Copyright © 1980 by Terence D. Parsons

Frege’s Hierarchies of Indirect Senses and the Paradox of Analysis
TERENCE D. PARSONS

One of the primary concerns of analytic philosophy has been with meaning and the analysis of meaning. Although much of this discussion is carried out informally, one often gets the impression that there is a fairly well worked out doctrine, or theory of meaning, lurking in the background. Frege’s theory of sense and reference is one such doctrine. This is a rich and powerful theory which deals in some detail with meanings as language-independent entities and with the relations between these entities and language. It also has certain peculiarities of its own—particularly the doctrine of indirect sense and reference—which set it apart from other theories of meaning, like Russell’s. Some have thought that this doctrine offers a unique and compelling solution to a central problem of analytic philosophy: the paradox of analysis. The goal of this paper is to articulate and assess Frege’s doctrine of indirect sense and reference and its relevance to the paradox of analysis. I will suggest that the best solution to the paradox of analysis is not one that employs the doctrine of indirect sense and reference.

1. FREGE’S THEORY: WHAT IS IT?

According to Frege, words have both sense and reference. Ordinarily a word expresses its customary sense and refers to its customary reference. But in certain contexts a word refers to its customary sense; these are called indirect contexts, and they include contexts like ‘Samantha believes that ____’ and ‘It is possible that ____’. Since (in any context) the reference of a word depends on what sense it expresses (in that context) a word in an indirect context expresses a sense that is different from its customary sense (it cannot express its customary sense because then it would refer to its customary reference, which it does not). This new sense is called its “indirect sense” [(S&R)].

Frege himself thought of this shift of reference as a defect of natural language, and he proposed that in a logically correct language different signs should be used
inside and outside indirect contexts, instead of a single sign which changes reference. The most impressive developments of Frege's theory of sense and reference, particularly Church [FLSD, RF] and Kaplan [FIL], follow this line of investigation—they explore the way language should work. The present paper is directed at the issue of how language does work.

According to many authors, the invocation of new senses in indirect contexts is just the first step toward an infinite hierarchy of senses for each word or phrase. On this view, if a sentence containing a word in an indirect context is itself embedded in an indirect context, then the word (which now occurs in an indirect context inside another indirect context) refers to its indirect sense, and so it expresses a new sense, its doubly indirect sense. Another reembedding requires a triply indirect sense, and so on.

This version of the theory of sense and reference is attributed to Frege by Carnap [M&N], Dummett [F], and others. It has such widespread identification with Frege's theory that it deserves to be called the "orthodox" theory; I will call it this. Ironically, the classical source for this theory is the essay "On Sense and Reference" [S&R], and the orthodox theory is at best an extension of the theory presented there. In fact, if [S&R] is read quite literally, the theory given there attributes to words exactly two senses: customary and indirect. I will refer to this as the "two-level" theory.

What theory is correct? Dummett ([F], pp. 267-68) objects to the orthodox theory and holds that Frege ought to have endorsed a theory that attributes to words only customary sense. Carnap claims to be able to show the two-level theory inconsistent ([M&N], p. 131). In this paper I hope to establish that the two-level theory is consistent and that it is equivalent (in a sense to be specified) to the one-level theory which Dummett says Frege should have adopted—a theory that is also equivalent to (at least one interpretation of) a theory that Carnap himself gave in [M&N]. Second, every orthodox (infinite-level) theory either is equivalent to the one-level theory or is not an adequate theory of language. (This last point will involve a discussion of the paradox of analysis.)

2. A MOMENTARY DIGRESSION

Before proceeding to a consideration of the point at issue, I want to distinguish it from a different though related claim. Carnap argues (correctly) that Frege is committed to a certain sort of infinite hierarchy associated with each word ([M&N], p. 130). Frege holds that to each sense there corresponds at most one entity; any word that expresses a sense (in some context) must refer (in that context) to that corresponding entity. I will use Frege's word 'present' to stand for the correspondence in question: i.e., I will say that each sense presents at most one entity. (My 'presents' is the same as Church's 'is a concept of'; I avoid Church's terminology because Frege used 'concept' to stand for a different notion.) Now consider the infinite list of phrases:
(1.1) the evening star
(1.2) the (customary) sense of 'the evening star'
(1.3) the (customary) sense of 'the (customary) sense of 'the evening star''

Our understanding of English requires that phrase \((1. n + 1)\) refer to the (customary) sense of phrase \((1.n)\), for each \(n\). Thus, the sense of phrase \((1.n + 1)\) must present the sense of phrase \((1.n)\). And this yields an infinite chain of senses, each related to the previous one by the presentation relation, and beginning with the sense of 'the evening star'.

Carnap found this consequence objectionable, though I do not know why. In fact, if the word 'sense' is replaced by 'intension' in the phrases above, you get a similar chain of intensions within Carnap's own theory. In any event, this issue is a different one than the one being discussed in this paper. The issue I want to discuss is whether each word itself has an infinite number of senses, each of which gets expressed by that word in some context. In the example above there is no reason to believe that any one word or phrase expresses all the senses in the infinite chain. There is no reason, for example, to think that the sense expressed by 'the evening star' in single indirect contexts is the same as the customary sense of 'the sense of 'the evening star'',' even according to the orthodox interpretation of Frege's theory.

3. THE ORTHODOX INTERPRETATION: SOME NOTATION

Here is the orthodox interpretation. Let us abbreviate 'the customary sense of \(w\)' by \('s_1 [w]\)', and 'the customary reference of \(w\)' by \('r_1 [w]\)'. In a singly embedded context—i.e., when the word is within an indirect context but is not in any indirect context which is itself in an indirect context—the word expresses a new, singly indirect sense, \(s_2 [w]\), and refers to a new, singly indirect reference, \(r_2 [w]\). The relations among these entities are as follows:

(2) Those due to the standard relation between the sense of a word in a given context and its reference there,
    namely: \(r_1 [w] = \text{the entity presented by } s_1 [w]\)
    and: \(r_2 [w] = \text{the entity presented by } s_2 [w]\).

(3) The principle of indirect reference:
    \(r_2 [w] = s_1 [w]\).

So far, this is Frege's view in [S&R], no matter how many levels of sense there are. The distinctive part of the orthodox interpretation comes next: namely, that when a sentence containing an indirect context is embedded in another indirect context, a similar sort of "elevation" of senses and references takes place. In a single such reembedding the word refers to its singly indirect sense:

(4) \(r_3 [w] = s_2 [w]\)
and there is a new sense, \( s_3[w] \), where

\[
(5) \quad r_3[w] = \text{the entity presented by } s_3[w].
\]

In general the view is that embedding an \( n \)-ly embedded context in a singly embedded context produces an \( n + 1 \)-ly embedded context in which the analogues of (4) and (5) hold, namely:

\[
(6) \quad r_n[w] = \text{the entity presented by } s_n[w], \quad \text{for all } n.
\]

\[
(7) \quad r_{n+1}[w] = s_n[w], \quad \text{for all } n.
\]

Further, \( s_n[w] \neq s_m[w] \) whenever \( n \neq m \), so each word actually expresses (in various contexts) an infinite number of different senses.

### 4. The Literal Interpretation

According to the orthodox interpretation, when a word is placed in a doubly embedded context, it refers to its (singly) indirect sense and expresses a new sense. This view is not stated in [S&R]. Instead we have the following quotation ([S&R], p. 59):

In reported speech one talks about the sense, e.g., of another person's remarks. It is quite clear that in this way of speaking words do not have their customary reference but designate what is usually their sense. In order to have a short expression, we will say: In reported speech, words are used indirectly or have their indirect reference. We distinguish accordingly the customary from the indirect reference of a word; and its customary sense from its indirect sense. The indirect reference of a word is accordingly its customary sense.

Now in sentence (8):

\[
(8) \quad \text{Mary said that John said that the evening star is a planet,}
\]

the phrase 'the evening star' occurs in reported speech. Frege said that in reported speech a word refers to its customary sense. So in (8) the phrase 'the evening star' apparently refers to its customary sense. This contradicts the orthodox view, which says that in (8) that phrase refers to its indirect sense. The orthodox view requires that in doubly embedded contexts we ignore what is said in [S&R].

Actually, I do not think that there is any strong reason to take the [S&R] statement of the theory literally. It seems plausible to me that when Frege stated his principle of indirect reference in [S&R], he just was not thinking of doubly embedded contexts, and it is really up to us to extend the theory he had in mind to account for such contexts. The orthodox interpretation gives a very elegant extension, an extension which Frege himself adopted ten years later in a letter to Russell ([F-R], p. 236). But in the next section I will explore another extension, the extension that you get by taking Frege's literal statement as a perfectly general principle: in any indirect context a word refers to its customary sense.
5. THE LITERAL (TWO-LEVEL) THEORY SPELLED OUT

In the Appendix to this paper I give a syntax and semantics for a formal system which incorporates intensional operators: phrases like ‘Samantha believes that _____’, ‘Herman said that _____’, ‘It is possible that _____’. The present section illustrates that theory and states some conclusions concerning it.

Let me use ‘R’ to stand for ‘Radium is harmless’, ‘B’ for ‘Madame Curie once believed that’, and ‘S’ for ‘History books say that’. Each of these phrases is complex, and in a complete Fregean theory this complexity would be dealt with; but it is irrelevant to the present issue and so I will ignore it here.

The semantics of the sentence R, occurring in isolation, is summed up in the following diagram:

```
#1  s1[R]   
    ↑    
    R    
    ↓    
  r1[R]  
```

The arrow pointing up represents the relation “expresses”; that pointing down represents the relation “refers to.” So the diagram indicates that the sentence occurring in isolation refers to its customary reference, r1[R] (a truth-value), and expresses its customary sense, s1[R] (a Thought).

The semantics of ‘Madame Curie once believed that radium is harmless’ is given by:

```
#2  s1[BR]  
    s1[B](s2[R])  
      \   /  
    BR     
      \  /  
  r1[B](s1[R])  
    \  /  
  r1[BR]  
```

Since B is not itself in an indirect context, it expresses its customary sense and refers to its customary reference. Since R is in an indirect context here, it refers to its indirect reference—i.e., its customary sense—and expresses its indirect sense. The reference of the whole sentence (a truth-value) is got by applying the reference of B (which is a function) to the reference of R (which is, in this context, a Thought); this is the significance of the curly brackets. I.e., $Z = X(Y)$ indicates that $Z = X(Y)$. Likewise, the sense expressed by the whole sentence (a Thought) is got by applying the customary sense of B (a function) to the indirect sense of R.

Everything that has been said so far is consistent with both the orthodox and the literal interpretation of [S&R]. The divergence comes at the next step. On the literal interpretation, the semantics of ‘History books say that Madame Curie once believed that radium is harmless’ is:
This diagram is got by a straightforward application of the principle that the sense (reference) of a whole is determined by a function-argument combination of the senses (references) of its parts, plus these two principles:

(9) any word or phrase that is not in an indirect context refers to its customary reference and expresses its customary sense,

(10) any word or phrase that is in an indirect context refers to its indirect reference (i.e., its customary sense) and expresses its indirect sense.

The orthodox interpretation would differ, in this case, only in having R refer to \( s_2 \) [R] and express \( s_2 \) [R]; the rest of the diagram would be exactly the same.

Now a surprising thing happens. According to the "expresses" side of the second diagram, the customary sense of BR is got by applying the customary sense of B to the indirect sense of R:

\[
(11) \quad s_1 [B] (s_2 [R]) = s_1 [BR]
\]

But according to the "refers to" side of the third diagram, the customary sense of BR is got by applying the customary sense of B to the customary sense of R:

\[
(12) \quad s_1 [B] (s_1 [R]) = s_1 [BR]
\]

So \( s_1 [B] \) maps both \( s_1 [R] \) and \( s_2 [R] \) to the same thing, namely \( s_1 [BR] \). This is not an inconsistency, but it is a surprise, and surprises of this sort often suggest that there may be an inconsistency lurking somewhere around. This is especially disturbing since Carnap ([M&N], p. 131), for example, claims to be able to prove that in Frege's theory a word has to express an infinite number of senses, a thesis that is inconsistent with the literal interpretation. (Carnap makes it clear that he is not confusing this claim with the irrelevant one mentioned in Section 2.) And many other authors argue to the same conclusion.

In fact, as the Appendix makes clear, this theory is not inconsistent. Carnap did not publish his proof, so I can only speculate about where and how it might be inadequate. And the argument by Dummett ([F], p. 267), for example, presupposes the orthodox interpretation of Frege's theory, so it is not relevant here. The import of (11) and (12) is not that the theory contains a hidden inconsistency...
but that there is a kind of triviality or redundancy in the notion of indirect sense (in this version of the theory). In anthropomorphic terms, (11) and (12) jointly show that the customary sense of B cannot tell the difference between the customary and the indirect sense of R. And this is just one instance of a general phenomenon: whenever a sense gets to "look at" both the customary and the indirect sense of a given word, it cannot tell them apart. More literally, whenever f is a sense that is also a function, if f is ever applied to the indirect sense of a word in the semantical analysis of a sentence, then f maps that indirect sense to the same thing to which it maps the customary sense of the word. And this suggests that the theory we are discussing is essentially a variant of a theory with only one level of sense. Indirect senses are almost just customary senses in disguise. They have to be literally different from customary senses because they have to present different references—but aside from this there is no difference in the way they work in the theory. This suggests that if we could merely alter that part of the theory which identifies the reference of a word (in a context) with the entity presented by what the word expresses (in that context), we could simply identify the indirect sense of a word with its customary sense.

In fact, we can do just that. Suppose we call the theory sketched above 'F2' (for 'Frege: 2 levels of sense'). We can convert F2 into a quasi-Fregean theory with only one level of sense, F1, by making the following alterations:

(13) a word always expresses its customary sense,
(14) a word in an indirect context refers to its customary sense; otherwise it refers to the entity that its customary sense presents.

F1 is the theory that Dummett says Frege should have given ([F], pp. 267-68). It is equivalent to F2 in the following sense: on at least one plausible model of what senses are, every isolated sentence expresses the same thought according to F2 as it expresses according to F1. (See Appendix for details. The proof that the two-level theory is consistent depends on an appeal to "nonactual situations," but the techniques employed are easily generalizable to other sorts of models—in particular, the techniques of Section 7 also apply.)

Carnap objected to Frege's notion of indirect reference. Roughly speaking, he held that a word ought to keep both its customary sense and customary reference in all contexts; where "indirect" contexts are concerned, we should simply give up the view that the references of the parts determine the reference of the whole. Suppose then that we alter F2 as follows:

(15) a word or phrase always expresses its customary sense and refers to its customary reference.
(16) if X is an "indirect" context, then the reference of X(Y) is the result of operating on the sense of Y by the reference of X, and the sense of X(Y) is the result of operating on the sense of Y by the sense of X.

Call this theory "C1" (For "Carnap: 1 level of sense"). Then C1 is equivalent to both F2 and F1 in the sense described above.
6. CRITIQUE OF THE ORTHODOX THEORIES

The orthodox theories attribute to each word an infinite sequence of senses; a word expresses the nth sense in its sequence when it is within n − 1 embeddings of indirect contexts. Theories of this sort fall into two kinds, depending on the relationship between the customary sense of an expression (the first sense in its associated sequence) and its indirect sense (the second sense in its associated sequence). The one kind of theory, which I will call "rigid," holds that the customary sense of an expression uniquely determines its indirect sense. That is, it holds that any two expressions that have the same customary sense also have the same "first-level" indirect senses. (It follows from this that any two expressions that share a customary sense also share their nth-level sense, for any n, and are thus interchangeable in all contexts without altering the sense of the whole.)

The other sort of (orthodox) theory I will call "libertine"; in this kind of theory the customary sense of an expression does not determine what the indirect sense of that expression will be, and two expressions may have the same customary sense while diverging in nth-level indirect sense, for some n. Expressions with the same customary sense will not in general be interchangeable while preserving the sense of the whole.

My discussion of the orthodox theories is aimed at making two points: rigid theories are equivalent to simpler theories that attribute to each expression only a single sense, whereas libertine theories do not have any interesting application to language.

In what follows I will use a specific language as a basis for evaluating the orthodox theories. It is as simple as possible, consistent with the principle of including all the sorts of contexts that are relevant to the issues to be discussed.

The language L consists of the following:

1. Names: \(a_1, a_2, a_3, \ldots\)
2. One-place predicates: \(P_1, P_2, P_3, \ldots\)
3. Extensional sentence operators: \(E_1, E_2, E_3, \ldots\)
4. Indirect sentence operators: \(O_1, O_2, O_3, \ldots\)

The sentences of L include any expression of the form \(P_{n} a_m\), and if A is a sentence, so are \(E_n(A)\) and \(O_n(A)\). An expression is said to be in an nth level indirect context just in case there are exactly n indirect sentence operators preceding it. For example, \(a_3\) is in a second level indirect context in \(O_1(E_2(O_2(P_3 a_3)))\).

I assume that associated with each simple sign, \(A\), of L is a sequence of senses \(s_1[A], s_2[A], s_3[A], \ldots\), and that \(s_{n+1}[A]\) presents \(s_n[A]\), for each n. All senses of predicates and operators are functions. Complex expressions also have sequences of senses associated with them, determined by the following rules:

1. \(s_n[P_{m} a] = s_n[P_m](s_n[a])\).
2. \(s_n[E_m(A)] = s_n[E_m](s_n[A])\).
3. \(s_n[O_m(A)] = s_n[O_m](s_{n+1}[A])\).
Reference is determined in the usual Fregean manner:

(iv) \( r_n[A] = \) the unique thing presented by \( s_n[A] \).

I assume that all of this is hooked up to language in the orthodox Fregean way, namely, that an expression, \( A \), that is not in an indirect context expresses \( s_1[A] \) and refers to \( r_1[A] \), and that an expression, \( A \), that is in an \( n \)-th level indirect context expresses \( s_{n+1}[A] \) and refers to \( r_{n+1}[A] \).

7. RIGID THEORIES

A rigid theory can be converted into a theory that associates with each expression only one sense, as follows. We define "the sense of \( A \)," i.e., \( \text{s}[A] \), as follows:

(i) If \( A \) is a name or a sentence or a predicate or an extensional sentence operator, then \( s[A] = s_1[A] \).

(ii) If \( A \) is an indirect sentence operator, then \( s[A] \) = that function which maps an arbitrary sense \( x \) to \( s_1[A] (s_2[B]) \), where \( B \) is an expression such that \( s_1[B] = x \). (If there is no expression \( B \) such that \( s_1[B] = x \), then let the function map \( x \) to some arbitrarily chosen object.)

The choice of \( B \) in clause (ii) need not be unique, for in rigid theories it will not matter which \( B \) we pick.

Now it can be shown (by induction on \( n \)) that:

For any sentence of \( L \) of the form \( A_1(A_2(\ldots A_n(Pa)\ldots)) \):
\[
s[A_1(A_2(\ldots A_n(Pa)\ldots))] = s[A_1] (s[A_2] (\ldots s[A_n] (s[P] (s[A]) \ldots))).
\]

That is, the sense of a whole is always a function of the senses of the parts (taken in the natural order). And since for any sentence \( A \), the sense of \( A = \) (by definition) \( s_1[A] \), this simplified theory attributes to every isolated sentence exactly the sense that that sentence expresses according to the original theory. It is also clear that if \( A \) and \( B \) have the same customary sense according to the original theory, they have the same sense according to this simplified theory, and, by the result above, they are interchangeable in all contexts without alteration of the sense (= the customary sense) of the whole.

I still have not given the "reference" side of this theory. There are two natural ways to do this. One is this:

(R1) An expression \( A \) in an extensional context refers to \( r_1[A] \), and in an indirect context (of any level) refers to \( s[A] \).

Thus fleshed out, this simplification of the rigid theory turns out to be the theory (F1) that Dummett says Frege should have given.

The other natural option is:

(R2) \( A \) always refers to \( r_1[A] \).

This variant is our (C1) of Section 5.
8. Libertine Theories; The Paradox of Analysis

Libertine theories permit phrases to agree in customary sense while diverging in higher-up sense. Does language work this way? Should it?

I know of no argument to the effect that language should work this way. Investigators who work on "logically correct" or "canonical" symbolisms typically do not address this issue at all—they tend to follow Frege's advice that in logically correct languages no word or phrase would ever change reference with context. I myself can think of no reason to want a language to behave in the libertine fashion unless there are things we want to say that cannot be said well in any other way. The place to look for what we want to say is probably in what we do say, so let me turn to this issue.

Does actual language behave in accordance with the libertine theory? There is at least one published argument to the effect that it cannot. This is Davidson's argument that if language were correctly described by a libertine theory, it would not be humanly learnable, since the sense of a given word or phrase in a high-level indirect context would not be predictable from its basic sense (this is a rough statement).

There is hardly any published argument that language does behave in accordance with the libertine theory. However, there seems to be a very widespread informal view to the effect that language does work in this way and that the paradox of analysis illustrates a linguistic phenomenon which can only be accommodated (within a Fregean framework) by a libertine theory.

I will focus on what I take to be the most popular and also most plausible version of the argument. It begins by considering:

(17) The concept brother = the concept male sibling.
(18) It is trivial that the concept brother = the concept brother.
(19) It is not trivial that the concept brother = the concept male sibling.

The claim is that (17)-(19) are all true, that in (17) the phrases of the form 'the concept φ' refer to the customary sense of φ, but that in (18) and (19) they refer to the indirect sense of φ. Since (17) is true, 'brother' and 'male sibling' have the same customary sense; they are not intersubstitutable in (18) or (19), and this is explained by the fact that in (18) and (19) they refer to their indirect senses, which are distinct. So the argument goes.

I think that this simple format conceals a host of issues; I will discuss them in stages. The first thing to be clarified is whether 'brother' is actually a semantically relevant constituent of 'the concept brother' or not. Italics are commonly used for mentioning words, and one way to understand (17)-(19) is to suppose that they are shorthand for:

(17) The concept ordinarily expressed by 'brother' = the concept ordinarily expressed by 'male sibling'.
(18) It is trivial that the concept ordinarily expressed by 'brother' = the concept ordinarily expressed by 'brother'.

(19) It is not trivial that the concept ordinarily expressed by 'brother' = the concept ordinarily expressed by 'male sibling'.
(19') It is not trivial that the concept ordinarily expressed by 'brother' = the
concept ordinarily expressed by 'male sibling'.

Read in this way, (17')-(19') are plausible, but it is not at all clear that they have
anything to do with our issue at all. For the words 'brother' and 'male sibling' are
mentioned, not used, and their indirect senses are not referred to in (18') and (19').
I am not supposing that anyone was ever confused on this issue, but (17')-(19') do
give one natural way to mentally process (17)-(19). So let me just remind the reader
to take care not to read (17)-(19) in this manner, and proceed.

Next, there are two (other) ways to interpret phrases like 'the concept
brother'. The first is a sort of "ordinary-language" way—though more properly I
should say "ordinary philosophical usage" way. Concepts are entities that are ap-
parently familiar to many working philosophers; they have a long and venerable
tradition, and within this tradition (17) is a paradigm example of concept identity.
I.e., (17) is not the sort of claim that is normally open to question—to deny it is
to give evidence of not understanding the concept "concept."

But if 'concept' is understood in terms of its popular philosophical tradition,
there is no prima facie reason to think that (17) has anything to do with whether
the customary senses of 'brother' and 'male sibling' are the same. It could easily
be that 'brother' and 'male sibling' are genuinely semantical parts of 'the concept
brother' and 'the concept male sibling', that 'brother' and 'male sibling' have dif-
ferent customary senses, and that (17) be true. This would simply mean that
(customary) senses are individuated more finely than the concepts of philosophical
tradition. It would also imply that much of the literature on the subject of the
paradox of analysis is simply irrelevant to the issue under consideration. I think
that this is the case. (Recall that necessary equivalence is often cited as a criterion
of concept identity. But in a Fregean theory, customary senses must be individuated
more finely than this, for otherwise necessarily equivalent sentences would always
be intersubstitutable in first-level belief contexts.)

The remaining way to understand the construction 'the concept Φ' is to sup-
pose that Φ forms a genuine constituent of the construction and that it does refer
to the sense of Φ (either because traditional concepts are senses, contrary to the
suggestion of the last paragraph, or because for present purposes we stipulate that
'concept' is to be understood in this way). The idea would be that 'the concept
_____' is an indirect-context forming operator, which refers to that function
which maps any given sense to itself, and which expresses a function that also maps
any given sense to itself. Then within any given n-level indirect context, 'the concept
Φ' would refer to s_n[Φ] and would express s_n + 1[Φ]. I see no reason why we
could not have a piece of language that does this, and if we understand 'the concept
_____' in this manner, then the paradox of analysis, as embodied in (17)-(19), is
directly relevant to the issue we are considering: whether 'brother' and 'male sib-
ling' have the same customary sense but different indirect senses. So let me urge the
reader to understand (17)-(19) with this (perhaps highly artificial) stipulation in
mind, and I will proceed from there.
I would now like to suggest that, so understood, (17) is just plain false. Recall that (17) is true just in case this is true:

(20) 'Brother' has the same customary sense as 'male sibling'.

But 'customary sense' is a technical term from a technical theory of meaning. It is entirely inappropriate to test the truth of (20) by direct appeal to intuition. Words have the same customary sense in the theory under consideration only if they are interchangeable \textit{salva veritate} in all first-level indirect contexts (or, interchangeable \textit{salva sense} in all extensional contexts).\(^8\) Whether this is true may or may not be testable, case by case, by appeal to speaker intuition. But (20) is not so testable, for we are not native speakers of that fragment of our language which contains 'customary sense' as a part. So the ordinary defense of (17) in terms of its being 'obviously true' is now beside the point.

(20) (and thus (17)) is false; it is false because 'brother' and 'male sibling' are not interchangeable \textit{salva veritate} in the context:

(21) Herman is certain that all and only brothers are ______'s.

This seems to me to be decisive. Uncertainty about whether all and only brothers are male siblings is a common phenomenon, experienced by practically everyone who first encounters the claim; these are typically people who are certain that all and only brothers are brothers. Other examples which make the same point are: 'Herman argued (claimed, proved) that all brothers are ______' and 'Herman wondered whether all brothers are ______'. I find these examples convincing, but I am aware that others do not, and so the rest of this section is devoted to a discussion of challenges to this conclusion.

First, some would claim, e.g., that anyone who argues that all and only brothers are male siblings is ipso facto arguing that all and only brothers are brothers since the latter is simply saying the same thing as the former, just in different words. Although Herman might insist that he was not arguing the 'trivial' claim that all and only brothers are brothers, we cannot take what he says at face value. He is arguing this, but just does not realize that he is.

I think that this challenge has some plausibility, but it gains its plausibility from a source that bypasses the point at issue. It is this: when we report on someone else's attitudes or beliefs, or on what they have asserted, we customarily take a certain amount of liberty in paraphrasing what they explicitly feel, believe, or say. For example, Herman says his knee hurts, and we report that he said his knee is bothering him. Normally no one objects, because the difference does not matter. But if Herman cares about the difference, he might protest that he never said his knee was bothering him; he said only that it hurts. Maybe he was giving an example of something that hurts without bothering, and the paraphrase does not do justice to his point. But whether it does or not, he is correct in denying that he said his knee bothers him.

The point is this: "close" paraphrases of statements are usually acceptable, even though literally incorrect. And phrases as close in meaning as 'brother' and
'male sibling' are often close enough to provide acceptable paraphrases of one another for ordinary purposes. But acceptability is relative to the context and purpose, and in some situations 'brother' and 'male sibling' are not close enough paraphrases to pass. Some of these situations will involve uses of (21).

A second challenge involves the claim that although 'brother' and 'male sibling' are not interchangeable in (21), this is because (21) is a disguised quotational context. The most sophisticated version of this claim goes as follows: "Although we are sometimes inclined to accept the interchangeability of 'brother' and 'male sibling' in (21), we also sometimes tend to reject it. The explanation for this is that in the latter case we instinctively treat (21) as implicitly having the form:

(22) Herman is certain that 'all and only brothers are male siblings' is true.

In all non-quotational versions of (21), the interchangeability is acceptable."

I do not know of any way to refute a claim that a context is covertly quotational. It is easy to say, and, in interesting cases, hard to argue either way. It seems clear that here is a situation in which the hard data about language is inadequate to decide the issue, and that considerations of theoretical simplicity and elegance come into play. Some, e.g., Quine, would have us think that all non-extensional contexts are implicitly quotational.

But perhaps we can decide the case at hand without having to face the general issue of how to tell quotational from non-quotational contexts. First, we are examining the viability of a libertine Fregean theory of language. If all non-extensional contexts are covertly quotational, then the theory is not viable, not because of falsity but because of vacuity. So I will assume that at least some non-extensional contexts are to be treated by the theory of indirect sense. Let me suppose even that the challenger is right in thinking that (21) is ambiguous, sometimes being an inducer of indirect contexts and sometimes being an inducer of quotational contexts. But if this is plausible, it is at least as plausible for sentences (17)-(19). That is, if we want to explain the failure of interchangeability of 'brother' and 'male sibling' in:

(23) It is trivial that the concept brother = the concept _____.

then it is just as plausible (perhaps even more so) to allude here to implicitly quotational readings as it is in (17) or (21). The trouble with the second challenge, in short, is that it does not show a difference in behavior (under substitution) of first and second level indirect contexts. The third challenge addresses itself to this issue.

The third challenge recognizes that there is no hard data to directly support the libertine view and addresses itself to a related phenomenon. It goes as follows: "There are facts about language that the libertine view can explain and which other (e.g., "rigid") views cannot. An example of such a fact is this: If you are a normal native speaker of English, then:

(i) you are certain that all and only brothers are brothers.

(ii) you are certain that necessarily all and only brothers are brothers.
(iii) after you have thought about it a while, you are probably certain that all
and only brothers are male siblings.

(iv) In the light of (i)-(iii) you may still be uncertain whether necessarily all
and only brothers are male siblings."

The trouble with this third challenge is that the "facts" are so easily explain-
able in terms of a rigid theory that holds 'brother' and 'male sibling' to diverge in
customary sense. We explain (i) and (ii) by noting that most people know simple
logical truths and know that they are necessary. We explain (iii) by noting that
most people, when pressed, find it difficult (perhaps impossible) to imagine a
brother who is not a male sibling, or vice versa. And we explain (iv) by noting that
most people are (properly) reluctant to jump to conclusions regarding what is
necessary, once they leave the safe confines of simple logical truth.

The fourth challenge is also theoretical. It goes like this: "The rejection of the
libertine view is only plausible if 'brother' and 'male sibling' have different custom-
ary senses, and if the same is true of many similar examples. But 'brother' is de-
 fined as 'male sibling', and so they must have the same sense."

Again, this challenge is relevant only if it yields a difference in the treatment
of customary and indirect senses, so the phrase 'same sense' in the last sentence
must be read as 'same customary sense'. But the idea behind the challenge seems to
be that the meaning of a defined term is the same as the meaning of the defining
phrase in the definition, and once this is given up for part of its meaning (its ind-
direct sense) the remaining view seems much less compelling. Still, it is compelling
to some, so it deserves a more detailed discussion. An adequate treatment of the
topic of definition would take at least a book; I will limit myself here to a sketch,
addressing myself to three sorts of definition:

(i) Ordinary dictionary definitions: Comment: these almost never yield
phrases which we are even tempted to think of as having the same customary senses
as the defined terms. (Besides, in most dictionaries 'brother' is not defined as "male
sibling," nor can their equivalence be deduced by the definitions given alone.)

(ii) Philosophical definitions of ordinary-language words: These have been
thought to preserve sense because it has been taken as a criterion of adequacy that
they do so. Comments: First, there are hardly any such definitions that anyone
thinks of as successful. Second, generally the criterion that they preserve sense has
not been that they preserve sense in the relevant sense of 'sense'. For example, it is
usually thought that a philosophical definition of 'A' as "B" will be such that most
people do not already believe that all and only A's are B's, and that it will take a
philosophical argument to convince them of this (certainly in practice this is what
happens). But if 'Most people do not already believe that all and only A's are B's'
is true, we have the makings of a first-level indirect context in which 'A' and 'B'
are not interchangeable salva veritate and thus a context that shows 'A' and 'B' to
diverge in customary sense. (A continuation of this discussion would return us to
the earlier discussion of the paradox of analysis.)

(iii) Stipulative definitions: These are common in some philosophical discus-
sions and in advanced logic and mathematics courses. These definitions are held to preserve sense because the defined term is held to be a mere abbreviation of the defining phrase, or because the meaning of the defined term is totally determined by the defining phrase. Comments: First, by two weeks into the semester anyone who truly understands the defined terms by "decoding" them into what they abbreviate will certainly be unable to comprehend the lectures. Second, if it were true that stipulative definitions produced abbreviations, or produced terms with exactly the same meaning as the defining phrase, then the defined terms would be eliminable within indirect contexts as well as within extensional ones, and they would not then yield examples of pairs of phrases that have the same customary senses but divergent indirect senses.

9. COMPLICATIONS AND CONCLUSIONS

In the last section I argued that 'brother' and 'male sibling' are not interchangeable salva veritate in all first-level indirect contexts; in the context of Frege's theory this suffices to show that these phrases differ in customary sense. One might reasonably worry that similar considerations would show that any two phrases differ in customary sense. This would be in opposition to Frege's observation that:

different expressions quite often have something in common, which I call the sense. . . . It is possible for one sentence to give no more and no less information than another; and for all the multiplicity of languages, mankind has a common stock of thoughts. If all transformation of the expression were forbidden on the plea that this would alter the content as well, logic would simply be crippled, for the task of logic can hardly be performed without trying to recognize the thought in its manifold guises (Frege [C&O], p. 46).

These remarks indicate that it would be undesirable if it turned out that different phrases rarely have the same sense. But they do not show that the theory in question lacks this consequence. Perhaps, contrary to Frege's own intentions, his theory of indirect sense and reference forces this conclusion.

Does it? That depends on how we correct a certain oversimplification in the application of the theory. In the discussion so far I have been ignoring the fact that a given word does not necessarily express the same customary sense in all direct contexts. For example, words that are lexically ambiguous express different customary senses in different direct contexts; the word 'bank' sometimes means a certain kind of financial institution, and sometimes the ground alongside a moving body of water. So on different occasions of use 'bank' expresses different customary senses. But lexical ambiguity is not the important phenomenon here; this is: ignoring lexical ambiguity altogether, practically all words change customary sense from (direct) context to (direct) context. Take the word 'loves' for example. When asked whether Mary loves Bill, the response may be "Well, she does and she doesn't; it depends on what you mean by 'love'." More pertinently, on one occasion the sentence 'Mary loves Bill' may be used to say something true, and on another oc-

occasion to say something false, even though the same people are being discussed (and the same time is at issue). According to Frege's theory, this can happen only if the reference of 'loves' is different in the two utterances, and, since sense determines reference, the customary sense of 'loves' must also be different in the two utterances.

So, the observation in the last section that 'brother' and 'male sibling' have different customary senses is only obvious if it is weakened to "On some occasions, 'brother' and 'male sibling' differ in customary sense." This leaves it open to maintain that on other occasions of utterance they might agree in customary sense, and this sort of speculation would allow us to preserve Frege's view about the same common stock of thoughts being variously expressed on different occasions, while still utilizing his theory of indirect sense and reference which forces "synonyms" to sometimes diverge in (customary) sense. I suspect that this sort of application is most faithful to the ordinary language of which the theory is supposed to be a theory.

Such a refinement in the application of the theory complicates the issue discussed in the last section. On the one hand, if words change sense from occasion to occasion, then the observation that 'brother' and 'male sibling' sometimes diverge in customary sense does not show that there is no version of the paradox of analysis in which sentence (17) is true. On the other hand, the evidence in favor of a libertine version of the theory is similarly weakened; there are now more ways than ever to explain the supposedly paradoxical phenomena, and the complications of the libertine view are less plausibly needed. Conceivably, the question of which sort of theory is better as a theory of natural language reduces to the question of which is simpler; if so, I doubt that the libertine view has much to recommend it. As for the paradox of analysis, there still appears to be little virtue in the libertine solution, but this could not be made conclusive without being clearer about the goals of traditional philosophical analysis. Since there has never been a clearly successful analysis of a philosophically important word or phrase that we could use as a paradigm and since there is less and less interest in the attempt to produce one, the issue may never be finally resolved.

Notes

1. Throughout the text abbreviations in square brackets will be used to refer to items in the bibliography. In discussing Frege's theory I will ignore denotationless names and senses that fail to "present" anything; this is for simplicity only; I think that none of my arguments depend on this idealization.

2. Some passages in which this view is articulated by Frege are [BLA], Section 11; [FD], Sections 58, 60; and especially [F-R], p. 236.

3. See Stegmuller [WIS], p. 149 and Dummett [F], p. 267; Linsky [R] considers both sides of the question.

4. For example, suppose \( s_1[A] = s_1[B] \), and suppose that \( F \) and \( G \) are indirect-context creating operators. We can show, e.g., that \( s_1[F(G(A))] = s_1[F(G(B))] \) as follows:

\[
\begin{align*}
  s_1[A] &= s_1[B] & \text{Given} \\
  s_2[A] &= s_2[B] & \text{Rigidity}
\end{align*}
\]
\[
s_i [G (s_i [A])] = s_i [G (s_i [B])]
\]
Subst. of Iden.

\[
s_i [G(A)] = s_i [G(B)]
\]
Fregian Analysis

\[
s_i [G(A)] = s_i [G(B)]
\]
Rigidity

\[
s_i [F (s_i [G(A)])] = s_i [F (s_i [G(B)])]
\]
Subst. of Iden.

\[
s_i [F (G(A))] = s_i [F (G(B))]
\]
Fregian Analysis

5. For a more sophisticated language including quantifiers we may want to quantify over senses that are not expressed by words of the language at hand. Probably we would then have to assume that every sense is the nth-level sense of some possible word, for some n, and then 'B' in clause (ii) should range over possible words. The principle of rigidity would then be extended to possible words. Notice that this does not commit us to the view that if a sense is the nth level sense of a word, \(W_n\), it is not also the mth level sense of some other word, \(W_m\), where \(n \neq m\).

6. See Davison [TMLL], [OST]. Davidson actually imposes conditions on learnability (= the formulability of a Tarski-like truth definition) which may reject even theories with only one level of sense, so in his view libertine theories may be only superficially worse off than rigid theories.

7. Church [Rev] proposes a Fregian solution to the paradox of analysis, though not clearly one that utilizes doubly indirect senses; Davidson [MEI] makes a similar suggestion about a Carnapian theory in which L-equivalence has been selected as the identity condition for customary senses. White [CFB] gives a clear statement of a solution to the paradox of analysis using doubly indirect senses. I am indebted throughout this section to conversation with and unpublished work of Herbert Heidelberger, who, however, does not necessarily endorse my conclusions.

8. This requires qualification regarding words that are lexically ambiguous, I will not attempt to spell out the qualification, instead I will just try to be careful not to beg any questions here.

9. Sellars seems to suggest something like this in [NFA].

References


[Rev] Church, A., review of four articles on the paradox of analysis, Journal of Symbolic Logic


[FD] ________, "Frege on Definitions," in Geach and Black, Translations.

[S&R] ________, "On Sense and Reference," in Geach and Black, Translations.

APPENDIX

In this appendix I will use the syntax of the language L from Section 6, namely:

- Names: $a_1, a_2, a_3, \ldots$
- One-place Predicates: $P_1, P_2, P_3, \ldots$
- Extensional Sentence Operators: $E_1, E_2, E_3, \ldots$
- Intensional Sentence Operators: $O_1, O_2, O_3, \ldots$

Any expression of the form $P_n a_m$ is a sentence, and if $S$ is a sentence so are $E_n(S)$ and $O_n(S)$.

A model is a quadruple $(w_0, W, I, s_1)$, where $w_0 \in W$, $I$ is a nonempty set, $1 \neq \{0,1\}$, $I \neq \{0,1\}^W$, and $W$ does not contain 0 or 1 or 2. In the intended application $w_0$ represents the "actual" situation, $W$ represents the set of all situations, and 0 and 1 are the truth-values False and True. ("Situations" are analogous to possible worlds, except that there is no commitment to their "possibility"; for example if we wish to mimic the behavior of words like "believes", then we may need to utilize a "situation" in which Fermat's last theorem is true and one in which it is false.) The number 2 will be used solely as a technical device to distinguish first- and second-level senses. $I$ is the set of "individuals"; it may contain anything at all—e.g., tables, chairs, truth-values, situations, senses, etc.—but, for purely technical reasons, it may not consist exactly of the two truth-values or of the functions from situations to truth-values.

It will be handy to have a special symbol $P$ for $\{0,1\}^W$, i.e., for the set of functions from members of $W$ to truth-values; I will refer to $P$ as the set of propositions. $s_1$ is going to be a function that assigns first-level senses to expressions of $L$; these first-level senses fall into four domains, which are defined below (in what follows I use the notation "f:A → B" as short for "f is a function defined on set A (and nowhere else) and yielding values in set B").

**Domains of First-Level Senses**

- $D_1[\text{Names}] = \text{df } \{ f | f:W \rightarrow I \}$
- $D_1[\text{Preds}] = \text{df } \{ f | f:D_1[\text{Names}] \rightarrow P, \text{ and for any } x,y \in D_1[\text{Names}] \text{ and } w \in W, \text{ if } x(w) = y(w) \text{ then } f(x)(w) = f(y)(w) \}$

(The second clause here ensures that predication is extensional; similarly for extensional operators below.)

- $D_1[\text{ExtOps}] = \text{df } \{ f | f:P \rightarrow P, \text{ and for every } x,y \in P \text{ and } w \in W, \text{ if } x(w) = y(w) \text{ then } f(x)(w) = f(y)(w) \}$
Before defining \( D_1 [\text{IntOps}] \), it is helpful to introduce some notation for the set of “constant individual concepts of propositions”:

\[
P^* = \text{df} \{ f : f : W \rightarrow P, \text{and for some } x \in P, f(w) = x \text{ for every } w \in W \}
\]

\[
D_1 [\text{IntOps}] = \text{df} \{ f : f : P \cup P^* \rightarrow P, \text{ and if } x \in P \text{ and } y \in P^* \text{ and if } y(w) = x \text{ for some } w \in W \text{ then } f(x) = f(y) \}
\]

Members of \( D_1 [\text{IntOps}] \) are defined on what will be the first-level senses of sentences (namely, \( P \)) and on what will be the second-level senses of sentences (namely, \( P^* \)); the complicated clause above ensures that in certain cases such functions cannot tell the first- and second-level senses apart.

**Domains of Second-Level Senses**

\[
D_2 [\text{Names}] = \text{df} \{ f : f : W \cup \{ 2 \} \rightarrow D_1 [\text{Names}] \}
\]

(The use of 2 here is merely to ensure that a second-level sense appropriate to a name is identifiable as such.)

\[
D_2 [\text{Preds}] = \text{df} \{ f : f : D_2 [\text{Names}] \rightarrow P^*, \text{ and there is some } x \in D_1 [\text{Preds}] \text{ such that for every } y \in D_2 [\text{Names}] \text{ and } w \in W, f(y)(w) = x(y(w)) \}
\]

\[
D_2 [\text{ExtOps}] = \text{df} \{ f : f : P^* \rightarrow P^*, \text{ and there is some } x \in D_1 [\text{ExtOps}] \text{ such that for every } y \in P^* \text{ and } w \in W, f(y)(w) = x(y(w)) \}
\]

\[
D_2 [\text{IntOps}] = \text{df} \{ f : f : P^* \cup \{ 2 \} \rightarrow P^*, \text{ and there is some } x \in D_1 [\text{IntOps}] \text{ such that for every } y \in P^*, f(y) = x(y) \}
\]

Having specified the domains, we now specify the requirement hinted at above: \( s_1 \) must map each name to a member of \( D_1 [\text{Names}] \), each predicate to a member of \( D_1 [\text{Preds}] \), each extensional operator to a member of \( D_1 [\text{ExtOps}] \), and each intensional operator to a member of \( D_1 [\text{IntOps}] \). Now suppose that we assume that both \( D_1 [\text{Names}] \) and \( P^* \) contain some arbitrarily selected “designated” member. Then we can define the function \( s_2 \) which assigns second-level senses to the vocabulary of \( L \) as follows:

\[
s_2 [a_n] = \text{df} \text{ the unique member of } D_2 [\text{Names}] \text{ which maps each } w \in W \text{ to } s_1 [a_n] \text{ and maps 2 to the designated member of } D_1 [\text{Names}] .
\]

\[
s_2 [P_n] = \text{df} \text{ the unique member } f \text{ of } D_2 [\text{Preds}] \text{ such that for every } y \in D_2 [\text{Names}] \text{ and } w \in W, f(y)(w) = s_1 [P_n] (y(w)).
\]

\[
s_2 [E_n] = \text{df} \text{ the unique member } f \text{ of } D_2 [\text{ExtOps}] \text{ such that for every } y \in P^* \text{ and } w \in W, f(y)(w) = s_1 [E_n] (y(w)).
\]

\[
s_2 [O_n] = \text{df} \text{ the unique member } f \text{ of } D_2 [\text{IntOps}] \text{ such that } f \text{ maps 2 to the designated member of } P^*, \text{ and for every } y \in P^*, f(y) = s_1 [O_n] (y).
\]

We can now extend \( s_1 \) and \( s_2 \) to the sentences of \( L \) by means of the stipulations:

\[
s_1 [A(B)] = \text{df} s_1 [A] (s_1 [B]), \text{ and }
\]
where $A$ is a predicate and $B$ a name, or $A$ is an operator and $B$ a sentence.

Next we will define the "presentation" relation. Notice that the set $D_1 [\text{Name}] \cup D_2 [\text{Name}] \cup P \cup P^*$ is disjoint from the other domains, and they are all disjoint from each other. We call any member of any of the domains, or of $P$ or $P^*$, a sense, and we specify, case by case, what each sense presents.

If $x \in D_1 [\text{Name}] \cup D_2 [\text{Name}] \cup P \cup P^*$ then $x$ presents $y$ iff $x(w_0) = y$.

If $x \in D_1 [\text{Preds}]$ then $x$ presents $y$ iff $y \in \{0,1\}^I$ and for every $z \in D_1 [\text{Name}]$, $x(z)(w_0) = y(z(w_0))$.

If $x \in D_2 [\text{Preds}]$ then $x$ presents $y$ iff $y \in D_1 [\text{Preds}]$ and for every $z \in D_2 [\text{Name}]$, $x(z)(w_0) = y(z(w_0))$.

(Similar clauses for $x \in D_1 [\text{ExtOps}]$ and $x \in D_2 [\text{ExtOps}]$.)

If $x \in D_1 [\text{IntOps}]$ then $x$ presents $y$ iff $y \in \{0,1\}^P$ and for every $z \in P^*$, $x(z)(w_0) = y(z(w_0))$.

If $x \in D_2 [\text{IntOps}]$ then $x$ presents $y$ iff $y \in D_1 [\text{IntOps}]$ and for every $z \in P^*$, $x(z) = y(z)$.

The systems $F_1, F_2$, and $C_1$ result from the above "ontology" by means of appropriate definitions of the semantical relations expressing and referring.

**System $F_2$ (= What Frege Said, Though Probably Not What He Meant)**

Say that an occurrence of $A$ in an extensional context in a sentence of $L$ refers to whatever $s_1 [A]$ presents, and expresses $s_1 [A]$, and an occurrence of $A$ that is in an indirect context in a sentence of $L$ refers to $s_1 [A]$ and expresses $s_2 [A]$. Then the following Fregean principles all hold:

1. If $A$ is complex, what it refers to in a given context is a function of what its parts refer to in that context.
2. If $A$ is complex, what it expresses in a given context is a function of what its parts express in that context.
3. In any context, $A$ always refers to the unique thing that is presented by what it expresses in that context.

**System $F_1$ (= What Dummett Says Frege Should Have Said)**

Say that an occurrence of $A$ in an extensional context refers to what $s_1 [A]$ presents and expresses $s_1 [A]$, and an occurrence of $A$ in an indirect context refers to and expresses $s_1 [A]$. Then:

1. If $A$ is complex, what it refers to in any given context is a function of what its parts refer to in that context.
2. If $A$ is complex, what it expresses in any given context is a function of what its parts express in that context.
(3') If an occurrence of $A$ is not in an indirect context, it refers$_D$ to whatever is presented by what it expresses$_D$ there.

Further, we have the following comparison of $F_1$ with $F_2$:

(4') Any isolated occurrence of a sentence refers to what it refers$_D$ to, and expresses what it expresses$_D$.

System $C1$ (= Carnap's System, Sort of)

Say that any occurrence of $A$ always refers$_C$ to whatever is presented by $s_1[A]$, and always expresses$_C s_1[A]$. Then:

(1'') If $A$ is complex and contains no intensional operators then what it refers$_C$ to in a given context is a function of what its parts refer$_C$ to in that context.

(2'') If $A$ is complex then in any given context what it expresses$_C$ is a function of what its parts express$_C$ in that context.

(3'') In any context, $A$ always refers$_C$ to whatever is presented by what it expresses$_C$ in that context.

Again we have a comparison with $F_2$:

(4'') Any isolated occurrence of a sentence refers to what it refers$_C$ to and expresses what it expresses$_C$. 