Notre Dame Journal of Formal Logic
Volume 32, Number 3, Summer 1991

A Simple Axiomatizable Theory of Truth

JODY AZZOUNI

Abstract A general framework for studying self-referential languages is given: by this is meant both a model theory and a complete set of axioms for that model theory. The generality of the approach is shown by exhibiting the wide range of pathological sentences it allows, and its model-theoretic compatibility with other approaches, such as semi-inductive ones. Philosophical motivation for some of the new technical moves is given, and an appendix supplies the completeness proof.

1 The most salient feature of the truth predicate is its apparent obedience to Criterion *T*. Criterion *T*, however, is so simple that one might have hoped that a predicate axiomatically duplicating this property of truth could be easily added to first-order logic much as equality has been. The Liar's Paradox dashes any thought that such a move is technically easy.

In any case, axiomatic treatments of the truth predicate have been largely ignored as a possibility in the standard literature.[1] One reason for thinking such approaches are ruled out is that it seems that a language cannot have a theory of truth without some capacity to describe its own syntax. But even a fairly weak syntactic capacity can breed incompleteness, and therefore (it seems) a theory of truth cannot be axiomatized either.

This problem can be solved by relativizing the syntactic capacity of the language to the model rather than fixing it once and for all for a class of interpreted languages. By this I mean, e.g., that instead of fixing a canonical mapping of constants to sentences (via a quotation function, say) that is to hold for all models of the language, one allows the constants to vary in what they are mapped to from model to model. Doing so results in a theory of truth that is valuable in two ways. First, it has been known since Tarski that syntactic theory and truth theory do not sit comfortably together. But it is the details—exactly how much syntax can sit together with exactly how much truth—that are of interest. There are some results (see Gupta [5], pp. 183–194 for some examples), but an approach that gives a general framework to study this question exhaustively is desirable. Such a framework should be one that contains a first-order theory of

Received August 3, 1989; revised April 2, 1990

truth together with constraints on how it will interact with whatever capacity the language has to describe its own syntax. The effects of this capacity on the theory of truth can then be studied by singling out, either model-theoretically or axiomatically, subcollections of interpreted languages with varied syntactic capacity.[2]

Second, a longstanding, although naive, objection to Tarski's approach is that the truth predicate in natural languages does not splinter into a hierarchy of predicates. It is almost a fact that something like a hierarchy of languages is unavoidable when "closed" languages are sought (Herzberger [8]), but perhaps the blame can be placed on the theory of syntax rather than on the truth predicate. It does seem, speaking naively again, that natural languages contain a theory of truth; should a hierarchy of languages enter the picture with the introduction of a (certain amount of) syntactic theory, this would not be as intuitively objectionable.

I now go on to present the foundations of a theory of truth in the spirit of the above paragraphs. I give a model theory and a set of axioms which are complete vis-à-vis the model theory. Along the way, I also give indications of what work has been left to do. But to motivate the technical moves of this theory, a bit of philosophical scaffolding is called for, and we turn to that in the following two sections.

2 The Liar's Paradox offers technical obstacles to a theory of truth, but in point of fact the less sensationalistic truth tellers do so as well. Should we approach a theory of truth by attempting to axiomatize a truth predicate via Tarski biconditionals, they will become vacuous on truth tellers. It is this that gives rise to the striking characteristic of a truth teller, that it is perfectly arbitrary what truth value such a sentence is assigned, regardless of "how the world is". That is, truth tellers put pressure on our intuition that truth is "Tarski-reducible" ([5], p. 196): that by specifying the truth values of all the sentences in which the truth predicate does not occur, we have described enough to fix the truth values for *all* sentences of the languages.[3]

The fact that lack of Tarski-reducibility is a characteristic of paradoxical sentences as well as truth tellers helps motivate Kripke-style theories of truth. Involved in all of them are techniques for recognizing "problematical" sentences by this lack and then neutralizing their effect on the theory of truth.[4] What makes these techniques more than just elegant constructions is how they exemplify intuitions about Tarski-reducibility—in particular, the intuition that there is something sematically wrong with a sentence that is not Tarski-reducible.

Truth tellers put pressure on the intuition that every sentence is Tarski-reducible only in tandem with another philosophic intuition, sentential bivalence,[5] the claim that *every* (open) sentence is either true or false (of an object) but not both. For languages rich enough to contain truth tellers, there is considerable tension between sentential bivalence and the constraint of Tarski-reducibility. One or the other must go, and, as it turns out, the constraint of sentential bivalence is the one that has yielded in Kripke-style truth theories.[6]

This is obvious in Kripke's approach.[7] It may not be in the Gupta/Herzberger approaches since they are explicitly two-valued. But what these approaches actually offer is a *collection* of models in which the truth predicate is bivalent. A philosophical interpretation of the role of this plethora of interpreted truth predicates must be given. Herzberger suggests picking out one model on the grounds that the construction used to derive the collection of models cycles around it. This certainly preserves bivalent intuitions, but with the danger that the criterion he uses to choose the model is arbitrary (Parsons [14], p. 264). Gupta's suggestion, by contrast, is really much more radical: truth should not be seen as a concept with an application procedure, but rather one with a revision procedure—in learning the meaning of "true" what we learn is a rule that enables us to improve on a proposed candidate for the extension of truth. Here it is quite clear that bivalence—of the truth predicate at least—has been given up, and given up in much the same spirit that it is given up in Intuitionism.[8]

There is no doubt that Tarski-reducibility is a very compelling intuition. This helps explain why gappy approaches arose first, whereas glut approaches, although technically similar, only arose later, and are still unpopular. But what is behind the intuition? At work, certainly, are motivations similar to those which drive philosophers away from bivalence. Quine [15] has noted that adopting bivalence saddles us with undecidables—and this is meant in the strongest sense possible. It is not merely that some human failing of time or energy prevents us from evaluating whether something is in the extension of a particular predicate, for often it is not even clear in principle how an idealized being could do the job, and in mathematics it is provably the case that some facts regarding the extensions of certain predicates are out of our hands, however idealized we make their grip.

If sentential bivalence is taken to apply to the truth predicate of rich languages, Tarski reducibility[9] must be given up, but only for a certain class of sentences. Doing so has impact on several other intuitions about the truth predicate which have played some role in the literature. I will illustrate this in the rest of this section by discussing two commonly offered constraints on a theory of truth: physicalism and Criterion T, and showing their connection to Tarski reducibility.

Field [4] has noted that Tarski expressed support for a certain reductionist doctrine, namely that there be no irreducibly semantic facts. This claim is just part of a broader physicalist program where all notions are to be "reduced" to notions of physics and set theory. Such a program has come under serious scrutiny in recent years and I don't want to dwell on its drawbacks now. Suffice it to say that bivalence about truth tellers amounts to jettisoning this program, at least "around the edges": the truth value of every sentence is fixed by what the facts are, but only if irreducibly semantic facts are included among "the facts".[10] Notice, however, that this is nothing more than a different way of describing how Tarski-reducibility is being violated.

Let us now turn to Criterion T. Criterion T is seen as so significant for truth that it has sometimes been offered as a *full* analysis of the meaning of "true" (and, by assumption, consequently determining totally the extension of the truth predicate). Of course it cannot be everything there is to say about the meaning of truth, since redundancy theories of truth do not work in the context of quantification. E.g., in Tarski-style definitions of truth, a great deal more than Criterion T is necessarily involved. In any case, the intuitive power of Criterion T is not a brute fact of semantic nature, but something that can be analyzed into other intuitions which can be questioned. In particular, the intuitive sway that Criterion *T* has over us lies partly[11] in that it codifies Tarski reducibility. In languages that violate Tarski-reducibility, Criterion *T* is sometimes vacuous. What this means, in turn, is that Tarski biconditionals cannot be seen as giving the whole truth about the reference of "true", even if they do give the whole truth about the meaning of "true".[12] Rather, they must function (in principle) as decision procedures for evaluating the truth values of certain (simple) sentences in which truth predicates appear—and they just fail for truth tellers. Such sentences have truth values, but there is no way for us (even in principle) to determine what they are.

Let me sum up where we are. There are a couple of independent-looking intuitions about theories of truth that seem quite compelling as a group, but ultimately owe much of their justification to Tarski-reducibility. Furthermore, there is no easy argument for Tarski-reducibility, any more than there is for sentential bivalence. The best thing to do would seem to be to design theories of truth with one or another intuition in control and see which theories work better both as far as applications to the phenomena observed in natural languages is concerned, and in terms of their research value regarding self-reference.

3 What about the Liar's Paradox? Tarski-biconditionals and sentential bivalence together force them to be both true and false. This is not acceptable. If we go bivalent, therefore, we must make it plausible that Tarski biconditionals do not force liar's paradoxes to have both truth values.

We rejected the claim that Tarski-biconditionals exhaust the meaning of the truth predicate, suggesting that the biconditionals, *when they work*, enable us to read off the truth values of the sentences T*a*, and F*a* from the referent of *a*. Let us say that a constant *a transparently refers* if it refers, and if the truth values of the sentences T*a* and −F*a* agree with the truth value of *S*.[13] Now consider the following argument: Let it be assumed 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".[14]

Bivalence requires a certain loss of extensionality. Consider a sentence F*a* to which *a* refers. We have already granted that F*a* has a truth value, but since *a opaquely refers* to F*a*, it is not a truth value we can determine by looking at the referent of *a*. This does not preclude, of course, another constant *b* referring transparently to F*a*. In that case, if, say, F*a* is true, then although *a* and *b* refer to the same sentence, F*b* will be false. Putting the matter another way, we may understand the Tarski biconditionals as implying the condition of extensionality for truth and falsity predicates, and, on this approach, that condition sometimes fails.

We are not in a position to see that Tarski-biconditionals owe their intuitive plausibility not just to the intuition of Tarski-reducibility. Another intuition somewhat independent of Tarski-reducibility, is *truth value compositionality*: the intuition that the truth value of a sentence is a function of its semantic structure and the referents of the terms occurring in that sentence.[15] This intuition gains its strength from the fact that its scope is not restricted to sentences containing truth predicates, and it is this intuition that the above suggestions about liar's paradoxes violate.

Notice what is *not* being given up on this approach. Although it is clear that the truth value of a sentence F*a* does not turn directly on Tarski biconditionals and what *a* refers to, this does not make the sentence a meaningless black box. After all, its semantic structure enters into the recognition that the Tarski biconditionals fail for it; merely the further traditional move of allowing the composition of the sentence to determine its truth value has been disallowed: this is what failure of extensionality comes to in this case.[16]

There is a general assumption that ordinary language intuitions about a term are somewhat binding on formal theories about such a term, especially if the formal theory is going to be used in "descriptive linguistics" (e.g., see Salmon's [16], pp. 82–83). However, there is also a general assumption that naive *theories* about a term, even if they are couched in natural languages, are not particularly binding on formal theories, although they may serve as a guide for initial theorizing. The reason is not hard to find. The former intuitions are raw data for any linguistic theory, the latter are not.

Unfortunately, this crisp distinction more or less collapses when one turns to languages (such as natural languages) which appear to be self-referential. Are intuitions such as Tarski reducibility and truth-value compositionality part of naive theories or actual semantic intuitions about the truth predicate in natural languages? Some of their tenacity as constraints for theories of truth is due to the fact that they are taken to be raw semantic intuitions. But they may not be, and we may make progress by dropping them as global constraints.[17]

4 Let us explore the ramifications of the distinction between transparent and opaque reference. Consider a Buridean Symposium:

(1) (2) is false

and

(2) (1) is true.

We must deny the transparency of either (1) or (2). Unfortunately, denying one alone seems arbitrary and denying both seems redundant. This issue is resolved by the following crude picture. Imagine that we name sentences in baptismal ceremonies—ceremonies which can fail however. Although we always succeed in referring in one sense (opaquely), we can fail to refer transparently; it can be that the Tarski biconditionals cannot hold. Further, what particular baptismal ceremonies succeed in giving us transparent reference often turns on what other successful baptismal ceremonies we have already carried out. In this sense, the baptismal ceremonies are holistic. For example, all things being equal, if we have already baptized "(2) is false", with the name "(1)", but have not yet utilized "(2)", our reference to (1) is transparent. But if we go on to baptize "(1) is false", with "(2)", "(2)" will not transparently refer.[18]

It may be unclear how the analysis just given should be extended to quantified liars. Consider a sentence of the form $(\exists x)(Px \,\&\, \mathrm{F}x)$, where x is an individual variable, F is understood to be the falsity predicate and P is a predicate holding of one item, namely the sentence $(\exists x)(Px \,\&\, \mathrm{F}x)$. 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 reductio used to show that the constant a in the liar's paradox Fa does not refer transparently, directly to the variables. That is, let x be a variable which in a particular interpretation refers to $(\exists x)(Px \,\&\, \mathrm{F}x)$. Then by a reductio, it refers opaquely, consequently the satisfaction conditions of Fx will not utilize what x refers to, and therefore Fx (and $(\exists x)(Px \,\&\, \mathrm{F}x)$) will be true or false via the standard satisfaction clauses and the presence (nonextensionally determined) of Fx in the extension of either the truth predicate or the falsity predicate.

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.[19] Thus, instead, we argue that $(\exists x)(Px \,\&\, \mathrm{F}x)$ is true if a constant a may be found such that $(Pa \,\&\, \mathrm{F}a)$ is true. But then a, as we have seen before, may refer opaquely; and in those cases where $(\exists x)(Px \,\&\, Fx)$ is a liar's paradox, it is easy to see that every constant referring to it must do so only opaquely.

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 taking constants in its first place, sentences in its second place, and its intuitive meaning is: the constant occurring in its first place refers transparently to the sentence in its second place. Necessary and sufficient conditions of its truth are: (a) that the constant, a token of which appears in its first place, is mapped by the model to the sentence, a token of which appears in its second place, and (b) that the Tarski biconditionals hold.[20] As an example of its application, let a be a constant. Then $-(a/\mathrm{F}a)$ is true by the kind of reductio used above. That is, it is a theorem that a can never refer transparently to Fa. Variables can appear in the first place of an ostensive, and well-formed formulas which are not sentences can appear in the second place. Quantification binding such variables is understood substitutionally. The ostensive encodes that bit of *metalanguage* needed for the theory of truth and, in fact, substitutional quantification into it supplies the minimal amount of syntax the theory of truth needs.[21]

5 Further discussion requires that I present the model theory. So, in what follows I give and motivate what I call *O*-model theory. An *O*-language (Ostensive-language) is an extension of the predicate calculus to include designed predicate constants: T, F, and the ostensive symbol: /.

Definition 1 A *domain* of an O-language Λ is a set of objects Δ which either contains all the sentences of Λ, or none of the sentences of Λ.

This definition excludes certain peculiarities which would arise if domains were admitted containing only *some* of the sentences of Λ.

The model theory is built in stages. First *premodels* are defined (the name is self-explanatory):

Definition 2 A *premodel* for an *O*-language Λ on a domain Δ of Λ is the set containing:

(a) A mapping from the set of *n*-place predicates to the set of subsets of Δ^n, for each *n* such that
   (i) T is mapped to a subset of the set of sentences in Δ, and
   (ii) F is mapped to the complement of T's extension relative to the set of sentences of Λ,
(b) a mapping from the set of constants to Δ.

Notice that it is perfectly arbitrary what sentences appear in the extensions of the truth and falsity predicates, provided the extensions are disjoint and exhaust the sentences of Λ. We go on to define *satisfaction* for premodels and then single out the desirable "honest premodels", that is, those premodels in which the extension of the truth predicate contains exactly those sentences which are satisfied according to the definition.

Now recall that quantified liars are to be handled substitutionally; thus, in defining satisfaction, we need to supplement the standard mappings of the variables onto the domain with an assignment of individual constants to the variables. An example will make it clear why.

Suppose we want to evaluate the satisfaction value of the wff Tx. Substitutionally speaking, it will be satisfied by an interpretation ι if a constant a is mapped to the same sentence x is mapped to by ι, and Ta is satisfied by ι. But because of the nonextensionality of the system, it is possible for two individual constants, a and b, to be mapped to the same sentence x is mapped to, and yet Ta be satisfied by ι while Tb is not. So to make sure the forthcoming definition of satisfaction is well-defined, an interpretation not only ties a variable to an item in the domain, but also ties it to some individual constant, should such exist, mapped to that item.

Definition 3 An *interpretation* ι on a premodel Θ is a mapping ι' from the set of individual variables to the domain of Θ, and a partially defined choice function λ from the set of individual constants and individual variables to the set of individual constants. For individual variable v, λ selects a member from the set of constants $\{a \mid \Theta(a) = \iota'(v)\}$, and is defined on v provided the above set is nonempty. On the set of individual constants λ is the identity function.

We also need the following:

Definition 4 If ι is an interpretation on a premodel Θ, $v_1, \ldots, v_n$ individual variables, $A(v_1, \ldots, v_n)$ a wff, then if λ, ι's choice function, is defined on all v_i, the sentence obtained by substituting $\lambda(v_i)$ for v_i free in A is an *ι-instance* of A.

At last we can give the definition of satisfaction:

Definition 5 Let ι be an interpretation on a premodel Θ.

(a) $Pv_1 \ldots v_n$, an *n*-place predicate variable followed by a string of length *n* of individual variables and/or individual constants, is satisfied if the sequence of items of Δ which are the images of the individuals under ι' and/or Θ respectively in the order of $v_1, \ldots, v_n$ is in *P's* extension.
(b) $-A$, where A is a wff, is satisfied if A is not satisfied.

(c) $(A \vee B)$, where A and B are wffs, is satisfied if either A or B is satisfied.

(d1) (a/A), where A is a wff and a is an individual constant, is satisfied if there is an ι-instance A' of A which is assigned to a by Θ, and either Ta and A' are both in T's extension or both in F's extension.

(d2) (x/A), where x is an individual variable, and A is a wff, is satisfied if $\lambda(x)$ is defined, and $((\lambda(x)/A)$ is satisfied.

(el) Ta, where a is an individual constant, is satisfied if there is a sentence A in T's extension, and (a/A) is in T's extension; or A is a sentence assigned by Θ to a, (a/A) is not in T's extension, and either Ta is in T's extension or both Ta and Fa are in F's extension.

(e2) Tx, where x is an individual variable, is satisfied if $\lambda(x)$ is defined, and T$\lambda(x)$ is satisfied.

(f1) Fa, where a is an individual constant, is satisfied if there is a sentence A in F's extension and (a/A) is in T's extension; or A is a sentence assigned by Θ to a, (a/A) is not in T's extension, Ta is not in T's extension and Fa is in T's extension.

(f2) Fx, where x is an individual variable, is satisfied if $\lambda(x)$ is defined, and F$\lambda(x)$ is satisfied.

(g) $(\exists v)A$, where v is an individual variable and A a wff, is satisfied if either
   (i) there is an interpretation ι^* on Δ which disagrees with ι on at most $\iota'(v)$ and $\lambda(v)$ (if defined), and ι^* satisfies A with respect to Θ or
   (ii) a is an individual constant and Θ satisfies the wff gotten from A by substituting a for all free occurrences of v in A.

Most of this is standard looking. Notice that the truth conditions (g) for quantified statements are both substitutional and objectual. The point of clause (d1) is precisely to give the truth conditions for ostensives we have attributed to them above. Where x is a variable, (d2), (e2), and (f2) give the truth conditions for wffs of the form Tx, Fx, and (x/A) substitutionally in terms of $\lambda(x)$. Where a is an individual constant, (el) and (f1) give the truth conditions of Ta and Fa in terms of the Tarski biconditionals when ostensives are satisfied, and give them in terms of their own presence in the extension of the truth or falsity predicate otherwise. Finally, we want to rule out cases where, for a constant a, Ta and Fa are *both* in T's extension or in F's extension. (el) and (f2) rule these cases out by preventing such premodels from being honest.

Definition 6 An *honest premodel* for an *O*-language Λ on a domain Δ is a premodel for Λ on Δ which satisfies the additional condition: if the extension of T or F is nonempty, the subset of sentences of Δ that T is mapped to is the set of all the sentences satisfied by the premodel.

A definition is not valuable if it fails to pick anything out. So we want to know that honest premodels exist. More than that, if they are going to serve as a basis for a model theory, we want to know that enough of them exist. The following theorem satisfies us on both counts:

Theorem 7 *Suppose Δ is an arbitrary domain of an O-language Λ, and let Θ' be a mapping from the set of constants into Δ and from the set of n-place predicate variables to the set of subsets of Δ^n. Then there is an honest premodel*

Θ on Δ which agrees with Θ on its mappings of the individual constants and predicate variables into Δ.

Theorem 7 tells us not only that honest premodels exist but that the extra structure they contain does not make the model theory forthcoming any less general than the model theory of the standard predicate calculus. This means that honest premodels contain no hidden restrictions on how much syntactic self-description is compatible with honesty. In fact, Theorem 7 coupled with the fact that the axioms and rules of derivation of the standard predicate calculus are valid in *O*-model theory, will show that we have a conservative extension of the standard predicate calculus.

6 This looks like a nice place to stop. The extension of the truth predicate in honest premodels contains precisely the sentences satisfied by the model, and the definition of honesty does not place any restrictions on how constants and predicate variables may be mapped into an arbitrary domain. However, there are two complications that make the choice of honest premodels not entirely satisfactory for the model theory. We take the easier one first. Generally, the set of models for a consistent set of sentences is supposed to be the set of possibilities (set-theoretically construed) consistent with those sentences. In a situation where the sentences of the language itself are among the items describable by a set of sentences, one kind of possibility currently ruled out is that where the language itself is augmented with additional vocabulary (individual constants). Thus it seems desirable to increase the range of models to include those where the *O*-language Λ has additional vocabulary.[22] The next two definitions carry this out.

Definition 8 Let Λ and Λ^* be two *O*-languages. We say that Λ^* is *at least as articulate as* Λ iff the set of individual constants of Λ is a subset of the set of individual constants of Λ^*.

Definition 9 An *O-model* for an *O*-language Λ on a domain Δ is an honest premodel on Δ for a language Λ^* which is at least as articulate as Λ. A *primary O-model* for an *O*-language Λ is an honest premodel of Λ.

Honest premodels are not entirely satisfactory for another reason. They do not, as it were, force a name to refer transparently to a sentence if it can so refer. For example, take a perfectly innocuous sentence such as *Pa*, where *P* is a one-place predicate variable, and consider a constant *b* which refers to *Pa*. There are honest premodels in which *Pa* is true but (*b/Pa*) is false, merely because T*b* is in the extension of the falsity predicate while F*b* is in the extension of the truth predicate. Nothing so far rules out such honest premodels. Notice that how "truthlike" the truth predicate is in an honest premodel turns exactly on how many ostensives are satisfied in the models: Tarski biconditionals containing sentences and constants mapped to such sentences do not hold if the ostensives containing those sentences and constants are not satisfied.

Insofar as what is being offered here is a general axiomatizable framework for studying languages with their own truth predicate, and the kinds of pathology that such languages give rise to, this is not a drawback for two reasons. First, the model theory being offered properly includes all the desirable models—if they can be picked out easily then it is a simple matter of supplying additional constraints (either model-theoretically or axiomatically) to honest premodels and then studying the more restricted theory.[23] Second, the purpose of this paper is to supply a general *axiomatizable* framework for studying the theory of truth in a sententially bivalent framework. I suspect that sharper characterizations of the desirable models simply are not axiomatizable[24]; more accurately, it may turn out that what is offered here is maximally satisfactory in first-order terms; that is, model-theoretic methods of restricting the class of honest premodels or axiomatic restrictions on the class of models may cost us either axiomatizability or model-theoretic generality.[25] As before, should this be false, the framework here is easily modified.

I shall illustrate exactly the worries of the last paragraph by offering one kind of intuitively desirable restriction and showing what costs it involves. We noted earlier that the truthlikeness of the truth predicate turns directly on the number of ostensives satisfied. Why not then place a maximality constraint on honest premodels? An honest premodel is not maximal if it is identical to another, except for additional ostensives that are satisfied (and some shifting of the truth values of sentences of the form Ta and Fa).

Definition 10 A *maximally honest premodel* Θ for an O-language Λ on a domain Δ is an honest premodel for Λ on Δ which satisfies the additional condition: let Θ^* be an honest premodel for Λ on the domain Δ which agrees with Θ on the mappings of the predicate variables and individual constants, and satisfies every ostensive sentence satisfied in Θ. Then every ostensive sentence satisfied in Θ^* is satisfied in Θ.

The maximal transparency condition, although intuitively plausible, may still seem not to do enough. For consider an innocuous sentence such as Pa, and suppose there is a model where b is mapped to Pa and Pa is true. Intuitively perhaps (if we reject temporal priority for a kind of semantic priority), nothing should stop (b/Pa) from being satisfied in that model. But should the sentence (c/Fb ∨ Fc) be satisfied in it, (b/Pa) will not be satisfied. Such models exist, and further conditions must be stipulated to rule them out.[26]

But self-referential pathology does still more damage to the maximality constraint. We need to know that this constraint, and indeed any constraint we might dream up, does not restrict the model theory in unacceptable ways. In particular, we want to know that a version of Theorem 7 holds for maximally honest premodels. Unfortunately, it does not:

Theorem 11 *There is an O-language Λ, a domain Δ, and a standard mapping Θ on Δ such that there is no maximal primary O-model which agrees with Θ on the mappings of the individual constants and predicate variables of Λ.*

Proof sketch: Consider an O-language Λ with countably many constants and countably many one-place predicate variables P_i, and consider a domain containing only the sentences of Λ. Single out a particular individual variable: x. A *nice sentence* of Λ is a sentence of the form $(\exists x)(P_i x \ \& \ \mathrm{F}x)$. A *nice mapping* is a bijection of the set of constants of Λ onto the set of nice sentences. Now consider the set M^* of nice sentences. A *nasty mapping* is a mapping of the predicate variables to the cofinite sets of sentences of M^* so that each cofinite set has one and only one predicate variable mapped to it *and* if P_i is mapped to a cofinite set, then $(\exists x)(P_i x \ \& \ Fx)$ is in that set. Such mappings can be easily seen to exist.

Now prove that given this domain, a nice mapping of the constants to the domain and a corresponding nasty mapping of the predicate variables to the domain, no maximally honest premodel exists which agrees with these mappings. First notice that for any finite set of nice sentences, there is a primary O-model agreeing with the nice and nasty mappings in which precisely the sentences of M^* containing those nice sentences are satisfied. Next, notice that in no primary O-model agreeing with these mapping can a countably infinite collection of ostensives be satisfied. This gives us the desired result.[27]

7 I now want to give the axioms, so a few definitions are needed. An implicit terminological convention here is that for wff A, and individual variables and/or individual constants $v_1, \ldots, v_n$, $u_1, \ldots, u_n$, $A[u_i{:}v_i]$ is the result of free-substituting v_i for u_i in A, for all i.

Definition 12 Two wffs $B(v_1, \ldots, v_n)$ and A are *alphabetic variants* if $u_1, \ldots, u_n$, $v_1, \ldots v_n$ are individual constants and/or individual variables, A is gotten from B by substituting u_i for free v_i in B, no instance of u_i so substituted is bound in B, and further for all i, not both u_i and v_i are individual constants

Definition 13 If $u_1, \ldots, u_n$ are individual variables then we call $(\exists u_1) \ldots (\exists u_n)$ an *existential prefix*. We say that $u_1, \ldots, u_n$ *occur in* $(\exists u_1) \ldots (\exists u_n)$.

Definition 14 If u and v are individual variables, $B_1, \ldots, B_n$ wffs, $u_1, \ldots, u_n$ individual variables such that u_i is u if B_i does not contain u free, $A_1, \ldots, A_m$ wffs in which u occurs freely, $v_1, \ldots, v_m$ individuals which differ from u, $\Sigma_1, \ldots, \Sigma_n$ (possibly null) existential prefixes in which u does not occur, and $\Sigma'_1, \ldots, \Sigma'_m$ existential prefixes such that u and v_i do not occur in Σ'_i although v does, then any wff of the form $(u)(\mathrm{T}u \vee \mathrm{F}u \vee \Sigma_1(u_1/B_1) \vee \ldots \vee \Sigma_n(u_n/B_n) \vee \Sigma'_i((v/A_1) \ \& \ (v/A_1[u:v_1])) \vee \ldots \vee \Sigma'_m((v/A_m) \ \& \ (v/A_m[u:v_m])))$ is a *covering wff.*

These are called covering wffs. because no closure of any such wff can hold in any model in which an item (sentence or otherwise) has no individual constant mapped to it. This is simply because of the substitutional nature of the satisfaction conditions for truth predicates and ostensives. Notice that the variable u_i in the closure of a covering wff may be captured either by the existential prefix Σ, the quantifier (u), or by a closure quantifier, depending on what it is. Covering wffs also require the domain to contain sentences, and in models satisfying the closure of a covering wff in which no ostensives are satisfied there can only be sentences in the domain.

Definition 15 If u and v are individual variables, $B_1, \ldots, B_n$ wffs such that none of the B_i's contain u free, $A_1, \ldots, A_m$ wffs in which u occurs freely, $v_1, \ldots, v_m$ individuals which differ from u, $\Sigma_1, \ldots, \Sigma_n$ (possibly null) existential prefixes in which u does not occur, and $\Sigma'_1, \ldots, \Sigma'_m$ existential prefixes such that u and v_i do not occur in Σ'_i although v does, then any wff of the form $(u)((-\mathrm{T}u \ \& \ -\mathrm{F}u) \vee \Sigma_1(u/B_1) \vee \ldots \vee \Sigma_n(u/B_n) \vee \Sigma'_1((v/A_1) \ \& \ (v/A_1[u{:}v_1])) \vee \ldots \vee \Sigma'_m((v/A_m) \ \& \ (v/A_m[u{:}v_m])))$ is an *uncovering wff.*

These are called uncovering wffs because in models containing sentences, and in which the closure of such a wff is satisfied, most of the sentences in the domain cannot have constants mapped to them. Notice that closures of uncovering wffs are trivially satisfied in models containing no sentences.

We can now give the axioms and inference rules.

Inference rules:

(R1) For wffs A and B, if a closure of A and closure of $(A \Rightarrow B)$ are theorems, then any closure of B is a theorem.

(R2) For wffs A and B and individual variable v, from any closure of $(A \Rightarrow B)$ we may deduce any closure of $((\exists v)A \Rightarrow B)$, provided that v does not occur free in B.

Axioms proper:

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

(A2) *For an individual variable v, and wffs A and B such that B differs from A in that every free occurrence of v has been replaced by an individual constant or individual variable u, and further, A has the same number of free occurrences of variables as B if u is an individual variable, then any closure of*

$$B \Rightarrow (\exists v)A$$

is an axiom.

(A3) *If A is a wff and v is an individual variable then any closure of*

$$(v)((v/A) \Rightarrow ((\mathrm{T}v \vee \mathrm{F}v) \ \& \ (\mathrm{T}v \Leftrightarrow A)))$$

is an axiom.

(A4) *If v is an individual variable then*

$$(v) - (\mathrm{T}v \ \& \ \mathrm{F}v)$$

is an axiom.

(A5) *If v is an individual variable, A and B distinct wffs which are not alphabetic variants of each other, then any closure of*

$$(v) - ((v/A) \ \& \ (v/B))$$

is an axiom.

(A6) *If u and v are individual variables, $u_1, \ldots, u_n$, n individual variables and/or individual constants among which u occurs, $v_1,.., v_n$, n individual variables and/or individual constants among which v occurs, and the u_i's and v_i's such that for all u_i which are not u, u_i is v_i and for u_i which are u, v_i is either u or v, A is any wff, and P is any n-place predicate variable, then any closure of*

$$((u/A) \ \& \ (v/A)) \Rightarrow (Pu_1 \ldots u_n \Leftrightarrow Pv_1 \ldots v_n)$$

is an axiom.

(A7) *If A, B are wffs in which the individual variable u appears free, v and w individual variables, Σ an existential prefix in which the variable x appears but not v and w, then any closure of*

$$(\exists v)((w)(A[u:w] \Rightarrow \Sigma((x/B[u:w]) \ \& \ (x/B[u:v]))) \ \& \ -A[u:v]) \Rightarrow -(\exists u)A$$

is an axiom.

(A8) *If A is a covering wff, and B is an uncovering wff, then any closure of*

$$-(A \ \& \ B)$$

is an axiom.

The meaning of the axioms, strictly speaking, must be read off the model theory. Indeed, once the motivations for the model theory have been established, and clearly understood, there should be no surprises in the content of the axioms. One must keep in mind, of course, that quantification into ostensives or semantic predicates must be understood substitutionally, since this contributes to the trickiness of interpreting the axioms. A substitutional reading is metalinguistic (the constant that goes into the slot gets mentioned) but an objectual reading is not (the constant that goes into the slot gets used). Let us take them one by one: (A3) tells us that necessary conditions for any constant a transparently referring to a sentence A are: either Ta or Fa is satisfied, and the Tarski biconditional with respect to a and A must also be satisfied. (A4) tells us that nothing is both true and false. (A5) tells us that no constant can transparently refer to more than one sentence. (A6) gives us a version of Leibniz's Law. It can fail, as already pointed out, for ostensives and for semantic predicates, but it cannot fail otherwise.[28] (A7) looks more complicated. We decompose it as follows. Consider the clause $\Sigma((x/B[u:w]) \ \& \ (x/B[u:v]))$, and consider it free only for w. Then it is satisfied only for those interpretations where the constant substituted for w is the same as that substituted for v. Thus the clause $(\exists v)((w)(A[u:w] \Rightarrow \Sigma((x/B[u:w]) \ \& \ (x/B[u:v]))) \ \& \ -A[u:v])$ tells us that $A[u:w]$ will be satisfied at most by interpretations where the same constant is substituted for w as is substituted for v. The entire axiom therefore says that if $A[u:w]$ is not satisfied by such a constant, then it cannot be satisfied by anything. (A8) is self-explanatory given the glosses for Definitions 14 and 15.

These axioms suffice to single out those sentences true in every *O*-model:

Theorem 16 *Let an O-language Λ be given. Inference rules* (R1) *and* (R2) *are truth preserving with respect to O-model theory, and each instance of axioms* (A1)–(A8) *hold in any O-model of* Λ.

Theorem 17 *If* Ω *is a consistent set of sentences of* Λ *with respect to* (R1), (R2), (A1)–(A8), *then there is an O-model* Θ *such that every sentence of* Ω *is satisfied in* Θ.

What these axioms do *not* do (nor have they been designed to do) is characterize honest premodels relative to premodels.[29] That is, there are premodels which are not honest in which all the axioms (A1)–(A8) are satisfied. This is simply because the satisfaction conditions *only* connect membership in the semantic predicates with satisfaction when ostensives are not satisfied, and the axioms are largely concerned with the circumstances under which ostensives are satisfied.

But therefore the following corollary follows from Theorem 17:

Corollary 18 *If* Θ *is a premodel in which all instances of axioms* (A1)–(A8) *are satisfied, then there is an O-model* Θ' *which is elementarily equivalent to* Θ.

8 An ideal hinted at in Kripke's approach and more or less adhered to by others in the field is a kind of representation requirement on a theory of truth. If one constructs a framework for self-referential languages, it seems that it should be possible to represent the full array of "peculiar" sentences in that framework. Not to be able to do so is *prima facie* evidence that the theory is self-referential only in a Pickwickian sense.

There are two sides to this issue for our approach. First, there is the model theory. That certain kinds of pathology are representable must be shown by consistency theorems—by showing that certain kinds of models exist. But in addition, the ostensives supply a bit of metalanguage in the language itself, and using them we can represent the kind of reasoning involved when pathology (or just certain kinds of virtuous circular reasoning) is present. I am going to focus on both kinds of phenomena here, although since consistency theorems are a bit more involved, I will tend to assert without proof the existence of models of various kinds.

Let us start with the Liar. I have stated that no constant a can refer transparently to Fa. This follows from the axioms:[30]

(1) $(a/\mathrm{F}a) \Rightarrow ((\mathrm{T}a \vee \mathrm{F}a)\ \&\ (\mathrm{T}a \Leftrightarrow \mathrm{F}a))$ (instantiation of a for v in (A3))

(2) $-(\mathrm{T}a\ \&\ \mathrm{F}a)$ (instantiation of a for v in (A4))

(3) $-(a/\mathrm{F}a)$ (application of tautology, m.p. to (1) and (2))

Liar's cycles are also popular examples of pathology. These are constructed with n constants $a_1, \ldots, a_n$ where each a_i, $i < n$, refers to $\mathrm{T}^*_{i+1}a_{i+1}$, a_n refers to $\mathrm{T}^*_1 a_1$, and where for each i, T^*_i may be either F or T. If an even number of T^*_i's are F's, the sentence $((a_1/\mathrm{T}^*_2 a_2)\ \&\ldots\&\ (a_n/\mathrm{T}^*_1 a_1))$ is consistent, otherwise, using reasoning similar to that applied to the Liar, we can show that $-((a_1/\mathrm{T}^*_2 a_2)\ \&\ldots\&\ (a_n/\mathrm{T}^*_1 a_1))$ is a theorem. Furthermore, it can be shown that for any $n-1$ of the ostensives, there is a model in which they are all satisfied although, of course, the remaining ostensive is not.

Contingent liars also occur here. For example, $(a/(\mathrm{F}a \vee (\exists x)Px))$ will be satisfied in some *O*-models[31] where something is a P and will not be satisfied in any model where nothing is a P.

Truth tellers are a phenomenon which are studied entirely through consistency proofs. There are models in which $((a/\mathrm{T}a)\ \&\ \mathrm{T}a)$ is satisfied and models in which $((a/\mathrm{T}a)\ \&\ -\mathrm{T}a)$ is satisfied. Similarly, there are models satisfying the infinite set of sentences: $\{(a_1/\mathrm{T}a_2), \ldots, (a_n/\mathrm{T}a_{n+1}), \ldots\}$ where, for all i, $\mathrm{T}a_i$ are all true or all false.

Finally, let us review Gupta's puzzle:[32] imagine that John is not running although Peter is running, and consider two sets of claims:

R's claims: (a) John is running.
(b) All of the claims made by *P* are true.
(c) At least one of the claims made by *P* is false.

P's claims: (d) Peter is running.
(e) At most one of the claims made by *R* is true.

Now Gupta has observed that we normally reason as follows. First, since (b) and (c) contradict each other, at most one can be true. Further, since (a) is false, (e) must be true. Finally, therefore, (c) is false, and (b) is true. To see how my approach treats this case, let us use "*a*", . . . ,"*e*" as names, and assume the following ostensives are satisfied: (a/Rj), $(b/(Td \,\&\, Te))$, $(c/(Fd \vee Fe))$, (d/Rp), $(e/((\mathrm{T}a \,\&\, \mathrm{F}b \,\&\, \mathrm{F}c) \vee (\mathrm{F}a \,\&\, \mathrm{T}b \,\&\, \mathrm{F}c) \vee (\mathrm{F}a \,\&\, \mathrm{F}b \,\&\, \mathrm{T}c))$.[33] The point of course is that the intuitive reasoning described above takes it for granted that there is no problem with transparent reference to any of these sentences by the constants chosen. If we let *A* through *E* indicate respectively each of the ostensives above, then

(1) $(A \,\&\ldots\&\, E \,\&\, -Rj \,\&\, Rp) \Rightarrow (\mathrm{T}b \,\&\, \mathrm{F}c \,\&\, \mathrm{T}e)$ is a theorem,

and

(2) $(A \,\&\ldots\&\, E \,\&\, -Rj \,\&\, Rp)$ is consistent.

What is nice is that the intuitive reasoning described by Gupta can be closely imitated by the proof of (1), suggesting that the axioms of the system are quite natural.

9 I have claimed above that *O*-model theory provides a general framework in which one can study the consequences of a variety of approaches to the logic of truth. On the other hand, it might seem that the ostension operator is too complicated and idiosyncratic (vis-à-vis other approaches) for this to be true. In point of fact, my claim is that the model theory is general enough to contain (pretty much) *all* the desirable classes of interpreted languages containing their own truth predicate in the classical setting. As an indication of this fact, I go on to sketch how the technique used in the Gupta/Herzberger approach can be employed here to pick out a natural subclass of *O*-models. Their approach turns on exploiting the differences between the set of sentences satisfied in a model and the set of sentences in the extension of the truth predicate in that model. Since *O*-models do not have this property, we cannot directly apply the semi-inductive construction to *O*-models. Instead we define a mapping from a suitably modified class of Gupta/Herzberger models to *O*-model theory.

Arbitrary premodels in *O*-model theory allow the predicates and the constants to be mapped to the sentences of the language in any way at all. In this sense, *O*-model theory is *impredicative*. But since the Gupta/Herzberger approach does not employ the ostensive, we will consider syntactic theories, and truth theories which exclude the ostensive:[34]

Definition 19 A *predicative premodel* for an *O*-language Λ on a domain Δ is defined similarly to a premodel except condition (b) is replaced by:

(b′) A mapping from the set of constants to Δ such that no constant is mapped to any sentence containing an ostensive. We call a mapping of this sort from the set of constants to Δ a *predicative mapping*.

Notice that this effectively bars the language from semantic recognition of sentences containing ostensives even though the truth predicates still (and must) contain such sentences. This is because all semantic reference occurs substitutionally via the constants. Since predicative premodels are a strict subclass of the class of premodels, the other definitions can remain the same.

Predicative O-model theory can be characterized by two additional axioms:

Definition 20

(A9) If A is a covering wff then any closure of

$$-A$$

is a *predicative axiom.*

(A10) If u is an individual variable, and A is any wff containing ostensive symbols then any closure of

$$-(\exists u)(u/A)$$

is a *predicative axiom.*

(A9) is necessary because whenever any sentences of an O-language are in the domain, all of them are. But in predicative O-models, no constants are mapped to the sentences containing ostensives and so covering wffs are never satisfied in such models. (A10), of course, just states that transparent reference to sentences containing the ostensive operator never holds.

We have the following:

Theorem 21 *Let an O-language* Λ *be given. Each instance of Axioms* (A9) and (A10) *holds in any predicative O-model.*

Theorem 22 *If* Ω *is a consistent set of sentences of with respect to* (R1), (R2), (A1)–(A10), *then there is a predicative O-model* Θ *such that every sentence of* Ω *is satisfied in* Θ. *Furthermore,* Θ *can be taken to be maximal.*

Notice the following. Although O-languages have the capacity to recognize when a constant transparently refers to a sentence in a domain, they have no capacity (on pain of contradiction) to mark out when a constant is merely mapped opaquely to a sentence in a domain. Consequently, although axioms (A9) and (A10) hold in predicative O-models and a completeness result regarding such models holds, it is *not* that case that every O-model in which such axioms hold is predicative. For example, there are honest premodels in which a constant a is mapped to an ostensive S (and a is the *only* such constant to be mapped to a sentence containing an ostensive operator) but (a/S) is not satisfied. (A9) and (A10) are satisfied in such a model. The proof of Theorem 22 therefore implies the following:

Corollary 23 *Let* Θ *be an O-model in which all instances of Axioms* (A9) *and* (A10) *are satisfied. Then there is a predicative maximal O-model* Θ' *such that* Θ *and* Θ' *are elementarily equivalent.*

There are several differences between O-models and the Gupta/Herzberger models that have to be ironed out before anything like an embedding theorem can be shown. First, since the truth predicate is substitutional in O-models, O-

model theory will not recognize a sentence if it does not have a constant mapped to it. So we are only going to pay attention to those *O*-models in which every sentence without an ostensive has at least one constant mapped to it:[35]

Definition 24 A *well-covered predicative premodel* for an *O*-language Λ on a domain Δ is defined similarly to a predicative premodel except that condition (b′) is replaced by:

(b″) A mapping from the set of constants to Δ such that no constant is mapped to any sentence containing an ostensive, although every other sentence of Λ has at least one constant mapped to it.

Because of two largely notional matters, well-covered predicative *O*-models are still not quite right for embedding the models generated on the Gupta/Herzberger approach. First, as the constructions are presented, a separate falsity predicate is not used—rather, "−T" suffices. Clearly, the construction can easily be modified to accommodate a falsity predicate. But there is an additional complication. Since either all sentences of the *O*-language are contained in the domain or none of them are, all predicative *O*-models act as if they contain nonlinguistic items.[36] Consequently we consider a slightly modified Gupta/Herzberger approach where: (a) two semantic predicates are defined: T and F, and (b) the domains contain a countable number of items which are in neither the extension of the truth predicate or the falsity predicate.

With these caveats, we have the following theorem:

Theorem 25 *Let Λ be a language without ostensives, but with the predicates:* T *and* F. *Let Θ be a model generated at any stage utilizing any semi-inductive construction in the Gupta/Herzberger approach, where Θ contains a countable number of nonsentences in its domain and where every sentence has at least one constant mapped to it. Then there is an O-model Θ' which has the same domain as Θ, and which agrees with Θ on its mappings of the constants and the predicates (including the truth predicates) to the domain.*

Proof Sketch: We need only expand the extension of the truth predicates to contain sentences with ostensives. This is done by defining the extension of the truth predicate along the lines of the satisfaction clauses in Definition 5.

We'll call the Gupta/Herzberger models mentioned in the above theorem *GH*-models and the *O*-models mentioned *GH′*-models. We make two observations:

(1) In general it is not the case that a sentence is satisfied in a GH-model Θ iff it is satisfied in the GH′-model Θ'. This is simply because every sentence in the extension of the truth predicate of an *O*-model is satisfied, whereas this is not true, in general, in a *GH*-model. However, in *GH*-models Θ which are *best candidates* (see Gupta's [5], p. 221), it is true that every stable sentence is satisfied in Θ iff it is in the truth predicate of Θ. Furthermore, it is true that every stable sentence in Θ is transparently referred to in Θ'. Actually, in general, usually a much larger class of sentences than the stable set of sentences is transparently referred to in Θ'.

(2) One might have hoped for a result something like this: For Θ a best candidate, Θ' in Theorem 25 may be taken to be maximal. Unfortunately, slightly modifying the proof of Theorem 11 shows that this is not true.

10 It is time to give an overview of what has been done here and draw a few tentative conclusions. Ultimately, I am interested in the grand project of providing formal models of self-referential languages. I understand such languages to be ones which can (largely) describe their own semantics and syntax. What such languages can look like, what restrictions must be placed on them, is completely open. What has been done here is something quite a bit narrower: the more manageable project of providing formal models for languages with their own truth predicates has been tackled instead. A broad framework (a model theory) and axioms that are complete with respect to the framework have been supplied.

How does this relate to the broader project? Those who like to use the standard predicate calculus as a base on which to build additional structure will find this approach congenial. Those who do not may still find much of interest here since it is possible that moves made here may be grafted onto approaches quite different otherwise.

How significant is the fact that the model theory is axiomatizable? At one time the answer would have been regarded as obvious, but in the current model-theoretic atmosphere, it is sometimes almost suggested that for a mathematical structure to be axiomatizable is a weakness.

Perhaps this is true here. But let us recall that it has been argued that there is a difference between a subject matter which is complete and one which is incomplete—and that this marks where logic ends and mathematics begins.[37] It has also been claimed that if our theorems for a subject matter are outrun by the subject matter, there is a sense in which we do not have a full grasp on what we are talking about.

Be that as it may, let me suggest a weaker position. Perhaps it *is* controversial to draw the border between *logic* and something else on the basis of completeness or incompleteness, especially if higher-order principles are employed in "reasoning". Nevertheless, the distinction marks something important and rather natural—perhaps the boundary line for concepts totally accessible to us (epistemologically), and certainly it could be that the concept of truth is simple enough to belong among these concepts. As I pointed out in Section 1, this possibility can be explored only if the capacity of a language to describe its own syntax is relativized to a model.[38] So, for example, it is not required that in every model every sentence can be referred to. But this requirement *really is* independent of the theory of truth: often self-referential reasoning involving truth presupposes only limited capacity to refer to sentences, e.g., the reasoning involved in Gupta's puzzle holds in models in which only the five sentences relevant are referred to. The point, of course, is not that such description is called for sometimes; rather, it is that the theory of truth does not presuppose it. This suggests that truth really is part of logic. The theory fits comfortably with standard logic and actually, it and the small bit of metalanguage it requires can be regarded as a natural extension of the first-order predicate calculus.[39]

It may actually be that no natural language has a built-in capacity for total syntactic self-description. If so, this can supply further motivation for the broadening of the model theory to include models with strictly richer vocabulary than that of the language they are models of. When needed, we forge additional vocabulary. We may even make assertions now that only hold if our language is augmented with additional vocabulary. In this odd sense natural languages may be regarded as openended. But none of this affects the truth predicate.

It might be argued that crucial to truth is the capacity for semantic ascent, and on this approach, precisely this natural element has been eliminated. For, in general, one needs to know that transparent reference to a sentence succeeds before one can make assertions about its truth. That is, if the sentence S is named by the constant a, one needs to assert not only Ta, but (a/S) as well. This in turn implies that the truth of sentences can only be asserted one at a time, and provided the sentences are listed explicitly.

In fact there are two issues that have been jumbled together in the above objection and they should be separated carefully. First off, there is the question of whether the respective notions of "stable", "grounded", and "transparently referred to by a constant" are expressible in the language. Generally, if the language is strengthened so that not only can a sentence's truth value be asserted, but also its unproblematic status, one faces the spector of the strengthened liar. For example, on Kripke's approach, one needs to know that one's assertions are *grounded*, but only if the object language is very weak can such a predicate be explicitly introduced into it (Gupta and Martin [6]). Similarly, on the Gupta/Herzberger approach, one needs to know that one's assertions are stable. Such a predicate can be introduced into the object language and treated much as truth is, but it will have problems similar to those that motivate us to introduce the predicate itself in the first place. With ostensives, one can assert the unproblematic status of sentences in the object language itself but only on an individual basis. Strengthening matters will result in inconsistency just as on other approaches. For example, it is easy to see that if a semantic predicate TR were introduced impredicatively, where TRa held only if a referred transparently to some sentence, then the resulting system would be inconsistent. Once one accepts languages which contain (something like) their own truth predicate, the world has been made unsafe for semantic ascent. Asserting facts about truth can get us in trouble (force us to assert problematical sentences).

But there is another point to be made here. Both the Gupta/Herzberger approach and Kripke's approach make semantic ascent safe insofar as, if one has managed somehow to get his hands on a (possibly infinite) collection of unproblematic sentences (grounded, stable, etc.), one can then go on to assert other sentences about the truth values of the definable[40] subclasses of these sentences and such sentences will be unproblematical as well. Of course, as I have noted immediately above, this kind of fact is not generally expressible in the language itself, but it is true for all that. However, in O-model theory, this fact is *not* true simply because it need not be the case that the additional sentences are transparently referred to.

As I have said before, O-model theory is a general framework, and I have resisted any further specific structural constraints on O-models simply because I am unsure which approach to this kind of problem is the best one. Certainly I have already (in Section 9) given indications that O-model theory is rich enough to supply sufficient models for more restricted model theories (such as those generated on the Gupta/Herzberger approach). But whether that is the best way to go here, and what proof theoretic properties such restricted model theories have is still open.[41]

As far as the value of *O*-model theory for research is concerned, I have already pointed out that *O*-model theory is a framework that supplies a variety of tools for studying the results of further constraints on primary *O*-models. This is simply because research can proceed along two fronts: axiomatically and model theoretically. This offers a variety of classification schemes for possible models.

I should also remind the reader that *O*-models also turn out to be quite general in two other respects. First, there is the matter of syntax,[42] and, secondly, there is the nonextensionality of the system. Regarding the second point, it is worth noting that perhaps the primary *O*-models corresponding to the best candidates on the Gupta/Herzberger approach are not the most interesting to be found in *O*-model theory. First of all, the theory of syntax in such a model does not recognize the existence of ostensives. Secondly, the truth predicate in such models is extensional, and generally the class of problematical sentences is smaller when the truth predicate is not extensional. On the other hand, they may be the broadest class of primary *O*-models where: (i) the truth predicate is extensional, and the syntax does not recognize ostensives, but where (ii) local determinability is not violated for unproblematical sentences.[43]

Acknowledgment—The work here is largely based on my dissertation and my debts are numerous. I am grateful to Melvin Fitting, Anil Gupta, Hans G. Herzberger, Jerrold J. Katz, Arnold Koslow, Mark Richard, David M. Rosenthal, and an anonymous referee for their helpful suggestions and thoughtful criticisms. I also want to extend a thanks for some helpful comments to the faculty at The University of Notre Dame and the faculty at Tufts University, where I read ancestors of this paper in the fall of 1984. Finally, let me acknowledge decade-old debts to both John Bacon and Raymond Smullyan with whom (on separate occasions) I had several conversations on the Liar's Paradox and self-reference.

NOTES

1. However, I should add that Herzberger has taken some steps in the direction of codifying the laws the truth predicate obeys on his approach (see [7]). Also, more recently, Michael Kremer has offered a logic of truth based on Kripke's approach (see [11]).

2. See Note 42 below, where I discuss this explicitly. Of course, the incompleteness of syntax is hardly the only obstacle to an attempt to axiomatize the truth predicate. If one does not fix the syntactic capacity of all the interpreted languages in advance, then one must have some way of expressing in the interpreted language itself something like the connection between the names in a model and the sentences they are names of (without, of course, landing in paradoxes as a result) so that, for example, one can express Tarski biconditionals. E.g., a look at Gupta's Definition 1 on p. 180 of [5] shows that he does not fix in advance the capacity of an interpreted language to describe its own syntax, but as a result a straightforward axiomatiza-

tion of, say, the *best candidates*, if possible, might not reveal anything particularly interesting about the truth predicate. There are also other complications pertaining to axiomatizibility which I discuss further in Sections 6, 9, and 10.

3. I have described Tarski-reducibility intuitively, and consequently, vaguely. In point of fact, the term "fix" can be cashed out in any number of ways which are not equivalent. For example, in certain languages without quantification, a *redundancy* theory of truth will do the job. Otherwise, if the language's capacity to describe its own syntax is sufficiently restricted, Tarski's approach will do. For still stronger languages, both the minimal and the largest intrinsic fixed points of Kripke's approach exemplify the intuition by excluding the sentences which violate Tarski reducibility. The latter fixed point focuses on the *arbitrariness* of the truth values (vis-à-vis the nonsemantic facts) of the sentences to be excluded whereas the former focuses on the largest class of sentences whose truth values are forced on them by a certain construction and the nonsemantic facts.

   On Kripke-type approaches, the intuition of Tarski-reducibility can be saved by recasting it as applying to all sentences of the language that "express propositions". Similarly, intuitions about bivalence may be saved by recasting them in terms of propositions. See Notes 5 and 7 below.

4. For the moment, I include under the rubric "Kripke-style", Gupta and Herzberger's approaches as well as Kripke's. I am being very sketchy in my discussion of all three approaches since they are so amply discussed elsewhere (e.g., Kripke's [12], Gupta's [5], Herzberger's [7] and [9], Parsons' [14], Yablo's [21], etc., etc.).

5. Several kinds of bivalence come into play here. First there is sentential bivalence and propositional bivalence, both taken in a narrow sense. They are to be distinguished in that the truth bearers are sentences in the former, whereas the truth bearers are propositions in the latter. However, I prefer to take sentential bivalence in the broader sense as Quine understands it in [15], that is, as requiring the satisfaction conditions on open sentences to be two-valued. Notice that it is easy to think of weird semantic rules which make any language that obeys them sententially bivalent in the broad sense, but not in the narrow; although this is impossible if the language contains its own truth predicate. Since I am concerning myself with the truth predicate alone, I speak of sentential bivalence in the text in the narrower sense but help myself to Quine's position, which applies, strictly speaking, to sentential bivalence construed widely.

6. Notoriously, Tarski sidestepped the issue by fiat. His languages, at least as traditionally construed, are both sententially bivalent and composed entirely of Tarski-reducible sentences. Of course, this is relative to the metalanguage.

7. An important caveat is necessary here. I have been arguing here that *truth tellers* exert pressure to give up either Tarski-reducibility or sentential bivalence in the classical setting. Nothing in the discussion here should suggest that doing so is sufficient to solve the *Liar's Paradox*, as the existence of the Strengthened Liar and the fact that the paradox is derivable in intuitionistic or minimal logic makes clear.

   These facts show more, of course. They show that it is relevant how *expressively complete* the languages is (vis-à-vis the set of truth value functions). This is no surprise since problems with the Liar's Paradox emerge via a derivation (of some sort). Sometimes this point is described in terms of possible restrictions on negation, but due to the (in general) holistic interplay of the connectives in derivations, this is not quite accurate. In any case, one may acquiesce in syntactic restrictions but still regard the semantics of the language as bivalent. This point clearly applies to what I have said in the text about Kripke's approach. First, I have laid the stress on sentential bivalence, but in doing so I am only being a kneejerk Quinean. If we take groundedness in Kripke's approach as criterial for *propositionhood*, it is doubtful that bivalence (in the serious philosophical sense) is lost. What value sentential bivalence has over propositional bivalence is a largely technical question which is still open. Second, sentential bivalence can be captured in Kripke's approach by "closing off" fixed points. There are objections to this, largely turning on the deviation of the resulting system from bivalent intuitions, but I forgo details since they are amply supplied in the literature (see, e.g., Parsons [14] and Gupta [5]).

8. It is odd, and worth noting, that Gupta's technical construction does not quite match the philosophical motivation he offers for it. The revision procedure really is a construction of a set of models in which the truth values of one set of sentences are eventually fixed while the truth values of the remaining sentences are left to cycle as long as the construction is continued. Notice how this is in striking contrast to, say, Intuitionism, where the provability concept is genuinely open to revision.

   Gupta, of course, is aware of this and seems to hang on to the revision picture anyway because of the appearance of the technique used to construct the desirable set of models. But perhaps the desirable models can be characterized in some other way, and in that case there really is no reason to regard the concept of truth as lacking an "application procedure".

   Philosophically I think there are roughly two options in interpreting these kinds of constructions. We can regard the resulting models as genuinely gappy (or *vague*), but nicer in certain respects than the Kripke models, or we can kick the question of which model "really" captures the truth predicate out of semantics and into epistemology, as I argue we should. Doing so does "result in a loss of connection between truth and assertion" as Gupta argues ([5], pp. 229–230), but only on the problematical sentences. There seems to be no gain in regarding the truth predicate as inherently vague on problematical sentences instead, nor does doing the latter result in a simpler semantics if it is not a requirement on the semantic rules for truth to supply a rule that picks out the "real" model in cases where Tarski-reducibility fails.

9. I should say that I am not the only one to make this kind of move. For very different reasons, Schiffer in [17] argues that what amounts to Tarski reducibility (he never uses the term) must be given up for propositional attitude sentences. See especially Chapter 6.

10. There is irony in the fact that pressing hard for the constraint of sentential bivalence in the context of self-referential languages forces one to give up this version of physicalism. By the way, this point runs fairly deep philosophically. I discuss it elsewhere [1]. I should also add that Barwise and Etchemendy's approach in [2] violates the reducibility constraint just as mine does by including "semantic facts" as possible elements in the model—and then making sure that the models containing semantic facts obey versions of the Tarski biconditionals as much as they can be made to. Here the resemblance ends. Truth bearers on their approach are propositions which contain as "constituents" the objects they are about (e.g., themselves, should they be self-referential). Although there are intuitions at work motivating this picture of propositions and even the nonstandard set theory used to construct them, there is no justification given for why semantic facts are included in the models. This is not to say, of course, that a justification could not be given. For example, motivation via bivalence, as in my approach, is not foreclosed for them (as it is for many who "solve" the paradoxes via a proposition-bearing theory of truth) since, e.g., liar's sentences *do* express propositions on their approach.

11. Criterion T draws its strength from the fact that it is the result of at least *two* powerful intuitions. I discuss the second one in Section 3.

12. In general, I understand the meaning of "true" to ultimately be given by the model theoretic apparatus, not by certain logical truths containing the truth predicate which are singled out because of their intuitive plausibility. But this caveat does not help since the apparatus does not, in general, single out a unique extension for the truth predicate either. Whatever we take as giving the meaning of the truth predicate will not determine its extension uniquely, except under certain extremely "nice" circumstances.

13. Let me say that I will use the term "opaque reference" to describe the relationship a term has to its referent when it does not refer transparently to that referent. Although my terminology is similar to talk of "transparent" and "opaque" contexts, and although the phenomenon described certainly is similar, my usage of "transparent" and 'opaque" should not be confused with that terminology (where, for example, a belief context is "opaque"—not opaque with respect to certain constants, although not necessarily with regard to others). Here, it is the term that can have opaque or transparent occurrences, not the context which is transparent or opaque (see, for example, p. 236 of Kaplan's [10] for a discussion of the distinction in another context). Also notice that here opaque reference is a necessary condition for transparent reference.

14. Of course this reductio is bogus. I have conveniently fixed the truth values of other assumptions to enable me to derive it. E.g., I assume sentential bivalence, e.g., I assume every sentence expresses a proposition, and, e.g., I assume any number of other assumptions that others in the literature have denied to "solve" the Liar's Paradox. That every name transparently refers in the context of a truth predicate is still another assumption that can be denied.

    Notice that the reductio is very similar to one Parsons gives on p. 16 of [14]. If what I subsequently call "truth value compositionality" is taken to be a necessary condition on a sentence's expressing a proposition, then my reductio can be taken to be quite similar in spirit to his. This is not an unnatural way to understand what I am up to, although, of course, my approach deviates seriously when we turn to quantification. Notice that on this version of my view, although liar's paradoxes do not express propositions (they do not get their truth values compositionally), that does not prevent them from being truth bearers. Incidentally, I am not sure what Parsons takes the philosophical significance of this reductio to be.

15. On some views there is redundancy in my formulation; in any case, what is clear is that no other factors should be involved than the ones I mentioned. Notice that allowing nonsemantic facts "in the world" allows a version of truth-value compositionality to be satisfied even if Tarski-reducibility is not.

    When Russell propounded his theory of descriptions to account for sentences of the sort: "The present King of France is bald", he was motivated, as has been remarked, by the desire to keep a sententially bivalent logic. *But* the intuitions we have been discussing were at work too: because of Tarski-reducibility, he did not feel comfortable claiming that "The present King of France is bald" is either true or false, but undetermined by the nonsemantic facts. Similarly, because of truth-value compositionality, he did not feel comfortable claiming the sentence has a truth value without giving an analysis of its meaning which enabled him to read off what its truth value should be. These are very natural assumptions, of course, and it is only the systematic gains vis-à-vis truth that drive me to give them up "around the edges". But once the apparatus is in place for handling truth the way I intend to, something similar may be done for the sentences Russell was concerned with. I cannot go further into the matter now.

16. A quick comparison with Skyrm's approach [18] is instructive here. He also suggests making the truth predicate nonextensional. But he also eschews bivalence. This enables him to honor truth-value compositionality. Thus, in self-referential languages, the presence of sentential bivalence forces truth-value compositionality to imply the extensionality constraint on the truth predicate.

17. Quine, early and late, has stressed the global aspect of meaning over the local. He has done so both in stressing the sentence over its parts, and in stressing the theory over the sentences which make it up. It has seemed to some, however, that Quine has made his epistemological insights do work in semantics where they do not belong. In fact, regimented languages on Quine's view do not look semantically global in any sense at all, for their semantics are strictly Tarskian. The global aspect comes in only in the process of regimentation, and it does so in much the same way that it does in theory construction. Once the epistemological points have been separated from the semantic ones, it begins to look perfectly wild to suggest that a sentence could have priority (semantically speaking) over its parts, or a theory over its sentences. Worse, it looks perfectly incoherent. *What* could such a theory look like? What kind of semantic rules could it possibly have? As it turns out, something like this idea is not incoherent at all. First, I will shortly give perfectly coherent semantic rules for the truth predicate which violate the truth-value compositionality. Second, the theory of truth here is made somewhat independent of a theory of syntax as I have already pointed out. In particular, the capacity to refer to whole sentences is called for, if the theory of truth is to do any work, but the capacity to refer to the parts that make them up is generally absent and unnecessary for semantic ascent.

18. The order in which opaque reference is established need not be a temporal one. We can imagine certain restrictions being placed upon constants so that certain ones are guaranteed transparent reference in cases of conflict. See Section 6 for further comments on this issue.

19. The approach I am not taking would call for extending the truth and falsity predicates (under an interpretation) to contain open sentences. I would also allow sentences of the form $(\exists x)(Px \,\&\, \mathrm{F}x)$ and $(\exists y)(Py \,\&\, \mathrm{F}y)$ to differ in truth value for distinct x and y. This, although perfectly understandable given that the variables act like constants, does violate intuitions about bound variables, and also complicates things quite a bit.

20. Is there anything like an ostensive in natural languages? This kind of question is ultimately hard to evaluate on its own. But certainly demonstratives occur, and the ostensive is pretty much a demonstrative pointing to the items within its parentheses.
    I should also add that the ostensive is designed to solve the problem mentioned in Note 5.

21. See Section 7 below where I give a reading of the axioms in English; this reading runs *both* a metalinguistic and a straight reading off the quantifier when the quantifier binds a variable occurring both in an ostensive and in a predicate *outside* an ostensive.
    I should add that substitutional quantification here has only a metalingustic interpretation. No sneaky moves to avoid ontological commitment are being contemplated.

22. Of course, I have a technical motivation for this move: the resulting model theory is axiomatizable. Whether such a motivation is desirable is something I discuss below. But it is of interest that a very natural and independent intuitive justification can be given for this move in the context of self-referential languages.

23. There is an analogy here to the characterization problem for total recursive functions. We are characterizing a broader, and consequently more manageable class of interpreted languages which, more or less, contain their own truth predicates, and hoping thereby to provide a framework for studying the narrower class.

24. I should add that Kremer's recent work [11] axiomatizing Kripke's approach suggests that there might be problems with finer characterizations of the desirable models. His approach takes as the models of the theory *all* fixed points, not just, say, intrinsic fixed points, or minimal fixed points, etc.

25. An example of what I have in mind as a model-theroretic restriction is defining a version of "grounded sentence" in this context and then giving such sentences priority over others or over each other according to a groundedness index. Another example which I discuss explicitly below is the notion of "maximally honest premodel." An example of an axiomatic restriction is to focus our attention on sets of models that satisfy certain collections of ostensive sentences so that every sentence (or every sentence obeying certain nice properties) is transparently referred to in such models.

26. What Gupta ([5], p. 196) calls local determinability is violated by maximally honest premodels. However, they are superior over honest premodels in that if transparent reference fails it must do so because of vicious reference elsewhere, not merely because the model misplaces sentences of the form T*a* and F*a*.

27. Of course, this is hardly the last word on the matter. The counterexample is pretty bizarre, especially in how the mappings of the individual constants onto the domain are restricted. In particular, it is not yet clear to me whether maximally honest premodels present much of a restriction in cases where the syntax is more standard looking. Thus, whether an interesting more restricted version of Theorem 11 for maximal models is true is still open.

28. Naturally enough, the nonextensionality of the system will emerge more strikingly when equality is added to the system. But all it will amount to are additional axioms which are perfectly natural when the metalinguistic interpretation for substitutional quantification is kept in mind. Of course, my approach is *not* conservative with respect to the predicate calculus with equality, except in the sense that all instances of the axioms of equality, in which ostensives or semantic predicates are not involved, hold.

29. To attempt this, of course, one would have to describe the conditions of satisfaction for sentences of an *O*-language in the *O*-language itself and connect these conditions axiomatically to the truth predicate. Since I am explicitly working in a more general setting where it is not presupposed that the language itself has enough syntactic resources to describe satisfaction, this move is ruled out. But as I point out in a slightly different context, one can decide to study that subclass of honest premodels which *do* have enough resources to describe satisfaction: *O*-model theory is a general framework for all such studies. I cannot pursue this topic further now.

One can also modify the satisfaction clauses so that presence in the truth or falsity predicate is necessary for satisfaction, e.g., (c) would read: "$(A \vee B)$, where A and B are wffs, is satisfied if $(A \vee B)$ is in the extension of T and either A or B is satisfied". Doing so would not change much except to rule out the possibility of anything but honest premodels satisfying the axioms. My approach takes up less ink.

30. Of course I am only giving a proof sketch. Let me add that a common (naive) reaction to the Liar is to claim it is a contradiction. Perhaps this can be taken to be intuitive recognition of the inconsistency of (a/Fa).

31. Provided, of course, there are no other pathological sentences which, due to violations of local determination, force the ostensive to be false.

32. [5], p. 210. I use a version due to Barwise and Etchemendy, [2], pp. 23–24.

33. I am avoiding representing these by means of quantification simply because it makes illustrating the point much harder. But the system treats the quantified version of the puzzle the same way.

34. This restriction is for illustrative purposes. The semi-inductive construction is easily modified to apply to languages containing an ostensive operator.

35. Of course we could also modify the satisfaction clauses used in the Gupta/Herzberger construction so that the truth predicate is substitutional there as well.

36. I briefly considered broadening the model theory so that arbitrary subsets of the language could be contained in the domain. Naturally, this simplifies the axiom system: i.e., (A8) is false in this context. Also, it makes embedding the models of the Gupta/Herzberger construction a little more direct. But these really are notational gains. On the other hand, (A8) does not pick out (in this context) only the *O*-models which contain at least all the sentences of the *O*-language they are models of—which are the only models we are interested in. This is for the same reason that we have Corollary 23 above: no *O*-language can describe the relation between constants and the items they are mapped to.

    Herzberger (in conversation) suggested I eliminate the falsity predicate in my approach to facilitate comparison of the two approaches. I have resisted this for the (perhaps shallow) reason that in *O*-models the distinction between sentence and nonsentence would be lost without additional apparatus.

37. Of course it has also been suggested that the distinction is artificial at best. See Barwise and Feferman's [3], Chapter 1.1.

38. The desire for a language to be able to describe its own syntax completely is not such a strange desire, but it really introduces irrelevancies when the logic of "truth" is at issue.

    Notice that Tarski gave the metalanguage the capacity to exactly specify the syntax of the object language and his doing so turns on his desire to *define* the truth predicate. Why is a *definition* so important as opposed to an axiomatization of it as a primitive? One possible reason I have already mentioned in Section 2. Otherwise, Tarski's answer ([20], pp. 405–406, [19], pp. 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. Our approach here may be taken as an attempt to overcome these difficulties.

39. Notice I am not making the grander claim that the semantics of the truth predicate *as a whole* is a natural extension of the first-order predicate calculus. What is being claimed here is that the *reasoning* involving truth is first order.

40. "Definable" here is relative to the resources of the language to describe its own syntax of course.

41. For example, the primary *O*-models corresponding to the best candidates on the Gupta/Herzberger approach are all *extensional*, that is, Ta has the same truth value as Tb if a and b are mapped to the same item in the domain. But there is reason to believe that "intensional" approaches have advantages. See the last paragraph of this paper.

Another approach to the problem of semantic ascent more in spirit with the added power of *O*-model theory can be described briefly: consider that restriction of *O*-model theory to those models where each n-place predicate P has associated with it an n-place predicate P^*, where P^* holds of every n-tuple n^* that P holds of provided that every sentence contained in n^* is referred to transparently by some constant. Let $(P \Leftrightarrow P^*)$ stand for the sentence that says that P and P^* hold of exactly the same objects. What it amounts to is the claim that every sentence in the extension of P (and P^*) is transparently referred to. If S is a sentence containing only the predicates P and P^* then the sentence $(P \Leftrightarrow P^*) \Rightarrow (\exists x)(x/S)$ asserts that S is referred to transparently if every sentence in the extension of P (and P^*) is referred to transparently. Now consider the class of models in which all sentences of the above sort (suitably generalized) are satisfied. This class of models satisfies the constraint that sentences about classes of sentences transparently referred to are themselves transparently referred to.

42. To be explicit, one way of measuring the effect a language's capacity to describe its own syntax has on its theory of truth is by considering what effects the mappings of the constants and predicate variables onto the domain has on the class of ostensives *forced* to be unsatisfied. So, for example, consider the set of axioms $(\exists x)(x/S)$, for all sentences S. It is obvious that not every mapping of the constants to the domain, and the n-place predicate variables onto the Cartesian nth product of the domain can correspond to an *O*-model in which these axioms hold. E.g., in any model in which the predicate P is mapped to the unit set of the sentence $(\exists x)(Px \ \& \ \mathrm{F}x)$, the above set of axioms is violated. The class of models in which these axioms hold characterize the amount of syntax compatible with a globally applicable theory of truth. Notice that these axioms do not rule out all cases of vicious reference. What they rule out are cases of vicious reference where the sentences in question cannot be transparently referred to by other constants, a somewhat more delicate matter. To study models ruling out *all* example of vicious reference, one must turn to systems with equality. I cannot go further into the matter now.

43. There are two complications in verifying this hunch. First, it is not at all easy to define *local determination*. Second, the notion of *unproblematic* sentence is hopelessly intuitive. At this point there is no intertheoretic notion; we have various notions: stable, grounded (in various senses), etc. This makes it hard to evaluate the significance of such a result.

REFERENCES

[1] Azzouni, J., "Realism and the autonomy of the special sciences," unpublished.

[2] Barwise, J. and J. Etchemendy, *The Liar*, Oxford University Press, New York, 1987.

[3] Barwise, J. and S. Feferman, eds., *Model Theoretic Logics*, Springer Verlag, New York, 1985.

[4] Field, H., "Tarski's theory of truth," *Journal of Philosophy*, vol. 69 (1972), pp. 347–375.

[5] Gupta, A., "Truth and Paradox," pp. 175–236 in *Recent Essays on Truth and the Liar Paradox*, edited by R. Martin, Oxford University Press, New York, 1984.

[6] Gupta, A. and R. Martin, "A fixed point theorem for the Weak Kleene Valuation Scheme," *Journal of Philosophical Logic*, vol. 13 (May 1984), pp. 131–135.

[7] Herzberger, H., "Naive semantics and the Liar Paradox," *The Journal of Philosophy*, vol. 79 (1982), pp. 479–497.

[8] Herzberger, H., "New paradoxes for old," *Proceedings of the Aristotelian Society* N.S., vol. 81 (1980–1981), pp. 109–123.

[9] Herzberger, H., "Notes on naive semantics," pp. 133–174 in *Recent Essays on Truth and the Liar Paradox*, edited by R. Martin, Oxford University Press, New York, 1984.

[10] Kaplin, D., "Opacity," pp. 229–289 in *The Philosophy of W. V. Quine*, Open Court, 1986.

[11] Kremer, M., "Kripke and the logic of truth," *The Journal of Philosophical Logic*, vol. 17 (1988), pp. 225–278.

[12] Kripke, S., "Outline of a theory of truth," pp. 53–81 in *Recent Essays on Truth and the Liar Paradox*, edited by R. Martin, Oxford University Press, New York, 1984.

[13] Martin, R., ed., *Recent Essays on Truth and the Liar Paradox*, Oxford University Press, New York, 1984.

[14] Parsons, C., "The Liar," pp. 221–267, in *Mathematics in Philosophy*, Cornell University Press, 1983.

[15] Quine, W., "What price bivalence?," pp. 31–37 in *Theories and Things*, Harvard University Press, Boston, 1981.

[16] Salmon, N., *Frege's Puzzle*, Bradford Books, MIT Press, Cambridge, 1986.

[17] Schiffer, S., *Remnants of Meaning*, Bradford Books, MIT Press, Cambridge, Massachusetts, 1987.

[18] Skyrms, B., "Intensional aspects of semantical self-reference," pp. 119–131 in *Recent Essays on Truth and the Liar Paradox*, edited by R. Martin, Oxford University Press, New York, 1984.

[19] Tarski, A., "The concept of truth in formalized languages," pp. 152–177 in *Logic, Semantics, Metamathematics*, 2nd edition, edited by J. Corcoran, Hackett, Indianapolis, 1983.

[20] Tarski, A., "The establishment of scientific semantics," pp. 401–408, in *Logic, Semantics, Metamathematics*, 2nd edition, edited by J. Corcoran, Hackett, Indianapolis, 1983.

[21] Yablo, S., "Grounding, dependence, and paradox," *Journal of Philosophical Logic*, vol. 11 (February 1982), pp. 117–137.

Tufts University
Department of Philosophy
Medford, Massachusetts 02155

APPENDIX The proof of the completeness of Axiom System R1, R2, A1–A8 with respect to *O*-model theory.

A1 Preliminaries An *O*-language Λ has the standard vocabulary: (,), −, ∨, ∃, &, ⇒, ⇔ (where the latter three symbols are defined notation), *m*-place predicate variables, $P_1, \dots, P_n, \dots$ for each *m*, individual constants, $a_1, \dots, a_n, \dots$, individual variables, $x_1, \dots, x_n, \dots$, two one-place predicates T, F, and the symbol /. We will speak of the predicate constants and predicate variables, in general, as predicates, and the individual constants and individual variables, in general, as individuals. Italicized versions of the individuals and predicate variables are generally metalinguistic variables for the syntactic category of their unitalicized counterparts. Where so used, implicit quasiquote conventions are in force to prevent use/mention errors. To avoid subscripts and superscripts other letters are occasionally used. Higher-order objects such as sets of sentences will be represented by terms such as "Θ", "Φ", "Λ", etc.

To the standard inductive clauses defining a well-formed formula (wff), we add:

Definition AP1 If *A* is a wff and *v* is an individual then (v/A) is a wff.

Given individual variables $v_1, \dots, v_n$ and a formula *A*, whenever we use the notation $A(v_1, \dots, v_n)$, we are implicitly assuming that the only variables that may have, and do have, free occurrences in *A* are the ones displayed: $v_1, \dots, v_n$. The formula obtained by free-substituting $w_1, \dots w_n$ for $v_1, \dots, v_n$ in the same order, and where the w_i may be individuals, is denoted by $A[v_1 : w_1, \dots, v_n : w_n]$, or sometimes by $A[v_i : w_i]$, when unambiguous. In a case where all the w_i are individual constants, we call $A[v_1 : w_1, \dots, v_n : w_n]$ an *instance* of $A(v_1, \dots, v_n)$.

Definition AP2 We understand *Axiom System* AP2 to be the Axiom System R1, R2, A1–A8. We understand *Axiom System* AP2* to be Axiom System AP2 without Axiom A5.

Definition AP3 Two wffs $B(v_1, \dots, v_n)$ and *A* are the same *predicate-type* if $u_1, \dots, u_n$, $v_1, \dots v_n$ are individuals, *A* is $B[v_i : u_i]$, and further, no instance of u_i substituted for v_i is bound in *B*.

Definitions AP4 Let Ξ be a set of axioms. *Ξ-derivation* and *Ξ-theorem* are defined as one expects. A sentence *A* is *Ξ-consistent* if $-A$ is not a theorem of Ξ.

Definition AP5 If a sentence *W* is of the form

$$(\exists v)A \Rightarrow B[v : a]$$

where *a* is a constant, and $(\exists v)A$ is a sentence in which *a* does not occur, then we call *W* an *∃-formula*. We say that *W* is an *∃-formula* with respect to *a*.

Definition AP6 An *∃-form* is a maximal collection of ∃-formulas which have the same antecedent modulo the bound variables of the antecedent.

Definition AP7 If a sentence *W* is of the form

$$-(\exists v_1) \dots (\exists v_n)(w/A)$$

where $A(v_1, \ldots, v_n)$ is a wff without individual constants, and w is an individual distinct from the v_i, then we call W an *O-formula* (also *O-sentence*, where w is an individual constant, or *O-wff* otherwise). We say that W is an O-formula with respect to w and A.

Definition AP8 An *O-form* is a maximal collection of O-formulas which differ at most in their variables free in A.

Definition AP9 Let a be a constant. If a sentence W containing a is of the form

$$-(\exists v_1) \ldots (\exists v_n)(w/A)$$

where A is a wff containing either no constants or the constant a, and w is either a or one of the variables v_i, then we call W an *a-formula*.

Definition AP10 An *A-form* is a maximal collection of a-formulas which differ at most in their variables free in (w/A).

Definition AP11 Let Ξ be a set of axioms and Λ an O-language. A set of sentences of Λ is *maximally Ξ-consistent* if it is Ξ-consistent and is not contained properly in an Ξ-consistent set of sentences of Λ.

Two reminders about the axiom (A5) are necessary now. Recall that (A5) is meant to convey the fact that a constant cannot refer to more than one sentence in a model. The alphabetic variant restriction (take note: the nomenclature is somewhat nonstandard here) on the substitution instances of the axiom prevents sentences like "$(x)(y)(z) - ((z/Px) \& (z/Py))$" from being instances of (A5), since such instances are not valid.

But this restriction complicates the completeness proof forthcoming since constants consequently do not behave in derivations quite as they do in the standard predicate calculus. In particular, from "(a/Pb)" we can derive "$-(a/Pc)$" using (A5) but we *cannot* derive "$(x) - (a/Px)$" from "(a/Pb)". We must work around this restriction in the completeness proof and we do so by discharging as assumptions all uses of (A5) in any derivation. Thus the special version of the lemma immediately below and the complicated form axioms (A7)–(A8) take.

Lemma AP12 *If Ω is a set of sentences in which a constant a does not occur, then if there is a derivation Ψ from Ω of a sentence W, where any instance of* (A5) *containing a in Ψ does not contain two ostensive formulas of the same predicate type in which a occurs in one formula while a different constant occurs in the other in place of a, then there is a variable v and a derivation from Ω of the sentence* $(v)W[a:v]$.

Lemma AP13 (Deduction Theorem) *Suppose Ω is a set of sentences and A and B are sentences. Then there is a derivation of B from A and Ω iff there is a derivation of* $(A \Rightarrow B)$ *from Ω.*

Overview We are working our way toward Theorem AP29 below. We essentially use a Henkin proof. We give two procedures for constructing maximally consistent sets out of O-consistent sets and show our results by judicious use of these procedures.

The first procedure (a Henkin 1procedure) can be carried out on a set of sentences if they do not require that every item in the domain have a constant mapped to it. (For example, "$(x)(\mathrm{T}x \vee \mathrm{F}x)$" holds *only* in models in which every object in the model has a constant mapped to it, and in fact, only in models containing only sentences.) Several complications arise in carrying out the proof. First, the standard move of supplying $\exists$-sentences with fresh constants to witness existential claims cannot be simply applied: in general it is false that $(\exists x)A(x) \Rightarrow A(a)$ is consistent with a set of sentences not containing a. So we must supply a messy case-by-case analysis. Secondly, the domain cannot merely contain *constants* as it can in the standard predicate calculus; to satisfy ostensives it may have to contain *sentences*. In a 1procedure, in general, it can be required for many of these sentences not to have constants mapped to them; but then using constants to witness them with $\exists$-sentences creates problems since if x is mapped to a sentence which no constants are mapped to, neither $\mathrm{T}x$ nor $\mathrm{F}x$ is satisfied, but *otherwise* one of them is. We circumvent this by utilizing nonstandard 'nice' satisfaction clauses for these special constants and then deleting the constants to construct the model we want.

Definition AP14 A *partner function* is a partial 1-1 mapping from the set of sentences in a language Λ into the set of individual constants of that language. We call the individual constant a sentence is mapped to *the partner* of it.

Definition AP15 Suppose Ω is an AP2-consistent set of sentences. Then the following is a Henkin 1procedure.

Let Λ^* be a language at least as articulate as Λ but containing in addition a countable number of new individual constants. Divide the new constants into a countable collection of countable sets of constants: $\xi_1, \ldots, \xi_n, \ldots$, and consider the collection of sublanguages of $\Lambda^*: \Lambda_0, \ldots, \Lambda_n, \ldots$, where Λ_0 is Λ, and the set of symbols of each Λ_i contains ξ_i plus the union of the sets of symbols of Λ_j for $j < i$.

We now construct a sequence of sets of sentences χ_i of the language Λ_i as follows: (Note: Ω is χ_0)

Step n (where n is odd): Let $A_1, \ldots, A_n, \ldots$, be all sentences of Λ_n which do not appear in the second place of an ostensive sentence in χ. For the sake of notational simplicity, call χ_n, $\chi_n + [a_0]$. We define $\chi_n + [a_i]$ inductively. Give each A_i a partner a_i, of ξ_{n+1}, and add $-\mathrm{T}a_i$, $-\mathrm{F}a_i$, and one member from every a_i-form to $\chi_n + [a_{i-1}]$. Call the resulting set $\chi_n + [a_i]$. Call the union of $\chi_n + [a_i]$ for all i, χ_{n+1}.

Step n (where n is even): First take χ_n and maximize it (as is done standardly in Henkin completeness proofs) with respect to the language Λ_n. Call the result χ_n^*. Then, let σ_{n+1} be some ordering of all the $\exists$-forms of Λ_{n+1}. Let τ_i be the ith $\exists$-form in σ_{n+1}. We define χ_{ni}^* for i inductively:

(1) χ_{no}^* is χ_n^*.
(2) Suppose the antecedent $(\exists v)B_i$ of τ_i is not in χ_n^*. Then χ_{ni}^* is $\chi_{n(i-1)}^*$.
(3) Suppose a member of τ_i, $(\exists v)B_i \Rightarrow B_i[v:a]$, with respect to a constant a of Λ_n is in χ_n^*. Then χ_{ni}^* is $\chi_{n(i-1)}^*$.
(4) Suppose the antecedent $(\exists v)B_i$ of τ_i is in χ_n^* although for no constant a is $B_i[v:a]$ in χ_n^*. Suppose further that for some constant b_j, where $j <$

i and $B_j[v:b]$ is in χ^*_{ni}, $-B_i[v:b_j]$ is not consistent with $\chi^*_{n(i-1)}$. Then χ^*_{ni} is $\chi^*_{n(i-1)}$ plus $B_i[v:b_j]$.

(5) Suppose the antecedent $(\exists v)B_i$ of τ_i is in χ^*_n but no member of τ_i is in Ω^*_n. Suppose further that for every $j < i$, $-B_i[v:b_j\}$ is consistent with $\chi^*_{n(i-1)}$. Then choose a constant b_i of ξ_{n+1} which has not been used before and let χ^*_{ni} be $\chi^*_{n(i-1)}$ plus $-B_i[v:b_j]$ for every $j < i$, and $B_i[v:b_i]$.

Let χ^{**}_n be the union of χ^*_{ni} for all i, and Henkin maximize it with respect to Λ_{n+1} as usual. Call the result χ_{n+1}.

Let χ be the union of the above sets, and let ϕ be the partner function defined above. We call the pair $\langle\chi, \phi\rangle$ the *result* of a Henkin 1procedure. Sometimes we will speak loosely of χ alone being the result of a Henkin 1procedure.

Theorem AP16 *Let $\langle\chi, \phi\rangle$ be the result of a Henkin 1procedure. Let u and v be individual variables. Let $A_1, \ldots, A_n$ be a set of existential generalizations of ostensive formulas in which u, and u only, occurs free. Suppose also that no constants appear in any of these ostensive-formulas. Let $B_1, \ldots, B_m$ be a set of wffs with at least u free, and let $\Sigma_1, \ldots, \Sigma_m$ be existential prefixes, where for each Σ_i, v and all the free variables of B_i, except u, occur in it. Let $b_1, \ldots, b_m$ be individual constants. Then if no sentence of the form* (*):

$$(u)(\mathrm{T}u \vee \mathrm{F}u \vee A_1 \vee \ldots \vee A_n \vee \Sigma_1((v/B_1) \ \& \ (v/B_1[u:b_1])) \vee \ldots \vee \Sigma_m((v/B_m) \ \& \ (v/B_m[u:b]))$$

follows from $\chi_n + [a_m]$ for all m and n, χ is a maximally AP2-*consistent set of sentences of* Λ^*.

Proof: We first show that every χ_i is consistent by induction. Certainly χ_0 is.

For n even: suppose χ_{n+1} is not consistent. Let b_i be the earliest constant with respect to which applying the procedure described in Step n results in an inconsistent set. Clearly we need only concern ourselves with Case 5, since $\chi^*_{n(i-1)}$ is O-consistent. Call the union of $\chi^*_{n(i-1)}$ with $-B_i[u:b_j]$ for all $j < i$, $\chi^{\sim}$. Notice that $\chi^{\sim}$ is AP2-consistent by assumption. It is easy to show in the standard way, using Lemma AP12, that χ_{n+1} is AP2*-consistent. Since it is not AP2-consistent, there is a sequence of instances of Axiom (A5): $C_1, \ldots, C_m$ so that $B_i[u:b_i] \Rightarrow -(C_1 \ \& \ldots \& \ C_m)$ is AP2*-derivable from $\chi^{\sim}$. Now notice that if it is the case that none of the conjuncts contain two ostensives of the same predicate type with the constant b_i occurring in the second place of one ostensive while some distinct constant c occurs in the same place of the other ostensive (and notice that the notion "same place" is well defined here since the two ostensives must be of the same predicate type), then use Lemma AP12 (and the alphabetic variance condition on (A5)), and note that $(u)(B_i \Rightarrow -(C_1 \ \& \ldots \& \ C_m)[b_i:u]$ follows from χ_n. But every conjunct of the consequent is still an instance of (A5) and this violates the fact that $(\exists u)B_i$ is in χ_n and $\chi^{\sim}$ is AP2-consistent.

Next, notice that for sentences $D_1, \ldots, D_p$, $B_i[u:b_i] \Rightarrow (D_1 \vee \ldots \vee D_p)$ iff $(B_i[u:b_i] \Rightarrow D_1) \vee \ldots \vee (B_i[u:b_i] \Rightarrow D_p)$. So consider one disjunct $-C_k$. Let $w_1, \ldots, w_o$ be individual variables among which is w, and let E be some wff in which u is free, so that $-C_k$ without loss of generality is of the form $-(w_1)\ldots(w_o) - (w/E[u:b_i]) \ \& \ (w/E[u:c]))$. Since, by assumption, $(\exists u)B_i$ and

$-B[u:c]$ are in $\chi^{\sim}$, if $(u)(B_i \Rightarrow (\exists w_1)\ldots.(\exists w_o)(w/E)\ \&\ (w/E[u:c]))$ is AP2*-consistent with $\chi^{\sim}$, we have, by the standard predicate calculus and (A7), $\chi^{\sim}$ AP2*-inconsistent contrary to assumption.

For n odd: Suppose that χ_{n+1} is not consistent. Let a_i be the earliest constant with respect to which the addition of a_i-sentences, $-\mathrm{T}a_i$, and $-\mathrm{F}a_i$ yields an inconsistent set. Then from $\chi_n + [a_{i-1}]$, we may derive, using Lemmas AP12 and AP13, a sentence of the form (*).

Since each χ_i is consistent, it is easy to see that χ is a maximally AP2-consistent set of sentences.

Lemma AP17 (O-maximality lemma) *Any maximal* AP2-*consistent set* χ *of sentences has the following properties*:

(1) *For any sentence* A, χ *does not contain both* A *and* $-A$.
(2) *For any sentence* A, χ *contains either* A *or* $-A$.
(3) *For any sentences* A *and* B, *if* A *and* $(A \Rightarrow b)$ *are in* χ, *then* B *is in* χ.

Definition AP18 Let $\langle\chi, \phi\rangle$ be the result of a Henkin 1procedure. Suppose the domain Δ of a model Θ of the standard predicate calculus is the union of all the sentences of Λ^* which do not contain any individual constants which are partners, and all constants of Λ^* which are defined by Clause 4 immediately below. Suppose the mappings of the individual constants, predicate variables, and predicate constants of Λ^* are defined by the following clauses:

1. Any constant a is mapped to a sentence A if (a/A) is in χ.
2. Any constant a which does not appear in the first place of an ostensive-sentence in χ, and for which either the sentence Ta or Fa is in χ, is mapped to Fa.
3. Order the constants which are partners. Any constant which is a partner is mapped to the sentence of which it is a partner
   i. provided none of the clauses above have procured the sentence a referring constant already,
   ii. provided the sentence it is a partner of is in Δ, and
   iii. provided no constant earlier in the ordering is mapped to the same sentence it is.
4. Any remaining constants are mapped to themselves.
5. Each n-place predicate variable P is mapped to that set of n-tuples $(\rho_1, \ldots, \rho_n)$ such that $Pa_1 \ldots a_n$ is in χ and a_i is mapped to ρ_i by the above mapping for constants.
6. The predicate T is mapped to the set of sentences contained in the intersection of χ and the domain Δ, and F is mapped to the remaining sentences of Δ.

Then we say that Θ is a *standard model* for $\langle\chi, \phi\rangle$. We extend the notion of *interpretation* to standard models in the natural way.

Definition AP19 Let Θ be a standard model for the result $\langle\chi, \varphi\rangle$ of a Henkin 1procedure. Consider the language $\Lambda^{\sim}$ gotten from Λ^* by deleting the partners from the alphabet of Λ^*, and the premodel Θ' of $\Lambda^{\sim}$ gotten from Θ by deleting the mappings for the partners, taking the extension of T in Θ' to be the extension of T in Θ intersected with $\Lambda^{\sim}$, and similarly for F. We call Θ' a *nice* premodel for χ.

Definition AP20 Let W be a wff and ι an interpretation on a standard model Θ. We say that ι *satisfies* W *nicely* with respect to Θ iff "satisfies nicely" obeys Clauses (a)–(c), (d2), (e2)–(g) of Definition 5, and

(dl′) W is (a/A), where a is a constant which is not a partner, A is a wff, there is an ι-instance A' of A, A' is a sentence assigned by Θ to a, and either Ta and A' are both in T's extension or both in F's extension.

(el′) W is Ta, a is a constant, there is a sentence A in T's extension, and (a/A) is in T's extension; or A is a sentence assigned by Θ to a, (a/A) is not in T's extension, and either: (i) Ta is in T's extension, or (ii) a is not a partner and both Ta and Fa are not in T's extension.

Lemma AP21 *Let* $\langle \chi, \phi \rangle$ *be the result of a Henkin* 1*procedure, let* Θ *be a standard model of* $\langle \chi, \phi \rangle$ *and* Θ' *the nice premodel gotten from* Θ. *Let W be a wff of* Λ^* *and let* $a_1, \ldots, a_n$ *be all the individual constants of W which are in* Λ^* *but not in* $\Lambda\sim$. *Let* $v_1, \ldots, v_n$ *be individual variables new to W. Let* ι *be any interpretation on* Θ. *Let* ι^* *be that interpretation on* Θ' *such that* ι^* *agrees with* ι *on all variables except perhaps for* $v_1, \ldots v_n$ *and for any variable* v, $\lambda_\iota(v)$ *is* $\lambda_{\iota^*}(v)$ *wherever* $\lambda_{\iota^*}(v)$ *is defined.* ι^* *maps each* v_i *to whatever* Θ *maps* a_i *to. Given the above conditions, W is satisfied nicely by* ι *if* $W[a_i : v_i]$ *is satisfied nicely by* ι^*.

Proof: This is shown by induction on the length of a wff.

Theorem AP22 *Let* $\langle \chi, \phi \rangle$ *of* Λ^* *be the result of a Henkin* 1*procedure,* Θ *a standard model of* $\langle \chi, \phi \rangle$, *and* Θ' *the nice premodel gotten from* Θ. *Then*
(1) *If W is a sentence of* Λ^*, *W is satisfied nicely by* Θ *iff it is in* χ.
(2) *If W is a sentence of* $\Lambda^{\sim}$, *W is satisfied nicely by* Θ' *iff it is satisfied by* Θ'.
(3) *If W is a sentence of* $\Lambda^{\sim}$, *W is satisfied nicely by* Θ *iff it is satisfied nicely by* Θ'.

Proof: These are shown by induction. Use Lemma AP21 for (3).

Theorem AP23 *Let* Θ' *be a nice premodel for the result of a Henkin* 1*procedure. Then* Θ' *is an honest premodel in which every sentence of* χ *is satisfied.*

Proof: This follows from Theorem AP22.

Corollary AP24 *Let* χ *be the result of a Henkin* 1*procedure which is maximally* AP2*-consistent. Then there is an O-model* Θ *which satisfies every sentence of* χ.

We now carry out a Henkin 2procedure for consistent sets of sentences which require that every item of the domain have a constant mapped to it.

Definition AP25 Suppose Ω is an AP2-consistent set of sentences. Then the following is a Henkin 2procedure.

Let Λ^*, ξ_i, Λ_i be as in a Henkin 1procedure. As before, we define a sequence of sets of sentences χ_i of the language Λ_i as follows (Note that Ω is χ_0):

Step n (where n is odd): Let $A_1, \ldots, A_n, \ldots$, be all sentences of Λ_n such that no A_i appears in the second place of an ostensive sentence in χ_n, and no A_i has a partner from an earlier stage. As before, we define $\chi_n + [a_i]$, for all i. Give each A_i a partner a_i, of ξ_{n+1} and add $\mathrm{T}a_i$ (or $\mathrm{F}a_i$, if the former is not consistent with $\chi_n + [a_{i-1}]$), and one member from every O-form with respect to a_i to $\chi_n + [a_{i-1}]$. Call the resulting set $\chi_n + [a_i]$. Call the union of $\chi_n + [a_i]$ for all i, χ_{n+1}.

Step n (where n is even): do as in the even steps of Definition AP15.

Let χ be the union of the above sets, and let ϕ be the partner function defined above. We call the pair $\langle \chi, \phi \rangle$ the *result* of a Henkin 2procedure. Again, we sometimes loosely refer to χ alone as being the result of a Henkin 2procedure.

Definition AP26 Let u and v be individual variables. Let $B_1, \ldots, B_n$ be wffs without constants in which u does not appear free, $A_1, \ldots, A_m$ wffs in which at least u occurs freely, $v_1, \ldots, v_m$ individual variables, $\Sigma_1, \ldots, \Sigma_n$ existential prefixes such that Σ_i contains all and only the free variables of A_i, and $\Sigma'_1, \ldots, \Sigma'_m$ existential prefixes such that Σ'_i contains v and all the free variables of A_i, except for u, and such that Σ'_i does not contain v_i. We call closures of wffs of the following sort *wffs of type* (1):

$$(\exists v_1) \ldots . (\exists v_m)(u)(-\mathrm{T}u \vee \mathrm{F}u \vee \Sigma_1(u/B_1) \vee \ldots \vee \Sigma_n(u/B_n) \vee \Sigma'_1((v/A_1) \;\&\; (v/A_1[u:v_1])) \vee \ldots \vee \Sigma'_m((v/A_m) \;\&\; (v/A_m[u:v_m])));$$

and we call closures of wffs of the following sort *wffs of type* (2):

$$(\exists v_1) \ldots (\exists v_m)(u)(-\mathrm{T}u \vee -\mathrm{F}u \vee \Sigma_1(u/B_1) \vee \ldots \vee \Sigma_n(u/B_n) \vee \Sigma'_1((v/A_1) \;\&\; (v/A_1[u:v_1])) \vee \ldots \vee \Sigma'_m((v/A_m) \;\&\; (v/A_m[u:v_m]))).$$

Theorem AP 27 *If $\langle \chi, \phi \rangle$ is the result of a Henkin 2procedure and either no sentences of type 1 are derivable from any $\chi_n + [a_m]$ or no sentences of type 2 are derivable from any $\chi_n + [a_m]$, then χ is a maximally AP2-consistent set of sentences.*

Proof: We first show that every χ_i is consistent by induction. Certainly χ_0 is.

For n even: this is the same as the analogous case for Henkin 1procedures.

For n odd: Suppose that χ_{n+1} is not consistent. Let a_i be the earliest constant such that the addition of O-formulas with respect to a_i, plus either $\mathrm{T}a_i$ or $\mathrm{F}a_i$, yields an inconsistent set. Then from $\chi_n + [a_{i-1}]$, we may derive, using Lemmas AP12 and AP13, a sentence of type 1 and a sentence of type 2.

Since each χ_i is consistent, it is easy to see that χ is a maximally AP2-consistent set of sentences.

Theorem AP28 *Let χ be the result of a Henkin 2procedure which is maximally AP2-consistent. Then there is an honest premodel Θ of Λ^* which satisfies every sentence of χ.*

Proof: We let the domain Δ be the union of all the sentences Λ^*, and all the individual constants of Λ^* which are defined by Clause 5 immediately below. We define the mappings of Θ as follows:

1. Any individual constant a is mapped to a sentence A if (a/A) is in χ.
2. Any individual constant a: (1) which does not appear in the first place of an ostensive-sentence in χ, (2) which is not a partner and (3) for which either the sentence $\mathrm{T}a$ or $\mathrm{F}a$ is in χ is mapped to $\mathrm{F}a$.

3. Consider the following sequence of sets of constants: Υ_1 is the set of partners of sentences to which constants are mapped by Clause 2. Υ_{n+1} is the set of partners of sentences of the form Fa where a is in Υ_n. Let Υ^* be the union of Υ_n for all n. Map each constant a of Υ^* to the sentence Fa.
4. Any constant which is a partner is mapped to the sentence of which it is a partner, provided none of the clauses above have already mapped it to something else.
5. All remaining constants are mapped to themselves.
6. Each n-place predicate variable P is mapped to that set of n-tuples $(\rho_1, \ldots, \rho_n)$ such that $Pa_1 \ldots a_n$ is in χ and a_i is mapped to ρ_i by the above mapping for constants.
7. The predicate T is mapped to the set of sentences of χ, and F is mapped to the remaining sentences of Λ^*.

It easy to see that the above clauses define a premodel. Easy inductions on the length of a wff show that Θ is honest and that a sentence is satisfied in Θ if it is in χ.

Theorem AP29 *If* Ω *is* AP2-*consistent, then there is an O-model* Θ *such that every sentence of* Ω *is satisfied in* Θ.

Proof: We start by applying the Henkin 1procedure to Ω. Either it yields us a maximally consistent set containing Ω, and we are done, or we are able to construct an AP2-consistent set of sentences containing Ω from which may be derived a sentence of type (*) as described in Theorem AP16. Now take this set and apply the Henkin 1procedure to it. If this procedure does not yield a maximally consistent set then at a certain point we have constructed an AP2-consistent set of sentences from which we are able to derive sentences of both types 1 and 2. But then it is easy to see, using (A4), that this violates (A8) and therefore the AP2-consistency of Ω.