Meaning and Metatheory

by

Jason Stanley

Submitted to the Department of Linguistics and Philosophy in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY in Philosophy at the

Massachusetts Institute of Technology

January 1995

© 1995 Jason Stanley
All rights reserved

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part.

Signature of Author

Department of Linguistics and Philosophy
January 10, 1995

Certified by

Robert Stalnaker
Thesis Supervisor

Accepted by

Ned Block
Chair, Departmental Committee on Graduate Studies

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 26 1995

LIBRARIES

ARCHIVES
MEANING AND METATHEORY

by

Jason Stanley

Submitted to the Department of Linguistics and Philosophy
on January 10, 1995 in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Philosophy

ABSTRACT

Semantic theory has been used for many different philosophical purposes. This thesis investigates
two such uses of semantic theory. The first is the use of semantic theory in providing a justification
for a formal theory. The second is the use of semantic theory in yielding an account of
understanding.

The first paper is "Truth and Metatheory in Frege". In this paper, it is contended, against much
recent work in Frege interpretation, that Frege should be credited with the first semi-rigourous
formulation of a semantic theory. In so doing, it attempts to show that many of the arguments
which purport to establish that Frege could not have engaged in semantic theorizing suffer from
two kinds of misconceptions. The first misconception involves the notion of truth which a
philosopher must accept in order to engage in such theorizing. The second misconception involves
the sort of justification which these arguments assume a semantic theory attempts to provide.

The second paper is "Frege's Thesis". For each primitive expression in a language, two questions
may be asked. The first is: what is it in virtue of which that expression has the semantic value it
does? The second is: what is it in virtue of which a speaker counts as understanding that
expression? Frege's Thesis states that one answer can be given to both of these questions: that the
account of what it is in virtue of which a speaker understands an expression can also serve as the
account of why that expression has the semantic value it does. In this paper, Frege's Thesis is
defended. First, it is argued that some objections against it result from a confusion between
semantics and metasemantics. Second, the Thesis is defended against traditional objections from
Kripke and Putnam.

The third paper is "Rigidity and Content". According to much recent philosophy of language, if
two utterances differ in modal semantic value, then they must express different things. In this
paper, it is argued that this view is false. In particular, it is argued that the notion of content which
results from this consensus is incompatible with fundamental philosophical principles relating
content to use.
My tenure at MIT occurred at a very special time in the history of MIT philosophy. I like to think that during my tenure at MIT, there was MIT philosophy—that strange mixture of the philosophies of language, logic, and mind that draws the best minds to the better half of Cambridge. One of my greatest debts is thus not to any particular person, but rather to a very special intellectual environment, one which I fear I may not encounter again.

After the MIT philosophy department, my next greatest debt to a non-person is to my thesis committee. This thesis is a direct result of literally hundreds of hours of conversations—and almost as many lengthy e-mail exchanges—I have had with Robert Stalnaker, Dick Cartwright, and George Boolos. Each one of them was prepared, from my first day in the department, to drop whatever they were doing to discuss my latest idea or projects, and to share with me their own current thinking. It is no exaggeration to say that one of my central goals in life is to prove to these three people that the time and work they invested in me was worthwhile.

Now to members of these groups. My greatest intellectual debt to a teacher of mine is to my thesis advisor Robert Stalnaker. Simply put, I have arrived at my views by struggling with his, both in conversation and in print. Bob’s manner of doing philosophy also has greatly affected me. There is a tendency in philosophy—and especially in philosophy of language—to embrace simplistic responses to difficult problems, out of a fear of the detail and complexity an adequate response to these problems may involve. Bob’s intellectual honesty in facing up to the depth and complexity of the problems which we face has given me hope that progress in philosophy is possible. I also owe Bob a tremendous personal debt. First of all, he has never imposed either his views or his methodology on me—virtually unheard of for a famous philosopher. Instead, he has helped me to develop my own views, even (and perhaps especially) when they depart from his. Secondly, during my years at MIT, he has patiently endured a variety of emotional explosions from my side, either stress induced (as with this year on the job market), or caused by frustrating encounters with the False (a terrifying and omnipresent being...). For all this, and more, I thank him.

When non-MIT philosophers encounter my views, they are often shocked to find out that Richard Cartwright is on my committee. “But doesn’t he tell you that your views are ridiculous?”, they all ask. The answer to this is: well, yes. But he doesn’t just tell me this; rather he tries to convince me that they are. As a result, Dick and I have spent a tremendous amount of time in extremely animated—and for me, extremely fruitful—discussion. My many discussions with Dick have been my greatest, and most rewarding, intellectual challenges so far in my career. Dick has also given me what is perhaps one of the greatest gifts a philosopher can give to their student: a philosophical conscience. It is now second-nature for me, whenever I am working on a problem, to attempt to formulate how Dick would challenge my approach, and what the best way to defend it against these challenges would be. Though my philosophical differences with Dick are quite large, this has never affected our working relationship, nor my feeling that we are both working on the same problems. Rather than causing me to feel that the problems have no solutions, our many clashes have brought me to a heightened awareness of their complexity and importance.

George Boolos, despite his repeated disavowals of philosophical aspirations, is the greatest active philosopher of logic. It has been a great honor (albeit often quite daunting) to be his student. I have learned to live with his twice-monthly outbursts decrying the possibility of progress in philosophy,
since they are punctuated by discussions of amazing profundity. It is still the case that George’s paper, “To Be is to be the value of a variable (or some values of some variables)”, with its marriage of philosophy of logic, philosophy of language, and metaphysics, is, after "Sense and Reference", the paper I wish most to have written in all of analytic philosophy. George’s understanding of logic, its history, and its philosophy, goes so far beyond what I—or anyone I know—could ever learn about these subjects, that it has almost discouraged me from pursuing these matters. Yet his ability, in print and in discussion, to convey the philosophical richness of these disciplines, has convinced me that to be a serious philosopher, I must pursue my interests in these fields.

I have also learned much from professors not on my committee. Chief among these are Jim Higginbotham, Ned Block, Judy Thomson, Tim Williamson, Sanford Shieh, Gabe Segal, Paul Horwich, Richard Larson, and Peter Ludlow.

In my time at MIT, I have been lucky to have some of the deepest young philosophers in the world as colleagues. My central conversation partner through the years among my MIT colleagues has been Daniel Stoljar. Daniel and I, from our first days at MIT, have been discussing (and arguing about) philosophy non-stop. I do not know how much we have resolved in our countless hours of debate. I have nonetheless learned a tremendous amount from these discussions, and am honored to count him as among my closest friends. Lenny Clapp, Kathrin Koslicki, and Zoltan Szabo also deserve very special mention. Though I have profound disagreements with each of them, my views in technical philosophy of language have been largely shaped through our numerous and energetic debates. In a way, I consider us to be the beginnings of a new perspective in the philosophy of language. I also have greatly profited from discussions with some of the members of MIT’s newest generation in philosophy of language: Delia Graff, Lisa Sereno, and Cara Spencer. Among the people with whom I had fruitful, albeit all too infrequent exchanges, are: Andrew Botterell, David Hunter, Darryl Jung, Rob Kermode, Joe Lau, Josep Macia-Fabrega, Rob Stainton, and Gabriel Uzquiano. Thanks also to the linguists, in particular, my year-mates Diana Cresti and Jon Bobaljik, as well as ex-linguist Sean Kern and Andrew Carnie.

Chief among my frequent conversation partners in other universities was my close friend Mike Thau. We differ on so much in philosophy that at times we find it hard to reach enough common ground to disagree. But, of course, we managed it anyhow. Michael Glanzberg was also always good for an evening of disputation. Alva Noe, Erin Kelly and Tamar Gendler deserve mention here as well.

All of my parents have provided me tremendous support over the years. My mother deserves, in particular, incredible thanks—without her support, I could not have made it. My father and step-mother also helped greatly with emotional support and advice. Thanks also to Njeri Thande for allowing me realize I do not need to suffer in order to produce philosophy.

For better or for worse, Kathrin Koslicki has been the central figure for most of my adult years. What I am today is due, in large part, to her influence.

I have saved my greatest debt for last. Richard Heck has been my most constant partner in my life’s quest, the attempt to deal with the philosophical problems. This thesis is dedicated to him.
TABLE OF CONTENTS

Truth and Metatheory in Frege .................................................. p. 9
Frege's Thesis ................................................................. p. 45
Rigidity and Content .......................................................... p. 69
Truth and Metatheory in Frege

According to much recent work on Frege, Frege lacked a semantical metatheory. Appearances to the contrary, such as Frege's seeming attempts, in Part I of the Basic Laws of Arithmetic, to provide an interpretation of his formal language as well as metatheoretical justifications of his rules of inference, are explained away, usually on the basis that to take them as such would conflict with foundational tenets of his philosophy. The purpose of this paper is to subject several of the underlying arguments of this recent work in Frege interpretation to critical scrutiny.

One of the most sophisticated examples of this line of thought is to be found in the work of Thomas Ricketts. Accordingly, the majority of this paper is devoted to countering his arguments. Contra Ricketts, I shall argue that Frege indeed crucially invokes a semantical metathecy in the Basic Laws, his central logical work, and that that there are no convincing reasons to believe that this conflicts with Frege's general philosophical programme.

In section I, a brief summary is given of those views of Ricketts that will concern us here. Section II is negative. In it, I argue that there is little reason to embrace Ricketts' conclusions. In Section III, I shall argue that some principle motivations for not taking Frege's apparently semantical remarks at face value stem from a misconception of the nature of semantical theorizing. Section IV is a brief attempt to stress the importance Frege attached to justifying his formal theory by classically semantical means. Finally, in Section V, I turn to the relation between Frege's works and model-theoretic semantics.
Section 1.

According to Ricketts, Frege's conception of logic is founded on his theory of judgement. As we shall see in the next section, Ricketts takes Frege's theory of judgement to imply that Frege ascribed to a view of truth equivalent to the conjunction of the following two theses:

(i) each occurrence of the word 'true' is eliminable in favor of a sentence not containing the word 'true'

(ii) Tarski's equivalence schema is the only principle governing occurrences of the word 'true'.

Let us call this theory *The Redundancy Theory of Truth*. By Thesis (i) of this theory, ineliminable occurrences of the truth-predicate are ruled out. But if we wish to express the validity of inference rules, or logical laws, we must use the truth-predicate ineliminably in generalizations.\(^1\) Thus, Ricketts takes Frege's theory of judgement to preclude expressions of the validity of inference rules or logical laws, and thereby to exclude "a genuine semantical metaperspective".

Ricketts argues further that "once truth is excluded as a property, Frege has no nonsyntactical metalogical vocabulary" ([1986b], p.83). Now, Ricketts concedes that Frege's talk of denotation in introducing his formal system could be "easily read as anticipations of the post-Tarskian interest in semantics" ([1986a], p.176). But Ricketts believes that to read it in this way is a grave error. Frege's use of expressions such as "denote" in the exposition of his formal system should not be interpreted as "incipient theorizing about a relation between words and things" ([1986a], p.176).

\(^1\)"...[A]ny attempt to account for the validity of an inference is bound to use the notion of truth ineliminably in generalizations" (Ricketts [1986a], p. 176). One might think that one could express the validity of inference rules by the use of corresponding logical laws. But, according to Ricketts, this would "obviate the distinction between logical laws and rules of inference" ([1986b], p. 83).
Unfortunately, Ricketts' interpretation of the term "denote" is never supported by textual evidence, since, as Ricketts himself points out, Frege certainly seems to use the term in a classically semantical manner. Rather, Ricketts argues indirectly for his view. First, he argues that Frege's conception of logic bars a semantical metaperspective. He then concludes from this that one cannot understand Frege's talk of reference semantically, as this would conflict with his conception of logic.

Two consequences flow from Ricketts position. The first of these is for Frege's position in the philosophy of language. If Frege's conception of judgement and logic rules out any sort of semantical metaperspective, then it also rules out the formulation of classical questions of the philosophy of language, which deal with the relations between words and their interpretation. A second, and related implication of Ricketts position is that if it is true, then the semantics of formal languages as developed by Tarski and others, would be utterly incomprehensible to Frege. In order to understand semantical results, one must have recourse to a metaperspective from which one can speak of relations between a formal language and its interpretation. But on Ricketts' view, Frege had no perspective from which he could separate a formal language from its interpretation. Hence, metatheoretic questions such as soundness or completeness would make no sense for Frege.

---

2"[For Frege] there is no standpoint from which to ask whether the thoughts expressed by the statements of language really represent reality, whether they are really true or false...Here there is only the work of disambiguation, amplification, and clarification. This task requires no metaperspective" ([1985], p. 8).

3E.g., "...anything like formal semantics, as it has come to be understood in the light of Tarski's work on truth, is utterly foreign to Frege" ([1986b], p. 67).

4Burton Dreben and Jean Van Heijenoort ([1985], p.44) agree with this consequence of Ricketts' position: For Frege, and then for Russell and Whitehead... metasystematic questions as such, for example the question of completeness, could not be meaningfully raised. We can give different formulations of logic, formulations that differ with respect to what logical
Thus, Ricketts' position has quite severe consequences. If it is true, then semantical questions, both in logic as well as the philosophy of language, would make no sense for Frege. It is worth mentioning that there are other, less severe interpretations. For instance, one might believe that there are certain semantical questions, such as completeness, which would have made no sense to Frege, but reject Ricketts' more powerful thesis. Be that as it may, for Ricketts, Frege's apparently semantical remarks serve only an expository, or elucidatory purpose. These elucidations should in no way be interpreted as invoking the type of semantical metaperspective implicit in modern semantics.

Section II.

On the face of it, non-syntactical terms occur frequently in Part I of the Grundgesetze. Most noteworthy is the term 'denote' [bedeuten], which seems to be a non-syntactical term par excellence. Another classically semantical term, the truth-predicate, also appears, but emended to fit Frege's conception of truth as an object. Instead of saying that a sentence is true, Frege says that a sentence denotes the True. Now, a central thesis of this paper is that one way in which Frege uses his truth-predicate is in giving a justification, in the metalanguage, of the rules of inference of the Begriffsschrift. Since this is precisely the sort of use of a metalanguage that Ricketts denies to Frege, this section is devoted to countering Ricketts' arguments.

The justification of the inference rule of modus ponens lies in the fact that if a conditional is true, and its antecedent is true, then so is the consequent. This fact gives some justification

constants are taken as primitive or what formulas are taken as formal axioms, but we have no vantage point from which we can survey a given formalism as a whole.

\(^5\)From now on, I use the expression 'is true' to express whatever is expressed by Frege's expression, 'denotes the True'.

12
to accepting modus ponens as a rule of inference in a logical system by pointing to its truth-preserving nature. But according to Ricketts, a justification of this sort is, for Frege, out of the question:

On the contemporary conception of logic, the acceptance of MP as a correct rule of inference is vouchsafed by our metalogical judgement that if a conditional is true and its antecedent is true, then so is the consequent. This judgement is not supposed to appear as a premise in every proof employing MP; it rather is a part of our reason for accepting derivations in some given formal system as proofs. But a statement of the validity of MP unavoidably involves taking truth to be flatly a property of thoughts. This treatment of truth is precluded by Frege's conception of judgement ([1986b], p.83).

A justification of a rule of inference requires the ineliminable use of a truth-predicate. According to Ricketts, however, Frege's conception of judgement rules out truth as an ineliminable property of thoughts; indeed according to Ricketts, truth is not, for Frege, a property of thoughts at all. Thus, on Ricketts' interpretation of Frege, no such justification could be given.

Now, there is a trivial sense in which Ricketts is correct in his assertion that truth, for Frege, is not a property of thoughts. There are many passages where Frege argues for this thesis. But in these passages, he is of course arguing that the relation of a thought to truth is not the relation of subject to predicate, but rather the relation of a sense to its referent. Thus, the relation of a thought to the True is not to be understood on the model of subject and predicate, but rather as parallel to the relation between the sense expressed by the singular term 'Bob', and the person, Bob. The issue here, however, is whether Frege ever appeals, in introducing his formal system, to a semantic predicate, 'denotes the true', which expresses a
property of thoughts, in the same manner as 'denotes Bob' is a semantic predicate true of
certain singular terms.\textsuperscript{6} It is this latter question to which Ricketts responds in the negative.

Initially, Ricketts' view may seem implausible. After all, before every section in which a
rule of inference is introduced, Frege gives a justification of the rule, which looks, for all
intents and purposes, exactly like the metalogical judgement discussed above. For instance, in
the beginning of §14, where Frege introduces Modus Ponens:

(1) From the propositions \"\(\alpha \supset \Gamma\)\" and \"\(\neg \alpha\)\" we may infer \"\(\neg \Gamma\)\"; for if \(\Gamma\) were not
the true, then since \(\alpha\) is the true \(\alpha \supset \Gamma\) would be the false.\textsuperscript{7}

A similar justification is offered for the other rules of inference.\textsuperscript{8} Here, it would seem, Frege's
truth predicate \textit{is} being used in a similar manner as in the contemporary metalogical assertion,
namely as an ineliminable property of thoughts. Hence, it seems natural to read (1) as an
attempt at a justification, in the metalanguage, of Frege's inference rule. Just as in the modern
case, (1) helps justify Frege's inference rule by pointing to the fact that it expresses the

\textsuperscript{6}From now on, I shall use the term 'denote' ambiguously. First of all, I shall use it in the
ordinary sense; that is, as expressing a relation between words and things. But I shall also use
it to express that relation between the sense(s) expressed by a term and the referent of that
term. Thus, to each relation between expressions and their denotations, there corresponds a
similar relation holding between the sense of the expression in question and the denotation of
that expression. I use the term 'denote' ambiguously as between these two relations.

\textsuperscript{7}Frege's rule is a generalized version of modus ponens. Note also the strange use Frege
makes of his capital Greek letters. The natural interpretation of the quotation marks occurring
in this passage is that they are devices of quasi-quotation. But if this were the case, then the
capital Greek letters \('\Delta' and '\Gamma'\) would have to be used as names of unspecified expressions.
But then the sentence \"\(\Delta\) is the true\" would be patently \textit{false}, for the True is a value-range, not
an expression.

\textsuperscript{8}Similar soundness type proofs are also given for the axioms, although Frege's defense of
the truth of axiom V is (with good reason!) more hesitant.
inference that we use when we infer from the truth of a conditional and its antecedent, the truth of its consequent.\(^9\)

According to Ricketts, however, we must reject this interpretation of Frege's remarks. For such a justification involves construing truth as an ineliminable property of thoughts. And such a use of the truth-predicate is, as we have seen, supposedly "precluded by Frege's conception of judgement". To evaluate this claim, let us now turn to a closer consideration of Ricketts' arguments.

Ricketts's central argument is a reinterpretation of an argument of Frege's, given in his 1918 paper, "The Thought", which purports to establish the indefinability of truth. As one may recall, in this paper, several arguments are given against the correspondence theory of truth. One of the arguments is then generalized as an attempted *reductio ad absurdum* of any possible definition of truth:

[Any attempt] to define truth breaks down. For in a definition, certain characteristics would have to be specified. And in application to any particular case the question would always arise whether it were *true* that the characteristics were present. So we would always be going round in a circle ([1984a], p. 353).

This argument is unsatisfactory, for reasons that are well-known.\(^10\) Ricketts, however, claims that in order to appreciate the argument, we must view it within the context of Ricketts' Frege's conception of judgement. Seen in this context, according to Ricketts, "The proper conclusion to the argument is not that truth is a primitive property, but that truth is not a property at all" ([1986b], p.79).

According to Ricketts, then, viewing Frege's argument within the context of Ricketts' Frege's conception of judgement transforms it into a valid argument for the conclusion that

\(^9\)In the next section we shall have more to say about *why* we should construe Frege in this manner.

\(^10\)See, for example, Michael Dummett's discussion in Chapter 13 of his [1981].
truth is not a property of thoughts. Laying aside objections to Ricketts' interpretation of Frege's conception of judgement, let us pause to evaluate this claim.

Three facts about Ricketts' Frege's conception of judgement, coupled with Frege's argument, are taken by Ricketts to imply that for Frege, truth is not a property of thoughts. First of all, according to Ricketts' Frege, in order for a speaker to be able to make a judgement, it must be in principle possible for the speaker to verify whether the judgement is correct. The second fact about Ricketts' Frege's conception of judgement is the following thesis about the nature of the act of judging:

Judgement is not itself a special recognitional capacity, a species of some genus. Judgement is itself the genus - to recognize anything to have a property is ipso facto to make a judgement ([1986b], p.77).

Thus, we have two facts about Ricketts' Frege's conception of judgement: (i) it must be in principle possible to verify the correctness of the judgements one makes, and (ii) to recognize anything to have any property is to make a judgement. The third fact about Ricketts' Frege's conception of judgement can be expressed as the following conditional; (iii) if truth is a property of thoughts, then in order to correctly make a judgement, I must antecedently acknowledge that the thought expressed by some sentence which represents that judgement is true. Now let us consider how these three facts about Ricketts' Frege's conception of judgement bear upon Frege's argument.

Let us assume (i), (ii), and (iii), as well as the thesis that truth is a property of thoughts. Using Frege's argument, we shall attempt to derive a contradiction. Now, if truth is a property of thoughts, then we can give some content to the notion of correctness, or assertiblity. A judgement is correct, or assertible, only if the thought expressed by a sentence which represents the judgement is true. So let us now say that we wish to assert that snow has the property of whiteness. In order to do so, by (iii), we must first ask ourselves whether the
thought expressed by the sentence "Snow is white" has the property of truth. But by (ii), the act of determining whether the thought expressed by the sentence "Snow is white" has the property of truth is itself an act of judgement. So we must then ask whether this act of judgement is correct. Thus, by (iii), we must ask whether the thought that the thought expressed by the sentence "Snow is white" is true is itself true. But this, in turn, is itself a judgement. Hence, the regress begins. Thus, if truth is a property, in order to determine whether I can assert that snow is, in fact, white, I must determine the correctness of an infinite number of other judgements. But since I am a finite being, this is impossible, contra (i).

Ricketts' argument is perfectly valid. Its central oddity is rather that it implies the opposite of what Frege actually seems to conclude. Rather than ending the argument, as Ricketts would have it, by asserting that truth is not a property of thoughts, Frege concludes that "the content of the word 'true' is *sui generis* and indefinable". Ricketts' rather weak explanation of Frege's closing remarks is that Frege's use of the term 'indefinable' here should not be taken to mean what Frege usually means by the word. Rather, we are to take Frege's claim that truth is indefinable as expressing the quite different proposition that "truth...cannot be understood to be a primitive unanalyzable property of thoughts" ([1986b], p.79).

Ricketts' reinterpretation of the moral Frege draws from his own argument becomes even stranger when one considers the other article of Frege's in which the same argument is given. In Frege's article "Logic", where we find, almost word for word, the same argument, we also find Frege drawing a conclusion that seems quite antithetical to Ricketts' interpretation:

---

11 I assume that the thought that the thought expressed by "Snow is white" is true is expressed by some appropriate sentence.

12 Frege's talk of indefinability in this context should be understood differently from his other uses of this notion" ([1986b], p.79).
Now it would be futile to employ a definition in order to make it clearer what is to be understood by 'true'. [For consider] any definition of the form 'A is true iff it has such-and-such properties or stands in such-and-such a relation to such-and-such a thing'. In each case in hand it would always come back to the question whether it is true that A has such-and-such properties, or stands in such-and-such a relation to such-and-such a thing. *Truth is obviously something so primitive and simple that it is not possible to reduce it to anything still simpler* ([1979a], pp. 128-129).

Just as in the previous case, it seems quite strange to take Frege as really concluding that truth is not a simple, unanalyzable property of thoughts. Rather it seems that Frege took the argument to show that truth is primitive, simple, and unanalyzable. Evidence that Frege, in addition, took truth to be a property of thoughts appears two pages after he gives the above argument:

> The sense of an assertoric sentence I call a thought. Examples of thoughts are laws of nature, mathematical laws, historical facts: all these find expression in assertoric sentences. I can now be more precise and say: The predicate 'true' applies to thoughts.\(^\text{13}\)

Although such passages are not conclusive, they do give the impression that Frege took the truth-predicate to express a property of thoughts. It thus seems dubious to take as the moral that Frege draws from his argument the thesis that the predicate 'true' does not express a primitive, simple property of thoughts.\(^{14}\)

But how does Frege's argument establish what Frege takes it to establish, namely that truth is indefinable? Let us call an "informative characterization of truth" any characterization which supplies a predicate, satisfaction of which constitutes a necessary and sufficient condition for a thought to denote the True, yet which expresses a different Sense than the ordinary truth-

---

\(^\text{13}\) [1979a], p. 131, and similarly, a few pages later:
Almost everything that we have said about the predicate 'true' holds for the predicate 'false' as well. In the strict sense it applies only to thoughts. Where it looks to be predicated of sentences or ideas, still at bottom it is being predicated of thoughts.

\(^\text{14}\) The seeming contradiction between these quotes and Frege's espousal of the doctrine that the word "true" denotes an object is, of course, illusory. In these passages, Frege is speaking of the predicate 'true', i.e. "denotes the true".
predicate, 'is true'. I suggest that Frege's argument shows truth to be indefinable by demonstrating the impossibility of any such characterization.

Obviously, if Frege's argument is to succeed, powerful premises are needed. In particular, some premise is needed that allows Frege to conclude that "it will always come back to the question of whether it is true that A has such-and-such properties". In other words, Frege is appealing to a principle which implies that in asserting that A is $\Phi$, we are actually asserting that it is true that A is $\Phi$. Such a principle is to be found in Frege's characterization of assertion as the manifestation of the act of acknowledging the truth of a thought.\(^{15}\) Assuming this principle, let us now turn to Frege's argument.

Assume that we have specified some condition, expressed by the predicate $\Phi$, which gives us an informative characterization of truth.\(^{16}\) Thus, our assumption entails that:

\[(2) \ p \text{ is true } \equiv p \text{ has the property expressed by } \Phi.\]

Now let us say that we wish, via our necessary and sufficient condition, to determine whether $p$ is true. By (2), we can accomplish this by determining whether $p$ has the property expressed by $\Phi$. But, by our premise, determining that $p$ has this property amounts to acknowledging that it is true that $p$ has this property. But we wish to determine that $p$ is true via the sense of $\Phi$ alone, and without the use of the predicate 'is true'. So in order to determine whether it is true that $p$ has the property expressed by $\Phi$ via our necessary and sufficient condition, we must determine whether the thought that $p$ has the property expressed by $\Phi$ itself has the property expressed by $\Phi$. And so the regress begins.

\(^{15}\)[1984a], p. 356. Richard Cartwright has pointed out to me the awkwardness of Frege's use of the factive verb "acknowledge" [anerkennen] in this context. If judgement is the acknowledgement of the truth of a thought, then we can never judge incorrectly. Similar remarks hold for Frege's description of judgement as the "inward recognition of the truth of a thought". For familiarity's sake, I will, however, retain Frege's terminology.

\(^{16}\)"$\Phi$" is here replaceable by names of predicates.
Truth, for Frege, is the most general logical notion. Some laws of logic are laws of identity, and others laws of generality, but all logical laws are laws of truth. It is a familiar point that a logical relation, such as the identity relation, cannot be explained in terms which themselves do not invoke the notion of identity. Frege's regress argument is a general argument for this conclusion for the case of the property of truth. But just as we do not, in the case of identity, conclude that identity is not a "real" relation, so we should not, in the case of truth, conclude that it is not a "real" property.

The argument, thus rendered, is an expression of Frege's belief that "the content of the word 'true' is *sui generis* and indefinable". Furthermore, the thesis that the truth-predicate expresses a primitive, simple, property of thoughts is fully consistent with this interpretation. Ricketts, on the other hand, must, in effect, either ignore or re-interpret most of what Frege himself has to say about the argument. Thus, not only do there exist other readings of Frege's argument which are consistent with the thesis that truth is a property of thoughts, but Ricketts' gloss of Frege's argument seems to contradict the very conclusions which Frege himself took the argument to imply.

If Frege had thought that the truth-predicate did not express a property, then Ricketts would be justified in ascribing the Redundancy Theory of Truth to Frege. As we have seen, however, not only does Ricketts fail to demonstrate that Frege believed the truth-predicate did not express a property, but the evidence bears strongly in favor of the opposite conclusion.

In assertions of the validity of inference rules, the truth-predicate occurs *ineliminably*. But such a use of the truth-predicate is inconsistent with the Redundancy Theory of Truth. So far we have shown Ricketts' central argument for ascribing this theory to Frege to be

---

17Henceforth, except where indicated, I will not use the term "validity" in the model-theoretic sense, but rather in the straightforward intuitive sense of "truth-preserving no matter what we take as the values of the schematic letters".

20
unpersuasive. But before we conclude that Frege's apparent statements of the validity of inference rules are what they appear to be, it would be good to have more evidence against the view that Frege believed the truth-predicate must be capable of elimination in all of its varied occurrences.

There are two contexts of interest to us in which the truth-predicate occurs ineliminably. The first context is in generalizations, of which assertions of the validity of inference rules form a sub-species. A simple example of such an occurrence, not involving rules of inference, is the following:

(3) Everything Socrates ever asserted was true.

Presumably, no one alive today can list every statement ever uttered by Socrates. Because of this unfortunate epistemic limitation, there is no way to avoid the use of the truth-predicate. But an even more serious situation arises in the case of a statement of the validity of an inference rule. For at least we can be sure that Socrates uttered only a finite number of statements. The variables occurring in the statement of the validity of an inference rule, however, range over an infinite number of sentences. Thus, our own mortality stands in the way of eliminating these occurrences of the truth-predicate.

The second context in which the truth-predicate occurs ineliminably is with sentences that lack a truth-value. For assume that the sentence "Pegasus swims" does not have a truth-value. Then

(4) "Pegasus swims" is true.

is false, whereas

(5) Pegasus swims.

has no truth-value. Hence, (4) and (5) are not equivalent.
Of the two contexts in which the truth-predicate occurs ineliminably, the second is by far the more philosophically contentious. Most contemporary philosophers sympathetic to the intuitions underlying the Redundancy Theory nonetheless accept that one function (or, as some would have it, the primary function) of the truth-predicate is to allow us to express generalizations. It seems that one would have a reason to deny that the truth-predicate serves such a function only if one believed that the truth-predicate did not express a property of thoughts. But it nonetheless could be maintained that Frege, for some other reason, believed the truth-predicate must always be capable of elimination.

One reason one might attribute the Redundancy Theory to Frege is because of the occurrences of passages such as the following in Frege's writings:

One can, indeed, say: "The thought that 5 is a prime number is true." But closer examination shows that nothing more has been said than in the simple sentence "5 is a prime number" ([1970a], p. 64).

Thus, according to Frege, when "S" is replaced by a simple, declarative sentence that does not lack a truth-value, the following sort of occurrences of the word 'true' are eliminable:

(6) The thought that S is true.

But of course such quotations do not imply that Frege held that the truth-predicate must always be eliminable. For nothing is here implied about the use of the truth-predicate in generalizations, or about the case in which S lacks a truth-value. That is, nothing is implied about the cases that are of interest to us. Such passages are therefore irrelevant to the issue at hand.19

18 E.g. philosophers who have advanced the "Disquotational Conception of Truth". For the classic statement of this view, see pp. 10-13 of Quine [1970].

19 Of course Quine is an example of a philosopher who agrees with Frege here, yet is perfectly content with ineliminable uses of the truth-predicate: "So long as we are speaking only of the truth of singly given sentences, the perfect theory of truth is what Wilfrid Sellars has called the disappearance theory of truth" ([1970], p. 11), and "By calling the sentence
We have not come across any convincing arguments for ascribing the Redundancy Theory to Frege. But even more damning is the fact that Frege's analysis of sentences that lack a truth-value commits him to the view that the truth-predicate occurs ineliminably in the second, and more disputed, context. For according to Frege, sentences containing non-referring singular terms lack a truth-value. And it is also the case for Frege that if all parts of a sentence have reference, then the sentence has a truth-value:

If a sentence has a meaning at all, this is either the True or the False. If a sentence can be split up into parts, each of which is meaningful [bedeutungsvoll], then the sentence also has a meaning.20

Thus, Frege held the following thesis:

Truth-Value Thesis: A sentence has a truth-value if and only if all of its parts have a reference.

Hence, (5) has no truth-value. But (4) must have a truth-value, since the singular term in subject position is a quote-name of a sentence, and hence is a referring expression.21 Therefore (4) and (5) are not equivalent. Furthermore, according to Frege, sentences with non-referring terms nonetheless express thoughts. Thus, we must also accept that:

(7) The thought that Pegasus swims is true

["Snow is white"] true, we call snow white" (Ibid., p. 12).

20Frege, [1979a], p. 194. Here "meaning" is used for "Bedeutung". I use the term "reference" instead.

21(4) must have a truth-value unless, of course, one wishes to deny that the predicate "is true" has a reference. But if one took this line, then sentences such as "'Snow is white' is true" would also lack a truth-value, since they also contain "is true"--which is, on this view, a non-referring expression.
has a truth-value, since the singular term "the thought that Pegasus swims" has a reference, namely the thought expressed by the sentence, "Pegasus swims". Hence, (7) is also not equivalent to (5).\footnote{Most of this polemic is of course familiar from Michael Dummett's [1978].}

One might, however, take Frege's characterizations of judgement and assertion as conflicting with Frege's Truth-Value Thesis. For, as we have seen, Frege's account of assertion entails that asserting $S$ is tantamount to asserting that $S$ is true. This might be taken to imply that (4), (5), and (7) are equivalent on the grounds that if (5) were asserted, then the thought thereby expressed would be the same as the thought expressed by (4) and (7). Yet in order to apply Frege's principles about assertion to conclude that (4), (5), and (7) are equivalent, we must assume that (5) is capable of being asserted. But for Frege, sentences containing non-referring singular terms, such as (5), cannot be asserted.

In order for a sentence to count as being asserted, it must be uttered with assertoric force. Utterances made by actors lack the appropriate force, and hence do not count as assertions.\footnote{As Frege puts the point: "As stage thunder is only sham thunder and stage fight only a sham fight, so stage assertion is only sham assertion." ([1984a], p. 356).} Yet for Frege, the fact that a sentence is uttered with assertoric force is not a sufficient condition for the utterance to count as an assertion. In addition, the sentence being uttered must have a truth-value. As Frege writes in his article "Logic":

Instead of speaking of 'fiction', we could speak of 'mock thoughts'. Thus, if the sense of an assertoric sentence is not true, it is either false or fictitious, and it will generally be the latter if it contains a mock proper name...Assertions in fiction are not to be taken seriously: they are only mock assertions ([1979a], p. 130).
Utterances of sentences that contain non-referring expressions, for Frege, are treated on a par with utterances that lack assertoric force. Both types of utterances, according to Frege, are instances of fiction \([Dichtung]\), and do not count as acts of assertion.\(^{24}\)

Sentences that lack a truth-value cannot, according to Frege, be asserted.\(^{25}\) Frege's remarks about assertion therefore do not imply the equivalence of (4), (5), and (7), and hence do not conflict with his Truth-Value Thesis. We can safely conclude that Frege was committed to ineliminable occurrences of the truth-predicate.

Now, it is standard interpretive practice to avoid attributing a contradictory position to a philosopher. If Ricketts had uncovered powerful, clear reasons why we must ascribe to Frege the Redundancy Theory of Truth, then we would be forced to conclude that Frege contradicted himself in his analysis of meaningless names. However, in the absence of such evidence, the interpretation which makes Frege's system a consistent, coherent whole must be chosen over the interpretation which makes him an internally inconsistent philosopher.

Ricketts argues that to take (1) as a justification of the validity of Frege's rule of modus ponens would commit Frege to denying the Redundancy Theory. Yet Ricketts fails to give sufficiently good reasons for ascribing the Redundancy Theory to Frege. We are therefore left without a convincing argument against taking (1) at face value, namely as a justification of the

\(^{24}\) As Frege writes: "But if my intention is not realized, if I only think I see without really seeing, if on that account the designation 'That lime tree' is empty, then I have gone astray into the sphere of fiction without knowing it or wanting to" ([1984a], p. 362), and similarly: "A sentence containing a meaningless proper name is neither true nor false; if it expresses a thought at all, then that thought belongs to fiction" ([1979a], p. 194).

\(^{25}\) Although Frege's view that sentences with non-referring singular terms cannot be asserted seems intuitively implausible, there are contemporary philosophers who have advocated similar views. Gareth Evans, for instance, has argued at length for the thesis that to utter a sentence containing a non-referring Russellian singular term (where this class, for Evans, includes most proper names) is not, in fact, to make an assertion. See his [1982] for details.
validity of modus ponens. Yet aside from noting that (1) certainly appears to provide such a justification, we have yet to give a positive reason for taking (1) in this manner. Although we shall focus more directly on this issue in the next section, a few preliminary remarks can be made here.

Logic, according to Frege, is the science of truth. As Frege writes, "...logic has much the same relation to truth as physics has to weight or heat. To discover truths is the task of all sciences; it falls to logic to discern the laws of truth." Thus, the purpose of a rule of inference is to encode a law of truth. Now, the goal of a justification of a logical system is to convince us that the rules of inference and axioms of the system serve the purposes for which they were intended. Thus, a Fregean justification of logic would have to provide a reason for us to take the rules of inference of the *Begriffsschrift* as reflecting the laws of truth. And calling our attention to the truth-preserving nature of the inference does just this.

"The laws of logic", asserts Frege, "ought to be guiding principles for thought in the attainment of truth" ([1964], p. 12). By showing the reader of the *Basic Laws* that the rules of inference of his formal system preserve truth, Frege is giving one reason to take the rules of inference and axioms of the *Begriffsschrift* as laws of logic. But if (1) is indeed an attempt at a metalinguistic justification of an inference rule, then Frege clearly resorted to the type of "semantic metaperspective" whose existence Ricketts denies.

---

26[1984a], p. 351. Also: "I assign to logic the task of discovering the laws of truth...the meaning of the word 'true' is spelled out in the laws of logic" (p.352). Similarly, "The word 'true' specifies the goal...The word 'true' characterizes logic" ([1979a], p. 126); "Logic is the science of the most general laws of truth" (p. 128).
Section III.

In the preceding section, we saw that there is no good reason for ascribing to Frege the Redundancy Theory of Truth. First of all, when Frege gives the argument Ricketts cites as the basis for this attribution, he actually draws the conclusion that truth is a primitive, simple property of thoughts, rather than the conclusion that truth is not a property of thoughts at all. Hence, it is dubious that this argument can be used in support of Ricketts' position. Secondly, if Ricketts were correct in his attribution, then, given Frege's commitment to ineliminable uses of the truth-predicate, Ricketts' interpretation would show a deep inconsistency in Frege's thought. Given standard interpretive practice, such attributions are to be avoided at all costs. In this section, I turn to another possible argument for the thesis that Frege had no metatheory, one stemming from considerations about the project of justifying logical laws.

As we have seen, Ricketts takes Frege's apparently semantical remarks to be what he calls "elucidations", rather than semantical assertions. On this interpretation, Frege does not present a semantic theory, but rather engages in a separate project, called elucidation. Joan Weiner presents this alternative interpretation as follows:

There are the early sections of the Begriffsschrift, the early sections of each volume of Basic Laws...all of which appear to consist of explanations or justifications of the work done in the language of Frege's systematic science. The justifications, in particular, look to be metatheoretic proofs...Yet my aim...will be to argue that it follows from Frege's general epistemological views that his discursive work has the status of elucidation rather than of objective statement of facts ([1990], p. 229).

As one sees from this passage, "elucidations" are to be contrasted with what Weiner calls "objective statements of fact". According to this alternative analysis, the purpose of elucidations is not to "establish truths", or to give "objective arguments", but rather to yield "some sort of prose gloss that will make the Bedeutung of the primitive term clear" (Weiner [1990], p. 140).
On this view, Frege's apparently semantical remarks are not intended to be *assertions*. This account of Frege's apparently semantical claims parallels the account of ethical sentences given by the metaethical thesis of emotivism. According to the emotivist, ethical utterances, despite appearances to the contrary, are not assertions, and do not have "factual meaning". On such an account of ethical sentences, "...in saying that a certain type of action is right or wrong, [one is] not making any factual statement" (Ayer [1942], p. 107). According to the elucidation account, Frege treats semantical sentences in a similar manner. In making what seem to be semantical assertions, Frege is actually not attempting to state semantical facts at all. Rather, he is elucidating the meaning of the *Begriffsschrift* signs.

I do not claim to fully grasp the argument for "semantical emotivism". Emotivists typically argue that ethical sentences cannot be taken as stating facts, since there are no ethical facts. But a parallel argument in the case of semantics would not aid the elucidation position. For even those dubious of the factivity of ordinary language talk of meaning and reference, such as Quine, are perfectly content to engage in explicitly semantical metatheoretic discourse. Indeed, I fail to see how serious, assertive uses of a semantical metatheory would commit one to any sort of ontologically or philosophically contentious position at all.

However, proponents of the elucidation position clearly believe otherwise. According to them, there is some non-innocuous philosophical position to which one is committed by serious, assertive uses of semantic theory. Their argument that Frege's remarks cannot be read as semantical flows from their belief that Frege did not accept the non-innocuous philosophical position in question, and hence could not have engaged in such theorizing. Though it is dangerous to speculate about motivations for a view, I would like to suggest that such a position results from a misconception about the nature of semantic theorizing.
What is this non-innocuous philosophical position to which one is allegedly committed when engaging in semantics? Those who advocate the elucidation interpretation are fond of pointing out that Frege did not believe that one could give a non-circular justification of logical laws, or infer such laws from truths of some other discipline. As Ricketts writes:

There is, as far as Frege is concerned, nothing to be said about the justification for our recognition of these basic laws of logic to be truths.... Moreover, the maximal generality of these laws precludes their inference on the basis of the truths of any other discipline ([1986b], p. 81).

In accordance with this, I hazard that those who advance the elucidation interpretation of Frege believe that the purpose of semantic theorizing is to provide some sort of non-circular justification of logical laws -- some sort of "inference" of logical laws from other, perhaps semantic, principles. Since Frege did not believe that such a justification was possible, then Frege could not have engaged in semantical theorizing.

However, if this is the motivation for the elucidation interpretation of Frege, then it does not even get off of the ground. For no one believes that it is possible to give completely non-circular justifications of logical laws, and least of all those who engage in classical semantics. A classical semantic theory does not help one to "infer" basic logical laws from truths of a different sort: one typically assumes the laws to be proven in the metalanguage.\(^{27}\) For instance, a classical semantics which employs "absolutely straightforward stipulations" in the sense of Dummett [1991a] does just this.\(^{28}\) This is why classical soundness proofs are so trivial. It is perfectly possible to engage in semantical reasoning without being at all committed to the view that non-circular justifications of logical laws are possible. Furthermore, such a project was never a motivation in the development of classical semantic theory.

\(^{27}\)See Prawitz [1974] for discussion here.

\(^{28}\)For instance, when Tarski proves Bivalence in Section III, theorem 2 of his [1983], he is assuming the excluded middle in the metalanguage.
Moreover, even proponents of non-classical semantic theories, such as Dummett, do not believe a completely non-circular justification of fundamental logical laws to be possible. Dummett, unlike many others, still holds out for the possibility of some sort of justification for fundamental logical laws. Dummett's position here is controversial, and certainly not necessary in order to engage in semantic theorizing. But even he thinks that the best possible justifications of fundamental logical laws will at the very least be what he calls "pragmatically circular". The debate here is not whether completely non-circular justifications of logical laws could be given, justifications which would completely convince one who did not antecedently accept them. It is rather about whether the circularity "deprives of all value any justification that displays it".29

Given the fact that one can engage in semantical theorizing without believing in the possibility of non-circular justifications for logical laws, or the possibility of "inferring" such laws from other, more basic truths, this argument for the elucidation interpretation of Frege collapses. Let us thus turn to another, related argument for the elucidation interpretation. This argument proceeds as follows. According to Frege, logic is applicable to any special science. In particular, inferences in the special sciences are to be understood as instantiations of logical laws to the vocabulary of that science. If there were a special science of semantics, then this science would have to itself appeal to the principles of logic. Rather than standing apart from, and justifying the system, it would have to appeal to the very features of the system it was attempting to justify.

However, like the previous argument for the elucidation interpretation, this argument too seems to suffer from a misconception about the nature of semantics. Contemporary semantics for set theory is carried out with the use of set theory. This fact provides no obstacle to

---

semanntical theorizing about set theory. Why, then, should the fact that a parallel situation may obtain in the case of semantical reasoning about Frege's system of logic prevent Frege from engaging in such an endeavor?

The elucidation interpretation is worrisome for another reason. The function of elucidations is to put the reader in a position to understand the *Begriffsschrift* notation (hence the term "elucidation"). If the only apparently metatheoretic remarks occurring in Part I were the soundness proofs for Frege's inference rules, then this construal might not be easily refutable. For these passages come close to remarks one might make in trying to explain an uninterpreted formalism to a layman. However, there are other apparently metatheoretic remarks in Part I that clearly do not serve anything close to an elucidatory purpose.

To give a famous example, in §31, Frege attempts "to show that the proper names, and names of first-level functions...always have a denotation." In order to show this, Frege gives a rather contorted argument, which involves an induction on the complexity of formulae of the *Begriffsschrift*. It is difficult to see how this argument is of any help whatsoever in aiding readers to come to a better understanding of the *Begriffsschrift*. To call such passages "elucidations" not only vastly misdescribes their clarity, but fails to respect their status as attempted arguments. Unless the proponents of the elucidation interpretation can arrive at a more satisfactory account of the purpose of such passages, we must ascribe to them the metatheoretic, justificatory purpose which they appear to embody.

---

30 Frege [1964], p. 87.

31 As Dummett [1991b], p. 218, points out, such an induction has no hope of success in the case of an impredicative second-order theory like that of the *Basic Laws*.

32 In recent, yet to be published work, Ricketts has taken up this challenge (see his [forthcoming]).
According to the elucidation interpretation of Frege, Frege could not have engaged in metatheory, because he did not believe that logical laws could be inferred from more basic truths. However, in order for this reasoning to hold, it must be the case that the purpose of metatheory is to infer logical laws from more basic truths. Since this is a false conception of the purpose of semantical reasoning, this argument for the elucidation interpretation fails.

If the purpose of a semantic metatheory is not to infer logical laws from more basic truths, then in what sense can a semantical metatheory serve to justify these laws? This is one of the most difficult and important questions of contemporary philosophy of logic. However, what I have been suggesting is that it is a question which would be just as relevant for Frege as it is for Tarski. In the next section, I shall turn to a brief discussion of the sort of answer Frege would give.

Section IV.

Now, there are occurrences of Frege's truth-predicate in the exposition of the Begriffsschrift apart from the sections in which inference rules are introduced. Typically, these other occurrences appear alongside uses of Frege's other apparently semantical term, "denote", in passages which seem to serve the purpose of giving an interpretation to the Begriffsschrift. In this section, I wish to argue briefly that one central reason Frege gave an interpretation to his formal language was to justify the axioms and rules of inference of his formal theory. I hope thereby to emphasize the importance Frege attached to providing a metatheoretical justification of his formal theory.

In Part III of the Basic Laws, Frege objects to "Formalist arithmetic", according to which no interpretation is given to the signs of a formal system. According to Frege, such an approach to arithmetic fails to account for its applications. Consider, for instance, a natural
scientist, who wishes to use some formal theory of arithmetic in her work. How can she be
sure that the theory of arithmetic in question will be useful in her application of it to natural
phenomena? According to Frege, the scientist must believe that the axioms of the formal
theory actually express true thoughts, and that the rules of inference are actually truth-
preserving. For only such a theory of arithmetic will be able to be usefully applied to natural
phenomenon.

But what justifies the scientist's belief? The formalist has not provided any reasons to take
the axioms of his formal theory as actually expressing true thoughts, or the rules as actually
truth-preserving. The sole concern of the formalist is derivability within a system with some
set of axioms and some rules of inference. The justification of the axioms and rules of
inference is, for the formalist, not at issue. And clearly, it is not the task of the natural
scientist to provide such a justification. Thus, the task of justifying the rules of inference and
axioms of the formal theory "falls into a void between the sciences" ([1970b], §92).

The justification of the rules of inference and axioms of a formal system of arithmetic,
according to Frege, must be carried out by the arithmetician himself. Only by providing such
a justification can the arithmetician take himself as accounting for the application of his formal
theory of arithmetic to the special sciences. And it is here that sense and reference gain their
importance:

In order to bridge [the gulf between arithmetical formulas and their applications], it is
necessary that formulas express a sense and that the rules be grounded in the reference
of the signs ([1970b], §92).

The role of sense and reference is thus to help justify the axioms and rules of inference of the
formal system, and thereby to account for the applications of the formal theory of arithmetic.
To fail to do so is to "exclude from arithmetic what it needs in order to be a science"
([1970b], §91).
Frege's remarks about arithmetic hold, of course, *mutatis mutandis* for logic. For Frege's discussion of formal theories of arithmetic concerns the necessity of justifying the *rules of inference* of a formal theory of arithmetic. But the rules of inference of any formal theory of arithmetic, such as the formal theory of arithmetic given in Frege's *Basic Laws*, are the same rules of inference that occur in any standard system of logic. Thus Frege's repeated insistence upon the need for a justification of a formal theory of arithmetic carries over directly to a demand for a justification of the basic inferences of any system of logic.

By assigning sense and reference to the expressions of a formal system, one helps determine whether the axioms of the system are true, and the rules truth-preserving. Furthermore, it is clear that Frege believed that assigning sense and reference was the only way to provide such a justification:

[The cognitive purpose of arithmetic] requires the character of the rules to be such that if in accordance with them a sentence is derived from true sentences, the new sentence will also be true. Whether the rules satisfy this condition can, of course, be determined only after the signs have been given a reference ([1970b], §104).

Frege's point is clear enough. Remarks such as (1) only demonstrate that the rule of inference is truth-preserving if the condition-stroke sign has been given the appropriate interpretation. Thus, we need to assign an interpretation to the condition-stroke in order to demonstrate the truth-preserving character of the relevant rule of inference.

We can thus take ourselves to have established the role of sense and reference, as well as the importance they play in Frege's conception of logic. Without assigning sense and reference to the expressions of the formal theory, there would be no way to tell whether the axioms are true, and the rules truth-preserving, and hence for remarks such as those found in (1) to accomplish their intended task. But in order to account for the applicability of logic, we must provide reasons to take the axioms and the rules of inference of our system of logic as true, or truth-preserving. Finally, since applicability is what lifts logic and arithmetic to the status of a
science, not to account for the applicability of logic and arithmetic is tantamount to not justifying its status as a science.

Part I of the *Basic Laws* corroborates the vital role we have granted a semantic theory in Frege's system. For it seems that Frege, in introducing his formal theory, is simultaneously giving its interpretation. Each introduction of a primitive symbol of the *Begriffsschrift* is always accompanied by an explanation of the conditions under which a sentence containing it denotes the true. One example is Frege's introduction of the condition-stroke in §12:

In order to enable us to designate the subordination of a concept under a concept, and other important relations, I introduce the function of two arguments:

\[ \xi \ \underline{\xi} \ \zeta \]

by stipulating that its value shall be the False if the True be taken as \( \zeta \)-argument and any object other than the True be taken as \( \xi \)-argument, and that in all other cases the value of the function shall be the true.

Another example is Frege's introduction of the sign for equality in §7 of the *Basic Laws*:

"\( \Gamma = \Delta \)" shall denote the True if \( \Gamma \) is the same as \( \Delta \); in all other cases it shall denote the False.

As with the occurrence of the truth-predicate in the description of the rules of inference, it is difficult to take these occurrences of the truth-predicate as anything but semantical.

Frege's use of the term 'denote' also seems to be in the service of giving a semantic theory for the formal language of the *Basic Laws*. For instance, in §11, where Frege introduces the \( \backslash \) operator:

"\( \backslash \chi \Phi(\epsilon) \)" denotes the object falling under the concept \( \Phi(\xi) \) if \( \Phi(\xi) \) is a concept under which falls one and only one object; in all other cases "\( \backslash \chi \Phi(\epsilon) \)" denotes the same as "\( \backslash \epsilon \Phi(\epsilon) \)".

He is evidently telling the reader what object corresponds to the expression "\( \backslash \chi \Phi(\epsilon) \)".

Furthermore, when Frege himself introduces the term in §2 of the *Basic Laws* - the very
beginning of the work- it is unambiguously characterized as a relation between words and things:

...the name "2^n denotes the number four. Accordingly I call the number four the denotation of "4" and "2^n"...

The view that Frege is laying down an interpretation of his formal system by giving a method for determining the truth-conditions of each of its sentences is further reinforced by Frege's own remarks about these passages:

...by our stipulations it is determined under what conditions [each name of a truth-value] denotes the true. The sense of this name -the thought- is the thought that these conditions are fulfilled. 33

Let us now return to the question we raised at the end of the last section. In what sense does giving an interpretation of one's formal language justify the rules of inference and axioms of a theory couched in that language? Now, Frege's conception of formal theory is, I believe, remarkably modern. As we saw above, Frege speaks of the project of giving signs reference. This suggests that Frege had the contemporary conception of a logic as a set of syntactic operations on strings of symbols. The project of giving signs reference justifies the syntactic manipulations performed on these symbols by ensuring that they express basic inference rules such as Modus Ponens, which Frege is already assuming we accept.

There is, of course, another sense of justification in which the use of a semantical metatheory can justify a particular system of logic. For instance, one can prove, with the use of semantical techniques, that a particular system possesses some desirable semantic property, such as categoricity, completeness, or consistency. In the next section, I shall turn to the

33[1964], p. 90. Note the remarkable similarity between Frege's comments about his semantic theory and some of Tarski's comments about his definition of Truth: "...through the theorems obtained, the meaning of the corresponding expressions of the type 'x∈Tr' become intelligible and unambiguous" ([1983], p. 197).
question of how Frege would have understood these methods of justifying a system by semantical means.

Section V.

In Part I of the *Grundgesetze*, Frege's truth-predicate is not relativized to a domain. Frege did not read his quantified *Begriffsschrift* sentences as true or false relative to varying domains of objects. Instead, Frege took the quantifiers in such sentences to range over all entities of the appropriate logical type. In this section, I will try to draw out the consequences of this feature of Frege's semantics.

The sentences in Frege's *Begriffsschrift* are not to be subject to varying interpretations. Their meaning is fixed from the outset, via the semantical stipulations laid down in Part I of the *Grundgesetze*. Tarski's semantic characterization of logical consequence, and its limiting case, the semantic characterization of logical truth, did not occur to Frege. This situation raises three questions. The first question is whether this fact implies that no interesting metatheoretical results could be proven within Frege's semantical framework. The second question is whether Frege would have understood model theoretic techniques. The final question is the hardest. It is whether Frege would have accepted the model-theoretic definition of consequence as a good definition of *logical* consequence, or whether there is some feature of his philosophy which would cause him to reject it.

The first question is easily answerable. Even within an "absolute" semantics, interesting semantic questions can be raised. For instance, if one gives the intended interpretation for the theory, and proves that the axioms of the theory are, in that interpretation, true, and the rules truth-preserving, then one has given a semantic proof of the consistency of the system. Indeed, it is just this that Frege is attempting in §31 of the *Grundgesetze*. This is a substantial, non-
trivial attempt at metatheory. Furthermore, one does not need varying interpretations in order to prove it. If the intended interpretation is a model for the theory, then it follows that the theory has a model, and is hence consistent.

One might object that Frege's universalist conception of logic would bar him from accepting such a result as a consistency proof. For the proof may have to be carried out either within the theory of the Grundgesetze, or by assuming the reasoning which the system codifies. However, such an objection would be confused. It was not until Gödel's second incompleteness theorem that it was known that a consistency proof for a system cannot be carried out within that system. And contemporary semantic consistency proofs for fragments of logic or set-theory also assume, in the metalanguage, the reasoning which they are attempting to justify. Frege's conception of logic thus places him in a situation no different from that of the contemporary logician.

In conclusion, though certainly not all metatheoretical results can be proven within the framework of an absolute semantics, there are some which can. What the limitations of such a semantics are seems to me to be an interesting question, one which can only be answered by investigating logico-semantical practice. What about the second question? Would Frege have understood model-theoretic techniques? The best way to see why the answer to this must be in the affirmative is to consider Frege's own logical and mathematical writings.

It is clear from even a cursory reading of the Grundgesetze that some of the techniques which we now would classify as model-theoretic are exploited by Frege. Consider, for example, Theorem 263 of the Grundgesetze. This theorem states that:

Endlos [i.e., the number of natural numbers] is the number which belongs to a concept, if the objects falling under this concept may be ordered in a series, which begins with a certain object and continues without end, without coming back on itself and without branching (Frege [1966], §144).
In order to prove this theorem, Frege constructs an isomorphism between any such series, and the series of natural numbers. If, as Heck [1993] has argued, it is correct to view these conditions as characterizing Frege's axioms of arithmetic, then Frege has essentially shown that any series satisfying these conditions could, in some sense, serve as the natural numbers. In other words, Frege here essentially proves the categoricity of his system of axioms, which is a classical model theoretic result.

Whether Frege took himself as having proven a result of this kind is another question. But it is obvious that Frege was aware of the relevant mathematical techniques. Furthermore, as has been stressed by Tappenden [forthcoming], Frege was also aware of the developments in geometry which we were taking place in the latter half of the 19th century, some of which involved arriving at reinterpretations of systems of geometrical axioms. Frege was hence quite familiar with what we now call model-theoretic techniques.

What about the third question? Would Frege accept the model-theoretic characterization of logical consequence? If not, then there are model-theoretic results which would not, for him, have the foundational importance that they do for some of us. Let me say at the outset that I think that the negative answer to this question, which has been espoused in unpublished work and lecture by Burton Dreben, is the only claim which can be retained from this recent trend in Frege-interpretation. However, in part, this is due to its vagueness. As far as I can tell, any interesting interpretation of it seems to make it massively underdetermined by the textual evidence, whereas the most uninteresting interpretation, that Frege was simply unfamiliar with the characterization, is trivially true.

Here is one reason one might think that the contemporary characterization of logical truth would be anathema to the project of logicism. Surely there are possible domains with only finite cardinality. However, in order to construct the series of natural numbers, and in
particular, for the proof that every number has a successor, one needs logical principles which
guarantee the existence of an infinite domain. If so, then logicism requires certain existence
statements to be considered as logical truths. This may be thought to be inconsistent with the
model-theoretic conception of logical truth, since no interesting existence claims can be logical
truths on this picture.

However, I do not find this consideration compelling. Suppose, \textit{per impossibile}, that naive
set theory were consistent. In this case, we might consider the axioms of set-theory as logical
axioms, and the vocabulary of set-theory as logical vocabulary, and hence not subject to
reinterpretation. If so, then we, too, would consider certain existence claims to be logical
truths. But we could still use the model-theoretic definition of consequence. The domains we
would consider would all contain the logical objects, but would differ from one another in the
nature and number of the other elements.

I am not aware of any other possible technical incompatibility between model theory and
logicism. However, there may be some aspect of Frege's thought which would cause him to
view the model theoretic characterization of validity as a poor analysis of the ordinary notion.
If so, then Frege would be in agreement with contemporary philosophers of logic such as
Etchemendy, though presumably for different reasons. This would be an important finding, if
ture. The discovery of any good reason for rejecting the model-theoretic characterizations of
informal metalogical notions would be an important contribution to philosophy. Since Frege's
reasons are generally particularly good, we can conclude that if scholarship can unveil reasons
why Frege would reject Tarski's characterization of logical consequence, then, modulo the
degree of persuasiveness of the reasons, this would certainly count as an important
contribution to the philosophy of logic.
One way to obtain the conclusion that Frege would have rejected the model-theoretic characterization of informal semantical notions is to argue that Frege would have rejected any form of semantics whatsoever, that semantical concepts are, for Frege, illegitimate, as are all semantical questions about a system. If this were true, it would be disappointing. For most of us, it is improbable that we will ever be convinced of the illegitimacy of semantic concepts. However, we are generally less-attached to the model-theoretic characterization thereof. Thus, if Frege would have rejected the model-theoretic characterization of semantical notions for the reasons given by Ricketts, then his rejection is less interesting, and less relevant, then had it been based on more model-theory specific ones.

I do not know whether there is some feature of Frege's philosophy which would cause him to reject the model-theoretic characterization of semantical notions. But if there is, it is, for the reasons given in this paper, certainly not that he viewed semantical endeavors as illegitimate. Of course, given that Frege was interested in semantical questions, he obviously would be impressed by the perspicuity with which model-theory treats them. Thus, I suspect that Frege, like us, would be inclined on pragmatic grounds to take the model-theoretic characterizations as legitimate representations of our informal semantic notions. However, a suspicion is not an argument, and a resolution of this question awaits further research.  

---

34Thanks especially to George Boolos, Richard Cartwright, and Richard Heck for numerous detailed discussions, criticisms, and helpful suggestions. Burton Dreben's input has aided immensely, and he also deserves special thanks. Discussions with Thomas Ricketts have occasioned substantial and important changes, and I wish to thank him for his intellectual objectivity, interest, and patience. Thanks also to Michael Glanzburg, Daryl Jung, Kathrin Koslicki, Sanford Shieh, and Jamie Tappenden.
REFERENCES


------. 1970b: "Frege Against the Formalists", in Geach and Black, pp. 182-233.


------. 1979b: "Introduction to Logic", in *Posthumous Writings*, pp. 185-196.

----- 1984b: "On Mr. Peano's Conceptual Notation and My Own", in Collected Papers, pp. 234-248.


Tappenden, J. Forthcoming: "Geometry and Generality in Frege's Philosophy of Arithmetic".

Tarski, A. 1983: "The Concept of Truth in Formalized Languages", in his Logic, Semantics, Metamathematics (Indianapolis: Hackett), pp. 152-278.

Frege's Thesis

From its introduction, in 1891, to Frege's final writings, the notion of sense plays several distinct roles. I will concentrate on what I take to be the two key roles Sense has in Frege's writings. The first is that of providing an explanation of terms' having the referents they do; or, as shall call it here, the reference-fixing role of sense. The second is that of serving as the object of understanding. The sense of an expression is what a speaker knows, in virtue of which she counts as understanding, or being linguistically competent with, that expression.

Let us call the thesis that one notion can play both of these roles, Frege's Thesis. Frege's thesis has been severely criticized. According to Kripke, for example, "Frege should be criticized for using the term 'sense' in two senses. For he takes the sense of a designator to be its meaning; and he also takes it to be the way its reference is determined". In particular, it is felt that considerations of proper names and natural kind terms provide clear refutations of Frege's thesis. I wish, in this paper, to respond to these worries.

In Section I, I discuss the nature of the two roles which Frege attributes to Sense. I then draw two major consequences of Frege's Thesis. In Section II, I defend the first of these consequences: that a term in the language of a given speech community refers to an object in virtue of true beliefs competent speakers in the community have about that object. In Section

---

1 I thus ignore the role of sense as providing a criterion for the success of analytic definitions in mathematics, most prominently displayed in Gottlob Frege, "Logic in Mathematics", in his Posthumous Writings, ed. by H. Hermes, et al., (Chicago: University of Chicago Press, 1979), pp. 203-250, first published in 1914. Though I will not argue the point here, I am in agreement with Dummett that this latter notion of sense is incompatible with other, more central ones.

III, I defend the second of these consequences: that full understanding of a term requires non-trivial, uniquely identifying knowledge of its referent.

I.

As is well known, Frege sharply distinguishes between the reference [Bedeutung] of an expression and its sense [Sinn]. The reference of a singular term, for Frege, was the object it denoted; the reference of a predicate, the concept it denoted; and the reference of a sentence, its truth-value.\(^3\) Quantifiers, sentential connectives and the like are, on this account, taken to refer to functions of the appropriate sorts.

When Frege gave what was arguably the first semantic theory, in Part I of his Basic Laws of Arithmetic, his semantic clauses did not link expressions to the senses they express. Rather, their purpose was to assign referents of the appropriate sort.\(^4\) Thus, for Frege, as for Davidson, a semantic theory is fundamentally a theory of reference. By the semantic value of an expression \(e\), I shall mean whatever entity is assigned to \(e\) via the axioms or theorems of some semantic theory. For Frege, truth-values, rather than thoughts, are the semantic value of sentences, and the person John, rather than the sense of the name “John”, is the semantic value of “John”. Though Frege does, in his informal philosophical writings, suggest that senses can

---

\(^3\)It has recently become common to use the term “concept” as a name for the mode of presentation of a property, rather than the property itself. Though I endorse this terminological move for contemporary discussions, it does have the propensity to engender nefarious misreadings of Frege, as this is certainly not how Frege used the term. See Frege’s letter to Husserl of 5/24/1891, in Gottlob Frege, Philosophical and Mathematical Correspondence, ed. by G. Gabriel, tr. by H. Kaal (Chicago: University of Chicago Press, 1980).

\(^4\)See my “Truth and Metatheory in Frege” for a defense of the claim that the theory outlined in Part I of Gottlob Frege, Grundgesetze der Arithmetik (Hildesheim: Georg Olms, 1966) is a semantic theory in the modern sense.
serve as semantic values in non-extensional contexts, nowhere in his technical presentations
does he exploit them as such.

This fact raises an important question about the place of Sense in Frege’s philosophy. If,
as I believe to be the case, senses, for Frege, are not fundamentally intended to serve as
semantic values, what then is their purpose? In the remainder of this section, I shall address
this difficult question. The purpose of this discussion is to arrive at a clearer understanding of
the two roles of sense invoked in Frege’s Thesis.

Before we begin our discussion of the two roles of Sense, I would like to make one
caveat. Frege’s Thesis is thought to be refutable on the basis of considerations stemming from
proper names and natural kind terms.5 Thus, to focus our discussion, I shall restrict myself, in
the remainder of this paper, to expressions of these types. The referents and senses of
descriptions, predicates, and sentences are thus not at issue here.

With David Kaplan, let us distinguish between semantical questions and metasemantical
questions.6 Semantics is the project of assigning semantic values to expressions. Accordingly,
semantical issues are those concerning which semantic values particular expressions, or
categories of expressions, should receive, and their relation to the semantic values of simpler
expressions (if any) of which they are composed. Metasemantical questions concern rather
foundational metaphysical or epistemological questions about semantic theory.

5“Proper name” is here used in the contemporary sense, rather than Frege’s. Thus,
descriptions and sentences will not count as proper names.

6See David Kaplan, “Afterthoughts”, in J. Almog, et al., eds., Themes From Kaplan
“Reference and Necessity” (forthcoming), calls this the distinction between descriptive and
foundational semantics, and Martin Davies, in his Meaning, Quantification, and Necessity
(Boston: Routledge & Kegan Paul, 1981), calls it the distinction between “Theories of
meaning 1” and “Theories of meaning 2”. Despite being less sonorous than Stalnaker’s
terminology, I prefer Kaplan’s to these alternatives, as metasemantics stands to semantics in
much the same way that metaethics stands to ethics.
Its use as a referent in propositional attitude contexts notwithstanding, the sense of an expression is not primarily a semantical notion. Sense, for Frege, is fundamentally a metasemantical notion, invoked to answer two metasemantical questions. The first, epistemological question concerns knowledge of semantical facts: given a semantical fact, such as that "water" refers to water, what are the conditions under which one knows that semantical fact? In more pre-theoretic terms, this is the question of what it is in virtue of which a speaker counts as understanding the relevant expression. The second, metaphysical question concerns the nature of semantical facts: given a semantical fact, such as that "water" refers to water, what explains this fact, i.e., what is it in virtue of which it is true? The first of these questions is how Frege would formulate the notion of linguistic meaning. The second is the foundational question of reference-fixing.

The distinction between semantics and metasemantics is analogous to the distinction between Ethics and Metaethics. Suppose, for argument's sake, that there are ethical facts, and that they correspond to true ethical utterances. Then presumably (1) states such a fact:

(1) Torturing babies for fun or profit is wrong.

---


8 Before we proceed, a brief caveat about the relation of Sense to metasemantic questions must be raised. As I will suggest, sense is intended to explain two metasemantical questions: what is it in virtue of which a speaker knows semantic facts, and what is it in virtue of which an expression has the reference it does. However, this leaves open two other questions: what is it in virtue of which a speaker knows that an expression has the sense that it does, and what is it in virtue of which an expression has the sense that it does. A complete answer to the foundational questions of understanding and reference-fixing must also incorporate an answer to these latter questions. In a more expanded version of this paper, I argue that there are no substantive philosophical issues raised by such metasemantical questions about sense.
Two metaethical issues arise here. First, there is the metaphysical metaethical issue: what is it in virtue of which (1) is true? That is, what is it in virtue of which torturing babies for fun or profit is wrong? Second, there is the epistemological metaethical question: what does one have to know in order to know (1)? That is, what does one have to know in order to know that torturing babies for fun or profit is wrong?

Let us now turn to the distinction between semantic and metasemantic issues. For Frege, a semantic theory is a theory of reference, a recursive assignment of referents to expressions which yields, for each sentence in the language, a statement of the conditions under which it refers to the True. Among the axioms of a Fregean semantic theory for English would be something of the form:

(2) Ref("Bill Clinton") = Bill Clinton

(2) states a semantical fact about English. Accordingly, two metasemantic questions arise. The first is the metaphysical metasemantic question: what is it in virtue of which (2) is true? That is, what is it in virtue of which the name "Bill Clinton" refers to Bill Clinton? The second is the epistemological metasemantic question: what does one have to know in order to know (2)? That is, what does one have to know in order to know that "Bill Clinton" refers to Bill Clinton?

We now must turn to the question how sense can resolve these two metasemantical issues. First of all, consider the metaphysical metasemantic issue: what is it in virtue of which (2) is true? For Frege, this question is answered by appeal to sense: (2) is true in virtue of the fact that "Bill Clinton" expresses the sense it does.9

Let us now turn to the epistemological metasemantic issue: what must a speaker know in order to know that "Bill Clinton" refers to Bill Clinton? According to Frege, what a speaker

---

9How sense answers the metaphysical question is an issue which cannot be addressed here.
must know in order to know that "Bill Clinton" refers to Bill Clinton, is the Sense of the name "Bill Clinton". Stated in more pretheoretic terminology, in order to count as understanding the name "Bill Clinton", or as being linguistically competent with this term, one must, according to Frege, know its Sense.

It is important to understand exactly what is controversial about Frege's Thesis. For instance, some may think that Frege's assumption that there is an epistemological metasemantic question at all in the case of proper names or natural kind terms is already contentious. According to this line of thought, there is nothing to a proper name or natural kind term over and above its referent. Let us pause here to evaluate this objection to Frege.

The first thing to note about this objection is its lack of clarity. What is it to claim that there is nothing to a proper name or natural kind term over and above its referent? One construal of this claim could be that there is nothing one needs to know in order to understand such an expression. However, if this is the import of the thesis, then it seems false. Someone who does not know that "Boston" refers to a city, or that "water" refers to a liquid, does not understand the terms. A denial of such facts is tantamount to a refusal to take talk of understanding or linguistic competence seriously.

Another construal of the claim that there is nothing to a proper name or natural kind term over and above its referent is that all one needs to know in order to understand such an expression is the relevant clause in the theory of reference. On this construal, the only explanation for why someone understands an expression such as "water" is that they know that "water" refers to water. However, this construal too misses the metasemantical questions.

Consider an analogy to ethics. We may explain the fact that John does not kill Harry, even if he loathes him, and could easily get away with the crime, by attributing to John the knowledge that to kill Harry would be wrong. However, in so doing, we have not addressed
any metaethical issues. Metaethics enters at the point where we ask what it is in virtue of which John counts as knowing the ethical fact that to kill Harry would be wrong. Similarly, metasemantics enters in at the point where we ask what it is in virtue of which someone counts as knowing the semantical fact that “water” refers to water.

A full consideration of the position that there is nothing to a proper name or natural kind term over and above its referent would require a separate paper. However, my suspicion is that some arguments in support of this position suffer from a failure to grasp the distinction between semantics and metasemantics. For instance, such arguments sometimes seem to be motivated by the following assumption: if there were something to a name or natural kind term over and above its referent, then a semantic theory for a language containing such expressions would have to assign them entities distinct from their referents as their semantic values. These arguments then establish that candidate semantic value assignments distinct from referents result in a flawed semantic theory.  

But such arguments clearly result from a failure to grasp the distinction between semantics and metasemantics. For they are motivated by the assumption that any answer to the epistemological metasemantical question for a particular term must state the semantic value of that term. But with the semantics/metase\-semantics distinction in mind, it is evident that this assumption is unwarranted.  

Fundamentally, then, what should be considered controversial about Frege’s Thesis is not that it assumes that there is some answer to the epistemological metasemantical question in the case of proper names and natural kind terms. Rather, what should be controversial about the Thesis is that it assumes that one and the same notion can specify what it is in virtue of which

---

10 The classic argument of this form is Kripke’s modal argument in Lecture I of NN.
a term has the reference that it does, as well as specify what a speaker must know in order to be linguistically competent with that term.

To see why Frege's Thesis is controversial, consider a widely held picture of the reference-fixing of proper names and natural kind terms: the causal theory. According to this view, the explanation for "Bill Clinton"'s referring to Bill Clinton is that there is a causal chain linking the name "Bill Clinton" to Bill Clinton. However, this picture is inconsistent with Frege's Thesis. For a causal chain could not possibly be what a speaker know in virtue of which they count as linguistically competent with the name "Bill Clinton". Causal chains are not intended to be objects of knowledge. Thus, Frege's view that one notion both explains reference-fixing as well as linguistic competence rules out the causal theory as a possible answer to the reference-fixing metasemantic question.

In fact, Frege's Thesis has quite substantive consequences for both the theory of reference-fixing and the theory of understanding. Consider first its major consequence for the theory of reference-fixing. The sense of a term specifies what speakers competent with it know in virtue of which they count as understanding that term. This implies that a fact about the referent of the term can only serve as part of its Sense if it plays some role in the explanation of what it is in virtue of which speakers count as understanding, or being linguistically competent, with that term. If no speaker knows the fact in question, then it is difficult to see how that fact enters into an account of how speakers successfully communicate with the use of that term. Thus, a constraint on a piece of information counting as specifying the Sense of a term is that the information be known by at least some of the competent users of that term in the community. But since the sense of a term also provides the explanation of a term's having the referent it does, this implies that the facts in virtue of which a term has the referent it does
must be known by at least some speakers in the community. Let us call this the first consequence of Frege’s Thesis.

The first consequence of Frege’s Thesis is a result of the pressure placed on the reference-fixing role of sense by the requirement that it serve the purposes of the role of sense as providing an explanation of linguistic competence. The second consequence, on the other hand, follows from the strain put on the latter role of sense by the requirement that it fix reference. Since the Sense of a term serves as the explanation of a term’s having the referent it does, in order to count as linguistically competent with that term, a speaker must know what it is in virtue of which the term has the referent it does. Thus, Frege’s Thesis implies that linguistic competence with a term requires non-trivial, uniquely identifying knowledge of the referent of the term. Let us call this the second consequence of Frege’s Thesis. In the next two sections, I will turn to a defense of these consequences.

II.

If central aspects of what fixes reference must serve in an account of what a speaker knows who understands an expression, then, given the plausible thesis that each aspect of the meaning of an expression is known by some speakers in the community, Frege’s Thesis implies that, for each term, there are speakers in the community who know the conditions something must satisfy in order to be its referent. Thus, if reference succeeds, it does so in virtue of true beliefs held by members of a speech community. In this section, I shall defend

---

11 Or, as, for example, Christopher Peacocke would put it, understanding a term amounts to knowing what it is for something to be its referent. See his *A Study of Concepts* (Cambridge: MIT Press, 1992; especially 1.3, 2.4.

12 See Kripke’s discussion of Condition C (p. 71ff. of NN) for a statement and explanation of the non-triviality requirement.
this consequence against several objections. I hope thereby to show that it is not as controversial as is traditionally believed.

Frege's Thesis implies that reference is fixed by speakers' beliefs. In itself, this is not controversial. For theories which presuppose that a word refers partly in virtue of the referential intentions of its competent users also have the consequence that reference is, at least partly, a matter of speaker's beliefs. In order to have the intention to do something, I must have certain beliefs. I cannot have the intention to buy something at the store at 2 p.m., unless I believe that it is open at 2 p.m.. Thus, such theories also presuppose that a correct account of the reference relation must invoke speakers' beliefs. However, Frege's Thesis implies something stronger than this. It implies that reference is fixed entirely by speakers' beliefs.

Objections to the first consequence of Frege's Thesis stem from the view that it results in a notion of reference that depends too heavily upon the beliefs of individual speakers. Phenomena such as linguistic deference are taken to show that a use of a word by a given speaker can refer to an object, despite the fact that that speaker's beliefs would "individuate" another object, or would fail to individuate an object at all. I shall first show that the existence of one type of linguistic deference, which I shall call social deference, is predicted by the identification of the two aspects of sense. I shall then turn to a discussion of another type of deference, namely deference to the world. This sort of deference is indeed incompatible with Frege's Thesis. I will suggest that, to the extent to which the phenomena which deference to the world is intended to capture exist, they are describable in terms consistent with Frege's Thesis.

Here is an obvious fact about language. A speaker of a language can use a term of that language to refer to an object despite that speaker's lack of individuating knowledge, or even minimally true beliefs, about the object in question. If a person who is quite misinformed
about Mars, believing it to be a star, utters the sentence “Mars is not a planet”, she has spoken falsely. We do not, in general, cast about for an alternative reference for her utterance of “Mars” which makes that utterance true. Similarly, someone who knows nothing about Mars may ask, “What is Mars?” That is, someone with no knowledge of the referent of a term may nonetheless use that term to refer to that object in speech acts such as questions and assertions. In sum, a speaker can refer to an object without knowing that her term designates an object of the appropriate type.

The phenomena which concern us are thus those in which speakers seem to refer to an object with a term, despite the lack of individuating knowledge of that object. We can divide the phenomena into two classes. The first class concerns cases in which the speaker in question does not count as an “expert” with the term she utters; that is, she does not count as knowing the meaning of the term. Instances in which non-experts with a term successfully refer to an object despite their lack of individuating knowledge will be called instances of social deference. The second class of cases concerns those in which experts with a given term seem to refer to an object, despite the fact that the total knowledge of experts fails to be individuating. Instances of this second class will be called instances of deference to the world.

Let us first consider the phenomenon of social deference. Social deference is a phenomenon that is consistent with the first consequence of Frege’s Thesis. For according to this consequence, what fixes the reference of a term is the meaning of that term. But the meaning of a term is constituted by a subset of commonly accepted beliefs held by competent users of that term in the community. Thus, the only beliefs that count as relevant to fixing the reference of a term are those held by competent users of that term in the community. The beliefs of a non-competent user of a term are simply irrelevant to deciding the reference of
that term. Frege's Thesis predicts that reference-fixing is not sensitive to the beliefs of non-competent users. Hence, it predicts the existence of social deference.\textsuperscript{13}

The case of deference to the world is more problematic for Frege's Thesis. Indeed, a genuine case of deference to the world would serve as a straightforward counterexample to the first consequence of Frege's Thesis. For if a term can refer to an object, despite the lack of individuating knowledge on the part of its competent users in the community, then certainly it cannot be that the beliefs of the competent users are what actually serve to fix the reference of that term.

The classic purported instance of deference to the world stems from Strawson's seminal discussion of massive reduplication in Chapter 1 of Individuals. The moral of Strawson's case can be taken to be that experts' knowledge about the referent of a term could fail to distinguish that object from another object, removed in space or time, without affecting the relation of reference. That is, such cases seems to suggest that a term t can refer to an object O, even when experts in the community do not have individuating knowledge of O.

Strawson expresses the worry quite clearly in the following passage:

...it may seem, in the non-demonstrative identification of particulars, we depend ultimately on description in general terms alone. Now one may be very well informed about a particular sector of the universe. One may know beyond any doubt that there is only one particular thing or person in that sector which answers to a certain general description. But this, it might be argued, does not guarantee that the description applies uniquely. For there might be another particular, answering to the same description, in another sector of the universe. Even if one enlarges the description so that it incorporates a description of the salient features of the sector of the universe concerned, one still lacks a guarantee that the description individuates. For the other sector might reproduce these features too. However much one adds to the description of the sector one knows about—its internal detail and its external relations—this possibility of massive reduplication remains open. No extension of one's knowledge of the world can eliminate this possibility. So, however extensive the speaker's

\textsuperscript{13}To my knowledge, this is first pointed out in Gareth Evans' "The Causal Theory of Names", in Collected Papers, pp. 1-24.
knowledge and however extensive the hearer's, neither can know that the former's identifying description in fact applies uniquely.\textsuperscript{14}

Strawson's version of the massive reduplication argument is given in terms of spatial location. But a similar argument could be given on the temporal dimension. For it could be that the current state of our world exactly reduplicates a previous time-slice of it. For example, it could have been that some natural disaster completely wiped off the face of the planet some very advanced civilization that mirrored our own in every respect. Any purportedly uniquely identifying description, given in purely qualitative terms, would then fail to individuate due to the existence of qualitatively identical particulars existing in the previous time-slice of our world.

However, as Strawson points out, such arguments only demonstrate deference to the world if it is allowed that the contents of beliefs of experts can only be individuated by purely qualitative, general descriptions.\textsuperscript{15} Now, one might believe this is the case, if one holds that mental states are qualitative ideas, or picture-like representations. But of course this is a conception of mental states that is completely abhorrent to Frege. Frege grants the existence of ideas in this sense—he calls them "Vorstellungen"—but he repeatedly denies that they are to be identified with elements of thoughts. Frege's anti-psychologism is a precursor to externalist conceptions of mental states.\textsuperscript{16} On such approaches, two beliefs may differ, even though no

\textsuperscript{14}Peter Strawson, Individuals (London: Methuen, 1959), p. 20.

\textsuperscript{15}See Strawson, Ibid., p. 22.

\textsuperscript{16}I am not here asserting that Frege was everywhere consistent in his externalism. For instance, in his unpublished 1906 "Introduction to Logic", he clearly endorses a picture of the senses of proper names according to which they are "object independent", that is, a proper name, such as "Odysseus", expresses the sense it does, irrespective of whether or not the name has a referent. See his Posthumous Writings, , ed. by Hans Hermes, et al. tr. by P. Long and R. White (Chicago: University of Chicago Press, 1979), pp. 185-96, at p. 191. This passage poses a difficulty for externalist interpretations of Fregean sense of the sort advocated here and in Chapter 1 of Gareth Evans' Varieties of Reference (New York: Oxford University Press,
distinction could be drawn between the two with the use of purely qualitative, general
descriptions.

The massive reduplication argument for deference to the world assumes that linguistic
meaning must be given in terms of purely qualitative, general description, and concludes from
this that no notion which serves as the linguistic meaning of a non-descriptive expression can
also fix its reference. Since Frege wished sense to serve as linguistic meaning, it could thus
not also fix reference. However, there is good reason to believe that linguistic meaning cannot
be given in terms of purely qualitative, general description. For suppose there is a symmetrical
universe with a substance different from water, yet qualitatively indistinguishable. Suppose
further that the denizens of this universe call this substance “water”. I think it is obvious in
this case that the linguistic meaning of our term “water” differs from the linguistic meaning of
their term “water”.

I do not think the above intuition is controversial. It is implied by the view that the
linguistic meaning of a natural kind term like “water” is its referent, which is a thesis held by
some. However, one does not need to embrace such a view in order to account for this
intuition. One can equally well account for the intuition without denying that “water” and
“H₂O” have different linguistic meanings.

There is no reason to think that if the linguistic meaning of “water” is not its referent, then
one must give up the thesis that, if it is used to refer to something other than water in some
other part of the universe, “water” there expresses a different sense. It is furthermore
uncontroversial that this is indeed Frege’s view. That is, Frege ascribed to the thesis that the
Sense of an expression was not its referent as well as to the thesis that a difference in
reference entailed a difference in Sense. As Frege writes:

1982). Thanks to Richard Cartwright for discussion here.
Sentences and parts of sentences with different meanings [Bedeutungen] also have different senses. If in a sentence or part of a sentence one constituent is replaced by another with a different meaning, the different sentence or part that results does not have to have a different meaning from the original; on the other hand, it always has a different sense.\(^\text{17}\)

Thus, while Frege would obviously think that the term “water” expresses a sense which is not its referent, he also would hold that, if “water” is used to refer to a substance different from water, it then expresses a different sense.

I do not mean to say here that massive reduplication, or “twin” cases, are philosophically unimportant. But the project which they threaten is a quite specific one. As twin cases show, two people could be in identical (intrinsic) brain states, but differing mental states. This poses a problem for the project of accounting for mind/brain supervenience.\(^\text{18}\) But this project would be of no interest to Frege. Frege never discusses the topic of how thoughts could be neurophysiologically implemented, and no feature of his philosophy would cause him to consider mind/brain supervenience issues as providing any sort of constraint whatever on the identity conditions of senses.

Massive reduplication arguments thus do not pose a threat to Frege’s Thesis.\(^\text{19}\)

Considerations of deference do not show that Frege’s Thesis places too many psychological constraints on the relation of reference. Arguments that it does are based upon flawed

\(^{\text{17}}\)Gottlob Frege, “Notes for Ludwig Darmstaedter”, in his Posthumous Writings, pp. 253-257, at p. 255.

\(^{\text{18}}\)See Jerry Fodor, Psychosemantics (Cambridge: Bradford Books, 1987), Ch. 2, for a clear discussion of the problems twin cases raise for this project.

\(^{\text{19}}\)The expression “twin-earth case” is often used to apply both to symmetrical universes, as well as “twin” different possible worlds. I have not here addressed the latter kind of twin earth case, because I believe modal considerations raise special concerns not present in the non-modal scenario described by Strawson. For a discussion of some of these problems, see my “Rigidity and Content”, forthcoming in R. Heck, ed., Language, Logic, and Reality: Essays in Honor of Michael Dummett (Oxford: Oxford University Press, 1995).
conceptions of linguistic meaning, or conceptions of mental states which issue from concerns foreign to Frege. Thus, such considerations do not refute the first consequence of Frege’s Thesis.

III.

Yet the first consequence of Frege’s Thesis is not the most controversial one. It is the second consequence of Frege’s Thesis that has often been taken to demonstrate conclusively the implausibility of the identification. For instance, in Language and Reality, Devitt and Sterelney claim that, while the purported counterexamples to other consequences of Frege’s Thesis are “serious but not catastrophic”, those leveled against its second consequence are the ones that truly demonstrate its falsity.\(^{20}\) In this section, I turn to a defense of this consequence.

The second role of the sense of a term is to state what its competent users know in virtue of which they count as understanding it. Thus, since knowledge of sense, or meaning, explains understanding, sense is the primary notion in the theory of understanding. Broadly speaking, a theory of understanding must explain our remarkable linguistic acumen. More narrowly conceived, the core set of phenomena for which a theory of understanding is accountable are our intuitions about linguistic competence. Stated in the formal mode, a theory of understanding, to achieve descriptive adequacy, must explain our intuitions about the truth-conditions of instances of the schema \(’x \text{ understands the utterance } u’\), in various counterfactual situations.

A theory of understanding is a theory of semantic competence. Like other linguistic theories, then, it is necessary to abstract from performance errors. Thus, the counterfactual

situations in which we test theories of understanding should be those in which there is no question of an error in performance. That is, in evaluating the data for such a theory, we must abstract from situations in which failure to understand is due to poor acoustics, or other contextual factors, such as the failure of context to disambiguate utterances. In what follows, discussion of such cases should be presumed to be occurring in such idealized circumstances.

According to the second consequence of Frege's Thesis, full competence with a term requires non-trivial, uniquely identifying information about its referent. I shall defend this consequence by arguing that a theory of the second role of sense, that is, a theory of understanding, must invoke a highly non-trivial, or "rich", notion of meaning. According to Hilary Putnam, natural kind terms raise special problems for theories of understanding which avail themselves of rich notions of meaning, of the sort required by Frege's Thesis. I shall begin by discussing Putnam's worries. I shall then argue that a recognition of the context-dependency of ascriptions of understanding provides a defense of rich notions of meaning against Putnam's arguments.

In his paper, "The Meaning of 'Meaning'", Putnam raises a famous problem for a rich notion of meaning. According to Putnam, a speaker, such as himself, who knows nothing that distinguishes elms from beeches, nonetheless is linguistically competent with the terms, and hence understands them. If so, then understanding a term does not require individuating knowledge of its referent, contra the second consequence of Frege's Thesis.

If Putnam is correct in maintaining that he is competent with the terms "elm" and "beech" despite his lack of individuating knowledge of elms and beeches, then the second consequence of Frege's Thesis imposes too stringent conditions for competence with a term. The interest of Putnam's case thus lies in the tension it reveals between the need for sense to play a central role in a theory of linguistic competence, and the second consequence of Frege's Thesis. As
we have seen, one result of the Thesis is that speakers who know the meaning of a term are predicted to have uniquely identifying knowledge of its referent. Furthermore, such knowledge is non-trivial; the fact that an account of the meaning of a term must also serve as an *explanation* of how the term gets the reference it does bars knowledge of trivial descriptions, such as "the reference of t", from conferring understanding of t to a speaker. Yet surely there is a sense of linguistic competence in which Putnam, despite his lack of individuating knowledge about elms, nonetheless counts as a competent user of the term. Compare Putnam with, for example, a monolingual speaker of Chinese, who cannot use the term "elm" at all. It seems intuitive to say that Putnam has certain knowledge concerning the term "elm" that the monolingual speaker of Chinese lacks, and it is in virtue of this knowledge that Putnam can use the term to utter grammatical sentences of English. But if we identify this knowledge with the *meaning* of the term "elm", then the resultant notion of meaning is far weaker than that required by Frege's Thesis.

Frege's strategy results in a conception of meaning according to which it is not a simple matter to attain knowledge of the meaning of a term. In order to count as knowing the meaning of a term, a speaker must be in possession of non-trivial individuating knowledge about its referent. Such a concept of meaning is thus similar in certain respects to a very intuitive characterization of meaning as dictionary definition. But cases such as Putnam's elm/beech example are intended to demonstrate that the "intuitive notion of meaning" is actually less demanding than the notion defined by Frege's strategy.

Let us grant that Putnam is correct in thinking that there is *some* notion of linguistic competence such that a speaker counts as knowing the meaning of a term in this sense without having the sort of knowledge required in order to know the meaning in the sense imposed by

---

21 This is, again, Condition (C) of NN.
Frege's Thesis. According to James Higginbotham, for instance, there are two different notions of linguistic knowledge. The first, which he calls "knowledge of reference", is less demanding than the second, which can, for our purposes, simply be called "knowledge of meaning". One counts as knowing the reference of a term, (or, in the limit case, the truth conditions of a sentence) simply in virtue of knowing its grammatical category and semantic type. However, such knowledge is only necessary, and not sufficient, for full-fledged knowledge of meaning. The latter requires, in addition, knowledge of something like the dictionary definition of the lexical items. Given this distinction, the Fregean could maintain that what lies behind elm/beech type cases is an equivocation between these two relatively clear notions of linguistic competence.

The Fregean's task is of course not an easy one. The pressure Putnam's examples place on the notion of meaning presupposed by Frege's strategy comes when one considers the meaning of a sentence as the object of communication. If knowledge of meaning is non-trivial enough to imply that speakers are often ignorant of meaning, and hence fail to understand the utterances of others, then it is difficult to account for our remarkable communicative success. A successful defense of the notion of meaning that results from Frege's Thesis must thus exploit a concept such as that of a speaker's partial knowledge of meaning in an analysis of the phenomenon of communication.

---


23 This is how I take Higginbotham's talk of "knowing the logical skeleton" of a sentence (Ibid., p. 166).

24 It is not clear that what Higginbotham calls "Elucidations of meaning" will result in something so informative as a dictionary definition, but it seems that this is a natural identification.
This point has been clearly recognized by those who defend a Fregean picture of sense, such as Dummett:

The notion of sense has to do with the speakers' understanding of their language, that is, with their grasp of meaning. Since Frege wrote, we have become highly conscious of a distinction he barely recognized, that between the meaning of an expression in the common language and an individual speaker's grasp of its meaning. This distinction is not forced on us merely by speakers' invariably imperfect knowledge of their language—their partial understanding of some words and misunderstanding of others: we need also to attend to the connections which a speaker makes between a name and its bearer which he does not suppose to underlie its use in the language and which are not of a kind to do so...[1]It is evident that an adequate account of how language functions cannot ignore the fact that speech is an activity of rational agents, whose reasons for saying what they do rest upon their admittedly limited awareness of the meanings of the words they use.25

Sense, conceived of as public linguistic meaning, still, on this picture, retains its primary role as the ultimate "object" of communication, since meaning in a public language can be taken to be conceptually prior to a speaker's partial knowledge of it.

Of course, the legitimacy of importing the notion of partial knowledge of meaning must be defended on explanatory grounds. More specifically, it must be argued that a theory that invokes a rich conception of meaning, together with partial knowledge, provides a better explanation of the phenomena under consideration, than competing frameworks. Given what we have said about the core data of a theory of understanding, we thus must demonstrate that ordinary ascriptions of understanding are best explained by such a theory.

As we have seen, for Frege, understanding an expression amounts to knowing what its sense is. Thus, understanding is a species of knowledge-what. But ascriptions of knowledge-what are context-dependent. Whether or not an utterance of "John knows what horses are" is true depends heavily on context. Suppose John was among people who could not distinguish horses from dogs, but that John could distinguish horses from dogs, though not from, say,

donkeys. If someone then pointed to a dog, and called it a horse, it seems to me John could
legitimately respond by asserting that he knew what horses were, and that they were not dogs. However, if John were among horse-experts, and ascribed horse-hood to a donkey, they could truly assert that he did not know what horses are.

It is a difficult question exactly what kind of knowledge-ascription ascriptions of understanding are. However, one feature they share with other ascriptions of knowledge-what is their context-dependency. Consider, for example, Putnam in a normal conversational situation, in which no botanists are present. If, in such a situation, a speaker utters the sentence, “An elm is a nice tree”, then it seems that we can attribute to Putnam an understanding of the occurrence of the term “elm”, despite his lack of knowledge of the full meaning of “elm”. However, suppose Putnam is in a situation in which expert botanists are holding a technical discussion about subtle differences between elms and other sorts of trees. In such a situation, Putnam could truly assert either that he does not know what elms are, or he does not really understand utterances of the word “elm” on that occasion.

Thus, in a situation in which we are comparing Putnam to ordinary non-elm experts, we might assert that he knows the meaning of “elm”. Among botanists, however, the claim that he knows the meaning of “elm” is more dubious. Thus, the truth of utterances of the form, ‘x understands the expression e’ depends upon the context in which they are uttered. One natural way of accounting for this phenomenon is via the notion of partial knowledge of meaning. On the Fregean conception, the (full) meaning of a term consists of some set of conditions which, taken together, apply uniquely to an object. The context-dependence of ascriptions of knowledge of meaning can be taken as providing evidence that the truth-conditions of such ascriptions invoke the concept of partial knowledge of meaning. That is, an instance of the
schema, \( f_x \) understands the expression \( e^1 \) is true just in case she knows those aspects of the (full) meaning of the term that are relevant in the context in which that instance is uttered.\(^{26}\)

I have argued that locutions of the form \( f_x \) understands the expression \( e^1 \) are context-dependent. The two "clear" notions of linguistic competence, Higginbotham's knowledge of reference, and knowledge of (full) meaning, can then be taken as the limit cases of the context dependency. Situations in which instances of the schema \( f_x \) understands the expression \( e^1 \) are true despite a potential lack of individuating knowledge of the referent of \( e \), are then accounted for via the notion of partial knowledge of meaning.

Now, the (full) meaning of a term, on a theory of understanding that recognizes the context-dependency of ascriptions of knowledge of meaning, will be perhaps richer than that required by Frege's identification. One worry about such a notion of meaning is that it may incorporate too much of what seems to be contingent information about the object into the meaning of a term. There are two related points to make about this worry. First of all, a sharp distinction between contingent information about an object and information about its meaning is in any case suspect. The second point is that, just as fallibilism in epistemology asserts that we can be certain about contingent propositions, so an epistemological view of meaning must be content with metaphysically contingent truths in virtue of meaning.\(^{27}\)

\(^{26}\)I am not here suggesting that the context-dependency of expressions such as \( f_x \) understands the expression \( \varphi^1 \), or \( f_x \) knows the meaning of \( N^1 \), cannot be accounted for in a theory that fails to invoke the concept of partial knowledge of meaning. One could, for instance, argue that the full meaning of an expression is itself a context dependent matter. That is, one could argue that the entire meaning of a term changes from context to context. One could then identify what is common to all utterances of an expression \( e \) with the function from contexts to full meanings. However, giving a coherent characterization of the function in question is not a simple task.

\(^{27}\)See my "Rigidity and Content" for discussion.
In a paper solely devoted to the theory of understanding, I would dwell more upon the sort of notion of meaning that I have argued such a theory must presuppose. Yet for our purposes, further investigation into the details is not required. For what we have seen is that a plausible theory of understanding can invoke a notion of meaning that, at the very least, determines reference.\textsuperscript{28}

\textsuperscript{28}Thanks to Robert Stalnaker and Richard Cartwright for discussion and criticism.
If d designates the object o, then, if we abstract from worlds in which o does not exist, d is a rigid designator just in case d has the same extension in all possible worlds.\(^1\) Rigid designation concerns the behavior of certain terms when evaluated with respect to counterfactual situations. Hence, one might think that the distinction between rigid and non-rigid expressions is one which is relevant only when the modal status of an utterance is in question. Yet many philosophers hold that this distinction is, demonstrably, one which has relevance for what is expressed by, or the truth-conditions of, utterances of ordinary unmodalized sentences.

In many of his writings, Michael Dummett has voiced objections to this view. However, his points have generally been misunderstood or ignored. Philosophers have continued to assume that rigidity has been shown, in Kripke's words, to affect "the truth conditions...of (the propositions expressed by) all sentences, including simple sentences".\(^2\) In this paper, I will attempt to uncover the possible arguments for this view. With Dummett, I shall argue that the thesis that rigidity affects the content of simple sentences should be viewed with more suspicion than recent philosophy of language suggests.

To set up the issue, we must first provide a characterization of what is asserted by an utterance of an ordinary unmodalized sentence. This I attempt in Section I. In Section II, I consider and reject several semantical arguments for the thesis that rigidity is somehow

---

\(^1\)I am using the definition of rigidity given in a letter from Kripke to Kaplan, cited on p. 569 of Kaplan's "Afterthoughts" (pp. 565-614 of Themes from Kaplan [Oxford: Oxford University Press, 1989]). Other forms of rigidity, such as temporal rigidity, are not at issue in this paper.

relevant for individuating what is asserted by utterances of unmodalized sentences. In Section III, I argue that this thesis depends upon an unsupported conception of the relation between assertoric content and metaphysical possibility.

Section I.

The focus of this paper is the thesis that rigidity affects the content of simple sentences. This thesis is a significant philosophical claim. It is significant because it sets significant constraints on the identity conditions of the objects of assertion. It is philosophical because an explication of the notion of the content of assertion is a central task of philosophy. By setting constraints on the identity conditions of the objects of assertion, the thesis constrains the class of possible explications of this notion. The goal of this paper is to evaluate whether it correctly constrains this class.

There are two general models of philosophical explication. The first is rational reconstruction. When one rationally reconstructs a concept, one does not concern oneself with faithfulness to some pre-theoretic notion. Only those features of the intuitive notion that can, in Carnap's words, be "rationally justified" are to be features of the reconstructed concept. A rational reconstruction is judged by how fruitful the reconstructed concept turns out to be. The second model of philosophical explication is a description of some intuitive, pre-theoretic notion. This should be judged on how much it matches our pre-theoretic intuitions about the concept. Every actual philosophical explication is a confusing mixture of both models.

The project of giving a philosophical explication of the objects of assertion is no different in this respect from other philosophical explications. Simply describing ordinary usage of locutions such as "what John said" will certainly not lead to any sort of theoretically fruitful concept. However, completely abandoning any pre-theoretic notion of content may leave us
with nothing about which to theorize fruitfully. In this section, I will attempt to flesh out this project more fully, with the goal of showing why the thesis that rigidity affects the content of assertions of simple sentences should not be blindly accepted.

Asserting that \( p \) is an act. As Richard Cartwright writes, "We sometimes contrast saying (asserting, stating) with doing; but in a wider sense to say something is to do something". As with any act, we must distinguish the act type from the act token. The act type is asserting that \( p \). A token of this type occurs when, on a particular occasion, someone asserts that \( p \). But aside from the act type, asserting that \( p \), and a given tokening of that type, we must also recognize the content of the assertion. The content of an assertion is not an act; it is not "done". It is, as Cartwright writes, "not the sort of thing that can be done".

At least one of the things which one accomplishes with the act of assertion is to distinguish possibilities. By asserting that \( p \), I am distinguishing the possible circumstances in which it is true that \( p \), from those in which it is not true that \( p \). There are two points to make about this observation. First of all, for many substitutions of different sentences for "\( p \)" and "\( q \)", asserting that \( p \) is a different action from asserting that \( q \). Now, kicking Paul is a different action from kicking Clem, if Paul is not Clem. Analogously, we account for intuitive distinctions between asserting that \( p \) and asserting that \( q \) by allowing the possible circumstances to differ in the two cases.

---

\(^3\)For instance, it has been argued that our ordinary intuition that "Hesperus is Phosphorus" and "Hesperus is Hesperus" say different things cannot be preserved on a systematic, reconstructed notion of meaning. But if it could be shown that our ordinary intuition that "Bill Clinton is President" and "2 + 2 = 4" say different things also cannot be preserved, we would view the result as casting doubt on the very project of systematically characterizing a notion of content.


\(^5\)Ibid., p. 36.
The second point is that the sense of possibility here is completely neutral. It is the sense of possibility invoked when one, in the course of a proof in mathematics, asserts that there are three distinct possibilities. On a straightforward account of mathematical assertion, one cannot take the possibility in question to be metaphysical. What is prima facie metaphysically impossible could very well, from the perspective of this notion, count as possible. Since to assert that $2 + 2 = 4$ is clearly to do something different from asserting that Peano Arithmetic is incomplete, in this neutral sense of possibility, they are true in different circumstances.

The unexplanatory nature of the sense of possibility in question provokes two opposing reactions. First of all, one could, with Robert Stalnaker, attempt to reduce this notion of possibility to metaphysical possibility. On this account, the content of an assertion that $p$ and an assertion that $q$ are the same just in case they distinguish different metaphysical possibilities. One must, however, reanalyze the content of apparent metaphysical impossibilities to make them come out metaphysically possible after all. The second route one could take is to deny that talk of possibility in this context is to be taken as explanatory at all. For if we tried to explain the difference in assertoric content between an utterance of "$2 + 2 = 4$" and "PA is incomplete" by invoking different possible circumstance in which they are true, there may be no other way to describe the different circumstances except by appealing to the difference in assertoric content to be explained. If this is correct, then one does not need to invoke metaphysical possibility at all in distinguishing the contents of assertions: some other account does the required work.

---

6 I owe this example to Richard Heck, who used it in a lengthy e-mail debate between Block, Heck, Stalnaker, and myself.


8 For instance, one could account for a difference in the content of two assertions by adverting to differing "Russellian propositions".

72
How does one assert that p? The typical method is to utter some sentence. Sentences are sequences of word-types. Utterances are tokens of these sequences produced by utterers. Now, one may utter some sentence, but fail to assert anything at all. For present purposes, we shall abstract away from this possibility, and assume that any utterance of a sentence on an occasion in fact asserts something. Furthermore, in what follows, we consider only utterances of "simple" sentences, that is, sentences not containing any modal vocabulary.

With these distinctions in place, we may now broach the issue of rigidity. I shall take rigidity to be a property of (non-sentential) word-types. The thesis under consideration is as follows. If one utters a sentence containing a rigid term, then one has asserted something different from what one would have asserted had one uttered, on that very same occasion, any sentence differing from the original one only in the substitution of a non-rigid term for the rigid one. Loosely following Dummett's terminological practice, let us say that the content of an assertion which is made by uttering a sentence S is the assertoric content of that utterance of S. Then the thesis under consideration is that if S and S' are two sentence which differ

---

9I am thus using "utterance" as synonymous with sentence-token rather than to refer to the act of uttering a sentence-token. I will occasionally also use it in this latter sense, leaving it to context to disambiguate.

10We also abstract away from sentences containing propositional attitude verbs.

11Strictly speaking, this characterization is inappropriate. For, as Charles Parsons has pointed out to me, there are (albeit rather concocted) examples of terms such that whether or not they are rigid depends upon context. Consider, for example, the description "the x such that x = me if I am Jason Stanley, or x = the present Prime Minister of Norway if not".

12One must be careful to distinguish this from the trivially true thesis that what one would have done would have been different. Of course, there is a sense in which what one would have done would have different, for one would be uttering different words.

13"Loosely", because Dummett typically takes assertoric content to attach to sentences.
only in that one contains a rigid term where the other contains a non-rigid term, then no two utterances of S and S' can have the same assertoric content.

It will be useful to have a notion of content for sub-sentential expressions. Since assertoric content is fundamentally a property of utterances in contexts, the content of a sub-sentential expression will also be taken relative to a context. We restrict our attention to terms, for simplicity's sake. Consider two terms (word-types), t and t'. We shall say that t has the same content, with respect to a context c, as t', just in case, for any sentence S which contains t as a constituent, an utterance u of S in c would have the same assertoric content as an utterance of a sentence which results from S by replacing t' for t. Then the thesis can be stated as: no rigid term ever has the same content as a non-rigid term. Let us call this thesis RT.

If RT is not to be trivially false, we must make a further distinction. For sentences could, in a quite intuitive sense, say the same thing in one context, though not in many others. Consider, for example, a context in which it is obvious to all participants that John is a man. In such a context, there may be no felt difference between asserting that John is a bachelor, and asserting that John is unmarried. However, in other contexts, in which John's sex is unknown, there would clearly be such a difference. Similarly, in a context in which it is obvious that John is the tallest man in the room, there may be no felt difference between asserting that John likes Mary, and asserting that the tallest man in the room likes Mary. Thus, we must come up with some notion of the content of an assertion, according to which such contexts do not count as refutations of RT.

Though this task requires several idealizations, I think it nonetheless can be accomplished. For even in contexts in which it is known that John is the tallest man in the room, there is a sense of content in which asserting that the tallest man in the room likes Mary and asserting that John likes Mary have different contents. Let us distinguish what is communicated by an
utterance of a sentence from what the utterance of that sentence expressed on that occasion.

We shall take both what is communicated as well as what is expressed by the utterance to be objects of the same-kind -- in vague parlance, "truth-conditions", or "propositions".

The notion of what an utterance of a sentence expresses and what an utterance of a sentence communicates are related by the following principle, which we may call the Expression-Communication Principle, or ECP.

If an utterance u of a sentence S expresses something different from an utterance u' of another sentence S', then, ceteris paribus, for any normal context c, had S and S' been uttered in c, they would have communicated different things.¹⁴

We shall take assertoric content as explicating the notion of what an utterance expresses. ECP reflects the obvious truth that differences in what is expressed should be reflected in differences in use. It is also worth emphasizing that the legitimacy of the Gricean attempt to reduce applied timeless meaning to utterer's-occasion meaning depends crucially on the truth of ECP. If ECP is false, then no reduction of literal meaning to intended meaning is possible. However, even one who lacks such reductive aspirations should nonetheless accept ECP, since even if literal meaning is not reducible to intended meaning, there surely must exist supervenience relations between the two.

Suppose RT is true. Then, if t is rigid, and t' is non-rigid, for any two sentences S and S' which differ only in that t occurs in the former where t' occurs in the latter, any utterance of S

¹⁴By a normal context I mean one in which the speakers are competent with all of the words in the sentences being uttered, and the sentence is used as it standardly is. The notion of a sentence being standardly used perhaps is characterizable with the use of something analogous to Grice's "central range of speech acts" -- see his "Utterer's Meaning, Sentence-Meaning, and Word-Meaning", pp. 117-137 of his Study in the Way of Words [Cambridge: Harvard University Press, 1989]. By "competent speaker" I mean someone who is considered authoritative with respect to that term in that community. Contexts thus count as normal relative to the sentences being uttered. From now on, I shall use "context" to mean "normal context".
expresses something different from an utterance of $S'$. RT is worrisome for the following two reasons.

The first reason why RT is worrisome is that there seem to be obvious counter-examples. Consider, for example, the two terms "the president", and "the actual president". Suppose Bill Clinton drops by for a surprise visit. I could convey this by uttering either:

(1) The President of the United States came by for a visit

or:

(2) The actual President of the United States came by for a visit

The difference between (1) and (2) does not seem to me to be a difference in assertoric content. For the difference between utterances of (1) and (2) is not reflected in truth-conditions. It is rather one of "coloring", or "tone", like the difference between typical unphilosophical usages of "truth" and "absolute truth". I would use "the actual President", rather than just "the President", not to distinguish different possibilities, but rather for emphasis, to convey my surprise. Nonetheless, one term is rigid, while the other is not, so, according to RT, sentences containing them must say different things.

---

15 Though let me emphasize that such intuitions of what is and is not truth-conditional are, of course, just as defeasible as similar intuitions about what is asserted. For they depend upon a certain conception of what truth-conditions are, which may turn out to be theoretically less motivated than another. However, I believe that no notion of truth-condition can incorporate everything which Frege called the associations of an expression, and the difference between (1) and (2) is only accountable in terms of this latter category (see p. 139 ff. of his "Logic" (pp. 126-151 of Posthumous Writings [Chicago: University of Chicago Press, 1979]).

16 The defender of RT might respond here by maintaining that (1) and (2) conventionally implicate different things, and therefore express different things. For (2) might be said to conventionally implicate that the utterer is surprised at the fact that the president came for a visit, whereas (1) carries no such implication. However, as Grice has often maintained, two utterances may conventionally implicate different things, yet say the same thing (c.f. Paul Grice, "Logic and Conversation", in his Studies in the Ways of Words (Cambridge MA: Harvard University Press, 1989), pp. 22-40, at pp. 25ff.). I believe the difference between utterances of (1) and (2) falls in this category. Even if the conventional implicature exists, it is not of the sort which affects assertoric content, since (2) would be true even if the utterer
The second reason why RT is worrisome is slightly more theoretical. As we have seen, the background knowledge of the participants in a context affects what is communicated by an utterance of a sentence. By invoking the distinction between what is expressed and what is communicated, these effects may be mitigated somewhat. Yet surely, what is expressed is linked to what is communicated at least via ECP. But even where S and S' are simple sentences, RT implies that speakers typically communicate different things by uttering them. But initially, there is no reason to believe that the distinction between rigidity and non-rigidity, by itself, provides for the existence of such a difference.

To take an example from Dummett, consider the name "St. Anne". Now, it is plausible that any speaker competent with this term knows that it refers to the mother of the Virgin Mary. However, "St. Anne" is rigid, whereas "the mother of the Virgin Mary" is not. But it is by no means clear that utterances of two sentences, differing only in that one of these terms is replaced by the other, can typically communicate different things. However, RT implies just this.

More generally, suppose that there is some non-rigid definite description DD such that anyone who uses the name N on a regular basis knows that it denotes the bearer of that name. Then, despite the fact that the name is rigid while the description is not, competent speakers typically would communicate the same thing by an utterance of a sentence containing the name as they would by uttering the sentence with the description replaced by the name. The fact that most proper names do not have the same content as any non-rigid expression seems not so much due to a modal distinction as to the fact that there is usually no one definite expected the visit.

17In every world in which the Virgin Mary exists, "the mother of the Virgin Mary" denotes St. Anne. However, there are worlds in which St. Anne exists, but "the mother of the Virgin Mary" does not refer.
description which all speakers who use a name on a regular basis know applies to its referent. But surely, this is a contingent fact.

There are two responses the defender of RT could give in reply to these worries. The first would be to bite the bullet, and maintain that despite intuitions to the contrary, rigidity and non-rigidity, by themselves, make a difference to what is communicated. The second response would be to claim that the occurrence of a rigid term in an utterance of one sentence, and a non-rigid term in an utterance of another, is sufficient to conclude that the assertoric contents are different, even if utterances of the sentences typically communicate the same things. This second response thus involves the denial of ECP. Both of these responses require motivation. The first requires an argument that the distinction between rigid and non-rigid expressions does, after all, always affect what is communicated. The second requires an argument that a notion for which ECP does not hold is a notion of content at all.

I shall not treat the first of these responses in this paper. It is certainly possible to introduce, as does Gareth Evans, a rigid designator whose reference is fixed by a single non-rigid description, and which is such that anyone who is competent with the term knows the reference-fixing stipulation. In such a case, simple sentences containing them will typically communicate the same thing.

Instead, I shall consider, in the rest of the paper, the second of these responses, that is, arguments which imply that utterances of two sentences could have different assertoric contents, even if utterances of them typically would communicate the same thing. One such argument might come from considerations of the semantics of natural language. For rigid terms and non-rigid terms are assigned different semantic values in a compositional semantic

---

theory for languages containing modal expressions. It might be thought that this difference in semantic value implies a difference in content, even if no difference in what is communicated exists. In the next section, we turn to an examination of this argument.

Section II.

In this section, I shall discuss how considerations from the formal semantics of natural language bear upon the evaluation of RT. I shall begin by introducing a thesis about content which implies RT, and underlies its acceptance. This thesis, as we shall see, follows from a trivial modal semantical fact, if there exist entailment relations between the notion of content discussed in the previous section, and modal semantic value. We then consider two theses which, if true, would establish the required entailment relations. The first of these proceeds from general considerations about semantic value, and its relation to content. The second identifies content with modal semantic value. Both of these theses, I shall argue, are false.

In Lecture I of Naming and Necessity, Saul Kripke considers two possible versions of a "cluster of descriptions" theory of names. The sole difference between these two theories lies in the fact that one theory, but not the other, accepts the following thesis

(*) The statement, 'If X exists, then X has most of the q's' expresses a necessary truth.

According to Kripke, the distinction between these two theories can be characterized as follow. On the theory which embraces Thesis (*), the cluster of descriptions gives the meaning of the

---

19 A terminological note: the expression "semantic value" is to be interpreted relative to a semantic theory. By the semantic value of an expression relative to a semantic theory T, I mean whatever entity T assigns to that expression via either its axioms or its theorems.

20 Naming and Necessity, p. 65.
name, whereas someone who rejects Thesis (*) "doesn't think that the cluster is part of the meaning of the name".  

Instead of "cluster-theories", let us suppose that these theories are "one-description" theories. The two theories then differ in that one accepts, while the other rejects:

(*) The statement, 'If X exists, then X is the φ' expresses a necessary truth.

Such "one-description" theories are certainly false, and no contemporary philosopher holds such a view. But using this theory will simplify our exposition considerably, without prejudicing the discussion.

Thus, we have two theories of proper names, which differ only in that one embraces, and the other rejects, Thesis (*). As we have seen, according to Kripke, the way to describe the difference between these theories is that according to the one that accepts Thesis (*), the description gives the meaning of the proper name, whereas according to the theory that rejects Thesis (*), it does not. Considering only possibilities in which φ refers, it thus is the case, on Kripke's view, that

If an expression φ has the same meaning as an expression ψ, then the sentence 'φ is ψ' must express a necessary truth.

Let us call this the Meaning Assumption, or MA.

What is relevant for our purposes is that a principle very similar to MA implies RT. This principle I shall call the content assumption, or CA. It is as follows:

If an expression φ has the same content as an expression ψ in a context c, then an utterance in c of the sentence 'φ is ψ' must express a necessary truth.

If CA is true, then RT is true. For let e be some rigid expression, and e' some non-rigid expression. Then utterances of 'e is e' are not necessary, and hence, by CA, e and e' do not have the same content relative to any context. Whether or not Kripke has, by his acceptance of

21 Ibid.
MA, endorsed CA, depends upon what Kripke intended to express by his use of the term "meaning".

A central concern of Kripke's was with expressions which had hitherto resisted treatment from the perspective of formal semantics. In particular, Kripke's semantics revolutionized the study of quantified modal logic. Kripke semantics revealed the semantic import of sentences of languages which combined modal operators with quantifiers. This showed why sentences such as the Barcan Formula and its converse, made controversial metaphysical claims. Finally, Kripke showed how to develop axiom systems which had the appropriate metaphysical neutrality.²²

The axiom system developed by Kripke in "Semantical Considerations" lacked designators -- it contained neither constants nor free-variables.²³ But axiom systems which lack constants seem inappropriate as representations of ordinary modal discourse. However, it was well-known that adding non-rigid designators led to rather drastic failures of traditional logical laws. If proper names could be shown to rigid, then the task of representing modal discourse with the use of quantified modal logic would become that much easier.²⁴

It is thus tempting to take his use of an expression such as "meaning" as semantic value in a standard modal semantics. This interpretation of Kripke's talk of meaning would also justify


²³Free variables were given the closure interpretation, to block the derivation of the Converse Barcan Formula.

²⁴It is often assumed that the failures of substitution are due to the nature of the terms in the language. However, Robert Stalnaker has persuasively argued that they are due instead to a failure to recognize the difference between an arbitrary formula and a predication. Though present conceptually in the extensional first-order theory, the difference becomes important when one adds intensional operators. If one follows Stalnaker, then there will be no violations of substitution, even if the language contains non-rigid terms. See his "Complex Predicates" and "The Interaction of Quantification with Identity and Modality", forthcoming.
Kripke's belief in the *obviousness* of MA. For if this is what Kripke had intended by his use of "meaning", then MA would be a trivial consequence of the standard definition of identity.

Let "Val" express the denotation function from expressions of the object-language and possible worlds to the extensions of those expressions in those worlds. In other words, 

\[ \text{Val}(e, w)^1 \] denotes the extension of \( e \) taken with respect to \( w \). If \( e \) is a singular term, then 

\[ \text{Val}(e, w)^1 \] denotes an individual; if \( e \) is a predicate-expression, then \( \text{Val}(e, w)^1 \) denotes a set of individuals; and if \( e \) is a sentence, then \( \text{Val}(e, w)^1 \) denotes a truth-value. Furthermore, assume the following as semantical axioms governing the object-language expressions, "\( \Box \)" and "\( = \).\(^{25}\)

(a) \( \Box p \) is true iff \( \forall w (\text{Val}(p, w) = t) \)

(b) \( \text{Val}(\Box e = e^1, w) = t \) iff \( \text{Val}(e, w) = \text{Val}(e', w) \)

If "meaning" in the statement of MA means semantic value in a standard modal semantics, then we may state it as follows:

\[ (\text{MA}^*) \forall w (\text{Val}(e, w) = \text{Val}(e', w)) \text{ iff } \Box e = e^1 \text{ is true} \]

MA* follows trivially from these standard axioms.

However, Kripke's talk of meaning almost certainly invokes a pre-theoretic concept. Thus, we should resist the temptation to identify his use of the term with semantic value in a modal semantics. One might, however, think that the triviality of MA* implies the triviality of CA. That is, one might think that CA is itself a direct consequence of the definition of identity in a traditional modal semantics. But this view is mistaken. For MA*, on our interpretation of it, employs the notion of semantic value, which is explicated in terms of functions from possible worlds into extensions. CA, on the other hand, alludes to the intuitive notion of content. One

\(^{25}\)Warning: I am also using "\( = \)" as the sign for the identity relation in the metalanguage.
can move from MA* to CA only if there exist entailment relations between the possible worlds notion of semantic value, and the notion of content discussed in Section 1.

It might be thought that the required entailment relations which allow one to move from MA* to CA stem, not from the particular notion of semantic value employed by modal semantic theories, but rather from a more general thesis about the relation of content to semantic value. In particular, one might hold the following thesis, which we shall call the Semantic Value Principle, or SVP:

Sameness of content (relative to a context c) implies sameness of semantic value in any interesting, true semantic theory.

Once the theoretical fruitfulness of modal semantics is accepted, SVP allows one to move from MA* to CA. For if two terms have the same content relative to some context, then the two terms must receive the same semantic value, and in particular, the same modal semantic value.

However, SVP is false. Consider, for example, Kaplan's theory of indexicals. According to Kaplan, utterances by me of "I study philosophy" and "Jason Stanley studies philosophy" express the same thing. Thus, relative to certain contexts, "I" and "Jason Stanley" have the same content. But these two expressions receive different semantic values according to the semantic theory -- they have different "characters". These differences in character do not show that the two expressions cannot, relative to some contexts, have the same content.

A denial of RT commits one to the thesis that, relative to some contexts, a rigid designator can have the same content as a non-rigid designator. Thus, a denial of RT commits one to the thesis that the distinction between rigidity and non-rigidity does not always affect content. Just as the semantic differences between "I" and "Jason Stanley" do not imply that in contexts in

which I use the expressions, they do not have the same content, so, modulo the truth of CA, the semantic differences between a rigid designator and a non-rigid designator do not imply a difference in content relative to simple sentences.

It is also worthwhile pointing out that even a "weakened" version of SVP, which states that if two sentences always have the same assertoric content, then they must have the same semantic values, is false. Consider the following counter-example. Sentences containing the terms "every" and "a" behave differently with respect to context. In particular, "a" seems to license discourse anaphora, whereas "every" does not. This is shown by the contrast between the following two discourses:

(3) A man is walking in the park. He is whistling

(4) * Every man is walking in the park. He is whistling

This contrast can be captured elegantly in Dynamic Logic, the semantic theory advanced by Groenendijk and Stokhof. In this theory, the differences between "every" and "a" are explained in terms of a difference in the type of semantic value they receive.

The details of Dynamic Semantics need not concern us here. The important point for our purposes is that the theory predicts that "∃xφx" does not receive the same semantic value as "¬∀x¬φx". However, this does not entail, and nor do Groenendijk and Stokhof believe it entails, that (5) and (6) do not have the same truth-conditions, or somehow assert different things:

(5) A man is walking in the park.

(6) Not every man is not walking in park.

---

Thus, in dynamic logic, the difference in semantic value between (5) and (6), though accounting for the interesting fact that one can follow (5) but not (6), with "He is whistling", does not entail a difference in assertoric content. The fact that (5) and (6), according to Dynamic Logic, have different semantic values, does not entail that they have different assertoric contents.\textsuperscript{28}

The explanation for the failure of weakened SVP is that semantic values are assigned in order to explain how an expression embeds in more complex constructions. However, two sentences could embed differently, yet nonetheless "say the same thing". As Dummett writes:

\begin{quote}
Someone who is able, for a given sentence, to classify specifications of possible states of affairs into those that are adequate for an assertion made by uttering it, as a complete sentence, on any given occasion, and then to classify the adequate ones into those that render it correct and those that render it incorrect, may be said to know the assertoric content of the sentence. It does not at all follow that he knows enough to determine its contribution to the assertoric content of complex sentences of which it is a subsentence. What one has to know to know that may be called its ingredient sense; and that may involve much more than its assertoric content. Ingredient sense is what semantic theories are concerned to explain.\textsuperscript{29}
\end{quote}

Semantic value is intended as an explication of what Dummett calls ingredient sense, rather than assertoric content.

Thus, SVP is false. Two expressions may have the same content, relative to some contexts, despite the existence of significant semantic differences between them. Hence, there must be something special about modal semantic value, such that any difference in it implies a

\textsuperscript{28}One particularly clear example of the failure of weakened SVP, due to Michael Dummett, is the difference between A and TA on a three-valued semantics. Utterances of these two sentences say the same thing. Yet they embed differently under negation. If A is truth-valueless, then so is ~A. But in this case, ~TA is true. See his \textit{The Logical Basis of Metaphysics} (Cambridge: Harvard University Press, 1991), p. 48. There are many other examples of failures of weakened SVP.

\textsuperscript{29}\textit{Ibid.}
difference in content. One particularly clear way of establishing this connection is via the following thesis:

The content of an occurrence of an expression is its semantic value according to a standard possible-worlds semantics.

Let us call this thesis, the modal account of content, or MAC for short. According to MAC, what is asserted by a sentence is the set of metaphysically accessible worlds in which that sentence is true, and the content of a complex expression is derived from the content of its parts, in a manner reflected by an appropriate semantical derivation.

MAC, together with MA*, entails CA. For assume MAC is correct, and consider uses of two expressions e and e'. If they are assigned the same content, then, by MAC, they express the same function from metaphysically possible worlds to truth-values. Thus, by MA*, an utterance of \( \square e = e' \) is true. But, by MAC, the semantic value of the utterance \( e = e' \) is the set of metaphysically possible worlds in which the sentence is true. Hence, the sentence \( e = e' \) expresses something necessary.

However, MAC is obviously false. No one believes that the assertoric content of an utterance of "2 + 2 = 4" is the same as the assertoric content of an utterance of "Peano Arithmetic is incomplete". The only view known to me which might be thought to support MAC is that of Robert Stalnaker. After all, Stalnaker takes what is expressed by a sentence to be a set of possible worlds. However, upon closer inspection, it can be seen that Stalnaker's theory does not actually embrace MAC. For Stalnaker believes that "both demonstrative expressions and proper names are rigid designators -- terms that refer to the same individual in all possible worlds". If Stalnaker embraced MAC, then he would first of all hold that two

co-referring rigid designators, such as "Hesperus" and "Phosphorus", have the same meaning. Thus, he would hold that:

(7) □ Hesperus is Phosphorus

is true. Yet Stalnaker believes that utterances of:

(8) Hesperus is Phosphorus

are often contingent; they express Stalnaker's "diagonal proposition".\textsuperscript{31} Thus, what is expressed by an utterance of (8) is not always --and in fact is rarely-- what it is predicted to be on a traditional modal semantics. Indeed, on Stalnaker's theory, MAC fails rather badly.

Does Stalnaker accept CA, despite his clear rejection of MAC? This is a difficult question. To answer it, we must distinguish between two possible interpretations of CA: a stronger, and a weaker one. According to the stronger interpretation of CA, if two expressions e and e' have the same content, then what is ordinarily expressed or asserted by the sentence \( e = e' \) is necessary. According to the weaker interpretation of CA, if two expressions e and e' have the same content, then the sentence \( □ e = e' \) is true.

I think it is plausible that insofar as Stalnaker would apply the term "content" to rigid designators at all, he would say that the content of a proper name is its bearer. But a typical utterance of, for example, (8), is contingent. Thus, only the second of these interpretations of CA is consistent with Stalnaker's views. On this second interpretation, an ordinary utterance of (8) can be contingent. Nonetheless, an utterance of (7) is true. Thus, the weaker reading of CA allows a theory such as Stalnaker's, where one utterance of a sentence can be necessary (say, when embedded under modal operators), but another utterance of the very same sentence contingent. Though Stalnaker does not --and cannot-- hold the stronger version of CA, he may very well hold the weaker one.

\textsuperscript{31}See his "Assertion" for details.
It is worth pausing to comment on Stalnaker's seemingly paradoxical view that, though an utterance of (8) is often contingent, (7) is nonetheless true. Surely the most natural reading of a sentence such as (7) is that it attributes necessity to ordinary utterances of (8). What is going on? Ironically, I believe that what is really behind his theory is a denial of at least the spirit of CA. What underlies endorsing the contingency of an utterance of (8), and the truth of (7), seems to be the recognition that in certain cases the metaphysical status of an utterance is irrelevant to its assertoric content. Stalnaker's treatment of (8) reveals a sensitivity to the deep connection between the notion of assertoric content, on the one hand, and informativity, on the other. But by accounting for this informativity in terms of metaphysical contingency, he seems to be conflating the very distinctions which Kripke correctly drew.

Be that as it may, though Stalnaker at least believes that a traditional modal semantics provides the right kind of meaning for an utterance --a set of possible worlds-- he still does not think that the semantic value of an utterance of any sentence is simply the set of metaphysically accessible worlds in which it is true. But most other philosophers reject MAC on different grounds. According to a typical view, for example, the correct semantic theory is one that assigns to utterances not sets of possible worlds, but rather structured propositions of one form or another. Thus, this view rejects MAC because traditional modal semantics does not provide the correct semantical paradigm for utterances. Nonetheless, on this view, CA is true.

If CA could be shown to be true, then ECP would be false. Thus, it is crucial to see whether or not an argument for CA can be provided. In the next section, we turn to more...

---


33 Of course, he is knowingly conflating these distinctions, in order to solve the problem of intentionality.
general considerations about the relation between assertoric content and modal status which could provide a basis for CA.

Section III.

In this section, I first consider an argument along different lines, attributable to Kripke, that CA follows from "ordinary language" considerations. I shall conclude that this argument is flawed and, more generally, that arguments for CA depend upon questionable assumptions about the relation between assertoric content and metaphysical possibility. Finally, I shall end with a positive suggestion about the proper place of the distinction between rigid and non-rigid expressions.

The following argument for CA can be culled from the preface to Naming and Necessity. If it is sound, the argument demonstrates the truth of CA by showing that if sentence of the form "t is t'" is not necessary, then t and t' must have different contents. Let t and t' be terms which differ in that in some metaphysically possible world w, Val(t,w) = Val(t',w). Now consider the following two simple sentences:

(9) t is t
(10) t is t'

Surely, what an utterance of (9) says is (metaphysically) necessary. But, given our assumptions, what an utterance of (10) says is not necessary. But then, by Leibniz's Law, what an utterance of (9) says and what an utterance of (10) says cannot be the same thing. Hence, the assertoric content of (9) is different from the assertoric content of (10). But (9) and (10) are simple sentences. Hence, t and t', by the definition of content in Section I, have different contents.

To evaluate this argument, consider a variation of it. Take the following two sentences:
(11) Mary is a motorcycle mechanic, but she is interested in anti-realism

(12) Mary is a motorcycle mechanic, and she is interested in anti-realism

Suppose John utters (11), and Bill utters (12). Now, what John said presupposed that motorcycle mechanics generally are not interested in anti-realism, but what Bill said did not presuppose this. Since John uttered (11) and Bill uttered (12), we thus may conclude, by Leibniz's Law, that what an utterance of (11) says and what an utterance of (12) says cannot be the same thing. Hence, the assertoric content of an utterance of (11) is different from the assertoric content of an utterance of (12).

However, for those who accept talk of presupposition, utterances of (11) do have the same assertoric content, or truth-conditions, as utterances of (12). The fact that different utterances give rise to different presuppositions does not imply a difference in truth-conditions. A presupposition theorist would not (and should not) be swayed in her conviction by the fact that English speakers regularly use such locutions as "what John said presupposes". Clearly, only a theoretical argument to the effect that no principled distinction between what is presupposed and what is asserted can be drawn would threaten the presupposition theorist's position.

More generally, once one begins to take ordinary uses of "what is said" at face value, no two distinct terms will have the same content. For instance, suppose John utters:

(13) Bachelors are unmarried men

Suppose that Bill does not know the meaning of the word "bachelor". A perfectly ordinary way of describing what occurred would be to say that what John said was news to Bill. Yet of course what (14) says would not be news to Bill:

(14) Bachelors are bachelors

Thus, since what (13) says is news to Bill, and since what (14) says is not news to Bill, we may conclude, by Leibniz's Law, that what (13) says and what (14) says are different; that is,
that they have different assertoric contents. Since (14) results from (13) by substituting the term "bachelors" for "unmarried men", the two terms hence have different contents.

Obviously, something is wrong with this sort of argument. The illegitimate step is the last one. Bill is not a competent user of the word "bachelor". Hence, whether an utterance containing that word is news to Bill or not has nothing to do with its assertoric content. Similarly, Kripke's argument is something of a non sequitur. What needs to be established is that the modal status of an utterance is relevant to its assertoric content. That is, what needs to be shown is that (metaphysical) necessity and contingency are properties of the assertoric content of an utterance. For it is only then that we can use Leibniz's Law to conclude that a difference in modal status entails a difference in assertoric content. But Kripke's argument assumes, rather than argues, for this thesis.

None of this would be news to Kripke. Kripke himself does not take his ordinary language argument as decisive, but rather relegates it to the status of "indirect evidence". The only reason I have dwelt upon it is because it seems to be the only argument advanced by Kripke which, if true, would demonstrate the truth of CA, and hence RT. Another consideration raised by Kripke which could be marshalled in support of RT is, in his words, "that we have a direct intuition of the rigidity of names, exhibited in our understanding of the truth-conditions of

---

34 Kripke's original argument proceeds by considering propositional anaphora, signaled by the use of "that". However, "that", if anything, is even looser than "what is said". Consider, for instance, the naturalness of such locutions as "That's news to me".

35 "...[V]arious secondary phenomena, about 'what we would say', such as the ones I mention in the monograph and others, give indirect evidence of rigidity", Naming and Necessity, p. 23.
particular sentences"). It would be difficult to account for this intuition, unless we suppose that the rigidity of a term affects what we understand (i.e. the assertoric content).

Kripke's appeal to intuition here raises difficult questions about the role of intuition in philosophical theorizing. There is no question that certain kinds of intuitions are crucial in philosophy. Examples of these include a speaker's intuitions about different possible interpretations of a given sentence, or, in the case of Ethics, clear intuitions about the rightness of a given act. However, Kripke's intuition is not an instance of either of these kinds.

The notion of truth-condition is one which itself requires a philosophical explication. Thus, the thesis that rigidity is relevant for individuating the truth-conditions of unmodalized sentences is a philosophical thesis. Now, a philosophical thesis can be justified by appealing to "lower-level" intuitions which the philosophical thesis helps explain. For instance, a philosophical thesis about content gains support if it accounts for "lower-level" intuitions of the form: an utterance u of a sentence S says something different from an utterance u' of a sentence S'. But Kripke's intuition is not a "lower-level" intuition. Rather, it is a "direct" intuition about the truth of a substantive philosophical claim.

Some might maintain that this fact automatically shows the illegitimacy of Kripke's appeal to intuition. However, I think it is incorrect to discount the intuition on these grounds alone. The correct response is rather to grant that there is some intuition which underlies Kripke's remarks, but to deny that it has the consequences which Kripke believes it to have. What

---

36 Naming and Necessity, p. 14. This consideration is advanced more in support of claims other than CA and RT.

37 Of course, these lower-level intuitions are, as I have been emphasizing throughout, nonetheless quite defeasible.

38 See, for instance, Michael Dummett, The Interpretation of Frege's Philosophy (Cambridge: Harvard University Press, 1981)], pp. 579 ff., which are part of the appendix on Kripke.
should be granted to Kripke is that a grasp of rigidity is required for some (perhaps quite important) uses of proper names. But if one wishes to use the intuition in support of CT, then one must provide an argument to the effect that the relevant use is the use of proper names in the speech-act of assertion.

One such argument might come from a consideration of the notion of a truth-condition. For surely, it might be argued, the notion of truth-condition is a modal notion. However it is explicated, it should fall out that knowing the truth-condition of a sentence implies knowing in what possible circumstances the sentence is true. Since the sentences:

(15) Aristotle was fond of dogs
(16) The last great philosopher of antiquity was fond of dogs
differ in truth-value in some metaphysically accessible world, then surely their truth-conditions are different.

The problem with this argument is that it presupposes that the notion of possibility which underlies our talk of truth-conditions is metaphysical possibility. But metaphysical possibility is only one of a plethora of different notions of possibility. This variety is clearly reflected in ordinary uses of modal terms. Indeed, if Angelika Kratzer is correct, which of these notions is expressed by an utterance of a modal term is hopelessly context-dependent. Given these diverse senses, what argument exists for the thesis that it is metaphysical modality which grounds our talk of truth-conditions?

The question becomes even more pressing, given that (15) and (16) would have the same truth-conditions on a number of different senses of possibility. For instance, if it is common knowledge that Aristotle is the last great philosopher of antiquity, then (15) and (16) will have

---

the same truth-conditions, if we construe the possibility in question as epistemic. More plausibly, however, the possibility in question is neither epistemic nor metaphysical possibility. Our assertions are made with a background of shared presuppositions, the most central of which are presuppositions about the meanings of our words. If the meaning of an expression is characterized as a set of core beliefs about the semantic properties of the expression, then perhaps the only possible circumstances relevant for determining the truth-conditions of the utterance are those in which these beliefs are true. On this conception of truth-condition, it is not possibility which allows us to individuate the content of terms, but rather content which allows us to say which possibilities are relevant.

I do not mean to imply that some version of the truth-condition argument can not be defended. However, simply stating that the notion of possibility in question is metaphysical does not constitute such a defense. What a friend of this argument must establish is that circumstances which are metaphysically possible, but not possible in any other sense, are involved in the individuation of the content of assertions. However, I am skeptical of the success of such an argument for the following reasons.

First of all, there is a well-supported notion of truth-condition according to which not every metaphysically possible circumstance is relevant for individuating assertoric content. Consider again the relation between "The F" and "The actual F". As we saw in Section I, utterances of simple instances of the schema, 'The F is G' and 'The actual F is G' seem to have the same assertoric content. Why do we seem to have the intuition that utterances of

---

40 If one takes assertoric content to be characterizable by sets of possible worlds, then one must, of course, also establish that any possible circumstance involved in individuating the content of an assertion is also a metaphorically possible circumstance.

41 A similar point holds for temporal rigidity, and the relation between the Noun Phrases, "the current President" and "the President". Consider:
(a) In the year 2001, the President will be Colin Powell.
such sentences say the same thing, despite the fact that they typically differ in truth-value in some metaphysically accessible world? One explanation is that the link between modality and truth-conditions only implies that sameness of truth-conditions entails that in whatever possible situation in which the words have the meaning they actually do, whenever one of these is uttered truly, the other could have been uttered truly instead. That is, if two utterances u and u' have the same assertoric content, then in any possible circumstance in which u is uttered, u' could have been uttered, without a change in truth-value.

Notice that "any possible circumstance" cannot be construed as any *metaphysically* possible circumstance. For there are metaphysically possible circumstances in which the meanings of the words differ. If this is indeed the constraint which modal notions place on sameness of truth-conditions, then only those possibilities are relevant in which the words have the meaning they actually do.\(^{42}\)

The second difficulty facing the defender of the truth-condition argument is that, even if the correct explication of truth-condition is not that just suggested, it is still difficult to see why every metaphysical possibility should be relevant for individuating assertoric content. When (15) is uttered (in a non-modal context), why should every esoteric metaphysical possibility be relevant in accounting for successful communication between competent speakers? If understanding an utterance of (15) requires being able to distinguish between

(b) In the year 2001, the current President will be Colin Powell.
Though "the current President" and "the president" have the same content (here, considering *temporally* simple sentences), (a) clearly says something different from (b).

\(^{42}\)Places in which it is argued that this is the correct modal constraint on truth-conditions include Dummett's appendix on Kripke in *The Interpretation of Frege's Philosophy* (see esp. p. 565), Evans's "Reference and Contingency" (see esp. p. 207), and, most explicitly, Martin Davies and Lloyd Humberstone, "Two Notions of Necessity", *Philosophical Studies* 38 (1980), pp. 1-30 (see esp. pp. 16-17).
possibilities, then surely only those possibilities that are "open" in some epistemic sense, are relevant.

That the argument from truth-conditions is often appealed to in support of CT is a touch ironic. For a central moral of Kripke's work is that many philosophical arguments, such as that for the contingency of identity, fail because they equivocate between different senses of the term "possible". Stemming, as it does, from pre-Kripkean times, it is difficult to avoid the suspicion that the alleged relation between the content of an assertion and metaphysical possibility is supported by precisely the type of equivocation between senses of possibility which Kripke warned us to avoid.\footnote{Ludwig Wittgenstein's \textit{Tractatus Logico-Philosophicus} (New York: Routledge, 1988) is generally thought to contain the first formulation of the truth-condition argument. Yet it is obvious from 4.464 of the \textit{Tractatus} (p. 99) that Wittgenstein fails to distinguish between epistemic and metaphysical necessity. Carnap, the other famous pre-Kripkean proponent of the truth-condition argument, of course completely rejects an independent category of metaphysical necessity.}

As we have seen, the ultimate motivation for RT is the idea that a necessary condition for two utterances to have the same assertoric content, or truth-conditions, is that they have the same truth-value when evaluated with respect to every metaphysically possible world. This is a significant, non-obvious thesis about the identity-conditions of assertoric contents. It thus requires either some independent justification, or the production of a philosophically interesting notion of content according to which it \textit{is} obvious. It remains to be seen whether either of these tasks can be accomplished.

Before we conclude, I would like to make one positive, albeit quite tentative, suggestion concerning the proper place of the distinction between rigid and non-rigid expressions. If, as I have suggested, modal properties do not hold of the contents of assertions, from what do the
modal properties of an utterance arise? Necessity and contingency are not merely properties which hold fundamentally of utterances. Rather, they stem from some non-syntactic source.

I believe it fruitful to locate semantic differences between expressions which do not always affect the content of assertions in the use of those expressions in other speech acts. In accordance with this, my suggestion is that the modal semantic value of an expression arises, not from its use in assertions, but rather fundamentally from its use in the speech act of counterfactual supposition. Our assertions are made with a background of shared presuppositions, some of which serve to fix the references of our expressions. A central purpose of supposition is to suppress these often quite fundamental presuppositions, and to imagine what the world would be like if they were false. Rigidity allows us to speak of the denotations of our expressions in counterfactual situations in which our shared assumptions do not hold. But this does not show that rigidity affects the content of ordinary assertions, where competent speakers may always employ these presuppositions.

Conclusion

Our discussion has left several questions open. Most markedly, we have not attempted a positive account of the relation between the intuitive notion of assertoric content and semantic value. The relation between these notions raises several issues. For instance, "The President is the president" and "The actual President is the president" have the same assertoric content, but the assertoric content of the two sentences that result from embedding them respectively under "It is (metaphysically) possible that" do not. Yet to say that two sentences can share the same assertoric content, but contribute different things to the assertoric content of sentences containing them, is simply to deny that there is a function which maps the assertoric content of the parts onto the assertoric content of the whole. Thus, the intuitive notion of assertoric
content is not compositional. Attempts to represent the assertoric content of an utterance as a semantic value, which I believe to be a central motivation behind Stalnaker's paper, "Assertion", are open to the kind of critique discussed in Section II, precisely because it seems that compositionality is a principle governing semantic values *tout court*. Whether we should follow Dummett in denying that the intuitive notion of assertoric content is to be directly accounted for by a compositional semantic theory, or follow Stalnaker in altering semantic theory to allow for a more direct expression of this notion, is unclear.44

It is sometimes maintained that, from a certain non-epistemological perspective, the differences between utterances of the sentences "Hesperus is Phosphorus", and "Hesperus is Hesperus" are unimportant.45 What I have been emphasizing in this paper is that from a perspective motivated by the attempt to give an account of assertoric content which has some relation to use, the differences between rigid and non-rigid terms are unimportant. Since there is no convincing argument to the contrary, I must conclude that the idea that rigidity provides

---

44On pp. 20-21 of *Naming and Necessity*, Kripke gives voice to a similar pessimism about representing the objects of belief as semantic values as Dummett does about so representing the objects of assertion: My view that the English sentence 'Hesperus is Phosphorus' could sometimes be used to raise an empirical issue while 'Hesperus is Hesperus' could not shows that I do not treat the sentences as completely interchangeable. Further, it indicates that the mode of fixing the reference is relevant to our epistemic attitude toward the sentences expressed. How this relates to the question what 'propositions' are expressed by these sentences, whether these 'propositions' are objects of knowledge and belief, and in general, how to treat names in epistemic contexts, are vexing questions. I have no 'official doctrine' concerning them, and in fact I am unsure that the apparatus of 'propositions' does not break down in this area.

45In particular, by advocates of "Russellian Singular propositions". I do not mean to endorse this position here.
some deep insight into the content of an expression remains an unsupported dogma of the philosophy of language.\textsuperscript{46}

\footnotesize
\textsuperscript{46}I am especially indebted to Richard Cartwright, Richard Heck, and Robert Stalnaker. Without their input, the paper simply could not have been written. Kathrin Koslicki, Sanford Shieh, and Timothy Williamson also merit special thanks. In addition, I have benefited from discussions with George Boolos, Noam Chomsky, Lenny Clapp, Michael Glanzberg, Joe Lau, Peter Ludlow, and Daniel Stoljar.