

### Handout #3

#### What is Radical Interpretation?

##### *What is the Project of Radical Interpretation*

Language is a social art. (Quine, *Word and Object*, 1960)

That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language. (Davidson, "The Structure and Content of Truth," 1990)

Two questions:

(Q1) What could we know that would enable us to interpret another speaker's utterances?

(Q2) How could we come to know it (on the basis of evidence that does not presuppose any knowledge of it)?

The aim is conceptual illumination: by seeing how the central concepts of a theory of interpretation can be empirically applied on the basis of evidence that does not presuppose already having such a theory.

Like many others, I wanted answers to such questions as "What is meaning?", and became frustrated by the fatuity of the attempts at answers I found in Ogden and Richards, Charles Morris, Skinner and others. So I substituted another question which I thought might be less intractable: What would it suffice an interpreter to know in order to understand a speaker of an alien language, and how could he come to know it? ("The Social Aspect of Language" (1994), p. 126)

[K]

[1] T is a truth theory for L whose axioms are ... and which is interpretive;

[2] Axiom ... of T means that ...; Axiom ... of T means that ...;

[3] ... is a canonical proof procedure for T.

Then an answer to Q2 would be a description of how one could confirm a theory of this form, i.e., come to know that a truth theory was interpretive.

This isn't Davidson's answer because he sought in T&M a substantive constraint that would suffice for a truth theory to be interpretive. He proposed there that a truth theory's being true would be adequate, but that proved to be mistaken. He retained this idea in "Radical Interpretation," namely, that the condition on the truth theory that would enable one to use it to interpret be substantive.

It might seem that there is no chance that if we demand so little of T-sentences, a theory of interpretation will emerge. And of course this would be so if we took the T-sentences in isolation. But the hope is that by putting appropriate formal and empirical restrictions on the theory as a whole, individual T-sentences will in fact serve to yield interpretations. (RI (1973), p. 134)

If we knew that a T-sentence satisfied Tarski's Convention T, we would know that it was true, and we could use it to interpret a sentence because we would know that the right branch of the biconditional translated the sentence to be interpreted. Our present trouble springs from the fact that in radical interpretation we cannot assume that a T-sentence satisfies the translation criterion. What we have been overlooking, however, is that we have supplied an alternative criterion: this criterion is that the totality of T-sentences should (in the sense described above) optimally fit evidence about sentences held true by native speakers. The present idea is that what Tarski assumed outright for each T-sentence can be indirectly elicited by a holistic constraint. If that constraint is adequate, each T-sentence will in fact yield an acceptable interpretation. (RI (1973), p. 139)

A T-sentence of an empirical theory of truth can be used to interpret a sentence, then, provided we also know the theory that entails it, and know that it is a theory that meets the formal and empirical criteria. (p. 139; emphasis added)

[K\*]

- [1] T is a truth theory for L whose axioms are ... and which is confirmable from the standpoint of a radical interpreter;
- [2] Axiom ... of T means that ...; Axiom ... of T means that ...;
- [3] ... is a canonical proof procedure for T.

### *The Standpoint of the Radical Interpreter*

(A) Positive Characterization. What the radical interpreter can appeal to:

- i. knowledge of the speaker's attitudes toward the truth of sentences in his language, such as what sentences the speaker believes true, hopes true, fears true, desires true, as well as the speaker's preference for the truth of one sentence over another;
- ii. knowledge of the speaker's interactions with his environment described in a way that does not entail anything about his attitudes or the meanings of his words; we can in addition, allow the interpreter as much knowledge of the physical world as he likes, as long as it is not described in a way that entails anything about the speaker's attitudes or meanings, though most of it will be irrelevant. Of the knowledge allowed by (ii), most of it will not be relevant to interpretation.

(B) Negative Characterization. What the radical interpreter cannot appeal to:

- i. knowledge of the meanings of any of the expressions of the object language;
- ii. knowledge of "the complex and delicately discriminated intentions with which the sentence is typically uttered" (RI (1973), p. 127), or, more generally, knowledge of any of the speaker's propositional attitudes other than what is allowed in (A).

Everyday linguistic and semantic concepts are part of an intuitive theory for organizing more primitive data, so only confusion can result from treating these concepts and their supposed objects as if they had a life of their own. ("Belief and the Basis of Meaning (1974), p. 143).

The interlocking of the theory of action with interpretation will emerge in another way if we ask how a method of interpretation is tested. In the end, the answer must be that it helps bring order into our understanding of behaviour. But at an intermediate stage, we can see that the attitude of holding true

or accepting as true, as directed towards sentences, must play a central role in giving form to a theory. ("Thought and Talk" (1975), p. 161)

Mental phenomena in general may or may not be private, but the correct interpretation of one person's speech by another must in principle be possible. A speaker's intention that her words be understood in a certain way may of course remain opaque to the most skilled and knowledgeable listener, but what has to do with correct interpretation, meaning, and truth conditions is necessarily based on available evidence. ... language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is "nothing but" observable behavior; but it does imply that meaning is entirely determined by observable behavior, even readily observable behavior. ("The Structure and Content of Truth" (1990), p. 314)

Therefore: ultimately the evidence in A(i) must be justified on the basis of that in A(ii).

### *Reformulating Q2.*

The point of adopting this position as the one from which an interpretation theory is to be confirmed is that in this way we relate the concepts of the theory to the most fundamental evidence we have for applying them.

(Q2') How could we come to know something knowledge of which would suffice to interpret another speaker from the standpoint of the radical interpreter?

### *Problems for Q1*

If the answer to Q2' is a description of the radical interpreter's procedure, then the answer to Q1 should be the result of replacing 'interpretive' in [K] with 'true', rather than [K\*]. But knowledge of this would not suffice for interpretation. If we answer Q2' by appeal to a description of how to confirm a theory is confirmable by a radical interpreter, what more information do we get than that it is true?

A theory that passes the empirical tests is one that in fact can be projected to unobserved and counterfactual cases, and this is apparent to anyone who knows what the evidence is and how it is used to support the theory. ("Reply to Foster" (1976), p. 174)

... when such a theory is treated as an empirical theory, T-sentences are laws which state the truth conditions not only of actual utterances but also of unspoken sentences. Laws formulated as universally quantified biconditionals convey far more than identity of truth value. This consideration, and the constraints that follow from the logical relations among sentences should, when coupled with the usual pressure for simplicity, ensure that contrived, gerrymandered theories are weeded out. ("Reply to Higginbotham (1999), p. 688)

### *Alternative suggestion*

If a truth theory is confirmable from the standpoint of the radical interpreter, its axioms and theorems are lawlike. Hence, replace 'is interpretive' in [K] with 'is lawlike', and treat Q2 as asking about how to confirm a simple true lawlike truth theory.

But we know already being true and lawlike is not adequate. And we need a reason to think being the simplest, if there is a simplest, theory would guarantee it is interpretive.

Appeal to the principle of charity! Surely confirmability by a radical interpreter, who is understood to employ the Principle of Charity (assume the interpretee is mostly right about his environment) can be seen to yield interpretative T-sentences.

First: it is not at all clear it does suffice. Second: it stands in need of justification itself. Third: we still need an independent characterization of the goal in order to ask whether the procedure is adequate to it.

### *Conclusion*

The right way to state what it is that the radical interpreter aims to confirm is expressed by [K], i.e., he is to confirm inter alia a truth theory and that it is interpretive. Without this independent, and uncontroversial, characterization of the target, there is no clear standard to hold the radical interpreter to. As we will see, it will enable us to isolate a fundamental difficulty that emerges from the point of view of the radical interpreter himself.

### *Presupposition of Q2 and Modest and Ambitious Project*

(P) One can come to know something sufficient to interpret a speaker from the evidential position of the radical interpreter.

If we accept (P), then we can argue from what is required for the possibility of radical interpretation to features of his subject matter.

Radical interpretation is possible iff (P) is true.  
If radical interpretation is possible, then p.

We make best sense of Davidson's overall philosophical project if we take him to be engaged in the ambitious project. Many of his conclusions about the nature of thought, language, meaning, truth and knowledge will not be attainable otherwise.

### *Some Contrary Indications*

[i] I do not think I have ever argued for the claim that radical interpretability is a condition of interpretability. [ii] Not only have I never argued that every language is radically interpretable; I have not even argued that every language can be understood by someone other than its employer, since it would be possible to have a private code no one else could break. [iii] I do not think, and have not argued, that radical interpretation of natural languages must be possible; I have argued only that it is possible. [iv] The point of the "epistemic position" of the radical interpreter is not that it exhausts the evidence available to an actual interpreter, but that it arguably provides sufficient evidence for interpretation. ("Radical Interpretation Interpreted" (1994), p. 121; roman numerals added)

Further support:

The problem of interpretation is domestic as well as foreign: it surfaces for speakers of the same language in the form of the question, how can it be determined that the language is the same? ... All understanding of the speech of another involves radical interpretation. ("Radical Interpretation" (1973), p. 125)

The point of the theory is not to describe how we actually interpret, but to speculate on what it is about thought and language that makes them interpretable. If we can tell a story like the official story about how it is possible, we can conclude that the constraints the theory places on the attitude may articulate some of their philosophically significant features" ("Could There be a Science of Rationality?" (1995), p. 128).

... the correct interpretation of one person's speech by another must in principle be possible. ... what has to do with correct interpretation, meaning, and truth conditions is necessarily based on available evidence. ... language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is "nothing but" observable behavior; but it does imply that meaning is entirely determined by observable behavior, even readily observable behavior. That meanings are decipherable is not a matter of luck; public availability is a constitutive aspect of language. ("The Structure and Content of Truth" (1990), p. 314)

### *The Radical Interpreter's Procedure*

#### First Stages

Identify hold true attitudes on the basis of behavior neutrally described, i.e., without presupposing any intentional or semantic facts about the subject. Correlate hold true attitudes which are prompted (indirectly) by what goes on in the subject's environment with conditions that they nomically co-vary with to arrive at (L) sentences:

(L) For all times  $t$ , ceteris paribus,  $s$  holds true  $S$  at  $t$  iff  $p$ .

Ex: For all times  $t$ , ceteris paribus, Pierre holds true 'Il y a un lapin sur la pelouse' iff there is a rabbit on the lawn.

We want to go from (L) sentences to (TF) sentences.

(TF) For all times  $t$ ,  $S$  is true for  $s$  at  $t$  iff  $p$

where we can be confident that this is a target theorem for an interpretive truth theory. Hold true attitudes are the vector sum of meaning and belief: to solve for either, we must hold one fixed.

### *The solution: the principle of Charity*

For present purposes, we can take the Principle of Charity to be defined as what licenses the move from

(L) to treating (TF) as a target for an interpretive truth theory.

Davidson has given three formulations, one of which is his considered formulation, but none of which are adequate.

Ceteris paribus, a speaker's hold true attitudes directed toward occasion sentences are true. (Veracity)  
Ceteris paribus, a speaker's beliefs about his environment are true. (Charity)  
Ceteris paribus, the speaker and interpreter agree about the speaker's environment. (Agreement)

Requiring that hold true attitudes be true is not prima facie adequate because this does not license taking them to be interpretive; at most we get that there is a nomic correlation between the truth of certain sentences and certain conditions.

Requiring agreement on the environment requires a further assumption before we arrive at the conclusion that we've identified conditions that correlate with truth, namely, that the interpreter or the speaker are right about the environment.

The second interpretation, the assumption that the speaker is mostly right about his environment, is the correct interpretation, and it alone make sense of the vector metaphor.

(Charity) For any speaker  $S$ , time  $t$ , belief  $b$ , ceteris paribus:  $b$  is a belief of  $S$ 's at  $t$  about and prompted by  $S$ 's environment iff  $b$  is true.

But Charity falls short:

- (1)  $s$  holds true  $S$  at  $t$  iff (and because)  $s$  believes that  $q$  and  $s$  means that  $q$ .
- (2)  $S$  is true for  $s$  at  $t$  iff  $q$ .
- (Charity):  $s$ 's belief that  $q$  is true at  $t$ .
- (3)  $s$  holds true  $S$  at  $t$  iff  $q$
- (L) Ceteris paribus,  $s$  holds true  $S$  at  $t$  iff  $p$ .
- (4) Ceteris paribus,  $s$  is true for  $S$  at  $t$  iff  $p$ .

But this doesn't suffice to show that what goes in for ' $p$ ' and ' $q$ ' are the same.

(Grace) Ceteris paribus, when we replace ' $p$ ' in (S)

(S)  $s$  believes at  $t$  that  $p$

with a sentence that expresses the content of an environmentally prompted belief of  $S$ 's, the sentence expresses also a condition in  $S$ 's environment that prompts that belief.

Interpretiveness follows with one more assumption:

(I) Ceteris paribus, the L-sentences that the interpreter confirms identify conditions under which a speaker holds true sentences and these conditions are also the prompting conditions, if any, which those beliefs, that are the basis on which the speaker holds those sentences true, are about.

- (1)  $s$  holds true  $S$  at  $t$  iff (and because)  $s$  believes that  $q$  and  $s$  means that  $q$ .
- (2)  $S$  is true for  $s$  at  $t$  iff  $q$ .
- (Grace):  $s$ 's belief that  $q$  is true at  $t$ .
- (3)  $s$  holds true  $S$  at  $t$  iff  $q$
- (L) Ceteris paribus,  $s$  holds true  $S$  at  $t$  iff  $p$ .

- (4) Ceteris paribus,  $s$  is true for  $S$  at  $t$  iff  $p$ .  
& (Grace):  
(5) Ceteris paribus,  $s$  is true for  $S$  at  $t$  iff  $q$ .

*Summary of the Stages of Radical Interpretation*

- 1) Justify L-sentences on the basis of behavior
- 2) Infer (TF) sentences with Grace and (I).
- 3) Formulate a truth theory that has as canonical theorems the TF-sentences arrived at stage (2).
- 4) Derive so far untested TF-sentences and use them to predict the corresponding L-sentences; make assignments to attitudes on that basis and check against predications of behavior.
- 5) Repeat stages (1)-(4), throwing out (L) as needed, until theory and observation coincide.

*Justifying the Principle of Charity*

Options:

- An a posteriori argument by induction
- An argument from the partial holism of attitude content
- An argument to the best explanation
- An argument from the necessity of radical interpretation

By induction, the idea would be to show by examining a speaker's beliefs and his environment independently we could show that by and large they were true. But this is useless in the context of radical interpretation because the interpreter has to assume the Principle of Charity to arrive at assignments of contents to beliefs.

*From Holism:*

We can ... take it as given that most beliefs are correct. The reason is 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. Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false) there must be endless true beliefs about the subject matter. False beliefs tend to undermine the identification of the subject matter; to undermine, therefore the validity of a description of the belief as being about the subject. And so, in turn, false beliefs undermine the claim that a connected belief is false .... It isn't that any one false belief necessarily destroys our ability to identify further beliefs, but that the intelligibility of such identifications must depend on a background of largely unmentioned and unquestioned true beliefs. To put it another way: the more things a believer is right about, the sharper his errors are. Too much mistake simply blurs the focus. What makes interpretation possible, then, is the fact that we can dismiss a priori the chance of massive error. A theory of interpretation cannot be correct that makes a man assent to very many false sentences: it must generally be the case that a sentence is true when a speaker holds it true. So far as it goes, it is in favor of a method of interpretation that it counts a sentence true just when speakers hold it to be true ... according to the method, the speaker holds a sentence true under specified conditions, and these conditions obtain, in the opinion of the interpreter, just when the speaker holds the sentence to be true. ("The Method of Truth in Metaphysics" (1975), pp. 168-69)

We have to distinguish between

- a) General beliefs, such as that the belief that guns are weapons, and
- b) Particular beliefs, such as the belief that this is a gun.

The argument from concept possession yields at best the conclusion that our general beliefs cannot be massively in error.

*To the best explanation:*

- 1) If the assumption that p is required in the best overall theory of the nature of human beings and their place in the natural world, then p.
- 2) The principle of charity is required in the best overall theory of the nature of human beings and their place in the natural world.
- 3) Therefore, the principle of charity is true.

What is the status of the best explanation principle? Is it a priori? If not, it is not help in the present context. But it is hard to see how it could be justified a priori. With respect to the second premise, we must evaluate it in the context of the overall evaluation of Davidson's account.

*From the Possibility of Radical Interpretation:*

- 1) Interpretation from the standpoint of the radical interpreter is possible.
- 2) If interpretation from the standpoint of the radical interpreter is possible, then the principle of charity is true.
- 3) Therefore, the principle of charity is true.

We postpone considerations of what arguments might be offered for premise 1, and remark only that an a priori argument is needed for it for the principle of charity to have a status that makes it usable by the radical interpreter for his project. The argument for the second premise is that some principle is required to get from the evidence to assignments of meanings and the only one that could be plausibly related to the nature of speakers is the Principle of Charity. Thus, a priori support for premise 1 would provide some support for premise 2.

*Additional Constraints*

- 1) The a priori theory of rational agency
- 2) Moderate holism about attitude content
- 3) Shared logic because of the logic of the metalanguage
- 4) Conformity in a linguistic community
  - a. Triangulation
- 5) Empirical Constraints?
  - a. Not available as starting assumptions if the idea is to start with purely behavior evidence
  - b. And not available at all if the object is to investigate what is necessary for meaning and thought as such.