8611324

Azzouni, Jawad N.

LANGUAGES CONTAINING THEIR OWN TRUTH AND FALSITY PREDICATES:
A NEW APPROACH

City University of New York PH.D. 1986

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

LANGUAGES CONTAINING THEIR OWN TRUTH AND FALSITY PREDICATES:

A NEW APPROACH

by

Jawad Azzouni

A dissertation submitted to the Graduate Faculty
in Philosophy in partial fulfillment of the
requirements for the degree of Doctor of Philosophy,
The City University of New York.

1986

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

1-6-86
date

Chairman of Examining Committee

1-6-86
date

Executive Officer

Melvin Fitting

Jerrold J. Katz

Arnold Koslow

Richard Mendelson
Supervisory Committee

The City University of New York

Abstract

LANGUAGES CONTAINING THEIR OWN TRUTH AND FALSITY PREDICATES

by

Jody Azzouni

Advisor: Professor Arnold Koslow

In the first chapter, I do two things. I discuss general philosophical motivations for studying languages which contain their own truth predicates: I discuss the history of the subject to some extent, starting with Tarski and examining briefly the work of Gupta, Herzberger, and Kripke. I also give an overview of the system itself, that is, I describe the language, the model theory and its axioms, and discuss some of the motivations, philosophical and technical for its properties.

The second chapter gives the model theory. I also give several "existence" theorems, that is, I show that certain kinds of desirable models exist so that the system is not any more restrictive than, say, standard model theory.

The third chapter gives the axiomatization for the model theory. Intuitively, it might come as somewhat of a surprise that languages with truth and falsity predicates are axiomatizable, but it does not follow, in the broad way this is understood, that languages with their own truth and falsity predicates are languages rich enough to do their own truth theory or syntax. Such special languages within the broader structure I have defined may be singled out model-theoretically or axiomatically and studied further.

I also give completeness and consistency proofs for the axioms in terms of the model theory of the second chapter. The completeness proof is an unusual variant of standard Henkin completeness proofs.

The last chapter describes a way of marking out the problematic sentences which have given rise to the literature on the paradoxes and languages with their own truth and falsity predicates in the first place. Such sentences are usually presented informally, but there is no way of recognizing them in a significant intertheoretic way. Since such sentences are problematical from a classical point of view I attempt to recognize them in a classical setting. The tools used here more closely resemble recent other work in the literature than anything earlier in the dissertation.

TABLE OF CONTENTS

CHAPTER 1

I

Russell's paradox did not arise in a void. Set theory had been developing for some time already; and a program to reduce mathematics to set theory and logic was already in progress. This program put strong constraints on solutions and also made the set theoretic paradox more than a curiousity. Any solution to Russell's paradox had to be placed within a global context in which the resulting set theory would be judged.[1]

Clearly, what are known as the semantic paradoxes did arise in something like a void. Until Tarski, semantic theory did not exist.[2] And it was born simultaneously with a solution to the Liar's Paradox and consequently could not act as the global factor within which solutions to the paradox would be evaluated. By this, I mean, of course, that Tarski's semantic theory came equipped with a way of handling the paradoxes which is fairly intrinsic to it and consequently could not be used as a background subject matter in which to place various solutions to the paradox.

These factors affected crucially the histories of the two kinds of paradoxes. Whereas Set Theory has settled fairly comfortably into one or two standard approaches (despite Quine's suggestions to the contrary), Semantic Theory has not. Much of this, certainly, has to do with issues such as intensions which are unconnected with the paradoxes. But a glance at the literature will show that much ink has spilled over solutions to the Liar's Paradox which are not set in the context of a general semantic theory, or at least are not evaluated in terms of what kinds of theories they would allow[3].

At this point I would like to describe a program associated with early work in Set Theory and Logic. This is Logicism. Now there seem to be good reasons why Logicism failed as a program.[4] I am going to suggest that there is a sense in which it has succeeded, although in a terribly half-hearted way. Mathematics up until the initiation of the program of Logicism had been carried out naively in natural languages. However, once the program is undertaken, there are two possibilities: 1) either the underlying semantic and syntactic structure of natural languages _is_ that of the Standard Predicate Calculus with Tarskian Semantics (in which case the structure need only be made explicit when difficulties due to the surface structure of natural languages obscure matters) or, 2) the Standard Predicate Calculus is to replace natural languages which are discarded as the appropriate languages for scientific work (in this case, the Standard Predicate Calculus functions as a kind of modern Latin).

Let me make some observations about these cases before going on. First, I don't think anyone at the time held either position I describe. Rather, a logical language was regarded as an ideal language and although some of its rules might be approximated by natural languages, they themselves weren't regarded as having an "exact logic".[5] That is, it was felt that there is no fact of the matter about what logical rules a natural language has. I think there is as little reason to believe this as there is that a natural language has no "exact syntax", although my hindsight has the benefit of Chomsky. Indeed, the former idea depends somewhat on the latter (at least historically), since the logical rules of a formal language were initially defined only syntactically. Semantics was informal until Tarski. What may have also motivated their view was that for a long time only a small number of formal languages were known, and it seemed clear none of these could model the properties natural languages seemed to have. The doctrine matured long before the rich proliferation of nonstandard formal languages which seem closer in spirit to natural languages. It is also possible that the epistemological problems facing a theorist of natural languages that I discuss below contributed support to this doctrine. What is actually hard was taken to be impossible.

Secondly, I should add that the first possibility is fairly unbelievable. Clearly, not <u>all</u> the resources of a natural language seem to be captured by the Standard Predicate Calculus. Terms like "Necessarily" and

"Possibly" occur with alarming frequency. So do counterfactuals. The connectives rarely seem truth-functional. And, of course, "true" does not seem to function in ways which are compatible with avoiding inconsistency, if the language is to be treated in a standard way. Of course these things can be explained away in various fashions. But, it is important to realize that the question of what formal language best captures the properties of natural languages is quite different from the question of which language (formal or otherwise) is the best for our purposes. Both questions are difficult. But the first is an empirical question of science and the second is fairly normative. They must not be confused. Separating them, I think, will make it clear that natural languages are rather different from the Standard Predicate Calculus (Considerations of the sort Kripke presents in NAN and OOTOT are pertinent here). However, these considerations do not exclude the possibility that a sufficiently restricted portion of the sentences of a natural language (say, those of Mathematics) have the underlying structure of the Standard Predicate Calculus. I will say something about this shortly.

The second case, on the other hand, may seem plausible. If true, it would entail that a large scale translation must be undertaken of Mathematics in natural languages to Mathematics in the Standard Predicate Calculus (presumably nothing will be lost: and possibly something gained--the former shown by the actual work in reducing classical Mathematics to Set Theory and Logic and the latter shown by the emergence of Metamathematics--the subject of which turns crucially on the syntax and semantics of the Standard Predicate Calculus, not on those of natural languages). But such a translation of mathematical statements from a natural language to a formal language is not easily evaluated, since natural languages are a natural phenomena. Consequently, no one can argue convincingly that their semantic structure is given intuitively or is obvious (any more so than their syntax).

Thus we are faced with a problem that clearly is not faced with artificial languages. Any semantic theory of natural languages is truly a theory in that it could be revised. In the sense similar to that in any science, we could be wrong about a natural language's semantics. Whatever scientific criteria there are for evaluating theories must be applied here as well. The labs (metaphorically speaking) in which experiments in semantics are carried out are artificial languages whose semantic "theories" are to be understood as given and transparent to us[6].

It might be thought that these considerations can be used to show that to leave the practice of mathematics in natural languages is to leave a largely empirical element in mathematics where it does not belong[7]. Mathematics must always be carried out in artificial languages, where the semantics and syntax are known. But at present, artificial languages themselves are studied via natural language metalanguages. This may supply a motivation for finding artificial languages which can serve as their own metalanguages or at least contain their own semantic theory.

Now I want to point out why in fact neither possibility is reasonable, even if we restrict ourselves to that part of natural languages which are used for mathematics. The problem is with the adoption of Tarskian Semantics for the uninterpreted language of the Standard Predicate Calculus. It is that formal languages strong enough to replace natural languages for scientific purposes do not contain their own semantic theory. This is not to say that they cannot: with appropriate restrictions they may[8]; but it is to say that in the standard case, they do not. I am alluding here to what is called the Tarskian Hierarchy and I will suggest that Logicism falls afoul of it.

Here is the argument. Consider a first order axiomatization of Number Theory and consider the proof that it is incomplete. This proof occurs in the metalanguage and involves the construction of a sentence of Number Theory which is true in the standard model for Arithmetic but not implied by the axioms. Now consider a set of axioms in the metalanguage gotten by adding to the axioms utilized in the proof of the Godel sentence those of a translation into the metalanguage of the original axioms of Number Theory. We can show (in a meta-meta-language) the incompleteness of this set of axioms, and so on. What, I ask, is the force of this "and so on"? Any mathematician, I submit, understands what I have just written. It is, suitably presented, a mathematical truth. But what language was it written in? Surely not one in the sequence of meta-languages I have just alluded to. At best (and not very believably), our mathematics is scattered up an indeterminate Tarskian hierarchy, and this without any apparent rigor about where we are located at a particular time. In point of fact, the whole practice goes on in natural languages and some are quite explicit about it[9]. Latter day mathematicians also use set theory naively. But set theorists are comfortable with this for they could easily reconstruct what is being done and used according to the lights of one or the other acceptable set theory. It would not be easy at all to reconstruct a mathematician's writing in formal Tarskian languages, for natural languages, even when restricted to mathematical practice, cannot be construed along the lines Tarski laid down for formal languages for the reason I have just suggested (e.g., they seem to take a viewpoint of mathematical results that lies outside of any hierarchy). To sum up: Mathematics is naively carried out in natural languages which seem not to have a hierarchy. This precludes the first possibility. On the other hand, no adequate translation of some of the most important mathematical statements into one or another sentence in a hierarchy of Tarskian languages is adequate. This precludes the second possibility. However, the success of Logicism lies in the fact that to the extent that the semantics of the language(s) of mathematics is taken consideration of in mathematics it is implicitly assumed to be Tarskian.[10]

These considerations motivate the search for formal languages which can contain their own semantic theory[11]. This is for two reasons: one is that one would like to show, if possible, that naive mathematical practice in natural languages is not dangerous. The other is that such formal languages are far better candidates than ones with Tarskian semantics to play the analogous role that Set Theory plays to naive mathematical practice with sets.

II

A glance at the literature in this area shows that the primary concern is not about languages which contain their own semantic theory but about languages which contain their own truth and falsity predicates. In particular, the concern (in the fairly recent literature) has been with languages which contain their own truth and falsity predicates and are, as Feferman puts it, type-free[1].

It may not seem at first glance, that this literature has much connection with the motivations I have suggested in (I).[2] We must ask two questions. First, why is it so important to construct a theory of truth for an artificial language, and second, why is it so important that the language itself contain this theory?

Well, why did Tarski think the first question was so important? Certainly, we know that he regarded the formalization of the semantic theory of a formal language important. But this involves more than the theory of truth. Indeed, the important notion is not "truth" but "satisfaction". And in formalizing this notion, we learn a great deal about the object language. For example, the fact that a recursive definition of satisfaction is forthcoming in certain cases tells us that the sentences are constructed in such a way that their semantic properties are interconnected in a strong sense.[3] We learn how the individual variables play a crucial and surprising role in how sentences are assigned truth values (what Quine calls "objectual quantification"), and so on. Indeed, all of this (and there is more) is much more interesting and informative about the object language than the fact that a predicate "true" can be defined, unless the notion of truth defined has independent philosophical interest. I would say something stronger: the focus on Criterion T and the fact that a definition of truth in accord with it is forthcoming looks trivial in comparison to the other aspects of the semantic theory of a language as Tarski sets it up.

Why, then, the focus on "truth" by Tarski himself? I conjecture that Tarski thought that in defining "truth" he had achieved something of philosophical importance. The importance lay in that he had given a precise definition (in a metalanguage) to the notion as it was understood according to "the correspondence theory of truth".[4] However, its philosophical importance is mitigated by the fact that what Tarksi thought he meant by it is stated as Criterion T.[5] The criterion is a trivial relation between the sentences of a metalanguage and the sentences of an object language that no one could comfortably deny, and the correspondence theory looks like it should be a substantive philosophical theory (to be contrasted with, say a coherence view, a conceptualist view, etc.,). Clearly, we should not confuse the correspondence theory, understood as

Criterion T, with more substantial philosophical theories (such as Realism, Idealism, etc). Tarski writes "Thus we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians--whatever we were before. The semantic conception is completely neutral towards all these issues."[6]

This, it must be said, empties the notion of truth of a great deal of philosophical content. If, then, "truth" is important solely as a device for semantic ascent, the only condition on the truth predicate is Criterion T (which is satisfied by certain trivial theories).[7] Thus, the notion, even from a semantic point of view is less significant than other notions (like satisfaction), which capture more aspects of the object language than it does, especially when, as in many formalized languages, it is divorced from them.

It does not follow of course, that the aims of Tarski need be shared by other writers in this area. For one thing, Tarski despaired of ever supplying a semantical theory for natural languages because, in particular, natural languages seemed to have no devices for avoiding liar's paradoxes.[8] Therefore, some workers have tried to show that natural languages are not as bad off as Tarski suggested with regard to the truth predicate.[9] Unfortunately, the problems that self-reference causes in the truth predicate is not restricted to it and therefore the problems can reemerge with other naturally defined predicates given the new semantics of the formal language.[10] Therefore, this motivation does not suffice for restricting one's studies of type-free languages to those containing the truth predicate alone. The truth predicate is only the most sensational member of a class of semantic terms. If it is important to study languages which contain a self-referential truth predicate, in order to show that natural languages are not inconsistent, it is equally important to study languages which contain their own semantic theory (that is, all the semantic terminology that is used to define the semantic theory of an object language).[11]

III

I want to examine briefly what it means for a language to contain its own semantic theory. It does so relative to a model. That is, a model: a domain of objects (including the sentences of the language), and a set of relations defined on these objects, is given. Further, a mapping from the constants of the language and the n-place predicate variables to the objects in the domain and the n-place relations defined on the model is given. If Tarski is to be a guide, then this is to suffice for the definition of the relations pertinent to semantic theory[1]. In particular note that a relation for the predicate "satisfaction" is not given. This is to be supplied either axiomatically or via a definition from the other predicates. Now of course if we succeed in doing this, there is still a great deal that might be expressible in a metalanguage although not in the object language. We have not, that is, eliminated the need for a metalanguage. For example, in general, not very much Set Theory will need to be expressed by the object language. So, the object language is understood against a background in full Set Theory (i.e., that is where the definition of a model comes from). Nonetheless, if a language could be constructed which could do this much for itself then it would be, in a very clear sense, a language which could do its own semantics.

By contrast consider a language which could do its own (Tarskian) model theory (or theory of validity as I call it below). Here it would be required to define, in the language, the notion of a model. Thus such a language would be required to contain some Set Theory[2]. The primary notion in this case would presumably be not satisfaction but satisfaction on a model.

All this is clear enough (although it is not possible to carry out). But when we turn to other approaches, matters are slightly more complicated. Consider Kripke's approach. What does a language which contains its own semantics mean in that case (as opposed to to one which contains its own validity)? Two possibilities present themselves. 1) The notion of a fixed point model is given, and the language contains the notions needed to define its own satisfaction. However, in this case, there is something empty in the requirement that the language contain the notions needed to define its own satisfaction. This is simply because _a fixed point language comes equipped with extensions for its truth and falsity predicates_ (or its satisfaction predicate), something that is clearly not the case above. It may be felt intuitively that the extension of the truth predicate should not be given in the model but only the extensions of the nonlogical predicates and individual constants (as in Tarski's case). This motivates the second possibility: 2) A model is given, but without extensions for the truth and falsity predicates. As before the language must have whatever resources needed to define a satisfaction (and nonsatisfaction) predicate (including, in this case, a bit of set theory).

Of course, which possibility we accept in cases where the extension of a truth predicate is the result either of an inductive construction or a set-theoretic definition depends on how we view the construction philosophically. If we take a Platonistic view of the extension of the truth predicate, it may be reasonable for us to take it for granted with the rest of the model; otherwise not. I will pursue this matter later in regard to my own approach.

Let me say something further about gappy approaches. Let us take Kripke's theory as an example. He points out that in his construction, instead of adding a truth predicate, one could just as well add a satisfaction predicate[3] and go through the construction. This is not always an example of a language with its own semantic theory on either understanding of a language containing its own semantic theory. For a central part of this approach is the fact that the language is gappy (I am leaving the closed off languages aside) and this cannot be expressed in every approach.[4] Where it is possible to add to the language a predicate "Gap" which holds of the truth-valueless sentences, the language has a better chance at the status of one which contains its own semantic theory (in the weak first sense).

IV

Let us, having considered the caveats above, restrict our attention to languages which contain their own truth and falsity predicates. It should be pointed out immediately that it is not merely the Liar's Paradox which offers technical obstacles to carrying out a project of this sort. Consider, for example, a sentence which says of itself that it is true, what is called a truth teller. If one thinks only for a few moments about Standard Model Theory as it was developed by Tarski, one sees that it is hard to accomodate sentences of this sort. Their striking characteristic is that it is perfectly arbitrary what truth value they are assigned, regardless of "how the world is."[1] That is, they violate our intuition that truth is redundant: that by specifying the truth values of all the sentences in which the truth predicate does not occur, we have described enough to fully determine the extension of the truth predicate. Thus, even if there were no Liar's Paradox, the natural move of defining the extension of a truth predicate in a model as the set of sentences satisfied in that model is blocked by the fact that the definition would fail to specify where the truth tellers belong. Kripke[2], among others, has observed that certain sentences are liar's paradoxes only in certain models. From this he draws the conclusion that an intrinsic criterion--one depending on the syntactic or semantic properties of the sentence alone--for a paradox is not forthcoming. No more so, and for exactly the same reason, is it forthcoming with truth tellers. This makes the question of how to modify the above natural definition of the extension of the truth predicate to handle truth tellers momentarily puzzling.

Gupta[3] has pointed out that the standard formalization of Tarski has the drawback of requiring that we know the extension of the truth predicate before we apply the formal definition to a language containing the predicate. This is just another way of alluding to the problem of the truth tellers: no difficulty need arise where the presence or absence of a sentence in the extension of the truth predicate turns ultimately on the application of the satisfaction clauses for sentences in which no truth predicates appear. It only does so when the presence or absence of a sentence in the extension of the truth predicate turns directly or indirectly on itself or leads to an infinite regress. I did not describe the drawback as a way of alluding to the Liar's Paradox because the problem with the latter is that the formal definition breaks down altogether. It is this, I presume, that accounts for the large literature on the Liar's Paradox as compared with the small literature on truth tellers. We are more upset at the fact that the formal definition breaks down than we are at the fact that it seems to fail to apply to certain sentences. Of course, even if there were no liar's paradoxes, we might still be puzzled as to how to modify Tarski's definition.

By way of motivating my solution to these problems, let me discuss certain aspects of previous solutions and the intuitions that motivated them. The approaches exemplified by Gupta, Herzberger, Kripke and Tarski all utilize a hierarchy of interpreted languages. This, of course, is clearest in Tarski and contributes to the unease his theory creates, especially since the hierarchy cannot (as it is standardly presented) be dispensed with. Kripke constructs the desirable minimal fixed point language by ascending through an ordinally indexed sequence of interpreted languages where the extension of the truth predicate in each language is a superset of the extension of the truth predicates earlier in the sequence. Similarly, Gupta and Herzberger start with some arbitrary extension for the truth predicate in a language, which is then revised through an ordinally indexed sequence of interpreted languages, until a collection of interpreted languages is reached, any of which will supply "the best candidate" for the extension of the truth predicate.[4]

Now, it seems to be a motivation of Gupta, Herzberger, and Kripke, to shed light on the concept of truth as it is utilized in natural languages. The question then naturally arises: how much are these constructions merely a technical tour de force designed to provide interpreted languages with highly desirable properties, and how much are they truly essential to the notion of truth in natural languages? Let us take first the Gupta/Herzberger construction. Does it imply that the truth predicate differs in some fundamental way from other predicates? In a certain sense, yes. The extension of the predicate cannot be defined axiomatically, given a standard kind of model theory. The desirable models with appropriate extensions must be constructed. However, this may reflect more on the descriptive powers of Logic than it does on the truth predicate (i.e., 'true' is not a first order concept). Consider, for contrasting sentiments on his approach, the following passage from Gupta[5]:

> "I propose...that we view the concept of truth as characterized by a revision procedure. I suggest that truth, unlike ordinary concepts such as red, blue, and sum, does not in general have an application procedure associated with it. Idealizing somewhat, we can say that underlying our use of words such as 'red' is an application procedure that divides objects into two classes: those objects to which the word applies and those to which it does not apply. The same holds for words such as 'sum'. Underlying them there is also an application procedure, though a logically more complex one: the procedure is recursive. In contrast, I am suggesting that underlying our use of 'true' there is not an application procedure but a revision procedure instead. When we learn the meaning of 'true' what we learn is a rule that enables us to improve on a proposed candidate for the extension of truth."

Now when Gupta says that an application procedure divides objects into two classes, clearly he has a theoretical division in mind. In practice, there are many objects which cannot be sorted out into the two classes since we cannot "get at them" to apply the application procedure. Also, there are the well-known problems with vagueness. But we assume there is a fact of the matter: any object is either in the extension of a predicate or its anti-extension. Now a possible motivation[6] for Gupta's treating the predicate 'true' differently is that the truth tellers cannot be sorted out in a satisfactory way. Any procedure that we specify will either fail to yield an answer when applied to them or will clearly do so in an arbitrary way. Another way of putting this is that the two classes mentioned in Gupta's passage are ill-defined when the predicate in question is 'true'. But put this way, the matter looks potentially trivial. For nearly every interesting term with an application procedure, there are cases where the application procedure fails. Quine, in a related context[7] has suggested that adopting bivalence saddles us with undecidables. Why can't we imagine that exactly the same thing occurs with sentences and the predicate 'true'. The application procedure in this case is eliminating instances of "true" in a sentence by use of Tarski biconditionals and it fails in some cases.

But this would overlook how important the technical construction Gupta uses is in motivating his view of the truth predicate. The important fact is that the construction involves an ordinally indexed sequence of interpreted languages which differ only in the extensions of the truth and falsity predicates. If the procedure

Gupta describes is essential to produce interpreted languages which have appealing extensions for the truth and falsity predicates, one might feel that we really fail to have an <u>application</u> procedure because the procedure carried out to construct the extension of the truth predicate involves changing it stage by stage; since all the other predicates are given outright in extension at the beginning of the construction, the procedures we use to determine their membership may be viewed as application procedures. If this is Gupta's motivation then it might seem to apply equally well to the procedure used by Kripke to construct the minimal fixed point. However, in the latter case, we can view the ordinally indexed sequence of languages as approximations to the fixed point language in question, and thus we can feel that we have an application procedure since the extensions of the truth predicates in the approximating languages are cumulative. In the Gupta/Herzberger case, this is not always so. As a result, in general, there is no unique best candidate in their approaches so that it is hard to view the technique as one using approximations[8].

Herzberger does not take this view of his very similar construction. Rather, it serves a double role. First, a diagnostic one:

> "...paradoxes are allowed to arise freely and to work their own way out. No semantic defenses are to be set up against them. No non-classical logic is to be introduced; no truth-value gaps, no indexed predicates,

> no impossible worlds are to be brought into the account. No effort will be made to eliminate the paradoxes, to suppress them, or in any way to interfere and take deliberate action against them. They are to unfold according to their own inner principles."[9]

This is certainly unobjectionable. But it is not clear to me that it motivates his particular technical approach over others that might be imagined. Why is a semi-inductive construction which forces every sentence to have one and only one truth value at every stage the natural habitat of the paradoxes? Herzberger says: "...one doesn't want to postulate Tarski's schema T as a basic assertion of naive semantics. That would render naive semantics itself an inconsistent and false theory."[10]

This remark is puzzling. I should say first that our understanding of natural languages in the pertinent literature is through the resources of logic. Thus "naive semantics" (understood colloquially) is precisely that utilizing the Tarski biconditionals. Doing so <u>does</u> saddle us with sentences with both truth values. To separate the reasoning that leads to this into separate stages (separated by jumps) so that a dynamic structure is achieved where at each stage every sentence has one and only one truth value seems artificial. At least, it does not seem to be the <u>natural</u> habitat of the paradoxes. The paradoxes are problematical. This is because a naive application of the Tarski biconditionals produces an assignment of <u>both</u> truth values to them. In chapter 4 I attempt to present a representation of the paradoxical, and more generally, the problematical sentences, from the classical point of view.

The second role is that a formal theory may be extracted from it. This turns on taking the alignment points as the models of such a theory. The stably true sentences are precisely the extensions of the truth predicate. The motivation for this choice is its technical centrality in the process (it is this point the truth predicate waxes and wanes about). But although this is a good motivation mathematically, it does not tell us why the stably true sentences should be the extension of the truth predicate. What have we learned about the problematical sentences that we didn't know before that enables us to make this decision? One thing that may be added is that if a good argument can be given for this choice, we may again, contra Gupta, regard "true" as having an application procedure.

Although Kripke's approach also allows us to view the meaning of the truth predicate as an application procedure, it may be thought to have other drawbacks arising in part from the non-standard logic in the construction and from the philosophical desirability of the minimal fixed point.[11] This desiribility turns on the fact that the construction of the minimal fixed point intuitively exemplifies the idea of groundedness. I turn now to an informal discussion of this.

On an approach which stresses groundedness, the way we determine the extension of the predicate 'true' is by starting from sentences in which truth predicates do not appear, and extending from these sentences to the rest of the language using the Tarski biconditionals and the other standard satisfaction clauses. Whichever sentences fail to be assigned truth values by this process are 'ungrounded' and therefore problematical. The move from this fact to the claim that 'ungrounded' sentences do not express propositions, or do not have meaning, is a non sequitor. Grounded sentences are ones to which the decision procedure (a theoretical one) using Tarski biconditionals can be applied. It need not be seen as a criterion either for the meaningfulness of a sentence (which in the Standard Predicate Calculus is merely well-formedness) or for its truth-value bearing capability (which is merely the condition of bivalence on all the sentences of a language). Consequently I claim it should have no role in a theory of truth.

The focus on groundedness gives a kind of logical priority to sentences in which truth predicates do not appear. One is justified in asking what arguments there are for this priority.[12] Perhaps it comes from observations on how we come to assert sentences as true (but if so, this might be seen as confusing assertion with truth). The picture would be that we observe the world, make assertions based on these observations (those sentences have no truth predicates in them) and then observing the assertions we have made, we go on to make assertions about these assertions, and so on. But epistemological priority is not logical priority so that nothing in our practice gives priority to sentences in which truth predicates do not appear. The desire to make groundedness part of the theory of truth is similar in motivation to programs attempting to reduce theoretical predicates to phenomenalistic ones. Actually, something stronger can be said, the motivation is the same: a desire to avoid undecidables. I have already alluded to Quine on bivalence above. All I need add here is that Quine has taken solace in the fact that terms which we have to regard "as true or false of objects even in the absence of what we in our bivalent way are prepared to recognize as objective fact"[13] are restricted to common sense classifications of physical objects. This solace vanishes when it is recognized that "true" is among them. In conclusion we may regard truth tellers as having truth values as other sentences do and our puzzles about what their truth values could be--in principle--are no more cause (or no less cause--this depends on your viewpoint) for concern than analogous problems with commonsense terms.[14]

To sum up, I have expressed skepticism about the philosophical motivations of Gupta, Herzberger, and Kripke. I have also sketched out an alternative position which takes bivalence seriously and regards every sentence as true or false, regardless of whether the Tarski biconditionals and the interpretations of the nonlogical predicates and constants supply a truth value for it or not. One more remark is in order. One can characterize the Kripkean fixed points without recourse to an inductive construction. Further, closing them off, we can take them as models in a model theory. Something analogous may be done in the Gupta/Herzberger approach[15]. Then all that is left to decide between different approaches are technical considerations. But this is really as it should be.

V

All of the above discussion must be qualified by taking the Liar's Paradox into account. After all, apart from the fact that liar's paradoxes are Tarski-ungrounded[1], they violate the Tarski biconditionals. They do so in such a way that, given bivalence and the natural association of truth value with assertability of truth, they must be regarded as both true and false. We have, I presume, as much desire for sentences to be either true and false as we do for them not to be both.[2] It will not do therefore to merely claim that liar's paradoxes are either true or false, but we don't know which. We must also be able to give some plausibility to the idea that they are not both.

Consider the following argument. Suppose it is asserted that the term "(1)" refers transparently to "(1) is false." in "(1) is false.". Then it follows that (1) is false if and only if (1) is true. By reductio ad absurdum it follows that "(1)" does not refer transparently to "(1) is false." in "(1) is false."[3]

I realize that some intuitions militate against this argument. The primary intuition is that "(1)" is a perfectly good name and that perfectly good names refer in perfectly good (transparent) contexts. The second, and equally important intuition is that "true" is a perfectly good (transparent) context. Skyrms[4] has claimed otherwise. To do so, he exploits the intuition that although a sentence like "(1) (1) is not true." is paradoxical and therefore problematical, a sentence like "(2) '(1) is not true.' is not true." is not paradoxical. This intuition gains force if one already is inclined to regard the paradoxical sentences as neither true nor false. The second sentence therefore is saying something true; but this intuition Skyrms feels, motivates intensionality[5]. Now several points need to be made. First, it is possible, when accepting the reductio above, to blame not only the truth predicate but the constant too. Theoretically, there are always two (not necessarily exclusive) possibilities: we can blame the context (an intensional context) or we can blame whatever is being substituted into the context. In the latter case we have to have an explanation of why what is being substituted into the context is not acting (tranparently referring, say) as it normally does; such an explanation must not turn on blaming the context alone. A theory allowing such an explanation can be called "intensional" if we wish, but care should be taken to distinguish it from traditional ones under that name. Generally speaking such an explanation seems not to be available with the most commonly studied intensional contexts such as belief or modal contexts, which is why it is not natural to think of the second possibility. The next thing that should be said is that in large measure the temptation to regard liar's paradoxes as neither true nor false turns on regarding them as not expressing a proposition or not having a meaning. In a way, this is peculiar. The sentence is certainly well-formed grammatically, and the name occurring in it refers (in some sense)[6] to an item as names normally do. The temptation is actually equivalent to regarding the Tarski biconditionals as exhausting the meaning of the term "true", an idea which I have discussed above. If we do not accept that idea, we have little reason to regard the Liar's Paradox as meaningless. Adopting an intensional theory need not cost us bivalence. If the Tarski biconditionals are taken as implying the condition of extensionality for truth and falsity predicates, then all that need follow is that for some sentences they are false.

Let us take a closer look at the kind of intensional theory I want to offer. Consider, instead of the Liar's Paradox, a Buridean Symposium: "(1) (2) is false.", and "(2) (1) is true." In this case, we can deny the transparency of either context. Unfortunately, denying one or the other alone seems arbitrary and denying both seems redundant. However, imagine the following very naive picture[7]. We name sentences in baptismal ceremonies. These baptismal ceremonies can fail, that is, although we always succeed in referring in one sense (opaquely), we can fail in another sense: it can be that the Tarski biconditionals cannot hold and consequently the name cannot refer transparently in the context of the truth and falsity predicates[8]. Why should I say that it is the name that does not refer transparently and not that the context is opaque? Because, in general, only some names will be "funny". What particular names refer transparently can turn on what other successful baptismal ceremonies we have already carried out. The baptismal ceremonies are holistic in the sense that they affect each other. For example, all things being equal, if we have already baptised "(2) is false." with the name "(1)", but have not yet utilized "(2)", we can refer transparently to "(2) is false.". However we cannot at this point transparently refer to "(1) is true." with the name "(2)". The notion of opaque reference can be understood to some extent the way opaque contexts are. If a constant A refers opaquely to FA in the context of a semantic predicate "F", it does not follow that "FA" is meaningless. Indeed, the process I have described in determing that "A" does not refer transparently turned on conditions for the predicate and the constant which constitute part of their meaning. Thus "A" refers, but opaquely, and consequently the Tarski biconditionals cannot be used to determine the truth or falsity of "FA" because they, together with opaque reference, imply the condition of extensionality. Traditional opaque contexts, such a belief contexts, are not meaningless. The truth values of sentences containing them must be determined in some nonextensional way.

Similarly, for "FA" that other nonextensional way is the presence or absence of the sentence "FA" itself in the extension of the truth or falsity predicate, an undecidable fact[9].

Now I have discussed a syntactically very simple example of the Liar's Paradox. Since such paradoxes can be sentences in which no constants appear, it may be unclear how the analysis given above should be extended to such cases. Take for concreteness, a sentence of the form (EA)(PA & FA), where A is an individual variable, F is understood to be the falsity predicate and P is a predicate holding of one item, namely the sentence (EA)(PA & FA). How should a sentence like this be handled? Recall that in Standard Model Theory, the quantifiers manage to range over all the objects in the universe of the model via the variables. More specifically, the variables are impressed into service as surrogate constants. Let us call a mapping of the variables to the universe of the model an interpretation. Since in an interpretation a variable acts like a constant, we could apply the above argument directly to the variables. That is, let A be a variable which in a particular interpretation refers to (EA)(PA & FA). Then by the kind of arguments we have offered, it refers opaquely and consequently the satisfaction conditions of FA will be disengaged from that of (EA)(PA & FA), and the latter sentence will be true or false via the standard satisfaction clauses and the truth value (nonextensionally determined) of FA.

Now for technical reasons, it is more appealing to use constants and go substitutional over the sentences of the language rather than use variables as above[10]. Thus instead, we argue that (EA)(PA & FA) is true, if a constant B may be found such that PB & FB is true. But then B as we have seen before may refer opaquely and in those cases where (EA)(PA & FA) is a liar's paradox, every constant referring to it must do so only opaquely as is relatively easy to see.

At this point it will do to discuss the particular technical issues in a little more detail. It is appealing to capture the kind of reasoning I have brought to bear on the paradoxes in the object language itself. To this end, a new piece of syntax is introduced. I call it an ostensive: /. It is a two place item with constants in its first place, sentences in its second place, and its intuitive meaning is: what is in its first place refers transparently in the scope of a semantic predicate to what is in its second place. Necessary conditions of its truth are a) that what occurs in its first place is mapped by the model to what occurs in its second place, and b) that the Tarski biconditionals hold[11]. As an example of its application, let A be a constant. Then -(A/FA) is true by the kind of reductio used above. That is, it is a theorem that A can never refer to FA transparently in the context of a semantic predicate. Let me add that, of course, variables can appear in the first place of an ostensive, and well-formed formulas which are not sentences can appear in the second place. All such free variables are to be understood substitutionally.

Let me make a quick observation or two about the contrast between an "intensional" approach of this sort to the paradoxes and an extensional approach. Recall that one ideal motivating us is a language which can express its own semantic concepts. Here, it is easy to see that "opacity" is restricted severely. If A is mapped to FA, then A must be opaque. But understanding FA to have a truth-value nevertheless, "intensionality" enables a different constant (B, say), to transparently refer to FA and, TB or FB's truth to express its truth value. This contrasts nicely with extensional approaches. For TB and FB are Tarski-ungrounded, and it is easy to see that they will be unstable in any of the Gupta/Herzberger approaches. Further a condition of a language containing its own truth predicate for every sentence of the language looks within reach. Namely, for each sentence S, there is a constant A mapped transparently to S so that TA <=> S. But this hope is dashed by a liar's paradox without constants such as discussed above, since any constant opaquely referring to it cannot do so transparently. Still, we have come closer than extensional approaches seem to[12].

VI

I will read the quantifiers as systematically ambiguous. That is, they are "objectual" over nonsentences in the domain of a model and "substitutional" over the sentences in the domain. "Substitutional" in this context means "metalinguistic"--for example, the existential quantifier is understood as "There is a constant such that" when quantifying into certain contexts[1]. A metalinguistic understanding of some of the apparatus of a language is natural when such languages have models which contain the languages themselves, although the particular technical devices I have may be unusual. I should add that a use/mention *error* should not be assumed any more than a set/object *error* should be assumed, in say, Quine's LBIA.

Although making the quantifiers (partially) substitutional evades certain technical problems, it may seem to create others. In particular, as has been observed in the literature, compactness can be lost. That is, in cases, where the language has, say, an infinite number of constants $\{A_j : j < O, O \text{ a cardinal}\}$, and one predicate P, $\{PA_j : j < O\}$ can imply (B)PB, even though no finite subset of it does. Now this makes substitutional quantification look worse than it actually is, since a standard assumption about human beings is that they can only use systems with proofs involving at most finitely many premises. But then since it looks like there is something inherently infinitistic about implication in such systems, and, unless an infinistic notion of derivation (costing us compactness) is introduced, there will be a gap between the notions of derivation and implication. But a simple modification of standard model theory will dispel this appearance by eliminating the validity of the infinistic inferences. Their acceptability is due to a certain picture of the relation between model theory and language. Roughly speaking, we enlarge the concept of a model satisfying a sentence of any language L to include models of languages L' where the set of constants and predicate variables of L' may properly contain that of L. With respect to this kind of model theory, substitutional systems can be shown to be complete and compact without recourse to infinitary proofs.

Traditionally, when models are defined, it is quite explicit that they are models of one and only one language. Typically, considerations of the consistency of a set of sentences or their completeness turns on the set of models of one particular language which is held constant. Thus, for example, the relation of a model satisfying a sentence holds only of models of a particular language we take the sentence to be in. In Standard Model Theory, this involves greater generality than might appear at first sight, as may be seen from the method of constants implicit in the Henkin completeness proof for the standard predicate calculus[2]. What this comes down to is the fact that given a set of sentences S of a language L which can be embedded via an identity mapping into a set of sentences S' of a language L' (which contains the alphabet of L plus additional constants) and there is a model M' of L' which satisfies S' then the reduct M of this model to L (that is, M contains the same domain as M' and the constants and predicate symbols of L have the same interpretation in M as in M') satisfies S. Of course, this makes it relatively a matter of indifference whether we put the stress on the sentences or on the languages. Fixing our attention to a particular language does not cost us anything in terms of, say, the consistency of a set of sentences since any model of a richer language that satisfies those sentences implies the existence of one in the language under question. But, if substitional quantification is used with a standard notion of model restricted to a particular language, we actually put a kind of holistic restraint on the notion of sentence which does not seem justified in this context. My alternate suggestion seems far more natural in regard to substitutional quantification than its competitor (after all, it gives rise to a more natural set of axioms than its competitor), and it seems to have no drawbacks. Of course the description of the set of models satisfying a set of sentences includes models of languages other than the one explicitly under consideration but such languages are as well defined as the collection of sets which are the domains in model theory. I should add that this picture may appear to fit the practices of natural languages better than the standard one. According to this picture, the introduction of new vocabulary is not seen either as always having been part of the language or as constituting a move to a new language (with a homophonic translation of the sentences of our original language into the new one), but merely as the introduction of new vocabulary. Not too much weight should be placed on this point however. The importance of the expanded model theory really lies in the fact that it is natural to substitional quantification.

The careful reader will have noticed that we have not merely broadened our relevent notion of model to include models of languages with more <u>constants</u> (as is all that would be required for substitutional quantification), but also to include predicate variables. The reason for this turns on the ostensive. Compactness is lost with substitutional quantification, recall, because the universal quantifier is not independent of the constants. We broke this dependence by allowing the universal quantifier to, as it were, rely on constants in other languages as well. The ostensive creates an analogous situation with regard to sentences (for these occur in the second place of ostensives) and we similarly regain compactness by allowing models of a set of sentences to include sentences not in the language of that set. It might be thought that additional constants generate <u>new</u>

sentences and this would be enough. But this is false because of the fact that the ostensive may be quantified into. A sentence such as (A)-(EB)(B/PA) in a language L denies that any sentence of the form PC, where C is a constant has a constant referring transparently to it (we read both quantifiers substitutionally). Any model of (A)-(EB)(B/PB) cannot allow sentences such as PC to have constants referring to them transparently even if PC is not a sentence of L. This is so because nothing in the satisfaction clauses of the quantifiers restricts them to a particular set of constants. Thus our broadening of the notion of a model causes a similar broadening in our interpretation of the quantifiers. They are metalinguistic in the sense that they refer to constants not only of the language the quantifiers appear in but also the constants of any language properly containing that language. So the only way to circumvent the loss of compactness is to introduce new predicate variables. This is essential to the completeness proof in the text as the reader can easily see.

VII

At this point we have motivated enough of the unusual aspects of my approach to enable us to now give an overview of the entire system. In discussing the model theory we will, for expository purposes, restrict ourselves to a language L. The models of L are like standard models except that they may contain the sentences of L itself, in which case the extensions of the truth and falsity predicates will not be empty. There must be special restrictions on the extensions of the truth and falsity predicates, since we want these extensions to have the following properties:

1) We want every sentence to fall either into the extension of the truth predicate or the extension of the falsity predicate. Further, these extensions are disjoint.

2) If every constant in a sentence transparently refers and further the sentence is Tarski-grounded, we want it to be in the extension of the truth predicate iff it has the value true.

3) Otherwise, we wish the presence of a sentence in the extension of a semantic predicate to be subject to the standard satisfaction clauses to the extent that the truth-value of the sentence is and, subject to the Tarski-biconditionals if its constants occur in the context of semantic predicates transparently.

(1) is the condition of bivalence. (2), although unobjectionable, looks weak. Suppose, for example, that S is a logical truth of the standard predicate calculus. Then S is grounded. Let A be a constant mapped to S. Then TA is grounded. But it is possible, on my approach, to have models in which A refers only opaquely to S. Suppose B is mapped to (-TB v -TA). Then, in models where (B/(-TB v -TA)) is satisfied, (A/S) cannot be. For if it were, TA would be true (since S is a logical truth) and -TB v -TA would reduce to a liar's paradox. But this is compatible with (2) since A refers opaguely and it can be seen to follow from the holistic nature of our assignment of names to sentences. But there is a strong intuition that models allowing this should not be admissible since what Gupta calls local determination[1] is violated. Of course a model theory which allows undesirable cases is superior to one which excludes desirable cases, for perhaps by either axioms or model-theoretic restraints, we may be able to exclude undesirable cases, or study the cost to the system of our restriction to this desirable subclass.[2]

(3) is due to two concerns which I will motivate by means of simple examples. Should a liar's paradox FA be in T's extension, we wish -FA to be in F's extension and (FA v S) for any sentence S to be in T's extension even though (for S false) both these sentences are Tarski-ungrounded. This is because of our understanding of negation and disjunction: namely our desire to guarantee that the

resulting system is conservative with respect to the standard predicate calculus. Secondly, should the constant B refer transparently to FA (which is possible), we want TB to be in T's extension; this is because of our understanding of the Tarski biconditionals, namely that where a constant transparently refers, Tarski biconditionals in which that constant appears on the right are satisfied.

The best way to make these issues clearer is to study the satisfaction clauses and see how they achieve the above three aims.

Another restriction must be placed on the extensions of the truth and falsity predicates. Unfortunately this restriction is not as easy as the others. For one thing, it is set-theoretic in nature. The conditions I have described above still allow the following possibility: it could be that a constant A refers to a Tarski-grounded sentence B (without constants, say) in T's extension, but FA is in T's extension. In this case (A/B) fails. Sometimes this is unavoidable, as the example in the discussion of (1) shows But often it is not. That is, it may be possible to modify the extensions of the truth and falsity predicates in such a way that they continue to obey the three constraints above, no examples of transparent reference that we had before now fails, and further, (A/B) is satisfied. The sense here is that the reasons for the failure of transparent reference should be stronger than the mere fact that the sentences TA and FA happen to be in the wrong places. Intuitively, what we need is a constraint which forces our model to contain the largest number of satisfied ostensives it can based on the set of ostensives already satisfied. (Alternatively, and more literally, we are interested in models whose extensions for the truth predicate contains a maximal set of ostensives. We might label this principle "the principle of maximal transparency".) This is the final condition we place on models of our system.

I have tried to be as general as possible in the construction of the model theory (for example, in many models there are not enough resources to do syntax). Given that the system is complete, certain kinds of specified situations can then be studied either model-theoretically or axiomatically. However, the last condition I've placed on the extensions of the truth and falsity predicates is quite strong. It is certainly imaginable that given an arbitrary domain containing (among other things, possibly) the sentences of a language, and given an arbitrary assignment of constants to the domain and n-place predicates to subsets of the nth Cartesian product of the domain, that no corresponding model exist. This turns out, by theorem 3 of chapter 2, to be a groundless fear.

Usually, the model theory supplies the motivation and intuitive justification for the complete set of axioms which characterize it (when such a set exists). So I will not say much about my set of axioms and inference rules. In any case, all are self-evident except axiom (A7): If $S_1, \ldots, S_n$ are wffs, A is an individual variable, $R_1, \ldots, R_m$ are wffs in which A occurs freely[3], and $B_1, \ldots, B_m$ distinct individual constants, then any closure of $-(A)((A/S_1) \vee \ldots \vee (A/S_n) \vee (B_1/R_1) \vee \ldots \vee (B_m/R_m))$ is an axiom. A simple example will motivate why (A7) must hold. Consider the sentence (A)((A/S) v (B/R)), where R is a wff with only the variable A free and S is a sentence. Let I be a mapping of the variables into the domain of a model which contains R and S such that the variable A is not mapped to S and (B/R[g(A):A]) is not satisfied (this is always possible because there are infinitely many sentences in a domain of a model if there is at least one, and if an ostensive with a constant B in its first place is satisfied than no other ostensive with B in its first place is satisfied). Then (A/S) is not satisfied, nor is (B/R). But therefore (A)((A/S) v (B/R)) cannot be satisfied. A simple generalization of this reasoning is used to show (A7).

Notes for Chapter 1

I

1) Here is an example of a "local" solution: the "paradox" is just a reductio ad absurdum of the existence of the set of all sets which are not members of themselves. This has some analogy to a "solution" of the Liar's paradox which states that such a sentence does not express a proposition, without saying further what criteria there are for sentences expressing propositions and what work the notion of proposition does in a general semantic theory.

2) Perhaps this may seem to be putting the matter too strongly. Isn't Criterion T (for which Tarski cites Aristotle, among others, as a source on page 155 of CTFL) part of a pre-existing semantic theory? Also, as Mostowski points out in ET, p. 78, the idea of a model and other related notions were already known to logicians and mathematicians. But these informal notions and the mere existence of Criterion T (which immediately, as it stands, seems refuted by the Liar's Paradox) hardly seem to be a body of semantic theory to compare to nineteenth century Mathematics. Semantic Theory before Tarski was in the same state as Set Theory before Russell. But there was nothing like nineteenth century Mathematics to act as a global restraint for Semantics.

3) Kripke made a complaint of this sort in OOTOT, pps. 698-699. But apparently dissatisfaction remains, since Feferman in UTT p. 77 repeats a form of the same complaint. I should add that often motivation for "solutions" of the paradox comes from fairly naive views of the properties of natural languages which are then pointed to for justification of the purported solution. I view the desire for a "universal" language in this light when it is understood as a request for a language that is so expressive that there is nothing a metalanguage could say about it that it couldn't say itself. This is a much stronger request than one for a language containing its own semantic theory or even one containing its own theory of validity! Although I can see (and try to give below) reasons why languages with <u>these</u> properties would be desirable, I have trouble seeing why a "universal" language is (apart from the claim that natural languages have this property).

4) Or at least failed in the aim its proponents claimed to have. Actually, since active areas of Mathematics lie outside Set Theory (e.g., Category Theory), it failed even broadly construed. For other problems with it, see Parsons, MFO, p. 196-7, and Benacerraf and Putnam POM, pps.

11-16.

By the way, I am understanding "Logicism" loosely as the reduction of Mathematics to Set Theory in a "collection" of formal languages with Tarskian Semantics. This bears a historical relationship with the traditional usage of the term.

5) Russell, MSOR, p. 126

6) I don't want to give the impression that the observations of this last paragraph are new. For example, similar sentiments may be found in Davidson's Semantics for Natural Languages, p. 59, TAI. But perhaps I take a broader view of the options than he does.

7) Alternatively, this might be seen as additional evidence for the empirical character of Mathematics.

8) For some details on the problem see Anil Gupta's TP, pps. 6-19.

9) For example, see Enderton's MIL, p. 122.

10) No one has tried to determine whether the general practice in mathematics of taking results from any area that may be applicable to one's work, is consistent with a rigorous hierarchy in Tarski's sense. If not, this would be a further difficulty, for this is a natural practice in a subject matter contained entirely in one language (or for which translation from other languages poses no threat of contradiction, e.g., from Russian to English) but must be carefully justified otherwise.

Two further caveats. It is true that this hierarchy can be captured entirely in the language of Set Theory. But that is no help. For Set Theory itself is obviously incomplete, and thus the same kind of argument applies to it as well. Also, it might be thought that the Godel sentence is produced by means of a proof schemata and it is only necessary that we know that such a proof be applicable to any language in an hierarchy. But then we would have knowledge of mathematical truths that transcends any language. This is peculiar.

11) Not quite. A language with a countable number of truth predicates may be regarded as one which contains its own semantics. (Consider a Tarski hierarchy and take the union of the languages) Such a language would contain its own semantic theory, in some sense. But, 1) mathematical truths describing the entire hierarchy would still elude translation, and 2) there would be type-restrictions on the quantifiers and on the constants and so the complaints of footnote 8 would still stand here. Clearly something more is needed than merely the requirement that the language

contain its own semantic theory. This something is that the language be, as Feferman calls it, type-free.

Two further points: I should also add that something more may be called for than that the language contain its own semantic theory. It is desirable to explore languages which can contain many semantic theories--not merely their own: the study of formal languages and their semantics is part of mathematics. Again, this is not quite a request for a "universal" language. Similarly, the desire to study languages which can contain all mathematical truths is not a desire for a formal system formalizing all mathematics. The latter is impossible, of course.

Next, I don't want to give the impression that no one is concerned with languages which contain their own semantics. Kripke in OOTOT, Gupta in TP, Gupta and Martin in FPTWK, and Herzberger in NPFO all address this issue from their vantage points. Gupta, in conversation, has pointed out to me that various concepts besides truth (e.g., stability, higher-level stability, etc.,) can be introduced via a rule of revision much as truth is in his approach. What this looks like in detail awaits further development. I should say probably, except under fairly peculiar (paraconsistent) circumstances, the best that can be hoped for is languages which "approximately" (from the perspective of a classical metalanguage) contain their own semantics. What "approximately" comes to, how it can be measured, and how much it satisfies the demands described above are highly nontrivial questions.

II

1) See Feferman's UTT, pps. 76-77.

2) See however, footnote 11 of section I above.

3) Why is a definition of satisfaction (and thus of truth) so important to have? Why not just introduce the predicate as a primitive? Kripke in OOTOT, p. 698, chides the literature (of that time, of course) for "abandon[ing] the attempt at a mathematical definition of truth, and tak[ing] it as an intuitive primitive." In so doing, "there is a sense in which [Tarski's approach] provides a theory while the alternative literature does not."

Now, clearly, merely the presence of an additional primitive predicate in a theory does not deprive it of the status of being a theory. But Kripke is getting at something. Tarski saw this particular task as important. He was not merely interested in constructing a formal metalanguage in which such a predicate occurred primitively. Of course it is always important to isolate the primitive predicates of a theory and reduce the nonprimitives to them by definitions. This is of mathematical interest and has long been traditional there. But, generally, there is no unique set of primitive predicates (any more so than a unique set of axioms) and there must be additional reasons for Tarski's desire to supply a definition for truth.

Well, clearly I am not the first to ask this question. Field asked it quite a while ago (TTOT) and the answer he gave is supported by a remark of Tarski's (ESS, p. 406). I am alluding, of course, to Tarski's desire that all semantical concepts be reduced to logical and physical concepts. If this is the aim, then I agree that Field has shown that it has not been done. Putnam, (MMS, p. 15-17) responds to Field's argument, essentially, by saying that one need (at least) for a truth predicate is semantic ascent. Why can't this need be satisfied by introducing 'true' as a primitive predicate with the appropriate axioms?

Tarski's answer (ESS, p. 405-406, CTFL pps. 154-163) is that introducing a primitive predicate "true"

1) involves technical problems with either a) intensional contexts, or b) the fact that Convention T cannot be given as an axiom in the particular case when a name of a sentence does not enable us to indicate the sentence (e.g., "The first sentence which will be printed in the year 2000"), and

2) makes the question of the consistency of the resulting semantical system harder to answer.

Let me add that being driven in this way to the approach Tarski takes makes him even more pessimistic about natural languages since the success of his approach depends crucially on the syntax of the object language and he feels, in regard to natural languages, that "[w]e are not able to specify structurally those expressions of the language which we call sentences." [CTFL, p. 164)

4) Actually, my suspicion is that Tarski thought he had defined a predicate which captured the viewpoint Putnam calls "Metaphysical Realism" in MMS, p. 123.

What evidence do I have for this claim? Unfortunately, not quite enough. In SCT p. 342, he writes:

"We should like our definition to do justice to the intuitions which adhere to the <u>classical conception of truth</u>--intuitions which find their expression in the well-known words of Aristotle's <u>Metaphysics</u>:

"<u>To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true</u>.

"If we wished to adapt ourselves to modern philosophical terminology, we could perhaps express this conception by means of the familiar formula:

"<u>The truth of a sentence consists in its agreement with (or correspondence to) reality</u>.

"...If, on the other hand, we should decide to extend the popular usage of the term "designate" by applying it not only to names, but also to sentences, and if we agreed to speak of the designata of sentences as "states of affairs," we could possibly use for the same purposes the following phrase:

"A sentence is true if it designates an existing state of affairs.

"...none of [these] can be considered a satisfactory definition of truth. It is up to us to look for a more precise expression of our intuitions."

In CTFL, p. 153, he writes:

"I would only mention that throughout this work I shall be concerned exclusively with grasping the intentions which are contained in the so-called classical conception of truth ('true--corresponding with reality') in contrast, for example, with the utilitarian conception ('true--in a certain respect useful')."

Of course, what he comes up with as a precise expression of our intuitions is Criterion T, which is trivial! I would like to draw the conclusion that he failed to see that Criterion T does not capture the classical conception of truth (understood realistically), as he described it. Certainly the last passage I cited seems to suggest that he did not think his definition captured the utilitarian view (which it does if the metalanguage is interpreted appropriately). But he may have regarded the classical notion as the core notion for the others, and then the fact that a pragmatist could use it would not surprise him. The first passage referred to in footnote 6 supports this idea. So I am not completely sure how Tarski saw what he did. However, I am relatively sure what others, for example, Popper, Carnap, and Hempel, made of it. See footnote 6 below.

5) For example, p. 345 in SCT or p. 187 of CTFL. Of course, Criterion T's constraints on a theory (of Truth) is exhausted by the condition that all equivalences of the form "X is true if, and only if, p" (where X is a name of p) follow from it. The fact that there are, in general, an infinite number of such equivalences, does not make it less trivial.

6) SCOT, p. 362. Now despite the explicit suggestion that the theory is neutral towards epistemological issues, his list of the particular positions the theory of truth is neutral towards suggests that it is also neutral toward metaphysical issues. This is the line that Putnam takes in MMS, p. 9. Putnam also writes on p. 3 in MMS that "...Carnap, Popper, Hempel, etc., all accepted the notion 'true' as a notion that even the most scrupulous empiricist could be happy with. But they accepted it because they thought that it did no philosophical work."

I am not sure that the matter is that simple. And I think this because of something due to Putnam himself! Putnam (MMS, pps. 25-33) notes that the axioms and inference rules governing the "classical" connectives are insufficient to "fix" them classically. For example (his example as a matter of fact), an "intuitionistic" reinterpretation of such axioms and inference rules is possible (this turns on a simple mapping of the intuitionistic system into the classical system). Defining truth a la Tarski in a metalanguage under such an interpretation gives us a nonrealist notion of truth! In particular, a language that looks bivalent actually isn't. So what? So, Carnap, Popper, and Hempel, etc., didn't know such a thing was possible. This is why they thought Tarski's definition (which looked trivial) actually captured a substantial philosophical notion. It was the best of all possible worlds: a definition based on Criterion T that no one would dare deny that yet does substantial philosophical work. (Actually, it is the metalanguage that does the work, and once this is recognized, the appearance that the definition captures a substantial notion vanishes.) So, for example, Popper writes in CAR, p. 223:

"So far I have spoken about science, its progress, and its criterion of progress without even mentioning truth. Perhaps surprisingly, this can be done without falling into pragmatism or instrumentalism. Indeed, it is even possible to argue in favour of the intuitive satisfactoriness of the criterion of progress in science without ever speaking about the truth of its theories. In fact, before I became acquainted with Tarski's theory of truth, it appeared to me safer and more economical to discuss the criterion of progress without getting too deeply in the highly controversial problem connected with the use of the word 'true'.

"My attitude at the time was this: although I accepted, as almost everybody did, the objective or absolute or correspondence theory of truth--truth as correspondence with the facts--I preferred to avoid the topic. For it appeared to me hopeless to try to understand clearly this strangely elusive idea of a correspondence between a statement and a fact.

"...All this was changed by Tarski's theory of truth and of the correspondence of a statement with the facts. Tarski's greatest achievement, and the real significance of his theory for the philosophy of the empirical sciences lies, I believe, in the fact that he re-established a correspondence theory of absolute or objective truth which showed that we are free to use the intuitive idea of truth as correspondence with the facts."

This hardly seems neutral. And look what Popper does to the competing theories on p. 224:

"Indeed, the three rivals of the correspondence theory of truth--the coherence theory which mistakes consistency for truth, the evidence theory which mistakes 'known to be true' for 'true', and the pragmatic or instrumentalist theory which mistakes usefulness for truth--these are all subjective (or 'epistemic') theories of truth, in contradistinction to Tarski's objective (or 'metalogical') theory."

The important phrase here is "in contradistinction to". Similarly, Hempel writes in ASE, p. 42:

"While...[once]...I argued in effect that the only possible interpretation of the phrase 'Sentence S is true' is 'S is highly confirmed by observation reports', I should now reject this view. As the work of A. Tarski, R. Carnap, and others has shown, it is possible to define a semantical concept of truth which is not synonymous with that of strong confirmation.... Thus, e.g., if S is any empirical sentence, then either S or its denial is true in the semantical sense, but clearly it is possible that neither S nor its denial is highly confirmed by available evidence."

And Carnap writes in TAC, p. 119:

"At times it was considered altogether impossible to establish an exact and consistent definition of truth (in its customary meaning); this has brought it about that the term 'true' was used in the sense of the entirely different concept 'confirmed'.... Thus one would find it necessary to abandon, e.g., the principle of the excluded middle. Tarski, however, succeeded in establishing an unobjectionable definition of truth...."

Clearly the importance of the (false) idea that Tarski's definition captures, in particular, a form of bivalence is at work here.

By the way, in ESS, p. 407, Tarski writes: "The very fact that it has been possible to define the semantical concepts, at least for formalized languages, in a correct and adequate manner seems to be not entirely without importance for the philosophical standpoint. The problem of the definition of truth, for example, has often been emphasized as one of the fundamental problems of *the theory of knowledge*." (italics mine). Clearly there had been a shift in opinion over the years.

7) See, in particular, Kit Fine and Timothy McCarthy, TWS, p. 401.

8) Tarski writes in CTFL, p. 164-5:

"A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality...If we are to maintain this universality of everyday language in connexion with semantical investigations, we must, to be consistent, admit into the language, in addition to its sentences and other expressions, also the names of these sentences and expressions, and sentences containing these names, as well as such semantic expressions as 'true sentence', 'name', 'denote', etc. But it is presumably just this universality of everyday language which is the primary source of all semantical antinomies, like the antinomies of the liar...These antinomies seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of logic hold, must be inconsistent...[N]o consistent language can exist for which the usual laws of logic hold and which at the same time satisfies the following conditions: (I) for any sentence which occurs in the language a definite name of this sentence also belongs to the language; (II) every expression formed from ["x is a true sentence if and only if p"] by replacing the symbol 'p' by any sentence of the language and the symbol 'x' by a name of this sentence is to be regarded as a true sentence of this language; (III) in the language in question an empirically established premiss having the same meaning as ["the sentence 'for all p, if c is identical with the sentence 'p', then not p' is identical with c."] can be formulated and accepted as a true sentence.

"If these observations are correct, then <u>the very possibility of a consistent use of the expression 'true sentence' which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression</u>."

But he writes in a milder vein some years later in SCT, p. 349:

"At first blush it would seem that [everyday language] satisfies both assumptions (I) and (II), [(I) the ordinary laws of logic hold; (II) A liar's paradox can be empirically proved] and that therefore it must be inconsistent. But actually the case is not so simple. Our everyday language is certainly not one with an exactly specified structure. We do not know precisely which expressions are sentences, and we know even to a smaller degree which sentences are to be taken as assertible. Thus the problem of consistency has no exact meaning with respect to this language. We may at best only risk the guess that a language whose structure has been exactly specified and which resembles our everyday language as closely as possible would be inconsistent."

9) For example, this seems largely Kripke's philosophical motivation in OOTOT. Similarly Gupta's in TP.

10) E.g., the Strengthened Liar.

11) It is important to study baby semantic theories which only contain their truth and falsity predicates as a preliminary study to the more difficult work. This is the spirit in which my own work here should be taken.

III

1) Of course, generally, the relations in the model and the predicates in the language must be rich enough to describe the syntax of the language.

2) It may be argued here that full Set Theory is still not needed since, by the Loweinheim-Skolem Theorem, countable domains are all that suffice.

3) OOTOT, p. 705

4) See Gupta and Martin's FPTWK.

IV

1) This must be qualified. It is possible to construct examples where a truth teller must have a particular truth value or another sentence will be paradoxical. But I am not considering these kinds of examples here.

2) OOTOT, p. 692.

3) TP. p. 37.

4) I have been very sloppy in my description of the various solutions here. The phrase "best candidate" is Gupta's and I don't mean to suggest, although it may appear that way, that Gupta and Herzberger's approaches are identical. The essential point for my purposes here is that the extension (and anti-extension) of the truth predicate accumulate sentences as we ascend up the hierarchy in Kripke's solution. This is not true of the extension of the truth predicate in the case of Gupta and Herzberger except in certain very favorable situations.

I want to note that from a mathematical point of view the inductive approach to the minimal fixed point language in Kripke is dispensable. On the details of how to show the existence of fixed point languages without recourse to an ordinally indexed sequence of approximations, see Fitting's NMT. It is reasonable to assume Kripke presented the step by step construction of the minimal fixed point because it seemed to be the most natural for the intuitive picture that he was trying to capture.

I draw my understanding of Herzberger's approach from NNS and NSL.

5) TP, p. 37.

6) I don't mean to suggest that this _is_ Gupta's motivation,

although it might contribute to it. I am merely running through the kinds of motivations that might fit this kind of approach.

7) For details, see Quine's WPB. A nice mathematical example of this are the sentences of the language of Number Theory which hold in the intended model.

8) One should not underestimate how radical Gupta's proposal is from a philosophical point of view. An intuitive picture of model theory is to see models as "worlds" in which sets of sentences are true or false by virtue of what the terms in them refer to in the "world" and what classes predicates hold of in the "world". This is not a natural model theory for Gupta. Rather, we need a set of such worlds (indexed ordinally) in which the revisable predicates are mapped to various sets.

There is something definitely peculiar about the revisibility thesis. It certainly doesn't correspond to normal intuitions about the term 'true' unless these intuitions are epistemological, in which case they apply to many terms. Further, on a sentence by sentence basis, the thesis makes no sense. If the truth-value of a sentence with the occurrence of the truth-predicate (e.g., "TA") intuitively depends on one without the truth predicate (e.g., "A"), then we feel we have an application procedure, namely, using the Tarski biconditionals (or an appropriate generalization such as Tarski satisfaction clauses) to reduce it to a sentence without such predicates. Otherwise, we are at a loss as to what to do. Notice that none of Gupta's best candidates is intuitively satisfying on paradoxes. And could they be? This is a most strange revision procedure. It is intuitively unnecessary on some sentences and it fails to gain us anything on the rest.

Something more must be added here. It may seem that I am implicitly relying on a intertheoretic notion of paradoxical sentence--one, that is, that applies to all the various approaches. Of course there is no such notion. Notions close to it can be defined relative to the different approaches but they are quite intratheoretic. Gupta, then, might respond that the revisibility process itself weeds out which sentences are troublesome and which are not and without it we would have no such notion (consequently, I cannot claim that it is useless in the way I have argued). However, we do have such a notion (although it does not appear in the literature), for the standard Tarskian case. See chapter 4 for details. This seems the appropriate notion to rely on when an intertheoretic notion is called for. We can then recognize that certain sentences paradoxical or problematic from the Tarskian point of view may not be problematic from other approaches. The sentences problematic from the Tarskian point of view are not merely problematic relative to that view; they are the sentences which motivated the study of alternative approaches in the first place.

9) NSL, p. 482

10) NSL, p. 481

11) For details on the drawbacks of Kripke's theory, see TP, pps. 33-7. Anticipating myself, I should say that as I reject the significance of groundedness for truth, I am not worried about Gupta's criticisms which arise from the drawbacks of the minimal fixed point language. Not much remains of his criticisms in this case. The first two vanish with appropriate choice of a valuation scheme (and the second doesn't hold in all fixed points). The fourth is based on the choice of the minimal fixed point. A modified version of his third criticism holds of all intrinsic fixed points. But as the reader will see below, I see no reason to find those superior choices either. Given the views I express below the only drawback I can find in this approach is that it is gappy (provided that really does make less simple than bivalent approaches), and that it seems to capture less of its semantics than bivalent approaches with their own truth predicates.

12) Consider the standard (recursive) satisfaction clauses of Tarski's. Do the well-formed parts of a sentence have logical priority over it? Is a theory of levels appropriate? I note that there is often an expressed interest in classifying sentences according to when they are assigned truth values in the inductive (or semi-inductive) construction of the desirable languages. This is something of a technical illusion. Part of the interest of the jump operator I define in chapter 4 is due to the fact that the resulting "theory of levels" would be quite different from the more standard jumps even though in the most interesting cases the fixed points are exactly the same. Perhaps it is thought that properties due to the operator are intrinsic, in some way, to the sentences themselves.

13) Quine, ibid, p. 36. Of course, it is our model theory that enables us to claim this. If we take the minimal fixed points of Kripke (not closed off), as our set of models, we will not regard any sentence as actually true or false but unverifiable (in principle) as such. So, the promissory note I must cash is to supply the actual model theory.

14) Consider the following passage from Kripke, p. 701:
"Suppose we are explaining the word 'true' to someone who does not yet understand it. We may say that we are entitled to assert (or deny) of any sentence that it is

true precisely under the circumstances when we can assert (or deny) the sentence itself. Our interlocutor then can understand what it means, say, to attribute truth to (6) ('snow is white') but he will still be puzzled about attributions of truth to sentences containing the word 'true' itself. Since he did not understand these sentences initially, it will be equally nonexplanatory, initially, to explain to him that to call such a sentence "true" ("false") is tantamount to asserting (denying) the sentence itself.

"Nevertheless, with more thought the notion of truth as applied even to various sentences themselves containing the word 'true' can gradually become clear. Suppose we consider the sentence,

(7) Some sentence printed in the New York Daily News, October 7, 1971, is true.

(7) is a typical example of a sentence involving the concept of truth itself. So if (7) is unclear, so still is

(8) (7) is true.

However, our subject, if he is willing to assert 'snow is white', will according to the rules be willing to assert '(6) is true'. But suppose that among the assertions printed in the New York Daily News, October 7, 1971, is (6) itself. Since our subject is willing to assert '(6) is true', and also to assert '(6) is printed in the New York Daily News, October 7, 1971', he will deduce (7) by existential generalization. Once he is willing to assert (7), he will also be willing to assert (8)."

Now this is just an intuitive sketch and so I should not lean on it too heavily for an understanding of Kripke's position. Nevertheless, clear differences between this and my approach can be seen. I feel that the meaning of the concept of truth is exhausted by certain model theoretic conditions placed on the predicate (such as that every sentence is either in the extension of the truth predicate or in the extension of the falsity predicate, etc.,). Thus to understand a usage of the predicate, I do not have to eliminate it in the way Kripke describes. This is so whether or not the Tarski biconditionals hold in a particular case. Grounded sentences are those for which a method (theoretical) can be given for deciding which sentence are in fact true or false; such a method is not a desiderata of whether I understand the occurrence of the term 'true' in a particular sentence.

15) As Herzberger suggests.

V

1) See chapter 4 for the definition of this notion.

2) More in fact. This is presumably another reason why the

Liar's Paradox is traditionally more upsetting than truth-tellers. It also (in my opinion) is why the gap literature developed before the glut literature despite the formal similarity of gluts and gaps.

3) Of course this reductio is bogus. I have conveniently taken other assumptions for granted to enable me to present the reductio this way. That is, I assume that every sentence is either true or false, that every sentence expresses a statement, and any number of other assumptions that others in the literature have doubted to "solve" the Liar's Paradox. This is still another assumption that can be doubted.

4) ISC.

5) He writes: "Intuitively, we want (1) to be neither true nor false, and (2) to be true. Since one can move between (1) and (2) by substitution of coreferential singular terms, such a theory must be <u>intensional</u>. Notice that although (1) and (2) refer to the same sentence, (1) is <u>self</u>-referential whereas (2) is not. This simplest example suggests that an adequate treatment of the semantical paradoxes will turn on intensional aspects of semantical self-reference."

Despite my partiality to "intensionality", I find this argument peculiar. First, I attribute the intuitive difference between (1) and (2) to an ambiguity about negation, in particular, external and internal negation. Consequently, this argument does not motivate an intensional approach for me but merely a desire to avoid ambiguity. Secondly, I cannot see what possible interest there is in the fact that although (1) and (2) refer to the same sentence, one is self-referential although the other is not. This could only motivate an intensional approach if one had forgotten that sentences are not individuated according to extensional equivalence, that "self-referential" is a predicate of sentences, and that therefore the notion "contains the constant A" should motivate an intensional theory as much as "self-referential".

6) I will say later that it "opaquely refers". For now, the point is that our recognition of paradox turns on treating the name appearing in the sentence as we would treat any other name.

7) I call the picture naive merely because I find reference puzzling. But my worries have nothing to do with the issues surrounding the Liar's Paradox so this is not the place to discuss them.

8) This is not traditional, but we will consider opaque

reference as a necessary condition of transparent reference.

9) Here is the picture. The baptismal ceremony for a constant A guarantees opaque reference in every context. If there are no further conditions for transparent reference in a context (such as with the nonlogical predicates) we have transparent reference immediately. With the truth predicates, the order in which opaque reference is established affects the possibility of transparent reference. One can imagine more general conditions upon different kinds of predicates (besides truth predicates) all depending on the order in which opaque reference is established. The interaction here is subtle and it seems to me that the traditional language of "opaque" and "transparent" contexts does not do justice to it. But the natural generalizations possible do seem to place the burden more on the constants opaquely or transparently referring than it does on the context being transparent or otherwise. My usage of "transparent" and "opaque" therefore should not be confused with traditional terminology (where, for example, a belief context is merely "opaque"--not opaque with respect to certain constants, although not necessarily with regard to others.) Let me finally say that the notion of the order in which the opaque reference is established need not be regarded as temporal. We can imagine certain restrictions being placed upon constants so that certain ones are guaranteed transparent reference in cases of conflict. See VII for further comments on this.

10) The approach I am not taking would allow a sentence of the form (EA)(PA & FA) to have a different truth value from (EB)(PB & FB). This, although perfectly understandable, given that the variables act like constants, does violate intuitions about bound variables. Also, the loss of alphabetic variance seems (to me) to complicate matters quite a bit. An ideal I have tried to stick to is that the resulting system be a conservative extension of the standard predicate calculus.

11) Thus, condition (a) states that at least opaque reference holds and with condition (b) we get transparent reference in the context of the truth predicates.

12) Solutions are imaginable (they always are but their cost must be measured carefully). One I have alluded to in (10). Another possibility would be to introduce a countable collection of types of variables restricted in range to parts of the domain.

VI

1) Actually, the matter is not that simple. It is possible to quantify simultaneously into contexts where the quantifier is understood substitutionally in one context and nonsubstitutionally in the other. Such sentences need a longer and more awkward translation of the quantifier into English.

2) Further details may be found in MT.

VII

1) See TP, p.21 and further.

2) Superficially, the competition scores higher on this. In Gupta's case, for example, this kind of violation will not survive the revision procedure. On the other hand, we can define the Tarski-grounded sentences as in chapter 4 (or an appropriate generalization for languages with ostensives), and restrict our attention to models in which any constant refers opaquely to such a sentence iff it refers transparently to it. I conjecture that this subclass of models will not be recursively axiomatizable and further that this result is not an idiosyncracy of my approach.

3) As may be seen in the definition of this notion, the variable must occur in a wff for it to occur freely there.

CHAPTER 2

CONVENTIONS

In Chapter 2 and Chapter 3, I present the formal results. Exposition will be at a minimum and solely for the purpose of orienting the reader and making clear the import of certain results. I will use certain conventions throughout. Capital letters from the beginning of the alphabet will be metalanguage variables, expressions following a colon will implicitly have quotes around them, contexts where metalanguage variables appear will implicitly have Quine's quasiquotes around them and "T" and "F" will be metalanguage variables for themselves.

In this chapter I will present the formal apparatus: the languages and the model theory. The remainder of the chapter will be dedicated to showing that certain kinds of models exist. In chapter 3, I will give axioms for the model theory presented in chapter 2 and show they are complete.

PRIMITIVE SYMBOLS OF OSTENSIVE-LANGUAGES (or O-languages)

Definition: An O-language has the following symbols:

1) A set of individual variables: x, y, z,...

2) A set CON of individual constants: a, b, c,...(this

set will vary from language to language).

3) A set of predicate variables: P, Q, R,...(This set also varies from language to language. Further, each predicate variable comes implicitly with a place-number).

4) A set of predicate constants: T and F (these both have place-number 1).

5) The symbols: -, v, (,), /, E.

We assume we have a denumerably infinite amount of the symbols of type 1, and at most denumerably many symbols of types 2 and 3.

FORMATION RULES FOR O-LANGUAGES

FR1 An expression consisting of a predicate followed by the predicate place-number of occurrences of individual variables and/or individual constants is a wff. All occurrences of individual variables in it are free occurrences.

FR2 If A is a wff, so is -A. The free occurrences of variables in -A are all and only the free occurrences of variables in A.

FR3 If A and B are wffs, so is (A v B). The free occurrences of variables in (A v B) are all and only the free occurrences of variables in A and B.

FR4 If A is a wff and B is any individual variable, (EB)A is a wff. The free occurrences of variables in (EB)A are all and only those of A if B does not occur freely in A.

If it does, then the free occurrences of variables in (EB)A are all and only those in A minus the occurrences of B. (If B occurs freely in A, then we say the _free_ occurrences of B in A are in the scope of the quantified B (the token of B in (EB)) or are bound by the quantified B.)

FR5 If A is a wff, and B is an individual variable or individual constant then (B/A) is a wff. If B is a constant, the free occurrences of variables in (B/A) are those occurring freely in A. Otherwise its free occurrences of variables also include the occurrence of B in the first place of (B/A) (Wffs of this last sort are hereafter called "ostensives").

FR6 Something is a wff if it is so only because of FR1-5.

Definition: An individual variable A is a free variable of a wff B if there is a free occurrence of it in B.

Definition: A wff is closed (a sentence) if it has no free variables (in particular, a closed ostensive will be called an "ostensive-sentence").

Definition: Let A be a wff, B an individual variable or individual constant, and C an individual variable or individual constant. Then A[B:C] is the result of replacing all free occurrences of C in A by B, if C is an individual variable, and the result of replacing all occurrences of C in A by B otherwise. If $B_1, \ldots, B_n$ are individual variables and $C_1, \ldots, C_n$ are individual variables or individual constants, then $A[B_1, \ldots, B_n : A_1, \ldots, A_n]$ is understood in a similar manner (the substitution of $B_1, \ldots, B_n$ for $A_1, \ldots, A_n$ is taken to be simultaneous).

MODEL THEORY O

Overview: A model for a language L is built up in several stages. First we define a premodel. For this, the only restrictions on the extensions of the truth and falsity predicates are that they be disjoint and exhaust the set of sentences contained in the domain of the premodel. A notion of satisfaction is defined for premodels. It is unusual for variables occurring in the scope of semantic predicates. The notion of an interpretation I, that is, a mapping of the variables to the objects of the domain is broadened so that any variable mapped to a sentence is associated through a function g_I with a constant mapped to that sentence, if such exists (of course g_I depends in part on I, but we will in practice suppress the subscript where no misunderstanding is possible). This is necessary since quantification over the sentences is substitutional. For recall that the theory is "intensional" when it comes to constants occurring in the scope of semantic predicates. That is, for constants A and B mapped to a sentence S, TA may be true and TB false. Therefore, since the quantifiers are substitional, variables cannot merely be mapped to sentences. They must be associated with constants and a wff such as TX, where X is a variable will be satisfied iff TB is, where B is the constant associated with X by g. The satisfaction clauses (0.3 and 0.4) for the atomic sentences with semantic predicates and for the ostensives follows the pattern described in chapter 1-V. It is here that the Tarski biconditionals are encoded. In general, in a premodel, the extension of the truth predicate does not contain all and only sentences which are satisfied according to the satisfaction clauses. However those sentences which intuitively function like the simple example of the liar are satisfied iff they are in the extension of the truth predicate. The others need not be satisfied since their satisfaction turns on other conditions besides their mere presence in the extension of the truth predicate. So we define an honest premodel as one for which the extension of the truth predicate coincides exactly with the sentences satisfied according to the satisfaction clauses. Finally, placing the maximality condition discussed in chapter 1-VII on the honest premodels gives us the primary O-models. The O-models proper are the primary O-models of any language whose set of constants and predicate variables is a superset of those of L. We note that if we stayed with the traditional notion of a model in this context, it would be the primary O-model.

Definition: A domain of an O-language LAN is a collection of objects d which either contain at least all the sentences of LAN or none of the sentences of LAN.

Definition: A premodel for an O-language LAN on a domain d of LAN is a collection of mappings from: the set of n-place predicate variables to subsets of d^n, a mapping from T to a subset of the set of sentences in d, a mapping from F to the complement of T's extension relative to the set of sentences of LAN in d (We call the image of an n-place predicate B under such a mapping "the extension of B"), and a mapping from the set of constants to d.

Definition: An interpretation I on a premodel M is a mapping from the set of individual variables to the domain of M, and a partially defined choice function g from the set of individual variables and individual constants to the set of individual constants. For individual variable A, g selects a member from the set of constants $\{B \mid M(B) = A\}$, provided the set is non-empty. On the set of individual constants g is the identity function. Thus, here, interpretations on a premodel are individuated not only on the basis of what they map the individual variables to, but also on the particular choice functions they contain.

<u>Definition</u>: Let S be a wff all of whose free variables are among $A_1, \ldots, A_n$. Then if $B_1, \ldots, B_n$ are individual constants, the sentence $S[B_1, \ldots, B_n : A_1, \ldots, A_n]$ is an <u>instance</u> of S. If I is an interpretation and $B_i = g(A_i)$ for all i, then the sentence $S[B_1, \ldots, B_n : A_1, \ldots, A_n]$ is an <u>I-instance</u> of S.

<u>Definition</u>: Let W be a wff and I an interpretation on a premodel M. We say that I <u>satisfies</u> W with respect to M ("satisfies" for short in the following clauses) iff

O.1) W is $BA_1 \ldots A_n$, an n-place predicate variable followed by a string of length n of individual variables and/or individual constants, and the sequence of items of d which are the images of the individual variables and/or individual constants under I and/or M respectively in the order of $A_1, \ldots, A_n$ is in B's extension.

O.2) W is -A, where A is an unsatisfied wff. Or W is (A v B) where A and B are wffs, and either A or B is satisfied.

O.3) W is (A/B) where A is an individual constant or individual variable, B is a wff, g(A) is defined, there is an I-instance B' of B, B' is a sentence assigned by M to g(A), (g(A)/B') is in T's extension, and either Tg(A) and B' are both in T's extension or both in F's extension.

O.4a) W is TA, A is an individual constant or individual variable, there is a sentence B in T's extension, g(A) is defined, and (g(A)/B) is in T's extension; or (g(A)/B) is not in T's extension, B is a sentence assigned by M to g(A), and Tg(A) is in T's extension.

0.4b) W is FA, A is an individual constant or individual variable, there is a sentence B in F's extension, g(A) is defined, and (g(A)/B) is in T's extension; or (g(A)/B) is not in T's extension, B is a sentence assigned by M to g(A), Tg(A) is not in T's extension, and Fg(A) is in T's extension.

0.5) A is an individual variable, Y a wff, W is (EA)Y, and either i) there is an interpretation I' on d which disagrees with I on at most A and g(A), and I' satisfies Y with respect to M or ii) B is an individual constant and M satisfies Y[B:A].

If a wff W is satisfied by every interpretation I with respect to a premodel M, we say that M _satisfies_ W. If W is a sentence, we say that W is _true_ in M.

Definition: An _honest premodel_ for an O-language LAN on a domain d is a premodel for LAN on d which satisfies the additional condition: If the extension of T or F is nonempty, the subset of sentences of d that T is mapped to is the set of all the sentences satisfied by the premodel.

Definition: A *maximally honest premodel* M for an O-language LAN on a domain d is an honest premodel for LAN on d which satisfies the additional condition: Let M* be an honest premodel on the domain d of M which agrees with M on the mappings of the predicate variables and individual constants to d. Then if A is an individual constant and B a sentence such that (A/B) is in T's extension in M*, (A/B) is in T's extension in M.

Definition: Let LAN and LAN* be two O-languages. We say that LAN* is *at least as articulate as* LAN iff the set of predicate variables of LAN is a subset of that of LAN* and the set CON of LAN is a subset of the set CON* of LAN*.

Definition: An *O-model* for an O-language LAN on a domain d is a maximally honest premodel on d for a language LAN* which is at least as articulate as LAN. A *primary O-model* for an O-language LAN is a maximally honest premodel of LAN.

O-Existence Theorem: It would be nice to know that there are such things as O-models. Let M be a model for the standard predicate calculus. That is, suppose we have: a domain d which contains no sentences of an O-language LAN, a mapping from the constants of LAN to the domain, and a mapping from the n-place predicate variables to subsets of d^n. We extend this to an O-model by supplying extensions for T and F. Simply, both are mapped to the null set. It is easy to see that the result is an O-model. We note that this coupled with the fact shown below that the axioms and rules of derivation of the standard predicate calculus are valid in O-model theory shows that we have a conservative extension in the sense that any sentence in the language of the standard predicate calculus is valid in O-model theory iff it is valid in the standard predicate calculus.

Definition: If a sentence W is true in every O-model, we say that W is O-valid.

Abbreviations: In an O-language LAN, instances of (A => B), instances of (A <=> B) and instances of (A)C abbreviate, respectively, instances of (-A v B), ((A => B) & (B => A)) and -(EA)-C.

CONSTRUCTING HONEST PREMODELS

Overview: O-models do not as obviously exist as traditional Tarskian models. However, the construction of various kinds of O-models turns out to be, although tedious, relatively easy. First, we present a general construction theorem whose content will be used repeatedly to construct honest premodels and from them O-models. The content of Theorem 1 turns on the satisfaction clauses and the fact that if the extensions of the truth and falsity predicates agree with them, the resulting model must be honest.

Definition: Let d be an arbitrary domain. Let M' be a mapping from the set of constants into d and from the set of n-place predicate variables into subsets of d^n. Let I be a pair of mappngs {I', g} where I' is a totally defined function from the set of individual variables of LAN into the domain d and g is a partially defined function from the set of individual variables and individual constants into the set of individual constants, where if A is an individual constant then g(A) = A and if A is an individual variable then g(A) is defined if there is at least one constant B mapped to the same item A* of d that A is mapped to, and in that case g(A) is a constant mapped to A*. We call I a content of M'.

Definition: We extend the notion of I-instance in the obvious way to contents.

Definition: Suppose for each content I(a), a set of wffs $S_{I(a)}$ is defined with the following properties (Note that a belongs to an index set IN indexing the set of contents):

1) If W is $BC_1...C_n$, an n-place predicate variable followed by n individual variables and/or individual constants, and the sequence of items of d which are the

images of the individual variables and/or individual constants under I(a) and/or M' respectively in the order of $C_1, \ldots, C_n$ is in the image of B under M', then W is in $S_{I(a)}$;

2) If A is a wff, then -A is in $S_{I(a)}$ iff A is not in $S_{I(a)}$. If A and B are wffs then (A v B) is in $S_{I(a)}$ iff either A or B is in $S_{I(a)}$;

3) For any individual variable or individual constant B and any wff C, if (B/C) is in $S_{I(a)}$ then, there is an I(a)-instance C' of C, g(B) is defined, g(B) is mapped to C' by M', (g(B)/C') is in $S_{I(a)}$, and either C' and Tg(B) are both in $S_{I(a)}$ or neither is in $S_{I(a)}$;

4) For any individual variable or individual constant B and any sentence C such that (g(B)/C) is in $S_{I(a)}$, T(B) is in $S_{I(a)}$ iff C is in $S_{I(a)}$ and FB is in $S_{I(a)}$ iff C is not in $S_{I(a)}$. Further, for all individual constants or individual variables B, if FB is in $S_{I(a)}$ then TB is not in $S_{I(a)}$;

5) If B is an individual variable, Y is a wff, W is (EB)Y, and either i) there is a content I(a') on M' which disagrees with I(a) on at most B and g(B), and Y is in $S_{I(a')}$, or ii) C is an individual constant, and Y[C:B] is in $S_{I(a)}$, then W is in $S_{I(a)}$;

suppose further that every $S_{I(a)}$ agrees on the sentences of the form (A/C), TA and FA, where A is any constant and C is any sentence. Then we say in such a case that the set of sets of wffs $\{S_{I(a)}$: a is in IN$\}$ is a <u>content set</u>. (In general we will abbreviate our notation for a content set by suppressing reference to the index set)

<u>Theorem 1</u>: Suppose a domain of d containing the sentences of LAN is given. Let M be a set of mappings: 1) from the set of n-place predicate variables to subsets of d^n, and 2) from the individual constants to d. Further, suppose the set of contents $\{I(a)\}$ is defined on d and a content set $\{S_{I(a)}\}$ exists. Let V be the set of sentences which occur in every $S_{I(a)}$ (note that V is nonempty since the $S_{I(a)}$ agree on the atomic sentences). If we augment M by adding a mapping from T to V and from F to the complement of V with respect to the sentences of LAN, then the resulting set of mappings M' is an honest premodel.

<u>Proof</u>: We show by induction on the number of connective and quantifier symbols in a wff that for all I(a), a wff S is in $S_{I(a)}$ iff it is satisfied in I(a) with respect to M'. The above formulation makes sense since we rely on the obvious fact that a content on M is equivalent to an interpretation on M'. This is simply because a content (or interpretation) turns only on the language LAN, the domain, and the mappings of the constants to the domain. These are the same for M and M'.

If S is an n-place predicate variable followed by n individual variables and constants then the result is immediate.

If S is -A then S is in $S_{I(a)}$ iff A is not in $S_{I(a)}$ iff A is not satisfied by I(a) iff -A is satisfied by I(a). If S is (A v B) then S is in $S_{I(a)}$ iff either A or B is in $S_{I(a)}$ iff either A or B is satisfied by I(a) iff (A v B) is satisfied by I(a).

If S is (A/B) where A is an individual constant or individual variable and B is a wff, then if S is in $S_{I(a)}$, the I-instance B' of B is defined, g(A) is defined and (g(A)/B') is in $S_{I(a)}$, and so (g(A)/B') is in V. Further, by assumption, either Tg(A) and B' are both in $S_{I(a)}$ (and therefore in V) or neither are. Therefore (g(A)/B') is satisfied by I(a) and so (A/B) is satisfied also. If S is satisfied, then g(A) is defined, the I-instance B' of B is defined, and (g(A)/B') is in V and therefore S is in $S_{I(a)}$.

If S is TA where A is an individual constant or individual variable then TA is in $S_{I(a)}$ iff Tg(A) is in $S_{I(a)}$ iff Tg(A) is in V. If there is no sentence B such that (g(A)/B) is in V, then TA is satisfied. If there is, then (g(A)/B) is in $S_{I(a)}$ and therefore, by definition, B is in $S_{I(a)}$ and therefore in V and so TA is satisfied.

If S is FA where A is an individual constant or individual variable, then FA is in $S_{I(a)}$ iff Fg(A) is in $S_{I(a)}$ iff Tg(A) is not in $S_{I(a)}$ iff Fg(A) is in V iff Tg(A) is not in V. If there is no sentence B such that (g(A)/B) is in V, then FA is satisfied. If there is, then (g(A)/B) is in $S_{I(a)}$ and therefore, by definition, B is not in $S_{I(a)}$

and therefore not in V and so FA is satisfied.

If S is (EB)Y, S is satisfied in I(a) iff: i) there is an interpretation I(a') which differs from I(a) on at most B and g(B) and in which Y is satisfied. But this holds iff Y is in $S_{I(a)}$ and that implies S is in $S_{I(a)}$. Or ii) there is a constant C such that Y[C:B] is satisfied in I(a). But this holds iff Y[C:B] is in $S_{I(a)}$ and that implies S is in $S_{I(a)}$. On the other hand, if S is in $S_{I(a)}$ then either there is a constant C such that Y[C:B] is in $S_{I(a)}$ or Y is in $S_{I(a')}$ for some I(a') which differs from I(a) on at most B and g(B). This gives us the converse direction by an argument similar to that immediately above.

Since, therefore, a sentence S is satisfied iff it is in the extension of T, we have shown what was desired.

EXISTENCE THEOREMS

<u>Overview</u>: We are interested here in showing Theorem 3. It is not clear initially that given an arbitrary domain and an arbitrary mapping of the constants into the domain d and the n-place predicate variables to subsets of d^n, that there is an O-model with this domain and with these mappings of predicates and constants. It might be, that is, that the requirements on the extensions of the truth and falsity predicates are such that they place constraints on possible mappings and domains. This would be an unpleasant result since it would cost the entire system

some generality. Theorem 3 shows no such thing occurs. The general idea of the proof is easy to grasp. The assumption that there is no O-model containing such a domain and such mappings would lead to an infinite chain of honest premodels each satisfying a strictly larger number of ostensives than ones earlier in the chain. We use such a chain to construct an honest premodel that satisfies a maximal number of such ostensives--an honest premodel, in fact, that is a limitpoint of the chain.

Lemma 1: Let d be an arbitrary domain of an O-language LAN containing the sentences of LAN, and let M' be a mapping from the set of constants of LAN into d and from the set of n-place predicate variables of LAN to subsets of d^n. There is an honest premodel of LAN which agrees with M'.

Proof: Let I(a) be a content on M'. We define a set $S_{I(a)}$ of wffs recursively:

Clauses 1, 2, and 5 are identical to clauses 1, 2, and 5 of the definition of a content-set.

Clause 3 reads: For any individual constant or individual variable B and any wff C, (B/C) is not in $S_{I(a)}$.

Clause 4 reads: For any individual constant or individual variable A, TA and FA are not in $S_{I(a)}$.

The defined set of sets of wffs $\{S_{I(a)}\}$ clearly is a content-set. Therefore, invoking O-theorem 1, we know there is an honest premodel M which agrees with M' and has as the extension of T the set of sentences in every $S_{I(a)}$. This gives us the desired result.

Theorem 2: Suppose d is an arbitrary domain of LAN containing the sentences of LAN, and let M' be a mapping from the set of constants of LAN into d which maps only a finite number of constants to sentences in d and a mapping from the set of n-place predicate variables to subsets of d^n. Then there is a primary O-model of LAN which agrees with M'.

Proof: By lemma 1 there is an honest premodel P of LAN which agrees with M'. If P is not an O-model this is because there is an honest premodel P' which agrees with M', satisfies all ostensives satisfied by P and further satisfies some additional ostensives of the language LAN. Since the set of constants mapped by M' to sentences is finite, any honest premodel of LAN agreeing with M' can only satisfy at most a subset of a finite set of ostensives of LAN. Therefore there must be an honest premodel Q which is maximal in the sense that no honest premodel of LAN agrees with M' but satisfies more ostensives than Q does. But then Q is the desired primary O-model of LAN.

Theorem 3: Suppose d is an arbitrary domain of an O-language LAN containing the sentences of LAN, and let M' be a mapping from the set of constants into d and a mapping from the set of n-place predicate variables to subsets of d^n. Then there is a primary O-model of LAN which agrees with M'.

<u>Proof</u>: Let P be an honest premodel of LAN which agrees with M'. Let R be an enumeration of all the ostensive-sentences of LAN. We define a <u>P-chain</u> as follows: P is the first member of the chain. Let N be the nth member of the P-chain, we define the n+1th member, if it exists, as follows. If N is a maximal premodel, we're done. Otherwise, there is an earliest member r of R such that r is not satisfied in N, but there is a premodel N+1 which agrees with M' and is such that r is satisfied in N+1 and every ostensive-sentence satisfied in N is satisfied in N+1. We let N+1 be the n+1th member of the P-chain. We note that the axiom of choice is used here to construct a P-chain. Further, we note that if a P-chain is finite, the last member of it is the desired maximal honest premodel. We consider the case where the P-chain is infinite.

Clearly any P-chain is well-ordered.

Claim: Any P-chain converges to a maximal honest premodel P*. We show this claim by using the P-chain to construct a set of sentences V which contain all ostensive-sentences satisfied by any member of the P-chain, using this as T's extension and showing that the resulting defined premodel P* is honest. The proof that P* is maximal is the following. Suppose there is an honest premodel which agrees with M' but satisfies ostensive-sentences not satisfied by P*. Choose one. This ostensive-sentence occurs somewhere in R and therefore at a certain point in the construction of the P-chain, an honest premodel was added in which this ostensive-sentence is satisfied. But every ostensive-sentence satisfied by any premodel in the P-chain is satisfied by P* and so this is impossible, P* is maximal and is therefore a primary O-model.

Now we construct P*. Clearly any infinite P-chain is well ordered and associates with each premodel in it an integer. For the duration of this proof if A is an individual constant, call a set of sentences of the form (TA, -FA), (-TA, FA), or (-TA, -FA) a _pair_.

A pair is _stubborn in H_ if for every integer j, there is an integer k, k > j such that the pair is satisfied in the premodel which is associated with k. We note that for any constant A, there is at least one pair for it which is stubborn. Now we construct our set V. We do so by constructing a sequence of subsets H(i) of H inductively. These subsets inherit the ordering of H. H(1) = H. Now assume R is an enumeration of all the individual constants of LAN. Consider H(n) and the nth constant A of R. Suppose there is a sentence B such that (A/B) is satisfied by some premodel in H(n). Choose a pair p* for A which is stubborn in H(n). H(n+1) is the set of premodels of H(n) which satisfy p* and (A/B). If no such sentence B exists for A then choose a pair p* for A which is stubborn in H(n). H(n+1) is just the set of premodels of H(n) which satisfy p*. We note by the stubbornness of p* that H(n+1) is nonempty. Let I be a content defined on M'.

Definition: Given I, a content, we say that a wff W settles down if there is an n such that W is satisfied by I with respect to every member of H(n). We say that it settles down satisfied when it is satisfied in those members and that it settles down unsatisfied otherwise.

Claim: Given I, every wff settles down. We show this by induction.

1) Suppose W is an n-place predicate variable followed by n individual variables and/or individual constants. Then since every member of H agrees with M', and the satisfaction of W turns only on M', W settles down as of H(1).

2) Suppose A is not TA or FA, then -A settles down when A does. When A and B both settle down then (A v B) settles down.

3) Let A be an individual constant or individual variable and B a wff. Let B' be the I-instance of B, if it is defined. Then (g(A)/B') is either defined and satisfied by a member of H or it is not. If so, then g(A) occurs at some point n in the enumeration R. Thus (A/B) settles down as of n. Otherwise it settles down as of 1.

4) If A is the nth member of R then TA, FA, -TA, -FA

settle down at n+1. If A is an individual variable then TA, FA, -TA, -FA settle down as of 1 if g(A) is not defined. Otherwise they settle down when Tg(A) does.

5) Suppose A is an individual variable and B a wff. Either there is a constant C such that B[C:A] settles down satisfied or there is an I' such that B settles down satisfied with respect to I' or neither of these case occur. If they do occur then (EA)B settles down satisfied. Otherwise (EA)B settles down unsatisfied at 1.

This fact motivates our definition of V.

Definition: V is all sentences which settle down satisfied with respect to every content I.

Now consider the following premodel P*: M' plus the mapping of T to V and F to the remaining sentences of LAN.

Claim: P* is honest.

Let I be an interpretation. We show a wff W is satisfied in P* iff it settles down satisfied with respect to I.

1) If W is an n-place predicate variable followed by n individual variables and/or individual constants then the result is obvious.

2) If W is -A then W is satisfied iff A is not satisfied iff A settles down unsatisfied iff W settles down satisfied.

If W is (A v B) then W is satisfied iff A or B is satisfied iff A or B settles down satisfied iff W settles down satisfied.

3) Suppose W is (A/B) and (A/B) is satisfied. Then B', the I-instance of B, is defined. Then (g(A)/B') is in T's extension, i.e., it is in V, i.e., it settles down. Therefore (A/B) settles down satisfied. Conversely, suppose (A/B) settles down unsatisfied. Then we show (A/B) is in T's extension. We need to show that (g(A)/B') is in T's extension and either Tg(A) and B' are both in T's extension or both not in T's extension. Well, suppose g(A) occurs in the nth place in R. Then at n+1 one pair (Tg(A), Fg(A)), (-Tg(A), Fg(A)), (Tg(A), -Fg(A)), (-Tg(A), -Fg(A)) is satisfied for all premodels in H(n+1). Further (g(A)/B') is also satisfied in every premodel in H(n+1). We also know by the fact that every sentence settles down that there is an m at which B' settles down. Therefore at H(max(n+1, m)) B' and Tg(A) have settled down. Since all members of H(max(n+1, m)) are honest, Tg(A) and B' are either both satisfied or both not in every member of H(max(n+1, m)). Therefore both are in V or both not. That is, they are both in T's extension or both not. So (g(A)/B') is satisfied (since it is in V also).

4) Suppose TA is satisfied. Then Tg(A) is in T's extension and therefore it settles down satisfied. Conversely, suppose it is in V. Then it settles down satisfied for some n. First suppose there is a B such that (g(A)/B) is satisfied. Then in H(n+1) B settles down satisfied also (since (g(A)/B) settles down satisfied at n

and all premodels in H(n+1) are honest). Therefore B is in T's extension in P* and thus Tg(A) is satisfied. If there is no B such that (g(A)/B) is satisfied, then since Tg(A) is in T's extension it is satisfied.

FA is handled similarly.

5) Suppose (EA)B is satisfied. Then there is either 1) an interpretation I' such that B is satisfied, but then B settles down satisfied by assumption and therefore (EA)B settles down satisfied, or 2) there is a constant C such that B[C:A] is satisfied and therefore by assumption, B[C:A] settles down satisfied and thus so does (EA)B. Conversely, suppose (EA)B settles down. Then (1) there is a constant C such that B[C:A] settles down. That is, B[C:A] is satisfied by I in P*, or (2) there is an interpretation I' such that B settles down in I'. That is, B is satisfied by I' with respect to P*, and therefore (EA)B is satisfied by I with respect to P*.

This shows that P* is honest and by the remarks above concludes the proof of O-theorem 3.

REFERENCE LEMMAS

<u>Overview</u>: We are working our way towards completeness and consistency proofs for this model theory in terms of axioms to be specified later. Some of these reference lemmas (such as 2 and 3) are used for this. Lemma 4 is an interesting result (it is actually a negative existence

theorem) showing that under the condition that every sentence has a name, there must be examples of transparent reference. It is also used below to give a quick proof that restricting our model theory to the primary O-models will cost us compactness.

Lemma 2: Suppose an interpretation I on a model M with domain d assigns to A, an individual variable or individual constant, and to B, an individual variable, the same object in d, and g(A) = g(B) if g(B) is defined. Then if C and E are wffs such that E = C[B:A], we have that C is satisfied by I iff E is satisfied by I.

Proof: By induction on the construction of the wffs.

It is not true, however, that if A and B are constants, and A and B are mapped by the model M to the same item of d, then if C and E are wffs such that E = C[B:A], that I satisfies C iff it satisfies E. For example, if we have (G/TA), we will not have (G/TB).

Lemma 3: Let Q be an individual constant. Let M be an O-model containing at least one nonsentence. Suppose further that Q is a constant mapped to a non-sentence. Suppose also that Q and M are such that no satisfied ostensive-sentence in which Q appears is one in which Q appears in the second place. Given an interpretation I, define a Q_I transform of a wff A to be a wff A' such that if $B_1, \ldots, B_n$ are all and only the free variables of A mapped to items of d which M maps no constants to, then A' is $A[Q, \ldots, Q{:}B_1, \ldots, B_n]$. Further, suppose that for every W and I such that if W is an n-place predicate followed by n individual variables and/or individual constants and I is an interpretation then W is satisfied by I iff its Q_I transform W' is satisfied by I. If S is a wff and I is an interpretation on M, then I satisfies S iff I satisfies its Q_I transform S'.

<u>Proof</u>: We show this for every interpretation I by a straightforward induction based on the satisfaction clauses and the formation rules:

If W is an n-place predicate variable followed by n individual variables or individual constants, the result is immediate.

Suppose A and B are wffs, W is (A v B) and I satisfies W. By our inductive assumption, the Q_I transforms A' and B' are such that I satisfies A iff it satisfies A' and I satisfies B iff it satisfies B'. But then by the satisfaction clauses, it satisfies (A v B) iff it satisfies (A' v B').

Suppose A is a wff, W is -A and I satisfies W. By our inductive assumption, the Q_I transform A' is such that I satisfies A iff it satisfies A'. But then the same holds of -A and -A'.

Suppose A is an individual variable or individual constant, B is a wff and W is (A/B). Then by the satisfaction clauses, if (A/B) is satisfied, there is an I-instance of it which is satisfied. But then the Q_I transform of W is just itself and thus it is satisfied trivially. On the other hand, if (A/B) is not satisfied, the Q_I transform of it is not either for no Q_I transform of any ostensive that is not the ostensive itself is satisfied.

Similarly, suppose A is an individual variable or individual constant. Then if TA or FA is satisfied, Tg(A) or Fg(A) is. But then the Q_I transforms of TA or FA are just TA or FA and these are satisfied trivially. On the other hand, if TA or FA is not satisfied, their Q_I transforms aren't either, since, in any case, TQ and FQ aren't satisfied.

Finally, suppose Y is a wff, A an individual variable, and W is (EA)Y. Suppose (EA)Y is satisfied. By the satisfaction clauses, either 1) there is an interpretation I' which differs from I on at most A and g(A) and satisfies Y or 2) there is a constant B such that Y[B:A] is satisfied by I. As regards (2), our inductive assumption then implies that the Q_I transform of Y[B:A] is satisfied. But then the Q_I transform of (EA)Y is satisfied also. As regards (1), by our inductive assumption, Y', which is the $Q_{I'}$ transform of Y, is satisfied by I'. But then the Q_I transform of (EA)Y is satisfied by I.

Conversely, suppose the Q_I transform of (EA)Y, (EA)Y', is satisfied. Then either 1) there is an interpretation I' which differs from I on at most A and g(A) and satisfies Y' or 2) there is a constant B such that Y[B:A]' is satisfied by I. As regards (2), our inductive assumption then implies that Y[B:A] is satisfied and therefore that (EA)Y is satisfied. As regards (1), if Y' is not the $Q_{I'}$ transform of Y, its $Q_{I'}$ transform Y'' is satisfied and is also the $Q_{I'}$ transform of Y. Therefore by our inductive assumption again, (EA)Y is satisfied.

This concludes the proof.

Lemma 4: Let LAN be an O-language. Let d be a domain containing the sentences of LAN. There is no primary O-model for LAN in which every sentence of d has a constant mapped to it yet no ostensive-sentence is satisfied.

Proof: We show that such an O-model cannot exist. Let a set of mappings M be given, where M maps the set of constants onto the sentences of LAN and maps the n-place predicate variables to subsets of d^n. Choose a sentence S which is a logical truth of the standard predicate calculus. Also choose a constant A which is mapped to it. We define a set of sets of sentences $\{S_{I(a)}\}$ for every content I(a) on M recursively:

Clauses 1, 2, and 5 are identical to clauses 1, 2, and 5 of the definition of a content-set.

Clause 3 reads: For any individual constant or individual variable B, and any wff C the ostensive (B/C) is in $S_{I(a)}$ only if the I-instance C' of C is defined, C' is S, and g(B) is A.

Clause 4a reads: TB is in $S_{I(a)}$ for all individual constants B, FB is not in $S_{I(a)}$ for any individual constant B.

Clause 4b reads: For TB, B an individual variable, TB is in $S_{I(a)}$ if Tg(B) is in $S_{I(a)}$ by clause 4a. For FB, B an individual variable, FB is in $S_{I(a)}$ if Fg(B) is in $S_{I(a)}$ by clause 4a.

The defined set of sets of sentences $\{S_{I(a)}\}$ is clearly a content-set. Invoking Lemma 1 gives us an honest premodel of LAN which agrees with M. Since for any primary O-model of of a language LAN, there is a mapping M which agrees with it on the domain, the predicate variables, and the mappings of the constants into the domain, it follows that there can be no O-model mapping a constant to every sentence in which no ostensive is satisfied (since on the assumption there is one, we can find an honest premodel agreeing with it which satisfies at least one ostensive-sentence).

CHAPTER 3

DEFINITIONS

Definition: Let W be a wff and let all the free variables of W be among $A_1, \ldots, A_n$, then $(A_1)\ldots(A_n)W$ is a closure of W.

Definitions: Let "p", "q",..., be sentential variables. If A is a sentential variable then A is a **sentential schema**. If A and B are **sentential schemas** then (A v B) and -A are **sentential schemas**. Anything is a **sentential schema** only if it is so by the above two conditions. A **tautology schema** is one which is true under all assignments of truth values to the variables of the schema. A **tautology-substitution instance** is a wff gotten by uniformly substituting wffs of LAN for the sentential variables of the tautology schema.

Definition: Two wffs A and B are alphabetic variants if $C_1, \ldots, C_n, D_1, \ldots, D_n$ are individual variables, A is $B[C_1, \ldots, C_n : D_1, \ldots, D_n]$, and further no C_i occurs in B.

AXIOM SYSTEM O

Inference Rules:

R1) If a closure of A and a closure of (A => B) are theorems, then any closure of B is a theorem.

R2) If A is any individual variable and B and C are any wffs, then from any closure of B => C we may deduce any closure of (EA)B => C, provided that A does not occur free in C.

Axioms:

A1) If A is a closure of a tautology-substitution instance, then A is an axiom.

A2) If A is any individual variable, B any wff and C a wff differing from B in that every free occurrence of A has been replaced by either an individual constant or individual variable E, and further B has the same number of free occurrences of variables as C if E is an individual variable, then any closure of C => (EA)B is an axiom.

A3) If A is a wff and B is an individual variable then any closure of (B)((B/A) => ((TB v FB) & (TB <=> A))) is an axiom.

A4) If A is an individual variable then (A)-(TA & FA) is an axiom.

A5) If A is an individual variable, B and C distinct wffs which are not alphabetic variants of each other, then any closure of (A)-((A/B) & (A/C)) is an axiom.

A6) If A and B are individual variables, $A_1,\ldots,A_n$, n individual variables and/or individual constants among which A occurs, $B_1,\ldots,B_n$ n individual variables and/or individual constants among which B occurs, and the A_i's and B_i's are such that for all A_i which are not A, A_i is B_i and for A_i which are A, B_i is either A or B, C is any wff, and P is any n-place predicate variable, then any closure of ((A/C) & (B/C)) => $(PA_1\ldots A_n$ <=> $PB_1\ldots B_n)$ is an axiom.

A7) If $S_1,\ldots,S_n$ are wffs, A is an individual variable, $R_1,\ldots,R_m$ are wffs in which A occurs freely, and $B_1,\ldots,B_m$ distinct individual constants, then any closure of $-(A)((A/S_1)$ v...v (A/S_n) v (B_1/R_1) v...v $(B_m/R_m))$ is an axiom.

CONSISTENCY OF AXIOM SYSTEM O RELATIVE TO MODEL THEORY O

We show that each of the axioms and inference rules hold in any O-model:

On (R1): Let A and B be wffs, and let $A_1,\ldots,A_n$, $B_1,\ldots,B_m$, $C_1,\ldots,C_p$, be individual variables such that all the free variables of A are among $A_1,\ldots,A_n$, all the free variables of A and B are among $B_1,\ldots,B_m$, and all the free variables of B are among $C_1, \ldots, C_p$. Suppose further that $(A_1)\ldots(A_n)A$ and $(B_1)\ldots(B_m)(A \Rightarrow B)$ are true in an O-model M and $(C_1)\ldots(C_p)B$ is false. We show this is impossible. We have $(EA_1)\ldots(EA_n)-A$ and $(EB_1)\ldots(EB_m)-(A \Rightarrow B)$ not satisfied in M and $(EC_1)\ldots(EC_p)-B$ satisfied in M. Therefore, there is an interpretation I and a (possibly null) sequence of individual constants $D_1, \ldots, D_q$ such that -B' (where B' comes from B via a substitution of $D_1, \ldots, D_q$ for $C_{i1}, \ldots, C_{iq}$) is satisfied in I with respect to M. But clearly, -A' and -(A' => B') (where A' is gotten from A by substituting those D_j for those C_{ij} which are free in A) are not satisfied by I (on pain of violating O.5) and this is impossible.

On (R2): Let $A_1, \ldots, A_n$, $B_1, \ldots, B_m$, and A be individual variables, and let B and C be as described in (R2) of axiom system O. Further, suppose all the free variables of B and C are among $A_1, \ldots, A_n$ and all the free variables of (EA)B and C are among $B_1, \ldots, B_m$. Finally, suppose that $(A_1)\ldots(A_n)(B \Rightarrow C)$ is true in an O-model M, and $(B_1)\ldots(B_m)((EA)B \Rightarrow C)$ is false in M. We show this is impossible. Clearly we have $-(EA_1)\ldots(EA_n)-(B \Rightarrow C)$ true in M and $(EB_1)\ldots(EB_m)-((EA)B \Rightarrow C)$ true in M. Therefore there is an interpretation I and a (possibly null) sequence of individual constants $C_1, \ldots, C_p$ such that -((EA)B' => C') (where (EA)B' and C' are gotten from (EA)B and C by substituting $C_1, \ldots, C_p$ for all free occurrences of $A_{i1}, \ldots, A_{ip}$ in (EA)B and C) is satisfied in I with respect to M. Therefore ((EA)B' & C') is satisfied in I. Therefore there is an I' differing from I on at most A and g(A) and in which (B'' & C') (where B'' may differ from B' in having an individual constant for free occurrences of A in B') is satisfied. (We know this since A does not occur free in C') But therefore $(EA_1)\ldots(EA_n)$-(B => C) is satisfied in M, which is impossible.

On (A1): Let W be a tautology substitution instance, $A_1, \ldots, A_n$, individual variables among whom are all the free variables of W, and let A be $(A_1)\ldots(A_n)$W. Suppose M is an O-model in which A is not true. Then $(EA_1)\ldots(EA_n)$-W is true in M. Thus there is an interpretation I in which -W' (where W' is gotten from W by substituting a possibly null sequence $B_1, \ldots, B_m$ of constants for $A_{i1}, \ldots, A_{im}$ in W) is satisfied. But W' is a tautology-substitution instance and by the satisfaction clauses this is impossible.

On (A2): Suppose that A, B and C are as described in (A2) of axiom system O, $A_1, \ldots, A_n$ are individual variables among whom are all the free variables of C => (EA)B, and suppose M is an O-model in which $(A_1)\ldots(A_n)$(C => (EA)B) is not true. We show this is impossible. We have $(EA_1)\ldots(EA_n)$-(C => (EA)B) is true in M. Therefore there is an interpretation I and a (possibly null) sequence of constants $B_1, \ldots, B_m$ such that -(C' => (EA)B') (where B' and C' are gotten from B and C by substituting $B_1, \ldots, B_m$ for free occurrences of $A_{i1}, \ldots, A_{im}$) is satisfied in I. Therefore (-(EA)B' & C') is satisfied in I. Case 1: either C differs from B in having every free occurrence of A replaced by a constant D, or all free occurrences of the individual variable A in B have been replaced in C' by a constant. Then (EA)B' is satisfied and this is impossible. Case 2: C' has an individual variable D replacing every free occurrence of A in B'. So choose an I' differing at most from I in that A is assigned the same element as D, and g(A) = g(D) if the latter is defined. Thus by lemma 2, since C' is satisfied so is B' and therefore (EA)B' is satisfied in I which is impossible.

On (A3): Suppose A and B are as stated in (A3) of axiom system O, $A_1, \ldots, A_n$ are individual variables among whom are all the free variables of A and suppose there is an interpretation I and a (possibly null) sequence of individual constants $C_1, \ldots, C_m$ such that (EB)-((B/A') => ((TB v FB) & (TB <=> A'))) (where A' is gotten from A by substituting $C_1, \ldots, C_m$ for free occurrences of $A_{i1}, \ldots, A_{im}$) is true in an O-model M. We show this is impossible. By O.3 we can assume without loss of generality that C_j is $g(A_{ij})$ and that A' is a sentence. Case 1: there is an individual constant C such that -((C/A') => ((TC v FC) & (TC <=> A'))) holds. We show this case is impossible by showing that if (C/A') holds then ((TC v FC) & (TC <=> A')) holds. We have by O.3 that TC <=> A' holds. Either A' is satisfied or it is not. If so, TC is satisfied by TC <=> A'. Further FC is not satisfied by O.4, and the fact that A' is satisfied. Suppose A' is not satisfied. Then A' is in F's extension. By O.4 FC is satisfied and TC is not. Thus in either case we have shown what was desired. Case 2: there is an interpretation I such that -((B/A) => ((TB v FB) & (TB <=> A'))) is satisfied under I. Therefore by O.3, g(B) is defined and (g(B)/A') is satisfied. By the reasoning in case 1 we have ((Tg(B) v Fg(B)) & (Tg(B) <=> A')). But this is impossible.

On (A4): This follows by O.4.

On (A5): this trivially follows from O.3 and the fact that an O-model assigns one and only one item of d to each constant.

On (A6): This follows from O.1 and O.3.

On (A7): Suppose there is an O-model in which a closure of (A)((A/S_1) v...v (A/S_n) v (B_1/R_1) v...v (B_m/R_m)) (which we abbreviate as W) holds. Then there is a sequence of constants C_1,..., C_p such that an instance W' of W gotten by substituting C_i for all free instances of the ith quantifier variable of the closure prefix of W in W is satisfied. Choose m+l sentences of LAN which are not instances of the S_i's and consider m+l interpretations I_j where A is mapped by I_j to the jth sentence of the above m+l. Clearly, in each of these interpretations, $(A/S_i)'$, for each i, cannot be satisfied since g(A) is not mapped to the I_j-instance of S_i. Further, g(A) with respect to I_j does not equal g(A) with respect to I_k for all j not equal to k, and there are m+l of them. Thus, suppose $(B_i/R_i)'$ is satisfied by some I_j. Then it can be satisfied by no other (on pain of violating axiom 6). But therefore W' is not satisfied, contrary to assumption.

Definition: An O-derivation is a finite sequence of sentences of some O-language LAN, each of which is either an axiom of O, follows from two sentences earlier in the sequence by (R1) of axiom system O or follows from one sentence earlier in the sequence by (R2) of axiom system O.

Definition: An O-theorem is any sentence A of an O-language LAN where a derivation exists which A is a member of.

Definition: A wff A is the same predicate-type as a wff B iff there is a set of individual constants or individual variables $A_1,\ldots, A_n, B_1,\ldots, B_n$, such that A is $B[A_1,\ldots, A_n:B_1,\ldots,B_n]$.

Overview: Theorem 4 gives us part of the motivation for not restricting our set of models for a language to the ones traditionally acceptable. It shows the failure of compactness when the model theory is so restricted.

Theorem 4: Let LAN be an O-language with a countable number of constants. Let S be the following set of sentences of LAN: (A)(TA v FA), and all closures of the wffs: (B)-(B/W), where W is any wff of LAN. Then there is no primary O-model of LAN in which every sentence of S is satisfied, although for any finite subset of S there is a primary O-model in which S is satisfied.

Proof: Suppose M is a primary O-model of LAN in which S is satisfied. Then, no ostensive-sentence in LAN is satisfied although every sentence has a constant mapped to it by LAN. But this is impossible by Lemma 4. On the other hand, suppose S' is a finite subset of LAN. Let d, a domain, be given containing all and only the sentences of LAN. Divide the constants of LAN into two infinite lists L' and L''. Map the constants L' onto the sentences of LAN in a 1-1 manner. Now, the wffs appearing in the second place of the ostensives in S' have only a finite number of types. In particular, we can find a wff among the following: FA, FA & FA, FA & FA & FA, etc., such that the type of the wff does not occur as a constituent of any sentence occurring in the second place of any ostensive in S' (In particular, just choose one with more occurrences of connectives than any sentence in S'. For abbreviatory purposes, call any sentence of that type with individual constant or individual variable A, F*A). Now let S_1,..., S_n,..., be a list of all the instances of the wffs that appear in the second place of ostensives in S', and let A_1,..., A_n,..., be the constants of L' mapped to each S_i. Put the constants of L' and L'' in a 1-1 correspondence and map B_i corresponding to A_i to $F*B_i$ & $((A_i/S_i) \vee -FA_i)$. Let M be the above mapping for the constants onto LAN and any mapping of the n-place predicate variables to subsets of d^n. Let I(a) be an interpretation on d. We define the set of sets of sentences $S_{I(a)}$ recursively:

Clauses 1, 2, and 5 are the same as those in the definition of a content-set.

Clause 3a reads: If A is a constant and B is a sentence, then (A/B) is in $S_{I(a)}$ iff A is B_i and B is $(F*B_i$ & $((A_i/S_i) \vee -FA_i))$.

Clause 3b reads: If A is an individual variable and B a wff, then (A/B) is in $S_{I(a)}$ iff (g(A)/B') where B' is the I-instance of B is defined, and is in $S_{I(a)}$ by clause 3a.

Clause 4 reads: FB is in $S_{I(a)}$ for all individual constants and individual variables B, TB is not in $S_{I(a)}$ for all individual constants and individual variables B.

It is easy to see that the resulting defined set of sets of sentences $\{S_{I(a)}\}$ is a content-set. Therefore, invoking Theorem 2, we have an honest premodel P of LAN which agrees with the above mapping M. Further, invoking Theorem 3 we have a primary O-model Q of LAN which agrees with M. Note that as the ostensives satisfied in P must be satisfied in Q, no ostensive of S' is satisfied. Further (A)(TA v FA) is satisfied since for the ostensives defined to hold, FA must be satisfied for all constants A.

STRONG COMPLETENESS FOR O-LANGUAGES: AXIOM SYSTEM O AND MODEL THEORY O

Overview: In general a strong completeness theorem is an existence theorem: given a consistent set of sentences, a model is shown to exist which satisfies those sentences. Recall that in this approach, models for a language L may contain sentences in their domain which are sentences of a language L' properly containing L. Since such models are necessary to exclude certain implications, unsurprisingly, such models must be constructed to satisfy certain sets of consistent sentences when a completeness proof is carried out. This makes even a Henkin-style proof more complicated than is the case with the standard predicate calculus. In fact, the our proof divides into two cases. The first is where a standard Henkin-style proof goes through and the second is where it fails. An interesting aspect of the second case is that the completeness theorem for the standard predicate calculus is used. This is one place where the fact that the axioms of the standard predicate calculus hold of model theory O plays a fundamental role.

Definition: Let M be a set of axioms. A sentence A is M-consistent iff -A is not a theorem of axiom system M.

Definition: Let M be a set of axioms. A set of sentences is M-consistent iff there is no finite list of sentences $A_1,\ldots, A_n$ contained in it such that $-(A_1 \& \ldots \& A_n)$ is a theorem of M.

Definition: If a sentence W is of the form (EA)B => B[C:A], where A is a variable, C is a constant, and (EA)B is a sentence in which C does not occur, then we call W an E-formula.

We say that W is an E-formula with respect to C.

Definition: An E-form is a collection of E-formulas which have the same antecedent.

The proof of strong completeness follows the standard pattern. We show that for any set of O-consistent sentences SEN, there is an O-model in which they are all true.

First let Z be an infinite list of new constants not in

LAN. Let LAN* be a language at least as articulate as LAN containing in its set CON* the constants of Z. Further, let LAN* contain the same predicate variables as LAN. Let X be a list of all the sentences of LAN*. Let Y be some ordering of all the E-forms of LAN*.

Definition: Let M be a set of axioms and L an O-language. A set of sentences of L is maximally M-consistent iff it is M-consistent and is not contained properly in an M-consistent set of sentences of L.

We construct a maximal O-consistent set containing SEN as follows. First utilizing a constant c* which does not appear in SEN, add the sentences -(c*/S) for every sentence S in LAN* and the sentences -(EA)(A/S) for every sentence S which c* appears in. We also add the sentences -Tc* and -Fc* to SEN. Next, augment SEN by adding, in the order of Y, one for each E-form, an E-formula with respect to a constant which does not appear in SEN, is not in any E-formula added already to SEN and is not c*. Let the resulting set be called V_1. Now;

Case 1: V_1 is consistent. Increase the set to a maximally O-consistent set by constructing a sequence of sets $V_1, \ldots, V_n, \ldots$, where V_{n+1} equals V_n plus the nth member of X if the resulting set is O-consistent, and V_{n+1} equals V_n otherwise. Finally, let V be the union of all the V_n's.

O-Maximality Lemma: Since axiom system O contains the standard sentential calculus, any maximal O-consistent set V has the following properties:

1) For any sentence A, V does not contain both A and -A.

2) For any sentence A, V contains either A or -A.

3) For any sentences A and B, if A and (A => B) are in V, then B is in V.

It is trivial (and standard) that V is consistent and maximal provided that V_1 is consistent. So we use V to construct a premodel M and show that, in fact, M is an O-model.

Let d be the union of all the sentences of LAN* and all constants A which do not appear in the first place of any ostensive-sentence in V and for which neither FA nor TA are in V. We define the premodel M as follows: any constant A is mapped to a sentence B if (A/B) is in V and is mapped to FA if either TA or FA are in V but no ostensive with A in its first place is in V; and is mapped to itself otherwise. We note that by construction, c* is mapped to itself. For each n-place predicate variable B, it is mapped to that set of n-tuples (R_1,..., R_n) such that $BA_1 \ldots A_n$ is in V and A_i is mapped to R_i by the above mapping for constants, or R_{i1},..., R_{in} are all members of the n-tuple which have no constants mapped to them by the above mapping for constants, but the n-tuple $(R^*_1, \ldots, R^*_n)$ gotten from $(R_1, \ldots, R_n)$ by substituting c* for each R_{ij} is such that B followed by the sequence of constants $C_1, \ldots, C_n$ where each C_i is mapped to R^*_i is in V. (We note that this definition for the predicate variables is well-defined even though it is possible for more than one constant to be mapped to the same item of d--i.e., if two constants are mapped to the same sentence. But by axiom 7, the predicates must agree on them.)

The predicate constant T is mapped to V and F is mapped to the remaining sentences of d.

This clearly defines a premodel. We want to show that in fact it defines an O-model. We shall do so as follows. First we will show that a sentence is satisfied by an interpretation I on M iff it is in V. From this we can show that M is an O-model. For V is precisely the extension of T, and this satisfies the first condition. Let R be an ostensive not satisfied in M. The constant A appearing in R is either mapped by M to itself or to FA. In the first case there certainly cannot be a premodel agreeing with M in which R is satisfied. The second case is taken care of by the following lemma:

Lemma 5: (A)-(A/FA) is an O-theorem.

Proof sketch:

1) (A)(B)((B/FA) => ((TB v FB) & (TB <=> FA)) Axiom A3, substituting FA for sentential variable A.

2) (A)((A/FA) => ((TA v FA) & (TA <=> FA)) (using

universal instantiation, which is easily shown from A2)

3) (A)-(A/FA)   by A4, A1 and R1.

So to show that a sentence is satisfied by M iff it is in V is to conclude this case. We do so by induction on the number of places at which connective or quantifier symbols appear.

1) Suppose W is $BA_1 \ldots A_n$, an n-place predicate variable followed by n individual constants. Then it is, by construction, satisfied iff it is in V.

2) W is -A or (A v B), where A and B are sentences. This is standard and follows from the O-maximality lemma.

3) W is (A/B) where A is an individual constant and B is a sentence. By construction (A/B) is in V iff (A/B) is in T's extension and B is assigned by M to A. By axiom (A3) (reversing the roles of A and B) and the O-maximality lemma, if (A/B) is in V, then (TA <=> B) is in V. Thus by the O-maximality lemma again, either a) TA and B are in V, in which case they are both in T's extension, or b) TA and B are not in V, in which case they are both in F's extension. In either case (A/B) is satisfied.

4) Suppose TA is in V. First suppose there is a sentence B such that (A/B) is in V. Then by axiom (A3) and the O-maximality lemma, B is in V. Since B is in T's extension and (A/B) is in T's extension, TA is satisfied. On the other hand, if (A/B) is not in V for any B, A is

mapped to FA, and therefore TA is satisfied. Suppose, conversely, that TA is satisfied. Then TA is in T's extension and therefore in V.

Suppose FA is in V. First suppose there is a sentence B such that (A/B) is in V. Then by axiom (A3) and the O-maximality lemma, B is not in V. Since B is in F's extension and (A/B) is in T's extension, FA is satisfied. On the other hand, if (A/B) is not in V for any B, A is mapped to FA and therefore FA is satisfied (since by axiom 9, TA is not in V). Suppose, conversely, that FA is satisfied. Then FA is in T's extension and therefore in V.

5) Suppose A is an individual variable and Y is a wff such that the sentence (EA)Y is in V. Then there is an E-formula ((EA)Y => Y[C:A]) in V. By our inductive assumption, Y[C:A] is satisfied. Therefore (EA)Y is satisfied by the satisfaction clauses. Suppose conversely, that (EA)Y is satisfied. Then either a) there is an individual constant B and Y[B:A] is satisfied or b) there is an interpretation I' and Y is satisfied in it with respect to M. As regards (a), Y[B:A] is in V by our inductive assumption and since (Y[B:A] => (EA)Y) is in V (this follows from the fact that O contains the predicate calculus and V is maximally consistent), we have by the O-maximality lemma that (EA)Y is in V. As regards (b), the conditions of lemma 3 apply to this model and therefore there is a $c^*_{I'}$ transform Y' which is satisfied. But Y' can differ from Y only on occurrences of A, that is, it is

of the form Y[B:A] for some constant B. But then again by the argument for (a) we have that (EA)Y is in V.

This concludes the proof for this case.

Case 2: V_1 is not consistent. In order to explore this case, we must present a few more lemmas:

Lemma 6: If U is a set of sentences in which a constant C does not occur, then if there is a derivation from U of a sentence B, there is a variable A and a derivation from U of the sentence (A)B[A:C].

Proof: Suppose $E_1, \ldots, E_n$ is a derivation of B. Suppose also A is an individual variable which does not appear in any of the E_i's. Then we show by induction that for each i there is a derivation of $(A)E_i[A:C]$. This will show the lemma.

Case 1: E_i is a sentence of U. Then C does not appear in E_i, and therefore E_i is the same as $E_i[A:C]$. The following derivation gives us what we desire:

1)	E_i	Assumption
2)	$-E_i => -E_i$	Tautology
3)	$(EA)-E_i => -E_i$	(R2) of axiom system O applied to (2) above.
4)	$E_i => (A)E_i$	Tautology, abbreviation of (A), m.p..
5)	$(A)E_i$	(R1) of O applied to (4)

and (1).

Case 2: E_i is an axiom of O. In this case it is easy to see that $(A)E_i[A:C]$ is also an axiom of O.

Case 3: E_i follows from E_j by an application of (R2) of O. In this case, $(A)E_i[A:C]$ follows from $(A)E_j[A:C]$ by an application of (R2) of O (it is essential here that A does not occur in the original derivation).

Case 4: E_i follows from E_j and E_k by an application of (R1) of O. In this case the same holds of $(A)E_i[A:C]$, $(A)E_j[A:C]$, and $(A)E_k[A:C]$.

<u>Lemma 7</u>: Suppose U is a set of sentences and A and B are sentences. Then there is a derivation of B from A and U iff there is a derivation of (A => B) from U. (Deduction Theorem)

<u>Proof</u>: We transform any derivation of B from U and A into a derivation of (A => B) from U.

Suppose C, a sentence in the derivation, is an axiom. Then we substitute the following in the derivation at the place of C:

n)	C	axiom
n+1)	C => (A => C)	Tautology
n+2)	(A => C)	Modus Ponens.

Suppose $(A_1)\ldots(A_n)C$, a sentence in the derivation, follows from $(B_1)\ldots(B_m)D$ and $(C_1)\ldots(C_p)(D \Rightarrow C)$ by R1. We sketch a derivation for $A \Rightarrow (A_1)\ldots(A_n)C$ from $A \Rightarrow (C_1)\ldots(C_p)(D \Rightarrow C)$ and $A \Rightarrow (B_1)\ldots(B_m)D$:

First, having derived the above, we can derive $(B_1)\ldots(B_m)(A \Rightarrow (B_1)\ldots(B_m)D)$ and $(C_1)\ldots(C_p)(A \Rightarrow (C_1)\ldots(C_p)(D \Rightarrow C))$. From these, using axiom (A2), axiom (A1), and (R1) several times, we can derive $(B_1)\ldots(B_m)(A \Rightarrow D)$ and $(C_1)\ldots(C_p)(A \Rightarrow (D \Rightarrow C))$. Using axiom (A1) and (R1) several times again, we get the desired result.

Suppose C, a sentence in the derivation, is of the form $(A_1)\ldots(A_n)((EG)D \Rightarrow H)$ where G does not occur free in H and it follows from $(B_1)\ldots(B_m)(D \Rightarrow H)$ by axiom (R2). In this case we have $A \Rightarrow (B_1)\ldots(B_m)(D \Rightarrow H)$. Utilizing axiom (A1) and (R1), we have $A \Rightarrow (B_1)\ldots(B_m)(G)(D \Rightarrow H)$. Since from $(B_1)\ldots(B_m)(G)(D \Rightarrow H)$ we can derive $(B_1)\ldots(B_m)((EG)D \Rightarrow H)$, we have $A \Rightarrow (B_1)\ldots(B_m)((EG)D \Rightarrow H)$. Now we want to replace $B_1, \ldots, B_m$, with $A_1, \ldots, A_n$. This can be easily done by applying axiom (A1) and (R1). This concludes the proof of the lemma.

<u>Lemma 8</u>: Axioms (A1), (A2) and inference rules (R1), (R2) are complete in the standard predicate calculus with respect to standard model theory.

Now, since V_1 is not consistent, we can derive a theorem of the following form from Sen: $-(R \,\&\, E_1 \,\&\ldots\&\, E_n \,\&\, G_1 \,\&\ldots\&\, G_m \,\&\, T\hat{}c^*)$, where each E_i is an E-formula with respect to a constant no other E-formula E_j is an E-formula with respect to, R is a conjunction of sentences of A (in particular, therefore, no new constants of the E-forms, or c* appear in them), each G_i is either a sentence of the form -(c*/S) for some sentence S, or a sentence of the form -(A/S) for some constant A and some sentence S in which c* appears, and finally T^ is either null, -T & -F, -T, or -F.

Claim: In actual fact, T^ cannot be null.

Proof: Utilizing lemma 6, we can universalize the above theorem with respect to c* and the constants the E-formulas are with respect to. Since none of these constants appear in R, we have something of the form: $-R \vee (A_1)\ldots(A_n)(-E'_1 \vee\ldots\vee -E'_n \vee -G'_1 \vee\ldots\vee -G'_m)$. But it is easy to see that the second disjunct implies $(A_1)\ldots(A_p)(-G'_1 \vee\ldots\vee -G'_m)$. This, in turn, violates axiom 8. So we're done, since therefore -R must hold, and this contradicts our assumption of the consistency of SEN.

So, utilizing the reasoning of the claim we have:

**) $R \Rightarrow (A_1)\ldots(A_n)(-G'_1 \vee\ldots\vee -G'_m \vee T\hat{}A)$,

where A is the variable substituted for c* in the application of lemma 6, and where each conjunct of R is in SEN.

<u>Definition</u>: Let A and B be constants, K a sentence and P a one place predicate variable. Then we call (A/(FA & (PA v -PA)) v -(B/K)) and (A/(FA & (PA v -PA)) & (B/K)) <u>blockers</u>.

In particular we say that <u>A</u> <u>blocks</u> <u>B</u> via the above blockers and we call the blockers respectively true and false.

We now consider a language LAN** which is at least as articulate as LAN* but contains in addition a countable number of new constants (call the set of such constants Z) and a countable number of new one-place predicate variables (Call the set of such predicate variables W). Divide Z into a countable collection of countable sets of constants: $CON_1, \ldots, CON_n, \ldots,$ and W into a countable number of countable sets of predicate variables $W_1, W_3, \ldots, W_{2n+1}, \ldots,$ and consider the collection of sublanguages of LAN**: $LAN_0, LAN_1, \ldots, LAN_n, \ldots,$ where LAN_0 is LAN*, the symbols of each LAN_i contains the union of the sets of constants of LAN_j for $j < i$ plus CON_i, and the sets of predicate variables of LAN_j for $j < i$ plus W_i when i is odd. We construct a sequence of sets of sentences V_i of the languages LAN_i as follows: (Note: SEN is V_0.)

Step n (where n is even): Let Y_{n+1} be some ordering of all the E-forms of LAN_{n+1}. For each E-form in the order of Y_{n+1}, add an E-formula to V_n with respect to a constant of CON_{n+1} which does not appear in any other E-formula added already to V_n, and make the resulting set a maximal one of LAN_{n+1} as usual.

Step n (where n is odd): Let $K_1, \ldots, K_n, \ldots,$ be all sentences of LAN_n which do not appear in the second place of an ostensive-sentence. Associate with each K_i two distinct constants c_i and q_i, of CON_{n+1}, and one one-place predicate P_i from W_i. In stages with respect to each c_i, add to V_n the true blocker (letting c_i block q_i) and Tc_i and $-Fc_i$ if these sentences are consistent with V_n and the sentences previously added and the false blocker and $-Tc_i$ and Fc_i otherwise. Call the resulting set of sentences $V_n+[c_i]$. Call the language of V_n with the addition of all constants $c_1,\ldots, c_i$ and all predicates $P_1,\ldots, P_i$ $LAN_n+[c_i]$. Call the union of $V_n+[c_i]$ for all i V_{n+1}.

Let V be the union of all the V_i.

Claim: V is a maximally O-consistent set of sentences of LAN**.

Proof: We first must show that every V_i is consistent. We do so by induction. Certainly V_0 is consistent.

For n even, the proof of consistency is the standard one used to show the addition of E-formulas with respect to new constants is innocuous.

For n odd: Suppose that V_{n+1} is not consistent. Let c_i be the earliest constant for which $V_n+[c_i]$ results in an inconsistency. Then there are O-derivations from $V_n+[c_{i-1}]$ of $-(-Fc_i \ \& \ Tc_i \ \& \ BL')$ and $-(Tc_i \ \& \ -Fc_i \ \& \ BL'')$ where BL' and BL'' are the true and false blockers respectively. If $a_1,\ldots, a_p$ are all usages of axioms A3-A7 in the O-derivations, then there is a derivation of:

$-(a_1 \ \&\ldots\& \ a_p \ \& \ (-Fc_i \ \& \ Tc_i \ \& \ BL'))$, and
$-(a_1 \ \&\ldots\& \ a_p \ \& \ (-Tc_i \ \& \ Fc_i \ \& \ BL''))$

from $V_n+[c_{i-1}]$ utilizing only axioms A1 and A2, and inference rules R1 and R2. We show this is impossible.

Given $V_n+[c_{i-1}]$, add to it the sentences $-(c_i/S)$ for every i and for every sentence S in the language $LAN_n+[c_i]$. Add also either Tc_i and $-Fc_i$, or Fc_i and $-Tc_i$ (if Tc_i and $-Fc_i$ are not consistent with the set). Call the result J. Suppose J is not consistent. Then from $V_n+[c_{i-1}]$ we may derive, utilizing lemma 6, $(A)-(TA \& -FA \& H_1 \&...\& H_p)$ and $(A)-(-TA \& FA \& H_{p+1} \&...\& H_{p+q})$ where each H_i is an ostensive of the form (A/S) for some S in $LAN_n+[c_i]$. But then, using **, (since any derivation in LAN* is also one in $LAN_n+[c_i]$), and axiom (A4), we actually can derive $(A_1)...(A_n)(-G'_1 v...v -G'_m v H_1 v...v H_{p+q})$, and this violates axiom (A7) and the consistency of $V_n+[c_i]$.

Now, let c be a constant of CON_{n+1} which has not been previously utilized. For the sake of the remainder of this proof call L the language containing c plus the alphabet of $LAN_n+[c_i]$. Add the sentences -(c/S) for every sentence S of L to J. Call the result J'. Suppose J' is not consistent. Then using lemma 6 again, we can derive from J $(A)-(H_1 \&...\& H_n)$ where each H_i is a sentence of the form -(A/S) for some S in L. But given axiom (A7), this contradicts the consistency of J. Next, add $-(c_i/S)$ for every sentence S with at least one occurrence of c in it to J'. Call the result J˜. That J˜ is consistent may be seen by the same kind of argument already offered.

Finally, add to J˜ all instances of axioms A3-A7 with respect to all instances of ostensives which appear in sentences of J˜. This set, which we will call J*, is obviously O-consistent.

Let $R_1, \ldots, R_m, \ldots,$ be all the predicate types of ostensives which appear in J*. (For expository purposes we write $R_k A_1 \ldots A_n$ for the ostensive with predicate type R_k and n free individual variables and/or individual constants). For any sentence X in J*, we define X˜ as follows: let G_k be a new n-place predicate symbol not in L for each R_k which has n occurrences of free individual variables and/or individual constants in it. For each $R_k A_1 \ldots A_n$, substitute $G_k A_1 \ldots A_n$ in X. Call the resulting set of sentences J^. We note that J^, which is a set of sentences in the standard predicate calculus (if we consider T and F one place predicate variables), is consistent. Otherwise derive a contradiction C˜. Then C is a contradiction also and it is derivable using axioms A1, A2 and inference rules R1, R2 from J* (which is impossible as J* is O-consistent). Therefore there is a model M of the standard predicate calculus for J^. Let R be the predicate type of the sentence BL' if Tc_i and $-Fc_i$ are in J^, and BL'' otherwise; and let G be a new predicate variable with 2 places (Note that the number of free places in either BL' or BL'' is 2). Also, let $Q_1, \ldots, Q_n, \ldots,$ be all the predicate types of ostensives which appear in the sentences a_j but do not appear in any sentence in J*. Let $H_1, \ldots, H_n$ be new predicate variables associated, as before, with the predicate types Q_k which appear in a_1,..., a_p but do not appear in any sentence of J*. We supply extensions for G and H_j's on the domain of M. Now, by definition, a couple of items of the domain of M are in G's extension iff c_i is mapped to the first item and q_i is mapped to the second item. The extension for each H_j is null. Let M' be the model with the domain of M and with the same extensions assigned to all predicates and constants that M assigns to them but in addition with the assignments to the new predicate variables G and H_j's that we have just defined. We now want to show that the ~-transforms of every sentence X in V_n+[c_i] and every a_j is satisfied in M and further that BL'~ (or BL''~) is satisfied in M'. This will show what is desired since if the transforms of a set of sentences are consistent with respect to axioms A1, A2 and inference rules R1, R2, the set itself also is.

1) Let S be any instance of axioms A3-A7 all of whose ostensive predicate types appear in J*. Then S~ is satisfied in M' by construction.

2) Let S be an instance of axiom (A3) in which the left-hand side is an ostensive whose ~-transform is either an H_j or G. If the former, then since no instance of the antecedent is satisfied, the axiom is satisfied. If the latter, we need only concern ourselves with the case when GAB is satisfied. This happens when A is the item c_i is

mapped to and B is the item q_i is mapped to. But we have that $-Fc_i$ and Tc_i (or Fc_i and $-Tc_i$) are satisfied and $-(q_i/K_i)^{\sim}$ is satisfied, and this gives us what we want.

3) Let S be an instance of axiom (A5) in which either ostensive occurring in it is of predicate type associated with an H_j or with G. Then since no instance of any H_j is satisfied and only one instance of G is satisfied with respect to c_i which occurs in the first place of no other predicate variable associated with an ostensive-type different from G, the ~-transforms of these axioms are satisfied also.

4) The reasoning concerning axiom (A6) is similar to that concerning axiom (A5).

5) Since no instance of a predicate variable associated with an ostensive type with c in its first place is satisfied in M', axiom (A7) is satisfied also.

This concludes our proof that each V_i is consistent. Therefore, V is an O-consistent set of sentences. To show that it is maximally O-consistent, it is sufficient to note that any sentence of LAN** is a sentence of LAN_n for some n. But at every even stage n, we expand V_n to a maximally consistent set V_n of LAN_{n+1}. We also note by construction, that any E-form of LAN** is an E-form of LAN_n for some n and therefore there is an E-formula in V for that form. Finally, every sentence S in LAN** is either in the second place of an ostensive in LAN** or there are constants A and B and one place variable P such that the ostensive (A/(FA &

(PA v -PA)) v -(B/S)) is in V. It is easy to see that in any model satisfying the sentences of V, which maps B to S cannot satisfy (B/S) on pain of contradiction.

As before, we use V to construct a premodel M and show that, in fact, M is an O-model.

Let d be the union of all the sentences of LAN** and all constants A which do not appear in the first place of any ostensive-sentence in V and for which neither FA nor TA are in V.

We define the premodel M as follows: any constant A is mapped to a sentence B if i) (A/B) is in V, or if ii) no ostensive-sentence with A in its first place is in V, but an ostensive-sentence of the form (C/(FC & (PC v -PC)) v -(A/B)) or (C/(FC & (PC v -PC)) & (A/B)) was added at step n for some odd n to V_n (We note that our construction is such that every sentence has a constant mapped to it). A constant A is mapped to FA if neither conditions (i) or (ii) hold, but either TA or FA are in V. It is mapped to itself otherwise. For each n-place predicate variable B, it is mapped to that set of n-tuples $(R_1,\ldots, R_n)$ such that $BA_1 \ldots A_n$ is in V and A_i is mapped to R_i by the above mapping for constants. (We note that this definition for the predicate variables is well-defined even though it is possible for more than one constant to be mapped to the same item of d--i.e., if two constants are mapped to the same sentence. But by axiom 7, the predicates must agree on them.)

The predicate constant T is mapped to V and F is mapped to the remaining sentences of d.

This clearly defines a premodel. We want to show that in fact it defines an O-model. We shall do so as follows. First we will show that a sentence is satisfied by M iff it is in V. From this we can show that M is an O-model. V is precisely the extension of T, and this satisfies the first condition. Let R be an ostensive (A/B) not satisfied in M, such that A is mapped by the above mapping to B. Either B is FA and (A/B) cannot be satisfied, or by construction, there is a constant C and some one-place predicate variable P such that, the ostensive-sentence (C/(FC & (PC v -PC)) v -(A/B)) (or (C/(FC & (PC v -PC)) & (A/B))) is in V. Using reasoning analogous to that used to show Lemma 5, we may see that there cannot be a premodel agreeing with M in which R is satisfied.

So to show that a sentence is satisfied by M iff it is in V is to conclude this case. We do so by induction on the number of places at which connective or quantifier symbols appear.

1) Suppose W is $BA_1 \ldots A_n$, an n-place predicate variable followed by n individual constants. Then it is, by construction, satisfied iff it is in V.

2) W is -A or (A v B), where A and B are sentences. This is standard and follows from the O-maximality lemma.

3) W is (A/B) where A is an individual constant and B is

a sentence. By construction (A/B) is in V iff (A/B) is in T's extension and B is assigned by M to A. By axiom (A3) and the O-maximality lemma, if (A/B) is in V, then (TA <=> B) is in V. Thus by the O-maximality lemma again, either a) TA and B are in V, in which case they are both in T's extension, or b) TA and B are not in V, in which case they are both in F's extension. In either case (A/B) is satisfied.

4) Suppose TA is in V. First suppose there is a sentence B such that (A/B) is in V. Then by axiom (A3) and the O-maximality lemma, B is in V. Since B is in T's extension and (A/B) is in T's extension, TA is satisfied. On the other hand, if (A/B) is not in V for any B, A is mapped to some sentence C (by construction), and therefore TA is satisfied. Suppose, conversely, that TA is satisfied. Then TA is in T's extension and therefore in V.

The reasoning is similar with FA.

5) Suppose A is an individual variable and Y is a wff such that the sentence (EA)Y is in V. Then there is an E-formula ((EA)Y => Y[C:A]) in V. By our inductive assumption, Y[C:A] is satisfied. Therefore (EA)Y is satisfied by the satisfaction clauses. Suppose conversely, that (EA)Y is satisfied. Then either a) there is an individual constant B and Y[B:A] is satisfied or b) there is an interpretation I and Y is satisfied in it with respect to M. As regards (a), Y[B:A] is in V by our

inductive assumption and since (Y[B:A] => (EA)Y) is in V (this follows from the fact that O contains the predicate calculus and V is maximally consistent), we have by the O-maximality lemma that (EA)Y is in V. As regards (b), we note that every item of d has a constant mapped to it. Therefore, for all interpretations I and for all individual variables A, g(A) is defined. Therefore there is a constant B such that Y[B:A] is satisfied. But this puts us back in case (a).

This concludes the proof of strong completeness.

CHAPTER 4

Conventions: in this chapter, I have borrowed terminology and notions freely from Fitting's NMT, Vissar's FVSL, and Woodruff's PTL.

There is a sense in which certain sentences are problematical or unproblematical from the classical point of view. Intuitively, if we imagine applying the standard Tarskian truth (satisfaction) definition to a Tarskian model which contains its own language, the unproblematic sentences are those which get one and only one truth value. The problematic sentences are those which get either both truth values or none of the truth values. Now, of course, these intuitions are based on very simple syntactic examples, and in fact with a little work it is easy to imagine more complicated types of problematic sentences. Nevertheless, it does seem that this notion of problematic sentence is one that can be made precise. Having done so, it gives us a further way of examining the various solutions in the literature--from the vantage point of the theory where the problems arose in the first place.

I diagnose the problematic sentences in this chapter using one set of tools and have offered a solution in Chapters 2 and 3 using another set. I am suspicious of approaches which purport to do both with the same set of tools. I make further specific complaints about the diagnostic aspects of the Gupta/Herzberger approach in Chapter 1--these are directed towards "Herzberger" merely because he seems to make more of the diagnostic aspect than does Gupta. I should say that by "diagnostic" I do not mean finding some deep pathology that lies at the root of the problematical (or paradoxical) sentences. I mean the more superficial notion of merely providing a way of recognizing the problematical sentences, that is, the ones which prevent us from simplemindedly applying Tarski's approach plus the biconditionals to a language containing its own truth and falsity predicates.

The notions "grounded" as Kripke uses it and "stable" as Gupta and Herzberger use it, are intratheoretic notions. One caveat: both Gupta and Herzberger speak of separating the diagnostic aspect of recognizing the problematical sentences from the normative question of how to solve them. They seem to give credit to their approaches doing the first in a paradigmatic way because the approaches are bivalent and it is in bivalent languages that the problem first arises. Of course, given the rich alternative technical possibilities, this is a weak reason. So I will persist in considering their notion of 'stable' as intratheoretic in the absence of stronger arguments.

We assume a language L of the standard predicate calculus with the additional one-place logical predicates "T" and "F", and further, a constant $\bar{e}$, for each sentence e of L.

<u>Definition</u>: A <u>Tarski pseudomodel</u> Ps is a standard Tarski model with a domain containing all the sentences of L and with interpretations in the domain for all the constants of L and all the predicate symbols of L (excepting "T" and "F") such that every element e in the domain has a constant mapped to it, and further, each sentence e has the constant $\bar{e}$ mentioned above, mapped to it.

<u>Definition</u>: Let St be a set of pairs of the form: {S,v}, where S is a sentence of L and v is a truth-value (t or f). We say St is a set of <u>truth-assignments</u>.

<u>Notation</u>: If v is a truth-value then v^{o} is the opposite truth value. If A and B are sets, then A U B is the union of A and B. Further, "A $\leq$ B" means that A is a subset of B, and "A $<$ B" means that A is a member of B.

<u>Definition</u>: Let St be a set of truth-assignments. We define the <u>limited-jump J(St)</u> as follows: it is the smallest set of pairs {S,v} such that:

1) St $\leq$ J(St)

2) If (A,t) $<$ St, then (-A,f) $<$ J(St); if (A,f) $<$ St, then (-A,t) $<$ J(St).

3) If (A,t) $<$ St, then {(A v B,t), (B v A,t)} $\leq$ J(St) for all sentences B. If {(A,f), (B,f)} $\leq$ St, then (A v

B,f) $\in$ J(St).

4) If (A(c),t) $\in$ St for any constant c, then ((EB)A(B),t) $\in$ J(St). If (A(c),f) $\in$ St for every constant c, then ((EB)A(B),f) $\in$ J(St).

5) If (A,t) $\in$ St then {(TĀ,t), (FĀ,f)} $\in$ J(St). If (A,f) $\in$ St then {(TĀ,f), (FĀ,t)} $\in$ J(St). If either (TĀ,t) or (FĀ,f) $\in$ St then (A,t) $\in$ J(St). If either (TĀ,f) or (FĀ,t) $\in$ St, then (A,f) $\in$ J(St).

Motivation: We use substitutional quantification in the second condition of 4 merely to avoid a detour through satisfaction. The model is a straightforward Tarskian one except for sentences with occurrences of truth or falsity predicates. The jump operator I have defined is not the common one in the literature but it follows naturally from our aims. We are interested in how the assignment of truth values trickles up syntactically, given the standard interpretation of the connectives. This justifies the jump definition (which takes us up, essentially, one step) for the connectives and the quantifier. Of course there is the possibility that more than one truth value can occur for a sentence (say, {(A,t), (A,f)} $\leq$ St for some A); in this case the standard definition would allow both truth values to trickle up, just as we have above. So, for example, if (B,f) $\in$ St then {(A v B,t), (A v B,f)} $\leq$ J (St). In cases where no truth value occurs for a sentence, the standard definition would let the gappiness trickle up, subject to the interpretation of the connectives, just as we have above. For the truth and falsity predicates, it gives us the expected interpretation (via the Tarski biconditionals) on sentences which have already been assigned truth values. Unsurprisingly, this should be fairly familiar: the conditions on the jump operator are strong four valued Kleene (true, false, neither and both, as in Vissar and Woodruff) and the Tarksi biconditional conditions on it are the ones that appear in the definition of the standard kind of jump (see below). Notice that condition 5 works both from the left to the right and from the right to the left of the Tarski biconditionals: that is, if $(A,t) < St$, then $(T\bar{A},t) < J(St)$, and if $(T\bar{A},t) < St$, then $(A,t) < J(St)$.

It might be thought that conditions which work down such as:

a) $\{(A,t), (-A \vee B,t)\} \leq St$ only if $(B,t) < J(St)$,

b) $(-A,t) < St$ only if $(A,f) < J(St)$,

c) $(A \vee B,f) < St$ only if $\{(B,f), (A,f)\} < J(St)$,

and so on, are desirable. What stops this are two considerations. First, we understand the Tarski satisfaction clauses as being "compositional" in a slightly broad sense. We are interested in the satisfaction conditions and how they break down. They work from the bottom up, so it is natural to immitate that. Secondly, adding the above conditions (and their natural companions) would make every sentence (and not just the ones which are broadly understood to be involved in the vicious reference) have both truth values whenever vicious reference arose[1] and that would defeat our diagnostic aim. We would then be able to diagnose only on a model-by-model basis rather than on a sentence-by-sentence basis.

Definition: Let Ps be a pseudomodel. We say that Ps generates the set of truth assignments St iff

1) All sentences occurring in any pair in St are atomic.

2) {S,v} is in St iff S is given the value v by the standard Tarskian definition of truth in Ps.

Definition: A set of truth assignments is normal if there is a pseudomodel Ps which generates it.

We note that the collection of all sets of truth assignments is a complete lattice with respect to set inclusion, union and intersection, and that our jump operator is monotonic, i.e., St $\leq$ St' only if J(St) $\leq$ J(St'). In fact, it is, as Woodruff would call it, "cumulative" (St $\leq$ J(St)). Thus, it is sound with respect to every set. Therefore given a set of truth assignments St, there is a smallest ordinal @ such that J applied to St @ times ($J^{@}$(St)) is a fixed point of J (see Yablo's GIP). Call $J^{@}$(St) a diagnostic fixed point of St. Intuitively, if St is a normal set of truth assignments, then its diagnostic fixed point tells us which sentences receive a truth value under the standard Tarskian definition of truth. We call such sentences *Tarski-grounded with respect to St*. Further definitions are possible. Consider a sentence S which is not Tarski-grounded with respect to St. If the diagnostic fixed point of (St U (S,t)) contains (St U (S,f)) and vice versa, S is *paradoxical with respect to St*. Otherwise, it is a *truthsayer with respect to St*. We note finally that refinements of this are possible. For example, a sentence may be a truthsayer with respect to a normal set B although paradoxical with respect to the union of St with a valuation for a truthsayer.

We want to relate our jump operator to the more standard ones appearing in the literature. To this end, given a set of truth assignments St, we define the following valuation V_{St}. This valuation is defined in terms of truth values as subsets of {T, F} as in Visser. Given St:

$(T\bar{c},t) \in V_{St}$ iff $(c,t) \in St$ iff $(F\bar{c},f) \in V_{St}$,

$(F\bar{c},t) \in V_{St}$ iff $(c,f) \in St$ iff $(T\bar{c},f) \in V_{St}$,

$(Pc_1 \ldots c_n,t) \in V_{St}$ iff $(Pc_1 \ldots c_n,t) \in St$,

$(Pc_1 \ldots c_n,f) \in V_{St}$ iff $(Pc_1 \ldots c_n,f) \in St$,

$(-A,t) \in V_{St}$ iff $(A,f) \in V_{St}$,

$(-A,f) \in V_{St}$ iff $(A,t) \in V_{St}$,

$(A \vee B,t) \in V_{St}$ iff (A,t) or $(B,t) \in V_{St}$,

$(A \vee B,f) \in V_{St}$ iff $\{(A,f), (B,f)\} \leq V_{St}$,

$(EA(B),t) \in V_{St}$ iff for some constant C, $(B[C:A],t) \in V_{St}$

$(EA(B),f) \in V_{St}$ iff for all constants C, $(B[C:A],f) \in V_{St}$.

Applying this to a set of truth assignments St, gives us a new set of truth assignments V_{St}. Thus as in Woodruff, we define J(St) as V_{St}.

<u>Theorem 1</u>: Let T be the set of sets of truth assignments. Then any fixed point of J is a fixed point of J.

<u>Proof</u>: Let R be a fixed point of J. Then (*) J(R) = R. We show that J(R) $\leq$ R.

1) Suppose $(Pc_1 \ldots c_n, v) \in J(R)$. Then it is in R by definition of J.

2) Suppose $(-A,v) \in J(R)$. Either $(-A,v) \in R$ (and we're done) or $(A,v_o) \in R$ (by definition of J). But by (*) therefore, $(A,v_o) \in J(R)$ and so, therefore, is (-A,v) (by definition of J). But then $(-A,v) \in R$ (by (*)) as desired.

3) Suppose $(A \vee B,t) \in J(R)$. Then either $(A \vee B,t) \in R$ (in which case we're done), or either (A,t) or $(B,t) \in R$ (by the minimality of J(R)). But then either (A,t) or $(B,t) \in J(R)$, and so $(A \vee B,t) \in J(R)$. Thus, $(A \vee B,t) \in R$ as desired.

4) Suppose $(A \vee B,f) \in J(R)$. Then either $(A \vee B,f) \in R$ (in which case we're done) or both $\{(A,f), (B,f)\} \leq R$ and by reasoning similar to that of (3) we have that $(A \vee B,f) \in R$.

5) Suppose $((EA)B,t) \in J(R)$. Then either $((EA)B,t) \in R$

or there is some constant c such that (B[c/A],t) ≺ R. But then (B[c/A],t) ≺ J(R) and so ((EA)B,t) ≺ J(R) and thus in R as desired.

6) Suppose ((EA)B,f) ≺ $\mathcal{J}$(R). Then either ((EA)B,f) ≺ R or for every constant c (B[c/A],f) ≺ R. Using reasoning similar to that of (5) we conclude that ((EA)B,f) ≺ R.

7) Suppose (T$\breve{c}$,t) ≺ $\mathcal{J}$(R). Then either (T$\breve{c}$,t) ≺ R or (c,t) ≺ R. But then (T$\breve{c}$,t) ≺ J(R), and thus in R.

The reasoning is similar for (T$\breve{c}$,f), (F$\breve{c}$,t) and (F$\breve{c}$,f).

Of course, fixed points of $\mathcal{J}$ are not necessarily fixed points of J. But we are interested in the smallest fixed points of $\mathcal{J}$ which contain certain "nice" sets and these fixed points are fixed points of J.

<u>Definition</u>: As in Woodruff, we introduce a "cumulative" jump J' such that J'(S) = J(S) U S.

<u>Definition</u>: Let St be a set of truth-assignments. St is <u>downward saturated</u> if

1) (A v B,t) is in St iff either (A,t) or (B,t) is in St.

2) (A v B,f) is in St iff both (A,f) and (B,f) are in St.

3) ((EA)B,t) is in St iff there is a constant c such that (B[c/A],t) is in St.

4) ((EA)B,f) is in St iff for every constant c,

(B[c/A],f) is in St.

<u>Definition</u>: A set of truth assignments St is <u>J-sound</u> iff St $\leq$ J(St).

<u>Definition</u>: A set of valuations St is <u>inconsistent</u> iff for some sentence A, it contains both (A,t) and (A,f).

We note that a downward saturated set can contain (TB,t) without containing (TC,t), where B = "TC" and C = "TC"; but such a set is not J-sound, since (TB,t) is not $<$ J(St). Also, inconsistent sets may be downward saturated although they cannot be J-sound. On the other hand, a J-sound set may contain PA and --(PA v PB) without containing (PA v PB). Such a set is clearly not downward saturated. But we have the following relation between the two notions:

<u>Theorem 2</u>: A set St of truth-valuations which is a fixed point of J and downwards saturated, is J-sound.

<u>Proof</u>: We show by induction that St $\leq$ J(St).

1) Let (PA_1...A_n,v) $<$ St, then it is $<$ J(St) by definition.

2) If (T$\bar{A}$,t) $<$ St, then (T$\bar{A}$,t) $<$ J(St). But (A,t) $<$ J(St) $\leq$ St. Therefore (T$\bar{A}$,t) $<$ J(St) as desired. The cases for (T$\bar{A}$,f), (F$\bar{A}$,t), and (F$\bar{A}$,f) are similar.

3) Suppose (A v B,t) $<$ St. So, by downwards saturation,

(A,t) or (B,t) $\in$ St $\leq$ J(St) (by our inductive assumption). But therefore, (A v B,t) $\in$ J(St) as desired. The case of (A v B,f) is similar.

4) Suppose (-A,v) $\in$ St. By downwards saturation, (A,v^{o}) $\in$ St $\in$ J(St) (by our inductive assumption). But therefore, by definition of J(St), (-A,v) $\in$ J(St).

5) Suppose ((EA)B,t) $\in$ St. By downwards saturation, (B[c/A],t) $\in$ St for some constant c. By our inductive assumption, (B[c/A],t) $\in$ J(St), and therefore, by the definition of J(St), ((EA)B,t) $\in$ J(St) too.

The case ((EA)B,f) is similar.

<u>Theorem 3</u>: Suppose St is a fixed point of J which is J-sound. Then St is a fixed point of J.

<u>Proof</u>: We show J(St) $\leq$ St. Soundness gives us the result. The proof is by induction.

1) Suppose $(PA_1 \ldots A_n, v)$ $\in$ J(St). Then it is in St by definition.

2) Suppose $(T\bar{A},t)$ $\in$ J(St). Then (A,t) $\in$ St. But (A,t) $\in$ J(St). Therefore $(T\bar{A},t)$ $\in$ J(St) $\leq$ St. The cases $(T\bar{A},f)$, $(F\bar{A},t)$, and $(F\bar{A},f)$ are similar.

3) Suppose (A v B,t) $\in$ J(St). Then (A,t) or (B,t) $\in$ J(St). But by the induction assumption, either (A,t) or (B,t) $\in$ St. Therefore, (A v B,t) $\in$ J(St) $\leq$ St, as desired. The case (A v B,f) is similar.

4) Suppose (-A,v) $\in$ J(St). Then (A,v^{o}) $\in$ J(St). By our induction assumption, (A,v^{o}) $\in$ St. But then (-A,v) $\in$ J(St)

$\leq$ St.

5) Suppose $((EA)B,t) < J(St)$. Then $(B[c/A],t) < J(St)$. By our induction assumption, $(B[c/A],t) < St$. But therefore $((EA)B,t) < \mathcal{J}(St) \leq St$ as desired.

In particular, any normal set St is downward saturated and therefore the smallest fixed point of J containing St is the same as the smallest fixed point of $\mathcal{J}$ containing St.

<u>Fact</u>: If $\bar{A}$ = "TA" and B = "TB", then (TA v TB,t) may be in a fixed point of $\mathcal{J}$ without (TA,t) or (TB,t) appearing there, but this is not possible in a fixed point of J.

Were I attempting anything other than diagnosis here, this fact would suggest a kind of peculiarity possible in fixed points. But we will, in general, be interested in normal sets, or normal sets augmented by problematic sentences (such sets can always be chosen to be downward saturated) and so this is no problem.

<u>Theorem 4</u>: If St is a fixed point of $\mathcal{J}$ then it is a fixed point of J'.

<u>Proof</u>: We show that $J'(St) \leq St$ by induction. Due to the nature of J', it is possible for a sentence to be in the J'-jump of a set because it is in the set itself. Clearly for these cases, the theorem is trivial. We thus

ignore them in the induction below:

1) Suppose $(PA_1 \ldots A_n, v)$ ∈ J'(St). Then it is in St by definition.

2) Suppose (TĀ,t) ∈ J'(St). Then (A,t) ∈ St. So (TĀ,t) ∈ J(St) ≤ St as desired. The cases (TĀ,f), (FĀ,t), and (FĀ,f) are similar.

3) Suppose (A v B,t) ∈ J'(St). Then either (A,t) or (B,t) ∈ J'(St). So, by our inductive assumption, either (A,t) or (B,t) ∈ St. But then (A v B,t) ∈ J(St) ≤ St. The case (A v B,f) is similar.

4) Suppose ((EA)B,t) ∈ J'(St). Then (B[c/A],t) ∈ J'(St). By our inductive assumption, (B[c/A],t) ∈ St and thus ((EA)B,t) ∈ J(St) ≤ St, as desired.

<u>Fact</u>: Suppose Ā = "FA". If St = {(FA,t)} then the smallest fixed point of J' containing St will contain (FA,t), but not (--FA,t). This is because St is not J-sound, and therefore J(St) contains neither (FA,t) nor (--FA,t), but the former is in St and therefore in J(St) U St although this is not the case with the latter. The smallest fixed point of J containing St will contain both.

I understand the motivation for a jump operator that is sound on every set of valuations, but J' seems particularly inappropriate in the light of this last fact. Intuitively, it saturates prematurely. I think J would do better in its place. It is easier to find appropriate sets for J (appropriate in this context meaning containing sentences whose behavior under the Tarski biconditional and satisfaction clauses we would like to study)--they merely have to be downward saturated and our intuitions will be satisfied with the smallest fixed point of J containing them. But no such appropriate ones seem available for J'.

Footnotes

1) Suppose we have (FA,t) and (FA,f) in St. Then by (2), (-FA,t) and (-FA,f) are in J(St). Then, using (3), (-FA v B,t) and (-FA v -B,t), for arbitrary sentence B, are in J(J(St)). Finally, using (a), we have (B,t) and (-B,t) in J(J(J(St))).

Bibliography

All references in the text and the footnotes shall be to abbreviations which follow the entries in parentheses.

Benacerraf and Putnam, _Philosophy of Mathematics_, second edition, Cambridge University Press 1983

Carnap, "Truth and Confirmation," in _Readings in Philosophical Analysis_, New York, 1949, (TAC)

Chang and Keisler, _Model Theory_, North Holland, second edition, 1977. (MT)

Davidson, Donald, _Inquiries into Truth and Interpretation_, Clarendon Press, 1985

Enderton, Herbert B., _A Mathematical Introduction to Logic_, 1972, Academic Press. (MIL)

Feferman, Solomon, "Useful Type-Free Theories," _The Journal of Symbolic Logic_ 49 (March 1984) (UTT)

Field, Hartry, "Tarski's Theory of Truth," _Journal of Philosophy_ LXIX (July 13, 1972) (TTOT)

Fine, Kit, and McCarthy, Timothy, "Truth without Satisfaction," _Journal of Philosophical Logic_ 13 (Nov. 1984): 397-421 (TWS)

Fitting, Melvin, "Notes on the Mathematical Aspects of Kripke's Theory of Truth," (forthcoming in _Notre Dame Journal of Formal Logic_) (NMT)

Gupta, Anil, "Truth and Paradox," _Journal of Philosophical Logic_ 11 (February 1982) (TP)

Gupta, Anil and Martin, Robert, L., "A fixed Point theorem for the Weak Kleene Valuation Scheme," _Journal of Philosophical Logic_ 13 (May 1984): 131-135 (FPTWK)

Hempel, Carl, "Studies in the Logic of Confirmation," in _Aspects of Scientific Explanation_, the Free Press, New York, 1965 (ASE)

Herzberger, Hans, "Notes on Naive Semantics" _Journal of Philosophical Logic_ (February 1982)

______. "New Paradoxes for Old" (unpublished) (NPFO)

______. "Naive Semantics and the Liar Paradox" _Journal of Philosophy_ LXXIX (September 1982) (NSL)

Kripke, Saul, Naming and Necessity Harvard University Press, 1980 (NAN)

______. "Outline of a Theory of Truth," Journal of Philosophy (Nov. 6, 1975) (OOTOT)

Mostowski, Andrzej, "Tarski, Alfred," The Encyclopedia of Philosophy Vol. 8 (ET)

Parsons, Charles, "Mathematics, Foundations of," The Encyclopedia of Philosophy (MFO)

Popper, Karl R., Conjectures and Refutations: The Growth of Scientific Knowledge, Harper torchbooks, 1965. (CAR)

Putnam, Hilary, Meaning and the Moral Sciences, International Library of Philosophy and Scientific Method, Routledge & Kegan Paul, 1978, (MMS)

______. Reason, Truth and History, Cambridge University Press 1981 (RTH)

Quine, Willard Orman Van, "What price Bivalence?," in Theories and Things, Harvard University Press, 1981: 31-37 (WPB)

______. "Logic Based on Inclusion and Abstraction," in Selected Logic Papers, Random House, 1966 (LBIA)

Russell, Bertrand, "Mr. Strawson on Referring," in Essays in Analysis George Braziller, New York, 1973 (MSOR)

Skyrms, Brian, "Notes on Quantification and Self-Reference" in The Paradox of the Liar, edited by Martin, Yale University Press, 1970, 67-77.

______. "The Intensionality of Self-Referential Contexts" in Truth and The Liar Paradox, edited by Martin, Clarendon Press, 1984 (ISC)

Tarski, Alfred, "The Concept of Truth in Formalized Languages," in Logic, Semantics, Metamathematics second edition, 1983 (CTFL).

______. "The Semantic Conception of Truth," Philosophy and Phenomenological Research IV (March 1944) (SCT)

______. "The Establishment of Scientific Semantics," in Logic, Semantics, Metamathematics, second edition, 1983 (ESS)

Visser, Albert, "Four Valued Semantics and the Liar," _Journal of Philosophical Logic_ 13 (May 1984) (FVSL)

Woodruff, Peter W., "Paradox, Truth and Logic. Part I: Paradox and Truth" _Journal of Philosophical Logic_ 13 (May 1984) (PTL)

Yablo, Steve, "Grounding, Independence, and Paradox," _Journal of Philosophical Logic_ 11 (Feb. 1982) (GIP)