

# Sketch of a theory of meaning

PHIL 93914  
Jeff Speaks

April 17, 2008

Let's call *mentalism* the view that what it is for an expression to have a given meaning is for users of the expression to be in certain propositional attitude states. One way to substantiate the mentalist claim: provide an analysis of meaning in terms of belief.

But: there's no analysis of meaning in terms of intention, and none in terms of belief. It is hard to see how an analysis could be given in terms of some other propositional attitude.

Second try for the mentalist: point out that meaning asymmetrically globally supervenes on the facts about the thoughts of agents.

But, first, there's no easy inference from asymmetric supervenience to priority. We need (at least) asymmetric supervenience + the claim that there's no class *A* of facts such that the *A* facts are independent of or prior to the facts about mental content & mental content supervenes on meaning + the *A* facts. However, it's hard to see what the *A* facts could be, so maybe (depending on how the philosophy of action turns out) the mentalist can assert with some confidence the stronger supervenience claim.

But we have to distinguish two supervenience claims: global supervenience over all times, and global supervenience at a particular time. It looks like (given the 'Pete Rose' example) the first holds, but the second does not.

What does this show us about the relationship between meaning and mental content? For one thing, it seems to distinguish the meaning/mental content case from standard 'nothing over and above' supervenience claims. For example:

The supervenience of the shapes presented on a TV screen on the colors of the pixels.

The supervenience of shape properties on microphysical properties.

The supervenience of moral properties on natural properties.

The supervenience of phenomenal character on intrinsic physical properties.

The supervenience of functional properties on physical properties.

The supervenience of dispositional properties on categorical properties.

The supervenience of (name some class of emergent properties) on (name the class they're supposed to emerge from).

One might of course dispute some of these supervenience claims; but what seems clear is that if these claims hold in the ‘at all times’ sense, they also hold in the ‘at any particular time’ sense.

(Are there any other examples of supervenience claims which hold in the ‘at every time’ sense but not in the ‘at a particular time’ sense? One plausible example seems to be the supervenience of mental content on physical properties — the ‘Pete Rose’ example can also be used to show this, as can lots of examples, esp. using names and kind terms.)

The failure of ‘at any particular time’ supervenience seems to show that we can’t think of the relationship between meaning and mental content as the former being nothing over and above the latter, or some sort of epiphenomenon relative to the latter. I.e., we should not think of the relationship between these two in the way that we might think of any of the above supervenience theses.

The above also shows that there’s no reduction of the meaning facts at  $t$  to the mental content facts at  $t$  since, if there were, the relevant supervenience claim would hold. (McLaughlin on argument by appeal to a FIST.)

So how should we think of it? One possibility is to think of the analysis of meaning in more Kripkean terms, by distinguishing between the conditions under which a term can be introduced with a given meaning and the conditions under which the term can be transmitted with a given meaning. Two arguments for this:

- If Kripke’s view of reference is correct, then, given that meaning determines reference, it follows that his distinction between introduction and transmission should also have relevance for views about the foundations of meaning.
- Facts about meaning introduction *do* supervene in the ‘at any particular time’ sense on the facts about mental content.

This then leads to two questions: what does it take to introduce a term with a given meaning? What does it take for that term to persist with that meaning over time?

On the question of term introduction: I suggest a two-pronged approach:

(1) In the case of many terms, a ‘baptism’ type theory seems appropriate. This seems particularly apt if the content of the term is the sort of content than can be a part of the content of a perceptual experience. It is controversial what sorts of contents are in this category: maybe external particulars; ‘sensible’ properties like colors, shapes, pitches, textures; natural kinds.

(2) This won’t cover every case. In the cases of terms with non-perceptible contents, a promising idea is that a Horwich-style use theory might be appropriate. To introduce a term is to adopt a certain explanatorily basic disposition to use that term in certain ways. E.g., to introduce a truth predicate is to be disposed to provisionally accept every instance of the relevant T-schema.

A Horwich-ian reply: the above argument against a unified theory of meaning only applied to mentalist theories. Why not simply adopt a Horwich-style use theory of meaning, and skip the distinction between introduction and transmission?

Two arguments:

- (a) It's not clear how the Horwich account is supposed to apply to names.
- (b) The same sort of FIST argument can be give here: the meaning facts at  $t$  do not globally supervene on the Horwichian use facts at  $t$ , and so can't be analyzed in terms of them. It's actually easier to argue this than to make the corresponding argument against the mentalist.

Horwichian reply: but what about deference? We could still say that all non-deferential users of terms satisfy the Horwich conditions, and that the others mean what they do by their words in virtue of their dispositions to defer to others who do satisfy those conditions.

Let's distinguish two different views of 'deference': on one trivial view of deference, talk about deference is just talk about using a term with a given meaning in virtue of the term's already possessing that meaning. I agree that the users who do not satisfy the Horwich conditions are deferential users in this sense; but this is just another way of saying that terms can be transmitted from user to user with a given meaning, even if neither user satisfies the conditions on introducing the term with that meaning.

On a second, more substantial view of deference, *deferring* is supposed to explain how a word can come to have a given meaning out of my mouth without appeal to unexplained facts about the meanings of words. So, for example, we might say that for every deferential user  $D$  of an expression  $x$ , there exists some expert  $E$  such that  $D$  stands in the deferring relation to  $D$  with respect to  $e$ , and  $E$  satisfies the Horwich conditions on meaning such-and-such by  $x$ .

The question can then be put this way: what is the deferring relation?

- Being disposed to accept correction from  $E$ . *But the experts might no longer exist, as in the case of Pete Rose.*
- Intending to use the word with the same meaning as  $E$ . *But  $D$  might have no de re intentions with respect to  $E$  at all – he might not know who  $E$  is.*
- Intending to use the word with the same meaning as whoever is a competent user of the term. *But this sort of thing is not sufficient for using a word with the same meaning as such a competent user. You might when learning a foreign language form such an intention — this wouldn't make you competent.*

One might then try something like this: it is to be a member of the same linguistic community (extended over time) as some person who did satisfy conditions sufficient for introducing the term with that meaning, and to be connected to them by a chain of meaning transmissions. But of course at this point we've lapsed back into the trivial view.

What conditions must be satisfied for a term to be passed from one language user to another with the same meaning? Clearly, they needn't satisfy the conditions for introducing the term with that meaning. So what conditions must they satisfy? I suggest that the only conditions are negative ones. The default case is that the term is transmitted

with the same meaning; only if the person acquiring the term satisfies the conditions for introducing the term with another meaning is it the case that meaning transmission fails to go as usual.

Two examples: color term transmitted to someone blind, and color term transmitted to someone spectrum inverted relative to you. In the first case, the person does not satisfy the conditions for re-introducing the term with a new meaning, and hence uses the term with the same meaning as you do. The spectrum invert, however, does use the name as a name for a particular color other than the one you use it for, and so the meaning is not transmitted in the usual way.

A hard case: Evans' example of Madagascar. Does this fit the above view?

The vexing puzzle of semantically significant abbreviations. Suppose we introduce a term via a Horwichian acceptance regularity, and suppose that it is a regularity of this form: we are disposed to accept 'that is  $F$  iff that is such-and-such'. Horwich's condition for 'water' is of this sort. The puzzle is: typically (as in the 'water' case we will not want to say that ' $F$ ' is synonymous with 'such-and-such.' But how can it fail to be? It just seems as though we are adopting a convenient abbreviation of 'such-and-such' – and how can abbreviations be semantically significant? If this is a genuine problem, this seems to be a serious limitation on the utility of Horwich-style term introductions, which leaves us no model of how terms with non-perceptual contents can be introduced. The worry is that every expression in the language would either have as its content something which can be a part of the content of perception, or be analyzable as a combination of such contents. But this does not look plausible. Possible reply: maybe the relevant acceptance conditions should be of some other form.

A stronger conjecture: is every content which you're capable of having thoughts involving one which either was a part of the content of your experience or is the content of some term with which you are competent?

Concluding question: on a view like this, is mental content prior to meaning? In one sense, yes, and in one sense, no.

Yes, because linguistic meaning is in large part explained in terms of perceptual representation, and perceptual representation is a kind of mental representation.

No, in two senses. If we think about a particular individual, and ask what makes them capable of having thoughts involving a given content, then answer will often be that the route is linguistic. And if we ask of everyone existing at  $t$  how they are able to have thoughts involving that content, the answer will also often be given in terms of the meaning of some natural language expression.

Further, the above sort of view leaves open an analysis of thought in terms of perception + language. This would be a kind of mixed view, on which perceptual experiences are prior to linguistic meaning, but both are prior to, e.g., facts about the contents of beliefs and judgements.