

Varieties of Coreference

KRISTA LAWLOR

Stanford University

Coreference: Internal, External and Strict

Udo says, “I need a haircut.” Victor agrees, “Yes, you do need a haircut.” There is same-saying here, and same-thinking. Looking at Udo, Wally says to a friend, “he needs a haircut”, unaware that Udo said any such thing. There’s same-saying here, too, but not the kind in the first case. In the first case, there is purposive same-saying, while in the second case, there isn’t. In the first case, we can say there is an *internal relation* of coreference—Udo’s utterance of “I” and Victor’s of “you” are supposed to corefer. The purposes of communication require coreference. In the second case of same-saying, there is coreference, but it is an *external relation*: Wally’s utterance of “he” is not supposed to corefer with anyone else’s utterance, although it happens to corefer with both Udo’s and Victor’s. No communicative purposes here require coreference. (Of course, Wally’s utterance *is* supposed to corefer with Wally’s own thoughts—so there is a second, internal, relation of coreference his utterance bears to his thought.)¹

As it turns out, it’s a wonderful fact about language and thought that there can be both internal and external coreference relations—wonderful and difficult, because it also turns out that things can go awry. Merely external coreference relations can be present when there should be internal relations. For instance, Udo looks in a mirror and says, “He needs a haircut even more than I do!” not realizing that it is his own reflection he sees. Udo’s utterances corefer, and he should realize as much, but he doesn’t. On the other hand, internal coreference relations can be present when there is no external coreference between terms. For instance, speakers can get confused and make mistakes: Wally says, “He needs a haircut” and his friend, thinking to agree, but looking at a different scruffy person, says, “he sure does.”

¹ See Perry (1988) for more about internal coreference.

Communicative purposes here, in part anyway, require coreference, but there is no actual coreference between terms.

These are no mere curiosities. Some of the deepest problems in the philosophy of language and mind rest on something going wrong with coreference relations. For instance, in Frege's puzzle about the informativeness of identity statements, and Kripke's puzzle about belief, subjects fail to recognize coreference—external relations of coreference hold while internal relations fail. (Call these Coreference Puzzles.) On the other hand, puzzles about confused thought arise when internal relations hold because subjects suppose coreference, even though external coreference fails. (Call these Confusion Puzzles.)²

In the philosophy of language, it is well known that Referentialists have difficulty handling these and related puzzles.³ Referentialists say that the semantic value of an occurrence of a name is only the object that it denotes. It seems then that Referentialists can talk about external coreference, and its absence, but not about supposed or purposed coreference. The Fregean on the other hand, can talk about supposed coreference, since the sameness of sense of two terms brings supposed coreference along with it. Fregeans can handle Coreference and Confusion puzzles fairly tidily. But how can Referentialists explain cases where external and internal coreference come apart?

Kit Fine offers a new solution for Referentialists. Briefly, the solution invokes a new primitive semantic relation—Fine calls it “strict coreference”—aimed at capturing a kind of on-purpose coreference. Two terms (in thought or language) *strictly corefer* if it is a semantic fact that they *corefer*; understanding the language requires understanding that the terms corefer. Fine then uses strict coreference to solve a raft of puzzles generated by the failure to recognize coreference. Frege's puzzle, to take one example, is resolved thus: In the statement “Hesperus is Hesperus”, the two tokens “Hesperus” and “Hesperus” strictly corefer, so in addition to their usual *referential* contribution to a *singular proposition*, they also make a relational contribution (namely, a *positively coordinated* sequence of individuals) to a *positively coordinated proposition*; on the other hand, in the statement “Hesperus is Phosphorous”, “Hesperus” and “Phosphorous” are not strictly coreferring tokens, and their contribution (namely, a *negatively coordinated* sequence) results in a *negatively coordinated proposition*.

² Defined thus, we can take Confusion puzzles to include puzzles about empty names.

³ Perry (2001) considers puzzles of “No Reference” or empty names and “Coreference” or cognitive significance; to these we add Confused Reference (see Camp (2002), Lawlor (2007); additionally the Referentialist faces problems about substitutivity in belief reports.

Thus the Referentialist, armed with different coordinated propositions to appeal to, can explain the differential informativeness of the two statements, even if the two identity statements express the same singular proposition.⁴

Cognitive Significance

Can strict coreference really do the job? Well, the proof of the pudding is in the eating, and Fine has tidy solutions to offer for a number of puzzles. But even so, I think we can still ask whether strict coreference is the right semantic primitive for the Referentialist, or whether something better might be found. In answering, we should remember what we seek, in trying to explain coreference phenomena quite generally.

First, in explaining coreference phenomena, we should bear in mind that it is sensitivity to the semantic properties of thoughts and utterances that explains our speaking, acting and reasoning as we do. One reasonably enough reacts differently on being told “Hesperus is Hesperus” than on being told “Hesperus is Phosphorus.” One also plans differently for the production of these utterances; and one reasons differently with the thoughts expressed by these utterances. In building a theory of coreference phenomena, ultimately, we want to identify that semantic property sensitivity to which explains and rationalizes our differential cognitive responses to referentially identical utterances.

Do we get an account with Fine’s view of that semantic property, sensitivity to which explains our differential treatment of utterances? This property is of course also known as an utterance’s *cognitive significance*. Fine’s candidate for the cognitive significance of an utterance or thought is a coordinated proposition. Though he never explicitly says as much, the evidence is pretty clear: Fine claims that we grasp standard Referentialist propositional contents by grasping coordinated propositional contents [58]; and a coordinated proposition is that content of belief in virtue of which we can differentiate a belief *that Cicero is Cicero* from a belief *that Cicero is Tully* [77]. It is also the presence or absence of coordinated propositions that accounts for whether a rational person will make an inferential transition or not [83]. Finally, relative differences in strict coreference of pairs of names is sufficient,

⁴ Fine’s relationism does share some broad features with existing referentialisms. Perry (2001) for instance, also accepts that non-intrinsic features of utterances (namely contextual features) help to determine propositions over and above the standard Referentialist proposition. And Fine shares with Perry the rejection of what Perry calls the “fallacy of misplaced information”—the idea that all the information of an utterance comes from the (singular) proposition expressed.

Fine says, to account for differences in cognitive significance [83]. So it seems safe to say that Fine intends the cognitive significance of an utterance to be located in coordinated propositions.⁵

Now we might wonder, is grasping a coordinated proposition grasping the cognitive significance of an utterance? Coordinated propositions are abstruse enough to seem unlikely objects of an agent's belief. Take a simple case: Udo thinks that "Cicero is Cicero." The relevant coordinated proposition $\langle p, @ \rangle$ is a Russellian proposition p with Cicero as a constituent occurring twice over, paired with a coordination scheme, $@$. Coordination schemes are "equivalence relations on the occurrences of individuals [in a given (sequence of) proposition(s)] such that two occurrences of individuals are related by $@$ only if they are occurrences of the same individual" [56]. In our example $@$ will positively coordinate the two occurrences of Cicero, because the token names "Cicero" and "Cicero" strictly corefer. But it seems that when Udo thinks "Cicero is Cicero", *the manner in which an equivalence relation treats the two occurrences of the individual Cicero in the proposition that is the object of his belief* is not what's on Udo's mind. There is good reason to make it a qualification for being the cognitive significance of an utterance that it be a proposition cognized or believed, since it is what makes rational a range of behaviors with respect to the utterance, from one's own point of view. So my first point is that strict coreference doesn't seem to yield a satisfying story about cognitive significance, and that might make us wonder if it is the right primitive for the Referentialist.

Confused Reference

A second concern about the choice of primitives is this. In explaining coreference phenomena, we should also bear in mind the full range of the phenomena: specifically, this means we need an account of Confusion puzzles. With *strict coreference* we get a tool designed for half the problem space—namely, Coreference puzzles, where external coreference exists, but goes unrecognized. Problems of Confusion, on the other hand, where coreference is supposed although none exists, pose a fundamental challenge for Fine's framework.

⁵ Fine also identifies cognitive significance with the cognitive impact an utterance has on a speaker's informational base, or the impact it has on any given informational base (also called "cognitive potential"). What I am calling "cognitive significance" is just that semantic property sensitivity to which rationalizes behavior toward an utterance. These are different ideas, but I think because of the evidence cited that Fine is actually trying to account for cognitive significance in this latter, standard, sense.

Suppose Wally says of Udo, “He needs a haircut”, and Zach, thinking to agree, but looking at another person, says, “he sure does.” Zach is confused. His communicative purposes require the coreference of his utterance and Wally’s, but there is no actual coreference between the utterances. Mistakes of all kinds generate confused reference—and in all such cases, we have *internal coreference without coreference*. The problem is that Fine’s *strict coreference* implies *coreference*. Zach’s utterance internally corefers with Wally’s but doesn’t strictly corefer, because it doesn’t corefer. So it is hard to see how to account for Confusion puzzles, if we start with strict coreference.⁶

In the Afterword, Fine suggests that Confusion puzzles can be handled by invoking *putative semantic requirements*. Although there is no semantic requirement that Zach’s utterance corefer with Wally’s, there is a putative semantic requirement of coreference. Moreover, in such cases, a *backup semantic requirement* takes effect, and so “instead of failed reference to two ordinary objects we have successful reference to some sort of amalgam of these objects.” [126] We might wonder how to make this suggestion work. What we want is to interpret confused utterances in such a way that we can see how, first, a confused utterance is not just a crazy or unintelligible utterance; and second, the subject’s reasoning with the proposition(s) expressed by confused utterances is often good reasoning. An amalgam might help with the first of these, but it is not clear how it helps with the second. Assessing reasoning requires assigning truth values to utterances/thoughts. So now we ask, if an amalgam of two people is the referent of Zach’s confused utterance, then is his utterance true? Perhaps a supervenient account is the natural way to go here—Zach’s utterance is true if true on any way of disambiguating his utterance, i.e. true if each of the people composing the amalgam needs a haircut. But then, what about Zach’s identity utterance: “the guy I said needs a haircut is the guy Wally said needs a haircut”? This utterance strikes us as false, but on a supervenient approach it is neither true nor false. That’s an awkward result.⁷ Moreover, Zach’s commitment to the truth of this identity is what rationalizes his subsequent reasoning. So it is hard to see how we’ll find an interpretation of Zach’s utterances that will rationalize his inferences from his point of view, if we go supervenient. Maybe a supervenient semantics isn’t the way to make good on Fine’s brief

⁶ Empty names create a challenge as well, since the relationist is also at pains to say how terms without reference strictly corefer.

⁷ See Camp (2002) for discussion of limits of the supervenient approach to confusion.

suggestion about amalgams, but I don't know offhand of another way.⁸

Another concern about Fine's suggestion, centering on the distinction between putative and actual semantic requirements, is this: before considering Confusion puzzles, Fine suggests that knowledge of actual semantic requirements is reliably available to competent language users in a position to understand a given utterance [60–64]. But once Fine introduces putative requirements, this knowledge cannot be so available. Imagine Zach isn't sure whether his use of "Udo" refers to one person—he's heard people speak of Udo, but is it one person they're all talking about? Here Zach won't know whether his utterance, "Udo sounds like an interesting person", is subject to a putative or a real requirement. And if he doesn't know this, he doesn't understand his own words.

Setting aside these concerns, we might think that invoking *putative* requirements is a step in the right direction; since putative requirements are *supposed* requirements, this step seems to be taking us closer to *internal coreference* (internal coreference is a *supposed*, though *not necessarily actual*, coreference relation). But now of course the question is, *What is a putative semantic requirement* of coreference? Certainly this seems to involve an agent's merely supposing that coreference obtains. *Supposing* coreference is a cognitive activity—broadly it is *thinking* coreference obtains. So what is it for someone to think that coreference obtains?

Coordination in Thought

Fine's Chapter 3 begins with this very question: "But what is it to represent an object as the same within one's thought?" [67] For instance, Udo thinks about Hesperus, recalling what he knows about it. Udo thinks *it is visible in the evening*; then he also remembers, *it is really a planet*. Here Udo's two thoughts represent their object (i.e. Venus) as the same. In what does this representational fact consist?

⁸ In a footnote Fine briefly suggests what might be a different approach to Confusion puzzles, saying "...one might allow the coordination scheme @ to relate occurrences of distinct individuals. This would correspond to the conflation of two individuals as one." [n14, p136] Perhaps Fine here just means to specify the coordination scheme that goes along with reference to an amalgam. But *perhaps* what Fine means is to allow coordination where no coreference occurs because no successful reference occurs. I suggest below that this is really what we want to do. It should also be noted, however, that there is a tension between this latter idea and the idea suggested in the Afterword: we don't need successful reference to an amalgam if we've *dropped* the requirement of coreference and so of successful reference.

It is important to understand how thoughts represent objects as the same. Not only is such representation pervasive in remembering and perceiving, but it is also of fundamental importance in reasoning. For reasoning requires trading on coreference.⁹ (I'll say more about reasoning below.) So we very much need an account of representing objects as the same in thought.

Fine, deftly to my mind, criticizes a range of alternative accounts. For instance, it won't do to say that what it is for thoughts to represent their objects as the same is for one to have an additional thought that the objects are the same; that just leaves the question of how the additional thought manages to represent its object as the same as the originals' objects. Nor will it do to say that for thoughts to represent their objects as the same is for the thoughts to somehow involve the same file of information; invoking a file isn't a way of giving an account as much as a metaphor that presumes the very notion to be explained.¹⁰

Fine's own account is simply stated: what it is to represent an object as the same within one's thought is for there to be an intentional or representational requirement that the object represented is the same. Two points are worth making about Fine's account: First, it is a thin explanation. In the language case we can say there is a semantic requirement that the object represented is the same when understanding the language in question requires understanding that the terms in question in fact corefer. In the case of thought, however, how do we cash out what it is for there to be a representational requirement that the object represented is the same? We could try an analog claim, saying "understanding the thoughts in question requires understanding that the elements of the thought in fact corefer." But understanding that the elements of the thought corefer was what we were trying to give an account of, wasn't it? It seems that being told that representational requirements hold doesn't move us far from the very phenomenon in need of explaining.

Second, and more importantly, even if we accept Fine's account of coreferential thinking, we don't get an account of what we really need: Recall, we needed an account of what is it for someone to think that coreference obtains *whether or not it does*. We needed this to answer the question *What is a putative semantic requirement?*, where this seems to involve an agent's merely supposing that coreference obtains *whether*

⁹ On this matter, see Campbell (1987), Boghossian (1992), Millikan (1996), Lawlor (2001), Gerken (2007).

¹⁰ It should be noted that not all accounts that make use of files to model coreference in thought are attempting a constitutive account. But I agree with Fine, and make a similar criticism of files as constitutive of coreferential thinking in Lawlor (2001).

or not it does obtain. Now, for Fine, “coordination within thought is taken to be a form of strict-corepresentation, in analogy to our previous account of coordination within language as a form of strict coreference.” [66] The problem is that we can’t get an account of *merely* supposing or thinking that coreference obtains if we start with the primitive notion of *strict coreference* in thought. For again, strict coreference entails coreference, and so is the wrong place to start in seeking to make sense of merely supposed coreference, which does not.

Maybe Fine could respond, “I’ll explain putative semantic requirements of coreference some other way, so we don’t need to account for them in terms of *merely supposed coreference*.” Perhaps this might be done. Even if it can be done, though, I think we’ll still need an account of merely supposed coreference. There are many reasons we might want such an account, beyond explaining Confusion cases. There is, for instance, the vexed issue of intentional identity, which is another case where utterances involve merely supposed coreference.¹¹

Moreover, there’s making sense of the epistemology of reasoning. As Fine notes, there are inferences that are truth-preserving that a subject is not warranted in making [119]. Conversely, we also find cases of inferences that a subject is warranted in making that are not truth-preserving. Two examples illustrate these points respectively. The first is Fine’s case: Imagine that Peter asserts “Paderewski is musical”, and then we derive our use of the name from him, and infer “Paderewski is musical.” Then Peter derives what he takes to be a new use of the name “Paderewski” from us, and infers “Paderewski is musical.” An inference from Peter’s first assertion to his second assertion is certainly truth-preserving, but it wouldn’t be warranted. The second example involves the sort of case discussed in the literature on externalism and self-knowledge. Imagine that Peter asserts “Paderewski is musical”, and then unwittingly switched to Twin Earth, asserts “Paderewski is musical.” Suppose he then makes the inference from the first to the second assertion. Peter’s inference is arguably not truth-preserving, involving as it does an equivocation; but the inference is in some sense warranted—it is rational from Peter’s point of view.

What we have here are cases (i) of inference that *is truth-preserving* because terms do corefer, but *not warranted*, because the subject does not take it that terms corefer; and (ii) of inference that *is warranted* because the subject takes it that terms corefer, but is *not truth-preserving*

¹¹ Geach (1967); Edelberg (1992). For instance, “John wants to catch a fish today, and have it for supper.” The relevant attitude attributions can be true without there being a fish that John wants.

because the terms merely taken to corefer do not corefer. Using a notion of coordination built on strict coreference, Fine can handle the first sort of case. But the second, resting as it does on confusion, again presents a difficulty. We need a notion of merely supposing coreference to handle such cases.

To recap, I think we want a more substantive answer to the question of what it is to represent an object as the same within one's thought than we get by appealing to representational requirements. More importantly, we want a substantive characterization that permits us to say how inferences might be warranted from the subject's point of view even if they're not truth-preserving.¹² Strict coreference is supposed to do the work of sameness of sense, which for the Fregean is, in part, the work of characterizing inferential warrant; but as we've seen, strict coreference doesn't work for confused inference. The epistemology of inference, especially, suggests that we might do well to cast further for a different semantic primitive than strict coreference.

Semantic Pro-forms

Let me briefly suggest a different starting place for a comprehensive theory of coreference in language and thought. Fine's primitive notion is strict coreference, and strict coreference implies or entails coreference. This feature turns out to be a limitation, when it comes to a general theory of coreference phenomena. A better relational semantic primitive might drop this feature.

In fact, we've already seen a relation without this feature—namely, *internal coreference*. We might understand internal coreference in terms of a chaining relation among token expressions in thought and language, with no implication of successful reference.¹³ This chaining is familiar in syntactic theory—pronouns and other pro-forms are supposed to share reference or semantic value, regardless of whether they actually refer at all. Semantic theory, it seems, might also do well to incorporate the notion of a pro-form.

Clearly, many questions would need answering, if we started with the notion of semantic pro-forms or chains. It is not immediately clear how we build an account of cognitive significance, for instance, from this primitive. That is just one issue. There are also issues about how

¹² This was my goal in Lawlor (2001) and (2007). See also Mikkel Gerken's (2007) .

¹³ Fine mentions the idea of tokens of a name being "internally linked" [107] where an internal link occurs "just in case the speaker takes them to have the same use." This might sound like what we need; but then Fine claims that tokens will be *coordinated if internally linked*. So again, since coordination implies actual coreference, the same problem infects the notion of being internally linked.

compositionality would look after we have included pro-forms into our semantics.

Finally, I note that what I am suggesting seems in keeping with broad ambitions of Fine's project. We would start theorizing about coreference phenomena with a relational semantic primitive. So perhaps the suggestion can be seen as a friendly amendment.

Conclusion

Does a comprehensive theory of language and cognition require primitive semantic relations? Fine thinks so, and I agree. Will incorporating primitive semantic relations have the result that traditional semantics will need to be dramatically re-written? Again, plausibly Yes—our understanding of compositionality, to take one instance, may have to be more subtle. Is strict coreference the right primitive in such a comprehensive theory? No, I think not—the semantic pro-form is a better primitive, because it affords a more general theory.

Naturally I've pressed criticisms here, but I greatly appreciate this book. We have Fine's work to thank for turning a powerful spotlight on coreference phenomena, illuminating how central they are in thought and language, and forcing us to think harder about how to understand them.¹⁴

Works Cited

- Boghossian, P. A. (1992). "Externalism and Inference." *Philosophical Issues* 2(Rationality in Epistemology): 11–42.
- Camp, J. (2002). *Confusion: a Study in the Theory of Knowledge*. Cambridge, MA, Harvard University Press.
- Campbell, J. (1987). "Functional Role and Truth Conditions: Is Sense Transparent?" *Aristotelian Society Supplementary Volume LXI*: 273–292.
- Edelberg, W. (1992). "Intentional identity and the attitudes." *Linguistics and Philosophy* 15: 561–596.
- Fine, K. (2007). *Semantic Relationism*. Oxford, Blackwell Publishing.
- Geach, P. (1967). "Intentional identity." *Journal of Philosophy* 64: 627–632.
- Gerken, M. (2007). *Epistemic Reasoning and the Mental*. Ph.D Dissertation, UCLA.
- Lawlor, K. (2001). *New Thoughts about Old Things: Cognitive Policies as the Ground of Singular Concepts*. *Studies in Philosophy*, Robert Nozick, ed. Garland Publishing, Inc. New York, NY.

¹⁴ I'm grateful to David Hills for wide-ranging discussion.

- (2005). “Confused Thought and Modes of Presentation.” *The Philosophical Quarterly* vol. 55, no.218, 21–37.
- (2007). “A Notional Worlds Approach to Confusion.” *Mind and Language*, vol.22 no.2 April pp. 150–172.
- Millikan, R. (1996). “Images of Identity.” *Journal of Philosophy*.
- Perry, J. (1988). “Cognitive Significance and New Theories of Reference.” *Nous* 22: 1–18.
- (2001). *Reference and Reflexivity*. Stanford, CSLI Publications.