Frege on Truth and Reference

PALLE YOURGRAU*

... I cannot help feeling that the problem raised by Frege's puzzling
conclusion has only been evaded by Russell's theory of descriptions and that
there is something behind it which is not yet completely understood.

Kurt Gödel, "Russell's Mathematical Logic"[1]

The History of the Argument   In "On Sense and Reference" Frege extended
his theory of the sense and reference of singular terms to cover as well concept-
expressions (or predicates) and sentences. He argued that in 'oblique' contexts,
typically governed by propositional attitude constructions, a sentence refers to
what is ordinarily its sense (which he called a 'thought'). Otherwise, he
maintained, a sentence refers to its truth-value. The latter fact is surprising in
at least three ways: (a) pre-theoretically one would not normally take sentences
as referring at all; (b) insofar as one accepted that sentences do refer, one would
hardly be inclined to take them as referring to their truth-values; and (c) given
the paradoxical character of (b), one would expect a far more convincing
argument in its favor than that in fact given by as rigorous a thinker as Frege.
To dwell on the latter consideration for a moment, it is well known that the
considerations actually advanced in "On Sense and Reference" in favor of his
thesis are so weak as to be hardly worth reproducing. And one looks in vain in
the rest of Frege's writings for an argument of sufficient persuasiveness to bear
the weight of his "puzzling conclusion".

In large part, perhaps, because of the latter fact, Frege's immediate suc-
cessors, Russell and Wittgenstein (in their atomistic youth), who rejected his doc-
trine and held that sentences refer to (or 'express' or 'indicate') *facts*, seem never
to have cast a backward glance at Frege's argument. The task of filling in the
embarrassing lacuna in the master's writings was finally taken up by his great
successor, Church, in whose review of Carnap there emerged a respectable ver-
sion of the argument (see [7]). Gödel repeated the favor ([13]). Once again,

---

however, the history of the argument took a curious turn. While some philosophers became strongly attached to the Church/Gödel/Frege argument (e.g., Quine [16] and Davidson [8], p. 753), others clung to the alternative doctrine of sentences referring to (or 'indicating') facts (e.g., van Fraassen [18]), this time refusing a backward glance at the argument even in its now respectable guise. To date, indeed, there still exists no satisfactory, brief, ideologically uncommitted discussion of the Frege argument.[2] In this essay I hope to fill this need, in a context free from ulterior motives as to how best to utilize sentence referents, if there be any, to solve this or that philosophical problem.

Why Look for Sentence-Reference in the First Place?   Pre-theoretically, matters seem thus: for singular terms, especially for the purest cases—proper names—that they *refer*, and that to which they refer, seems relatively clear. Contrariwise, for sentences, that they express a *sense* (i.e., a thought) is a familiar notion. What stands in need of elucidation is the idea that names have senses and that sentences refer. Church's 'truth and reference argument' can be seen as an attempt to present an account that will unify this double division, sense/reference and singular term/sentence. The intuitions just expressed can then be seen as minimum conditions of adequacy on any such theory. It must emerge that any singular term, e.g., "Kurt", refers to Kurt, and that any sentence, e.g., "Kurt is wise", expresses the sense (thought) that Kurt is wise. The developed account will then help us see what to take as the sense of a singular term and the reference of a sentence.

    For Frege, every significant semantic unit can have a reference—including singular terms, concept expressions, and sentences. If one reads for "reference" here, "semantic value", this may seem slightly more palatable, and perhaps even more so if we recognize that the *way* in which, for Frege, each of these three categories of expression refers to its appropriate referent is quite different.[3] Indeed, Russell held that sentences 'indicate' rather than refer to facts, and one can see this as a refinement, rather than abandonment, of Frege's idea. Further, as Church has emphasized ([6], p. 24), in treating sentences as singular terms only *insofar* as maintaining that they both have referents, one need not deny that sentences have also a function not shared by other singular terms—viz., they can be used to make assertions. As Church notes, when sentences are embedded in larger sentences, they occur nonassertively, and when they are used to make assertions we could, like Frege, mark this out by using the 'assertion-sign' (⊢).

    But the above points are mainly negative. More positively, one might cite Frege's Principle of Compositionality: the reference (sense) of a term (any term) is a function of the reference (sense) of its parts. *This* principle is highly intuitive. One might then suggest that for sentences that are truth-functional compounds of other sentences, the truth of the whole is a function of the truth of the parts. Does this support the thesis that sentences have referents? Unfortunately no, not unless one assumes that the *referent* of the containing sentence is precisely its truth-value. For unless one does so, he will have shown something about the compositionality of the truth-value of sentences, but not about the issue of sentence *reference*. One can do a little better than this, however. A Fregean might offer the following two arguments: (a) When a sentence follows, for example "believes", it serves to indicate or refer to the thought, or sense, that

the sentence expresses when it occurs alone; so at least here sentences refer. And if they refer here, why not when they occur in isolation? (b) Sentences sometimes are embedded in singular terms which themselves clearly *do* refer: for example: "$\imath x(x = \text{Reagan} \cdot P)$"; here the truth-value of "P" seems to contribute to the reference of the containing term; and if we accept the Principle of Compositionality, this suggests we take the truth value of "P" to be its referent.[4]

The above considerations will be found persuasive to some, unpersuasive to others. What is important about the Church/Frege argument, however, is that it remains of interest regardless of one's position on the question of sentential reference. Insofar as one has an interest in facts, or situations, or events (and many philosophers have such an interest), there remains the question of the relationship of these entities to sentences. If one believes in the fact that P, for example, it is hard to see how one can deny that "P" is true iff it is a fact that P. Whether one says, then, that "P" expresses this fact, indicates it, is true in virtue of it, or refers to it, is not in this context of crucial importance. What is important, though, is that if the Church/Frege argument is sound, it is very difficult to come up with a convincing theory of facts (or events, or situations), especially in regard to their relationship to sentences. It is worthwhile, therefore, from a broad philosophical perspective, to look closely at Church's version of Frege's argument.

The Church/Frege Argument   Church takes two (Fregean) principles as fundamental: **PI**, logically equivalent sentences are co-referential, and **PII**, the Principle of Compositionality. The following argument is a reconstruction and simplified version of the one Church offered in his review of Carnap [7]. We begin with any two true (false) sentences, P and Q, and show that they are co-referential ("coref" for short). The natural conclusion is that their reference is their truth-value.

1. P (assume)
2. Q (assume)
3. "P" L-equiv. to "$\phi = \imath x(x = \phi \cdot P)$" (logic)
4. "Q" L-equiv. to "$\phi = \imath x(x = \phi \cdot Q)$" (logic)
5. "P" coref. "$\phi = \imath x(x = \phi \cdot P)$" (3, PI)
6. "Q" coref. "$\phi = \imath x(x = \phi \cdot Q)$" (4, PI)
7. $\imath x(x = \phi \cdot P) = \imath x(x = \phi \cdot Q)$ (1, 2, logic)
8. "$\imath x(x = \phi \cdot P)$" coref. "$\imath x(x = \phi \cdot Q)$" (7, logic)
9. "P" coref. "$\phi = \imath x(x = \phi \cdot Q)$" (subst. "$\imath x(x = \phi \cdot Q)$" for "$\imath x(x = \phi \cdot P)$" in 5, given 8 and PII.)
10. "P" coref. "Q" (6, 9, trans. of coref.)

This argument is obviously formally valid, so that any critique must focus on PI and PII.[5] Before we engage in such a critique, however, it is worth noting that there is a slightly different version of this argument offered by Church elsewhere (see [6], pp. 24, 25) and also by Gödel ([13], p. 129).[6] In this version, in addition to PII, the assumption is made that (a) every sentence "P" is synonymous with a sentence of the form "Fa", and (b) any sentence "Fa" is synonymous with "$a = \imath x(x = a \cdot Fx)$". This alternative version is of interest because PI

is the principle that we will soon find unacceptable, while the alternative argument only needs PIII: Synonymous sentences are co-referential; and PIII is beyond dispute. To cast doubt on the revised version, then, one of the premises themselves must be rejected.

　　To turn our attention for the moment to this revised version of the argument, we note that indeed one of the premises is suspect — namely that "$F(a)$" is synonymous with "$a = \imath x(x = a \cdot Fx)$". Now, pre-theoretically, this premise may see quite plausible, and some might argue that it would have been accepted by Frege. In "On Concept and Object" [10], for example, he says that the sense of a sentence, a thought, may be broken up in many ways — with now one term serving as subject, now another. The issue is, however, somewhat delicate. Dummett, for example, speaks of the 'decomposition' vs. the 'analysis' of a thought (see [9]). In any event, recent theorists have realized that to have a plausible theory of sense and synonymy, we need a tighter criterion of sameness of sense — Carnap [3] used 'intensional isomorphism', Putnam [15] 'intensional-isomorphism-cum-logical-form', and Church himself 'synonymous isomorphism' in [5]. But the above proposed synonymy *fails* under each of these stiff versions. So, in the more precise development in recent years of a criterion of sameness of sense, the crucial premise of the revised argument must be rejected.[7] This still leaves us, however, with the primary version of the argument presented in full above. What, if anything, can be said in criticism of it?

The Flaw in the Argument　　　Some (e.g., Sober [17]) have challenged PII. Given the plausibility of PII we remarked on earlier, however, it would be preferable to put some pressure on PI instead. But PI too may seem at first to be sound. If two sentences are logically equivalent, then, to use a current image, they are true in exactly the same possible worlds. How can this be, one may well ask, unless they refer to the same things in every possible world in which they do refer? Nevertheless, PI is not an acceptable first principle, and Church's argument therefore fails. To see why PI is not an appropriate principle, note first that it practically begs the question at issue. Why should we believe that *logically* equivalent sentences are co-referential? Logical equivalence has to do with shared *truth*-values — but what has this got to do with sentence *reference*? Indeed, it is precisely the purpose of the whole Church–Frege argument to show us the relationship between truth and sentential reference, while PI assumes before the argument even begins a crucial, indeed a decisive, relationship between these two. For this reason alone PI is suspect as a premise for Church's truth and reference argument.

　　The question is: what are l-equivalent sentences, once we consider sentences to be a kind of singular term (at least insofar as they too are held to have referents)? It is reasonable to hold that true sentences refer to facts, leaving it open whether, as Frege (Church) in effect has it, there is one big fact — the true — which all true sentences refer to. Following Russell,[8] we can reasonably say that false sentences do not refer (to the facts). At a minimum, then, we can see that l-equivalent sentences are *l-together-referring* (whenever one refers, the other does). The question that remains is: are they *l-co-referential* (do they always have the same referent)? Church's argument provides an answer to this question, not in its conclusion, but as one of its premises. For PI in effect just states that l-

equivalent sentences, which at a minimum are 1-together-referring, are 1-co-referential. What was to be shown has been presupposed.

That 1-together-referring terms need not always be 1-co-referential is easily shown. Consider "the father of Kurt" and "Kurt, a child of the father of Kurt". Clearly these terms are 1-together-referring, but in no world are they co-referential. If the case of sentential singular terms is an exception, here, then we need to see the argument for this thesis. But, as we have seen, Church presupposes rather than argues for the special status of sentences. The Church–Frege argument, therefore, does not achieve its purpose.

Consequences of the Failure of the Argument   Where does this leave us? The failure of Church's argument means that if one is persuaded, by the reasons given earlier or by different reasons, that sentences do have a referent, the door is open to semantic engineering to come up with alternative sentential referents. The only purely semantic constraint on this enterprise is that the logical structure of the referents be isomorphic with the logical structure of the sentences. For example, one might choose as the referent of "Reagan is President" the ordered pair containing Reagan and the property of being President: $\langle$Reagan, $\ulcorner$is President$\urcorner\rangle$, and for "Reagan $= \imath x(x =$ Reagan $\cdot x$ is President)" choose the entirely different referent: $\langle$Reagan, $\ulcorner = \urcorner$, Reagan$\rangle$. We see here a clear mirroring of ontology in our logical grammar. This should come as no surprise—it was insisted on by the early logical atomists who countenanced a theory of facts. Whether or not one finds this result palatable will depend on broader philosophical considerations. The failure of Church's argument is a purely semantic result, and given only that one satisfy the condition of isomorphism, theorists who find a need for sentential referents are free to construct them to suit their ideological promptings. Note, however, that our results serve not (as in [1]) to (help) refute Frege's semantics, but rather to improve it. For (as I argue at length in [21]) a revised Fregean theory, in which sentences express thoughts and refer to 'facts', is stronger than ever (especially with the further 'improvements' suggested in [21]).

NOTES

1. In the present paper I discuss the Frege argument that Gödel recast in precise form and puzzled over. Gödel's dissatisfaction with Russell's approach stemmed in part from his misperception of the theory of descriptions. In particular, he maintained (see [13], 130) that for Russell definite descriptions don't really denote at all. Recently, Kaplan has renewed this interpretation of Russell (see [14]). I have criticized this approach to the theory of descriptions in [22].

2. The only possible exception to this claim is Barwise and Perry's in [1]. Though their analysis has strong affinities with ours (although our results were reached independently) there remain serious disagreements: (a) While Barwise and Perry view their results as (part of) a general refutation of Frege's semantics, ours are not seen in that light. (For a specific argument that Perry has not elsewhere refuted Frege, see [20], as well as my forthcoming [22] (in which I incorporate the results of the present paper with a 'new and improved' version of Frege's theory).) (b) I do not agree with Barwise and Perry that one can pinpoint the problem with the Church–Frege ar-

gument in its failure to pay sufficient heed to Donnellan's referential/attributive distinction for definite descriptions. (In fact, I am suspicious of the relevance of Donnellan's distinction for semantics. See my [22], note 1, for some thoughts on this question.)

3. For an extensive discussion of this point for the first two categories, see [12].

4. Lest one too hastily conclude from the fact that in "$\imath x(x = \text{Reagan} \cdot P)$" the truth-value of "$P$" contributes to the reference of the whole, that this truth-value is the *referent* of "P", consider the term: "$\imath x(x = \text{Reagan} \cdot (\exists x)(x = \text{Carter}))$". In the latter term, the referent of the whole remains invariant under substitution of any singular term for "Carter" that succeeds in referring to any (existent) object. But we should hardly conclude that "Carter" here refers to 'Existence', or some such object.

5. Cummins and Gottlieb in [4] deny that "P" is 1-equivalent with "$\phi = \imath x(x = \phi \cdot P)$", since the sentences have different ontological commitments. If necessary, therefore, we could restrict the argument to atomic sentences, "Fa", since "Fa" is 1-equivalent to "$a = \imath x(Fx \,.\, x = a)$" (this is in effect what Gödel did; see [19]). Or we could leave the original argument alone, and adopt some form of 'free logic', where identities like "$\phi = \imath x(x = \phi)$" are preserved, whether or not "ϕ" succeeds in denoting an object in the universe of discourse.

6. Church says more precisely that "$F(a)$" and "$(\exists x)(x = a \,.\, F(x))$" are, if not synonymous, close enough to synonymy to ensure coreferentiality. And Gödel says that the two sentences "mean" the same thing, where it appears that this signifies synonymy.

7. Part of the problem here arises from the multiple roles Frege expected from his notion of thought: (a) on the one hand, thought captures the information value of a sentence (in which case Frege's observations in "On Concept and Object" seem plausible); (2) on the other hand, insofar as a thought is the sense of a sentence and so represents *belief* contents, it must be sensitive to the discriminatory powers of the believer (and the Carnap/Putnam/Church refinement seems necessary). Finally, (3) a thought is the primary bearer of truth-value (see [11]). This final aspect of thought may explain Frege's otherwise too easy acceptance of his argument that a sentence refers to its truth-value. For more on the many roles of Fregean sense, see Burge's [2] and my [20], note 5, and forthcoming [22].

8. See [13], p. 129: "So, according to Russell's terminology and view, true sentences "indicate" facts and correspondingly false ones indicate nothing".

REFERENCES

[1] Barwise, J. and J. Perry, "Semantic innocence and uncompromising situations," pp. 387–404 in *Midwest Studies in Philosophy*, vol. VI, *The Foundations of Analytic Philosophy*, ed., P. A. French et al., University of Minnesota Press, Minneapolis, 1981.

[2] Burge, T. "Sinning against Frege," *The Philosophical Review*, vol. 88 (1979), pp. 398–432.

[3] Carnap, R., *Meaning and Necessity*, University of Chicago Press, Chicago, 1956.

[4] Cummins, R. and D. Gottlieb, "On an argument for truth-functionality," *American Philosophical Quarterly*, July 1972, pp. 265–269.

[5] Church, A., "Intensional isomorphism and identity of belief," *Philosophical Studies* (Oct. 1954), pp. 65–73.

[6] Church, A., *Introduction to Mathematical Logic*, Princeton University Press, Princeton, New Jersey, 1956.

[7] Church, A. J., "Review of *Introduction to Semantics*," pp. 298–304 in *The Philosophical Review*, vol. LII, no. 3, 1943.

[8] Davidson, D., "True to the facts," *The Journal of Philosophy* (1969), pp. 748–764.

[9] Dummett, M., "An unsuccessful dig," *Philosophical Quarterly*, vol. 34, no. 136 (1980), pp. 377–401.

[10] Frege, G., "On concept and object," pp. 42–55 in *Selections from the Writings of Gottlob Frege*, eds., P. Geach and M. Black, Rowman and Littlefield, Totowa, 1980.

[11] Frege, G., "The thought," pp. 17–38 in *Philosophical Logic*, ed., P. Strawson, Oxford University Press, Oxford, 1967.

[12] Furth, M., "Two types of denotation," pp. 9–45 in *Studies in Logical Theory*, ed., N. Rescher, *American Philosophical Quarterly*, Monograph Series, #2 (Oxford, Basil Blackwell, 1968).

[13] Gödel, K., "Russell's Mathematical Logic," pp. 125–153 in *The Philosophy of Bertrand Russell*, ed., P. A. Schilpp, The Library of Living Philosophers, Evanston, Illinois (New York: The Tudor Publishing Company).

[14] Kaplan, D., "What is Russell's theory of descriptions?" pp. 277–288 in *Physics, Logic, and History*, eds., W. Yourgrau and A. D. Breck, Plenum Press, New York, 1970.

[15] Putnam, H., "Synonymy and the analysis of belief sentences," *Analysis* (1954), 14, no. 5, pp. 114–122.

[16] Quine, W. V. O., "Three grades of modal involvement," pp. 158–176 in *Ways of Paradox*, Cambridge University Press, Cambridge, 1979.

[17] Sober, E., "Sentence reference, aboutness, and the ontology of science," unpublished.

[18] van Fraassen, B., "Facts and tautological entailments," *The Journal of Philosophy*, vol. XVI, no. 15 (1969), pp. 477–487.

[19] Widerker, D., "The extensionality argument," *Noûs* (Sept. 1983), pp. 457–468.

[20] Yourgrau, P., "Frege, Perry and demonstratives," *Canadian Journal of Philosophy*, vol. XII, no. 4 (1982), pp. 725–752.

[21] Yourgrau, P., "The Path Back to Frege," forthcoming (December, 1986), Proceedings of the Aristotelian Society.

[22] Yourgrau, P., "Russell and Kaplan on denoting," *Philosophy and Phenomenological Research* (Dec. 1985), pp. 315–321.

Department of Philosophy
Barnard College
Columbia University
New York, NY 10027-6598