Radical Interpretation, Part II: Philosophical Logic.

Lecture III, *Davidson on Meaning and Radical Interpretation*, 20th November.

Christopher J. Masterman (cm789@cam.ac.uk, christophermasterman.com)

1. *Davidson on Meaning and Interpretation*

1.1. Much of Donald Davidson's work in the philosophy of language has been concerned with a very broad question: 'What is it for words to mean what they do?'. Davidson approaches this question, and correlate questions, in a distinctive way. The opening paragraph of Davidson's "Radical Interpretation" demonstrates:

> Kurt utters the words 'Es regnet' and under the right conditions we know that he said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. What could we know that would enable us to do this? How could we come to know it? (Davidson, 1973: 125)

First, there is the **emphasis on *utterances***. For Davidson, fundamental questions in the philosophy of meaning should be tackled with a great sensitivity to our principal experience of language—everyday communication. Thus, much like Quine, we should not *start* by focusing on questions about, say, words—these are abstractions. Second, there is the **emphasis on *interpretation***. For Davidson, interpretation is simply the process of trying to understand what the utterances of another means.

1.2. Davidson (above) distinguishes two specific questions which are worth keeping apart.

Q1. What could we know that would enable us to interpret another?

Q2. How could we come to know whatever it is which is required for interpretation?

Focus on (Q1.). For Davidson, what we need to know is a theory which models our ability to interpret the utterances of others. For Davidson, such a theory falls under the label of 'Theory of Meaning'. That is, what we need to know to interpret a particular language is a theory of meaning for that language.

1.3. Such a theory should of course be comprehensive—it should be able to interpret, assign meaning, to an arbitrary utterance of the language. For Davidson, theories of meaning have a very particular form. They consist of finitely many axioms and inference rules which together determine an interpretation for every utterance. (Finite because we should be able to *know* the theory and this knowledge should allow us to interpret some language.) How *does* a successful theory of meaning pin down interpretation? Crucially, a theory of meaning for some language $\mathcal{L}$ will have as *theorems* true biconditionals which state, for each sentence of $\mathcal{L}$, what is necessary and sufficient for the sentence to be true, i.e., theorems of the form:

(T) S is true-in-$\mathcal{L}$ if and only if p
(Where, 'S' names a sentence of the target language $\mathcal{L}$ and 'p' is a sentence in some metalanguage.)

So, e.g., the theory of meaning for Norwegian in the metalanguage of English contains as a theorem:

(T′) 'Han elsker edderkoppene' is true-in-Norwegian if and only if he loves the spiders.

1.4. This approach raises a number of questions. For one, why the focus on truth? In particular, why focus on theories which generate theorems of the form (T)? Davidson's approach here combines two ideas.

Truth-Conditional Semantics: Any sentence's meaning is intimately connected with its truth conditions. We can outline what a sentence means by outlining the conditions under which that sentence is true.

- Wittgenstein: "To understand a proposition means to know what is the case if it is true" (TLP4.024).

– Propositions as sets of possible worlds. That is, a proposition—the meaning of a sentence, or what the sentence *expresses*—is the totality of states of the world at which it is true.

Tarskian Definitions of Truth: Loosely, for any language $\mathcal{L}$, a definition of truth-in-$\mathcal{L}$ is adequate if it entails all sentences (in the metalanguage) of the form '*S* is true if and only if *p*'.

It's important to note that there are significant differences between Davidson's approach and the original approach taken by Tarski. First, Tarski is interested only in formal languages. Second, Tarski's explicit that '*p*' is a *translation* of *s* in the metalanguage. Davidson is interested in theories of meaning for natural languages and does not think we can presuppose a notion of translation, and hence synonymy, when addressing foundational questions in the philosophy of language. Instead, we just focus on all *true* sentences of the form (T): "Our outlook inverts Tarski's: we want to achieve an understanding of meaning or translation by assuming a prior grasp on the concept of truth" (Davidson, 1974: 150).

1.5. So, a Davidsonian theory of meaning for $\mathcal{L}$ defines truth in $\mathcal{L}$ and thus defines what the relevant utterances of that language mean. This approach raises a number of worries. A natural worry is whether such theories of meaning can capture *all* of natural language: what about the meaning of questions? commands? orders? promises? intensional contexts? After all, Tarski focuses on formal languages because natural languages were *semantically closed*, and so problematically contain their own truth predicate. We will ignore these questions because there are pressing ones relating particularly to the role Davidson's theories of meaning play in radical interpretation.

2. Constraining Theories of Meaning and Radical Interpretation

2.1. A theory of meaning for $\mathcal{L}$ is supposed to be *interpretive*: by knowing a particular theory of meaning, we are supposed to be able to know what the sentences in $\mathcal{L}$ *mean*. Are there enough constraints on theories of meaning to single out interpretive theories? Some constraints are needed. Contrast:

(T1) 'Schnee ist weiss' is true in German iff snow is white

(T2) 'Schnee ist weiss' is true in German iff grass is green

Simply requiring *true* biconditionals of the form (T) is not enough to distinguish (T1) from (T2).

2.2. Davidson proposes several constraints for adequate theories of meaning. For one, theories of meaning are supposed to deliver *law-like* biconditionals, i.e., biconditionals which support counterfactuals, even though the logic of the theories of meaning is purely extensional. That is, biconditionals like (T1) and (T2), if they are part of an adequate theory of meaning, must hold true even in situations where 'Snow is white' and 'Grass is green' diverge in their truth value, as they might (Davidson, 2005: 54)

2.3. Second, theories of meaning are *holistic*. Theories of meaning shouldn't be judged by assessing their biconditional theorems in isolation, but assessed by how well they fit with the totality of data. A theory of meaning of $\mathcal{L}$ is adequate just in case it entails a true biconditional of the form (T) for any sentence. (It's unlikely that a theory with (T2) will get related sentences right, i.e., 'Das ist weiss', see (Glüer, 2011: 54)

2.4. But most importantly of all, given our present concerns, Davidson thinks that theories of meaning are constrained by both formal and *empirical* constraints. In particular, he held that theories of meaning should be knowable by a r*adical interpreter*—someone who is attempting to interpret some unknown language with only the physical facts available to them. How does this help? For one, since the radical interpreter must be able to come to know the relevant theory of meaning, the theory must be finite and contain axioms governing sub-sentential parts which jointly entail (along with the inference rules) all the true biconditionals.

2.5. For example, the radical interpreter would formulate a theory containing at least axioms (G1)–(G2) and an axiom like (G3) for how such subsentential parts could be combined and how their combinations should be understood, where $\cap$ is for something like concatenation.

(G1) 'Schnee' refers in German to snow

(G2) $\forall x$(x satisfies 'ist weiss' iff x is white)

(G3) '$\alpha \cap$ 'ist' $\cap \beta$' is true in German iff whatever 'α' refers to satisfies the following: 'ist' $\cap \beta$'.

Noteably, (G1)–(G3) jointly entail (T1), but do not entail (T2).

2.6. Of course, the natural worry is that since the data available to the radical interpreter is so slim, it's unclear that they would successfully be able to formulate a theory of meaning, let alone an *interpretive* one. Davidson rejects this and holds that radical interpretation can be done. Importantly, the radical interpretation is guided by two important further constraints.

1. The radical interpreter is able, and ought, to incorporate as part of the physical data a special kind of belief on the basis of the physical data: *the attitude of holding a sentence true (at a time)*.

In general, Davidson does not allow appeal to mental states in radical interpretation, since these are too closely associated with the semantic concepts which the interpreter is attempting to theoretically capture. However, *holding true* is a special kind of mental state: 'an interpreter may plausibly be taken to be able to identify [such a state] before he can interpret'. That is, the interpreter 'may know that a person intends to express a truth in uttering a sentence without having any idea *what* truth'.

2. The radical interpreter should be guided by The Principle of Charity: assign truth conditions to alien sentences that make native speakers right when plausibly possible (Davidson, 1973).

The idea: we should try to hold fixed and find consistency between our own beliefs and the beliefs of the other or, failing that, find consistency between the other's beliefs over time. Why? One idea is that (PC) is a constitutive constraint on what it is to have a set of beliefs, it is essential to what it means to have beliefs.

> If we cannot find a way to interpret the utterances ... of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, or having beliefs, or as saying anything. (Davidson, 1973: 137)

It is not simply a 'rule of thumb to suppress our romantic inclinations' (Hacking, 1975: 147ff)(Hacking, 1975).

3. Successful Radical Interpretation and Indeterminacy

3.1. Where does this leave us? For Davidson, radical interpretation is achievable, since the data available to the radical interpreter includes access to a special kind of belief, *holding true*, and radical interpretation is guided by the non-optional (PC). Like Quine, for Davidson, the data available to the radical interpreter is all the data there is. So interpretation of utterances is secured in a purely naturalistic framework.

3.2. The success of radical interpretation, for Davidson, gives an answer to (Q2.) above: how could we come to *know* whatever it is which is required for interpretation. This is because:

> 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 acts of another involves radical interpretation (Davidson, 1973: 125)

3.3. There is, however, indeterminacy on this picture, particularly indeterminacy at the level of reference. Davidson appeals to Quine's idea of proxy functions, see (Davidson, 1979). That is, supposing everything has a shadow, we could instead formulate the theory of meaning including (G1^S)–(G2^S) as follows:

(G1^S) 'Schnee' refers in German to the **shadow of snow**

(G2^S) $\forall x$(x satisfies 'ist weiss' iff x is **the shadow of something white**)

A theory of meaning which made all the appropriate adjustments to our understanding of subsentential expressions along the lines of (G1^S) and (G2^S) would entail all the same true biconditionals as a theory including (G1) and (G2) instead—snow is white iff the shadow of snow is the shadow of something white.

3.4. For Davidson, this reflects the fact that notions like *reference* are the analogue of theoretical posits in physics. Theories of meaning are naturalist theories of public, linguistic behaviour like utterances. We test them by testing the relevant biconditionals. A part of these theories are semantic notions like reference, but these notions are theoretical posits. It should be unsurprising, on the Davidsonian picture, that the data underdetermines the theory at this level. As Davidson writes:

> Within the theory, the conditions of truth of a sentence are specified by adverting to postulated structure and semantic concepts like that of satisfaction or reference ... The analogy with physics is obvious: we explain macroscopic phenomena by postulating an unobserved fine structure. But the theory is tested at the macroscopic level. Sometimes, to be sure, we are lucky enough to find additional, or more direct, evidence for the originally postulated structure; but this is not essential to the enterprise. I suggest that words, meanings of words, reference, and satisfaction are posits we need to implement a theory of truth. They serve this purposes without needing independent confirmation or empirical basis (Davidson, 1977: 222)

References

Davidson, Donald (1973). Radical Interpretation. *Dialectica* 27, 313–328.

Davidson, Donald (1974). Belief and the Basis of Meaning. *Synthese* 27, 309–323.

Davidson, Donald (1977). Reality Without Reference. *Dialectica* 31, 247–258.

Davidson, Donald (1979). The Inscrutability of Reference. *The Southwestern Journal of Philosophy* 10, 7–19.

Davidson, Donald (2005). *Truth and Predication*. Ed. by Donald Davidson. Cambridge, Mass.

Glüer, Kathrin (2011). *Donald Davidson: A Short Introduction*. New York, US: Oup Usa.

Hacking, Ian (1975). *Why Does Language Matter to Philosophy?* New York: Cambridge University Press.