Frege on Identity

Michael Weisberg
Language Core Seminar First Paper
Fall 2000

Frege’s most celebrated paper “On Sense and Reference”\(^1\) begins with a puzzle. The sentences “a=a” and “a=b” superficially strike the reader as being different kinds of claims. Among other things, “a=a” seems to be trivially true, even self-evident. On the other hand, “a=b” is an informative sentence. It is of a kind important to many episodes in the history of science. Deeper reflection, however, reveals that the situation is more complicated. Both sentences express identity or sameness relations. If we think that sameness is a relation between an object and itself, both sentences have the same content. So the puzzle is to explain identity in a way that accounts for the difference between these two sentences.

At the beginning of “On Sense and Reference,” as the fictional philosopher McWood notes, Frege suggests two possible accounts of identity. The first possibility is that sameness is a relation between objects. Frege’s second suggestion is that identity sentences express a relation between the names (or signs) of objects. This second suggestion is the view that Frege claims to have endorsed in his earlier book the \textit{Begriffsschrift}. McWood notes that Frege does not actually tell us which of these models he has chosen. She also notes that the two models have the same truth conditions, so some of the

\(^1\)All references and translations are from “On Sense and Nominatum” by Gottlob Frege. Trans. Herbert Feigl, reprinted in A.P. Martinich, \textit{The Philosophy of Language}, p. 186-198. I have substituted the word ‘reference’ for Feigl’s ‘nominatum’ according to customary practice.
remarks he makes at the end of the paper cannot help us adjudicate between
them. One cannot argue with McWood’s interpretation taken literally. Frege
does not explicitly tell us which of the models that he endorses. The final
statements of “On Sense” are not of much help since either model of identity
can account them for them. However, if one closely examines the two models
and considers what problems Frege’s account of sense can actually fix, we will
find that Frege has taken a side on the dispute. Specifically, I will argue that
Frege has built his new model of identity by adding sense to the object model.
Before arguing this, I will explain the two models and Frege’s objections to
them.

1 The Two Models

Most of what Frege says about the object model comes by way of critique.
Here, however, I will just deal with the intuitive, underlying idea. The
sentence “Hesperus=Phosphorus” gets decomposed into three parts on this
model. It contains the sign ‘Hesperus’ the sign ‘Phosphorus’ and the ‘=’
symbol. The sign ‘Hesperus’ stands for the object Venus as does the sign
‘Phosphorous’. The ‘=’ then expresses a relation between the object that
each sign stands for. Each sign stands for the planet Venus. Thus an iden-
tity relation is truthfully expressed, for the object Venus is identical to itself.
In Frege’s words: “This would express a relation of a thing to itself, namely,
a relation such that it holds between everything and itself but never between
one thing and another.” (186)

Contrasting the object model is the coreference model, which is also dis-
cussed by Frege in the first paragraph of the essay. Consider once again the
sentence “Hesperus=Phosphorus”. As with the object model, the sentence
gets decomposed into three parts: ‘Hesperus,’ ‘Phosphorus,’ and ‘=’. Unlike
the object model, ‘Hesperus’ does not stand for the object Venus. Rather,
in an identity sentence, the sign ‘Hesperus’ stands for the sign ‘Hesperus’
and ‘Phosphorus’ stands for the sign ‘Phosphorous’. The identity symbol expresses a relation between these signs. Specifically, it express the relation of coreference between the two signs. This means that each of the signs refers to the same thing (outside the context of identity sentences). So on this account, the signs refer to themselves and a special relation of coreference is asserted to hold between them when the identity sentence is true.

2 Problems for the Models

I already alluded to the major problem for the object model in discussing Frege’s opening puzzle. If the sentence “a=b” expresses an identity between the objects a and b, then if it is true, it expresses the same thing as the sentence “a=a.” Because “a=b” can only be true if a=b, i.e. a is the same object as b, it has the same content as “a=a.” Let us call this the difference problem:

(DP) According to the object model, there is no difference in content between “a=b” and “a=a” if “a=b” is true.

Frege thinks that there is obviously a difference between “a=a” and “a=b” and the difference is in what he calls the cognitive content of the sentences. He gives us a few bits of evidence that there is, in fact, a difference. First, he points out that one of these sentences is knowable a priori and the other is not. We need to go out into the world to discover that “Chelsea’s father=Bill Clinton,” but we can conclude that “Chelsea’s father=Chelsea’s father” from the comfort of our recliners. A closely related piece of evidence is that some discoveries of identity sentences like “a=b” have been considered astonishing scientific achievements. Frege mentions the discovery that the same star (the sun) rises in our sky every morning as an example of such a sentence. Although closely related to the issue of a priority, examples like these force philosophers skeptical of a priori or analytically true knowledge to acknowledge the problem.
Having devoted half of a paragraph to criticizing the object model, Frege’s initial conclusion is that maybe we really need the coreference model. That is, we want our identity sentences to express the coreference relation between signs. However, he immediately mentions two problems with this view. The first problem arises because connections between signs and the world are arbitrary. There is nothing to prevent “the use of arbitrarily produced process[es] or object[s] as a sign for something else.” This is a subtle point. On the coreference model, identity sentences are true because the signs stand in the relation of coreference to one another. As long as signs the sign ‘Hesperus’ and the sign ‘Phosphorus’ are both attached to Venus, then the coreference model and the identity model will reach the same conclusions about identity statements. However, if we only examine our signs, not the world, there is no way to be sure that the signs are referring to the correct objects.

The problem of arbitrariness is a completely general one when we take the reference of a sign to be the sign itself. It creates an especially interesting problem for the coreference model. Because signs can be arbitrarily attached to objects, the relation of coreference must have resources in its definition to prevent this possibility. An obvious choice would be that corefers(‘a’, ‘b’) is true if ‘a’ refers to a, ‘b’ refers to b, and a=b. This means, however, that the coreference model and the object model have the same truth conditions. This might make us suspicious about the claim that signs are referring to themselves in identity sentences. Not defining coreference in this way, however, forces us to accept the possibility of arbitrary assignments of signs to objects, in which identity sentences are nonsense. Formally, let us say that the arbitrariness problem is:

(AP) The possibility of an arbitrary assignment of signs to objects prevents the coreference model from justifiably expressing identity claims, unless one makes the truth conditions of coreference the same as those of the object model.
The second problem for the coreference model can be thought of as a criticism based on the use/mention distinction. If “a=b” is true in virtue of a relation between the signs ‘a’ and ‘b’, the fact expressed by the sentence is a fact about our language, not about the world. Casting it as a distinction between use and mention, we might say that the coreference model mistakes the mention of the signs ‘a’ and ‘b’ for their use (in natural language), which is to pick out the objects a and b. Frege puts the same idea more obscurely when he says that in the coreference model, “a sentence like “a=b” would no longer refer to a matter of fact but rather to our manner of designation.” (186) Let us call this second objection the use/mention problem:

(UMP) The coreference model mistakenly asserts that identity is a relation between signs, mistaking the mention of a sign for its use.

3 Which Problems are Solved by Sense?

Frege’s solution to the opening puzzle, of course, involves the distinction between sense and reference. The meaning of a name, according to Frege, really has two components. First, there is the referent—the object that is picked out by the name’s sign. The sign ‘Clinton’, for example, picks out Bill Clinton, the President. We therefore say that the reference of ‘Clinton’ is Clinton. Reference allows us to account for the fact that signs do not map to objects one-to-one. The referent of ‘Chelsea’s father’ is also Clinton as is the referent of ‘the President of the United States.’ The difference between ‘Chelsea’s father’ and ‘Clinton’ is accounted for by the sense. Frege never fully characterizes the notion of sense in this paper, be he gives us some descriptive clues. He variously refers to the sense of a term as its mode of presentation, its connotation, and sometimes (although circularly, I think) as its cognitive significance.

---

With an account of the sense/reference distinction, we can see how the identity puzzle is solved. The signs ‘a’ and ‘b’ can express different senses, while still denoting the same object. Since ‘Chelsea’s father’ and ‘Clinton’ have different senses, it is informative to assert the identity sentence “Clinton=Chelsea’s father.” On the other hand, “Clinton=Clinton” is not informative because both signs referring to President Clinton have the same sense.

We can now return to McWood’s claim that Frege never actually chooses between the object and the coreference models. We should note that whichever model Frege chooses will have sense built in. The object model will be object +sense model (OS) and the coreference model would be coreference+sense model (CS). So I will assume that McWood wants to know if Frege chooses OS or CS. Since Frege adopts neither “officially,” I will adjudicate the issue by investigating how OS fairs with the DP and how CS fairs with AP and UMP.

First let us consider CS. The referent of a sign, according to CS, is still the sign itself. Additionally, the sign expresses it sense. This sense, or mode of presentation, also determines the reference of the sign, which (in this case) is the sign itself. Before we added sense, the sign referred to itself. Upon adding sense, the sign refers to itself indirectly vis-à-vis its sense. Recall that the UMP is that coreference models mistake the mention of a sign with its use. More simply, it says that identity is posited between the wrong kinds of things—words instead of objects. Does CS either fix this problem or explain it away? We can safely answer no! By adding sense, we are actually adding another layer of indirectness to our model of identity. UMP tells us that identity is being asserted between the wrong kinds of things. Now we only compound the problem by making a sign refer to itself indirectly, with the reference remaining the sign itself. Of course adding sense probably doesn’t make the problem worse, it simply doesn’t help. Thus, the UMP is also a valid criticism of CS.
The CS model fairs slightly better with the AP. It is possible that sense could be used to ensure that signs are not arbitrarily attached to objects, but this involves going beyond the text. In “On Sense and Reference,” Frege doesn’t give us the metaphysics of senses. If they are abstract objects that are publicly accessible and graspable, as he is known elsewhere to argue, then it may be that senses ensure non-arbitrary connections between signs and objects. For example, ‘Clinton’ would still refer to ‘Clinton,’ but there is a sense expressed by ‘Clinton’ that in turn points to the President of the United States. This would certainly ensure that names pointed to the correct objects, but it has two major costs.

First, one has to accept a very robust metaphysics. Senses, to be useful in this way, have to be publicly accessible objects. I cannot give a detailed critique of the metaphysics of abstract objects, but their existence as publicly accessible objects is certainly not uncontroversial. Frege clearly accepts their existence, so the second problem is more salient. The second problem with this kind of solution to AP is that the coreference part of CS is no longer necessary to do any semantic work. The central claim of the pure coreference model was that the signs in identity sentences refer to themselves, Frege’s first attempt to account for differences in cognitive significance. On our new model CS, the sign is still referring to itself, but the sense is determining the connection between the sign and the world. The sense is both solving the problem of conceptual content (as I will discuss below) and hooking the signs, indirectly, onto objects in the world. The coreference part of the model merely hangs off to the side, no longer doing any semantic work. This does not bode well for CS.

While the coreference model is not helped much by the addition of sense, it seems to be just the sort of addition needed by the object model to solve the DP. Whatever the difference in cognitive content is between “a=a” and “a=b”, it is not captured by the object model, hence the model gives rise to the DP. Conveniently, however, by carefully distinguishing between two
aspects of meaning—sense and reference—Frege is able to solve this problem. In identity sentences, the sign ‘b’ and the sign ‘a’ have the same referent, but different senses. When we append the sense/reference distinction to the object model (OS), we get the best of both worlds. The signs connect to objects in the world via the referents and we can continue to say that identity is a relation between objects. Differing senses between ‘a’ and ‘b’, however, allow us to account for the difference in cognitive significance between these signs. We bring our account of semantic content in alignment with our account of cognitive significance. Appending sense to the object model solves the DP in a straightforward way.

Let us now return to McWood’s claim. While Frege does not explicitly choose the object model of identity, I think we now have gathered the evidence necessary to conclude that this is his choice. The addition of sense to the object model (OS) is exactly what is needed to solve the DP. The semantic content of sentences is matched up with the cognitive content of thoughts by this move. This addition to our original model, however, comes with no theoretical cost. Adding sense renders no part of the original object model superfluous. In contrast, when we add sense to the conference model we get a disjointed theory with unnecessary components, as mentioned above. More importantly, adding sense does not solve the problems Frege identifies for the conference model. While we cannot be absolutely sure that the OS model was Frege’s intention, the principle of charity seems to demand that we interpret him as such.