KIT FINE AND TIMOTHY McCARTHY

TRUTH WITHOUT SATISFACTION*

In his famous paper [7], Tarski gave a definition of truth for a formalized language. Unable to perform a direct recursion on the concept itself, he gave a definition in terms of satisfaction. This makes it natural to ask if such an indirect procedure is necessary or whether a definition of truth can be given without using or somehow invoking the concept of satisfaction.

The question, as it stands, is vague; and later we shall be concerned to make it more precise. But even as it stands, it has an obvious technical interest. The situation that Tarski found himself in is common in mathematics. We wish to define a certain concept, but unable to perform a direct recursion on the concept itself we perform a recursion on a related concept of which the given concept is a special case. It would therefore be desirable to know when the related concept is necessary, both in the case of truth and in general.

The question may also have some philosophical interest. There is a fundamental difference between the concepts of truth and of satisfaction. The former merely applies to certain linguistic units; the latter connects language to an ontology of objects, typically extra-linguistic. A negative result on defining truth without satisfaction may perhaps constrain formal attempts to implement non-referential conceptions of truth. In the present paper, however, we will not be concerned in detail with the philosophical aspects of our question, although we will from time to time mention some points of contact between our discussion and the philosophical literature.

Interest in our question dates back to Wallace [9]; and the topic was subsequently taken up by Tharp [8] and Kripke [3] (especially Section 10). We have made our presentation self-contained, though the reader may consult the earlier work for general background and for elucidation of particular points.

The plan of our paper is as follows. Section 1 sets out the general framework in which our question and its cognates are posed. Section 2 solves the questions in case the meta-theory is not required to be finitely axiomatized;

Journal of Philosophical Logic **13** (1984) 397–421. 0022–3611/84/0134–0397$02.50

and Section 3 gives partial solutions in case finite axiomatizability is required, thereby answering a question of Kripke's [3] and of Tharp's [8]. Finally, Section 4 considers the question under other provisos on the meta-theory.

1. FRAMEWORK

Our question is: can truth be defined without invoking the notion of satisfaction, and, if so, under what circumstances? We shall here set up the general framework in which these and related questions can be made precise.

Following Tarski [7], we shall suppose that truth is defined for sentences of an object-language and that the definition itself is given in a meta-language. We take an *object language* to be a classical one-sorted first-order language L, with finitely many predicates, including identity, and with finitely many individual constants and function symbols. We take a *meta-language for* L to be a classical many-sorted first-order language L_M. It contains finitely many sorts, but at least two: the object-sort, with variables $x_1, x_2, \ldots$; and the number-sort, with variables $\alpha_1, \alpha_2, \ldots$. It is with the number variables $\alpha_1, \alpha_2, \ldots$ that we talk, via a Gödel numbering, about the expressions of L. L_M contains three kinds of descriptive constants: those of L, but restricted to the object-sort; the standard arithmetical constants, restricted to the number-sort; and finitely many additional "semantic" constants, without restriction of sort. It will be supposed that the identity-predicate only relates objects of the same sort, unless otherwise stated.

It is merely from convenience that we use a many-sorted rather than a single-sorted meta-language and that we use variables $\alpha_1, \alpha_2, \ldots$ for numbers in place of variables for expressions. Our results could be restated under the alternative conventions.

We take a *theory* T to consist of a language L_T and a set of sentences (the axioms) from that language. A theory T for an object-language L will be called an *object-theory*; and a theory M for a meta-language L_M will be called a *meta-theory*. It will always be supposed that a meta-theory contains the axioms of the system R of Robinson's arithmetic. This will secure a finite basis for a reasonable amount of elementary number theory.

The meta-theory M will be said to be *for* the object-language L if the meta-language L_M of M is for L; and M will be said to be *over* the

object-theory T for L if, in addition, any theorem of T is a theorem of M. We shall always use the symbols 'M', 'T', 'L_M' and 'L' in such a way that L_M is the language of M, L the language of T, and M is a meta-theory over T.

A standard meta-theory for L is the *Tarski theory M_0 of satisfaction*. This adds to the minimal meta-theory for L a third sort of variable, for infinite sequences of objects, a two-place function symbol taking sequences and numbers into objects, and a satisfaction-predicate relating sequences to numbers. The theory M_0 also contains, as axioms, the standard recursive clauses for satisfaction and some statements that specify elementary properties of sequences.

Having specified the meta- and the object-theories, let us say when a meta-theory defines truth. We shall discuss two notions of definability, one proof-theoretic, and other model-theoretic. In the proof-theoretic sense, we shall say that a meta-theory M for L *characterizes truth for* L, or is a *truth-theory for* L, if for some formula $T(\alpha)$ of L_M with one free variable α, each formula

$$(T) \qquad T(\overline{\ulcorner\phi\urcorner}) \equiv \phi,$$

is a theorem of M for any sentence ϕ of L. A statement of the form (T) will be called a *T-sentence*.

In the model-theoretic sense, we suppose given an intended interpretation or model $\mathfrak{A}$ for the object-language L. Any model $\mathfrak{B}$ for L_M contains, in an obvious sense, an object-language part $\mathfrak{B}_L$ and an arithmetical part $\mathfrak{B}_R$, obtained by restricting the language to L or to arithmetic, respectively. Call a model $\mathfrak{B}$ for L_M *arithmetically standard* if $\mathfrak{B}_R$ is a standard model for arithmetic, and say that $\mathfrak{B}$ *respects* the model $\mathfrak{A}$ *for* L if $\mathfrak{B}_L = \mathfrak{A}$. Then we may say that the meta-theory M *implicitly defines truth for* $\mathfrak{A}$ if (i) there is an arithmetically standard model for M that respects $\mathfrak{A}$ and (ii) there is a formula $T(\alpha)$ of one free variable α such that each T-sentence with $T(\alpha)$ as truth-predicate is true in any arithmetically standard model for M that respects $\mathfrak{A}$. We might say, in place of (ii), that for any arithmetically standard model $\mathscr{C}$ for M that respects $\mathfrak{A}$, $\mathscr{C} \models T(\overline{\ulcorner\phi\urcorner})$ iff $\mathfrak{A} \models \phi$ for all sentences ϕ of L. A weaker notion might be obtained by dropping the requirement in (i) and (ii) that the models for M be arithmetically standard; but this is not a case we shall consider.

These accounts of what it is to define truth are familiar from the literature. The account of implicit definability requires no motivation. However,

the account of characterizability is more problematic. If M characterizes truth, then, in any arithmetically standard model for M, the formula $T(\alpha)$ must receive the intended interpretation. However, it is easy to see that the converse does not hold. For example, let M be a finitely axiomatized meta-theory for L which does characterize truth for L, and consider any true arithmetical statement Φ which is independent of R. Suppose that M has an arithmetically standard model. Then if $\mathfrak{A}$ is any arithmetically standard model of M, the theory axiomatized by $R \cup \{\Phi \rightarrow (A_1 \wedge \ldots \wedge A_n)\}$, where $A_1, \ldots, A_n$ are the axioms of M, implicitly defines truth for $\mathfrak{A}$, but fails to characterize truth for L. However, we believe that characterizability may be of interest in its own right, independently of its connection to the notion of implicit definability. (See, for example, [7]: 187–188.)

Similar definitions may be given for satisfaction, though in this case we must take account of the fact that a different number of objects may be said to satisfy a formula. For the proof-theoretic sense, we say that, when $n > 0$, *M characterizes n-satisfaction for L*, or is an *n-satisfaction theory*, if there is a formula $S_n(x_1, \ldots, x_n, \alpha)$ of L_M containing $n + 1$ free variables $\mathbf{x} = x_1, \ldots, x_n, \alpha$ such that each "S-sentence"

$$(S) \qquad \forall \mathbf{x}(S(\mathbf{x}, \overline{\ulcorner \phi(\mathbf{x}) \urcorner}) \equiv \phi(\mathbf{x}))$$

is a theorem of M for any formula $\phi(\mathbf{x})$ of L with at most n free variables $x_1, \ldots, x_n$. For the model-theoretic sense, we say that *M implicitly defines n-satisfaction for $\mathfrak{A}$* if (i) there is an arithmetically standard model for M that respects $\mathfrak{A}$, and (ii) there is a formula $S_n(\mathbf{x}, \alpha)$ in the free variables $\mathbf{x}, \alpha$ for which each S-sentence is true in any such model. We say that *M characterizes (implicitly defines) satisfaction for L* (for $\mathfrak{A}$) if it characterizes (implicitly defines) n-satisfaction for L (for $\mathfrak{A}$) for each natural number n.

With these definitions, our general question can be made precise: when does there exist a meta-theory M that characterizes (implicitly defines) truth without characterizing (implicitly defining) n-satisfaction for a particular n? Our main interest is in characterizability; and in that case, we shall call a truth-theory *Tarskian$_n$* (*Tarskian*) if it characterizes n-satisfaction (n-satisfaction for any n).

In dealing with this question, we shall consider the effect of imposing two further conditions on the meta-theory M. The first is that it should contain the theorems of a given object-theory T. The second is that it should be finitely axiomatized.

The rationale for the first condition is two-fold. First, it is usual to define a truth-theory over a given object-theory, and so the condition just restricts us to the usual situation. But second, and more importantly, the condition may be treated, not as a restriction on a truth-theory, but as a way of getting at whether it commits us to a satisfaction theory. Our question, it may be argued, concerns our knowledge of (the axioms of) a satisfaction theory: of whether, given knowledge of a truth-theory for a language, one can thereby know a satisfaction theory for that language. Let us suppose that the axioms of the object-theory represent our non-semantical information about the world. Then we wish to ask, given a particular truth-theory, whether its axioms, in conjunction with our non-semantical information, lead to a satisfaction theory.

The requirement that the meta-theory be finitely axiomatized can also be motivated in different ways. One motivation is that it is a necessary (and perhaps a necessary and sufficient) condition for the intuitive requirement that the theory provide a recursive theory of truth. Another motivation sometimes claimed for the finiteness requirement is that finiteness is required in order to explain how a finite mind can come to know an infinity of T-sentences (see, for example [1]). Certainly, finiteness does serve this purpose; although it is not clear why, in the absence of additional constraints, a perspicuous recursive axiomatization should not serve this purpose as well as a finite axiomatization.

2. WITHOUT FINITENESS

Given an object theory T, we wish to know when there exists a non-Tarskian$_n$ truth-theory (without further restriction) over T. Although this case is comparatively trivial, its study will prove fruitful in consideration of the finite case.

Relative to the object-theory T, there is a *trivial meta-theory* M_T *for* T, obtained by introducing a new one-place predicate $T(\alpha)$ (applying only to numbers) and adjoining to T the axioms of Robinson's arithmetic and all T-sentences for sentences of L_T. It is clear that M_T characterizes truth for L_T; it is, in an obvious sense, the least committal of all truth-theories for T. Our question is now as to when the trivial truth theory for T is non-Tarskian$_n$.

Rather than answer this question directly, we shall find it helpful to

consider a more general question concerning "segregated" truth-theories. Call an n-place predicate P of a many-sorted language *uniform* if it only applies to n variables $v_1, \ldots, v_n$ to yield a formula $Pv_1 \ldots v_n$ when all of $v_1, \ldots, v_n$ are of the same sort, and call an n-place function symbol f *uniform*, $n \geqslant 0$, if it only applies to n variables $v_1, \ldots, v_n$ to form a term $fv_1 \ldots v_n$ when $v_1, \ldots, v_n$ and $fv_1 \ldots v_n$ are all of the same sort. We then say that a many-sorted language (or a theory based thereon) is *segregated* if all of its predicates (including identity) and all of its function symbols are uniform. The intuitive significance of the notion of uniformity for meta-theories is that a segregated meta-theory does not enable one to express, at the level of the primitives of the language, any connection between the world and language, i.e., between the objects as represented by the variables $x_1, x_2, \ldots$ and the expressions as represented by $\alpha_1, \alpha_2, \ldots$. It should be clear that the trivial meta-theory is segregated, since the non-logical constants of the object-language and of arithmetic are uniform and since the sole semantic primitive $T(\alpha)$ is also uniform.

Call a formula ϕ *single-sorted* if all of the terms of ϕ are of the same sort. Then a fundamental fact concerning segregated languages is the following:

LEMMA 1 (SEGREGATION). Any formula ϕ of a segregated language is logically equivalent to a truth-functional compound of single-sorted formulas.

Proof. By induction on the complexity of ϕ. The basis is given by the assumption that each predicate and function symbol is uniform. The truth-functional cases are trivial. Now suppose that ϕ is of the form $\exists x \psi$. By the induction hypothesis, ψ is equivalent to a truth-functional compound ψ' of single-sorted formulas. Put ψ' in disjunctive normal form and distribute $\exists x$ across the disjunction. Then ψ' is equivalent to a disjunction of formulas of the form $\exists x \bigwedge_{i=1}^{n} \chi_i$, with each χ_i single-sorted. Suppose that $\chi_1, \ldots, \chi_m$, $m \leqslant n$, are the χ_i's that contain free occurrences of x. Then $\exists x \bigwedge_{i=1}^{n} \chi_i$ is equivalent to $\exists x \bigwedge_{i=1}^{m} \chi_i \wedge \bigwedge_{j=m+1}^{n} \chi_j$. But it is clear that $\exists x \bigwedge_{i=1}^{m} \chi_i$ is single-sorted, and so the result is proved.

It should be noted that the logical equivalent produced by the proof contains the same free variables as ϕ.

The result may be strengthened in two directions. As it stands, the

identity predicate is required to be uniform. However, the result still holds with a universal identity-predicate (applying to objects of all sorts), as long as there are axioms that insure that the different domains are disjoint. This means then that the result can be transferred to a single-sorted theory under the standard translation.

Second, a modified conclusion can be reached without full segregation. Say two sorts are *directly connected* if they are connected, in the obvious sense, by a predicate or function symbol, and say two sorts are *connected* if they are linked by a sequence of direct connections. Call a formula *connected* if any two of its terms have connected sorts. Then it may be shown, in the same way as before, that each formula is equivalent to a truth-functional compound of connected formulas.

Let us say that the object theory T is finitely n-typed with respect to the formulas $\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \ldots, \phi_n(\mathbf{x})$ in the free variables $\mathbf{x} = x_1, \ldots, x_n$ iff, for any formula $\psi(\mathbf{x})$ of L with the free variables $\mathbf{x}$, the sentence $\bigvee_{i=1}^{n} \forall \mathbf{x}[\psi(\mathbf{x}) \equiv \phi_i(\mathbf{x})]$ is a theorem of T. It should be clear that the condition of being finitely n-typed is equivalent to the standard model-theoretic condition of admitting only finitely many n-types.

Let $M|L$ be the theory with language L and with the closed theorems of M in the language L as axioms. Then a necessary condition for a segregated meta-theory to characterize n-satisfaction is:

LEMMA 2. Let M be a segregated meta-theory. Then M characterizes n-satisfaction, $n > 0$, only if $M|L$ is finitely n-typed.

Proof. Suppose that M is segregated and characterizes n-satisfaction by means of the formula $S_n(\mathbf{x}, \alpha)$. Since M is segregated, there is a truth-functional form Φ such that:

$$(1) \qquad \vdash_M \forall \mathbf{x} \forall \alpha [S_n(\mathbf{x}, \alpha) \equiv \Phi(\psi_1(\mathbf{x}), \ldots, \psi_k(\mathbf{x}), \chi_1(\alpha), \ldots, \chi_l(\alpha))],$$

where the $\psi_i(\mathbf{x})$ are formulas with only object language variables and the $\psi_j(\alpha)$ are formulas with only variables of other sorts. Let T be the formula $\forall x_1(x_1 = x_1)$ of L and let $\perp$ be its negation. Now let $\phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x})$ be the result of substituting all possible distributions of T and $\perp$ for the $\chi_j(\alpha)$'s in $\Phi(\psi_1(\mathbf{x}), \ldots, \psi_k(\mathbf{x}), \chi_1(\alpha), \ldots, \chi_l(\alpha))$ (2^l cases in all). Then we show that M is finitely typed with respect to the formulas $\phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x})$. For

choose any formula $\phi(\mathbf{x})$ of L with Gödel number g. Since M characterizes n-satisfaction,

(2) $\quad \vdash_M \forall \mathbf{x}[S_n(\mathbf{x}, \bar{g}) \equiv \phi(\mathbf{x})].$

So from (1) and (2),

(3) $\quad \vdash_M \forall \mathbf{x}(\phi(\mathbf{x}) \equiv \Phi(\psi_1(\mathbf{x}), \ldots, \psi_k(\mathbf{x}), \chi_1(\bar{g}), \ldots, \chi_l(\bar{g}))).$

But the $\mathbf{x}$ do not occur in the $\psi_j(\bar{g})$. So from (3) by quantificational logic:

(4) $\quad \vdash_M \bigvee_{i=1}^{m} \forall \mathbf{x}(\phi(\mathbf{x}) \equiv \phi_i(\mathbf{x}));$

and we are done.

By examining the proof, we see that the hypothesis of the lemma can be weakened to the requirement that the object- and number-sorts be unconnected.

Under the assumption that M is a truth-theory, a converse to Lemma 2 can be established:

LEMMA 3. Let M be a truth-theory. Then M characterizes n-satisfaction if $M|L$ is finitely n-typed.

Proof. Let the truth-predicate for M be $T(\alpha)$, and suppose that $M|L$ is finitely typed with respect to the formulas $\phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x})$ of L, so that for any formula $\phi(\mathbf{x})$ of L with at most n free variables $\mathbf{x} = x_1, x_2, \ldots, x_n$,

(1) $\quad \vdash_M \bigvee_{i=1}^{m} \forall \mathbf{x}(\phi(\mathbf{x}) \equiv \phi_i(\mathbf{x})).$

Let $U(\alpha, \beta)$ be a term of arithmetic (or a conservative extension thereof) that represents in R a primitive recursive function $u(m, n)$ such that, for any formulas ϕ and ψ of L, $u(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$ is the Gödel number of the closure of $\phi \equiv \psi$. Let $g_1, \ldots, g_m$ be the Gödel numbers of $\phi_1(\mathbf{x}), \ldots, \phi_m(\mathbf{x})$, and let $S_n(\mathbf{x}, \alpha)$ be the formula $\bigvee_{i=1}^{m} [T(U(\alpha, \bar{g}_i)) \wedge \phi_i(\mathbf{x})]$. Choose any formula $\phi(\mathbf{x})$ of L with free variables from $\mathbf{x}$; suppose that its Gödel number is g. Then by definition of $S_n(\mathbf{x}, \alpha)$:

(2) $\quad \vdash_M S_n(\mathbf{x}, \bar{g}) \equiv \bigvee_{i=1}^{m} [T(U(\bar{g}, \bar{g}_i)) \wedge \phi_i(\mathbf{x})].$

From the derivability of the T-sentences and the fact that U represents the recursive function u, we have:

$$(3) \qquad \vdash_M S_n(\mathbf{x}, \bar{g}) \equiv \bigvee_{i=1}^{m} [\forall \mathbf{x}(\phi(\mathbf{x}) \equiv \phi_i(\mathbf{x})) \wedge \phi_i(\mathbf{x})].$$

But then from (1) and (3), it follows that:

$$(4) \qquad \vdash_M S_n(\mathbf{x}, \bar{g}) \equiv \phi(\mathbf{x}),$$

as required.

It may be of help to think of this result in the following way. Given disjunction over countable collections of formulas, satisfaction for a finitary language L may be defined in terms of truth. For let $\{\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \ldots\}$ be an enumeration of all the formulas of L in the free variables $\mathbf{x} = x_1, \ldots, x_n$. We may define n-satisfaction by the formula $\bigvee_{\xi < \omega} [\phi_\xi(\mathbf{x}) \wedge \alpha = \overline{\ulcorner \phi_\xi(\mathbf{x}) \urcorner}]$. Such an infinitary expression is not available in our meta-language L_M. But the hypothesis of the lemma shows how it may be replaced by a finitary disjunction of the same sort. (We are indebted to Alasdair Urquhart for this observation.)

Lemmas 2 and 3 may be put together to give a necessary and sufficient condition for a segregated truth-theory to characterize n-satisfaction.

THEOREM 4. Let M be a segregated truth-theory (or one in which the object and number sorts are unconnected). Then M characterizes n-satisfaction iff $M|L$ is finitely n-typed.

As an immediate consequence of Theorem 4, we obtain:

COROLLARY 5. Let M be a segregated truth-theory. Then M characterizes satisfaction iff $M|L$ is finitely n-typed for each n.

According to a result of Ryll-Nardzewski's [5] (see also [0], pp. 101–103), a theory is $\aleph_0$-categorical iff it is consistent, complete and admits only finitely many n-types for each n. Now it is readily shown that a theory admits only finitely many n-types iff all of its complete and consistent extensions admit only finitely many n-types. Therefore the condition in Corollary 5 may be replaced by the condition that all of the consistent and

complete extensions of $M|L$ are $\aleph_0$-categorical. It is interesting that the condition that arises naturally in the formulation of the corollary should connect up so well with a condition of quite independent interest in model theory.

Let us now apply these results to the trivial meta-theory M_T. It is clear that M_T is segregated and is a truth-theory. The other important feature of M_T in the present context is:

LEMMA 6. M_T is a conservative extension of the object-theory T, i.e., for any sentence ϕ of L, if $\vdash_{M_T} \phi$ then $\vdash_T \phi$.

Proof. Suppose not $\vdash_T \phi$. Then for some model $\mathfrak{A}$ of T, not $\mathfrak{A} \models \phi$. Extend $\mathfrak{A}$ to a structure $\mathfrak{C}$ for the language of M_T by adjoining the standard arithmetical part and the intended interpretation for $T(\alpha)$. Then $\mathfrak{C}$ is a model for M_T, but not $\mathfrak{C} \models \phi$. Therefore not $\vdash_{M_T} \phi$.

A simple syntactic proof may also be given.

This result means that for the case of $M = M_T$, $M|L$ in the condition of Corollary 5 can be replaced with T. That is:

COROLLARY 7. The trivial meta-theory M_T characterizes n-satisfaction iff T is finitely n-typed.

We can ow give a necessary and sufficient condition for there to exist a non-Tarskian$_n$ truth-theory over a given object-theory.

THEOREM 8. There is a non-Tarskian$_n$ truth-theory over T iff T is not finitely n-typed.

Proof. For the left-to-right direction, use Lemma 3; and for the right-to-left direction, let M be M_T and use Corollary 7.

It may be shown that, for each $n \geqslant 1$, we may obtain a truth-theory that characterizes n-satisfaction but not $(n + 1)$-satisfaction. In the light of Corollary 7, it suffices to produce a theory T^n that is finitely n-typed but not finitely $(n + 1)$-typed. But this follows from the result of [4], according to which there is, for each $n > 0$, a structure $\mathfrak{A}^n$ that realizes finitely many n-types yet infinitely many $(n + 1)$-types. For later purposes, we may note that the theory $T^n = \mathrm{Th}(\mathfrak{A}^n)$ may be taken to be decidable.

When we apply Theorem 8 to the case in which T is a logic (no non-logical axioms), we get conditions concerning truth-theories for a language. The critical question is: When is a logic for a given language finitely n-typed? Call a language *purely monadic* if its only non-logical constants are monadic predicates and individual constants. Then the answer to one direction of our question is given by:

LEMMA 9. For any $n > 0$, the logic T for a language L is finitely n-typed if L is purely monadic.

Proof. Straightforward. Use, for example, a normal form theorem for the monadic predicate calculus. Note, however, that it is essential to use our underlying assumption that the stock of non-logical primitives of L is finite.

For the other direction we shall establish, for later purposes, a stronger result than is required here:

LEMMA 10. If the language L is not purely monadic, then there is an axiomatized, consistent and complete theory phrased in L that is not finitely n-typed for any $n > 0$.

Proof. Let S be the theory of successor, i.e., of the structure $\langle \omega, f \rangle$ for f: $\omega \to \omega$ the successor function: L_S contains the single non-logical function symbol s. We observe that S can be axiomatized (cf. [0], pp. 159–160). Also, S is not finitely 1-typed and hence not finitely n-typed for any n. For let $\psi_0(x)$ be the formula $\forall y(x \neq \mathrm{s}y)$ and let $\psi_{n+1}(x)$ be the formula $\exists y(\psi_n(y) \wedge x = \mathrm{s}y)$. Then $\psi_n(x)$ is satisfied by n alone in $\langle \omega, f \rangle$ and so $\psi_m(x)$ and $\psi_n(x)$ are not provably co-extensive in S if $m \neq n$.

We now show how a theory T in L can be mutually interpreted in S. Suppose first that L contains an n-place function symbol, g, $n > 0$. We set up two translations τ and σ, from L_S to L and from L to L_S, respectively. For any formula ϕ of L_S, let $\tau(\phi)$ be the result of replacing each occurrence of $\mathrm{s}t$ in ϕ with $g\overbrace{t \ldots t}^{n}$; and for any formula ψ of L, let $\sigma(\phi)$ be the result of replacing each term $gt_1 \ldots t_n$, $n > 0$, with $\mathrm{s}t_1$, each individual constant a with the individual description $\iota x \forall y(x = \mathrm{s}y)$ (subsequently eliminated from the resulting formula in the usual way), and each atomic formula

$Pt_1 \ldots t_n$, for P a non-logical predicate of L, with Γ. Let T be the theory whose axioms are (i) the translations $\tau(\phi)$ of the axioms from a recursive basis for S, (ii) $\forall x_1 \ldots x_n P x_1 \ldots x_n$, for any n-place non-logical predicate P of L, (iii) $\forall x \forall x_2 \ldots x_n \forall y_2 \ldots y_m (hxx_2 \ldots x_n = gxy_2 \ldots y_m)$, for any m-place function symbol h, $m > 0$, and (iv) $\forall x_1 \ldots x_n (a \neq fx_1 \ldots x_n)$ for any individual constant a. Then it may be shown, either model- or proof-theoretically, that the theories S and T are mutually interpretable with respect to the translations σ and τ, i.e., that $\vdash_S \phi$ implies $\vdash_T \tau(\phi)$, $\vdash_T \psi$ implies $\vdash_S \sigma(\psi)$, and that $\vdash_S \phi \equiv \sigma(\tau(\phi))$, $\vdash_T \psi \equiv \tau(\sigma(\psi))$ for any formulas ϕ and ψ of L_S and L, respectively. From this, it readily follows that T is consistent, complete and not finitely 1-typed, given that the same is true of S.

In case L contains no n-ary function symbols, for $n > 0$, but only n-adic predicates, for $n > 1$, the proof may be modified by interpreting the function symbol g above as a predicate in the usual way.

From Lemmas 9 and 10, we obtain:

LEMMA 11. For any $n > 0$, the logic T for L is finitely n-typed iff L is purely monadic.

Theorem 8 with Lemma 11 now gives:

THEOREM 12. There is a non-Tarskian$_n$ (or non-Tarskian) truth-theory for the language L iff L is not purely monadic.

Although we have concentrated on the concept of characterizability, similar results can be proved for implicit definability. For any structure $\mathfrak{A}$, let Th($\mathfrak{A}$) be the theory with the language $L_{\mathfrak{A}}$ of $\mathfrak{A}$ and with the sentences ϕ in $L_{\mathfrak{A}}$ such that $\mathfrak{A} \models \phi$ as axioms. Then in analogy to Theorem 4 (and by an analogous proof) we have:

THEOREM 13. Let $\mathfrak{A}$ be a structure for the language L and let M be a segregated meta-theory for L that implicitly defines truth for $\mathfrak{A}$. Then M implicitly defines n-satisfaction for $\mathfrak{A}$ iff Th($\mathfrak{A}$) is finitely n-typed (i.e., iff $\mathfrak{A}$ realizes only finitely many n-types).

From Theorem 13 the analogues of Corollary 7 and Theorem 8 then readily follow:

THEOREM 14. Let T be the logic for the language L. Then the trivial meta-theory M_T implicitly defines n-satisfaction for the interpretation $\mathfrak{A}$ of L iff $\mathfrak{A}$ realizes finitely many n-types.

THEOREM 15. There is a meta-theory M that implicitly defines truth but not n-satisfaction for $\mathfrak{A}$ iff $\mathfrak{A}$ realizes infinitely many n-types.

We see then that there is no essential difference in the conditions concerning implicit definability for a structure $\mathfrak{A}$ and characterizability over the theory $\mathrm{Th}(\mathfrak{A})$.

3. WITH FINITENESS

We now consider the case in which the meta-theory is required to be finitely axiomatizable. In this case, the trivial meta-theory M_T will not do to prove our results. Indeed, it is easily shown that M_T, for a consistent theory T, has no finite axiomatization on the basis of T. Therefore another method must be used to obtain the non-Tarskian truth-theories. The underlying idea here is to produce a provability interpretation for the notion of truth.

First an auxiliary result:

LEMMA 16. Suppose that the object-theory T has a consistent, complete and axiomatizable extension. Then there is a *finitely* axiomatized theory T^* in an expanded language whose restriction $T^*|L$ to the language L of T is a consistent, complete (and, of course, axiomatizable) extension T^+ of T.

Proof. We distinguish two cases. (i) T has a finite model $\mathfrak{A}$. A single sentence ϕ then describes $\mathfrak{A}$ up to isomorphism. Let T^* be the theory in the language L whose sole axiom is ϕ. Then it is clear that T^* is a finitely axiomatized, consistent and complete extension of T. (ii) T has only infinite models. Choose an axiomatizable, consistent and complete extension T^+ of T. Then T^+ has only infinite models. Therefore by a result of Kleene's [2] (proved using Tarski's theory of truth!), T^+ can be finitely axiomatized by a theory T^* in an expanded language; and so we are done.

We now give a sufficient condition for a finitely axiomatizable meta-theory to characterize truth.

LEMMA 17. Suppose that the object-theory T has a consistent, complete and axiomatizable extension (i.e., T is not essentially undecidable). Then there is a consistent, finitely axiomatized and segregated truth-theory M over T.

Proof. By the supposition and Lemma 16, there is a finitely axiomatizable theory T^* in an expanded language whose restriction to L is a consistent and complete extension T^+ of T. Let G be the set of Gödel numbers of theorems of T^+, and let $T(\alpha)$ be a formula of the language of R that numeral-wise represents G. Then M has as axioms those of T^*, R and the consistency statement A for the provability predicate $T(\alpha)$. The result will be consistent as long as a new sort of variable is used to formulate A and the axioms of R.

Clearly M is segregated and finitely axiomatized. So it remains to show that M characterizes truth. Pick a sentence ϕ of L. By the completeness of T^+, either $\vdash_M \phi$ or $\vdash_M \sim\phi$. If $\vdash_M \phi$, then $\vdash_M T(\overline{\ulcorner\phi\urcorner})$ by representability for $T(\alpha)$, and so $\vdash_M T(\overline{\ulcorner\phi\urcorner}) \equiv \phi$. If $\vdash_M \sim\phi$, then $\vdash_M T(\overline{\ulcorner\sim\phi\urcorner})$, again by representability, $\vdash_M \sim T(\overline{\ulcorner\phi\urcorner})$, from the consistency axiom A, and so $\vdash_M T(\overline{\ulcorner\phi\urcorner}) \equiv \phi$. So, in either case, the T-sentence is derivable.

We may note that for this truth-theory the proof of the T-sentences is a kind of cheat, with the truth-value of either side of the equivalence established independently of the other.

A converse to Lemma 17 can also be proved.

LEMMA 18. Suppose that there is a consistent, finitely axiomatizable and segregated truth-theory for T. Then T has a consistent, complete and axiomatizable extension T^+.

Proof. Let M be the given meta-theory for T and $T(\alpha)$ the truth-predicate or formula for M. Then M may be obtained from R by the addition of a single "semantical" axiom ϕ. By the Segregation Lemma, ϕ and $T(\alpha)$ are equivalent to truth-functional compounds of single-sorted formulas $\Phi(\psi_1, \ldots, \psi_k, \chi_1, \ldots, \chi_l)$ and $\Phi'(\psi_{k+1}, \ldots, \psi_m, \chi_{l+1}(\alpha), \ldots, \chi_n(\alpha))$, respectively, where the ψ_i's contain object-language terms and the χ_j's do not. It is then clear that for some distribution $\pi_1, \ldots, \pi_m$, $\pi_{m+1}, \ldots, \pi_{m+l}$ of blanks or negation-signs, the result of adding the axioms $\pi_1\psi_1, \ldots, \pi_m\psi_m, \pi_{m+1}\chi_1, \ldots, \pi_{m+l}\chi_l$ is a consistent extension M^* of M.

Let T^* be the result of adding $\pi_1\psi_1, \ldots, \pi_m\psi_m$ to T (note that the ψ_i's may contain new non-logical constants); and let T^+ be the restriction of T^* to the language L. It is clear that T^+ is consistent, since M^* is. Also, T^+ is complete. For suppose not $\vdash_{T^+} \phi$ and not $\vdash_{T^+} \sim \phi$, for some sentence ϕ of L. Then not $\vdash_{T^*} \phi$ and not $\vdash_{T^*} \sim \phi$. So for some models $\mathfrak{A}$ and $\mathfrak{A}'$ of T^*, not $\mathfrak{A} \models \phi$ and not $\mathfrak{A}' \models \sim \phi$. Since M^* is consistent, there is a model $\mathfrak{C}$ for the axioms $\pi_{m+1}\chi_1, \ldots, \pi_{m+l}\chi_l$ and Robinson's arithmetic R. Let $\mathfrak{B}$ and $\mathfrak{B}'$ be the result of combining $\mathfrak{A}$ and $\mathfrak{A}'$, respectively, with $\mathfrak{C}$. Then $\mathfrak{B}$ and $\mathfrak{B}'$ are both models for M^*. Since $\mathfrak{B}$ and $\mathfrak{B}'$ share $\mathfrak{C}$ and satisfy $\pi_{k+1}\psi_{k+1}, \ldots, \pi_m\psi_m$, they both assign the same truth-values to $T(\overline{\ulcorner\phi\urcorner})$; yet they assign different truth values to ϕ. Therefore they assign differing truth-values to the T-sentence $T(\overline{\ulcorner\phi\urcorner}) \equiv \phi$, contrary to the fact that the extension M^* of M characterizes truth.

Finally, we may show that T^+ is axiomatizable. For:

(1) $\qquad \vdash_{M^*} T(\overline{\ulcorner\phi\urcorner})$ iff $\vdash_{M^*} \phi$,

since M^* characterizes truth. Also,

(2) $\qquad \vdash_{M^*} \phi$ iff $\vdash_{T^+} \phi$

by the completeness of T^+ and the consistency of M^*. So

(3) $\qquad \vdash_{M^*} T(\overline{\ulcorner\phi\urcorner})$ iff $\vdash_{T^+} \phi$.

But since M^* is finitely axiomatizable, the set $\{\ulcorner\phi\urcorner: \vdash_{M^*} T(\overline{\ulcorner\phi\urcorner})\}$ is recursively enumerable; and so T^+ is axiomatizable.

It should be clear from the proof that the condition of M's being segregated can be replaced by the weaker condition of the number- and object-sorts being unconnected. Philosophically, the result is suggestive; for it shows that a connection must be made between language and the world if truth is to be defined over theories, such as formal arithmetic, that admit of no consistent, complete and axiomatizable extension, i.e., theories which are essentially undecidable. However, it is far removed from showing that language must be linked to the world by anything recognizable as a satisfaction-predicate.

Putting Lemmas 17 and 18 together gives:

THEOREM 19. There is a consistent, finitely axiomatizable and segregated

truth-theory M for T iff T enjoys a consistent, complete and axiomatizable extension.

The "provability" theory M of Lemma 17 is not as useful as the trivial theory M_T in constructing non-Tarskian truth-theories; for whereas M_T is a conservative extension of T, M in general is not. However, we may still construct a non-Tarskian example under an additional supposition on T:

THEOREM 20. Let $n > 0$ be given. Suppose that T has a complete and axiomatizable extension T^+ that is not finitely n-typed. Then there is a non-Tarskian$_n$ finitely axiomatized truth-theory M over T.

Proof. Let the T in Lemma 17 be the T^+ here. Choose the M of Lemma 17. Then M is finitely axiomatized and characterizes truth. Since M consistently extends T^+ and T^+ is complete, M is a conservative extension of T^+. So since T^+ is not finitely n-typed, the restriction $M|L$ of M to L is also not finitely n-typed. But M is segregated; and so by Lemma 2, it does not characterize n-satisfaction.

There exist numerous theories that satisfy the hypothesis of Theorem 20. Examples include logics that are not purely monadic, the theory of linear order, and certain algebraic theories, such as the theories of real closed and algebraically closed fields. Therefore the theorem settles a question of Tharp [8] and Kripke [3], p. 401, as to whether there exists a finitely axiomatizable non-Tarskian truth-theory. We may also note, in the light of the observation following Theorem 8, that there exists, for each $n = 1, 2, \ldots$, a finitely axiomatized truth theory that characterizes n-satisfaction but not $(n + 1)$-satisfaction.

The condition on T in Theorem 20 is not very perspicuous; but it is hard to see how it can be simplified. Certainly, it is not equivalent to the condition that T not be finitely n-typed but enjoy a maximally consistent and axiomatizable extension. For let T be the theory that results from disjoining the conjunction ϕ of axioms for dense linear ordering without end-points with an axiom ψ (in the same predicates) for a theory T_ψ that is essentially undecidable and not finitely n-typed. Then T has a maximally consistent and axiomatizable extension, viz. the result T_ϕ of adding the axiom ϕ. Also, T is not finitely n-typed, since the stronger theory T_ψ is not. However, any consistent, complete and axiomatizable extension

T^+ of T will be deductively equivalent to T_ϕ and hence be finitely n-typed.

The non-Tarskian theory M constructed in Theorem 20 may seem in one respect rather odd. For if we follow through the proof of the Kleene result used in Lemma 16 and hence in Lemma 17, we see that a Tarskian theory of truth may be embedded in the theory M, but with the object-language variables doubling up as the variables for Gödel numbers. Thus the S-sentences may be provable in the meta-theory once the names for the formulas of L are taken to be terms from L_M of the same sort as the object variables. This is not a contradistinction to our claim that M fails to be a satisfaction theory, since that requires that the names in the S-sentences be taken from the language of R. So we see that the notions of a truth or satisfaction theory are very sensitive to the system adopted for naming the sentences of the object-language.

It might be thought that this relativity somewhat detracts from the interest of our example in Theorem 20. After all, the S-sentences can be derived, though under a different system of naming. But two points need to be set against this objection. First, as long as it is assumed that the extension T^+ is itself finitely axiomatizable, the appeal to Kleene's result can be avoided. Second, our interest in constructing a truth-theory may concern a particular system of naming; for under the intended interpretation of the language, it is only certain terms that will be regarded as naming expressions and so it is with reference to these that we would like to derive the T-sentences.

All the same, there may be some interest attaching to the concept of being a truth or satisfaction theory under *some* system of naming sentences. One would then like to know when there exists a truth theory that characterizes truth under one system of naming and yet fails to characterize satisfaction under *any* system of naming (the same or not). The negative part of such a claim, though, would seem to present peculiar difficulties; since, for suitably weak object-theories, there may be a system of naming from the ontology of the theory itself for which the S-sentences could be derived.

Theorem 20 does not give necessary and sufficient conditions. However, the result can be reversed in case T itself is consistent, complete and axiomatizable:

THEOREM 21. For T consistent, complete and axiomatizable, there is a non-Tarskian$_n$ finitely axiomatized truth-theory over T iff T is not finitely n-typed.

Proof. $\Leftarrow$. By Theorem 20.

$\Rightarrow$. By Lemma 3.

Even when T is finitely n-typed for each $n = 1, 2, \ldots$, there may be no uniform method of obtaining a formula $S_n(\mathbf{x}, \alpha)$ that characterizes n-satisfaction. For it follows from the general result of [6], that there exists a structure $\mathfrak{A}$ such that (i) the theory $\mathrm{Th}(\mathfrak{A})$ is decidable, (ii) $\mathfrak{A}$ realizes only finitely many n-types for each n (i.e., $\mathfrak{A}$ is $\aleph_0$-categorical) and yet (iii) there is no effective function f taking each n into the number $f(n)$ of n-types realized by $\mathfrak{A}$. Given such a structure $\mathfrak{A}$, we can use the proof of Theorem 21 to produce a finitely axiomatized truth-theory M over $\mathrm{Th}(\mathfrak{A})$ that characterizes n-satisfaction for each n and yet for which there exists no effective function f taking each n into the Gödel number $f(n)$ of a formula $S_n(\mathbf{x}, \alpha)$ that characterizes n-satisfaction in M.

As in the non-finite case, we may obtain necessary and sufficient conditions for a language by letting the object-theory be a logic:

THEOREM 22. There is a non-Tarskian finitely axiomatized truth-theory for a language L iff L is not purely monadic.

Proof. $\Leftarrow$. By Lemma 10 and Theorem 20.

$\Rightarrow$. By the same direction of Theorem 15.

We note that the condition in Theorem 2 is the same as for Theorem 12. Thus the requirement that the meta-theory be finitely axiomatizable makes no difference as to whether there exists a non-Tarskian truth-theory for a given language.

Turning to the concept of implicit definability, we may establish the natural analogue of Theorem 20, and by essentially similar methods:

THEOREM 23. Suppose that $\mathrm{Th}(\mathfrak{A})$ is axiomatizable and not finitely n-typed. Then there exists a finitely axiomatized meta-theory M that implicitly defines truth but not n-satisfaction for $\mathfrak{A}$.

Indeed, in this case, a much stronger result can be established. Call a

structure $\mathfrak{A}$ *arithmetical* if $|\mathfrak{A}| = \omega$ and each relation or function of $\mathfrak{A}$ is arithmetically definable. Then:

THEOREM 24. Suppose that $\mathfrak{B}$ is elementarily equivalent to an arithmetical structure $\mathfrak{A}$ and that $\mathfrak{B}$ realizes infinitely many n-types. Then there exists a finitely axiomatized theory M that implicitly defines truth but not n-satisfaction for $\mathfrak{B}$.

Proof. Without loss of generality, we may suppose that $\mathfrak{B}$ contains no distinguished elements or functions. For any n-ary predicate R of the language L of $\mathfrak{B}$, let $\phi_R(x_1, \ldots, x_n)$ be an arithmetical formula that defines the arithmetical relation corresponding to R in $\mathfrak{A}$. Let L^+ be obtained from L by adding the numerals $\bar{0}, \bar{1}, \ldots$ obtained from $\bar{0}$ by iterated application of a function symbol for successor. Construct a meta-theory M for L^+ as follows. L_M is the result of combining L^+ with the language of R and adding the truth predicate $T(\alpha)$. The axioms of M are those of R together with the arithmetized versions of the following, where in (i) R represents any n-ary predicate of L:

(i) for any $k_1, \ldots, k_n, R\bar{k}_1 \ldots \bar{k}_n$ is true iff $\phi_R(k_1, \ldots, k_n)$;

(ii) for any sentence ϕ of L^+, $\sim\phi$ is true iff ϕ is not true;

(iii) for any sentences ϕ and ψ of L^+, $(\phi \vee \psi)$ is true iff ϕ is true or ψ is true;

(iv) for any sentence $\exists x\phi(x)$ of L^+, $\exists x\phi(x)$ is true iff $\phi(\bar{n})$ is true for some n.

The axioms (i)–(iv) constitute a substitutional characterization of truth in $\mathfrak{A}$.

It is readily checked that M implicitly defines truth for $\mathfrak{A}$, and hence for $\mathfrak{B}$. However, since M is segregated and $\mathrm{Th}(\mathfrak{A})$ is not finitely n-typed, it follows by Theorem 13 that M does not implicitly define n-satisfaction for $\mathfrak{B}$.

We see from the proof that the analogue of Lemma 18 fails. For let $\mathfrak{A}$ be the standard model for arithmetic and let M be the result of adding to the minimal meta-theory the recursive axioms for the substitutional interpretation of truth. Then M is finitely axiomatized and segregated; it

implicitly defines truth for $\mathfrak{A}$; and yet $\mathrm{Th}(\mathfrak{A})$ is not axiomatizable. It is because of this difference that the method of using segregated meta-theories has greater application in regard to implicit definability.

It would be nice to establish full necessary and sufficient conditions for the existence of theories defining truth without satisfaction, as in the previous section. A natural conjecture in this direction is:

(I) There exists a finitely axiomatized non-Tarskian$_n$ truth theory M over T iff T is not finitely n-typed.

(II) There exists a finitely axiomatized meta-theory M that implicitly defines truth without implicitly defining n-satisfaction for $\mathfrak{A}$ iff $\mathrm{Th}(\mathfrak{A})$ is not finitely n-typed.

The conditions here are the same as for Theorems 8 and 15. Thus if the conjectures are correct, it means that the finiteness condition can do no extra work in forcing a theory of truth to be a theory of satisfaction.

Although we have not been able to settle either conjecture, we have been able to find a construction that may prove the first conjecture and help with certain cases of the second. We add, to the minimal meta-theory, variables for the following extra sorts of entities: restricted or F-sequences of objects; unrestricted sequences of objects; unrestricted sequence indices; and classes of unrestricted sequences. We also add extra non-logical constants for the following notions: the "inner domain" F of objects; membership between unrestricted sequences and classes; an unrestricted value function (Val), taking unrestricted sequences and their indices into objects; a restricted value function, taking restricted sequences and numbers into objects; an unrestricted successor function on the unrestricted sequence indices; and satisfaction between restricted sequences and numbers. Finally, we add the following finitely many extra axioms:

(i) theory of successor for unrestricted successor;

(ii) closure of classes of unrestricted sequences under extensions of object-language relations, complementation, union, projection, and interchange;

(iii) embedding: $\forall X \forall s(s \in X \supset \exists s' \in X \wedge \forall i(F\mathrm{val}(s', i)) \wedge \forall i(F\mathrm{val}(s, i) \supset \mathrm{val}(s', i) = \mathrm{val}(s, i)))$;

(iv) a Tarskian theory of truth for F-sequences and F-relativized formulas.

The resulting meta-theory M characterizes truth. For from (i) and (ii), it follows that, for each formula of the object language, there is a class that is its extension. With the help of (iii), it then follows that each sentence ϕ of L is provably equivalent to its F-relativization ϕ^F. From (iv) we may characterize truth for F-relativized sentences. So truth for arbitrary sentences ϕ of L may be characterized as truth of its relativization ϕ^F. It seems plausible, for all of the theories under (I) and for many of the structures under (II), that M should fail to characterize or to implicitly define n-satisfaction; but we have not been able to prove this.

However, it can be shown that it is not necessary that T have a consistent, complete and axiomatizable extension in order to enjoy a non-Tarskian truth theory. For let T be a theory whose non-logical constants are the usual arithmetical predicates, a two-place ordering predicate R, and a one-place predicate N ("is a number"), and whose axioms are those of Robinson's arithmetic relativized to N, those of an infinite discrete order with first element relativized to $\sim N(x)$, and certain disjointness assumptions stating that the arithmetical predicates apply only to numbers and that the ordering predicate R applies only to non-numbers. Let M be the meta-theory over T that combines a Tarski satisfaction theory for the formulas of L that are relativized to $N(x)$ with a "provability" truth-theory for the sentences of L that are relativized to $\sim N(x)$. Clearly, T is essentially undecidable and M is finitely axiomatized. Since a segregation result holds for the formulas of L, M will characterize truth for L. But, for essentially the same reasons as before, M will not be able to yield the S-sentences for formulas of L that are relativized to $\sim N(x)$.

4. OTHER CONDITIONS

We have shown that there exist non-Tarskian truth-theories that (a) are finitely axiomatizable and (b) contain a given object-theory. The conditions (a) and (b) may be regarded as requirements on a reasonable theory of truth. The question then arises as to how stable are our results over the satisfaction of further requirements of this sort.

One such requirement is that the meta-theory M be a *conservative*

extension of the object theory T, i.e., that ϕ is a theorem of M, for ϕ a sentence from the language L of T, only if ϕ is a theorem of T. It is readily shown that Tarski's satisfaction theory is a conservative extension of the object-theory; for any model of the latter can be extended in a standard way to a model of the former. On the other hand, the provability meta-theory of Theorem 17 will not in general be a conservative extension of the object-theory, since it will be complete over the sentences from the language of that theory.

What happens, then, when the requirement of conservative extendibility is made? The answer turns out to be very simple:

THEOREM 25. Suppose there exists a non-Tarskian$_n$ truth-theory M over T Then there also exists a non-Tarskian$_n$ truth-theory M' that is a conservative extension of T and that is finitely axiomatized if M is.

Proof. Let M_0 be the Tarski theory over T. We may suppose that the semantical languages of M and M_0 contain only a primitive truth predicate in common. Now let M' have as its language the union of the languages of M and M_0, and have as its axioms the disjunction of any conjunction of axioms from M with any conjunction of axioms from M_0 (deleting any repetition of conjuncts).

Then M' is a truth-theory for the language L of T. For let its truth-predicate $T(\alpha)$ be the common truth-predicate of M and M_0. Since M and M_0 are truth-theories:

(1) $\quad \vdash_M T(\overline{\ulcorner\phi\urcorner}) \equiv \phi$; and

(2) $\quad \vdash_{M_0} T(\overline{\ulcorner\phi\urcorner}) \equiv \phi$,

for any sentence ϕ of L. So since M' contains the common theorems of M and M_0:

(3) $\quad \vdash_{M'} T(\overline{\ulcorner\phi\urcorner}) \equiv \phi$.

M' is non-Tarskian$_n$. For suppose all S-sentences were provable in M' relative to the formula $S_n(x, \alpha)$. Let M^+ be the expansion of M to the language of M'. Then clearly all of the above S-sentences are provable in M^+. But on a suitable replacement of the non-logical constants occurring in $S_n(\mathbf{x}, \alpha)$ but not in the language L_M of M, a formula $S_n'(\mathbf{x}, \alpha)$ of L_M can be found relative to which all S-sentences can be proved in M, contrary to our supposition.

M' is an extension of T, since both M and M_0 are. It is also a conservative extension; since if $\vdash_{M'} \phi$, for ϕ a sentence of L, then $\vdash_{M_0} \phi$, and so $\vdash_T \phi$, by the conservative extension result for M_0.

Finally, it is clear from the construction that M' is finitely axiomatized if M is.

From Theorem 25 it follows that:

COROLLARY 26. Under the supposition of Theorem 20, there is a non-Tarskian$_n$ finitely axiomatized truth-theory that is a conservative extension of T. It also follows, if the earlier conjecture (I) is true, that (I) remains true under the additional requirement that the non-Tarskian truth-theory be conservative over T.

Another requirement concerns the standard interpretation of the object-language and of the arithmetical portion of the meta-language. If M is conservative over T, then M will be compatible with whatever are the truths of L under some intended interpretation of T. But we may also require M to permit an interpretation that is compatible with the intended interpretation of the non-semantical parts of its language. Accordingly, let us say that M is *semantically conservative over* T if (a) M extends T and (b) for any model $\mathfrak{A}$ of T there is an arithmetically standard model $\mathfrak{C}$ of M that respects $\mathfrak{A}$ (cf. the definition of implicit definability). Using the theory M' from the proof of Theorem 25, it may be shown that:

THEOREM 27. Under the supposition of Theorem 25, there is a non-Tarskian$_n$ truth-theory M' that is semantically conservative over T and that is finitely axiomatized if M is.

Proof. It suffices to show that M' satisfies condition (b) above. Pick a model $\mathfrak{A}$ for T. Now the Tarski theory M_0 is semantically conservative over T. So there is an arithmetically standard model $\mathfrak{C}$ for M_0 that respects $\mathfrak{A}$. But since each theorem of M_0 is a theorem of M', $\mathfrak{C}$ can be expanded to a similar model for M'.

In the light of this result, "semantically conservative" may be substituted for "conservative" in Corollary 26 and in the revised formulation of conjecture (I).

A rather different requirement on a truth-theory concerns the strength of its arithmetical part. After all, it would be a kind of deductive freak if it were only the absence of certain commonly acceptable arithmetical principles that prevented the derivation of the S-sentences. A weak requirement of this kind is that the induction scheme hold for all formulas of the meta-language, not just the arithmetical part; and a very strong requirement is that all arithmetical truths be at the disposal of the truth-theory. We do not know to what extent such requirements might effect the existence of non-Tarskian truth theories. A result with full induction is given in Kripke ([3], p. 400); but his other assumptions are rather special and involve identifying the object-theory with the arithmetical part of the meta-theory.

There are no doubt other requirements that might be imposed upon a truth-theory. Some may limit the scope of our results or conjectures. But we are inclined to think, contrary to the drift of Wallace [9], that there is no set of general and formal desiderata whose presence will exclude the possibility of a non-Tarskian truth theory in a large number of interesting cases.

NOTES

* This paper had its origin in 1980 in discussions between the authors of some technical questions raised by Kripke in [3]. We would like to thank a referee of this Journal for pointing out an error in the formulation of Lemma 10. We also thank the proprietors of Drake's Tea Shop for providing an environment suitable for research.

REFERENCES

[0] Chang, C. C. and Keisler, H. J., *Model Theory*, North-Holland, Amsterdam, 1973.
[1] Davidson, D., 'Truth and meaning', *Synthese* 17 (1967) 304–323.
[2] Kleene, S. C., 'Finite Axiomatizability of theories in the predicate calculus using additional predicate symbols', *Two Papers on the Predicate Calculus, Memoirs of the American Mathematical Society*, No. 10 (1952), 27–68.
[3] Kripke, Saul, 'Is there a problem about substitutional quantification?' in G. Evans and J. McDowell (eds.), *Truth and Meaning: Essays in Semantics* (Oxford, 1976), 325–419.
[4] Rosenstein, J. G., 'Theories which are not $\aleph_0$-categorical', *Proceedings of the Summer School in Logic, Leeds '67*, Lecture Notes in Mathematics, Springer-Verlag, pp. 273–278.
[5] Ryll-Nardzewski, C., 'On categoricity in power $\aleph_0$', *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys.* 7 (1959), 545–548.
[6] Schmerl, J. H., 'A decidable $\aleph_0$-categorical theory with a non-recursive Ryll-Nardzewski function', *Fundamenta Mathematicae* 98 (1978), No. 2, 121–125.

[7] Tarski, A., 'The concept of truth in formalized languages', in A. Tarski, *Logic, Semantics and Metamathematics* (Oxford, 1956), 152–278.
[8] Tharp, Leslie H., 'Truth, quantification, and abstract objects', *Nous* V (1971), 363–372.
[9] Wallace, J., 'On the frame of reference', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, D. Reidel, 1972, pp. 219–252.

Kit Fine,
Department of Philosophy,
The University of Michigan,
Ann Arbor,
MI 48109, U.S.A.

Timothy McCarthy,
Department of Philosophy,
University of Illinois,
Urbana,
IL 61801, U.S.A.