

The Mechanics of Meaning

Pierluigi Miraglia

Abstract: In this talk I intend to elucidate a certain conception of meaning that emerges primarily from the work of W.V.O. Quine and Donald Davidson, two of the most important American representatives of what is generally termed analytic philosophy. This philosophical current arose from developments of post-Kantian thought in the mid-to-late nineteenth century, and is commonly associated with an approach to philosophical problems that emphasizes linguistic analysis of concepts and empirical methods.

A central concept in this approach is meaning, as an essential feature of both linguistic communication and mental representation, and both Quine and Davidson have made essential contribution to a theory of meaning. I will focus most closely on the work of Davidson, and a family of notions that are core to his understanding of meaning: radical interpretation and conceptual schemes. The interesting thing is that Davidson arrives at a conception of meaning that may seem quite distant from the original positivist methodology of analytical philosophy: meanings, representations and in general intentional content are social and communicative features, not easily reducible to behavioristic or physicalist constructions.

Preamble: Why the turn toward language

I'll start with a potted history of several centuries of philosophy – in a very very small pot. The scale is going to be that of an aerial view from dizzying elevations.

In retrospect, there appear to be two major shifts in the history of Western philosophy, particularly if we look at what has been its central component, the theory of knowledge.

The first separated “modernity” from antiquity, and it brought the focus of philosophy on the role of subjectivity and world-subject interaction. It became explicit with Descartes or perhaps Galileo. At its center is a new construct, the mind and its intellectual faculties, sensory perception, imagination, reason: the primary question is how the mind can acquire knowledge of the (presumably external) world. The divide between the Rationalists (Descartes, Spinoza, etc.) and the Empiricists (Locke, Berkeley, Hume) has a common root in this shift of focus.

While these thinkers arrive at sharply contrasting positions on many or most issues, there is an important area of overlap in how they divide up certain types of intellectual labor: they maintain a distinction between knowledge of truths that derive from purely

rational analysis, such as the truth of logic or mathematics, and knowledge of experiential truths, i.e., truths deriving from the facts and events coming “into the mind”, so to speak, from the external world. The former are presumed to be **necessarily** true, and in a certain sense unquestionable. There is more disagreement about the latter: Empiricists claim that most of our knowledge – and all of our knowledge of the world itself – must be of this latter kind, that is, acquired through experience. Rationalists attempt to expand the scope of truths “of reason” to include more portions of external world knowledge (some of the laws of physics, for instance).

The figure of Immanuel Kant stands at the culmination of this confrontation: his philosophy and particularly his theory of knowledge are often seen as an attempt to cast the debate on a foundation that preserves the role of rational analysis in delivering universal and necessary truths with the productive role of experience in informing us about the world. Kant thought scientific inquiry exemplifies this “optimal” knowledge, which he categorizes as **synthetic a priori**, and saw his mission as that of bringing philosophy itself in line with the rigor of the sciences.

He purported to achieve this by his so-called “Copernican revolution,” an inversion of perspective he describes here:

Thus far it has been assumed that all our cognition must conform to objects. On that presupposition, however, all our attempts to establish something about them a priori by means of concepts through which our cognition would be expanded, have come to nothing. Let us, therefore, try to find out by experiment whether we shall not make better progress in the problems of metaphysics if we assume that objects must conform to our cognition. This assumption already agrees better with the demanded possibility of an a priori cognition of objects—i.e., a cognition that is to ascertain something about them before they are given to us.

[Critique of Pure Reason, pref. Bxvi.]

Practically all philosophical thought after Kant (at least within the confines of the Western tradition) springs from responses to Kant’s position. One major issue emerged not directly from philosophy, but from developments in the sciences themselves. In particular, the discovery of the possibility of non-Euclidean geometries demonstrated that the structure of Euclidean space couldn’t be presumed to be the a priori (universal, necessary) structure of the subjective experience of space: maybe our perception of space does conform to Euclidean three-dimensional space, but that means that our knowledge of geometrical truths **is not a priori**.

Aside from the consequences for Kant’s own theory, the way this discovery had come about – through a rigorous attention to axiomatic systems of geometry, with reliance on rigorous formulation of postulates and proof methods – shifted the locus where one might hope to find necessary truths, or “truths of reason.” Similar developments took

place in other fundamental branches of mathematics, such as arithmetic, the theory of natural numbers and set theory (Gottlob Frege's introduction of modern predicate logic was related to this).

The focus of philosophical epistemology shifted, too, from attempts to individuate concepts and structures within the subject of cognition (the mind) toward the proper way of constructing and defining concepts in linguistic systems – what some have called “the semantic tradition” (A. Coffa, *To the Vienna Station: the semantic tradition from Kant to Carnap*).

To be sure, I am not making the claim that all philosophical projects turned to a focus on language. That would be patently false, as most people would regard Hegel's idealism, Schopenhauer and Nietzsche's “negative” or existentialist thought, Husserl's phenomenology with its ushering of Heidegger (just to make a quick laundry list) as responses, direct or indirect, to the fraying of the Kantian synthesis.

The most prominent representatives of the semantic tradition at the beginning of the 20th century were a group of thinkers, many of whom had started as practicing scientists, clustering around Vienna and the philosophy of Rudolf Carnap and Moritz Schlick (also strongly influenced by the early work of Ludwig Wittgenstein, Frege and Russell, and in contact with the so-called neo-Kantians). They gave rise to the philosophy generally labeled as **logical positivism** or logical empiricism: their “positivism” manifested itself in their view that the methods of empirical and mathematical sciences ought to be the model for a well-founded theory of knowledge; the foundation of epistemology would come from the materials of sense perception and the rest from constructions in terms of the (then relatively new) formal logic.

The catastrophes of continental Europe in the Twenties and Thirties and the subsequent intellectual diaspora brought almost all of these thinkers to Britain or the United States (Carnap at Chicago and UCLA, Grünbaum at Pitt, etc.) and analytic philosophy became predominantly associated with the English-speaking world. W.V.O. Quine, who had studied with Carnap early on in Vienna, became the most prominent analytic philosopher in the U.S., and Donald Davidson was one of his Ph.D. students.

Representations and intentional content

In the rest of my talk the primary focus will be on Davidson's approach to the salient issues about (generally) meaning and semantics introduced by Quine's work. But before diving in, I'd like to introduce a couple of thought experiments made famous by H.

Putnam that give us a sense of the semantic notions we encounter: When does something represent or mean something else? And what is the role of a representational or cognitive agent in meaning something?

Here is the first (the first is a direct quote; the second is my reconstruction; both are taken from Putnam's *Reason Truth and History*, 1981, the first two essays, "Brains in a vat" and "A problem about reference"):

An ant is crawling on a patch of sand. As it crawls, it traces a line in the sand. By pure chance the line that it traces curves and recrosses itself in such a way that it ends up looking like a recognizable caricature of Winston Churchill. Has the ant traced a picture of Winston Churchill, a picture that *depicts* Churchill?

Most people would say, on a little reflection, that it has not. The ant, after all, has never seen Churchill, or even a picture of Churchill, and it had no intention of depicting Churchill. It simply traced a line (and even that was unintentional), a line that we can 'see as' a picture of Churchill.

We can express this by saying that the line is not 'in itself' a representation of anything rather than anything else.

Putnam presents this as the more intuitive of the two hypothetical cases: it highlights the fact that the hypothetical drawing does not have an intrinsic representational quality of pointing to or referring or meaning something else. Even if it "looks like" Churchill, it's not a depiction of Churchill no matter the visual resemblance:

- There is no **causal link** between the presumptive object of representation (W. Churchill) and the ant's tracing.
- The ant does not **intend** to point or refer to, or describe Churchill.

The second thought experiment is often tagged as the "Twin-Earth case" (the label is used by Putnam as well as many other authors who have commented on it). This is my reconstruction of the case (trying to simplify some things):

Imagine a planet just like the one we live in (i.e., Earth), inhabited by people who happen to be near-perfect copies or doppelgänger of the people on Earth. There is someone just like me, a physical molecular copy of myself, on this Twin-Earth; we are exactly alike in our internal states, our beliefs, desires, intentions, feelings, etc. On both Earth and Twin-Earth people who speak English use the word 'water' to refer to a certain substance they reach for in its liquid state when thirsty. Except that on Twin-Earth that substance is different: it is not H₂O but some other compound, call it XYZ. (We may suppose that XYZ looks, smells and tastes to our Twins exactly like water does to us.)

The point of this case is the following: English speakers on Earth (that is, we) refer to water (the substance) when they say 'water'. Twin-Earthers, however, don't have water around at all, but rather water-XYZ. When I speak (or think) of water I refer to, well, water; but when my Twin-Earth copy, call him me*, speaks (or think) using the term 'water', he refers to XYZ.

The moral Putnam wants to drive here is what might be called a certain **externalist** intuition: the reference of our thoughts and other mental states is partly determined by the real-world context of our linguistic practice – the references of our concepts and names – and not just by our internal cognitive content.

Quine and the indeterminacy of meaning

Quine started from a position of rigorous empiricism, as a close student of logical positivism would. His goal was that of analyzing concepts such as linguistic meaning, as much as possible, within the methodology of empirical science. We should stay away, he thought, from metaphysical and psychological presuppositions that aren't supported in fact or logic.

From this point of view, even speaking of "meanings" as an explanatory device in a theory of language use would be highly suspect. It helps to keep in mind that Quine's major philosophical work on language developed around the heyday of behaviorism (in psychology and philosophy), which attempted to construe any talk of mental states and internal representations in empirically verifiable, strictly behavioral terms. Quine is not a behaviorist, however (and Davidson even less so): actually he arrives at a conception that deeply undermines any behavioral theory of meanings.

How should we establish a theory of language use that is rooted in observable evidence and at the same time gives an accurate description of what language speakers mean? Quine's device is to imagine the condition of an hypothetical linguist, dropping in on an isolated community. Members of the community speak a language unknown to him, and devoid of similarities to any language he knows. If he is to understand these people, his task is to build a manual for **radical translation**: that is, to figure out what the community members say with no prior evidence in the way of vocabularies, grammars or such.

Here is Quine's description of the linguist's predicament:

The utterances first and most surely translated in such a case are ones keyed to present events that are conspicuous to the linguist and his informant. A rabbit scurries by, the native says 'Gavagai', and the linguist notes down the sentence 'Rabbit' (or 'Lo, a rabbit') as tentative translation, subject to testing in further cases. The linguist will at first refrain from putting words into his informant's mouth, if only for lack of words to put. When he can, though, the linguist has to supply native sentences for his informant's approval, despite the risk of slanting the data by suggestion.

Otherwise he can do little with native terms that have references in common. For, suppose the native language includes sentences [...] really translatable respectively as 'Animal', 'White', and 'Rabbit'. [...] How then is the linguist to perceive that the native would have been willing to assent to 'Animal' in all the situations where he happened to volunteer 'Rabbit', and in some but perhaps not all of the situations where he happened to volunteer 'White'? Only by taking the initiative and querying combinations of native sentences and stimulus situations so as to narrow down his guesses to his eventual satisfaction.

So we have the linguist asking 'Gavagai?' in each of various stimulatory situations, and noting each time whether the native assents, dissents, or neither. But how is he to recognize native assent and dissent when he sees or hears them? Gestures are not to be taken at face value; the Turks' are nearly the reverse of our own. What he must do is guess from observation and then see how well his guesses work. Thus suppose that in asking 'Gavagai?' and the like, in the conspicuous presence of rabbits and the like, he has elicited the responses 'Evet' and 'Yok' often enough to surmise that they may correspond to 'Yes' and 'No', but has no notion which is which. Then he tries the experiment of echoing the native's own volunteered pronouncements. If thereby he pretty regularly elicits 'Evet' rather than 'Yok', he is encouraged to take 'Evet' as 'Yes'. Also he tries responding with 'Evet' and 'Yok' to the native's remarks; the one that is the more serene in its effect is the better candidate for 'Yes'. However inconclusive these methods, they generate a working hypothesis. If extraordinary difficulties attend all his subsequent steps, the linguist may decide to discard that hypothesis and guess again.

Let us then suppose the linguist has settled on what to treat as native signs of assent and dissent. He is thereupon in a position to accumulate inductive evidence for translating 'Gavagai' as the sentence 'Rabbit'. The general law for which he is assembling instances is roughly that the native will assent to 'Gavagai?' under just those stimulations under which we, if asked, would assent to 'Rabbit?'; and correspondingly for dissent.

(From Quine, *Word and Object*, 1960 pp. 28-29.)

What we see here is the description of the kind of evidence one might collect to corroborate translation "hypotheses," as Quine calls them. It's clear that the basic step must come from atomic observations of stimulus (a rabbit scurries by) and response

('Gavagai!'). One can imagine this sort of translation manual extending to other types of seemingly perceptual reports, such as colors or sounds or noises in the observable environment.

But obviously languages and speakers are not limited to perceptual reports; there are a lot of more conceptual or abstract contents that the sentences of the community's language will convey. A satisfactory translation manual must cover the whole language: but Quine believes this implies that if there is one correct translation for the most basic evidentiary sentences, **then there can be many** which differ in other ways, obtained by modifying in a coordinated way other multiple translation hypotheses that have been advanced along the way. Even for relatively simple perceptual sentences such as 'Gavagai', the linguist may be able to accommodate different translations such as 'lo, a rabbit' or 'lo, attached parts of a rabbit' or 'lo, rabbit fur!', etc.

Quine famously believed this proved one of his more famous theses, that **meaning is necessarily indeterminate**, at least by linguistic evidence. Many have observed that Quine's case is less than a knock-down, conclusive argument (how much less, of course, has been the subject of innumerable papers and theses). But the conception of radical translation had a profound influence on analytic philosophy, to the point that talk of meaning itself came at one time under the same suspicion as talk of metaphysical notion suffered among logical positivists.

Davidsonian interpretation

Donald Davidson was a graduate student with Quine, and was profoundly influenced by him. 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]

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.

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

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. E.g., (an example by S. Shoemaker) in our translation the sentence “I had steak last night” is 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 translation manual 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.)

An objection is often raised at this point that goes like this: 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.

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.

What of the indeterminacy of meaning, which Davidson accepts from Quine? It is also connected to Davidson's argument for the Principle of Charity. The principle entails that an interpretive model for another language (or another conceptual scheme) should maximize agreement over the interpreted beliefs of the other speakers. But there will be different ways of maximizing agreement – or, in other words, different ways of localizing disagreements. This also suggests, according to Davidson, that there is not one correct interpretation across two languages: if there is one coherent way to reconstruct meanings, there will be others.

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 the *INquiries*.

I found this very useful to bring several themes together:

W. Taschek, "[Making Sense of Others: Donald Davidson on Interpretation](#)", *Harvard Review of Philosophy*.