

Davidson on the theory of meaning, holism and interpretation

(The Mechanics of Meaning, pt. II)

PL. Miraglia

The linguistic triangle

Words $\longleftrightarrow$ **Meaning** $\longleftrightarrow$ **Reference (world)**

It is neither *necessary* nor *sufficient* to grasp the *atomic* meaning of sentence parts to understand the meaning of sentences.

Ramberg:

If meaning lies in the difference something's being true or false make, then it is in the attribution of truth-value that the connection to the non-linguistic must be made. Independently meaningful bits of language are what have empirical impact, and such bits are sentences (or words used as sentences). So it is in sentences being true or false that the connection between semantic concepts and other concepts must be made. [...] Sentences are meaningful only embedded in a larger structure, a language.

(B. Ramberg, *Donald Davidson's Philosophy of Language*, 34)

This is a *holistic* view of meaning and language: there is no independent abstraction at work in the triangle, it is the totality of language that is both the medium and the object of understanding. Davidson believes this captures the true import of Frege's so-called *context principle*:

Frege said that only in the context of a sentence does a word have meaning; in the same vein he might have added that *only in the context of a language does a sentence (and therefore a word) have meaning*.

(Davidson, "Truth and Meaning," *Inquiries*)

The distinctive aspect of Davidson's theory of meaning is therefore that it is a *theory of truth*: we know what someone's linguistic expressions mean if we are able to specify the truth conditions of the sentences s/he believes.

(Against this, other theorists, influenced by Wittgenstein, argue that a theory of meaning must consist principally of a theory of *verification*: to know the meaning of a sentence is to have a notion of what would constitute a method of confirming or proving the sentence.)

Davidsonian interpretation

Donald Davidson was profoundly influenced by Quine. Still, Davidson ultimately built an approach to meaning and linguistic semantics that is quite original. And apparently, late in his career Davidson realized that his theories had brought him close to certain views of Hans-Georg Gadamer, who came from a very different school of thought (hermeneutics):

Four basic agreements stand out: (1) that all understanding is linguistic; (2) that all understanding is interpretive; (3) that the principle of charity is a necessary principle of interpretation; and (4) that we should reject the distinction [between] scheme and content (or [between] worldview and world).

[R. Dostal, "In Gadamer's neighborhood", in: *Dialogues with Davidson*, ed. J. Malpas 2011]

I know very little about Gadamer, but these are genuinely the core insights one finds in Davidson's theory. It's good to keep this summary in mind as one explores these themes in his work.

First, Davidson accepts Quine's indeterminacy thesis in a broad sense but recasts it within a different framework of **radical interpretation**. The difference, in his view, is that the latter puts at the center the goal of semantic **understanding** in a way that translation tends to obscure. One can conceivably obtain a translation manual between two languages in a quasi-mechanical way, by listing pairs of sentences without understanding them.

When interpretation is our aim, a method of translation deals with a wrong topic, a relation between two languages, where what is wanted is an interpretation of one (in another, of course, but that goes without saying since any theory is in some language). [Davidson, "Radical Interpretation", *Inquiries into Truth and Interpretation*, 1985, p. 129]

Subtly, the task of Quine's linguist in the radical translation setting is reframed as that of the interpreter that aims at understanding the community's speakers.

Now, what does understanding require? For Davidson, understanding someone else's actions requires attributing to them **beliefs** and **desires** that would justify their intention to perform that action. And to understand such beliefs also requires understanding the words that can be used in the sentences expressing such beliefs. A belief for Davidson is something that can be formulated in a sentence, not just any utterance. All these factors are weaved together in the project of interpretation:

...the central difficulty is that we cannot hope to attach a sense to the attribution of finely discriminated intentions independently of interpreting speech. The reason is not that we cannot ask necessary questions, but that **interpreting an agent's intentions, his belief and his words are part of a single project, no part of which can be assumed to be complete before the rest is.** ["Radical Interpretation," p. 127]

Again, what Davidson is describing here is the fundamentally **holistic** nature of any solution to the challenge of radical interpretation. Note also that Davidson does not reject Quine's claim that meaning is inherently "indeterminate" (or that there cannot be resolute facts of the matter about meaning), and the above passage contains an indication of the place indeterminacy will have in Davidson's approach.

Charity: a presupposition of holism in interpretation

The key to that will be in the third "basic agreement" cited above: that the "principle of charity" is a necessary principle of interpretation.

Davidson expresses this clearly from his early paper on "Truth and Meaning", 1967:

Just as we must maximize agreement, or risk not making sense of what the alien is talking about, **so we must maximize the self-consistency we attribute to him, on pain of not understanding him.** [...] We do not know what someone means unless we know what he believes [i.e., holds true]; we do not know what someone believes unless we know what he means. ("Truth and Meaning", 27)

Let's delve a bit more deeply into the way in which we can make sense of others' beliefs, thoughts and intentional actions, of course including verbal utterances. In the typical case, we try to explain a person's action by attributing to her an intention to produce a certain outcome, and a belief that by acting in such a way she would, more likely than not, succeed in reaching her goal. If she believes that there is an ice cream shop in front of us and she wants to get some ice cream, that would explain her

stepping into the shop. We would think that the action **makes sense to her**, that is, she would have reason enough to go into the shop (unless she has some other countervailing beliefs or motives). But by the same token, it also **makes sense to us** that this would be sensible for her.

This is part of a shared understanding that is essential for us to say that we have made sense of her action. If, for example, we had – perhaps through a linguistic misunderstanding – come to believe that she dislikes ice cream intensely, then the above explanation of her action would not make sense to us. We could not understand her reason for going into a shop unless we also had grounds to attribute to her other, countervailing beliefs. Otherwise, her action would look to us completely irrational.

And that's the point: at the basis of explanations of this sort – which are part and parcel of the most ordinary “making sense of others” we engage in everyday – there must be a large common ground of shared norms and beliefs, even just to be able to **interpret** others.

But there is an even sharper consequence that emerges here. Suppose that our interpretation model, when applied to real-life speech and conversation, resulted in attributions of systematically **false** beliefs to the speakers we aim at making sense of. Suppose, e.g., (an example by S. Shoemaker) that in our translation the sentence “I had steak last night” was routinely uttered by confirmed vegetarians but very rarely by people we know ate steak last night. Clearly we would have every reason to regard our interpretation as incorrect, everything else being equal.

In other words, **in order to make sense of others' actions, thoughts and speech we must construe our interpretation in such a way that it entails that their beliefs are by and large true**, in addition to being intelligible to us. To put it in another way, the task of radical interpretation presupposes that we are able to maximize agreement so that we can understand each other. Here is a quote from Wittgenstein expressing this thought concisely:

If language is to be a means of communication, there must be agreement not only in definitions, but also (queer as this may sound) in judgments.
[*Philosophical Investigations*, 242]

This is the content of the Principle of Charity according to Davidson: Interpret another **so as to minimize error and irrationality**. (Note that this is not meant as an ethical or deontological norm for interpreters: rather, it's a constitutive principle of the activity of interpreting others without which interpretation doesn't get off the ground, so to speak.)

Propositional attitudes and rational animals

We have seen how often Davidson's approach refers to *beliefs, desires, intentions* (thoughts, more generically): these concepts are all examples of a category of items that linguists and philosophers call "propositional attitudes." They are the kind of things captured by 'that' sentences in English, and use a verb representing the "attitude" a subject has to the sentence:

I **believe** that [Brazil won their match against Norway]
She **feared** that [the cat might be ill]

This type of construct is used, in a word, to describe the content of someone's thoughts, and therefore also enter any explanation of someone's actions, intent, plans etc. One key fact, according to Davidson, is that *"to have one [such attitude] is to have a large complement. One belief requires many beliefs."* ("Rational Animals," 96)

My thesis is rather that a creature cannot have a thought unless it has language. In order to be a thinking, rational creature, the creature must be able to express many thoughts, and above all, be able to interpret the speech and thought of others. (Ibid. 100).

On conceptual schemes and the "very idea" of the distinction between scheme and content

Consider these pairs of "concepts in opposition":

- Spontaneity **versus** Receptivity (Kant)
- Logical or conceptual construction **versus** objects of experience, sense data (empiricism)
- Representational apparatus **versus** the "world itself"

I want to urge that this second dualism of scheme and content, of organizing system and something waiting to be organized, cannot be made intelligible and defensible. It is itself a dogma of empiricism... ("On the very idea of a conceptual scheme," 72)

Granted that making sense of others requires the apparatus of interpretation as described, it still seems possible for someone else to have a radically different **conceptual scheme** in terms of both beliefs and norms of rationality that genuine understanding is precluded. If that is so, it would present a genuine challenge to Davidson's view that attribution of thought and making sense of others are inextricably linked.

A “conceptual scheme” here would be similar to what Davidson views as a “language,” that is, the whole complex of rules, definitions, beliefs and attitudes that can be the object of interpretation – the translation between languages is Davidson’s model for the interpretation of a conceptual scheme (into another). We might then think of radical conceptual differences as being grounded in disagreement about particular important or foundational beliefs.

But both to have beliefs and disagreements – and indeed to have the concept of beliefs – requires to understand the possibility of being mistaken, for Davidson: to believe something is to be ready to make an assertion. And an assertion is the expression of a commitment that something is true, i.e., that the current state of affairs is a certain way and not another. Essentially, Davidson is saying that we could not understand each other as having a disagreement without having an interpretive framework at hand to explain to ourselves what we are in disagreement about. And that we cannot have, by the principle of charity, without being again subject to the constraint that most of the beliefs on the other side of the conceptual fence, so to speak, are true.

If translation succeeds, we have shown that **there is no need to speak of two** conceptual schemes, while if translation fails, **there is no ground for two**. If I am right then, there never can be a situation in which we can intelligibly compare or contrast divergent schemes, and in that case we do better not to say that there is one scheme, as if we understood what it would be like for there to be more.

What Davidson concludes is not that **there is a single, universal scheme** in which all others can be translated, but that the very distinction between schemes is a mirage: it is a radically anti-foundationalist view of both language and thought.

OPEN: How does Davidson’s philosophical framework bear on artificial agents with linguistic capabilities in AI?

A list of readings with links when possible:

Most of Davidson's major papers are collected in book form in a relatively small number of books. Particularly recommended: *Inquiries into Truth and Interpretation*, Oxford UP, 1984.

Many individual papers are now available from public sources. These are directly relevant to my reading:

"Radical Interpretation", <https://www.jstor.org/stable/42968535>

"Rational Animals", <https://www.jstor.org/stable/42969807>

"On the Very Idea of a Conceptual Scheme", in the *Inquiries*, above.

"Thought and Talk", in *Inquiries*.

I found these very useful to bring several themes together:

- W. Taschek, "[Making Sense of Others: Donald Davidson on Interpretation](#)", *Harvard Review of Philosophy*.
- B. Ramberg, *Donald Davidson's Philosophy of Language*, Blackwell 1989.