson, Quine thinks it makes no sense to apply a truth predicate to sentences that we have not translated and do not understand.

[9] Davidson *Inquiries*, 24.

[10] W.V. Quine, "Truth by Convention" (1935), reprinted in W.V. Quine, *The Ways of Paradox*, revised and enlarged edition (Cambridge, Mass.: Harvard University Press, 1979), 77–106; quotation from 82.

[11] Davidson *Inquiries*, 21–22, my emphasis.

[12] Davidson *Inquiries*, 23.

[13] Davidson *Inquiries*, 62.

[14] Davidson *Inquiries*, 135.

[15] Davidson *Inquiries*, 152.

[16] Davidson *Inquiries*, 137.

[17] This is the main reason that I have emphasized (3) and (4), which we (theorists describing Al's situation) can state without taking Al's word 'arthritis' at face value, not the practical judgments that Al would express by affirming sentences (a), (b), and (c), the meanings of which, according to Davidson, we cannot take ourselves to know in advance of interpretation.

[18] Davidson *Inquiries*, 137.

[19] Davidson, *Inquiries*, 136, 153, 168–169.

[20] Davidson *Inquiries*, 196, my emphasis.

[21] In "A Nice Derangement of Epitaphs," published in E. LePore, editor, *Truth and Interpretation* (Oxford: Blackwell, 1986), Davidson emphasizes that "... if the speaker is understood he has been interpreted as he intended to be interpreted." (436). In "A Nice Derangement of Epitaphs" Davidson endorses some of H. P. Grice's views about the relationship between a speaker's intentions and the literal meanings of her words. But Davidson thinks that a speaker's intentions to be interpreted in a certain way cannot have the status that Grice attributes to them. In "Belief and the Basis of Meaning" (reprinted in *Inquiries*, 141–154) Davidson argues that "... making detailed sense of a person's intentions and beliefs cannot be independent of making sense of his utterances. If this is so, then an inventory of a speaker's sophisticated beliefs and intentions cannot be the evidence for the truth of a theory for interpreting his speech behavior." (144)

[22] Davidson, "The Second Person," in Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, *Midwest Studies in Philosophy*, volume XVIII (Notre Dame: University of Notre Dame Press, 1992), 255–267, quotation from 261. Davidson elaborates on this attitude towards "incorrect" usage in "A Nice Derangement of Epitaphs," where he writes that "... error or mistake of this kind, with its associated notion of correct usage, is not philosophically interesting. We want a deeper notion of what words, when spoken in context, mean ..." (434).

[23] One might think that Davidson can avoid this conclusion by claiming that 'arthritis' is a theoretical term. Davidson himself has claimed that "Disagreement about theoretical matters may (in some cases) be more tolerable than disagreement about what is more evident …" (Donald Davidson, *Inquiries*, 169). One might try to use this common-sense observation to argue that Davidson could allow both that Al's word 'arthritis' does not denote the ailment in Al's thigh, and that Al believes that he has arthritis in his thigh.

There are two main problems with this objection. First, given Davidson's conception of the task and test of a theory of interpretation, he has no grounds for thinking that Al's word 'arthritis' is a theoretical term. Second, even if we did have some reason to regard Al's word 'arthritis' as theoretical, that would not show that the error that we would be attributing to him if we were to take his word 'arthritis' to denote *x* if and only if *x* is arthritis is, by Davidson's standards, explicable. Recall that to explain a given false belief of a speaker, according to Davidson, it is not enough simply to attribute other false beliefs to the speaker in light of which her mistake makes sense. Despite Davidson's occasional suggestions to the contrary, his principle of charity apparently implies that it is no easier to accept error among theoretical beliefs than among observational ones.

Simon Evnine thinks that a distinction between theoretical beliefs and observational beliefs can be invoked to defend Davidson against the charge that on his view error is impossible. See Simon Evnine, *Donald Davidson* (Stanford, Calif.: Stanford University, 1991), chapter 6, especially section 6.5. At the crucial point, however, Evnine simply quotes Davidson's commonsense remark about the likelihood of error among our theoretical beliefs, and concludes that Davidson can accommodate error. For the reasons I just sketched, I don't see how Davidson's remark can help him to avoid the consequence that according to his theory of interpretation, Al can't be mistaken about whether he has arthritis in his thigh.

[24] This consequence of Davidson's view—that we have no choice but to regard ourselves as ultimate authorities on truth—does not imply that for Davidson our confident applications of our words *make true* the assertions that we express by using them. My point here is epistemological, not metaphysical: it follows from Davidson's theory of interpretation that we are never justified in interpreting another's assertions in such a way that those assertions conflict with our own beliefs. It may be tempting to try to explain this by saying that our confident applications of our words *make true* the assertions that we express by using them. But even if Davidson's principle of charity implies that I must take myself to be an *authority* on whether not my word 'arthritis' denotes a given object *x*, it does not follow that Davidson is committed to the metaphysical claim that my *confidence* that a given object *x* is an instance of arthritis *makes true* my judgement that *x* is an instance of arthritis. At most the point about authority raises the question of whether Davidson is entitled to the distinction between belief and truth. This is an important and interesting question, but I need not settle it to show that Davidson's principle of charity precludes learning from others. I grant for the sake of argument that Davidson is entitled to the logical distinction between belief and truth.

[25] Davidson *Inquiries*, 169.

[26] This shows what is wrong with C. A. J. Coady's attempt (in *Testimony*, chapter 9) to use Davidson's principle of charity to *justify* our practice of taking ourselves to learn from others. Coady does not realize that Davidson's principle of charity presupposes a radical epistemological individualism.

[27] Thus I disagree with some critics of Davidson's principle of charity, such as Richard Grandy (in "Reference, Meaning, and Belief," *The Journal of Philosophy 70* (1973): 439–452) and David Lewis (in "Radical Interpretation," *Synthese 23* (1974): 331–344), who think they can avoid counterintuitive consequences of that principle by reformulating it slightly, without questioning Davidson's conception of the problem of interpretation.

[28] My approach to interpretation is in some ways similar to the approach Rudolf Carnap recommends in his paper "Meaning and Synonymy and Natural Languages," reprinted in *Meaning and Necessity*, second edition (Chicago: University of Chicago Press, 1956) 233–247. At one point,

Carnap recommends that to determine what a speaker's words mean, we must ask him whether he would apply it in various different possible circumstances—circumstances that we describe by using words of his, the translations of which we take as settled. At the same time, however, Carnap also assumes (wrongly, in my view) that the extensions and intensions of a speaker's words are fixed by her linguistic dispositions.

[29] Sometimes two speakers will disagree about the proper definition of the word without concluding that they are actually using different words with different denotations. This possibility is crucial to the attempt to provide good definitions of words already in use. See chapter 8 of my book *Rule-Following and Realism*, and Tyler Burge, "Intellectual Norms and Foundations of Mind," *The Journal of Philosophy* 83 (1986): 697–720.

[30] Although I think it is typical that speakers take each other's words at face value, under certain unusual circumstances we may come to the conclusion that we should *not* take most of a given English speaker's words at face value. Still, if we arrive at this conclusion, we will have worked our way toward it by asking the speaker questions, taking her words at face value until we find reasons not to take her words at face value.

[31] My approach also allows us to describe Al as having the sort of self-knowledge (or "first-person authority") that goes with minimal competence in the use of language, even though he does not know that arthritis afflicts the joints only. Davidson's conception of the problem of interpretation and principle of charity prevents him from accepting this description of Al. As Davidson points out, his principle of charity implies that "A belief is identified by its location in a pattern of beliefs; it is this pattern that determines the subject matter of the belief, what the belief is about... false beliefs tend to undermine the identification of the subject matter; to undermine, therefore, the validity of the description of the beliefs as being about that subject." (Davidson, *Inquiries*, 168). To take Al to believe that one can't have arthritis in one's thigh would "tend to undermine the validity of the description of the belief as being about that subject," arthritis. In short, as I stressed earlier, the principle of charity implies that each individual is the ultimate authority on whether or not one of his words denotes a contextually salient object, such as a boat sailing by, or a painful condition in one's thigh. According to Davidson, to take Al to believe that one can't have arthritis in one's thigh is in effect to deny that Al is an authority on whether or not his word 'arthritis' denotes the painful condition in his thigh, and thereby to deny that Al has "first-person authority" about what thoughts he expresses when he uses the word 'arthritis'. Once again, however, if we reject Davidson's conception of the task and test of a theory of interpretation, we are not committed to his principle of charity, so we are not committed to his assumption that each individual is the ultimate authority on whether or not one of his words denotes a contextually salient object. We are free to accept that Al has self-knowledge or "first-person authority," even if he believes falsely that he has arthritis in his thigh.

[32] Ebbs, "The Very Idea of Sameness of Extension Across Time," *American Philosophical Quarterly*, Volume 37, Number 3 (July 2000): 245–268.

[33] Recall that this does not imply that for Davidson our confident applications of our words *make true* the assertions that we express by using them. See note 24.

[34] W. V. Quine, *Word and Object* (Cambridge, Mass.: MIT Press, 1960), 257–262.

[35] For helpful comments on previous drafts, I thank members of the audience at my talk for the Philosophy Department at the University of Illinois at Chicago in October 2000, students in my graduate seminar on testimony at the University of Illinois at Urbana-Champaign in Fall 2000, and two anonymous referees for this journal.