THE CONCEPTION OF SEMANTIC TRUTH

by
Hoke ROBINSON

In this paper I would like to examine the sense in which Tarski's "Semantic Conception of Truth" ([1]) contains a conception of truth. I shall first give my interpretation of Tarski's argument in the article of that title (supplemented by his "The Concept of Truth in Formalized Languages" ([2]) and other works), in order to make clear what it is I am discussing, and then follow this with observations on certain problems raised by Tarski's notion of truth which I am unable to clear up by reference to that text alone.

Tarski begins his attempt to "catch the meaning of an old notion" ([3]) by stating the general criteria a definition of truth must satisfy: the definition must be *materially adequate*, in the sense that it must be applicable without violating unnecessarily our intuitions about truth; and it must be *formally correct*, i.e. it must not violate formal criteria of correct language-usage. The question, however, is: what are the *conditions* of material adequacy; and what are the *grounds* of the criteria (concepts, rules, etc.) of formal correctness.

Tarski begins by stipulating that the term "true" will initially be applied to *sentences*, not to either "propositions" nor to any other objects – though the possibility of extension to other objects remains open. As restricted for the moment to sentences, however, truth will be relative to the language of which the sentence is a sentence. As a starting-point in setting conditions of material adequacy, Tarski appeals to Aristotle's formula – "... to say of what is that it is ..., is

([1]) Alfred TARSKI, "The Semantic Conception of Truth and the Foundations of Semantics," in *Readings in Philosophical Analysis* (Herbert Feigl and Wilfrid Sellars, eds.). New York: Appleton-Century-Crofts, Inc., 1949, pp. 52-84. Reprinted from *Philosophy and Phenomenological Research*, Vol. IV, 1944. Hereinafter referred to as SCT.

([2]) Alfred TARSKI, "The Concept of Truth in Formalized Languages," in Tarski, *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press, 1957. Translated by J. H. Woodger from "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, Vol. 1, 1937. Hereinafter referred to as CTFL.

([3]) SCT, p. 53.

true'' – and refines this into a restatement of the correspondence-theory: a sentence is true if it agrees with / corresponds to / designates an existent state-of-affairs / reality. In order to avoid the vagueries of the slash-connected locutions, however, he starts with a concrete example: ''The sentence 'snow is white' is true if, and only if, snow is white.'' ([4])

The prima facie triviality of this definition is vitiated by the explanation that the quoted expression is the *name* of the expression unquoted, and not that expression itself. If, then, this statement expresses the material conditions of the truth of the quoted expression, then *a definition of truth is materially adequate only if from it all expressions of the same form as that statement are derivable*. That form, isolated, is: if p is any sentence, and X is the *name* of that sentence, then:

(T) *x is true if, and only if, p.*

Any substitution-instance of this equivalence is an equivalence of the form (T)([5]).

This ''semantic'' conception of truth allies it with other semantic notions, such as ''satisfaction,'' ''designation'' and ''definition,'' and thus brings it within range of the family of antinomies known as the semantic paradoxes. As these can be dealt with only insofar as the language used is unambiguous in the relevant respects, and as ordinary language is generally *not* thus unambiguous, Tarski begins his discussion of the *formal correctness* of the semantic definition of truth by detailing the notion of a language with a *specified structure*. Such a specification must determine (1) what expressions are to be considered meaningful; (2) what expressions are primitive (undefined); (3) what rules introduce defined terms; (4) which expressions are sentences; (5) when sentences are to be asserted([6]); (6) which *sentences* are primitive (axioms); and (7) what the rules of inference or proof are to be. Granted a language whose structure is thus

([4]) CST, p. 54.

([5]) Thus each instance is a *partial* definition of truth, i.e. a definition of the truth of *that* sentence; the definition of truth in general would be the totality of these.

([6]) Note that, having abandoned ordinary language, we are dealing with a *formalized* language (see below).

specified, Tarski attacks the paradox of Epimenides the Liar.

Epimenides the Cretan said, "All Cretans are liars." If we represent the quoted expression as c, then to suppose c true is to deny what Epimenides said, and hence to suppose c false. On the other hand, to suppose c as false is to affirm what Epimenides said, and hence to suppose c true. We should scarcely call a definition of truth "formally correct" which allowed the inference of the denial of a sentence from its affirmation, and conversely. Tarski finds two essential([7]) presuppositions leading to the paradox, one of which must thus be altered if the paradox is to be avoided. One of these is the assumption that the laws of logic hold; this is clearly to be abandoned only under utmost duress. The other is that the (now, specified-structure) language in question is "semantically closed."

To assume that the language producing the antinomy is semantically closed is to assume that it contains its expressions, the names of each expression, the term "true", and all expressions determining the use of "true". If we are not to reject the laws of logic, we must reject this paradox-producing assumption, and avoid languages which are semantically closed. This can be done by using *two* languages, an "object-language", and a "meta-language"([8]) which "talks about" the object-language. The definition of truth will obviously be in this meta-language, and hence so will all the equivalences of form (T) which it must imply (i.e., in order to be materially adequate). Yet the "p" of (T) is obviously from the object-language. Hence the meta-language must contain *every* sentence of the object-language, together

([7]) Tarski's third assumption – that empirical premises such as "Epimenides is a Cretan" are possible – is, he holds (in footnote 11 to SCT), inessential here, since the paradox can be constructed from the first two premises alone, as follows: S is any sentence of the form, "Every sentence _____." S* is the sentence correlated with each sentence S so that, where S is "Every sentence is ϕ", S* is "The sentence 'every sentence is ϕ' is ϕ." S is "self-applicable" if and only if S* is true, non-self-applicable if and only if S* is false. Then let ϕ be "non-self-applicable". Then if S is ϕ, S* is false (i.e., the sentence "The sentence 'Every sentence is non-self-applicable' is non-self-applicable" is false), and hence S is *not* ϕ. Contrariwise, if S is not ϕ, S* is true, and hence S *is* ϕ. See also the Grelling-Nelson paradox of heterological words, which also appears to require no empirical statements.

([8]) This distinction is relative; discussion about the meta-language is carried on in a higher-level meta-language, relative to which the original meta-language is the object-language.

with the name of each sentence (i.e., X)([9]), logical operators, and semantic terms, especially "true". If the semantic terms, especially "true", are introduced into the meta-language *by definition* (rather than as primitive)([10]), we may consider them *adequate*.

Note that the meta-language definition's formal correctness requires that the meta-language be "essentially richer" than the object-language, i.e. it must contain every term in the object-language,([11]) and some the object-language does *not* contain (e.g., the *names* of object-language sentences X, and "true", etc.). If it were not essentially richer, it could be interpreted in the object-language, and Epimenides' Paradox could be reconstructed at the meta-level.([12])

([9]) In CTFL, Tarski uses a variety of Polish logic symbolism for the object-language, a variety of the calculus of classes symbolism for the *translation* of an object-language sentence into the meta-language (so that the meta-language "contains" the object-language in the sense that it includes a translation of every object-language sentence), and another notation (perhaps based on Boole) for the *names* of these sentences. Thus the sentence "given any two classes, one contains the other" is in the object-language:

$$\Pi x \Pi x \; A I x x \; I x x \quad ,$$

in its meta-language translation:

$$\text{for any classes a and b we have } a \subseteq b \text{ or } b \subseteq a \quad ,$$

and has as its *name* in the meta-language:

$$\cap_1 \cap_2 (\iota_{1,2} + \iota_{1,2}) \quad .$$

(See CTFL, p. 187.)

([10]) As Tarski points out (in "The Establishment of Scientific Semantics" [ESS], *Logic, Semantics, Metamathematics*, p. 405, and elsewhere, introduction of the term "true" by axiom has two disadvantages: there is always the "accidental" or contrived character that attaches to stipulated axioms, violating our intuitions of the material adequacy of the term; and there is the possibility that further axioms could render the meta-language inconsistent. Both these problems are avoided if "true" is introduced by definition.

([11]) See note 8.

([12]) That is to say, unless the meta-language is richer than the object-language, the object-language remains "closed", that is, everything that can be said in the meta-language can be said, *mutatis mutandis*, in the object-language, and the semantic paradoxes are reconstructable. See especially CTFL, pp. 247-251, in which it is shown that where a richer meta-language is not possible (in this case, because the object-language is of infinite order, and transfinite ordinals are not considered), both a sentence and its negation are derivable (following Gödel). See also CTFL, p. 274.

Within this framework of language and meta-language, Tarski defines truth via the semantic notion of *satisfaction*. This is defined as a relation sometimes obtaining between objects and sentential functions. Sentential functions are defined "recursively", that is, the simplest cases are described, together with rules which may be applied recursively to these simple cases to generate all and only sentential functions. The notion of satisfaction, according to Tarski, may also be defined recursively: "We indicate which objects satisfy the simplest sentential functions; and then we state the conditions under which given objects satisfy a compound function – assuming that we know which objects satisfy the simpler functions from which the compound one has been constructed." ([13]) Given this definition of satisfaction and of sentential function, Tarski defines "truth" as applying to those sentences (sentential functions with no free variables) which are satisfied by *all* objects, all other sentences being false.

"The Semantic Conception of Truth" was published in English in 1944; in 1951 R. M. Martin, noting that "after reading the relevant passage the reader may feel that he knows no more about the truth concept than he did before," attempted to clarify this passage([14]) (CTFL had not yet appeared in English). Why is it the case, asks Martin, that "for a sentence only two cases are possible: a sentence is either satisfied by all objects, or by no objects"?([15]) Tarski, he says, talks of functions being satisfied, not by objects one-at-a-time, but rather by *infinite sequences* f of objects f_1, f_2, ... In this way a function of one argument is satisfied by sequence f if f_1 satisfies it; a two-argument function is satisfied by f if f_1 and f_2 respectively satisfy it; etc. Then the partial definition of satisfaction – e.g., "satisfaction for 2-place function (relation) R" – is the correlating of it to a sequence f such that $f_1 R f_2$.

Martin then quotes Lamma A of CTFL:

LEMMA A. If the sequence f satisfies the sentential function x, and the infinite sequence g... satisfies the following condition: for every

([13]) SCT, p. 63.

([14]) R. M. Martin, "Discussion on Tarski's 'Semantic Conception of Truth'", *Philosophy and Phenomenological Research*, Vol. XI (Mar. 1951), pp. 411-12.

([15]) SCT, p. 63.

k [where k is some positive integer, and members of sequences, as well as variables in functions, are ordered – say, alphabetically – and numbered], $f_k = g_k$ if v_k is a free variable of the function x; then the sequence g also satisfies the function x.([16])

This lemma presents very little problem as it stands; it is clear that we are dealing with infinite sequences in order to avoid having to construct a different notion of satisfaction for one-argument functions, two-argument functions, etc.; but whether or not a sequence satisfies a function depends only on those members of the sequence corresponding to variables in the functions; all others are irrelevant. Thus only f_1 and f_2 are relevant in deciding whether sequence f satisfies two-argument function R; f_3, f_4, etc. are immaterial. Hence, obviously, if some other sequence g has members identical with those of f in all the relevant places (i.e., those with indices corresponding to the indices of the free variables of the function under consideration), then if f satisfies the function, so will g, no matter how much the sequences may differ in the irrelevant places.

Since a sentence is a function with *no* free variables, there are *no* members of any sequence which are relevant to the satisfaction of this function, and hence Lemma B follows trivially:

LEMMA B. If x [is a sentence] and at least one infinite sequence ... satisfies the sentence x, then every infinite sequence ... satisfies x.([17])

This exposition does indeed clear up the sense of Tarski's remark on "satisfaction by all objects"; but if Martin thinks that, by virtue of his exposition alone, the reader (or, to speak for myself, *this* reader) knows any more about the truth-concept than he did before, he is sadly mistaken. The feeling persists that a great deal of logical slight-of-hand has been performed, and that, rabbit to the contrary notwithstanding, there is something funny about that hat.

To begin with, what can it mean to speak of an "object" "satisfying" a sentence? We do speak of terms, perhaps loosely of objects, satisfying a sentential *function*; but in *PM* and elsewhere a function is

([16]) P. 198; I use Woodger's translation rather than Martin's.
([17]) CTFL, p. 198.

a sentence with one or more *vacant places*, and the function *is* satisfied *(gesättigt)* when these places are filled. Thus, it would seem, all objects (terms) of all sequences would neither satisfy nor fail to satisfy all sentences, which would consequently be of indeterminate truth-value. If the truth of, "We are now under nuclear attack" depends upon the sequence "Hamster, yak, abominable snowman, . . .", we are in sad shape.

But Tarski clearly does not mean this. In company with every serious thinker since Kant, he holds that a criterion of *material* truth is not possible:

In fact, the semantic definition of truth implies nothing regarding the conditions under which a sentence like (1):

(1) snow is white

can be asserted. It implies only that, whenever we assert or reject this sentence, we must be ready to assert or reject the correlated sentence (2):

(2) the sentence "snow is white" is true.

Thus, we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians – whatever we were before. *The semantic conception is completely neutral toward all these issues.* ([18])

We are not, therefore, using the semantic conception to determine the assertability of an object-language sentence such as (1) above. But then, what is the purpose of the semantic conception, if it is neutral toward epistemological attitudes?

Notice that in Lemmas A and B quoted above there was a deletion following the term "sequence". In Martin's version of these lemmas, sequences are explained as sequences *of objects*,([19]) and the key

([18]) SCT, p. 71 (emphasis mine).
([19]) Martin, p. 412.

sentence in SCT also says, "A sentence is true if it is satisfied by all *objects*, and false otherwise."([20]) But the original Lemmas A and B appear in the subsection of CTFL entitled, "The Concept of True Sentence in the Language of the Calculus of Classes," and the lemmas there speak of sequences of *classes*. Thus in this *particular* case, a sentence in the formalized language of the calculus of classes is true if it is satisfied by all infinite sequences of *classes*, and false otherwise.

We may, however, still ask: when is this the case? A closer look at the notion of a *formalized* language may help us here. Tarski first excludes ordinary or colloquial language from consideration: since this *is* semantically closed, no "formally correct" definition is possible: the antinomy of the Liar is always constructable in such languages.([21]) Formal correctness is only possible in *formalized* languages, i.e. languages with a specified structure.([22]) We noted above([23]) that one condition of the specifying of the structure of a language is that the conditions under which sentences in the language are assertable be specified. As it is obviously *not* possible to specify the conditions under which material (empirical) sentences are to be asserted within a formalized language, we are dealing either with equivalences of the kind that, e.g., *if* sentence (1) above is assertable, so is sentence (2), and conversely; or we are dealing with "logical truths" of the formalized language in question: primitive sentences (axioms) stipulated to be assertable *without* proof or evidence, together with those sentences (theorems) the evidence for which consists in the axioms and the "truth-preserving" inference-rules which produce these theorems from the axioms. It is clearly sentences of the second kind which Tarski refers to as "true if satisfied by every infinite sequence of objects."

Let us examine the formalized language of Section 3 of CTFL, the language of the calculus of classes. The definition of truth for this language which is held to result from Lemmas A and B is:

DEFINITION 23. x is a *true sentence* – in symbols $x \in Tr$ – if and

([20]) SCT, p. 63 (emphasis mine).
([21]) CTFL, pp. 157-165, and SCT, pp. 57-60.
([22]) SCT, p. 57; ESS, p. 403.
([23]) P. 94, number (5).

only if $x \in S$ [i.e., "x is a sentence"; see Def. 12, p. 178] and every infinite sequence of classes satisfies x.([24])

The move to classes has not, of course, solved our problem as to how a sentence can either be satisfied or fail to be satisfied by classes or by any other "objects". The notion of "satisfaction" is defined in Def. 22, but contains the term "function", the definition for which is:

DEFINITION 10. x is a *sentential function* if and only if x is an expression which satisfies one of the four following conditions: (α) there exists natural numbers k and l such that $x = \iota_{k,l}$ [$\iota_{k,l}$ reads, "the k-th class-variable is included in the l-th class-variable," or "class l includes class k."]([25]); (β) there exists a sentential function y such that $x = \bar{y}$ [$\bar{y}$ reads $\sim y$.]; (γ) there exist sentential functions y and z such that $x = y + z$ [$y + z$ read $y \vee z$.]; (δ) there exists a natural number k and a sentential function y such that $x = \cap_k y$ [read, "the universal quantification of the k-th variable over the function y.].([26]) (β), (γ) and (δ) all presuppose the notion of sentential function: it is (α) which must be examined. In the calculus of classes, the most elementary form of function is that in which one variable is related to another by *inclusion*, and other functions must be defined in terms of this. Tarski goes on to define "sentence" (function without free variables, Def. 12)([27]), stipulates the axioms([28]), and provides the rules for producing theorems from the axioms.([29]) These constitute the conditions under which sentences in this formalized language (i.e., of the calculus of classes) are to be asserted (condition (5), p. 94 above), and thus contribute to specifying its structure.

The definition of satisfaction itself, then, is as follows:

DEFINITION 22. The sequence f *satisfies* the sentential function x if f is an infinite sequence of classes and x is a sentential function and these satisfy one of the following four conditions: (α) there exist

([24]) CTFL, p. 195.
([25]) CTFL, p. 175, Def. 1.
([26]) CTFL, p. 177, Def. 10.
([27]) CTFL, p. 178.
([28]) CTFL, p. 179 Def. 13.
([29]) CTFL, p. 180-82, Defs. 14-16.

natural number k and a sentential function y such that $x = \cap_k y$ and x = 'the k-th class variable is included in the l-th class-variable, and the k-th member of sequence f is included in the l-th member of sequence f' ''] ; (β) there is a sentential function y such that $x = \bar{y}$ and f does not satisfy the function y ; (γ) there are sentential functions y and z such that x = y + z and f either satisfies y or satisfies z ; (δ) there is a neutral number k and a sentential function y such that $x = \cap_k y$ and every infinite sequence of classes which differs from f in at most the k-th place satisfies the function y.([30])

Once again, (β), (γ) and (δ) contain the term ''satisfies'' already, so it is to (α) that we confine our interest. If we restrict our consideration to functions of one and two variables only, we may talk, not of infinite sequences of objects (classes), but of individual classes and pairs of them (since, in this case, other members of the sequence would be irrelevant to the satisfaction of the functions being considered). We then read (α): ''x is satisfied by f in case x is a two-place function stating that the first free variable is included in the second, and f is a pair of objects (classes) of which the first is included in the second.'' If we universally-quantify the second variable, we get a one-place function, which states: f satisfies x, where $x = \cap_2 \iota_{1,2}$ [read, ''x = _____ is included in all classes''], in case f is an object (class) such that it is included in all classes, i.e. where f = the null class. Note that in both cases, whether or not x is satisfied depends upon properties of f. The question, then, in both cases is: *is* there such a pair, or such an object (class) f, such that it satisfies the function x ? The answer to this question is found in the symbols, axioms and rules of the calculus of classes, the formalized language under consideration here: according to these rules, there *is* a pair of classes such that one includes the other, and there *is* a null class. Hence satisfaction here is determined by which ''objects'' ''exist'' in the language under consideration.

Now it is clear from Lemmas A and B that either *all* objects ''satisfy'' a sentence, or none do ; the question was, what is ''satisfaction'' for sentences. Tarski gives an example([31]) of a true sentence in

([30]) CTFL, p. 192, Def. 22. (See Def. 10 quoted above for interpretation of the symbolism.)

([31]) CTFL, p. 196.

the sense of Def. 23. First, the function $\cap_2 \overline{\iota_{1,2}}$ [read, "the first class-variable is *not* included in any class"] is false for all the first class-variables (and hence its negation, $\cup_2 \iota_{1,2}$ [read, "there is a second class including the first"] is true for all first class-variables) in case there *is* some class including all classes. Tarski's calculus of classes assures us that this is the case, and that furthermore we may universally quantify the first class-variable. Hence we arrive at a true sentence (in the language of the calculus of classes) of form (T):

$\cap_1 \cup_2 \iota_{1,2} \in \mathrm{Tr}$ [read, " 'There is a class which includes all classes' is true"] if and only if for all classes a there is a class b such that $a \subseteq b$.([32])

Thus the truth of $\cap_1 \cup_2 \iota_{1,2}$ is guaranteed by the "existence" of a class b including all classes a, which existence (together with the existence of any classes at all) is guaranteed by the stipulated axioms, rules, etc. of the calculus of classes (including such existential assumptions as the axiom of infinity([33])). A simpler way of putting this definition of the truth of a sentence might be, "x is a true sentence if and only if it asserts something already assumed by the axioms and rules of the formalized language in question." (We ignore undecidable sentences for the moment.) Should the axioms and rules allow the construction of objects (here, classes) which could compose sequences *not* "satisfying" some sentence (i.e., incompatible with it), then (modus tollens on Lemma B), *no* sequence "satisfies" (is compatible with) it. And this result is strange only if we forget that we are dealing with formalized languages here, and thus that there are no "contingencies", only "logical truths" and "logical falsities".

It seems, then, that a true sentence is satisfied by all objects in the sense that no object (as stipulated in the specified structure of the language under consideration) may be other than the sentence describes it as being. If, in the language of the calculus of classes, a sentence says that there is a null class, then this sentence is "satisfied" by all sequences of "objects" in case there *is* a null class (since, as specified by the language, no sequence will contain any object

([32]) CTFL, p. 196.

([33]) See remarks on existential assumptions, CTFL, pp. 183-85.

incompatible with there being a null class). And this same situation would obtain in any other language whose structure could be specified, principally because in specifying the structure, the "objects" this specification stipulates as existent are obtained. The philosophical question of the "truth" of statements is transformed into the question of the existence of certain objects of certain kinds in certain languages; and even if this transformation is helpful logically, it is scarcely illuminating epistemologically. Thus we see the sense of Tarski's note that this conception of truth is epistemologically neutral, that it solves none of the issues between realists, idealists, etc.; it is a formal notion.

One is tempted to ask: then what good is it? As a formal notion, Tarski has used his conception of truth to provide a solution to the age-old semantic paradoxes; and in so doing has arrived at logical results paralleling those of Gödel, but in a more perspicuous form. These results are far from trivial. Tarski is able to show that, since the meta-language must be "essentially richer" (here read, "must contain objects of a higher logical type") than the object-language, then when the object-language is of infinite order, no definition of truth is constructable from which a contradiction is not derivable([34]), and that there are true sentences which are not decidable.([35])

Memphis State University Hoke ROBINSON
Dept. of Philosophy
Memphis, TN 38152
USA

([34]) CTFL, pp. 247-252.
([35]) CTFL, pp. 274-76.