CrossMark

Sense, reference and substitution

Jeremy Goodman[2] · Harvey Lederman[1]

Abstract We show that, contrary to conventional wisdom, Frege's distinction between sense and reference does not reconcile a classical logic of identity with apparent counterexamples to it involving proper names embedded under propositional attitude verbs.

Frege (1892) held that, when names are embedded under propositional attitude verbs, they refer not to their ordinary referents but to their ordinary senses. For example, consider:

1. If Hesperus is Phosphorus and Hammurabi knew that Hesperus is Hesperus, then Hammurabi knew that Hesperus is Phosphorus.

According to Frege, asserting this sentence would involve something like equivocation: the first occurrence of "Hesperus" would refer to Hesperus, while the second occurrence of "Hesperus" would refer not to Hesperus but to the ordinary sense of "Hesperus".

It is often thought that, by accepting this doctrine, Fregeans can hold that, although someone who asserted 1 would speak falsely, 1 is still valid, as is the schema:

✉ Harvey Lederman
harvey.lederman@princeton.edu

Jeremy Goodman
jeremy.goodman@usc.edu

[1] Princeton University, Princeton, NJ, USA

[2] University of Southern California, Los Angeles, CA, USA

Substitution: If a is b and Φ, then $\Phi[b/a]$.

Instances of this schema, such as 1, are obtained by replacing Φ with a declarative English sentence φ, replacing a and b with proper names n and m, and replacing $\Phi[b/a]$ with a sentence obtained from φ by replacing an occurrence of n that is not within quotation marks with an occurrence of m.

For example, in what is arguably the *locus classicus* of contemporary Fregeanism, Kaplan (1968) writes:

> Frege's main idea, as I understand it, was just this. [...A]pparent failures of substitutivity and the like [are] due to confusion about what is denoted by the given [term's] occurrence. (p. 183)

> So we require no special non-extensional logic, no restrictions on Leibniz' law, on existential generalization, etc., except those attendant upon consideration of a language containing ambiguous expressions. (p. 184)

To illustrate Kaplan's idea, consider the sentence:

2. If Jane has tenure, then Jane has tenure.

If we equivocate, we can use 2 to speak falsely; for example, by using the different occurrences of "Jane" to refer to different people with that name. But it would clearly be misguided to deny the validity of 2 on this basis. Kaplan's idea is that it would be similarly misguided for Fregeans to deny the validity of 1 on the basis of the fact that we can use 1 to speak falsely.[1]

Both proponents and detractors of Fregean senses (although not Frege himself) have taken the reconciliation of Substitution with the falsity of sentences like 1 to be the principal virtue of postulating such entities. For example, Kaplan (1968, p. 185) writes that his "own view is that Frege's explanation, by way of ambiguity, of what appears to be the logically deviant behavior of terms in intermediate contexts [e.g., complement clauses of attitude ascriptions] is so theoretically satisfying that if we have not yet discovered or satisfactorily grasped the peculiar intermediate objects in question, then we should simply continue looking". And Carnap (1947, p. 136), a detractor of senses, writes:

---

[1] Our 2 plays the role of Kaplan's unnecessarily confusing (11) involving two uses of "F.D.R.". The sort of "restrictions [...] attendant on consideration of a language containing ambiguous expressions" that Kaplan has in mind are *not* on which sentences to count as instances of schemas, but rather on which sentence *uses* are such that their falsity threatens the validity of schemas of which the used sentence is an instance. He would not deny, for example, that 2 is an instance of the schema "If Φ, then Φ" (as his distinction between ambiguity-based and what he calls "mono-denotationalist" ways of thinking about such sentences makes clear). He writes: "The natural analysis of (11) involves pointing out that the name 'F.D.R.' is ambiguous, and that in the second clause it denotes a television show rather than a man. Substitutions or any other logical operations based on the assumption that the name has here its usual denotation are pointless and demonstrate nothing [..., unlike] transformations based on a *correct analysis* of the name's denotation *in this context* [...]" (Kaplan 1968, p. 183, emphasis his).

The idea that Fregeans should hold that sentences like 1 are valid is not an idiosyncratic suggestion of Kaplan's. On the contrary (and to our surprise), it has been a common reaction, both in talks and in conversation, whenever we have claimed that the validity of such sentences marks an important contrast between Fregean and Millian treatments of the semantics of names.

> It seems that Frege was aware of the fact that [Substitution] would lead to a contradiction if the ordinary nominata of names were ascribed also to their oblique occurrences and that the contradiction does not arise if different nominata are ascribed to these occurrences. [...] It is true that Frege does not speak explicitly of the necessity of avoiding a contradiction; he gives other reasons for his distinction between the ordinary nominatum and the oblique nominatum of a name. His reasoning gives the impression that this distinction appeared to him natural in itself, without regard to any possible contradiction. However, I think that to many readers it will scarcely appear very natural and that they, like myself, will see the strongest argument in favor of Frege's method rather in the fact that it is a way of solving the antinomy.

In this note we will argue that Kaplan and Carnap are mistaken. Whatever its other merits, Frege's doctrine that names refer to senses when embedded under propositional attitude verbs does not reconcile the validity of Substitution with the threat posed to it by sentences like 1.

Consider the sentence:

3. If Kripke knows that Hesperus is Phosphorus, then Hesperus is Phosphorus.

Assume for the moment that this sentence is valid. (It is, after all, an instance of the schema "If S knows that Φ, then Φ", perhaps the most basic principle of epistemic logic.) Assume, moreover, that the set of valid sentences is closed under classical propositional logic. So if 1 and 3 are both valid (and we interpret "if ..., then ..." as material implication), then the following sentence must also be valid:

4. If (Kripke knows that Hesperus is Phosphorus) and Hammurabi knew that Hesperus is Hesperus, then Hammurabi knew that Hesperus is Phosphorus.

Now according to Frege, in 4 every word has the same reference in each of its occurrences.[2] So Kaplan's strategy for holding that 1 is valid despite expressing

---

[2] This is not to say that, according to Frege, "Hesperus" and "Phosphorus" refer only to their ordinary senses in all of their occurrences in 4. Frege held that the complement clauses of factive attitude verbs need "to be taken twice over, with different referents, of which one is a thought, the other a truth value" (Frege 1892, p. 228). In other words, expressions in the complement clauses of ascriptions involving factive attitude verbs (i.e., those which create a presupposition of the truth of their complement clauses) refer to their ordinary senses for the purpose of determining those ascriptions' truth-values and refer to their ordinary referents for the purpose of determining those ascriptions' presuppositions.

It is worth pausing on this point because a popular approach to theorizing about validity in languages with vocabulary that create presuppositions has the consequence that validities are not closed under classical consequence. Call a sentence *Strawson-valid* (von Fintel 1999) if it is guaranteed to be true if its presuppositions are satisfied. Now consider "The king of France is a king", "If the king of France is a king, then France has a king" and "France has a king": the first two sentences are Strawson-valid, and they classically imply the third sentence, but the third sentence is not Strawson-valid. However, notice that any classical consequence of some Strawson-valid sentences will itself be Strawson-valid provided that the sentence's presuppositions are at least as strong as the presuppositions of the sentences that classically imply it. Since the presuppositions of 4 are at least as strong as the presuppositions of 1 and 3 (given standard assumptions about presupposition projection) the pattern of presuppositions of sentences in our argument does not raise trouble for our appeal to the closure of validities under classical logic. We could also sidestep this issue by replacing 3 in our argument with "If the thought that Hesperus is Phosphorus is a true thought, then Hesperus is Phosphorus", since "is a true thought", unlike "know",

something false cannot be applied to 4. By his and Carnap's lights, Fregeans should deny that 4 is valid, since they should think that it can be used falsely without equivocating. So given our assumptions that 3 is valid and that classical consequences of valid sentences are themselves valid, Fregeans should also deny that 1 is valid.[3]

Some Fregeans might deny that 3 is valid, on the grounds that "knows" is not a 'logical constant'. But our mode of argument does not essentially rely on 3 being valid. Let *schmalidity* be that good status, however precisely it is understood, typified by 3 and at which systematic theorizing about knowledge aims. Assume (i) that all valid sentences are schmalid, (ii) that no sentences used falsely without equivocating are schmalid, and (iii) that the set of schmalid sentences is closed under classical propositional logic. Given that 3 is schmalid and 4 can be used falsely without equivocating, (i)–(iii) imply that 1 is not valid.

A common response to this argument has been to grant its conclusion but claim that it misses the point. According to this response, Fregeans should think that there is a theoretically important status that the schema Substitution enjoys despite having non-valid instances like 1. In particular, they should think that *weak validity* is such a status, where a schema is weakly valid just in case all of its *embedding uniform* instances are valid, and a sentence is embedding uniform just in case every expression occurring in it is embedded under the same number of attitude verbs in each of its occurrences. Since 1 is not embedding uniform, its invalidity does not threaten the weak validity of Substitution.

Footnote 2 continued

does not generate presuppositions. (Frege's own views about truth as a predicate of thoughts are complex, but in respects that do not threaten the modified version of our argument; see Heck and May (2018).)

[3] One might resist this argument by claiming that, despite the fact that every word in 4 would have the same reference in each of its occurrences if 4 were used to speak falsely, such uses would nevertheless involve a kind of equivocation because they would involve a mid-sentence shift in context. Compare "If today is Christmas, then tomorrow is boxing day": if the clock struck midnight on December 26th at the moment "then" was uttered during a use of this sentence, then the falsity of that use would involve a kind of equivocation (since it would turn on the fact that "today" and "tomorrow" were uttered in different contexts on different days) despite the fact that every word in the sentence would have the same reference in each of its occurrences. Here is one way in which this contextualist proposal might be fleshed out. Suppose that context supplies a function from entities to senses, and that, when a name is used embedded under a single propositional attitude verb (as in the consequents of 1 and 4) it refers to the sense assigned by context to its ordinary referent. On this view, when 1 and 4 are used to speak falsely, the occurrences of "Hesperus" and "Phosphorus" in "Hammurabi knew that Hesperus is Phosphorus" must be in different contexts, since they refer to different senses despite having the same ordinary referent.

This proposal is Kaplanian to the extent that it upholds the validity of 1 and 4 by convicting their false uses of a kind of equivocation, and Fregean to the extent that it deploys a sense/reference distinction in the semantics of propositional attitude ascriptions. But it does not invoke the sort of equivocation Kaplan had in mind, nor does it fit with Frege's way of thinking about the connection between sense and reference. Regarding Kaplan, note that the equivocation is not between two differently embedded occurrences of the same name (one of which refers to an ordinary object and the other to a sense) but rather between two occurrences of different names embedded under the same attitude verb. Regarding Frege, he did not think that the senses of proper names are context-sensitive in the way the proposal requires—he would not think that, as long as context is held fixed, there *is* a backward road from reference to sense. (Dorr (2014) argues that proponents of a Fregean semantics of attitude ascriptions should, *pace* Frege and most of his followers, accept something like the above contextualist proposal.)

We think that this response is half right. Denying that 1 is valid clearly does not bar Fregeans from accepting some version of the idea that true identities license the intersubstitution of their flanking terms, and the claim that Substitution is weakly valid is a natural way for them to make this commitment precise. But this is *not* because weak validity is a theoretically important status. It isn't. For consider the schema:

Anti-Factivity: If S knows that Φ, then not-Φ.

This schema is weakly valid, since it has no embedding uniform instances. But it is clearly not a 'good schema' in any interesting sense. The claim that Substitution is weakly valid is interesting because it is equivalent to the claim that Weak Substitution (the schema whose instances are all and only the embedding uniform instances of Substitution) is *valid*, and Weak Substitution is itself an interesting schema. By contrast, Weak Anti-Factivity (the schema whose instances are all and only the embedding uniform instances of Anti-Factivity) has no instances, and the claim that it is valid is therefore trivial. Fregeans (like everyone else) should consider all of a schema's substitution instances in assessing its good standing.

We have argued that Fregeans should think that 1 is not a valid sentence of English, and hence that Substitution is not a valid schema. The way in which, according to them, the two occurrences of "Phosphorus" in 1 have different referents is not the kind of equivocation capable of reconciling a false reading of a sentence with that sentence nevertheless being valid.[4] It is not like the failure of the two occurrences of "Jane" to co-refer on the relevant reading of 2.[5] This conclusion is not intended as a *criticism* of Fregean treatments of attitude ascriptions. We hope rather that it will encourage further investigation into the possibilities for systematic theorizing in settings where Substitution is given up.[6]

References

Bacon, A., & Russell, J. S. (2017). The logic of opacity. *Philosophy and Phenomenological Research*

---

[4] How should Fregeans characterize the validity-relevant notion of non-equivocation? The simplest proposal would be to count a reading of a sentence not containing quotation marks as equivocation-free just in case, for every elementary expression in the sentence, there is a sense such that every occurrence of the expression embedded under an attitude verb (or in some other 'indirect context') refers to that sense and every occurrence of the expression not embedded under an attitude verb refers to the referent determined by that sense. A different proposal (for those who think that iterated attitude ascriptions require a hierarchy of senses) would be to say instead that, for every elementary expression in the sentence, there is a hierarchy of senses $s_0, s_1, \ldots$ such that each s_n is the referent determined by s_{n+1} and every occurrence of the expression in the scope of n attitude verbs refers to the referent determined by s_n.

[5] One salient dissimilarity here is that the pattern of distinct referents of "Phosphorus" in 1 is conventionalized, obviating the need for the sort of explicit indication of the intended reading that we just made above regarding 2; see Kripke (2011 [2008], p. 262, f.31).

[6] For some recent work in this direction, see Bacon and Russell (2017) and Caie et al.(forthcoming).

Caie, M., Goodman, J., Lederman, H. (forthcoming). Classical opacity. *Philosophy and Phenomenological Research*.

Carnap, R. (1947). *Meaning and necessity: A study in semantics and modal logic*. Chicago: University of Chicago Press.

Dorr, C. (2014). Transparency and the context-sensitivity of attitude reports. In M. Garcia-Carpintero & G. Martí (Eds.), *Empty representations: Reference and non-existence* (pp. 25–66). Oxford: Oxford University Press.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik, 100*, 25–50. Translated as "On sense and reference" by Max Black in *Philosophical Review, 57*(3), 209–230 (1948).

Heck, R. K., & May, R. (2018). Truth in Frege. In M. Glanzberg (Ed.), *The Oxford handbook of truth* (pp. 193–217). Oxford: Oxford University Press.

Kaplan, D. (1968). Quantifying in. *Synthese, 19*, 178–214.

Kripke, S. A. (2011). Frege's theory of sense and reference: Some exegetical notes. In *Philosophical troubles* (pp. 254–291). Oxford University Press, 2011 [2008]. First published in *Theoria, 74*(3), 181–218.

von Fintel, K. (1999). NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics, 16*(2), 97–148.