Extensionality and the Composition of Thoughts

Nicholas Koziolek
University of Chicago

Draft of
July 8, 2014

This is a draft. Please do not cite or circulate without the author's permission (but do feel free to ask for permission). Comments are both welcome and appreciated.

Abstract

In recent years, a consensus has emerged, among philosophers of mind and language, that there is no consistent Fregean theory of propositional attitude ascriptions which dispenses with Frege's infamous hierarchy of senses. In this paper, I show that this consensus is premature: even the most persuasive arguments for the hierarchy rest on the undefended, and questionable, assumption that any genuinely Fregean theory must be purely extensional. On the alternative theory I propose, both the sense and reference of an expression will remain the same in every linguistic context, but the truth-value of a propositional attitude ascription will nonetheless depend on the customary senses of some of its parts. The result is a consistent, plausibly Fregean, but non-extensional theory of propositional attitude ascriptions.

Introduction

For those sympathetic to Frege's general approach to issues in the philosophy of mind and language, his account of propositional attitude ascriptions was, for a long time, a source of embarrassment. Carnap, for example, initially complained that the account involved positing entities—"indirect senses"—of which Frege provided no account. To make matters worse, Davidson later argued that, even if Frege had provided an account of these entities, any language which answered to his view would have to be unlearnable anyway.[1] For a time, it seemed that the Fregean had no response. More recently, however, a number of philosophers, led by Tyler Burge, have developed a sophisticated defense of Frege's position, one which, they claim, is immune to the charges pressed against it by Carnap and Davidson.[2] Indeed, as Burge sees things, Frege's position is not only not an embarrassment, it is actually a source of great pride.[3] There has thus been something of a revival, in the contemporary literature, of Frege's account of propositional attitude ascriptions.

At the same time, Frege's theory—his theory of indirect contexts, as I will usually call it—arguably runs afoul of another famous objection of Davidson's, the objection from semantic innocence. On Frege's view, the sense and reference of an expression "shift" when the expression occurs in an indirect context. In particular, an expression in an indirect context refers to the sense it ordinarily expresses and expresses a new sense, which Frege calls its indirect sense. Davidson objects:

> If we could recover our pre-Fregean semantic innocence, I think it would seem to us plainly incredible that the words 'The earth moves', uttered after the words 'Galileo said that', mean [read: express] anything different, or refer to anything else, than is their wont when they come in other environments. (1968: 109)

The naive (or "innocent") view, Davidson thus suggests, is that neither the sense nor the reference of an expression varies from one context to another. An expression in an indirect context must therefore refer to its customary reference and express its customary sense, just as it would in any other context.

[1] See, respectively, Carnap 1947: 129 and Davidson 1965: 15.

[2] See Burge 1979 and 2004, Kripke 2008, Peacocke 2009, and Parsons 2009.

[3] Burge says: "I see Frege's account of attribution of thought as a contribution to a scientific attempt to account for thought, including actual propositional attitudes, as well as abstract thinkable *Gedanken*. I believe that at present there is no superior account of thought, or the attribution of thought" (2004: 167–8).

My aim in this paper is twofold: first, to show that neither Frege nor his contemporary defenders have provided us with any principled reason to reject the intuition which grounds Davidson's plea for semantic innocence; and, second, to show that it is possible to construct a consistent theory of indirect contexts which employs the concepts of sense and reference, respects the basic claim of Frege's own theory of indirect contexts, and yet is semantically innocent in Davidson's sense. The theory I will propose does, however, involve an important revision to the formal framework within which Frege develops his more general theory of sense and reference. In particular, my theory, unlike Frege's own, permits exceptions to what I will call the Principle of Extensionality, according to which the truth-value of a sentence remains unchanged when a part of the sentence is replaced by an expression with the same reference. I conclude that, without a principled and non-question-begging defense of this Principle, we have no reason to prefer Frege's own way of developing his theory of sense and reference over the neo-Fregean one adopted here. Moreover, whatever reason we have to prefer a theory which is semantically innocent to one which is not would seem to give us reason to prefer the neo-Fregean approach to Frege's own.

1 The objection from semantic innocence

It is tempting to respond to Davidson's objection from semantic innocence by pointing out that it completely ignores Frege's reasons for rejecting the naive view, that is, the view on which both the sense and the reference of an expression remain everywhere the same. After all, if there are sound arguments in favor of Frege's own theory—as he and his contemporary defenders claim there are—then it is just irrelevant that someone innocent of his influence should think the theory "plainly incredible." Davidson's rhetoric aside, however, there is a real question here: how convincing *are* the reasons Frege gives for rejecting the naive view?

On Frege's own presentation, the case against the naive view is apparently quite direct. He says:

> If words are used in the ordinary way, what one intends to speak of is their reference. It can also happen, however, that one wishes to talk about the words themselves or their sense. [...] 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. (1892: 58–59)

The general thought in the background here is that the reference of an expression (token) is what we mean to talk about in using that expression (token). Frege adds that, when we report the speech (or the propositional attitude) of another person, we mean to refer to the customary senses of the very words we use in doing so. It follows that an expression in an indirect context refers to its customary sense, and thus that the naive view is false.

But notice that what seems "quite clear" to Frege here is the very thing that strikes Davidson as "plainly incredible." If there is nothing more to say about the matter, we will, I think, be forced to conclude that each philosopher is simply begging the question against the other. Indeed, we ourselves may feel pulled in both directions on this issue. We might want to say: in one sense, words in indirect speech refer to the same things they do in ordinary speech; but, in another sense, they refer to something else, to what is meant or expressed or said by them in ordinary speech—i.e., their customary senses. In any case, without some account either of the conflict between Davidson's remarks and Frege's, or of the ambiguity that gives the appearance of conflict, we do not understand what is at stake between these two philosophers. Or, at best, we are left to represent the conflict as a bare conflict of intuitions: Davidson thinks that an expression in an indirect context must "stand for" or "be about" its customary reference, while Frege thinks it must "stand for" or "be about" its customary sense. The conflict turns out to be nothing more than idle wrangling over some purportedly pre-theoretical concept of aboutness.

There is, however, another way of seeing the disagreement, one which makes it a substantive one, and which, at the same time, makes it seem more likely to admit of a resolution. On this way of seeing things, Frege's case against the naive view begins with a simple observation: ordinarily, the truth-value of a sentence depends on the references of its parts. The truth-value of the sentence 'Hesperus is a planet', for example, depends on the reference of the name 'Hesperus' (that is, on the object it, in this case uncontroversially, "stands for" or "is about"): it is true just in case the reference of 'Hesperus'—i.e., Hesperus/Phosphorus/Venus—is a planet.

So far, Davidson need have no quarrel. What rules out the naive view is a subsequent generalization of this initial observation; that is, the stronger claim that the truth-value of a sentence *always* depends on the references of its parts (and not on any other of their semantic features). For example, given the generalization, it follows that the truth-value of the sentence 'John believes that Hesperus is a planet' depends on the reference of the name 'Hesperus' and that the truth-value of the sentence 'John believes that Phosphorus is a planet' depends on the

reference of the name 'Phosphorus'. But these two sentences may *differ* in truth-value (if John does not know or believe that Hesperus is Phosphorus). So we are forced to conclude that, in these more complex sentences, 'Hesperus' and 'Phosphorus' do not refer to Hesperus/Phosphorus/Venus.[4] This conclusion, of course, is one that Davidson thinks is "plainly incredible." Whether or not we feel similarly, we might ask, in a spirit of cautious sympathy: why think that Frege's initial observation generalizes? This question, I suggest, is what is at the heart of the disagreement between Frege and Davidson. An adequate response to the objection from semantic innocence must therefore provide a justification for Frege's generalization.

A typical response to the question is to point out that, for Frege, the generalization itself is an analytic truth. Thus Kripke, for example, says:

> Why shouldn't the reference of the whole [sentence] be allowed to depend on other features of the parts [than their references]? Indeed, why not just say that exceptions to this principle [i.e., the principle that the reference of a sentence is a function of the references of its parts] obviously exist, e.g., in indirect contexts and possibly in quotation? [...] [I]t could be argued that for Frege the principle is analytic. (2008: 270, note 46)

Now, if our aim is only to understand how Frege came to hold his view of indirect contexts (as, I should say, I think Kripke's aim is), this response will perhaps suffice. But if our concern is to determine whether Frege's theory is true, the relevant question is whether the generalization really is analytic (or, more simply, true), not merely whether Frege thought it was.

A central thesis of the present essay is that the generalization is *not* analytic, at least not in a sense of 'analytic' strong enough to ground the rejection of a semantically innocent theory of indirect contexts. That is, while one might hold the generalization is simply *stipulated*, by Frege, to be true of his concept of reference, a satisfactory defense of Frege's view against Davidson's objection from semantic innocence cannot rest on such a stipulation. For, as we will see, in the context of Frege's other views, the generalization constitutes a substantive philosophical thesis, and therefore stands in need of a substantive defense.

My strategy in the remainder of the paper is as follows. First, I will present a "neo-Fregean" theory of indirect contexts—a theory which makes central use of

[4]Strictly, all that follows is that 'Hesperus' and 'Phosphorus' do not *both* refer to Hesperus/Phosphorus/Venus. But since there is no obvious principled way of deciding which of the two should keep its ordinary reference, it seems reasonable to conclude that *neither* does.

Frege's concepts of sense and reference, but on which Frege's generalization about the determination of truth-value is false, i.e., admits of exceptions—and then to ask where, exactly, that theory goes wrong. With the neo-Fregean theory laid out clearly before us (in §2), we will be able to provide a precise statement of two different objections which have been offered as decisive against any semantically innocent, broadly Fregean view, and thus as at least indirect evidence for the truth (though not, I must stress, the analyticity) of Frege's generalization. The first, it turns out, is an objection to the neo-Fregean theory's principle for determining the *references* of sentences containing indirect contexts, while the second is an objection to the theory's principle for determining the *senses* of those sentences. I treat the first objection in §3, and the second in §4. As I have already said, I find that neither objection is decisive. I conclude that the neo-Fregean theory is consistent, and so represents a legitimate way of applying Frege's general philosophical framework to sentences of indirect discourse and ascriptions of propositional attitude. While further work is required in order to decide whether it is preferable to Frege's own theory, it at least deserves to be taken seriously.[5]

2 A semantically innocent (neo-)Fregean theory of indirect contexts

My aim in this section is to introduce and explain the fundamental tenets of a semantically innocent (neo-)Fregean theory of indirect contexts, and to do so in a way that will make the view seem at least initially plausible. My presentation of the view will thus be both brief and relatively informal. I will try to provide enough detail to make the shape of the view clear, and to allow me to state and respond to the two central objections to it; but I will try to avoid providing so much detail that the discussion is bogged down by distracting technicalities. For those who are interested, however, a more complete formal statement of the view is provided in an Appendix.

To begin, I take it that the basic claim of Frege's theory of indirect contexts is

[5]I should mention that a theory closely related to the one introduced in §2 was once recommended by Michael Dummett (1973: 267–268), though he never developed the theory in any detail. There are, however, important differences between my view and the one Dummett suggests. In particular, I claim that neither the sense nor the reference of an expression shifts when it occurs in an indirect context, while he claims that the reference does shift, though the sense does not. Certain objections which have been made to Dummett's view thus do not apply to the view proposed below.

just this:

> **Frege's Basic Claim:** the truth-value of a sentence of the form '*S* ϕs that *p*' depends on the customary sense of the embedded sentence '*p*'.

For example, the truth-value of the sentence 'Plato believes that Socrates is wise' depends on the customary sense of the embedded sentence 'Socrates is wise', i.e., on the thought that Socrates is wise. Since the sense of a complex expression is composed of the senses of its parts, Frege's Basic Claim automatically generalizes: the truth-value of a sentence containing an indirect context depends on the customary senses of the expressions embedded in that context. Thus, the truth-value of the sentence 'Plato believes that Socrates is wise' depends, too, on the customary senses of the name 'Socrates' and the predicate 'is wise'. For these are simply parts of the thought expressed by the whole embedded sentence 'Socrates is wise', on which the truth-value of the containing sentence has explicitly been said to depend.

A Fregean theory which is semantically innocent—if such a thing is indeed possible—will need to respect not only Frege's Basic Claim, but also the following condition:

> **Semantic Innocence:** both the sense and the reference of an expression remain everywhere the same.

Thus, the name 'Socrates' refers to Socrates and expresses the relevant way of thinking of Socrates (the customary sense of the name 'Socrates') not only in the simple sentence 'Socrates is wise' but also in the more complex sentence 'Plato believes that Socrates is wise'—even though, in the latter, it occurs in an indirect context, while, in the former, it occurs in an ordinary or (as I will say) transparent context.

How, then, would a semantically innocent Fregean theory work? Well, like any semantic theory, it will need to tell us how the semantic features of sentences are determined by (i) their syntactic structures and (ii) the semantic features of their parts. Since the theory is Fregean, the semantic features of both sentences and their parts are their senses and references. And since the theory is semantically innocent, the principles it employs in executing this task will need to make appeal only to (what Frege called) the *customary* senses and references of the parts. (Since, on a semantically innocent theory, both the sense and the reference of an expression remain everywhere the same, we can ultimately dispense with Frege's label. For the sake of clarity and explicitness, however, I will often continue to employ it.)

We thus need two rules, one which tells us how the *senses* of sentences are determined by their syntactic structures and the senses and references of their parts; and one which tells us how the *references* of sentences are so determined. The first of these rules can be stated, roughly, in the following way:

> **SENSE-COMP** the sense of a sentence is composed of the senses of its parts.

What is rough about this statement of the rule for the composition of sense is that the notion of composition it employs is, so far, left undetermined. There are various things it could mean to say that one way of thinking is composed of others, and we have not yet chosen any particular one of these. I will postpone detailed discussion of SENSE-COMP until §4, for reasons I will explain there. I want to focus first on the second rule, the rule for the composition of reference, which can be stated as follows:

> **REF-COMP** The reference of a sentence is a function of the *references* of those of its parts that occur in *transparent* contexts and of the *senses* of those of its parts that occur in *indirect* contexts.

For present purposes, an indirect context can simply be defined as a context created by the complementizer 'that', so that an expression occurs in an indirect context just in case it occurs within the scope of the complementizer. Thus, in the sentence 'Plato believes that Socrates is wise', the name 'Socrates' and the predicate 'is wise' both occur in an indirect context, while the name 'Plato' and the predicate 'believes', as well as the complementizer 'that' itself, occur in a transparent context.

I will assume, with Frege, that the sense of a sentence is the thought it expresses, and its reference is its truth-value; that the reference of a name is an object, and its sense is a way of thinking of that object; and that the reference of a predicate is a function from objects to truth-values, and its sense is a way of thinking of that function. The result is that the reference of the sentence 'Plato believes that Socrates is wise' will be the value of the function *Believes(x,y)* for *Plato* and *the thought that Socrates is wise* as arguments. For the function Believes(x,y) is the reference of the predicate 'believes', the man Plato is the reference of the name 'Plato', and the thought that Socrates is wise is the sense of the sentence 'Socrates is wise'.

One feature of this neo-Fregean theory which will become important later is that the complementizer 'that' itself is assigned neither a sense nor a reference. Its job is solely to mark the indirect context it creates, and so to ensure, in the first

place, that the reference of any sentence in which it occurs depends on the sense, rather than the reference, of the sentence which occurs within its scope. (The effect of 'that' on the *sense* of the sentences in which it occurs is something we will consider in detail in §4.) One consequence of this fact about the complementizer is that the compositional machinery of the theory will assign neither a sense nor a reference to whole that-clauses, like 'that Socrates is wise'. The theory is nonetheless fully compositional, in the sense that both the sense and the reference of any *sentence* is determined by its syntactic structure and the senses and references of its parts (i.e., of those of its parts which have senses and references). And it is not that 'that' plays no semantic role in the theory. It is just that its semantic role is completely described by the compositional rules for reference and for sense. In particular, with respect to REF-COMP, its job is to mark contexts in which what an expression contributes to the truth-conditions of sentences in which it occurs is its sense rather than its reference.

Such is the general shape a semantically innocent Fregean theory of indirect contexts might take. The theory is semantically innocent because both the sense and the reference of an expression remain everywhere the same. That is, every expression has the same sense and reference in an indirect context that it has in an ordinary, transparent context. The theory is Fregean because it respects Frege's Basic Claim: the reference—i.e., the truth-value—of a sentence containing an indirect context depends on the senses of the expressions which occur in that context. But where Frege's own theory respects the Basic Claim by positing shifts in the sense and reference of expressions occurring in indirect contexts, the theory proposed here respects the same Claim instead by providing compositional rule for reference which is sensitive to the difference between indirect and transparent contexts.

3 First objection: the principle of extensionality

So what is the answer to our question? Why *shouldn't* the reference, the truth-value, of a whole propositional attitude ascription be allowed to depend on the senses of the expressions that occur in its that-clause? In other words, what is wrong with the neo-Fregean theory just presented?

A first objection to the neo-Fregean theory of §2 rests on an allegedly fundamental Fregean principle:

> **the Principle of Extensionality** The truth-value of a sentence remains unchanged when a part of the sentence is replaced by an expression with

the same reference.[6]

This principle can be shown to be incompatible with the neo-Fregean theory presented in the preceding section.

The argument runs as follows. On the neo-Fregean theory, as on the classical one, the sentences

(1) John believes that Hesperus is a planet

and

(2) John believes that Phosphorus is a planet

may differ in truth-value (if John does not believe that Hesperus is Phosphorus). On the neo-Fregean theory, though, 'Hesperus' and 'Phosphorus' will have the same reference, even in these indirect contexts. For, on that theory, remember, both the sense and the reference of an expression remain unchanged, no matter where (in what kind of linguist context) the expression appears. The problem is that the Principle of Extensionality tells us that, if two expressions have the same reference, then the substitution of one for another in a given sentence cannot affect the truth-value of the whole sentence. Something must give. If the neo-Fregean view is to respect Frege's Basic Claim, it cannot take the neo-Russellian route and conclude that sentences (1) and (2) do not differ in truth-value after all. If it is to remain semantically innocent, and not collapse back into the classical Fregean view, it cannot conclude that the occurrences of 'Hesperus' and 'Phosphorus' in (1) and (2) differ in reference. Apparently, the only remaining conclusion is that indirect contexts constitute an exception to the Principle of Extensionality.

[6]The work being done here by the Principle of Extensionality could also be done, more directly, by the sorts of remarks Frege makes in the passage from "On Sense and Reference" quoted in the Introduction. I focus on the Principle of Extensionality because I think it, and not Frege's intuition about aboutness, is what sustains the impression, amongst contemporary readers of Frege, that the neo-Fregean theory is untenable. It is worth stressing, again, that, while accepting Frege's intuition would allow us to sidestep our question—since the intuition tells us that the reference of an expression in an indirect context just isn't what the neo-Fregean theory says it is—it is not clear why we should accept this intuition rather than Davidson's. Appeal to the Principle of Extensionality is tempting, in this context, in part because it at least seems to provide a principled reason for favoring Frege's intuition. My point in the present section is that this appearance is reality only if we have independent reason to accept the Principle of Extensionality itself, and that, at least at present, no such reason has been produced.

So why not just reject the Principle of Extensionality? What reason do we have to insist that this principle must be exceptionless? An objection to the neo-Fregean view which rests on a bare appeal to this principle can only beg the question against that view. It is, of course, true that conjoining this principle with the neo-Fregean view, as presented here, results in contradiction. But the above presentation of the view (in §2) did not involve any mention of the Principle of Extensionality. So the contradiction is not internal to the theory. If it were, the theory would be unacceptable. Since it is not, some defense of the principle itself is in order, if the rejection of the neo-Fregean theory is to be justified.

There is, however, a tendency—one which is pervasive in contemporary discussions of Frege's theory of indirect contexts—to assume that the Principle of Extensionality *is*, that it *must* be, internal to the neo-Fregean view, at least if the view is to have any claim to be a Fregean one—even a "neo"-Fregean one. Let me put it this way: it is tempting to assume that the principle does no more than characterize the concept of reference; that is, that it does no more than give expression to an essential feature of that concept. If that is what it does, then it would seem that to reject it is already to abandon the theoretical framework of sense and reference, and so to abandon Frege's whole approach to the study of language and thought.

What makes these assumptions so tempting is that, in Frege's own writings, the very framework of sense and reference *is* in fact shaped by his acceptance of the Principle of Extensionality, and by some closely related principles. In "On Sense and Reference," for example, he cites the Principle of Extensionality explicitly: "If we now replace one word of the sentence by another having the same reference, but a different sense, this can have no bearing upon the reference of the sentence" (1892: 62).[7] Since he provides no justification, it is natural to take the principle

[7]It is perhaps worth noting that Frege uses the Principle of Extensionality, in the context from which I have taken it, to justify his claim that the reference of a sentence cannot be the thought it expresses; the reason being, of course, that, to use his example, the two sentences 'The morning star is a body illuminated by the sun' and 'The evening star is a body illuminated by the sun' express different thoughts, despite the fact that 'the morning star' and 'the evening star' have the same reference. He does not here connect the principle explicitly to the issues raised by indirect contexts. He does do so in a 1902 letter to Russell (see Frege 1980: 149); but, even there, it does not seem that he cites the claim as a *justification* for his theory of indirect contexts—in fact, the order of explanation seems to be the reverse. Of course, the obvious alternative theories would conflict with the Principle of Extensionality. But, if Frege's reason for taking his view is given by his remarks in the passage quoted in the Introduction, then it may be a real question how he would have responded if the view from the earlier passage were challenged. That is, Frege might well think that, if we share Davidson's intuition that expressions in indirect contexts

to be true as a matter of definition. Importantly, however, he also says, a couple pages earlier, that "[t]he reference of a proper name is the object itself which we designate by its means" (1892: 60); and this too seems to be definitional. Since Frege himself thinks that "the object which we designate by means of" an expression in an indirect context is its customary sense, these definitions seem to him to be consistent.

Davidson, on the other hand, is convinced that "the object which we designate by means of" an expression in an indirect context is what Frege would call its customary *reference*. So, on his view, the two definitions come into conflict: the second (the one which mentions "designation") implies that an expression in an indirect context refers to its customary reference, while the first (the one which deals with substitution) implies that it refers to its customary sense. A Fregean who is sympathetic to Davidson should thus be led to wonder why we should accept both definitions. The neo-Fregean theory is one on which we accept the second definition and reject the first. Or rather, more precisely, it is a view on which we *revise* the first definition, so that it allows for exceptions in special cases. For, even on the neo-Fregean view, there is truth in the Principle of Extensionality: though it admits of exceptions in special cases, it does apply to atomic sentences, and serves thereby to get our semantic theory off the ground.

The neo-Fregean theory thus serves as an ideal test case for deciding the issue at stake between Frege and Davidson. For both of these philosophers take seriously the idea that the reference of an expression is the object which we designate by its means. And they agree about the result of applying this idea in ordinary cases, that is, to expressions in transparent contexts (and so, crucially, to atomic sentences). They just disagree about the result of applying it to expressions in indirect contexts. The neo-Fregean theory provides us with something which fits Davidson's intuition on this matter, something to stand to it in the way Frege's own theory stands to *his* intuition. The question of semantic innocence thus becomes a question of the relative merits of these two theories.

What is crucial here is that it does not suffice, to decide this matter in favor of the classical Fregean view, to say that the Principle of Extensionality is definitionally true of Frege's concept of reference. For, if that is all we can say, we will be left without a substantive reason for rejecting the neo-Fregean view. Granted,

---

are "about" their customary references, then the appropriate thing for us to do is to admit that indirect contexts are exceptions to the Principle of Extensionality. At any rate, we are looking for independent justification for Frege's theory, and the present question is whether the Principle of Extensionality can supply it. For these purposes, it is significant that Frege himself seems to provide no justification for the principle.

the latter view employs a slightly different concept of reference than Frege himself does, and so there is at least something to the thought that a defense of the neo-Fregean view cannot simply rest on Frege's laurels. But the force of this last objection should not be overestimated. After all, there *is* significant overlap between the classical and the neo-Fregean concepts of reference. In the first place, on both conceptions, the reference of a name is the object which we designate by its means, and this idea gives us a place to start in constructing a semantic theory. But the overlap is even greater than that. For, even if we employ the neo-Fregean concept of reference, the Principle of Extensionality will still apply to those sentences whose treatment is fundamental for the Fregean approach: atomic sentences. The difference is only that the neo-Fregean permits exceptions to the principle in special cases. And, since these exceptions can be given a systematic account, there is no reason to think that permitting them is objectionable in principle.

4 Second objection: extensionality and the composition of thoughts

4.1 The new principle of extensionality

On the usual understanding of Fregean thoughts, the structure of a thought is mirrored, in a particular way, in the structure of a sentence that expresses it (if the sentence is a sentence of a proper *Begriffsschrift*, anyway). In particular, every logically significant expression in a sentence expresses a sense which is a component part of the thought expressed by the sentence. A part of a thought is complete if the part of the sentence which expresses it is complete, and it is incomplete if the part of the sentence which expresses it is incomplete. Just as the sentence is the result of completing its incomplete parts with its complete parts, so the thought is the result of completing *its* incomplete parts with its complete parts.

Now thoughts, like sentences, have truth-values. And the truth-value of a sentence, we have assumed, is determined by the references—and, on the neo-Fregean view, sometimes the senses—of the component parts of the sentence, and by the way in which those parts are combined (and, of course, by the way the world is). This fact is reflected in principles like Ref-Comp and Ref-Comp-T. These principles tell us, again, that the truth-value of a sentence is the value of a given function for given arguments. In the basic case, that of atomic sentences, the function is given as the reference of the sentence's predicate, and the arguments

are given as the references of its terms. The further syntactic structure of the sentence tells us which term supplies an argument to which place in the function. On the neo-Fregean theory, things sometimes work a bit differently. But REF-COMP-T makes it clear enough how we should understand these cases.

Since thoughts also have truth-values, and since the structure of a thought is reflected in the structure of a sentence that expresses it, it would seem that we should be able to tell a similar story about the way in which the truth-value of a *thought* is determined by its parts and their composition. Indeed, if, as we have assumed, the sense of an expression is a way of thinking of its reference, then, since the sense of an expression is also a part of the thought expressed by sentences in which the expression occurs, it would seem that the story we have told about the truth-values of sentences ought to be straightforwardly applicable to thoughts. The idea would be that the truth-value of a thought, just like the truth-value of the sentence which expresses it, is the value of a given function for given arguments. The function and arguments are no longer given as the references of expressions, however. Instead, the function is given by the incomplete part of the thought, the part which corresponds to the sentence's predicate: the function we want is the one of which that incomplete thought-component is a way of thinking. Similarly, the arguments are given by the complete parts of the thought, the parts which correspond to the sentence's terms: the arguments we want are those of which the complete thought-components are ways of thinking. And the further structure of the thought—which, again, simply mirrors the syntactic structure of the sentence which expresses it—tells us which complete part of the thought supplies an argument to which place in the function.

This idea can be captured more precisely by the following principle:

> **TRUTH-COMP** The truth-value of a thought $\mathrm{W}[F(a_1, \ldots, a_n)]$ is the result of applying the function presented by the incomplete way of thinking $\mathrm{W}[F(x_1, \ldots, x_n)]$ to the objects presented by the complete ways of thinking $\mathrm{W}[a_1]$, …, $\mathrm{W}[a_n]$ (in the appropriate order),

What I have called the new principle of extensionality is, in effect, the claim that TRUTH-COMP *suffices* as an account of the way in which the truth-value of a thought depends on its parts and their composition. Or, to state it in a form which parallels our statement of the (old) principle of extensionality:

> **the new principle of extensionality** The truth-value of a thought remains unchanged when a part of the thought is replaced by a part which is a way of thinking of the same object or function.

As I have already said, the neo-Fregean theory of indirect contexts is incompatible with the new principle of extensionality, just as it was incompatible with the old. And the reasons are essentially the same. What we saw in §3 was that, if the (old) principle of extensionality is true, then the truth-value of a sentence can never depend on features of its parts other than their references. Similarly, if the *new* principle of extensionality is true, then the truth-value of a *thought* can never depend on features of *its* parts other than the objects and functions of which they are ways of thinking.

As we will see, it is rather more difficult to make room for exceptions to the new principle of extensionality than it was to make room for exceptions to the old. But our earlier discussion of the latter at least suggests how we ought to proceed. Our way of making room for exceptions to the principle of extensionality in §3 was to exploit the fact that the sentences we wanted to treat as exceptional all involved an expression, the operator 'that ρ', which we could treat as the source of the anomalousness of sentences containing it. We thus introduced a principle, Ref-Comp-T, which applied only to sentences containing the operator 'that ρ', and which laid it down that the truth-value of such a sentence is to depend on the sense of the sentence which completes the operator (and so, by extension, on the senses of its parts), rather than on its reference (or, by extension, the references of its parts). What we need, then, is to find a way of applying the same strategy to thoughts themselves. That is, we need a principle which stands to Truth-Comp as Ref-Comp-T stands to Ref-Comp: a supplementary principle for the determination of truth-value which applies to all and only the thoughts we wish to treat as anomalous.

But here we run into an apparent problem. Ref-Comp-T exploits the fact that sentences like 'Plato believes that Socrates is wise' contain the operator 'that ρ'. It is the presence of this operator which, in effect, explains why this occurrence of 'Socrates is wise' is anomalous, relative to its occurrences in other contexts (for example, its occurrence in 'If Socrates is wise, then Plato is wise', in which the truth-value of the whole apparently depends only on the reference(s) of the relevant part(s)). Our construal of the thought expressed by such a sentence, however—the thought that Plato believes that Socrates is wise—involves no mention of anything corresponding to the operator 'that ρ', and so contains no mention of anything we might exploit in formulating a principle which applies to thoughts as Ref-Comp-T applies to sentences. That is, our official way of referring to the thought that Plato believes that Socrates is wise was 'W[Believes(Plato, Wise(Socrates))]'. If we apply Truth-Comp to this thought, so characterized, it tells us that its truth-value is Believes(Plato, Wise(Socrates)). And there is noth-

ing to indicate that Truth-Comp should not apply to this thought.

This is a serious problem for the neo-Fregean theory. As that theory was spelled out in §2, the sentence 'Plato believes that Socrates is wise' refers to the truth-value Believes(Plato, W[Wise(Socrates)]),[8] *not* to the truth-value Believes(Plato, Wise(Socrates)). The theory also implies that 'Plato believes that Socrates is wise' expresses the thought W[Believes(Plato, Wise(Socrates))];[9] and the truth-value of this thought, as we just saw, is Believes(Plato, Wise(Socrates)). So it turns out that the truth-value of the *sentence* 'Plato believes that Socrates is wise' is Believes(Plato, W[Wise(Socrates)]), while the truth-value of the *thought* it expresses is Believes(Plato, Wise(Socrates)). And these truth-values are not necessarily the same: they will differ, in particular, in the case in which Plato has some true belief, but does not believe specifically that Socrates is wise (assuming, that is, that Socrates is indeed wise). For the one applies the function Believes(x, y) to Plato and *the thought* that Socrates is wise, while the other applies it to Plato and *the truth-value* obtained by applying the function Wise(x) to Socrates (that is, I assume, to Plato and the True). The conclusion thus violates the principle that the sense of a sentence is a way of thinking of its reference. Worse, it means that, on the neo-Fregean view, a sentence may have a different truth-value than does the thought it expresses.

As far as I can see, this is the point at which contemporary defenders of the classical Fregean view rest their case against the neo-Fregean view.[10] But it seems rash, to me, to break off the investigation here. One of the crucial assumptions of the argument just presented was that the sentence 'Plato believes that Socrates is wise' expresses the thought W[Believes(Plato, Wise(Socrates))]. The expression 'W[Believes(Plato, Wise(Socrates))]' was introduced on the supposition that it was a harmless piece of shorthand for the ordinary English expression 'the thought that Plato believes that Socrates is wise'. But, in the context of the neo-Fregean theory and the principle Truth-Comp (which, I should stress, I think the neo-Fregean ought to accept), it ends up expressing a substantive view about the nature of the thoughts expressed by sentences containing the operator 'that ρ'. If the neo-Fregean can reject that view of the nature of such thoughts, the objection can be avoided. In other words, for all we have seen, it is possible that the real problem which is revealed by the above argument resides in our acceptance of the claim that 'Plato believes that Socrates is wise'

[8] See the derivation at the end of §2.

[9] Again, see the derivation at the end of §2.

[10] I take the argument I have just given to be a version of the "metalanguage argument" given by Burge (2004: 190–98) and Parsons (2009: 53–56).

expresses the thought W[Believes(Plato, Wise(Socrates))]. Of course in order to reject that claim, we do need to answer some questions: If 'Plato believes that Socrates is wise' does not express W[Believes(Plato, Wise(Socrates))], then what thought does it express? And, for that matter, which English sentence expresses the thought W[Believes(Plato, Wise(Socrates))], if this one does not? More generally, how are we to understand the thought W[Believes(Plato, Wise(Socrates))], if not through our prior understanding of (the English expression) 'the thought that Plato believes that Socrates is wise'?

What is clear is that more needs to be said about the understanding of thoughts which informs our W notation. Without it, we can have little confidence that we understand principles which employ it, principles like Truth-Comp, Sense-Comp, Sense-Comp-T, and even W-Comp. The task of the remainder of this paper (especially of §4.3) is thus to work out a more satisfying account of the nature of thoughts, and to be sure that our W notation is adequate to the task of describing them. Interestingly, it will turn out that our conception of thoughts and their composition rests more heavily on the analogy between thought and language than we might have expected. At the same time, however, that analogy suggests a way of treating the new principle of extensionality which closely parallels the above way of treating the (old) principle of extensionality. The treatment also turns out to leave precious little room for the classical Fregean to object.

In §4.2 I take a brief detour from my main line of argument to show that the neo-Fregean theory cannot be saved by assigning a sense and a reference to the operator 'that ρ'. The upshot of this section is that, if we wish to save the neo-Fregean theory of indirect contexts, we must show that having a sense and a reference is not the only way for an expression to be logically significant. Since the argument is rather dense, and since the details of it are not essential to the overall argument of this paper, some readers may wish to skip directly to §4.3, where I explain and defend my view of the logical significance of 'that ρ', and introduce a set of principles which provides a consistent elaboration of the neo-Fregean theory of indirect contexts.

4.2 A sense and reference for 'that ρ'?

What we want, again, is for the thought expressed by the sentence 'Plato believes that Socrates is wise' to have a feature which can account for its anomalousness in the way in which the presence of 'that ρ' accounts for the anomalousness of the sentence itself, on the neo-Fregean view of it. That is, we need a principle that stands to Truth-Comp as Ref-Comp-T stands to Ref-Comp, and so we need

the thought expressed by 'Plato believes that Socrates is wise' to have a feature which can be exploited in that principle in basically the way in which Ref-Comp-T exploits the presence of the expression 'that ρ' in the sentence itself. The general strategy must therefore be, in essence, to find something in the thought which corresponds to the operator 'that ρ' in the sentence which expresses it.

An obvious proposal, then, is that the second objection to the neo-Fregean theory rests on the mistaken assumption that the operator 'that ρ' does not express a sense. If it did, then perhaps we could formulate the principle we want by means of a simple adaptation of Ref-Comp-T. First, here is Ref-Comp-T again:

> **Ref-Comp-T** The reference of a sentence '$F(a_1, \ldots, a_n$, that $p)$' is the result of applying the reference of the predicate '$F(\xi_1, \ldots, \xi_{n+1})$' to the references of the terms 'a_1', ..., 'a_n' and the sense of the sentence 'p' (in the appropriate order).

The new principle would be obtained by replacing talk of expressions with talk of the ways of thinking they express. Let T(x) be the reference of 'that ρ'. (It may aid comprehension, here, to assume that T(x) is the identity function, that is, that it is a function that maps any given thought onto itself.) Its sense, then, is W[T(x)]. Then the adaptation of Ref-Comp-T is:

> **Truth-Comp-T** The truth-value of a thought W$[F(a_1, \ldots, a_n, \mathrm{T}(p))]$ is the result of applying the function presented by W$[F(\xi_1, \ldots, \xi_{n+1})]$ to the objects presented by W$[a_1]$, ..., W$[a_n]$ and the thought W$[p]$ (in the appropriate order).

Unfortunately, the assumption that 'that ρ' expresses a sense renders the principle Truth-Comp-T incompatible with other neo-Fregean commitments. In particular, if 'that ρ' expresses W[T(x)], it would seem that the sense of the sentence 'Plato believes that Socrates is wise', for example, is W[Believes(Plato, T(Wise(Socrates)))]. And, while Truth-Comp-T tells us that the truth-value of this thought is Believes(Plato, W[Wise(Socrates)]), as we wanted, the earlier principle Truth-Comp is *also* applicable to this thought, and tells us that its truth-value is Believes(Plato, T(Wise(Socrates))). For W[T(Wise(Socrates))] is a complete way of thinking, and so W[Believes(Plato, T(Wise(Socrates))] has the form W$[F(a_1, \ldots, a_n)]$, which is the form of thought to which Truth-Comp applies.[11]

Since Truth-Comp now ends up applying to thoughts expressed by sentences containing the operator 'that ρ', we might think that the thing to do is simply to

[11] It is important, here, that I assume that 'a_1', ..., 'a_n' may be *complex* terms, and so that W$[a_1]$, ..., W$[a_n]$ may be complex complete ways of thinking.

drop Truth-Comp-T and let Truth-Comp do all the work. Doing so would allow us to avoid the direct conflict just adduced. But it will not save the theory. For, again, Truth-Comp tells us that the truth-value of the thought W[Believes(Plato, T(Wise(Socrates)))] is Believes(Plato, T(Wise(Socrates))). At the same time, the neo-Fregean theory tells us that the truth-value of the sentence 'Plato believes that Socrates is wise' is Believes(Plato, W[Wise(Socrates)]). This is a problem, because W[...] is not, and cannot be, a function, while T(x) *is* a function. Thus, for the very reason that W[...] cannot be a function, T(x) cannot do the work the neo-Fregean needs it do. That is, no matter what function T(x) is—no matter what function we assign as the reference of 'that ρ'—Believes(Plato, T(Wise(Socrates))) is potentially distinct from Believes(Plato, W[Wise(Socrates)]). (Again, consider the case in which Plato believes some true thought, but does not believe specifically that Socrates is wise. In that case, the former truth-value will be the True, while the latter will be the False.)

The crucial issue here is just this. We have assumed a conception of sense on which the sense of an expression is a way of thinking of its reference. In executing this assumption, we took the more specific view that expressions refer to objects, functions, and truth-values, and thus that they express ways of thinking of objects, functions, and truth-values. Since thoughts, finally, are then taken to be composed of such ways of thinking, it turns out that the only way in which two thoughts can differ is for them to have different ways of thinking as parts. Once all of these views are in place, the only way of seeing the operator 'that ρ' as playing a logical or semantic role in our language is to claim that it has a sense and a reference. Since it is an incomplete expression, however, its reference must be a function, and so its contribution to determining the truth-value of a sentence or thought can only be functional. But if the truth-value of the thought that Plato believes that Socrates is wise is to depend on the thought that Socrates is wise—as every Fregean agrees it does—then we cannot consistently claim, as the neo-Fregean does, that the thought that Socrates is wise is a component part of the thought that Plato believes that Socrates is wise. For, if it were, then the truth-value of the whole thought could only depend on the truth-value of the thought that Socrates is wise. On our assumptions, all that the thought that Socrates is wise can contribute to the determination of the truth-values of thoughts in which it occurs is its truth-value. And whatever function we apply to that truth-value—whatever function we take to be the sense of 'that ρ'—the value of that function will never be the *thought* that Socrates is wise. There may, of course, be a function from the True to the thought that Socrates is wise. But this cannot be the reference of 'that ρ'. If it were, every propositional attitude ascription of the form 'Plato believes that p' in which

'p' is a true sentence would have the same truth-conditions. More specifically, to revert to an earlier example, 'John believes that Hesperus is a planet' and 'John believes that Phosphorus is a planet' would be guaranteed to have the same truth-conditions.

Again, the neo-Fregean theory seems hopeless. Even here, however, there is still room to wonder: *is* the only way of seeing the operator 'that ρ' as playing a logical or semantic role in our language to assume that it has a sense and a reference? In the next section, I give some reasons for thinking that it is not.

4.3 The logical role of 'that ρ'

We saw in the last section that the logical role which the neo-Fregean implicitly assumes is played by the operator 'that ρ' cannot be explained by assuming that it has a sense and a reference.[12] The classical Fregean claims, in effect, that there is no other way to explain how it could play that logical role, and thus that the neo-Fregean is mistaken in assuming that it does so.

Of course, even on the classical theory, the operator 'that ρ' must be seen as playing an important role in *English* sentences in which it occurs. Its role, on that theory, is to indicate that the expressions which occur within its scope are used with their indirect sense and reference. This role, however, is seen as inessential. In a proper *Begriffsschrift*, in which all ambiguity is eliminated, there will be no distinctions among linguistic contexts. Occurrences of, for example, English words in indirect contexts will be replaced, in the corresponding *Begriffsschrift* sentences, with expressions which are distinct from those which replace occurrences of the same words in transparent contexts. As Frege himself puts, in a letter to Russell: "[t]o avoid ambiguity, we ought really to have special signs in indirect speech" (1980: 153). Since these "special signs" already have the senses and references they need to have to ensure that the relevant *Begriffsschrift* sentences—those which ascribe propositional attitudes—have the right senses and references, there is no work left for 'that ρ' to do. So it can be omitted from the *Begriffsschrift* entirely. To be sure, a *Begriffsschrift* constructed along these lines would be an unwieldy thing, since there would not necessarily be any systematic connection between the two expressions which replaced, say, the English name 'Socrates' in each of its relevant uses: we might as well use 'Moe' for the first and 'Larry' for

[12] At least not if its reference is a function, and its sense a way of thinking of that function. Perhaps there is a possible view on which the sense of 'that ρ' is something other than a way of thinking of a function. But I confess I do not know what else it could be. At any rate, since such a view would represent a radical departure from Frege's, I refrain from considering it further here.

the second. So Frege, wisely, immediately adds: "though their connection with the corresponding signs in direct speech[13] should be easy to recognize."

What Frege's reasonable addition threatens to obscure, however, is the fact that the relation between an occurrence of an expression in a transparent context and an occurrence of the same expression in an indirect context is nothing like the relation between, say, the use of 'bank' to refer to stretches of river-proximate land and its use to refer to financial institutions. The ambiguity in question is not mere lexical ambiguity. If it were, our lexicon would need to include multiple entries for each word of the language: one which gives its transparent sense and reference, one which gives its indirect sense and reference, and so on—all the way up Frege's hierarchy. This is, of course, Davidson's learnability objection again. Since contemporary defenders of the classical Fregean theory of indirect contexts claim to have a response to it—a response which, for all I have said here, is successful—it may seem odd for me to bring in the objection just now. But it is worth mentioning in this context because it suggests that English (and, no doubt, very many other natural languages) have already managed to solve the problem for which the classical Fregean needs to introduce a whole theory of indirect senses and references. And that solution lies in the role played, in English, by the word 'that', a solution which seems to work quite well (at least in practice), despite the fact that the majority of us know nothing at all about these supposed indirect senses.

Somewhat less polemically: it is agreed, by the defenders of the classical Fregean view, that the word 'that', in English, does in some way mark a logical distinction. They just think that the logical distinction itself is to be explained by an account of indirect senses. But it is worth asking whether there is an explanation of the logical distinction which does not require an account of indirect senses. That is, it is worth asking whether the role played by 'that' in English might itself provide an explanation of the logical distinction in question. The fundamental conviction of those who yearn for semantic innocence is that it can. Davidson himself provided one such explanation.[14] And it seems to me that something more like Davidson's explanation can be made to fit with many of Frege's most important insights. So let me say, finally, how I think the explanation should go.

To begin with, there is a sense in which the principle Ref-Comp-T itself provides an explanation of the logical function of the operator 'that ρ'. The challenge

[13] I take it that Frege means to be referring to transparent contexts, and not to quotation, though the latter phenomenon is sometimes called "direct discourse" (as opposed to "indirect discourse," which is my topic here).

[14] Cf. Davidson 1968.

was only to see how that explanation could be adapted to apply to thoughts, rather than to the sentences that express them. In this context, it is worth noting that we have already assumed, and applied, a solution to a parallel problem in the case of sentences involving asymmetric binary predicates, and the thoughts they express.

Consider the sentences 'Romeo loves Juliet' and 'Juliet loves Romeo'. These sentences clearly have different truth-conditions. Although, in the play, both sentences are true, they need not have been: Juliet might have preferred a suitor more acceptable to her family, and the whole tragedy might thereby have been averted. Fortunately, REF-COMP bears this out: it tells us that the reference of the first sentence is the result of applying the function Loves(x,y) to the arguments Romeo and Juliet, *in that order*, and that the reference of the second is the result of applying the same function to Juliet and Romeo, in *that* order. Our familiar functional notation takes care of the rest: the first sentence refers to the truth-value Loves(Romeo, Juliet), while the second refers to the truth-value Loves(Juliet, Romeo), and we all already know that these are potentially distinct.

But now suppose that we were given the relevant thought-components—W[Romeo], W[Juliet], and W[Loves(x,y)]—without being given the sentences. And suppose we were then asked to give the result of completing W[Loves(x,y)] by W[Romeo] and W[Juliet]. The appropriate response, no doubt, is to ask in which order W[Romeo] and W[Juliet] are supposed to complete W[Loves(x,y)]. Is W[Romeo] supposed to go in the *x*-gap or the *y*-gap? And yet, what justifies us in thinking that this question is a good one? What justifies us in thinking that it has an answer?

What justifies us in thinking that the question is a good one (if anything does) is that we know to what use these ways of thinking—W[Romeo] and the rest—are to be put. They figure in an explanation of the way in which the semantic features of a sentence depend on the semantic features of its parts. But, of course, they provide only part of the explanation. The remainder is supplied by other features of the sentence, in particular by its syntactic structure, by the way in which its semantically relevant parts are *combined* in the whole. Our principle REF-COMP gave the correct answer to our initial question (about the respective truth-conditions of 'Romeo loves Juliet' and 'Juliet loves Romeo') because it exploits the syntactic structure of the sentence. In particular, it is sensitive to the order in which terms occur in the sentence. Parallel points explain the success of TRUTH-COMP, as applied to the thought W[Loves(Romeo, Juliet)]. That principle can provide an answer to the question what the truth-value of this thought is only because it exploits the ordering of signs in our notation, and transfers this idea of order from the signs to the ways of thinking they express.

The exploitation of syntactic structure in the case of an explicitly linguistic (semantic) principle like Ref-Comp is rightly uncontroversial. Semantic theory begins from the assumption that both the meanings of words and the syntactic structure of the sentences in which they occur are relevant to determining the meanings of those sentences. What is potentially objectionable is only the analogical application of the idea of syntactic structure to Fregean thoughts. My point here is not that there is definitely nothing objectionable in this application. I wish only to make two weaker points. The first is that, if there *is* something objectionable about the idea that thoughts have structure in a sense analogous to the sense in which sentences have structure, that would threaten to undermine the *whole* of Frege's theory of thoughts (*Gedanken*). So even the *classical* Fregean needs this idea; without it, she will have no adequate account of the thoughts expressed by sentences containing asymmetric binary predicates. My second point is that, because the classical Fregean must assume that there is nothing objectionable in the idea that thoughts have a structure analogous to the structure of the sentences that express them, it is open to the neo-Fregean to exploit that idea in other ways.

In order to see that the idea *can* be exploited in other ways, we first need to recognize that the relevant idea of *order*, even in the case of straightforwardly linguistic (semantic) principles like Ref-Comp, is not the idea of a spatial relation. It is not the fact that 'Romeo' stands to the left and 'Juliet' to the right of 'loves' in the sentence 'Romeo loves Juliet' that makes Ref-Comp work. What makes Ref-Comp work is rather the fact that the spatial relations between these expressions is given a particular logical and semantic significance. And many other facts, including non-spatial facts, could have had the same significance. We might adopt a convention whereby the spatial location of the names is completely irrelevant, and the "order" of the names, in the sense relevant to Ref-Comp, is indicated by the color or the font in which the name is written, or by the *temporal* order in which we write them, or whatever. What is essential is only that there be *some* difference between these sentences which can be interpreted as indicating logical ordering, as indicating which terms supply arguments to which places in the given function, and which our compositional principles can therefore exploit.[15]

At this level of abstraction there is no obvious reason not to make a parallel move in the case of sentences like 'Plato believes that Socrates is wise'. Just as we see the spatial ordering of terms as having logical or semantic significance, so we can see the presence of the operator 'that ρ' as having logical or semantic

[15]I am indebted, in this paragraph, and in my thinking on this topic more generally, to Diamond 2012.

significance. In fact, the basic idea is already given by our principle Ref-Comp-T. This principle tells us that the effect of the presence of the operator 'that ρ' is that the truth-value of the sentence depends on the senses of the words which occur within its scope, rather than on their references. That is, Ref-Comp-T *already* provides an explanation of the logical or semantic role played by 'that ρ'. The problem that has been plaguing the neo-Fregean theory since the beginning of this section is caused by the fact that nothing in the corresponding principle Sense-Comp-T reflects that role. The latter principle treats an occurrence of an expression within a that-clause as if it were an ordinary occurrence. And while there is a tempting understanding of the neo-Fregean theory of indirect contexts on which its central thesis is precisely that every occurrence of an expression *is* an ordinary occurrence—because an expression always has its customary sense and reference—what I am claiming is that this understanding is fundamentally mistaken. On the contrary, there *is* something unusual about the occurrence of an expression in an indirect context: namely, that it occurs within the scope of the operator 'that ρ'. But if that is what makes the occurrence unusual, then it had better be the presence of the operator that explains the anomalousness of the resulting sentences—and so, too, that of the thoughts they express.

I have claimed that Ref-Comp-T is exactly right, but that we need to make some kind of adjustment to Sense-Comp-T. I have also claimed that an analogous problem—that concerning sentences like 'Romeo loves Juliet' and 'Juliet loves Romeo'—was already solved by the appeal to the logical or semantic ordering of terms in the principles Ref-Comp and Sense-Comp. The fact that *both* of these principles appeal to such ordering suggests that, despite initial appearances to the contrary, Sense-Comp-T does not actually exploit the presence of the operator 'that ρ'. And this proves to be the case. Compare again the two principles:

> **Sense-Comp** The sense of a sentence '$F(a_1, \ldots, a_n)$' is the result of completing the sense of the predicate '$F(\xi_1, \ldots, \xi_n)$' by the senses of the terms 'a_1', ..., 'a_n' (in the appropriate order).
>
> **Sense-Comp-T** The sense of a sentence '$F_1(a_1, \ldots, a_n$, that $p)$' is the result of completing the sense of the predicate '$F_1(\xi_1, \ldots, \xi_{n+1})$' with the senses of the terms 'a_1',..., 'a_n' and the sense of the sentence 'p' (in the appropriate order).

Although the syntax provided in §2 ensures that Sense-Comp does not apply to sentences containing the operator 'that ρ' (because terms of the form 'that p' do not have a sense, and so the principle can, as it were, return no response to our query), the result that it *would* give if we simply ignored the occurrence of the

operator 'that ρ' in the relevant sentences is precisely the result that SENSE-COMP-T gives.

The point becomes slightly clearer if we suppose, for a moment, that sentences are (complex) terms, and so that 'Plato believes Socrates is wise' (without the operator 'that ρ'), for example, is a well-formed sentence. On this supposition, in conjunction with the semantic axioms given in §2, SENSE-COMP tells us that the sense of 'Plato believes Socrates is wise' is W[Believes(Plato, Wise(Socrates))]. But this is just what SENSE-COMP-T tells us is the sense of the distinct sentence 'Plato believes *that* Socrates is wise'. So SENSE-COMP-T is, in effect, insensitive to the presence of the operator 'that ρ'.

We can reformulate SENSE-COMP-T to be sensitive to the presence of the operator 'that ρ' if we distinguish two different ways in which one thought can occur in another. This distinction is analogous to the distinction between different orders in which two different complete ways of thinking W[a] and W[b] can complete the same incomplete way of thinking W[F(x,y)]: either so as to result in W[F(a,b)], or so as to result in W[F(b,a)]. Both distinctions, I have argued, are grounded in syntactic features of relevant sentences, features which have been imbued (by our linguistic practice) with a logical significance which is captured in the principles REF-COMP and REF-COMP-T. I thus propose that, just as we can say that W[Romeo] occurs "to the left" and W[Juliet] "to the right" of W[Loves(x,y)] in W[Loves(Romeo, Juliet)], so we can say that the thought that Socrates is wise occurs *indirectly* in the thought that Plato believes that Socrates is wise, and that it occurs *transparently* in the thought that if Socrates is wise then Plato is wise.

All we need now is to introduce into the W notation a way of indicating whether an occurrence of a thought in another thought is an indirect one or a transparent one. (Our way of handling order in the W notation was much easier, since it was already taken care of by the ordering implicit in our use of function–argument notation.) Thus, to refer to the thought that Plato believes that Socrates is wise—in which, on the neo-Fregean theory, the thought that Socrates is wise occurs indirectly—I will use:

$$\text{W[Believes(Plato, }\uparrow\text{[Wise(Socrates)])]}$$

(using square brackets, as with W, to indicate scope; and also to remind us that $\uparrow$, again like W, is not a function).[16] We then update and supplement W-COMP, our

[16]Notice that, on this notation, we will refer to the thought that Aristotle believes that Plato believes that Socrates is wise by

$$\text{W[Believes(Aristotle, }\uparrow\text{[Believes(Plato, }\uparrow\text{[Wise(Socrates)])])].}$$

principle for the combination of ways of thinking, accordingly (revising it slightly to ensure that the principles work smoothly together):

> **W-Comp′** If a way of thinking W[$F(x_1, \ldots, x_i, \ldots, x_n)$] is completed *transparently* (at the appropriate place) by a way of thinking of an object W[a_i], the result is the way of thinking W[$F(x_1, \ldots, a_i, \ldots, x_n)$].
>
> **W-Comp-T** If a way of thinking W[$F(x)$] is completed *indirectly* by a way of thinking of a truth-value (i.e., by a thought) W[p], the result is the way of thinking W[$F(\uparrow[p])$].

Our old principle Sense-Comp-T is then replaced by a principle which indicates that the sense of a sentence which occurs within the scope of the operator 'that ρ' occurs indirectly in thoughts expressed by sentences which contain the whole that clause. More precisely (including, again, the appropriate revisions of Sense-Comp):

> **Sense-Comp-T′** The sense of a sentence 'F(that p)' is the result of completing the sense of the predicate '$F(\xi)$' *indirectly* by the sense of the sentence 'p'.
>
> **Sense-Comp′** The sense of an expression '$F(\xi_1, \ldots, a_i, \ldots, \xi_n)$' is the result of completing the sense of the predicate '$F(\xi_1, \ldots, \xi_i, \ldots, \xi_n)$' (at the appropriate place) *transparently* by the sense of the term 'a_i'.

We can also finally state the principle which stands to Ref-Comp-T as Truth-Comp stands to Ref-Comp:

> **Truth-Comp-T′** The truth-value of a thought W[$F(\uparrow[p])$] is the result of applying the function presented by the incomplete way of thinking W[$F(x)$] to the thought W[p].

Application of the principles given below shows that the iteration of ↑ is redundant: W[Socrates], for example, both is a component of the whole thought and contributes *itself* to the determination of the truth-value of the thought, just as it does in the case of the simpler thought W[Believes(Plato ↑[Wise(Socrates)])]. But this is exactly what we should have expected, since we have not given any sense to the idea that a thought can occur "doubly" indirectly in another thought. So if a thought occurs indirectly in a thought which occurs indirectly in another thought, then the first thought occurs indirectly in the third. We do not need a separate principle to tell us this about ↑, because ↑[Wise(Socrates)] is not a part of the above thought about Aristotle. The relevant part of the latter is, rather W[Wise(Socrates)]. This reflects the fact that '↑' is not conceived of as picking out a *part* of the thoughts it is used to talk about (any more than 'W' itself is: recall that the only parts of W[Wise(Socrates)] are W[Wise(x)] and W[Socrates], and that these are *simple* thought-components).

Sentences like 'Plato believes that Socrates is wise' now come out as legitimate exceptions to the new principle of extensionality. Since we have kept Ref-Comp-T intact, the reference of this sentence is still Believes(Plato, W[Wise(Socrates)]). According to the background theory of thoughts captured by the (newly extended) W notation, its sense is the thought W[Believes(Plato, ↑[Wise(Socrates)])]. And this thought is precisely what Sense-Comp′ and Sense-Comp-T′ tell us is the sense of the sentence. Finally, if we apply Truth-Comp-T′, we learn that the truth-value of this thought is the result of applying the function Believes(x, y) to Plato and the thought W[Wise(Socrates)] (rather than the truth-value of which that thought is a way of thinking, as the new principle of extensionality would require): namely, Believes(Plato, W[Wise(Socrates)]). At long last, the sense of our sentence is a way of thinking of its reference! Just as important, the truth-value of our sentence is the same as the truth-value of the thought it expresses. So, on this final way of developing it, the neo-Fregean theory is consistent after all.

Conclusion

As I said in the Introduction, there is a growing consensus that the only genuinely Fregean theory of propositional attitude ascriptions is one which, like Frege's own, maintains that embedding an expression in an indirect context shifts its sense and reference. While Frege's theory has, indeed, recently been shown to be much more plausible than his early opponents claimed,[17] the arguments of the present essay suggest that the consensus rests mainly on some unquestioned Fregean dogmas. In particular, it rests on what I have called the new principle of extensionality, and on the conception of thoughts and their composition which supports it.

What I have shown here is that there is a way of developing the neo-Fregean theory on which it is both consistent and plausible. Perhaps it nonetheless false—I do not claim to have demonstrated that the classical Fregean theory is false, and the two theories cannot both be true. But those who wish to reject the neo-Fregean theory in favor of the classical Fregean theory face a challenge. In order to reject the neo-Fregean theory, they need to show that the presence of the word 'that' in sentences like 'Plato believes that Socrates is wise' *cannot* be given the logical or semantic significance the neo-Fregean claims it has. That is, they need to show

[17] Such, I take it, is the upshot of the account of indirect (or canonical) senses recently proposed by Burge 2004: 170–76, Parsons 2009: 56–57, Peacocke 2009: 164–169, and Kripke 2008: 268–72. (Based on what Kripke says in that paper, his unpublished Whitehead Lectures are also highly relevant, though I have not yet been lucky enough to read them.)

that there can be no feature of a thought like the thought that Plato believes that Socrates is wise which can play a logical role in that thought which is analogous to the role played by 'that' in sentences like 'Plato believes that Socrates is wise'. And they need to do all this without *also* showing that the order of the names 'Romeo' and 'Juliet' in a sentence like 'Romeo loves Juliet' cannot be given the logical or semantic significance they themselves claim it has. That is, they need to avoid undermining their own view that there is a feature of a thought like the thought that Romeo loves Juliet which plays a logical role in that thought which is analogous to the role played by the order of the names 'Romeo' and 'Juliet' in the sentence 'Romeo loves Juliet'. If their objection to the neo-Fregean theory, as I have developed it here, undermines their own appeal, in the theory of Fregean thoughts, to the logical or semantic significance of the order of terms in sentences which express those thoughts, then the objection becomes, not an objection to an idiosyncratic and questionably Fregean attempt to avoid Frege's hierarchy, but, instead, an objection to Fregeanism itself.

Appendix

The question I wish to address here is this: Why shouldn't the reference—that is, the truth-value—of a whole propositional attitude ascription be allowed to depend on the *senses*, rather than the references, of some of its parts, namely, those of its parts which occur in its that-clause?

The question is pressing in part because it is a relatively simple matter to state a view on which the truth-value of a propositional attitude ascription does depend on the senses of those of its parts which occur in its that-clause. I begin, then, by stating such a view. To keep matters simple, and reasonably intuitive, I state the view for a tiny fragment of English.[18]

Vocabulary. Consider a language that has only the following vocabulary:

Terms: Socrates, Plato, Aristotle

Predicates: ξ is wise, ξ believes ζ

We want, of course, to provide an account of sentences like 'Plato believes that Socrates is wise', hence, in effect, of the expression 'ξ believes that ρ' (where 'ξ'

[18]The method I follow here is based on that employed by Parsons (2009).

marks a place for a term and 'ρ' marks a place for a sentence). To that end, we need to introduce the expression 'that' as well. We thus add to our vocabulary:

Operators: that ρ

Syntax. The rules for the construction of complex expressions are simple. There are only two.

SYNT-1 A predicate completed by an appropriate number of terms is a sentence.

SYNT-2 An operator completed by an appropriate number of sentences is a term.

Sentences of our language fragment thus include 'Socrates is wise', 'Aristotle is wise', 'Plato believes Socrates', 'Plato believes that Socrates is wise', and 'Aristotle believes that Plato believes that Socrates is wise'.

Semantics. The semantics proceeds in two stages. First, we assign both a sense and a reference to each basic expression of the language.[19] Second, we provide compositional rules which determine the sense and reference of any sentence, given (i) the senses and references of its component parts and (ii) the way in which those parts are combined. I begin by providing a semantics for the core of the language fragment introduced above, i.e., that part of it which excludes

[19] Alternatively, we could say that the semantics assigns an *intension* and an *extension* to each expression of the language. On this construal, the extension of a term is an object, its intension a way of thinking of an object; the extension of a predicate is a function, its intension a way of thinking of a function; and the extension of a sentence is a truth-value, its intension a way of thinking of a truth-value. Stating the view this way may make it look somewhat less Fregean, since it makes no use of the concepts of sense and reference. But it still shares quite a lot with Frege's view, since, at least for sentences containing no indirect contexts, the compositional semantics of the two views make reference to exactly the same entities: the entities Frege assigns as senses, this view assigns as intensions; the entities Frege assigns as references, this view assigns as extensions. I think that this similarity is enough to justify treating the view presented in this section as at least neo- or quasi-Fregean. Others may see things differently. Either way, the present section demonstrates the possibility of such a view, and the possibility of the view makes our question pressing. However we wish to describe the view presented here, the question is, why not accept it, rather than Frege's view? Or, in the terms employed in the present footnote: why shouldn't the *extension* of a propositional attitude ascription be allowed to depend on features of its parts other than their *extensions*? In particular, why shouldn't it be allowed to depend on their *intensions*? Putting the question this way may help to free it from some of the historical baggage carried by Frege's terminology.

the operator 'that ρ'. The semantics of 'that ρ' (more precisely: of sentences containing it) will be considered shortly.

First, then, the senses and references of the basic expressions are as follows:

	Sense	**Reference**
'Socrates'	<Socrates>	Socrates
'Plato'	<Plato>	Plato
'Aristotle'	<Aristotle>	Aristotle
'ξ is wise'	<Wise(x)>	Wise(x)
'ξ believes ζ'	<Believes(x, y)>	Believes(x, y)

The metalanguage here is just philosophical English, with two additions. First, I use function–argument notation (of the form 'F(x)', 'G(x, y)', etc.) to refer to Fregean functions. Second, I use expressions of the form '<...>' to refer to ways of thinking. More specifically, an expression of the form '<...>' will refer to a particular way of thinking of the object, function, or truth-value denoted by the metalanguage expression '...'. The particular way of thinking in question can be given by saying that <...> is the unique way of thinking of ...*as* .... For example, <Socrates> is the unique way of thinking of Socrates *as* Socrates. It is also, I assume, the customary sense of the object-language term 'Socrates'.

For simplicity, I assume a uniform account of Fregean senses as ways of thinking of entities of appropriate kinds. Thus, in every case, the sense of an expression is a way of thinking of its reference. We could, no doubt, raise questions about the view that the sense of a predicate is a way of thinking of a function, or that the sense of a sentence—a thought—is a way of thinking of a truth-value. But I will not consider such questions here. In the present context, the chosen way of speaking serves merely to require that the sense of an expression determines its reference, in the specific sense that, if two expressions have the same sense, then they also have the same reference. I take it that thoughts do determine truth-values in this sense; i.e., two sentences which express the same thought have the same truth-value. I also take it that, whatever the sense of a predicate is, it determines a Fregean function in this sense; i.e., two predicates with the same sense will be true of, or satisfied by, the same things.

Finally, we need to provide compositional rules which fix the sense and reference of each sentence on the basis of the senses and references of its parts and the way in which those parts are combined. There will be a rule for the composition of reference and a rule for the composition of sense.

The rule for the composition of reference is already suggested by our choice of references: since the reference of a predicate is a function, we can take the relevant composition to consist in function–argument application, taking the references of appropriate terms as corresponding arguments to the function. So (giving the rule in full generality) the reference of a sentence '$F(a_1, \ldots, a_n)$' will be the result of applying the function $F(x_1, \ldots, x_n)$ to the references of the terms 'a_1', ..., 'a_n', with the reference of each 'a_i' serving as the x_i argument to the function. Since the reference of a predicate is, I suppose (with Frege), a function from an object or objects to a truth-value, the rule implies that the reference of a sentence will be a truth-value.

In order to give a rule for the composition of sense, we can treat ways of thinking of objects and ways of thinking of functions as being, like the expressions that express them, complete and incomplete, respectively. So a way of thinking of an object <*a*>, which is the sense of a term, is complete; while a way of thinking of a function <*F(x)*>, which is the sense of a predicate, is incomplete. We can then state a rule of combination for ways of thinking: the result of completing a way of thinking of a function by an appropriate number of ways of thinking of objects is a *thought*.[20] That is, to use the angle bracket notation, the result of completing <*F(x)*> by <*a*> is <*F(a)*>.

I should stress, here, that the incomplete way of thinking <*F(x)*> and the complete way of thinking <*a*> are component parts of the thought <*F(a)*>—just as the predicate 'ξ is wise' and the term 'Socrates' are component parts of the sentence 'Socrates is wise'. Our angle brackets are thus functioning in roughly the way that quotation marks function in philosophical English, except that they are used to mention ways of thinking rather than expressions.

[20]This way of stating the rule implicitly restricts it to ways of thinking of functions from objects to truth-values, hence to (ways of thinking that serve as) senses of predicates. We will not want to say that the result of completing a way of thinking of a function from objects to objects results in a thought, since the resulting (complete) way of thinking will be the sense of a (complex) term, and so will be a way of thinking of an object rather than a truth-value. But since the semantics for our sample language fragment will require no functions other than functions from objects to truth-values, I will ignore these complications here. (It is, in any case, easy enough to apply the same ideas to ways of thinking of functions from objects to objects.) The same remarks go for the statement of W-Comp below.

With these points in mind, the general rule for combining ways of thinking can be stated as follows:

> **W-Comp** If an incomplete way of thinking $<F(x_1, \ldots, x_n)>$ is completed (in the appropriate order) by complete ways of thinking $<a_1>, \ldots, <a_n>$, the result is the thought $<F(a_1, \ldots, a_n)>$.

We can use this rule for combining ways of thinking to state a closely related rule for composing senses: the sense of a complex expression is the result of combining the ways of thinking that are the senses of its component parts.

We thus have the following two rules of composition for the core language:

> **Ref-Comp** The reference of a sentence '$F(a_1, \ldots, a_n)$' is the result of applying the reference of the predicate '$F(\xi_1, \ldots, \xi_n)$' to the references of the terms 'a_1', ..., 'a_n' (in the appropriate order).
>
> **Sense-Comp** The sense of a sentence '$F(a_1, \ldots, a_n)$' is the result of completing the sense of the predicate '$F(\xi_1, \ldots, \xi_n)$' by the senses of the terms 'a_1', ..., 'a_n' (in the appropriate order).

These principles allow us to calculate the sense and reference of the complex expressions of our language fragment. For example, we can calculate the sense and reference of the sentence 'Socrates is wise' as follows.

The reference of 'Socrates is wise'

1. The reference of 'Socrates is wise' is the result of applying the reference of 'ξ is wise' to the reference of 'Socrates'. (By Ref-Comp)
2. The reference of 'ξ is wise' is Wise(x). (Axiom)
3. The reference of 'Socrates' is Socrates. (Axiom)
4. So the reference of 'Socrates is wise' is the result of applying Wise(x) to Socrates, viz., Wise(Socrates). (From 1–3 and function–argument application)

The sense of 'Socrates is wise'

1. The sense of 'Socrates is wise' is the result of completing the sense of 'ξ is wise' by the sense of 'Socrates'. (By Sense-Comp)
2. The sense of 'ξ is wise' is <Wise(x)>. (Axiom)
3. The sense of 'Socrates' is <Socrates>. (Axiom)
4. So the sense of 'Socrates is wise' is the result of completing <Wise(x)> by <Socrates>, viz., <Wise(Socrates)>. (From 1–3 and W-Comp)

In other words, the reference of the sentence 'Socrates is wise' is whatever truth-value is the value of the function Wise(x) for the argument Socrates, and its sense is the thought <Wise(Socrates)>, that is, the thought that Socrates is wise.

The semantics of 'that ρ'. To provide a semantics for 'that ρ'—and so to provide a semantics for propositional attitude ascriptions—we need to introduce another pair of principles for the composition of the senses and references of sentences which contain it. On the approach I take here, these principles do not assign a sense and reference to 'that ρ', and so do not assign a sense and reference to complex terms of the form 'that p' (for example, 'that Socrates is wise'). Nor will I introduce any other principles which do so.[21] Instead, we can introduce principles which determine the sense and reference of any *sentence* in which 'that ρ' occurs. The principles are these:[22]

> **Ref-Comp-T** The reference of a sentence '$F(a_1, \ldots, a_n, \text{that } p)$' is the result of applying the reference of the predicate '$F(\xi_1, \ldots, \xi_{n+1})$' to the references of the terms 'a_1', ..., 'a_n' and the sense of the sentence 'p' (in the appropriate order).[23]
>
> **Sense-Comp-T** The sense of a sentence '$F(a_1, \ldots, a_n, \text{that } p)$' is the result of completing the sense of the predicate '$F(\xi_1, \ldots, \xi_{n+1})$' with the senses of the terms 'a_1',..., 'a_n' and the sense of the sentence 'p' (in the appropriate order).[24]

We can then carry through the following derivations of the sense and reference

[21] In §4.2, I will briefly consider another approach, on which we do assign a sense and reference to 'that ρ'. Importantly, however, the original rule for the composition of reference, Ref-Comp, will *never* suffice as the only rule for the composition of reference, so long as our theory of indirect contexts is semantically innocent. This suggests that one motivation for Frege's view is to have just a single, perfectly general, rule of composition of reference, one that works for every single sentence of the language. Again, however: why think that we absolutely must make do with only one such rule?

[22] These principles are not perfectly general. In particular, they do not cover sentences which contain multiple that-clauses, except where each that-clause is embedded within another. But since the only sentences containing multiple that-clauses which are relevant to my purposes here are of the latter kind, we can make do with the principles given in the text.

[23] Note that this principle is not in competition, as it were, with the earlier principle Ref-Comp. For the earlier principle will simply fail to provide a reference if it is applied to a sentence which contains a term of the form 'that p'. The reason is that a term of this form does not have a reference, which means that there will be no argument to provide to the relevant function, and the rule can therefore return no value. In such a case, we, in effect, instruct our semantics to try the other rule, Ref-Comp-T, which *will* apply to a sentence containing a that-clause, and will return the appropriate truth-value. (These same remarks apply, *mutatis mutandis*, for the rules Sense-Comp and Sense-Comp-T.)

[24] As I said in the Introduction, this principle has been the target of an important objection, one which has seemed to a number of philosophers to constitute a decisive refutation of the sort of view presented in this section. I consider the objection, and provide a response, in §4.

of the ascription 'Plato believes that Socrates is wise'. (The semantics of 'that ρ' can also be applied recursively to allow us to derive the sense and reference of the iterated ascription 'Aristotle believes that Plato believes that Socrates is wise'.)

The reference of 'Plato believes that Socrates is wise'

1. The reference of 'Plato believes that Socrates is wise' is the result of applying the reference of 'ξ believes ζ' to the reference of 'Plato' and the sense of 'Socrates is wise'. (By Ref-Comp-T)
2. The reference of 'ξ believes ζ' is Believes(x,y). (Axiom)
3. The reference of 'Plato' is Plato. (Axiom)
4. The sense of 'Socrates is wise' is <Wise(Socrates)>. (Previous result)
5. So the reference of 'Plato believes that Socrates is wise' is the result of applying Believes(x,y) to Plato and <Wise(Socrates)>, viz., Believes(Plato, <Wise(Socrates)>). (From 1–4 and function–argument application)

The sense of 'Plato believes that Socrates is wise'

1. The sense of 'Plato believes that Socrates is wise' is the result of completing the sense of 'ξ believes ζ' by the sense of 'Plato' and the sense of 'Socrates is wise'. (By Sense-Comp-T)
2. The sense of 'ξ believes ζ' is <Believes(x,y)>. (Axiom)
3. The sense of 'Plato' is <Plato>. (Axiom)
4. The sense of 'Socrates is wise' is <Wise(Socrates)>. (Previous result)
5. So the sense of 'Plato believes that Socrates is wise' is the result of completing <Believes(x,y)> by <Plato> and <Wise(Socrates)>, viz., <Believes(Plato, Wise(Socrates))>. (From 1–4 and W-Comp)

In other words, the reference of 'Plato believes that Socrates is wise' is the value of the function Believes(x,y) for the arguments Plato and the thought that Socrates is wise, namely, a truth-value; and its sense is the relevant way of thinking of that truth-value, namely, the thought that Plato believes that Socrates is wise.

This theory works in precisely the way relevant to the question raised at the beginning of this section. Within the sort of system just described, the reference, i.e., the truth-value, of a propositional attitude ascription depends on the senses of the expressions that occur within its that-clause. Our question thus becomes: What is wrong with this view? (For convenience, I will hereafter refer to the view as 'the neo-Fregean view'. I will sometimes refer to Frege's own view as 'the classical Fregean view' or just 'the classical view'.)

References

Almog, Joseph and Paolo Leonardi (eds.) 2009. *The Philosophy of David Kaplan*. Oxford: Oxford University Press.

Burge, Tyler. 1979. Frege and the Hierarchy. In Burge 2005, 153–166.
— 2004. Postscript to "Frege and the Hierarchy". In Burge 2005, 167–210.
— 2005. *Truth, Thought, Reason*. Oxford: Oxford University Press.

Carnap, Rudolf. 1947. *Meaning and Necessity*. Chicago: The University of Chicago Press.

Davidson, Donald. 1965. Theories of Meaning and Learnable Languages. In Davidson 2001a, 3–15.
— 1968. On Saying That. In Davidson 2001a, 171–183.
— 2001a. *Inquiries Into Truth and Interpretation*. Oxford: Oxford University Press.

Diamond, Cora. 2012. What can you do with the general propositional form? In *Wittgenstein's Early Philosophy*. Ed. by José Zalabardo. Oxford: Oxford University Press.

Dummett, Michael. 1973. *Frege: Philosophy of Language*. London: Duckworth.

Frege, Gottlob. 1892. On Sense and Reference. In Frege 1960, 56–78.
— 1960. *Translations from the Philosophical Writings*. Ed. by Peter Geach and Max Black. Oxford: Basil Blackwell.
— 1980. *Philosophical and Mathematical Correspondence*. Ed. by Gottfried Gabriel et al. Trans. by Hans Kaal. Chicago: The University of Chicago Press.

Kripke, Saul. 2008. Frege's theory of sense and reference: some exegetical notes. In Kripke 2011, 254–291.
— 2011. *Philosophical Troubles*. Oxford: Oxford University Press.

Parsons, Terence D. 2009. Higher-Order Senses. In Almog and Leonardi 2009, 45–59.

Peacocke, Christopher. 2009. Frege's Hierarchy: A Puzzle. In Almog and Leonardi 2009, 159–186.