1 Inferential Roles and Semantics

If semantics is the study of meanings, and the meaning of an expression \( e \) is understood as what speakers of the language must know use \( e \) competently, then a natural first guess is that the correct semantics of names will involve their referents. One might, initially, make the further supposition that the meaning of a name is just its referent. However, this sort of ‘Fido’-Fido view immediately runs into the difficulties posed by Frege: although the referent(s) of the names in ‘Hesperus is Hesperus’ is (are) identical to the referent(s) of the names in ‘Hesperus is Phosphorus’ (assuming the latter is true), it is easy to imagine that the latter is informative, while it is difficult to imagine the same about the former. If we suppose, plausibly, that the difference in informativeness between the two comes from some aspect of the meaning of the names in the two sentences, then it is clear that a semantic theory that appeals only to reference lacks the resources to accommodate the difference Frege found. Frege’s well-known solution to this puzzle was to introduce the notion of sense into his semantics: even though ‘Hesperus’ and ‘Phosphorus’ share a reference, their failure to share a sense will explain how the true identity statement ‘Hesperus is Phosphorus’ can be informative. For present purposes, it will be sufficient to note the following moral of Frege’s proposal: to cope with Frege’s problem, semanticists could no longer understand meaning solely in terms of reference, but now had to invoke a psychological component of meaning to explain how expressions figure in our cognitive economies.\(^1\)

IRS theories are proposals for how this cognitive component of meaning should be understood. Roughly, according to such theories, the cognitive meaning of an expression is identified as the role that that expression plays in a lan-

\(^1\)Putnamian Twin-Earth puzzles give further reason for thinking that meaning doesn’t collapse onto reference, but must involve a cognitive component as well. However, here the failure is in the opposite direction: Frege puzzles show that identity of reference is compatible with difference of cognitive meaning, while Putnam puzzles show that identity of cognitive meaning fails to ensure identity of reference. Putnam and Burge (among others) have convinced many that reference can’t be explained without adverting to things outside the head of the speaker; hence, what an expression refers to (/denotes) is sometimes called its wide (/broad) meaning. In contrast, the cognitive meaning of an expression can be fixed by what’s in the head of the speaker, and is often called its narrow meaning.
guage, theory, or system of thought. There are several flavors of these theories. For example, the conceptual role of an expression is explained by one proponent of IRS as “the causal role of the expression in reasoning and deliberation and, in general, the way the expression combines and interacts with other expressions so as to mediate between sensory inputs and behavioral outputs. A crucial component of a sentence’s conceptual role is a matter of how it participates in inductive and deductive inferences. A word’s conceptual role is a matter of its contribution to the role of sentences” ([Block, 1986], 93). Another conceptual role theory, developed in [Field, 1977], characterizes conceptual role in terms of a probability function defined over all the sentences of a person’s language that specifies her commitments concerning how she will change her degrees of belief upon discovering new information. Inferential role theories define cognitive meaning in terms of the roles of expressions in inferences. Other versions of these theories have been proposed by Loar, Schiffer, McGinn, Harman, Lycan, Fodor, Sellars, and others. Obviously, there are many interesting points of difference between these alternative formulations, but I cannot compare them here. However, we can notice that all these theories are holistic in the sense that they specify the cognitive meaning of each expression in terms of its role in some larger network. And since the nodes in this network are individuated by their cognitive meanings (possibly *inter alia*), IRS theories specify the cognitive meaning of a given expression in the language (/system of thought) only by adverting to the cognitive meanings of all other expressions in the language (/system of thought).

2 Advantages of IRS

IRS theories have been claimed to have many advantages. Although I can’t discuss them all here, it will be useful to rehearse some of the more important advantages to show that IRS theories would be quite useful if they could be sustained.

First, it is claimed, IRS theories offer a solution to Frege’s puzzle that is psychologically better motivated and less ontologically extravagant than Frege’s own solution. To wit, since a thinker could maintain distinct conceptual roles

---

2 Many advocates of IRS regard the theory as a semantics for mental expressions in a language of thought, not expressions of natural language. For some such theorists, the picture is supposed to be that for every expression of natural language there is a corresponding expression of Mentalese, and that the former inherits its semantics from the latter, while the latter has a semantics given by some flavor of IRS. For others, IRS offers a picture of the semantics of thought as well as a separate semantics for language. I can’t examine the plausibility of these pictures here. I will, however, switch back and forth between talking about IRS as a semantics for spoken languages and as a semantics for Mentalese. Nothing I will say about these matters will hang on the difference.

3 Below I shall consider the suggestion that the holism of IRS theories can be delimited.

4 Psychological motivation for the IRS solution comes from the intimate connections between IRS theories and both stereotype theories of conceptual structure in cognitive psychology and procedural semantics in artificial intelligence. In contrast, it’s far from obvious how a Fregean notion of sense is to be understood in psychological terms. The IRS solution is
for the expressions ‘Hesperus’ and ‘Phosphorus’ (despite their referring to the same astronomical body), she might well find it informative to know that the referents of the two expressions coincide. In other words, if the conceptual roles of the two expressions differ, it’s easy to see how a notion of meaning composed of both conceptual role and referential factors would contain the resources to explain the informativeness of the (true) identity statement ‘Hesperus is Phosphorus’.

Second, the cognitive aspect of meaning specified by an IRS seems roughly the right sort of thing to supply psychological explanations (but see §3.3). For example, conceptual role offers a plausible explanation for the difference between the beliefs I express by uttering the sentences ‘I am on fire’, and ‘Jonathan Cohen is on fire’. Clearly, referential aspects of meaning are, by themselves, impotent to distinguish between what is expressed by my utterances of these two sentences, while conceptual role, tied as it is to the role of each expression within my language/system of thought, seems to be perfectly suited to make this distinction. Moreover, the difference in what is expressed by the two sentences above can make for significant causal differences in action (cf., [Perry, 1977], [Perry, 1979], and ([Kaplan, 1989], §XVII)). Therefore, the differences made available by an IRS theory would seem to be causally relevant, and hence relevant to causal explanation in psychological terms as well.

Third, IRS theories offer a simple conception of (inter- and intra-linguistic) sameness and difference of meaning in terms of sameness and difference of conceptual role. So for example, the reason that ‘brilliant’ means the same as ‘fabulous’ to speakers of British English is that the two expressions share a conceptual role for speakers of British English (or so my informants tell me). The same pair of expressions do not share a conceptual role for speakers of American English, so the theory makes the correct prediction that such speakers will not find the two to have the same meaning. The same idea makes available a notion of translation as a process of finding expressions of the target language whose (references plus) conceptual roles match those of source language expressions. Presumably this picture sameness of meaning could also be used to ground understandings of analyticity, language understanding, and other useful semantic notions along similar lines.

For these reasons, IRS theories have much to offer; unfortunately, they suffer from serious drawbacks as well. It is to these drawbacks that I now turn.

3 Drawbacks of IRS Theories

I’ll discuss three kinds of worries about IRS theories: those connected with holism, accusations that IRS theories cannot explain the compositionality of meaning, and questions concerning IRS theories and interpretation.

claimed to be more ontologically parsimonious than Frege’s solution because the ontology of inferential roles seems more compatible with a naturalistic metaphysics than does Frege’s exotic third realm.

But see section 3.1, where I’ll be raising problems for the notion of same conceptual role.
3.1 Holism

As mentioned, IRS theories specify the cognitive meaning of a given expression in a language/system of thought only in terms of the system of cognitive meanings of all the other expressions in that language/system of thought. Hence, it is impossible to specify the cognitive meaning of a single expression without adverting to the cognitive meanings of all the other expressions. A consequence of this commitment is that a change in the cognitive meaning of a single expression will change the cognitive meanings of every other expression in the language/system of thought. But this extreme sensitivity of the whole structure to small changes in the meanings of any single expression in the entire system makes it highly unlikely that any two expressions ever have the same cognitive meaning for distinct individuals at a given time or for a single individual at distinct times. In other words, the holism entailed by IRS makes it unlikely that there is ever interpersonal or intrapersonal sharing of cognitive meaning.

This result apparently brings with it some unpleasant consequences. For example, many philosophers have thought that without a coherent notion of interpersonal and intrapersonal sameness of cognitive meaning, the intelligibility of such notions as interpersonal communication and intrapersonal endurance of meanings over time (which are usually understood as requiring distinct expression-tokens to share a meaning) is in serious jeopardy. Moreover, as argued in ([Fodor and Lepore, 1992], 11–17) the kind of holism entailed by IRS theories poses a prima facie threat to both an attractive metaphysical realism in the philosophy of science and the possibility of a scientific semantics. Finally, this sort of holism threatens the intuition that some meanings are encapsulated from some others — that my learning to infer from ‘x is an aardvark’ to ‘x is nocturnal’ seems not to affect the cognitive meaning of the expression ‘carburetor’ in my language/system of thought.

There are two standard responses to these worries. The first is to admit that holism would bring with it at least some of the negative consequences just adduced, but to deny that IRS theories are holistic in the way they are claimed to be. This response holds that, although the cognitive meaning of an expression e is determined by the cognitive meanings of some other expressions, it is not determined by the cognitive meanings of all the other expressions in the language/system of thought. It is hoped that this response, which

---

6 There is an analogous worry concerning the possibility of intrapersonal rational inference justification: the success of standard rationalizations of my inference from p to p ∨ q rests on the assumption that the two occurrences of the expression ‘p’ have the same cognitive meaning — that there has been no equivocation between the two expressions.

7 Realism in philosophy of science: if cognitive meaning is holistic, then ancient Greek astronomy and contemporary astronomy don’t agree or disagree, so astronomy can’t be progressive in any useful sense; this would sap the strength of standard arguments for scientific realism based on the progressiveness of science. Possibility of scientific semantics: if meanings are proprietary to time-slices of individuals, it’s hard to see how there could be any true and intrapersonally or intrapersonally applicable generalizations about the ways in which meanings are related to each other or the world.

8 This response is advocated in, e.g., [Boghossian, 1993] and [Block, 1993], and discussed
seeks to delimit the ramifications of the cognitive meaning of $e$ on the rest of the language/system of thought, will block the unpalatable consequences of holism. Unfortunately, the plausibility of this response seems to depend on the intelligibility of a principled analytic/synthetic distinction: to delimit the class of expressions whose cognitive meanings determine the cognitive meaning of $e$ from those which are irrelevant to the cognitive meaning of $e$ is tantamount to distinguishing the expressions linked to $e$ by meaning alone. Since [Quine, 1953] has convinced many philosophers that no principled and appropriately general analytic/synthetic distinction is possible, the prospects for this way of saving IRS theories appear rather dim.

The second standard response to holist worries about IRS theories is to accept that IRS theories fail to provide a robust notion of identity of meaning, but to hold that all the necessary work can be done by a notion of similarity of meaning. So, for example, an IRS theorist who favored this response might suggest that, although you and I attach different cognitive meanings to the English expression ‘aardvark’, we manage to communicate by using this expression insofar as our cognitive meanings are sufficiently similar.

The difficulty with this suggestion is that it is unclear how to understand the relevant notion of similarity if we lack a robust notion of identity of cognitive meaning. For if the cognitive meaning I attach to $e$ is determined by its connections to some set of cognitive meanings of other expressions $M$, while the cognitive meaning you attach to $e$ is determined by its connections to a set of cognitive meanings of other expressions $M'$, then presumably the similarity between the cognitive meanings we attach to $e$ is going to be a function of the extent of overlap between $M$ and $M'$. However, the IRS theorist we are imagining has agreed to forego a notion of identity of cognitive meaning, so there’s no obvious way to compare the members of $M$ against the members of $M'$. As Fodor and Lepore comment with respect to an IRS for objects of belief,

The colloquial senses of “similar belief” presuppose some way of counting beliefs, so they presuppose some notion of belief identity. If you have most of the beliefs that I have, then a fortiori, there are (one or more) beliefs that we both have. And if there is a proposition that you sort of believe and that I sort of believe strongly, then, a fortiori, there is a proposition that is the object of both of our beliefs. But precisely because these colloquial senses of belief similarity presuppose a notion of belief identity, they don’t allow us to dispense with a notion of belief identity in favor of a notion of belief similarity ([Fodor and Lepore, 1992], 18–19, emphasis in original).

Therefore, pending some clarification of the notion of similarity of cognitive meaning (that doesn’t appeal to a notion of identity of cognitive meaning), this objection is a decisive one against the response suggested.\textsuperscript{10}

\textsuperscript{9}This suggestion is particularly ubiquitous in the literature; e.g., see ([Harman, 1993], 169–170), ([Smith et al., 1984], 268), and ([Churchland, 1989], 102).

\textsuperscript{10}[Churchland, 1998] proposes to answer an analogous problem that arises for his own “state-
If this is right, then the holism entailed by IRS theories creates serious problems for these approaches.

3.2 IRS and Compositionality

A further objection against IRS theories, pressed in ([Fodor and Lepore, 1992], chapter 6), among other places, is that they cannot account for the compositionality of meaning. Compositionality can be understood for present purposes as demanding that (excepting finitely many idiomatic expressions such as ‘kicked the bucket’) the meanings of complex expressions must be determined exclusively by the meanings of their simple constituents plus logico-syntactic apparatus.\footnote{To use Fodor’s toy example, one expects that the meaning of the complex expression of English ‘brown cow’ should be in some way built from just the meanings of the constituents ‘brown’ and ‘cow’ together with information about the syntactic configuration in which these constituents fall in the complex expression.}

Why do Fodor and Lepore think IRS theories run afoul of compositionality? They point out that the cognitive meanings of complexes often involves (non-logico-syntactic) information apparently not available from any of their constituents; e.g., I may be willing to infer from ‘x is a pet fish’ to ‘x lives in a bowl’ even though I would infer from neither ‘x is a pet’ nor ‘x is a fish’ to ‘x lives in a bowl.’ In this sense, the additional information in the cognitive meaning of the complex expression ‘pet fish’ is emergent from the combination of the two simpler expressions ‘pet’ and ‘fish’, but is not part of the cognitive meaning of either on its own.

One response to this accusation, offered in [Block, 1993], is to hold that, in fact, the emergent information is encoded in a latent form in one of the constituents. So, on this line, the cognitive meaning of ‘x is a fish’ involves the information ‘x lives in a bowl if x is a pet’. If this were correct, it would be easy to see how thinkers might infer from ‘x is a pet fish’ to ‘x lives in a bowl’ on the space” semantics in terms of a topological relationship between regions of spaces in different dimensionalities. Transposing this proposal to the present idiom (this effort necessarily elides some of the details, I’m afraid), the idea would be, roughly, that $e$ and $e'$ count as similar in meaning just in case $M$ (the set of expressions connections to whose cognitive meanings constitute the meaning of $e$) properly contains $M'$ (the set of expressions connections to whose cognitive meanings constitute the meaning of $e'$) or vice versa.

Unfortunately, it seems pretty clear that this proposal won’t suffice. First, proper containment, once again, presupposes a way individuating the nodes in the network, and this is precisely what is at issue. Second, it would seem that proper containment is insufficient to ensure communication and the like: if all I believe about beer is that it is delicious and wet, and if you believe this but also that beer is a conscious substance, that beer is the President of the U.S., that beer is an abstract object, . . . (we can expand this set as much as we like without violating the containment constraint, once we make sure to include my single relevant conceptual connection in your larger set), then we satisfy the containment criterion, but I think we would say that my pronouncements using the expression ‘beer’ before you are likely to result in miscommunication.

Once again, there are many subtle issues I’ll ignore about e.g., what ‘determined by’ demands, what logico-syntactic apparatus is acceptable, just what the simple/complex distinction amounts to, and so forth.
basis only of the cognitive information of the simple constituent expressions plus logico-syntactic apparatus.\textsuperscript{12} There are several problems with this suggestion. First, it requires that the competency/possession conditions for ‘pet’ (or ‘fish’) involve knowing something about ‘bowl’; but that requirement is implausible for exactly the reasons that make the inference under consideration strongly emergent from the meanings of ‘pet’ and ‘fish’. Second, assuming constituency relations among cognitive meanings underwrite \textit{a priori} inferences, this proposal would make the inference from ‘\(x\) is a pet fish’ to ‘\(x\) lives in a bowl’ \textit{a priori}. Again, this seems doubtful. Finally, in view of the generality of the problem we are considering, it seems psychologically implausible that all such emergent information is encoded conditionally in the cognitive meanings of constituents.

Thus, if Fodor and Lepore are right in supposing that cognitive meanings are compositional, then the realization that what IRS theories offer as cognitive meanings aren’t compositional would stymie IRS aspirations.

### 3.3 IRS and Interpretation

A further challenge to semantic theories that include as distinct components (i) a theory of reference and truth and (ii) a separate characterization of cognitive meanings in terms of the conceptual roles of expressions comes from the criticism in [Lepore and Loewer, 1987] that such theories cannot serve interpretation in the way that they should. Lepore and Loewer argue in [Lepore, 1982], [Loewer, 1982], and [Lepore and Loewer, 1981] (along roughly Davidsonian lines) that an adequate theory of meaning for a language \(L\) should contain information such that a listener possessing this information should be able to understand \(L\) (given the rest of her cognitive capacities). In particular, they argue that a semantic theory should be sufficient to explain how listeners justifiably acquire beliefs about what a speaker believes on the basis of that speaker’s assertions. But, they claim in [Lepore and Loewer, 1981] that dual aspect theories such as IRS theories cannot supply such explanations.\textsuperscript{13}

To see this, first notice that, under an IRS theory, if \(S\) utters the expression \(p\), then \(H\) can conclude that \(p\) means \(q\) only if the conceptual role assigned by \(S\) to \(p\) is identical to that assigned by \(H\) to \(q\), because sameness of meaning is analyzed by such theories as identity of conceptual role. So \(H\) can conclude that \(S\) meant \(q\) by uttering \(p\) only if the role assigned by \(S\) to \(p\) is identical to that assigned by \(H\) to \(q\). But since, as discussed earlier, interpersonal iteration of the very same conceptual role is very rare, this makes it highly unlikely that \(H\) could ever reach such a conclusion about what \(S\) meant. Furthermore, if belief

\textsuperscript{12} One embarrassment associated with this suggestion is that it’s not clear whether we should choose to encode ‘\(x\) lives in a bowl if \(x\) is a pet’ in the cognitive meaning of ‘\(x\) is a fish’, or to encode ‘\(x\) lives in a bowl if \(x\) is a fish’ in the cognitive meaning of ‘\(x\) is a pet’. Perhaps this difficulty is not damning.

\textsuperscript{13} They also demand that semantic theories explain how listeners justifiably acquire beliefs about the world on the basis of speakers’ assertions, and that dual aspect theories make these explanations unavailable as well. However, their case for this last claim is given in different forms for different IRS theories, and I don’t have the space to enter into such detailed discussions here.
is construed in terms of a relation to expressions of a language of thought whose semantics is given by an IRS (a picture advocated by many IRS theorists), then all of H’s attributions of the form ‘S believes that p’ will be false unless H’s conceptual role for the linguistic expression p matches S’s conceptual role for the language of thought expression to which she stands in the belief relation — and once again, we have reason to think that such a match will be extremely rare.

This objection against IRS theories is especially significant because it has the consequence that these theories cannot supply rationalizing psychological explanations for behavior — the very explanations claimed as a driving motivation for the framework in the wake of Frege’s problem. To see this point, suppose that, for reasons just discussed, it is generally impossible for me to make true attributions of the contents of your beliefs of the form ‘S believes that p’, where ‘p’ in my mouth specifies the content of your belief. Of course, I might still be able to specify the conceptual roles of your mental representations. But it is hard to see how such specifications of conceptual roles, if (as we are supposing) they do not license conclusions about what content is believed, will by themselves serve to rationalize your actions. That is, if I cannot say what it is that you believe and desire, citing the conceptual roles of your mental representations leaves it entirely open why in the world you do what you do: I will have no rational explanation of the connection between having a mental representation with a particular conceptual role and the ensuing action. If this is right, then IRS won’t make available the sorts of psychological explanations that we want from an acceptable understanding of content.

References


