The Semantic Conception of Truth and the Foundations of Semantics

ALFRED TARSKI

This paper consists of two parts: the first has an expository character, and the second is rather polemical.*

In the first part I want to summarize in an informal way the main results of my investigations concerning the definition of truth and the more general problem of the foundations of semantics. These results have been embodied in a work which appeared in print several years ago.[1] Although my investigations concern concepts dealt with in classical philosophy, they happen to be comparatively little known in philosophical circles, perhaps because of their strictly technical character. For this reason I hope I shall be excused for taking up the matter once again.[2]

Since my work was published, various objections, of unequal value, have been raised to my investigations; some of these appeared in print, and others were made in public and private discussions in which I took part.[3] In the second part of the paper I should like to express my views regarding these objections. I hope that the remarks which will be made in this context will not be considered as purely polemical in character, but will be found to contain some constructive contributions to the subject.

In the second part of the paper I have made extensive use of material graciously put at my disposal by Dr. Marja Kokoszyńska (University of Lwów). I am especially indebted and grateful to Professors Ernest Nagel (Columbia University) and David Rynin (University of California, Berkeley) for their help in preparing the final text and for various critical remarks.

Reprinted from *Philosophy and Phenomenological Research*, IV (1944), with permission of Jan Tarski.

*Part II is not reprinted in this volume.

1. EXPOSITION

1. The Main Problem—A Satisfactory Definition of Truth.

Our discussion will be centered around the notion[4] of *truth*. The main problem is that of giving a *satisfactory definition* of this notion, i.e., a definition which is *materially adequate* and *formally correct*. But such a formulation of the problem, because of its generality, cannot be considered unequivocal, and requires some further comments.

In order to avoid any ambiguity, we must first specify the conditions under which the definition of truth will be considered adequate from the material point of view. The desired definition does not aim to specify the meaning of a familiar word used to denote a novel notion; on the contrary, it aims to catch hold of the actual meaning of an old notion. We must then characterize this notion precisely enough to enable anyone to determine whether the definition actually fulfills its task.

Secondly, we must determine on what the formal correctness of the definition depends. Thus, we must specify the words or concepts which we wish to use in defining the notion of truth; and we must also give the formal rules to which the definition should conform. Speaking more generally, we must describe the formal structure of the language in which the definition will be given.

The discussion of these points will occupy a considerable portion of the first part of the paper.

2. The Extension of the Term "True."

We begin with some remarks regarding the extension of the concept of truth which we have in mind here.

The predicate "*true*" is sometimes used to refer to psychological phenomena such as judgments or beliefs, sometimes to certain physical objects, namely, linguistic expressions and specifically sentences, and sometimes to certain ideal entities called "propositions." By "sentence" we understand here what is usually meant in grammar by "declarative sentence"; as regards the term "proposition," its meaning is notoriously a subject of lengthy disputations by various philosophers and logicians, and it seems never to have been made quite clear and unambiguous. For several reasons it appears most convenient to *apply the term "true" to sentences*, and we shall follow this course.[5]

Consequently, we must always relate the notion of truth, like that of a sentence, to a specific language; for it is obvious that the same expression which is a true sentence in one language can be false or meaningless in another.

Of course, the fact that we are interested here primarily in the notion of truth for sentences does not exclude the possibility of a subsequent extension of this notion to other kinds of objects.

3. The Meaning of the Term "True."

Much more serious difficulties are connected with the problem of the meaning (or the intension) of the concept of truth.

The word "*true*," like other words from our everyday language, is certainly not unambiguous. And it does not seem to me that the philosophers who have discussed this concept have helped to diminish its ambiguity. In works and discussions of philosophers we meet many different conceptions of truth and falsity, and we must indicate which conception will be the basis of our discussion.

We should like our definition to do justice to the intuitions which adhere to the *classical Aristotelian conception of truth*—intuitions which find their expression in the well-known words of Aristotle's *Metaphysics*:

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

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

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

(For a theory of truth which is to be based upon the latter formulation the term "correspondence theory" has been suggested.)

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

> A sentence is true if it designates an existing state of affairs.[6]

However, all these formulations can lead to various misunderstandings, for none of them is sufficiently precise and clear (though this applies much less to the original Aristotelian formulation than to either of the others); at any rate, none of them can be considered a satisfactory definition of truth. It is up to us to look for a more precise expression of our intuitions.

4. A Criterion for the Material Adequacy of the Definition.[7]

Let us start with a concrete example. Consider the sentence "*snow is white*." We ask the question under what conditions this sentence is true or false. It seems clear that if we base ourselves on the classical conception of truth, we shall say that the sentence is true if snow is white, and that it is false if snow is not white. Thus, if the definition of truth is to conform to our conception, it must imply the following equivalence:

The sentence "snow is white" is true if, and only if, snow is white.

Let me point out that the phrase "*snow is white*" occurs on the left side of this equivalence in quotation marks, and on the right without quotation marks. On the right side we have the sentence itself, and on the left the name of the sentence. Employing the medieval logical terminology we could also say that on the right side the words "*snow is white*" occur in *suppositio formalis*, and on the left in *suppositio materialis*. It is hardly necessary to explain why we must have the name of the sentence, and not the sentence itself, on the left side of the equivalence. For, in the first place, from the point of view of the grammar of our language, an expression of the form "*X is true*" will not become a meaningful sentence if we replace in it '*X*' by a sentence or by anything other than a name—since the subject of a sentence may be only a noun or an expression functioning like a noun. And, in the second place, the fundamental conventions regarding the use of any language require that in any utterance we make about an object it is the name of the object which must be employed, and not the object itself. In consequence, if we wish to say something about a sentence, for example, that it is true, we must use the name of this sentence, and not the sentence itself.[8]

It may be added that enclosing a sentence in quotation marks is by no means the only way of forming its name. For instance, by assuming the usual order of letters in our alphabet, we can use the following expression as the name (the description) of the sentence "*snow is white*":

the sentence constituted by three words, the first of which consists of the 19th, 14th, 15th, and 23rd letters, the second of the 9th and 19th letters, and the third of the 23rd, 8th, 9th, 20th, and 5th letters of the English alphabet.

We shall now generalize the procedure which we have applied above. Let us consider an arbitrary sentence; we shall replace it by the letter '*p*.' We form the name of this sentence and we replace it by another letter, say '*X*.' We ask now what is the logical relation between the two sentences "*X is true*" and '*p*.' It is clear that from the point of view of our basic conception of truth these sentences are equivalent. In other words, the following equivalence holds:

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

We shall call any such equivalence (with '*p*' replaced by any sentence of the language to which the word "*true*" refers, and '*X*' replaced by a name of this sentence) an "*equivalence of the form* (T)."

Now at last we are able to put into a precise form the conditions under which we will consider the usage and the definition of the term "*true*" as adequate from the material point of view: we wish to use the term "*true*" in such a way that all equivalences of the form (T) can be asserted, and *we shall call a definition of truth "adequate" if all these equivalences follow from it.*

It should be emphasized that neither the expression (T) itself (which is not a sentence, but only a schema of a sentence) nor any particular instance of the form (T) can be regarded as a definition of truth. We can only say that every equivalence of the form (T) obtained by replacing '*p*' by a particular sentence, and '*X*' by a name of this sentence, may be considered a partial definition of truth, which explains wherein the truth of this one individual sentence consists. The general definition has to be, in a certain sense, a logical conjunction of all these partial definitions.

(The last remark calls for some comments. A language may admit the construction of infinitely many sentences; and thus the number of partial definitions of truth referring to sentences of such a language will also be infinite. Hence to give our remark a precise sense we should have to explain what is meant by a "logical conjunction of infinitely many sentences"; but this would lead us too far into technical problems of modern logic.)

5. Truth as a Semantic Concept.

I should like to propose the name "*the semantic conception of truth*" for the conception of truth which has just been discussed.

Semantics is a discipline which, speaking loosely, *deals with certain relations between expressions of a language and the objects* (or "states of affairs") *"referred to" by those expressions.* As typical examples of semantic concepts we may mention the concepts of *designation, satisfaction,* and *definition* as these occur in the following examples:

the expression "the father of his country" designates (denotes) George Washington;

snow satisfies the sentential function (the condition) "x is white";

the equation "$2 \cdot x = 1$" defines (uniquely determines) the number ½.

While the words *"designates" "satisfies,"* and *"defines"* express relations (between certain expressions and the objects "referred to" by these expressions), the word *"true"* is of a different logical nature: it expresses a property (or denotes a class) of certain expressions, viz., of sentences. However, it is easily seen that all the formulations which were given earlier and which aimed to explain the meaning of this word (cf. Sections 3 and 4) referred not only to sentences themselves, but also to objects "talked about" by these sentences, or possibly to "states of affairs" described by them. And, moreover, it turns out that the simplest and the most natural way of obtaining an exact definition of truth is one which involves the use of other semantic notions, e.g., the notion of satisfaction. It is for these reasons that we count the concept of truth which is discussed here among the concepts of semantics, and the problem of defining truth proves to be closely related to the more general problem of setting up the foundations of theoretical semantics.

It is perhaps worth while saying that semantics as it is conceived in this paper (and in former papers of the author) is a sober and modest discipline which has no pretensions of being a universal patent-medicine for all the ills and diseases of mankind, whether imaginary or real. You will not find in semantics any remedy for decayed teeth or illusions of grandeur or class conflicts. Nor is semantics a device for establishing that everyone except the speaker and his friends is speaking nonsense.

From antiquity to the present day the concepts of semantics have played an important rôle in the discussions of philosophers, logicians, and philologists. Nevertheless, these concepts have been treated for a long time with a certain amount of suspicion. From a historical standpoint, this suspicion is to be regarded as completely justified. For although the meaning of semantic concepts as they are used in everyday language seems to be rather clear and understandable, still all attempts to characterize this meaning in a general and exact way miscarried. And what is worse, various arguments in which these concepts were involved, and which seemed otherwise quite correct and based upon apparently obvious premises, led frequently to paradoxes and antinomies. It is sufficient to mention here the *antinomy of the liar,* Richard's *antinomy of definability* (by means of a finite number of words), and Grelling-Nelson's *antinomy of heterological terms.*[9]

I believe that the method which is outlined in this paper helps to overcome these difficulties and assures the possibility of a consistent use of semantic concepts.

6. Languages with a Specified Structure.

Because of the possible occurrence of antinomies, the problem of specifying the formal structure and the vocabulary of a language in which definitions of semantic concepts are to be given becomes especially acute; and we turn now to this problem.

There are certain general conditions under which the structure of a language is regarded as *exactly specified*. Thus, to specify the structure of a language, we must characterize unambiguously the class of those words and expressions which are to be considered *meaningful*. In particular, we must indicate all words which we decide to use without defining them, and which are called "*undefined* (or *primitive*) *terms*"; and we must give the so-called *rules of definition* for introducing new or *defined terms*. Furthermore, we must set up criteria for distinguishing within the class of expressions those which we call "*sentences*." Finally, we must formulate the conditions under which a sentence of the language can be *asserted*. In particular, we must indicate all *axioms* (or *primitive sentences*), i.e., those sentences which we decide to assert without proof; and we must give the so-called *rules of inference* (or *rules of proof*) by means of which we can deduce new asserted sentences from other sentences which have been previously asserted. Axioms, as well as sentences deduced from them by means of rules of inference, are referred to as "*theorems*" or "*provable sentences*."

If in specifying the structure of a language we refer exclusively to the form of the expressions involved, the language is said to be *formalized*. In such a language theorems are the only sentences which can be asserted.

At the present time the only languages with a specified structure are the formalized languages of various systems of deductive logic, possibly enriched by the introduction of certain non-logical terms. However, the field of application of these languages is rather comprehensive; we are able, theoretically, to develop in them various branches of science, for instance, mathematics and theoretical physics.

(On the other hand, we can imagine the construction of languages which have an exactly specified structure without being formalized. In such a language the assertability of sentences, for instance, may depend not always on their form, but sometimes on other, non-linguistic factors. It would be interesting and important actually to construct a language of this type, and specifically one which would prove to be sufficient for the development of a comprehensive branch of empirical science; for this would justify the hope that languages with specified structure could finally replace everyday language in scientific discourse.)

The problem of the definition of truth obtains a precise meaning and can be solved in a rigorous way only for those languages whose structure has been exactly specified. For other languages—thus, for all natural, "spoken" languages—the meaning of the problem is more or less vague, and its solution can have only an approximate character. Roughly speaking, the approximation consists in replacing a natural language (or a portion of it in which we are interested) by one whose structure is exactly specified, and which diverges from the given language "as little as possible."

7. The Antinomy of the Liar.

In order to discover some of the more specific conditions which must be satisfied by languages in which (or for which) the definition of truth is to be given, it will be advisable to begin with a discussion of that antinomy which directly involves the notion of truth, namely, the antinomy of the liar.

To obtain this antinomy in a perspicuous form,[10] consider the following sentence:

The sentence printed in this paper on p. 190, l. 35, is not true.

For brevity we shall replace the sentence just stated by the letter '*s*.'

According to our convention concerning the adequate usage of the term "*true*," we assert the following equivalence of the form (T):

(1) '*s*' is true if, and only if, the sentence printed in this paper on p. 190, l. 35, is not true.

On the other hand, keeping in mind the meaning of the symbol '*s*,' we establish empirically the following fact:

(2) '*s*' is identical with the sentence printed in this paper on p. 190, l. 35.

Now, by a familiar law from the theory of identity (Leibniz's law), it follows from (2) that we may replace in (1) the expression "*the sentence printed in this paper on p. 190, l. 35.*" by the symbol " '*s*.' " We thus obtain what follows:

(3) '*s*' is true if, and only if, '*s*' is not true.

In this way we have arrived at an obvious contradiction.

In my judgment, it would be quite wrong and dangerous from the standpoint of scientific progress to depreciate the importance of this and other antinomies, and to treat them as jokes or sophistries. It is a fact that we are here in the presence of an absurdity, that we have been compelled to assert a false sentence (since (3), as an equivalence between two contradictory sentences, is necessarily false). If we take our work seriously, we cannot be reconciled with this fact. We must discover its cause, that is to say, we must analyze premises upon which the antinomy is based; we must then reject at least one of these premises, and we must investigate the consequences which this has for the whole domain of our research.

It should be emphasized that antinomies have played a preëminent rôle in establishing the foundations of modern deductive sciences. And just as class-theoretical antinomies, and in particular Russell's antinomy (of the class of all classes that are not members of themselves), were the starting point for the successful attempts at a consistent formalization of logic and mathematics, so the antinomy of the liar and other semantic antinomies give rise to the construction of theoretical semantics.

8. The Inconsistency of Semantically Closed Languages.[7]

If we now analyze the assumptions which lead to the antinomy of the liar, we notice the following:

(I) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "*true*" referring to sentences of this language; we have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called "*semantically closed.*"

(II) We have assumed that in this language the ordinary laws of logic hold.

(III) We have assumed that we can formulate and assert in our language an empirical premise such as the statement (2) which has occurred in our argument.

It turns out that the assumption (III) is not essential, for it is possible to reconstruct the antinomy of the liar without its help.[11] But the assumptions (I) and (II) prove essential. Since every language which satisfies both of these assumptions is inconsistent, we must reject at least one of them.

It would be superfluous to stress here the consequences of rejecting the assumption (II), that is, of changing our logic (supposing this were possible) even in its more elementary and fundamental parts. We thus consider only the possibility of rejecting the assumption (I). Accordingly, we decide *not to use any language which is semantically closed* in the sense given.

This restriction would of course be unacceptable for those who, for reasons which are not clear to me, believe that there is only one "genuine" language (or, at least, that all "genuine" languages are mutually translatable). However, this restriction does not affect the needs or interests of science in any essential way. The languages (either the formalized languages or—what is more frequently the case—the portions of everyday language) which are used in scientific discourse do not have to be semantically closed. This is obvious in case linguistic phenomena and, in particular, semantic notions do not enter in any way into the subject-matter of a science; for in such a case the language of this science does not have to be provided with any semantic terms at all. However, we shall see in the next section how semantically closed languages can be dispensed with even in those scientific discussions in which semantic notions are essentially involved.

The problem arises as to the position of everyday language with regard to this point. At first blush it would seem that this language satisfies both assumptions (I) and (II), and that therefore it must be inconsistent. But actually the case is not so simple. Our everyday language is certainly not one with an exactly specified structure. We do not know precisely which expressions are sentences, and we know even to a smaller degree which sentences are to be taken as assertible. Thus the problem of consistency has no exact meaning with respect to this language. We may at best only risk the guess that a language whose structure has been exactly specified and which resembles our everyday language as closely as possible would be inconsistent.

9. Object-Language and Meta-Language.

Since we have agreed not to employ semantically closed languages, we have to use two different languages in discussing the problem of the definition of truth and, more generally, any problems in the field of semantics. The first of these languages is the language which is "talked about" and which is the subject-matter of the whole discussion; the definition of truth which we are seeking applies to the sentences of this language. The second is the language in which we "talk about" the first language, and in terms of which we wish, in particular, to construct the definition of truth for the first language. We shall refer to the first language as "*the object-language*," and to the second as "*the meta-language*."

It should be noticed that these terms "object-language" and "meta-language" have only a relative sense. If, for instance, we become interested in the notion of truth applying to sentences, not of our original object-language, but of its meta-language, the latter becomes automatically the object-language of our discussion; and in order to define truth for this language, we have to go to a new meta-language—so to speak, to a meta-language of a higher level. In this way we arrive at a whole hierarchy of languages.

The vocabulary of the meta-language is to a large extent determined by previously stated conditions under which a definition of truth will be considered materially adequate. This definition, as we recall, has to imply all equivalences of the form (T):

(T) X is true if, and only if, p.

The definition itself and all the equivalences implied by it are to be formulated in the meta-language. On the other hand, the symbol 'p' in (T) stands for an arbitrary sentence of our object-language. Hence it follows that every sentence which occurs in the object-language must also occur in the meta-language; in other words, the meta-language must contain the object-language as a part. This is at any rate necessary for the proof of the adequacy of the definition—even though the definition itself can sometimes be formulated in a less comprehensive meta-language which does not satisfy this requirement.

(The requirement in question can be somewhat modified, for it suffices to assume that the object-language can be translated into the meta-language; this necessitates a certain change in the interpretation of the symbol 'p' in (T). In all that follows we shall ignore the possibility of this modification.)

Furthermore, the symbol 'X' in (T) represents the name of the sentence which 'p' stands for. We see therefore that the meta-language must be rich enough to provide possibilities of constructing a name for every sentence of the object-language.

In addition, the meta-language must obviously contain terms of a general logical character, such as the expression "if, and only if."[12]

It is desirable for the meta-language not to contain any undefined terms except such as are involved explicitly or implicitly in the remarks above, i.e.: terms of the object-language; terms referring to the form of the expressions of the object-language, and used in building names for these expressions; and terms of logic. In particular, we desire *semantic terms* (referring to the object-language) *to be introduced into the meta-language only by definition*. For, if this postulate is satisfied, the definition of truth, or of any other semantic concept, will fulfill what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal. And, moreover, we have then a kind of guarantee that the use of semantic concepts will not involve us in any contradictions.

We have no further requirements as to the formal structure of the object-language and the meta-language; we assume that it is similar to that of other formalized languages known at the present time. In particular, we assume that the usual formal rules of definition are observed in the meta-language.

10. Conditions for a Positive Solution of the Main Problem.

Now, we have already a clear idea both of the conditions of material adequacy to which the definition of truth is subjected, and of the formal structure of the language in which this definition is to be constructed. Under these circumstances the problem of the definition of truth acquires the character of a definite problem of a purely deductive nature.

The solution of the problem, however, is by no means obvious, and I would not attempt to give it in detail without using the whole machinery of contemporary logic. Here I shall confine myself to a rough outline of the solution and to the discussion of certain points of a more general interest which are involved in it.

The solution turns out to be sometimes positive, sometimes negative. This depends upon some formal relations between the object-language and its meta-language; or, more specifically, upon the fact whether the meta-language in its logical part is "*essentially richer*" than the object-language or not. It is not easy to give a general and precise definition of this notion of "essential richness." If we restrict ourselves to languages based on the logical theory of types, the condition for the meta-language to be "essentially richer" than the object-language is that it contain variables of a higher logical type than those of the object-language.

If the condition of "essential richness" is not satisfied, it can usually be shown that an interpretation of the meta-language in the object-language is possible; that is to say, with any given term of the meta-language a well-determined term of the object-language can be correlated in such a way that the assertible sentences of the one language turn out to be correlated with assertible sentences of the other. As a result of this interpretation, the hypothesis that a satisfactory definition of truth has been formulated in the meta-language turns out to imply the possibility of reconstructing in that language the antinomy of the liar; and this in turn forces us to reject the hypothesis in question.

(The fact that the meta-language, in its non-logical part, is ordinarily more comprehensive than the object-language does not affect the possibility of interpreting the former in the latter. For example, the names of expressions of the object-language occur in the meta-language, though for the most part they do not occur in the object-language itself; but, nevertheless, it may be possible to interpret these names in terms of the object-language.)

Thus we see that the condition of "essential richness" is necessary for the possibility of a satisfactory definition of truth in the meta-language. If we want to develop the theory of truth in a meta-language which does not satisfy this condition, we must give up the idea of defining truth with the exclusive help of those terms which were indicated above (in Section 8). We have then to include the term "*true*," or some other semantic term, in the list of undefined terms of the meta-language, and to express fundamental properties of the notion of truth in a series of axioms. There is nothing essentially wrong in such an axiomatic procedure, and it may prove useful for various purposes.[13]

It turns out, however, that this procedure can be avoided. *For the condition of the "essential richness" of the meta-language proves to be, not only necessary, but also sufficient for the construction of a satisfactory definition of truth*; i.e., if the meta-language satisfies this condition, the notion of truth can be defined in it. We shall now indicate in general terms how this construction can be carried through.

11. The Construction (in Outline) of the Definition.[14]

A definition of truth can be obtained in a very simple way from that of another semantic notion, namely, of the notion of *satisfaction*.

Satisfaction is a relation between arbitrary objects and certain expressions called "*sentential functions*." These are expressions like "*x is white*," "*x is greater than y*," etc. Their formal structure is analogous to that of sentences; however, they may contain the so-called free variables (like '*x*' and '*y*' in " *x is greater than y*"), which cannot occur in sentences.

In defining the notion of a sentential function in formalized languages, we usually apply what is called a "recursive procedure"; i.e., we first describe sentential functions of the simplest structure (which ordinarily presents no difficulty), and then we indicate the operations by means of which compound functions can be constructed from simpler ones. Such an operation may consist, for instance, in forming the logical disjunction or conjunction of two given functions, i.e., by combining them by the word "*or*" or "*and*." A sentence can now be defined simply as a sentential function which contains no free variables.

As regards the notion of satisfaction, we might try to define it by saying that given objects satisfy a given function if the latter becomes a true sentence when we replace in it free variables by names of given objects. In this sense, for example, snow satisfies the sentential function "*x is white*" since the sentence "*snow is white*" is true. However, apart from other difficulties, this method is not available to us, for we want to use the notion of satisfaction in defining truth.

To obtain a definition of satisfaction we have rather to apply again a recursive procedure. 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. Thus, for instance, we say that given numbers satisfy the logical disjunction "*x is greater than y or x is equal to y*" if they satisfy at least one of the functions "*x is greater than y*" or "*x is equal to y*."

Once the general definition of satisfaction is obtained, we notice that it applies automatically also to those special sentential functions which contain no free variables, i.e., to sentences. It turns out that for a sentence only two cases are possible: a sentence is either satisfied by all objects, or by no objects. Hence we arrive at a definition of truth and falsehood simply by saying that *a sentence is true if it is satisfied by all objects, and false otherwise*.[15]

(It may seem strange that we have chosen a roundabout way of defining the truth of a sentence, instead of trying to apply, for instance, a direct recursive procedure. The reason is that compound sentences are constructed from simpler sentential functions, but not always from simpler sentences; hence no general recursive method is known which applies specifically to sentences.)

From this rough outline it is not clear where and how the assumption of the "essential richness" of the meta-language is involved in the discussion; this becomes clear only when the construction is carried through in a detailed and formal way.[16]

12. Consequences of the Definition.

The definition of truth which was outlined above has many interesting consequences.

In the first place, the definition proves to be not only formally correct, but also materially adequate (in the sense established in Section 4); in other words, it implies all equivalences of the form (T). In this connection it is important to notice that the conditions for the material adequacy of the definition determine uniquely the extension of the term "true." Therefore, every definition of truth which is materially adequate would necessarily be equivalent to that actually constructed. The semantic conception of truth gives us, so to speak, no possibility of choice between various non-equivalent definitions of this notion.

Moreover, we can deduce from our definition various laws of a general nature. In particular, we can prove with its help the *laws of contradiction and of excluded middle*, which are so characteristic of the Aristotelian conception of truth; i.e., we can show that one and only one of any two contradictory sentences is true. These semantic laws should not be identified with the related logical laws of contradiction and excluded middle; the latter belong to the sentential calculus, i.e., to the most elementary part of logic, and do not involve the term "*true*" at all.

Further important results can be obtained by applying the theory of truth to formalized languages of a certain very comprehensive class of mathematical disciplines; only disciplines of an elementary character and a very elementary logical structure are excluded from this class. It turns out that for a discipline of this class *the notion of truth never coincides with that of provability*; for all provable sentences are true, but there are true sentences which are not provable.[17] Hence it follows further that every such discipline is consistent, but incomplete; that is to say, of any two contradictory sentences at most one is provable, and—what is more—there exists a pair of contradictory sentences neither of which is provable.[18]

13. Extension of the Results to Other Semantic Notions.

Most of the results at which we arrived in the preceding sections in discussing the notion of truth can be extended with appropriate changes to other semantic notions, for instance, to the notion of satisfaction (involved in our previous discussion), and to those of *designation* and *definition*.

Each of these notions can be analyzed along the lines followed in the analysis of truth. Thus, criteria for an adequate usage of these notions can be established; it can be shown that each of these notions, when used in a semantically closed language according to those criteria, leads necessarily to a contradiction;[19] a distinction between the object-language and the meta-language becomes again indispensable; and the "essential richness" of the meta-language proves in each case to be a necessary and sufficient condition for a satisfactory definition of the notion involved. Hence the results obtained in discussing one particular semantic notion apply to the general problem of the foundations of theoretical semantics.

Within theoretical semantics we can define and study some further notions, whose intuitive content is more involved and whose semantic origin is less obvious; we have in mind, for instance, the important notions of *consequence*, *synonymity*, and *meaning*.[20]

We have concerned ourselves here with the theory of semantic notions related to an individual object-language (although no specific properties of this language have been involved in our arguments). However, we could also consider the problem of developing *general semantics* which applies to a comprehensive class of object-languages. A considerable part of our previous remarks can be extended to this general problem; however, certain new difficulties arise in this connection, which will not be discussed here. I shall merely observe that the axiomatic method (mentioned in Section 10) may prove the most appropriate for the treatment of the problem.[21]

NOTES

1. Compare Tarski [2] (see bibliography at the end of the paper). This work may be consulted for a more detailed and formal presentation of the subject of the

paper, especially of the material included in Sections 6 and 9–13. It contains also references to my earlier publications on the problems of semantics (a communication in Polish, 1930; the article Tarski [1] in French, 1931; a communication in German, 1932; and a book in Polish, 1933). The expository part of the present paper is related in its character to Tarski [3]. My investigations on the notion of truth and on theoretical semantics have been reviewed or discussed in Hofstadter [1], Juhos [1], Kokoszyńska [1] and [2], Kotarbiński [2], Scholz [1], Weinberg [1], *et al.*

2. It may be hoped that the interest in theoretical semantics will now increase, as a result of the recent publication of the important work Carnap [2].

3. This applies, in particular, to public discussions during the 1st International Congress for the Unity of Science (Paris, 1935) and the Conference of International Congresses for the Unity of Science (Paris, 1937); cf., e.g., Neurath [1] and Gonseth [1].

4. The words "notion" and "concept" are used in this paper with all of the vagueness and ambiguity with which they occur in philosophical literature. Thus, sometimes they refer simply to a term, sometimes to what is meant by a term, and in other cases to what is denoted by a term. Sometimes it is irrelevant which of these interpretations is meant; and in certain cases perhaps none of them applies adequately. While on principle I share the tendency to avoid these words in any exact discussion, I did not consider it necessary to do so in this informal presentation.

5. For our present purposes it is somewhat more convenient to understand by "expressions," "sentences," etc., not individual inscriptions, but classes of inscriptions of similar form (thus, not individual physical things, but classes of such things).

6. For the Aristotelian formulation see Aristotle [1], (Γ), 7, 27. The other two formulations are very common in the literature, but I do not know with whom they originate. A critical discussion of various conceptions of truth can be found, e.g., in Kotarbiński [1] (so far available only in Polish), pp. 123 ff., and Russell [1], pp. 362 ff.

7. For most of the remarks contained in Sections 4 and 8, I am indebted to the late S. Leśniewski who developed them in his unpublished lectures in the University of Warsaw (in 1919 and later). However, Leśniewski did not anticipate the possibility of a rigorous development of the theory of truth, and still less of a definition of this notion; hence, while indicating equivalences of the form (T) as premisses in the antinomy of the liar, he did not conceive them as any sufficient conditions for an adequate usage (or definition) of the notion of truth. Also the remarks in Section 8 regarding the occurrence of an empirical premiss in the antinomy of the liar, and the possibility of eliminating this premiss, do not originate with him.

8. In connection with various logical and methodological problems involved in this paper the reader may consult Tarski [6].

9. The antinomy of the liar (ascribed to Eubulides or Epimenides) is discussed here in Sections 7 and 8. For the antinomy of definability (due to J. Richard) see e.g., Hilbert-Bernays [1], Vol. 2, pp. 263 ff.; for the antinomy of heterological terms see Grelling-Nelson [1], p. 307.

10. Due to Professor J. Łukasiewicz (University of Warsaw).

11. This can roughly be done in the following way. Let *S* be any sentence beginning with the words "*Every sentence.*" We correlate with *S* a new sentence *S** by subjecting *S* to the following two modifications: we replace in *S* the first word, "*Every*," by "*The*"; and we insert after the a second word, "*sentence*," the whole sentence *S* enclosed in quotation marks. Let us agree to call the sentence *S* "(self-) applicable" or "non-(self)-applicable" dependent on whether the correlated sentence *S** is true or false. Now consider the following sentence:

Every sentence is non-applicable.

It can easily be shown that the sentence just stated must be both applicable and non-applicable; hence a contradiction. It may not be quite clear in what sense this formulation of the antinomy does not involve an empirical premiss; however, I shall not elaborate on this point.

12. The terms "logic" and "logical" are used in this paper in a broad sense, which has become almost traditional in the last decades; logic is assumed here to comprehend the whole theory of classes and relations (i.e., the mathematical theory of sets). For many different reasons I am personally inclined to use the term "logic" in a much narrower sense, so as to apply it only to what is sometimes called "elementary logic," i.e., to the sentential calculus and the (restricted) predicate calculus.

13. Cf. here, however, Tarski [3], pp. 5f.

14. The method of construction we are going to outline can be applied—with appropriate changes—to all formalized languages that are known at the present time; although it does not follow that a language could not be constructed to which this method would not apply.

15. In carrying through this idea a certain technical difficulty arises. A sentential function may contain an arbitrary number of free variables; and the logical nature of the notion of satisfaction varies with this number. Thus, the notion in question when applied to functions with one variable is a binary relation between these functions and single objects; when applied to functions with two variables it becomes a ternary relation between functions and couples of objects; and so on. Hence, strictly speaking, we are confronted, not with one notion of satisfaction, but with infinitely many notions; and it turns out that these notions cannot be defined independently of each other, but must all be introduced simultaneously.

To overcome this difficulty, we employ the mathematical notion of an infinite sequence (or, possibly, of a finite sequence with an arbitrary number of terms). We agree to regard satisfaction, not as a many-termed relation between sentential functions and an indefinite number of objects, but as a binary relation between functions and sequences of objects. Under this assumption the formulation of a general and precise definition of satisfaction no longer presents any difficulty; and a true sentence can now be defined as one which is satisfied by every sequence.

16. To define recursively the notion of satisfaction, we have to apply a certain form of recursive definition which is not admitted in the object-language. Hence the "essential richness" of the meta-language may simply consist in admitting this type of definition. On the other hand, a general method is known which makes it possible to eliminate all recursive definitions and to replace them by normal, explicit ones. If we try to apply this method to the definition of satisfaction, we see that we have either to introduce into the meta-language variables of a higher logical type than those which occur in the object-language; or else to assume axiomatically in the meta-language the existence of classes that are more comprehensive than all those whose existence can be established in the object-language. See here Tarski [2], pp. 393 ff., and Tarski [5], p. 7.

17. Due to the development of modern logic, the notion of mathematical proof has undergone a far-reaching simplification. A sentence of a given formalized discipline is provable if it can be obtained from the axioms of this discipline by applying certain simple and purely formal rules of inference, such as those of detachment and substitution. Hence to show that all provable sentences are true, it suffices to prove that all the sentences accepted as axioms are true, and that the rules of inference when applied to true sentences yield new true sentences; and this usually presents no difficulty.

On the other hand, in view of the elementary nature of the notion of provability, a precise definition of this notion requires only rather simple logical devices. In most cases, those logical devices which are available in the formalized discipline itself (to which the notion of provability is related) are more than sufficient for this purpose. We know, however, that as regards the definition of truth just the opposite holds. Hence, as a rule, the notions of truth and provability cannot coincide; and since every provable sentence is true, there must be true sentences which are not provable.

18. Thus the theory of truth provides us with a general method for consistency proofs for formalized mathematical disciplines. It can be easily realized, however, that a consistency proof obtained by this method may possess some intuitive value—i.e., may convince us, or strengthen our belief, that the discipline under consideration is actually consistent—only in case we succeed in defining truth in terms of a meta-language which does not contain the object-language as a part (cf. here a remark in Section 9). For only in this case the deductive assumptions of the meta-language may be intuitively simpler and more obvious than those of the object-language—even though the condition of "essential richness" will be formally satisfied. Cf. here also Tarski [3], p. 7.

The incompleteness of a comprehensive class of formalized disciplines constitutes the essential content of a fundamental theorem of K. Gödel; cf. Gödel [1], pp. 187 ff. The explanation of the fact that the theory of truth leads so directly to Gödel's theorem is rather simple. In deriving Gödel's result from the theory of truth we make an essential use of the fact that the definition of truth cannot be given in a meta-language which is only as "rich" as the object-language (cf. note 17); however, in establishing this fact, a method of reasoning has been applied which is very closely related to that used (for the first time) by Gödel. It may be added that Gödel was clearly guided in his proof by certain intuitive considerations regarding the notion of truth, although this notion does not occur in the proof explicitly; cf. Gödel [1], pp. 174 f.

19. The notions of designation and definition lead respectively to the antinomies of Grelling-Nelson and Richard (cf. note 9). To obtain an antinomy for the notion of satisfaction, we construct the following expression:

The sentential function X does not satisfy X.

A contradiction arises when we consider the question whether this expression, which is clearly a sentential function, satisfies itself or not.

20. All notions mentioned in this section can be defined in terms of satisfaction. We can say, e.g., that a given term designates a given object if this object satisfies the sentential function "*x is identical with T*" where '*T*' stands for the given term. Similarly, a sentential function is said to define a given object if the latter is the only object which satisfies this function. For a definition of consequence see Tarski [4], and for that of synonymity—Carnap [2].

21. General semantics is the subject of Carnap [2]. Cf. here also remarks in Tarski [2], pp. 388 f.