VIII

THE CONCEPT OF TRUTH IN FORMALIZED LANGUAGES†

Introduction

The present article is almost wholly devoted to a single problem—*the definition of truth*. Its task is to construct—with reference to a given language—*a materially adequate and formally correct definition of the term 'true sentence'*. This problem, which belongs to the classical questions of philosophy, raises considerable difficulties. For although the meaning of the term 'true sentence' in colloquial language seems to be quite clear and intelligible, all attempts to define this meaning more precisely have hitherto been fruitless, and many investigations in which this term has been used and which started with apparently evident premisses have often led to paradoxes and antinomies (for which, however, a more or less satisfactory solution has been found). The concept of truth shares in this respect the fate of other analogous concepts in the domain of the semantics of language.

The question how a certain concept is to be defined is correctly formulated only if a list is given of the terms by means of which the required definition is to be constructed. If the definition is to fulfil its proper task, the sense of the terms in this list must admit of no doubt. The question thus naturally arises: What terms are we to use in constructing the definition of truth? In the course of these investigations I shall not neglect to clarify this question. In this construction I shall not make use

† Bibliographical Note. This article was presented (by J. Łukasiewicz) to the Warsaw Scientific Society on 21 March 1931. For reasons beyond the author's control, publication was delayed by two years. The article appeared in Polish in Tarski, A. (73). A German translation was published under the title 'Der Wahrheitsbegriff in den formalisierten Sprachen', in *Studia Philosophica*, vol. 1 (1936) (reprint dated 1935), pp. 261–405; it is provided with a Postscript in which some views which had been stated in the Polish original underwent a rather essential revision and modification. The present English version is based upon the German translation. For earlier publications and historical information concerning the results of this work see p. 154, footnote, p. 247, footnote, and the concluding remarks of the Postscript.

of any semantical concept if I am not able previously to reduce it to other concepts.

A thorough analysis of the meaning current in everyday life of the term 'true' is not intended here. Every reader possesses in greater or less degree an intuitive knowledge of the concept of truth and he can find detailed discussions on it in works on the theory of knowledge. I would only mention that throughout this work I shall be concerned exclusively with grasping the intentions which are contained in the so-called *classical* conception of truth ('true—corresponding with reality') in contrast, for example, with the *utilitarian* conception ('true—in a certain respect useful').[1]

The extension of the concept to be defined depends in an essential way on the particular language under consideration. The same expression can, in one language, be a true statement, in another a false one or a meaningless expression. There will be no question at all here of giving a single general definition of the term. The problem which interests us will be split into a series of separate problems each relating to a single language.

In § 1 colloquial language is the object of our investigations. The results are entirely negative. With respect to this language not only does the definition of truth seem to be impossible, but even the consistent use of this concept in conformity with the laws of logic.

In the further course of this discussion I shall consider exclusively the scientifically constructed languages known at the present day, i.e. the formalized languages of the deductive sciences. Their characteristics will be described at the beginning of § 2. It will be found that, from the standpoint of the present problem, these languages fall into two groups, the division being based on the greater or less stock of grammatical forms in a particular language. In connexion with the 'poorer' languages the problem of the definition of truth has a positive solution: there is a uniform method for the construction of the required

[1] Cf. Kotarbiński, T. (37), p. 126 (in writing the present article I have repeatedly consulted this book and in many points adhered to the terminology there suggested).

definition in the case of each of these languages. In §§ 2 and 3 I shall carry out this construction for a concrete language in full and in this way facilitate the general description of the above method which is sketched in § 4. In connexion with the 'richer' languages, however, the solution of our problem will be negative, as will follow from the considerations of § 5. For the languages of this group we shall never be able to construct a correct definition of the notion of truth.† Nevertheless, everything points to the possibility even in these cases—in contrast to the language of everyday life—of introducing a consistent and correct use of this concept by considering it as a primitive notion of a special science, namely of the theory of truth, and its fundamental properties are made precise through axiomatization.

The investigation of formalized languages naturally demands a knowledge of the principles of modern formal logic. For the construction of the definition of truth certain purely mathematical concepts and methods are necessary, although in a modest degree. I should be happy if this work were to convince the reader that these methods already are necessary tools even for the investigation of purely philosophical problems.[1]

§ 1. The Concept of True Sentence in Everyday or Colloquial Language

For the purpose of introducing the reader to our subject, a consideration—if only a fleeting one—of the problem of defining truth in colloquial language seems desirable. I wish especially

[1] This was communicated to the Society of Sciences in Warsaw by J. Łukasiewicz on 21 March 1931. The results it contains date for the most part from 1929. I have reported on this, among other things, in two lectures which I gave under the title 'On the Concept of Truth in relation to formalized deductive systems' at the logical section of the Philosophical Society in Warsaw (8 October 1930) and at the Polish Philosophical Society in Lwów (15 December 1930), a résumé of which appeared in Tarski, A. (73). For reasons unconnected with me the printing of this work was much delayed. This enabled me to supplement the text with some rather important results (cf. p. 247, footnote). In the meantime a résumé of the chief results has appeared in Tarski, A. (76).

† Regarding this statement compare the Postscript.

to emphasize the various difficulties which the attempts to solve this problem have encountered.[1]

Amongst the manifold efforts which the construction of a correct definition of truth for the sentences of colloquial language has called forth, perhaps the most natural is the search for a *semantical definition*. By this I mean a definition which we can express in the following words:

(1) *a true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.*[2]

From the point of view of formal correctness, clarity, and freedom from ambiguity of the expressions occurring in it, the above formulation obviously leaves much to be desired. Nevertheless its intuitive meaning and general intention seem to be quite clear and intelligible. To make this intention more definite, and to give it a correct form, is precisely the task of a semantical definition.

As a starting-point certain sentences of a special kind present themselves which could serve as partial definitions of the truth of a sentence or more correctly as explanations of various concrete turns of speech of the type 'x is a true sentence'. The general scheme of this kind of sentence can be depicted in the following way:

(2) *x is a true sentence if and only if p.*

In order to obtain concrete definitions we substitute in the

[1] The considerations which I shall put forward in this connexion are, for the most part, not the result of my own studies. Views are expressed in them which have been developed by S. Leśniewski in his lectures at the University of Warsaw (from the year 1919/20 onwards), in scientific discussions and in private conversations; this applies, in particular, to almost everything which I shall say about expressions in quotation marks and the semantical antinomies. It remains perhaps to add that this fact does not in the least involve Leśniewski in the responsibility for the sketchy and perhaps not quite precise form in which the following remarks are presented.

[2] Very similar formulations are found in Kotarbiński, T. (37), pp. 127 and 136, where they are treated as commentaries which explain approximately the classical view of truth.

Of course these formulations are not essentially new; compare, for example, the well-known words of Aristotle: '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.' (Aristotle, *Metaphysica*, Γ, 7, 27; *Works*, vol. 8, English translation by W. D. Ross, Oxford, 1908.)

place of the symbol '*p*' in this scheme any sentence, and in the place of '*x*' any individual name of this sentence.

If we are given a name for a sentence, we can construct an explanation of type (2) for it, provided only that we are able to write down the sentence denoted by this name. The most important and common names for which the above condition is satisfied are the so-called *quotation-mark names*. We denote by this term every name of a sentence (or of any other, even meaningless, expression) which consists of quotation marks, left- and right-hand, and the expression which lies between them, and which (expression) is the object denoted by the name in question. As an example of such a name of a sentence the name "it is snowing" will serve. In this case the corresponding explanation of type (2) is as follows:

(3) *'it is snowing' is a true sentence if and only if it is snowing.*[1]

Another category of names of sentences for which we can construct analogous explanations is provided by the so-called *structural-descriptive names*. We shall apply this term to names which describe the words which compose the expression denoted by the name, as well as the signs of which each single word is composed and the order in which these signs and words follow one another. Such names can be formulated without the help of quotation marks. For this purpose we must have, in the language we are using (in this case colloquial language), individual names of some sort, but not quotation-mark names, for all letters and all other signs of which the words and expressions of the language are composed. For example we could use '*A*', '*E*', '*Ef*', '*Jay*', '*Pe*' as names of the letters '*a*', '*e*', '*f*', '*j*', '*p*'. It is clear that we can correlate a structural-descriptive name with every quotation-mark name, one which is free from quotation marks and possesses the same extension (i.e. denotes the same expression) and vice versa. For example, corresponding to the name "'snow'" we have the name 'a word which consists of the four letters: *Es*, *En*, *O*, Double-*U* following one another'. It is thus clear that we can construct partial definitions of the type (2) for structual-descriptive names of sentences. This is illustrated by the following example:

(4) *an expression consisting of three words of which the first is composed of the two letters I and Te (in that order) the second of the two letters I and Es (in that order) and the third of the seven letters Es, En, O, Double-U, I, En, and Ge (in that order), is a true sentence if and only if it is snowing.*

Sentences which are analogous to (3) and (4) seem to be clear and completely in accordance with the meaning of the word 'true' which was expressed in the formulation (1). In regard to the clarity of their content and the correctness of their form they arouse, in general, no doubt (assuming of course that no such doubts are involved in the sentences which we substitute for the symbol '*p*' in (2)).

But a certain reservation is nonetheless necessary here. Situations are known in which assertions of just this type, in combination with certain other not less intuitively clear premisses, lead to obvious contradictions, for example the *antinomy of the liar*. We shall give an extremely simple formulation of this antinomy which is due to J. Łukasiewicz.

[1] Statements (sentences) are always treated here as a particular kind of expression, and thus as linguistic entities. Nevertheless, when the terms 'expression', 'statement', etc., are interpreted as names of concrete series of printed signs, various formulations which occur in this work do not appear to be quite correct, and give the appearance of a widespread error which consists in identifying expressions of like shape. This applies especially to the sentence (3), since with the above interpretation quotation-mark names must be regarded as general (and not individual) names, which denote not only the series of signs in the quotation marks but also every series of signs of like shape. In order to avoid both objections of this kind and also the introduction of superfluous complications into the discussion, which would be connected among other things with the necessity of using the concept of likeness of shape, it is convenient to stipulate that terms like 'word', 'expression', 'sentence', etc., do not denote concrete series of signs but whole classes of such series which are of like shape with the series given; only in this sense shall we regard quotation-mark names as individual names of expressions. Cf. Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 661–6 and—for other interpretations of the term 'sentence'—Kotarbiński, T. (37), pp. 123–5.

I take this opportunity of mentioning that I use the words 'name' and 'denote' (like the words 'object', 'class', 'relation') not in *one*, but in many distinct senses, because I apply them both to objects in the narrower sense (i.e. to individuals) and also to all kinds of classes and relations, etc. From the standpoint of the theory of types expounded in Whitehead, A. N., and Russell, B. A. W. (90) (vol. 1, pp. 139–68) these expressions are to be regarded as systematically ambiguous.

For the sake of greater perspicuity we shall use the symbol '*c*' as a typographical abbreviation of the expression 'the sentence printed on this page, line 5 from the top'. Consider now the following sentence:

c is not a true sentence.

Having regard to the meaning of the symbol '*c*', we can establish empirically:

(α) *'c is not a true sentence' is identical with c.*

For the quotation-mark name of the sentence *c* (or for any other of its names) we set up an explanation of type (2):

(β) *'c is not a true sentence' is a true sentence if and only if c is not a true sentence.*

The premisses (α) and (β) together at once give a contradiction:

c is a true sentence if and only if c is not a true sentence.

The source of this contradiction is easily revealed: in order to construct the assertion (β) we have substituted for the symbol '*p*' in the scheme (2) an expression which itself contains the term 'true sentence' (whence the assertion so obtained—in contrast to (3) or (4)—can no longer serve as a partial definition of truth). Nevertheless no rational ground can be given why such substitutions should be forbidden in principle.

I shall restrict myself here to the formulation of the above antinomy and will postpone drawing the necessary consequences of this fact till later. Leaving this difficulty on one side I shall next try to construct a definition of true sentence by generalizing explanations of type (3). At first sight this task may seem quite easy—especially for anyone who has to some extent mastered the technique of modern mathematical logic. It might be thought that all we need do is to substitute in (3) any sentential variable (i.e. a symbol for which any sentence can be substituted) in place of the expression 'it is snowing' which occurs there twice, and then to establish that the resulting formula holds for every value of the variable, and thus without

further difficulty reach a sentence which includes all assertions of type (3) as special cases:

(5) *for all p, 'p' is a true sentence if and only if p.*

But the above sentence could not serve as a general definition of the expression '*x* is a true sentence' because the totality of possible substitutions for the symbol '*x*' is here restricted to quotation-mark names. In order to remove this restriction we must have recourse to the well-known fact that to every true sentence (and generally speaking to every sentence) there corresponds a quotation-mark name which denotes just that sentence.[1] With this fact in mind we could try to generalize the formulation (5), for example, in the following way:

(6) *for all x, x is a true sentence if and only if, for a certain p, x is identical with 'p' and p.*

At first sight we should perhaps be inclined to regard (6) as a correct semantical definition of 'true sentence', which realizes in a precise way the intention of the formulation (1) and therefore to accept it as a satisfactory solution of our problem. Nevertheless the matter is not quite so simple. As soon as we begin to analyse the significance of the quotation-mark names which occur in (5) and (6) we encounter a series of difficulties and dangers.

Quotation-mark names may be treated like single words of a language, and thus like syntactically simple expressions. The single constituents of these names—the quotation marks and the expressions standing between them—fulfil the same function as the letters and complexes of successive letters in single words. Hence they can possess no independent meaning. Every quotation-mark name is then a constant individual name of a definite expression (the expression enclosed by the quotation marks) and in fact a name of the same nature as the proper name of a man. For example, the name "*p*" denotes one

[1] For example, this fact could be formulated in the following way:

(5′) *for all x, if x is a true sentence, then—for a certain p—x is identical with 'p';*

from the premisses (5) and (5′) the sentence (6) given below can be derived as a conclusion.

of the letters of the alphabet. With this interpretation, which seems to be the most natural one and completely in accordance with the customary way of using quotation marks, partial definitions of the type (3) cannot be used for any significant generalizations. In no case can the sentences (5) or (6) be accepted as such a generalization. In applying the rule called the rule of substitution to (5) we are not justified in substituting anything at all for the letter '*p*' which occurs as a component part of a quotation-mark name (just as we are not permitted to substitute anything for the letter '*t*' in the word '*true*'). Consequently we obtain as conclusion not (5) but the following sentence: '*p*' *is a true sentence if and only if it is snowing*. We see at once from this that the sentences (5) and (6) are not formulations of the thought we wish to express and that they are in fact obviously senseless. Moreover, the sentence (5) leads at once to a contradiction, for we can obtain from it just as easily in addition to the above given consequence, the contradictory consequence: '*p*' *is a true sentence if and only if it is not snowing*. Sentence (6) alone leads to no contradiction, but the obviously senseless conclusion follows from it that the letter '*p*' is the only true sentence.

To give greater clarity to the above considerations it may be pointed out that with our conception of quotation-mark names they can be eliminated and replaced everywhere by, for example, the corresponding structural-descriptive names. If, nevertheless, we consider explanations of type (2) constructed by the use of such names (as was done, for example, in (4) above), then we see no way of generalizing these explanations. And if in (5) or (6) we replace the quotation-mark name by the structural-descriptive name '*pe*' (or '*the word which consists of the single letter Pe*') we see at once the absurdity of the resulting formulation.

In order to rescue the sense of sentences (5) and (6) we must seek quite a different interpretation of the quotation-mark names. We must treat these names as syntactically composite expressions, of which both the quotation marks and the expressions within them are parts. Not all quotation-mark

expressions will be constant names in that case. The expression "'*p*'" occurring in (5) and (6), for example, must be regarded as a function, the argument of which is a sentential variable and the values of which are constant quotation-mark names of sentences. We shall call such functions *quotation-functions*. The quotation marks then become independent words belonging to the domain of semantics, approximating in their meaning to the word 'name', and from the syntactical point of view, they play the part of functors.[1] But then new complications arise. The sense of the quotation-function and of the quotation marks themselves is not sufficiently clear. In any case such functors are not extensional; there is no doubt that the sentence "*for all p and q, if p if and only if q, then* '*p*' *is identical with* '*q*'" is in palpable contradiction to the customary way of using quotation marks. For this reason alone definition (6) would be unacceptable to anyone who wishes consistently to avoid intensional functors and is even of the opinion that a deeper analysis shows it to be impossible to give any precise meaning to such functors.[2] Moreover, the use of the quotation functor exposes us to the danger of becoming involved in various semantical antinomies, such as the antinomy of the liar. This will be so even if—taking every care—we make use only of those properties of quotation-functions which seem almost evident. In contrast to that conception of the antimony of the liar which has been given above, we can formulate it without using the expression 'true sentence' at all, by introducing the

[1] We call such words as 'reads' in the expression '*x* reads' functors (this is a sentence-forming functor with *one* individual name as argument); also 'sees' in the expression '*x* sees *y*' (a sentence-forming functor with *two* name arguments), and 'father' in the expression 'the father of *x*' (a name-forming functor with *one* name argument), as well as 'or' in the expression '*p* or *q*' (a sentence-forming functor with two sentence arguments); quotation marks provide an example of a name-forming functor with *one* sentence argument. The term 'functor' we owe to T. Kotarbiński, the terms 'sentence-forming functor' and 'name-forming functor' to K. Ajdukiewicz; cf. Ajdukiewicz, K. (3).

[2] I shall not discuss the difficult problem of extensionality in more detail here; cf. Carnap, R. (8) where the literature of the problem is given, and especially Whitehead, A. N., and Russell, B. A. W. (90), vol. I, pp. 659–66. It should be noted that usually the terms 'extensional' and 'intensional' are applied to sentence-forming functors, whilst in the text they are applied to quotation marks and thus to name-forming functors.

quotation-functions with variable arguments. We shall give a sketch of this formulation.

Let the symbol 'c' be a typographical abbreviation of the expression '*the sentence printed on this page, line* 6 *from the top*'. We consider the following statement:

for all p, if c is identical with the sentence 'p', then not p

(if we accept (6) as a definition of truth, then the above statement asserts that c is not a true sentence).

We establish empirically:

(α) *the sentence 'for all p, if c is identical with the sentence 'p', then not p' is identical with c.*

In addition we make only a single supplementary assumption which concerns the quotation-function and seems to raise no doubts:

(β) *for all p and q, if the sentence 'p' is identical with the sentence 'q', then p if and only if q.*

By means of elementary logical laws we easily derive a contradiction from the premisses (α) and (β).

I should like to draw attention, in passing, to other dangers to which the consistent use of the above interpretation of quotation marks exposes us, namely to the ambiguity of certain expressions (for example, the quotation-expression which occurs in (5) and (6) must be regarded in certain situations as a function with variable argument, whereas in others it is a constant name which denotes a letter of the alphabet). Further, I would point out the necessity of admitting certain linguistic constructions whose agreement with the fundamental laws of syntax is at least doubtful, e.g. meaningful expressions which contain meaningless expressions as syntactical parts (every quotation-name of a meaningless expression will serve as an example). For all these reasons the correctness of definition (6), even with the new interpretation of quotation marks, seems to be extremely doubtful.

Our discussions so far entitle us in any case to say that *the attempt to construct a correct semantical definition of the expression 'true sentence meets with very real difficulties.* We know of no

general method which would permit us to define the meaning of an arbitrary concrete expression of the type 'x is a true sentence', where in the place of 'x' we have a name of some sentence. The method illustrated by the examples (3) and (4) fails us in those situations in which we cannot indicate for a given name of a sentence, the sentence denoted by this name (as an example of such a name 'the first sentence which will be printed in the year 2000' will serve). But if in such a case we seek refuge in the construction used in the formulation of definition (6), then we should lay ourselves open to all the complications which have been described above.

In the face of these facts we are driven to seek other methods of solving our problem. I will draw attention here to only *one* such attempt, namely the attempt to construct a *structural definition*. The general scheme of this definition would be somewhat as follows: *a true sentence is a sentence which possesses such and such structural properties* (i.e. properties concerning the form and arrangement in sequence of the single parts of the expression) *or which can be obtained from such and such structurally described expressions by means of such and such structural transformations.* As a starting-point we can press into service many laws from formal logic which enable us to infer the truth or falsehood of sentences from certain of their structural properties; or from the truth or falsehood of certain sentences to infer analogous properties of other sentences which can be obtained from the former by means of various structural transformations. Here are some trivial examples of such laws: *every expression consisting of four parts of which the first is the word 'if', the third is the word 'then', and the second and fourth are the same sentence, is a true sentence; if a true sentence consists of four parts, of which the first is the word 'if', the second a true sentence, the third the word 'then', then the fourth part is a true sentence.* Such laws (especially those of the second type) are very important. With their help every fragmentary definition of truth, the extension of which embraces an arbitrary class of sentences, can be extended to all composite sentences which can be built up from sentences of the given class by combining them by means

of such expressions as 'if . . . then', 'if and only if', 'or', 'and', 'not', in short, by means of expressions belonging to the sentential calculus (or theory of deduction). This leads to the idea of setting up sufficiently numerous, powerful, and general laws for every sentence to fall under one of them. In this way we should reach a general structural definition of a true sentence. Yet this way also seems to be almost hopeless, at least as far as natural language is concerned. For this language is not something finished, closed, or bounded by clear limits. It is not laid down what words can be added to this language and thus in a certain sense already belong to it potentially. We are not able to specify structurally those expressions of the language which we call sentences, still less can we distinguish among them the true ones. *The attempt to set up a structural definition of the term 'true sentence'—applicable to colloquial language is confronted with insuperable difficulties.*

The breakdown of all previous attempts leads us to suppose that there is no satisfactory way of solving our problem. Important arguments of a general nature can in fact be invoked in support of this supposition as I shall now briefly indicate.

A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language'. If we are to maintain this universality of everyday language in connexion with semantical investigations, we must, to be consistent, admit into the language, in addition to its sentences and other expressions, also the names of these sentences and expressions, and sentences containing these names, as well as such semantic expressions as 'true sentence', 'name', denote', etc. But it is presumably just this universality of everyday language which is the primary source of all semantical antinomies, like the antinomies of the liar or of heterological words. These antinomies seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of

logic hold, must be inconsistent. This applies especially to the formulation of the antinomy of the liar which I have given on pages 157 and 158, and which contains no quotation-function with variable argument. If we analyse this antinomy in the above formulation we reach the conviction that no consistent language can exist for which the usual laws of logic hold and which at the same time satisfies the following conditions: (I) for any sentence which occurs in the language a definite name of this sentence also belongs to the language; (II) every expression formed from (2) by replacing the symbol 'p' by any sentence of the language and the symbol 'x' by a name of this sentence is to be regarded as a true sentence of this language; (III) in the language in question an empirically established premiss having the same meaning as (α) can be formulated and accepted as a true sentence.[1]

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

§ 2. Formalized Languages, especially the Language of the Calculus of Classes

For the reasons given in the preceding section I now abandon the attempt to solve our problem for the language of everyday life and restrict myself henceforth entirely to *formalized languages*.[2] These can be roughly characterized as artificially con-

[1] The antinomy of heterological words (which I shall not describe here—cf. Grelling, K., and Nelson, L. (24), p. 307) is simpler than the antinomy of the liar in so far as no empirical premiss analogous to (α) appears in its formulation; thus it leads to the correspondingly stronger consequence: there can be no consistent language which contains the ordinary laws of logic and satisfies two conditions which are analogous to (I) and (II), but differ from them in that they treat not of sentences but of names, and not of the truth of sentences but of the relation of denoting. In this connexion compare the discussion in § 5 of the present article—the beginning of the proof of Th. 1, and in particular p. 248, footnote 2.

[2] The results obtained for formalized language also have a certain validity for colloquial language, and this is owing to its universality: if we translate into colloquial language any definition of a true sentence which has been constructed for some formalized language, we obtain a fragmentary definition of truth which embraces a wider or narrower category of sentences.

structed languages in which the sense of every expression is unambiguously determined by its form. Without attempting a completely exhaustive and precise description, which is a matter of considerable difficulty, I shall draw attention here to some essential properties which all the formalized languages possess: (α) for each of these languages a list or description is given in structural terms of all the *signs with which the expressions of the language are formed*; (β) among all possible expressions which can be formed with these signs those called *sentences* are distinguished by means of purely structural properties. Now formalized languages have hitherto been constructed exclusively for the purpose of studying *deductive sciences* formalized on the basis of such languages. The language and the science grow together to a single whole, so that we speak of the language of a particular formalized deductive science, instead of this or that formalized language. For this reason further characteristic properties of formalized languages appear in connexion with the way in which deductive sciences are built up; (γ) a list, or structural description, is given of the sentences called *axioms* or *primitive statements*; (δ) in special rules, called *rules of inference*, certain operations of a structural kind are embodied which permit the transformation of sentences into other sentences; the sentences which can be obtained from given sentences by one or more applications of these operations are called *consequences* of the given sentences. In particular the consequences of the axioms are called *provable* or *asserted sentences*.[1]

It remains perhaps to add that we are not interested here in 'formal' languages and sciences in one special sense of the word 'formal', namely sciences to the signs and expressions of which no material sense is attached. For such sciences the problem here discussed has no relevance, it is not even meaningful.

[1] The formalization of a science usually admits of the possibility of introducing new signs into that science which were not explicitly given at the outset. These signs—called *defined signs* (in contrast to the *primitive* signs)—appear in the science in the first instance in expressions of a special structure called *definitions*, which are constructed in accordance with special rules—the *rules of definition*. Definitions are sometimes regarded as asserted sentences of the science. This feature of the formalization of languages will not be considered in the sequel.

We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the languages we shall consider.[1] The expressions which we call sentences still remain sentences after the signs which occur in them have been translated into colloquial language. The sentences which are distinguished as axioms seem to us to be materially true, and in choosing rules of inference we are always guided by the principle that when such rules are applied to true sentences the sentences obtained by their use should also be true.[2]

In contrast to natural languages, the formalized languages do not have the universality which was discussed at the end of the preceding section. In particular, most of these languages possess no terms belonging to the theory of language, i.e. no expressions which denote signs and expressions of the same or another language or which describe the structural connexions between them (such expressions I call—for lack of a better term—*structural-descriptive*). For this reason, when we investigate the language of a formalized deductive science, we must always distinguish clearly between the language *about* which we speak and the language *in* which we speak, as well as between the science which is the object of our investigation and the science in which the investigation is carried out. The names of the expressions of the first language, and of the relations between them, belong to the second language, called the *metalanguage* (which may contain the first as a part). The description of these expressions, the definition of the complicated concepts, especially of those connected with the construction of a deductive theory (like the concept of consequence, of provable sentence, possibly of true sentence), the determination of the properties of these concepts, is the task of the second theory which we shall call the *metatheory*.

For an extensive group of formalized languages it is possible

[1] Strictly speaking this applies only to the signs called constants. Variables and technical signs (such as brackets, dots, etc.) possess no independent meaning; but they exert an essential influence on the meaning of the expressions of which they form parts.

[2] Finally, the definitions are so constructed that they elucidate or determine the meaning of the signs which are introduced into the language by means of primitive signs or signs previously defined (cf. p. 166, note 1).

to give a method by which a correct definition of truth can be constructed for each of them. The general abstract description of this method and of the languages to which it is applicable would be troublesome and not at all perspicuous. I prefer therefore to introduce the reader to this method in another way. I shall construct a definition of this kind in connexion with a particular concrete language and show some of its most important consequences. The indications which I shall then give in § 4 of this article will, I hope, be sufficient to show how the method illustrated by this example can be applied to other languages of similar logical construction.

I choose, as the object of my considerations, the language of a deductive science of the utmost simplicity which will surely be well known to the reader—that of the *calculus of classes*. The calculus of classes is a fragment of mathematical logic and can be regarded as one of the interpretations of a formal science which is commonly called the *algebra of logic*.[1]

Among the signs comprising the expressions of this language I distinguish two kinds, *constants* and *variables*.[2] I introduce only four constants: the *negation* sign 'N', the sign of *logical sum* (*disjunction*) 'A', the *universal quantifier* 'Π', and finally the *inclusion* sign 'I'.[3] I regard these signs as being equivalent in

[1] Cf. Schröder, E. (62), vol. 1 (especially pp. 160–3) and Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 205–12.

[2] By making use of an idea of Łukasiewicz I avoid introducing any technical signs (like brackets, dots, etc.) into the language, and this is due chiefly to the fact that I always write the functor before the arguments in every meaningful expression; cf. Łukasiewicz, J. (51), especially pp. v and 40.

[3] Usually many other constants occur in the calculus of classes, e.g. the existence sign, the sign of implication, of logical product (conjunction), of equivalence, of identity, as well as of the complement, the sum, and the product of classes (see p. 168, note 1); for that reason only a fragment of the calculus of classes can—formally speaking—be constructed in the language under consideration. It is, however, to be noted that all constants of the calculus of classes could be introduced into this language as defined terms, if we complete its formalization by making the introduction of new signs possible by means of definitions (see p. 166, note 1). Owing to this fact our fragmentary language already suffices for the expression of every idea which can be formulated in the complete language of this science. I would also point out that even the sign of inclusion 'I' can be eliminated from our language by interpreting expressions of the type 'xy' (where any variables occur in the place of 'x' and 'y') in the same way in which in the sequel we shall interpret the expression 'Ixy'.

meaning respectively with the expressions 'not', 'or', 'for all' (in the sense in which this expression was used in statement (6) of § 1, for example) and 'is included in'. In principle any arbitrary symbols could be used as variables, provided only that their number is not limited and that they are distinct in form from the constants. But for the further course of our work it is technically important to specify the form of these signs exactly, and in such a way that they can easily be ordered in a sequence. I shall therefore use as variables only such symbols as 'x_1', 'x_2', 'x_{13}', and analogous signs which consist of the symbol 'x' and a number of small strokes added below. The sign which has k such small strokes (k being any natural number distinct from 0) will be called *the k-th variable*. In the intuitive interpretation of the language, which I always have in mind here, the variables represent names of classes of individuals. As *expressions* of the language we have either single constants and variables or complexes of such signs following one another, for example: 'x_1Nx_2', 'NIx_1x_2', '$AIx_1x_2Ix_2x_1$', 'Πx_1', '$\Pi x_1Ix_2x_{13}$', 'Ix_2x_{13}' and so on. Expressions of the type 'Np', 'Apq', 'Πxp', and 'Ixy', where in the place of 'p' and 'q' any sentences or sentential functions (this term will be explained below), and in the place of 'x' and 'y' any variables, appear, are read: 'not p' (or 'it is not true that p'),[1] 'p or q', 'for all classes x we have p', and 'the class x is included in the class y', respectively. Regarding composite expressions, i.e. those which are not signs, we can say that they consist of two or more other simple expressions. Thus the expression 'NIx_1x_2' is composed of the two successive expressions 'N' and 'Ix_1x_2' or of the expressions 'NI' and 'x_1x_2' or finally of the expressions 'NIx_1' and 'x_2'.

But the proper domain of the following considerations is not the language of the calculus of classes itself but the corresponding metalanguage. Our investigations belong to the *metacalculus of classes* developed in this metalanguage. From this springs the need to give the reader some account—if only a very brief

[1] For stylistic reasons we sometimes use the expression 'it is not true that' instead of the word 'not', the whole expression being regarded as a single word, no independent meaning being given to the separate parts, and in particular to the word 'true', which occur in it.

one—of the structure of the metalanguage and of the metatheory. I shall restrict myself to the two most important points: (1) the enumeration of all the signs and expressions which will be used in the metalanguage, without explaining in more detail their importance in the course of the investigation, and (2) the setting up of a system of axioms which suffices for the establishment of the metatheory or at least will form a foundation for the results obtained in this article. These two points are closely connected with our fundamental problem; were we to neglect them, we should not be able to assert either that we had succeeded in correctly defining any concept on the basis of the metalanguage, or that the definition constructed possesses any particular consequences. But I shall not attempt at all to give the metatheory the character of a strictly formalized deductive science. I shall content myself with saying that—apart from the two points mentioned—the process of formalizing the metatheory shows no specific peculiarity. In particular, the rules of inference and of definition do not differ at all from the rules used in constructing other formalized deductive sciences.

Among the expressions of the metalanguage we can distinguish two kinds. To the first belong *expressions of a general logical character*, obtainable from any sufficiently developed system of mathematical logic.[1] They can be divided into primitive expressions and defined expressions, but this would be pointless in the present case. First we have a series of expressions which have the same meaning as the constants of the science we are considering; thus '*not*' or '*it is not true that*',[2] '*or*', '*for all*', and '*is included in*'—in symbols '$\subseteq$'. Thanks to this circumstance we are able to translate every expression of the language into the metalanguage. For example, the statement 'for all a (or for all classes a) $a \subseteq a$' is the translation of the expression '$\prod x, Ix, x$'. To the same category belongs a series of analogous

[1] For example, from the work Whitehead, A. N., and Russell, B. A. W. (90). (But I do not intend to use here any special logical symbolism. Apart from the exceptions which I shall explicitly mention I shall use expressions of colloquial language.) For the meaning of the general logical expressions given below see Carnap, R. (8).

[2] See p. 169, note 1.

expressions from the domain of the sentential calculus, of the first order functional calculus and of the calculus of classes, for example, '*if . . ., then*', '*and*', '*if and only if*', '*for some x*' (or '*there is an x such that . . .*'), '*is not included in*'—in symbols '$\not\subseteq$', '*is identical with*'—in symbols '$=$', '*is distinct from*'—in symbols '$\neq$', '*is an element of*'—in symbols '$\in$', '*is not an element of*'—in symbols '$\notin$', '*individual*', '*class*', '*null class*', '*class of all x such that*', and so on. We also find here some expressions from the domain of the theory of the equivalence of classes, and of the arithmetic of cardinal numbers, e.g. '*finite class*', '*infinite class*', '*power of a class*', '*cardinal number*', '*natural number*' (or '*finite cardinal number*'), '*infinite cardinal number*', '0', '1', '2', '$<$', '$>$', '$\leqslant$', '$\geqslant$', '$+$', '$-$', ... Finally I shall need some terms from the logic of relations. The class of all objects x, to which there corresponds at least *one* object y such that xRy (i.e. x stands in the relation R to y) will be called the *domain of the binary or two-termed relation* R. Analogously, the *counter domain of the relation* R is the set of all objects y for which there is at least *one* object x such that xRy. In the case of many-termed relations we do not speak of domain and counter domain, but of the *1st, 2nd, 3rd,..., n-th domain of the relation*. The relation having *only one* element x in its domain and *only one* element y in its counter domain (a relation which thus holds only between x and y and between no other two objects) is called an *ordered pair, where x is the first and y the second member*. Analogously using many-termed relations we define *ordered triples, quadruples*, and in general *ordered n-tuples*. If, for every object y belonging to the counter domain of a two-termed relation R, there is *only one* object x such that xRy, then the relation R is called *one-many*. The concept of *sequence* will play a great part in the sequel. An *infinite sequence* is a one-many relation whose counter domain is the class of all natural numbers excluding zero. In the same way, the term '*finite sequence of n terms*' denotes every one-many relation whose counter domain consists of all natural numbers k such that $1 \leqslant k \leqslant n$ (where n is any natural number distinct from 0). The unique x which satisfies the formula xRk (for a given sequence R and a given natural number k) is called the

k-th term of the sequence R, or the *term of the sequence R with index k*, and is denoted by 'R_k'. We say that *the sequences R and S differ in at most the k-th place*, if any two corresponding terms of these sequences R_l and S_l are identical with the exception of the kth terms R_k and S_k which may be distinct. In the following pages we shall deal with sequences of classes and of natural numbers, i.e. with sequences all of whose terms are either classes of individuals or natural numbers. In particular, a sequence all of whose terms are classes which are included in a given class a, will be called a *sequence of subclasses of the class a*.

In contrast to the first kind of expression, those of the second kind are *specific terms of the metalanguage of a structural-descriptive character*, and thus names of concrete signs or expressions of the language of the calculus of classes. Among these are, in the first place, the terms *'the negation sign'*, *'the sign of logical sum'*, *'the sign of the universal quantifier'*, *'the inclusion sign'*, *'the k-th variable'*, *'the expression which consists of the expressions x and y following one another'* and *'expression'*. As abbreviations of the first six terms I shall use the symbols '*ng*', '*sm*', '*un*', '*in*', 'v_k', and '$x^\frown y$' (the sign 'v' thus denotes a sequence, the terms of which are the successive variables $v_1, v_2, v_3, \ldots$). These terms have already been used in introducing the reader to the language of the calculus of classes. I hope that, thanks to the explanations already given, no doubt will remain concerning the meaning of these terms. With the help of these terms (and possibly general logical terms) all other concepts of the metalanguage of a structural-descriptive kind can be defined. It is easy to see that every simple or composite expression of the language under investigation has an individual name in the metalanguage similar to the structural-descriptive names of colloquial language (cf. pp. 156 and 157). For example, the symbolic expression '$((ng^\frown in)^\frown v_1)^\frown v_2$' can serve as a name of the expression '$NIx_1 x_n$'. The fact that the metalanguage contains both an individual name and a translation of every expression (and in particular of every sentence) of the language studied will play a decisive part in the construction of the definition of truth, as the reader will see in the next section.

As variables in the metalanguage I shall use the symbols (1) 'a', 'b'; (2) 'f', 'g', 'h'; (3) 'k', 'l', 'm', 'n', 'p'; (4) 't', 'u', 'w', 'x', 'y', 'z'; and (5) 'X', 'Y'. In this order they represent the names of (1) classes of individuals of an arbitrary character,[1] (2) sequences of such classes, (3) natural numbers and sequences of natural numbers, (4) expressions, and (5) classes of expressions.

We turn now to the axiom system of the metalanguage. First, it is to be noticed that—corresponding to the two kinds of expressions in the metalanguage—this system contains two quite distinct kinds of sentences: the *general logical axioms* which suffice for a sufficiently comprehensive system of mathematical logic, and the *specific axioms of the metalanguage* which describe certain elementary properties of the above structural-descriptive concepts consistent with our intuitions. It is unnecessary to introduce explicitly the well-known axioms of the first kind.[2] As axioms of the second kind we adopt the following statements:[3]

AXIOM 1. *ng, sm, un, and in are expressions, no two of which are identical.*

AXIOM 2. *v_k is an expression if and only if k is a natural number distinct from 0; v_k is distinct from ng, sm, un, in, and also from v_l if $k \neq l$.*

AXIOM 3. *$x^\frown y$ is an expression if and only if x and y are expressions; $x^\frown y$ is distinct from ng, sm, un, in, and from each of the expressions v_k.*

AXIOM 4. *If x, y, z, and t are expressions, then we have $x^\frown y = z^\frown t$ if and only if one of the following conditions is satisfied: (α) $x = z$ and $y = t$; (β) there is an expression u such that $x = z^\frown u$ and $t = u^\frown y$; (γ) there is an expression u such that $z = x^\frown u$ and $y = u^\frown t$.*

AXIOM 5. (The principle of induction.) *Let X be a class which satisfies the following conditions:* (α) $ng \in X$, $sm \in X$, $un \in X$

[1] Although in the cases (1) and (4) I use distinct variables I here treat expressions as special classes of individuals, namely as classes of concrete series of printed signs (cf. p. 156, note 1).

[2] They may again be taken from Whitehead, A. N., and Russell, B. A. W. (90), cf. p. 156, note 1.

[3] As far as I know the metatheory has never before been given in the form of an axiomatized system.

and $in \in X$; (β) *if* k *is a natural number distinct from* 0, *then* $v_k \in X$; (γ) *if* $x \in X$ *and* $y \in X$; *then* $x^\frown y \in X$. *Then every expression belongs to the class* X.

The intuitive sense of Axs. 1–4 requires no further elucidation. Ax. 5 gives a precise formulation of the fact that every expression consists of a finite number of signs.

It is possible to prove that the *above axiom system is categorical.* This fact guarantees to a certain degree that it will provide a sufficient basis for the construction of the metalanguage.[1]

Some of the above axioms have a pronounced existential character and involve further consequences of the same kind. Noteworthy among these consequences is the assertion that the class of all expressions is infinite (to be more exact, denumerable). From the intuitive standpoint this may seem doubtful and hardly evident, and on that account the whole axiom-system may be subject to serious criticism. A closer analysis would restrict this criticism entirely to Axs. 2 and 3 as the essential sources of this infinite character of the metatheory. I shall not pursue this difficult problem any further here.[2] The con-

[1] I use the term 'categorical' in the sense given in Veblen, O. (86). I do not propose to explain in more detail why I see in the categoricity of an axiom system an objective guarantee that the system suffices for the establishment of the corresponding deductive science; a series of remarks on this question will be found in Fraenkel, A. (16).

Regarding the interpretation of the term 'categorical' there are certain, although not especially important, differences of opinion. Without going into details I may mention that in the case of one of the possible interpretations the proof that the system is categorical would require the addition of two further axioms to the system given in the text. In these axioms (which otherwise are not of great importance) the specific conception of expressions as classes would occur (cf. p. 156, note 1). The first axiom would state that two arbitrary expressions are disjoint classes (i.e. have no element in common), in the second the number of elements of every expression would be stipulated in some way.

[2] For example, the following truly subtle points are here raised. Normally expressions are regarded as the products of human activity (or as classes of such products). From this standpoint the supposition that there are infinitely many expressions appears to be obviously nonsensical. But another possible interpretation of the term 'expression' presents itself: we could consider all physical bodies of a particular form and size as expressions. The kernel of the problem is then transferred to the domain of physics. The assertion of the infinity of the number of expressions is then no longer senseless and even forms a special consequence of the hypotheses which are normally adopted in physics or in geometry.

sequences mentioned could of course be avoided if the axioms were freed to a sufficient degree from existential assumptions. But the fact must be taken into consideration that the elimination or weakening of these axioms, which guarantee the existence of all possible expressions, would considerably increase the difficulties of constructing the metatheory, would render impossible a series of the most useful consequences and so introduce much complication into the formulation of definitions and theorems. As we shall see later this will become clear even in the present investigations. For these reasons it seems desirable, at least provisionally, to base our work on the axiom system given above in its initial unweakened form.

Making use of the expressions and symbols of the metalanguage which have now been enumerated, I shall define those concepts which establish the calculus of classes as a formalized deductive science. These are the concepts of *sentence, axiom* (*primitive sentence*), *consequence* and *provable sentence.* But first I introduce a series of auxiliary symbols which will denote various simple types of expression and greatly facilitate the later constructions.

DEFINITION 1. *x is an* inclusion *with* v_k *as first and* v_l *as second term—in symbols* $x = \iota_{k,l}$*—if and only if* $x = (in^\frown v_k)^\frown v_l$.

DEFINITION 2. *x is a* negation *of the expression y—in symbols* $x = \bar{y}$*—if and only if* $x = ng^\frown y$.

DEFINITION 3. *x is a* logical sum (disjunction) *of the expressions y and z—in symbols* $x = y+z$*—if and only if* $x = (sm^\frown y)^\frown z$.

DEFINITION 4. *x is a* logical sum *of the expressions* $t_1, t_2, \ldots, t_n$ (*or a logical sum of a finite n-termed sequence t of expressions*)*—in symbols* $x = \sum_k^n t_k$*—if and only if t is a finite n-termed sequence of expressions which satisfies one of the following conditions:* (α) $n = 1$ *and* $x = t_1$, (β) $n > 1$ *and* $x = \sum_k^{n-1} t_k + t_n$.[1]

[1] As will be seen, Def. 4 is a recursive definition which, as such, raises certain methodological misgivings. It is, however, well known that with the help of a general method, the idea of which we owe to G. Frege and R. Dedekind, every recursive definition can be transformed into an equivalent normal definition

DEFINITION 5. *x is a* logical product (conjunction) *of the expressions y and z—in symbols $x = y.z$—if and only if* $x = \overline{\bar{y}+\bar{z}}$.

DEFINITION 6. *x is a* universal quantification *of the expression y under the variable v_k—in symbols $x = \bigcap_k y$—if and only if* $x = (un^\frown v_k)^\frown y$.

DEFINITION 7. *x is a* universal quantification *of the expression y under the variables $v_{p_1}, v_{p_2}, \ldots, v_{p_n}$—in symbols $x = \bigcap_{p_k}^{k \leqslant n} y$—if and only if p is a finite n-termed sequence of natural numbers which satisfies one of the following conditions*: (α) $n = 1$ *and* $x = \bigcap_{p_1} y$, (β) $n > 1$ *and* $x = \bigcap_{p_k}^{k \leqslant n-1} \bigcap_{p_n} y$.

DEFINITION 8. *x is a* universal quantification *of the expression y if and only if either $x = y$ or there is a finite n-termed sequence p of natural numbers such that* $x = \bigcap_{p_k}^{k \leqslant n} y$.

DEFINITION 9. *x is an* existential quantification *of the expression y under the variable v_k—in symbols $x = \bigcup_k y$—if and only if* $x = \overline{\bigcap_k \bar{y}}$.

We have thus introduced three fundamental operations by means of which compound expressions are formed from simpler ones: negation, logical addition, and universal quantification. (Logical addition is, of course, the operation which consists in forming logical sums of given expressions. The terms 'negation' and 'universal quantification' are thus used to refer both to certain operations on expressions and to expressions resulting from these operations.) If, beginning with the inclusions $\iota_{k,l}$, we apply to them the above operations any number of times we obtain an extensive class of expressions which are called *sentential functions*. We obtain the concept of *sentence* as a special case of this notion.

(cf. Dedekind, R. (15), pp. 33–40, and Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 550–7, and vol. 3, p. 244). This, however, is unpractical in so far as the formulations so obtained have a more complicated logical structure, are less clear as regards their content, and are less suitable for further derivations. For these reasons I do not propose to avoid recursive definitions in the sequel.

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 exist natural numbers k and l such that* $x = \iota_{k,l}$; (β) *there exists a sentential function y such that* $x = \bar{y}$; (γ) *there exist sentential functions y and z such that* $x = y+z$; (δ) *there exists a natural number k and a sentential function y such that* $x = \bigcap_k y$.[1]

The following expressions will serve as examples of sentential functions according to Def. 10: '$Ix_n x_m$', '$NIx_n x_m$', '$AIx_n x_m Ix_m x_n$', '$\prod x_n NIx_n x_m$', and so on. On the other hand the expressions 'I', 'Ix_n', '$AIx_n x_m$', '$\prod Ix_n x_m$', etc., are not sentential functions. It is easily seen that for every sentential function in the language we can automatically construct a structural-descriptive name of this function in the metalanguage, by making use exclusively of symbols which were introduced in Defs. 1, 2, 3, and 5. For example, the following symbolic

[1] Def. 10 is a recursive definition of a somewhat different type from that of Def. 4 since the usual 'transition from $n-1$ to n' is lacking in it. In order to reduce this to an ordinary inductive definition we must first inductively define the expressions '*x is a sentential function of the nth degree*' (inclusions $\iota_{k,l}$ would then be functions of the 0th degree, the negations and logical sums of these inclusions, as well as their generalizations for any variable, functions of the 1st degree, and so on), and then simply stipulate that '*x is a sentential function*' means the same as '*there is a natural number n such that x is a sentential function of the nth degree*'. Def. 10 could also be transformed into an equivalent normal definition in the following way:

x is a sentential function *if and only if the formula $x \in X$ holds for every class X which satisfies the following four conditions*: (α) *if k and l are natural numbers distinct from* 0, *then* $\iota_{k,l} \in X$; (β) *if $y \in X$, then* $\bar{y} \in X$; (γ) *if $y \in X$ and $z \in X$ then* $y+z \in X$; (δ) *if k is a natural number distinct from* 0 *and $y \in X$, then* $\bigcap_k y \in X$.

It should be emphasized that recursive definitions of the type of Def. 10 are open to much more serious methodological objections than the usual inductive definitions, since in contrast to the latter, statements of this type do not always admit of a transformation into equivalent normal definitions (see p. 175, note 1). The fact that such a transformation is possible in the present case is owing to the special nature of the concepts occurring in the definition (to the fact, namely, that every expression consists of a finite number of signs and that the operations given in conditions (β)–(δ) always lead from shorter to longer expressions). If, nevertheless, I sometimes give definitions of this kind in the present article in the place of equivalent normal definitions (Defs. 10, 11, 14, 22, and 24), I do so because these definitions have important advantages of quite another kind: they bring out the content of the concept defined more clearly than the normal definition does, and—in contrast to the usual recursive definition—they require no previous introduction of accessory concepts which are not used elsewhere (e.g. the accessory concept of a sentential function of the nth degree).

expressions function as names of the above examples of sentential functions: '$\iota_{1,2}$', '$\overline{\iota_{1,2}}$', '$\iota_{1,3}+\iota_{3,1}$', and '$\bigcap_1 \overline{\iota_{1,2}}$'.

DEFINITION 11. *v_k is a free variable of the sentential function x if and only if k is a natural number distinct from 0, and x is a sentential function which satisfies one of the following four conditions: (α) there is a natural number l such that $x = \iota_{k,l}$ or $x = \iota_{l,k}$; (β) there is a sentential function y such that v_k is a free variable of y and $x = \bar{y}$; (γ) there are sentential functions y and z such that v_k is a free variable of y and $x = y+z$ or $x = z+y$; (δ) there is a number l distinct from k and a sentential function y such that v_k is a free variable of y and $x = \bigcap_l y$.*

Variables which occur in a sentential function but are not free variables of this function, are usually called *bound* (*apparent*) *variables*.[1]

DEFINITION 12. *x is a* sentence (*or a* meaningful sentence)—*in symbols $x \in S$—if and only if x is a sentential function and no variable v_k is a free variable of the function x.*

Thus the expressions: $\bigcap_1 \iota_{1,1}$, $\bigcap_1 \bigcap_2 \iota_{1,2}$, $\bigcap_1 \bigcup_2 \iota_{1,2}$, $\bigcap_1(\iota_{1,1}+\bigcap_1 \bigcup_2 \iota_{2,1})$ are sentences, but the functions: $\iota_{1,1}$, $\bigcap_2 \iota_{1,2}$, $\iota_{1,1}+\bigcap_1 \bigcup_2 \iota_{2,1}$ are not sentences because they contain the free variable v_1. By virtue of the above definition the symbol 'S' denotes the class of all meaningful sentences.

The system of primitive sentences of the calculus of classes will contain two kinds of sentences.[2] The sentences of the first kind are obtained by taking any axiom system which suffices as a basis for the sentential calculus and contains the signs of negation and logical addition as the only constants. For example, the axiom system consisting of the following four axioms:

'*ANAppp*', '*ANpApq*', '*ANApqAqp*', $\sim(p \vee q) \vee (q \vee p)$

and '*ANANpqANArpArq*'.[3]

[1] Cf. Hilbert, D., and Ackermann, W. (30), pp. 52–54.

[2] Concepts which I shall discuss in the further course of § 2 do not occur in the definition of true sentence itself. I shall, however, make use of them in the preparatory discussions at the beginning of § 3 which establish the definitive form of the definition. I shall also use them in the formulation of certain consequences of this definition (Ths. 3–6 of § 3) which express characteristic and materially important properties of true sentences.

[3] This axiom system is the result of a modification and simplification of the axiom system which is found in Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 96–97; cf. Hilbert, D., and Ackermann, W. (30), p. 22.

In these axioms we replace the sentential variables 'p', 'q', and 'r' by any sentential functions, and then to the expressions thus obtained, if they are not already sentences, we apply the operation of universal quantification a sufficient number of times until all the free variables have disappeared. The following will serve as examples:

'$ANA \prod x, Ix, x, \prod x, Ix, x, \prod x, Ix, x,$',

'$\prod x, \prod x, ANIx, x, AIx, x, Ix, x,$', etc.

In order to obtain the sentences of the second kind we shall take as our starting-point some axiom system of the as yet unformalized calculus of classes which contains the inclusion sign as the only undefined sign,[1] and we then translate the axioms of this system into the language of the present article. Naturally we must first eliminate all constants which are defined by means of the inclusion sign, as well as all terms belonging to the sentential calculus and the functional calculus which are distinct in meaning from the universal quantifier, the negation sign and the sign of logical addition. As examples of sentences of this second kind we have

'$\prod x, Ix, x,$' and '$\prod x, \prod x, \prod x, ANIx, x, ANIx, x, Ix, x,$'.

DEFINITION 13. *x is an* axiom (primitive sentence) *if and only if x satisfies one of the two following conditions: (α) $x \in S$ and there exist sentential functions y, z, and u such that x is a universal quantification of one of the four functions $\overline{y+y}+y$, $\bar{y}+(y+z)$, $\overline{y+z}+(z+y)$, and $\overline{\bar{y}+z}+(\overline{u+y}+(u+z))$; ($\beta$) x is identical with one of the five sentences*

$$\bigcap_1 \iota_{1,1}, \qquad \bigcap_1 \bigcap_2 \bigcap_3 (\overline{\iota_{1,2}}+\overline{\iota_{2,3}}+\iota_{1,3}),$$

$$\bigcap_1 \bigcap_2 \bigcup_3 (\iota_{1,3} \cdot \iota_{2,3} \cdot \bigcap_4 (\overline{\iota_{1,4}}+\overline{\iota_{2,4}}+\iota_{3,4})),$$

$$\bigcap_1 \bigcap_2 \bigcup_3 (\iota_{3,1} \cdot \iota_{3,2} \cdot \bigcap_4 (\overline{\iota_{4,1}}+\overline{\iota_{4,2}}+\iota_{4,3})),$$

and

$$\bigcap_1 \bigcup_2 (\bigcap_3 \bigcap_4 ((\overline{\iota_{3,1}}+\overline{\iota_{3,2}}+\iota_{3,4}) \cdot (\overline{\iota_{1,3}}+\overline{\iota_{2,3}}+\iota_{4,3})) \cdot \bigcap_5 (\iota_{5,2}+\bigcup_6 (\iota_{6,1} \cdot \overline{\iota_{6,2}} \cdot \iota_{6,5}))).$$

[1] I have chosen here the system of postulates which is given in Huntington, E. V. (32), p. 297 (this system has, however, been simplified by the elimination of, among others, certain assumptions of an existential nature).

In the formulation of the definition of the concept of consequence I shall use, among others, the following expression: *'u is an expression obtained from the sentential function w by substituting the variable v_k for the variable v_l'*. The intuitive meaning of this expression is clear and simple, but in spite of this the definition has a somewhat complicated form:

DEFINITION 14. *x is an expression obtained from the sentential function y by* substituting *the (free) variable v_k for the (free) variable v_l if and only if k and l are natural numbers distinct from* 0, *and x and y are sentential functions which satisfy one of the following six conditions*: (α) $x = \iota_{k,k}$ *and* $y = \iota_{l,l}$; (β) *there exists a natural number m distinct from l, such that* $x = \iota_{k,m}$ *and* $y = \iota_{l,m}$ *or* $x = \iota_{m,k}$ *and* $y = \iota_{m,l}$; (γ) v_l *is not a free variable of the function y and* $x = y$; (δ) *there exist sentential functions z and t such that* $x = \bar{z}$, $y = \bar{t}$, *and z is an expression obtained from t by substituting the variable v_k for the variable v_l*; (ε) *there exist sentential functions z, t, u, and w, such that* $x = z+u$, $y = t+w$, *where z and u are obtained from t and w respectively by substituting the variable v_k for the variable v_l*; (ζ) *there exist sentential functions z, t and a natural number m distinct from k and l such that* $x = \bigcap_m z$, $y = \bigcap_m t$, *and z is obtained from t by substituting the variable v_k for the variable v_l*.[1]

For example, it follows from this definition that the expressions $\iota_{1,1}$, $\bigcap_3 (\iota_{3,1}+\iota_{1,3})$ and $\iota_{1,3}+\bigcap_2 \iota_{2,3}$ are obtained from the functions: $\iota_{2,2}$, $\bigcap_3 (\iota_{3,2}+\iota_{2,3})$ and $\iota_{2,3}+\bigcap_2 \iota_{2,3}$ respectively by substituting v_1 for v_2. But the expression $\bigcap_1 \iota_{1,3}$ cannot be obtained in this way from the function $\bigcap_2 \iota_{2,3}$ nor the expression $\bigcap_1 \iota_{1,1}$ from the function $\bigcap_2 \iota_{2,1}$.

[1] The following is a normal definition which is equivalent to the above recursive one (cf. p. 177, note 1):

x is an expression obtained from the sentential function y by substituting the variable v_k for the variable v_l if and only if k and l are natural numbers distinct from 0 *and if the formula xRy holds for every relation R which satisfies the following six conditions*: (α) $\iota_{k,k} R \iota_{l,l}$; (β) *if m is a natural number distinct from* 0 *and l, then* $\iota_{k,m} R \iota_{l,m}$ *and* $\iota_{m,k} R \iota_{m,l}$; (γ) *if z is a sentential function and v_l is not a free variable of z, then zRz*; (δ) *if zRt, then $\bar{z}R\bar{t}$*; (ε) *if zRt and uRw, then* $z+uRt+w$; (ζ) *if m is a natural number distinct from* 0, *k, and l and zRt, then* $\bigcap_m zR \bigcap_m t$.

The definitions of substitution in Leśniewski, S. (46), p. 73 (T.E. XLVII), and (47), p. 20 (T.E. XLVII°) depend on a totally different idea.

Among the consequences of a given class of sentences we include first all the sentences belonging to this class, and all the sentences which can be obtained from these by applying, an arbitrary number of times, the four operations of *substitution, detachment*, and *introduction* and *removal of the universal quantifier*.[1] If we had wished to apply these operations not only to sentences, but to arbitrary sentential functions, obtaining thereby sentential functions as results, then the meaning of the operation of substitution would be completely determined by Def. 14, the operation of detachment would correlate the function z with the functions y and $\bar{y}+z$, the operation of introduction of the universal quantifier would consist in forming the function $y+\bigcap_k z$ from the function $y+z$ (provided that v_k is not a free variable of the function y), the operation of removal of the universal quantifier would proceed in the opposite direction—from the function $y+\bigcap_k z$ to the function $y+z$.[1]

In order to simplify the construction I first define the auxiliary concept of *consequence of the n-th degree*.

DEFINITION 15. *x is a* consequence of the *n*th degree *of the class X of sentences if and only if* $x \in S$, $X \subseteq S$, *n is a natural number and either* (α) $n = 0$ *and* $x \in X$, *or* $n > 0$ *and one of the following five conditions is satisfied*: (β) *x is a consequence of the* $n-1$*th degree of the class X*; (γ) *there exist sentential functions u and w, a sentence y and natural numbers k and l such that x is the universal quantification of the function u, y is the universal quantification of the function w, u is obtainable from the function w by substituting the variable v_k for the variable v_l, and y is a consequence of the class X of the* $n-1$*th degree*; (δ) *there exist sentential functions u and w as well as sentences y and z such that x, y, and z are universal quantifications of the functions u, $\bar{w}+u$, and w respectively, and y and z are consequences of the class X of the* $n-1$*th degree*; (ε) *there exist sentential functions u and w, a sentence y and a natural number k such that x is a universal quantification of the function* $u+\bigcap_k w$, *y is a universal quantification of the function* $u+w$, *v_k is not a free variable of u, and y is a consequence of the class X of the* $n-1$*th*

[1] Cf. Łukasiewicz, J. (51), pp. 159–63; IV, p. 56.

degree; (ζ) *there exist sentential functions* u *and* w, *a sentence* y *and a natural number* k, *such that* x *is a universal quantification of the function* $u+w$, y *is a universal quantification of the function* $u+\bigcap_k w$ *and* y *is a consequence of the class* X *of the* $n-1$*th degree.*

DEFINITION 16. x *is a* consequence *of the class* X *of sentences—symbolically* $x \in Cn(X)$*—if and only if there is a natural number* n *such that* x *is a consequence of the* n*th degree of the class* X.[1]

DEFINITION 17. x *is a* provable (accepted) sentence *or a* theorem*—in symbols* $x \in Pr$*—if and only if* x *is a consequence of the set of all axioms.*

From this definition, it is easy to see that we shall have, among the provable sentences, not only all the sentences which can be obtained from the theorems of the sentential calculus in the same way in which the axioms of the first kind (i.e. those satisfying the condition (α) of Def. 13) were obtained from the axioms of the sentential calculus, but also all known theorems of the unformalized calculus of classes, provided they are first translated into the language under investigation. In order to become convinced of this we imitate in the metatheory, in every particular case, the corresponding proof from the domain of the sentential calculus or of the calculus of classes. For example, it is possible in this way to obtain the sentence $\bigcap_1 (\overline{\iota_{1,1}}+\iota_{1,1})$ from the well-known theorem '*ANpp*' of the

[1] The concept of consequence could also be introduced directly (i.e. without the help of consequence of the nth degree) in the following way:

$x \in Cn(X)$ *if and only if* $X \subseteq S$ *and if the formula* $x \in Y$ *holds for every class* Y *which satisfies the following conditions*: (α) $X \subseteq Y$; (β) *if* $y \in S$ *and is a universal quantification of the function* u, z *is a universal quantification of the function* w, u *is obtainable from the function* w *by substituting the variable* v_k *for the variable* v_l *and* $z \in Y$, *then* $y \in Y$; (γ) *if* $y \in S$, y, z, *and* t *are universal quantifications of the functions* u, $\bar{w}+u$, *and* w *respectively and* $z \in Y$ *and* $t \in Y$, *then* $y \in Y$; (δ) *if* $y \in S$, u *and* w *are sentential functions,* y *is a universal quantification of the function* $u+\bigcap_k$, z *a universal quantification of the function* $u+w$, v_k *is not a free variable of the function* u *and* $z \in Y$, *then* $y \in Y$; (ε) *if* $y \in S$, u *and* w *are sentential functions,* y *is a universal quantification of the function* $u+w$, z *a universal quantification of the function* $u+\bigcap_k w$ *and* $z \in Y$, *then* $y \in Y$.

It is, however, to be noted that by transformation of the definition just given into a recursive sentence of the type of Def. 10 we obtain a sentence which is equivalent neither with the above definition nor with any other normal definition (cf. p. 177, note 1).

sentential calculus. Translating the proof of this theorem,[1] we show successively from Def. 13 that

$$\bigcap_1(\overline{\iota_{1,1}+\iota_{1,1}}+\iota_{1,1}+\iota_{1,1}), \qquad \bigcap_1(\overline{\iota_{1,1}}+(\iota_{1,1}+\iota_{1,1})),$$

and

$$\bigcap_1(\overline{\overline{\iota_{1,1}+\iota_{1,1}}+\iota_{1,1}+(\overline{\iota_{1,1}}+(\iota_{1,1}+\iota_{1,1})+(\overline{\iota_{1,1}}+\iota_{1,1}))})$$

are axioms; consequently by Def. 15

$$\bigcap_1(\overline{\overline{\iota_{1,1}}+(\iota_{1,1}+\iota_{1,1})}+(\overline{\iota_{1,1}}+\iota_{1,1}))$$

is a consequence of the 1st degree and $\bigcap_1(\overline{\iota_{1,1}}+\iota_{1,1})$ is a consequence of the second degree of the class of all axioms. Hence by Defs. 16 and 17 $\bigcap_1(\overline{\iota_{1,1}}+\iota_{1,1})$ is a provable sentence.

From examples of such inferences the difficulties can be imagined which would at once arise if we wished to eliminate from the axioms of the metatheory the assumptions which are of an existential nature. The fact that the axioms would no longer guarantee the existence of some particular sentence, which we wish to demonstrate, is not of much consequence. Serious importance attaches only to the fact that, even assuming the existence of some concrete sentence, we could not establish its provability; since in the proof it would be necessary to refer to the existence of other, as a rule more complicated, sentences (as is seen in the proof of the theorem '$\bigcap_1(\overline{\iota_{1,1}}+\iota_{1,1}) \in Pr$' which was sketched above). So long as we are dealing with special theorems of the type '$x \in Pr$', we can take measures to provide these sentences with premisses which guarantee the existence of the sentences necessary for the proof. The difficulties would increase significantly if we passed to sentences of a general character which assert that all sentences of a certain kind are provable—or, still more generally, are consequences of the given class of sentences. It would then often be necessary to include among the premisses general existential assumptions which would not be weaker than those which, for intuitive reasons, we had eliminated from the axioms.[2]

[1] Cf. Whitehead, A. N., and Russell, B. A. W (90), vol. 1, p. 101, *2.1.

[2] This is easily seen from the examples of Ths. 11, 12, 24, and 28 in § 3.

For these reasons the standpoint might be taken that Def. 17, in case the existential assumptions are rejected, would no longer embrace all the properties which we ascribe to the concept of *theorem*. The problem of a suitable 'correction' of the above definition would then arise. More precisely expressed, it would be a question of constructing a definition of *theorem* which would be equivalent to Def. 17 under the existential assumptions and yet—quite independently of these assumptions—would have as consequences all theorems of the type '*if the sentence x exists, then* $x \in Pr$', provided the corresponding theorem '$x \in Pr$' could be proved with the help of the existential assumptions. I shall give here a brief sketch of an attempt to solve this problem.

It can easily be shown that the axiom system adopted in the metatheory possesses an interpretation in the arithmetic of the natural numbers. A one-one correspondence can be set up between expressions and natural numbers where operations on numbers having the same formal properties are correlated with the operations on expressions. If we consider this correspondence, we can pick out, from the class of all numbers, those which are correlated with sentences; among these will be the 'primitive' numbers. We can introduce the concept of a 'consequence' of a given class of numbers, and finally define the 'accepted' numbers as 'consequences' of the class of all 'primitive' numbers. If we now eliminate the existential assumptions from the axioms, the one-one correlation disappears: to every expression a natural number still corresponds, but not to every number, an expression. But we can still preserve the concept of 'accepted' number previously established and define the theorems as those which are correlated with 'accepted' numbers. If we try, on the basis of this new definition, to prove that a concrete sentence is a theorem, we shall no longer be compelled—as is easily seen—to refer to the existence of any other sentences. Nevertheless the proof will still require—and this must be emphasized—an existential assumption, the assumption, namely, that there exist sufficiently many natural numbers or—what amounts to the same thing—sufficiently many distinct individuals. Thus

in order to derive all desired conclusions from the new definition, it would be necessary to include in the metatheory the *axiom of infinity*, i.e. the assumption that the class of all individuals is infinite.[1] I know of no method, be it even less natural and more complicated than the one just discussed, which would lead to a satisfactory solution of our problem which is independent of the above axiom.

In connexion with the concepts of consequence and of theorem I have mentioned rules of inference. When we have in mind the construction of a deductive science itself, and not the investigation of such a science carried out on the basis of the metatheory, we give, instead of Def. 17, a rule by which we may add to the science as a theorem every consequence of the axioms. In our case this rule can be divided into four rules—corresponding to the four operations which we use in the construction of consequences.

By means of the concepts of sentence and of consequence all the most important methodological concepts can be introduced into the metatheory, in particular the concepts of *deductive system*, of *consistency* and of *completeness*.[2]

DEFINITION 18. *X is a* deductive system *if and only if*

$$Cn(X) \subseteq X \subseteq S.$$

DEFINITION 19. *X is a* consistent *class of sentences if and only if* $X \subseteq S$ *and if, for every sentence x, either* $x \overline{\in} Cn(X)$ *or* $\bar{x} \overline{\in} Cn(X)$.

DEFINITION 20. *X is a* complete *class of sentences if and only if* $X \subseteq S$ *and if, for every sentence x, either* $x \in Cn(X)$ *or* $\bar{x} \in Cn(X)$.

In the sequel yet another concept will prove useful:

DEFINITION 21. *The sentences x and y are* equivalent *with respect to the class X of sentences if and only if* $x \in S$, $y \in S$, $X \subseteq S$ *and both* $\bar{x}+y \in Cn(X)$ *and* $\bar{y}+x \in Cn(X)$.

A more detailed analysis of the concepts introduced in this section would exceed the limits of the present work.

[1] Cf. Whitehead, A. N., and Russell, B. A. W. (90), vol. 2, p. 203.
[2] Cf. pp. 70, 90, and 93 of the present volume.

§ 3. The Concept of True Sentence in the Language of the Calculus of Classes

I pass on now to the chief problem of this article—the construction of the definition of *true sentence*, the language of the calculus of classes still being the object of investigation.

It might appear at first sight that at the present stage of our discussion this problem can be solved without further difficulty, that 'true sentence' with respect to the language of a formalized deductive science means nothing other than 'provable theorem', and that consequently Def. 17 is already a definition of truth and moreover a purely structural one. Closer reflection shows, however, that this view must be rejected for the following reason: no definition of true sentence which is in agreement with the ordinary usage of language should have any consequences which contradict the principle of the excluded middle. This principle, however, is not valid in the domain of provable sentences. A simple example of two mutually contradictory sentences (i.e. such that one is the negation of the other) neither of which is provable is provided by Lemma E below. The extension of the two concepts is thus not identical. From the intuitive standpoint all provable sentences are without doubt true sentences (the Defs. 13–17 of § 2 were formulated with that in mind). Thus the definition of true sentence which we are seeking must also cover sentences which are not provable.[1]

[1] The fact must also be taken into consideration that—in contrast to the concept of true sentence—the concept of provable sentence has a purely accidental character when applied to some deductive sciences, which is chiefly connected with the historical development of the science. It is sometimes difficult to give objective grounds for narrowing or widening the extension of this concept in a particular direction. For example, when we are dealing with the calculus of classes the sentence $\overline{\bigcap_1 \bigcap_2 \iota_{1,2}}$, which stipulates the existence of at least two distinct classes, is not accepted on the basis of the definitions of § 2—which will be expressed in Lemma E. Moreover, this sentence cannot be derived from the formal hypotheses upon which the work of Schröder is based, although in this case the matter is not quite clear (cf. Schröder, E. (62), vol. 1, pp. 245 and 246; vol. 2, Part 1, p. 278; vol. 3, Part 1, pp. 17 and 18); but in many works this sentence occurs as an axiom of the algebra of logic or forms an obvious consequence of these axioms (cf. Huntington, E. V. (32), p. 297, Post. 10). For quite different reasons, which will be discussed below in connexion with Th. 24 (cf. especially p. 207, footnote), it would be desirable to include the sentence $\bigcap_1(\bigcap_2 \iota_{1,2} + \bigcup_3(\iota_{2,1} \cdot \bigcap_3(\bigcap_4 \iota_{3,4} + \iota_{3,3} + \iota_{3,3})))$ among the

Let us try to approach the problem from quite a different angle, by returning to the idea of a semantical definition as in § 1. As we know from § 2, to every sentence of the language of the calculus of classes there corresponds in the metalanguage not only a name of this sentence of the structural-descriptive kind, but also a sentence having the same meaning. For example, corresponding to the sentence '$\prod x_1 \prod x_2 A I x_1 x_2 I x_2 x_1$' is the name '$\bigcap_1 \bigcap_2(\iota_{1,2}+\iota_{2,1})$' and the sentence 'for any classes a and b we have $a \subseteq b$ or $b \subseteq a$'. In order to make clear the content of the concept of truth in connexion with some one concrete sentence of the language with which we are dealing we can apply the same method as was used in § 1 in formulating the sentences (3) and (4) (cf. p. 156). We take the scheme (2) and replace the symbol 'x' in it by the name of the given sentence, and 'p' by its translation into the metalanguage. All sentences obtained in this way, e.g. '$\bigcap_1 \bigcap_2(\iota_{1,2}+\iota_{2,1})$ *is a true sentence if and only if for any classes a and b we have $a \subseteq b$ or $b \subseteq a$*', naturally belong to the metalanguage and explain in a precise way, in accordance with linguistic usage, the meaning of phrases of the form 'x is a true sentence' which occur in them. Not much more in principle is to be demanded of a general definition of true sentence than that it should satisfy the usual conditions of methodological correctness and include all partial definitions of this type as special cases; that it should be, so to speak, their logical product. At most we can also require that only sentences are to belong to the extension of the defined concept, so that, on the basis of the definition constructed, all sentences of the type '*x is not a true sentence*', in which in the place of 'x' we have the name of an arbitrary expression (or of any other object) which is not a sentence, can be proved.

Using the symbol 'Tr' to denote the class of all true sentences, the above postulate can be expressed in the following convention:

Convention T. *A formally correct definition of the symbol*

theorems, although this is not usually done. In the course of this work I shall have several occasions to return to the problem of the mutual relations of these two concepts: of theorem and of true sentence.

'Tr', formulated in the metalanguage, will be called an adequate definition of truth *if it has the following consequences*:

(α) *all sentences which are obtained from the expression '$x \in Tr$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage*;

(β) *the sentence 'for any x, if $x \in Tr$ then $x \in S$' (in other words '$Tr \subseteq S$')*.[1]

It should be noted that the second part of the above convention is not essential; so long as the metalanguage already has the symbol '*Tr*' which satisfies the condition (α), it is easy to define a new symbol '*Tr'*' which also satisfies the condition (β). It suffices for this purpose to agree that *Tr'* is the common part of the classes *Tr* and *S*.

If the language investigated only contained a finite number of sentences fixed from the beginning, and if we could enumerate all these sentences, then the problem of the construction of a correct definition of truth would present no difficulties. For this purpose it would suffice to complete the following scheme: *$x \in Tr$ if and only if either $x = x_1$ and p_1, or $x = x_2$ and p_2,... or $x = x_n$ and p_n*, the symbols 'x_1', 'x_2',..., 'x_n' being replaced by structural-descriptive names of all the sentences of the language investigated and 'p_1', 'p_2',..., 'p_n' by the corresponding translation of these sentences into the metalanguage. But the situation is not like this. Whenever a language contains infinitely many sentences, the definition constructed automatically according to the above scheme would have to consist of infinitely many words, and such sentences cannot be formulated either in the metalanguage

[1] If we wished to subject the metalanguage and the metatheory expressed in it to the process of formalization, then the exact specification of the meaning of various expressions which occur in the convention T would present no great difficulties, e.g. the expressions '*formally correct definition of the given symbol*', '*structural-descriptive name of a given expression of the language studied*', '*the translation of a given sentence (of the language studied) into the metalanguage*'. After unimportant modifications of its formulation the convention itself would then become a normal definition belonging to the metatheory.

or in any other language. Our task is thus greatly complicated.

The idea of using the recursive method suggests itself. Among the sentences of a language we find expressions of rather varied kinds from the point of view of logical structure, some quite elementary, others more or less complicated. It would thus be a question of first giving all the operations by which simple sentences are combined into composite ones and then determining the way in which the truth or falsity of composite sentences depends on the truth or falsity of the simpler ones contained in them. Moreover, certain elementary sentences could be selected, from which, with the help of the operations mentioned, all the sentences of the language could be constructed; these selected sentences could be explicitly divided into true and false, by means, for example, of partial definitions of the type described above. In attempting to realize this idea we are however confronted with a serious obstacle. Even a superficial analysis of Defs. 10–12 of § 2 shows that in general composite sentences are in no way compounds of simple *sentences*. Sentential functions do in fact arise in this way from elementary functions, i.e. from inclusions; sentences on the contrary are certain special cases of sentential functions. In view of this fact, no method can be given which would enable us to define the required concept directly by recursive means. The possibility suggests itself, however, of introducing a more general concept which is applicable to any sentential function, can be recursively defined, and, when applied to sentences, leads us directly to the concept of truth. These requirements are met by the notion of the *satisfaction of a given sentential function by given objects*, and in the present case by given classes of individuals.

Let us try first to make clear by means of some examples the usual meaning of this notion in its customary linguistic usage. The way in which we shall do this represents a natural generalization of the method which we have previously used for the concept of truth.

The simplest and clearest case is that in which the given sentential function contains only *one* free variable. We can then

significantly say of every single object that it does or does not satisfy the given function.[1] In order to explain the sense of this phrase we consider the following scheme:

for all a, a satisfies the sentential function x if and only if p

and substitute in this scheme for 'p' the given sentential function (after first replacing the free variable occurring in it by 'a') and for 'x' some individual name of this function. Within colloquial language we can in this way obtain, for example, the following formulation:

for every a, we have a satisfies the sentential function 'x is white' if and only if a is white

(and from this conclude, in particular, that snow satisfies the function 'x is white'). A similar construction will be familiar to the reader from school algebra, where sentential functions of a special type, called *equations*, are considered together with the numbers which satisfy these functions, the so-called *roots* of the equations (e.g. 1 is the only root of the equation '$x+2 = 3$').

When, in particular, the function belongs to the language of the calculus of classes, and the corresponding explanation of the expression 'a satisfies the given sentential function' is to be formulated wholly in the terms of the metalanguage, then in the above scheme we insert for 'p' not the sentential function itself, but the expression of the metalanguage having the same meaning, and for 'x' we substitute an individual name of this function which likewise belongs to the metalanguage. For example, this method gives the following formulation in connexion with the function '$\prod x_n Ix_1 x_n$':

for all a, a satisfies the sentential function $\bigcap_2 \iota_{1,2}$ *if and only if for all classes b we have* $a \subseteq b$

(whence it follows at once that the only class which satisfies the function '$\prod x_n Ix_1 x_n$' is the null class).

In cases where the sentential function has two distinct free variables we proceed in an exactly analogous manner. The only

[1] Provisionally I ignore problems connected with semantical categories (or logical types); these problems will be discussed in § 4.

difference is that the concept of satisfaction now refers not to single objects but to pairs (more accurately to ordered pairs) of objects. In this way we reach the following formulations:

for all a and b, a and b satisfy the sentential function 'x sees y' if and only if a sees b; for all a and b, a and b satisfy the sentential function $\iota_{2,3}$ *(i.e. '*$Ix_n x_m$*') if and only if* $a \subseteq b$.

Finally we pass to the general case, where the given sentential function contains an arbitrary number of free variables. For the sake of a uniform mode of expression we shall from now on not say that given objects but that *a given infinite sequence of objects satisfies a given sentential function*. If we restrict ourselves to functions from the calculus of classes, then the establishment of an unambiguous explanation of this expression is facilitated by the fact that all the variables which occur in the language of this science are ordered (enumerated) in a sequence. In considering the question of which sequences satisfy a given sentential function, we shall always have in mind a one-many correspondence of certain terms of a sequence f with the free variables of the sentential function, where with every variable corresponds the term of the sequence with the same index (i.e. the term f_k will be correlated with the variable v_k). No account will be taken of the terms which are not correlated with any variable.[1] We can explain the procedure best by means of concrete examples. Consider the function $\bigcap_2 \iota_{1,2}$ already mentioned. This function contains only *one* free variable v_1, so that we consider only the first terms of sequences. We say that the *infinite sequence f of classes satisfies the sentential function* $\bigcap_2 \iota_{1,2}$ *if and only if the*

[1] This is a simplification of a purely technical nature. Even if we could not order all the variables of a given language in a sequence (e.g. because we used symbols of arbitrary shapes as variables), we could still number all the signs, and thus all the variables, of every given expression, e.g. on the basis of the natural order in which they follow one another in the expression: the sign standing on the extreme left could be called the first, the next the second, and so on. In this way we could again set up a certain correlation between the free variables of a given function and the terms of the sequence. This correlation (in contrast to the one described in the text) would obviously vary with the form of the function in question; this would carry with it rather serious complications in the formulation of Def. 22 given below and especially of conditions (γ) and (δ).

class f_1 satisfies this function in the former sense, i.e. if for all classes b, we have $f_1 \subseteq b$. In an analogous way the *infinite sequence f of classes satisfies the sentential function $\iota_{2,3}$ if and only if the classes f_2 and f_3 satisfy the function in the previous sense, i.e. if $f_2 \subseteq f_3$.* This process may be described in general terms as follows:

We consider the following scheme:

The sequence f satisfies the sentential function x if and only if f is an infinite sequence of classes and p. If we have a sentential function from the calculus of classes, then in the above we replace the symbol 'x' by an individual (structural-descriptive) name of this function formulated in the metalanguage, but 'p' by a translation of the function into the metalanguage, where all free variables v_k, v_l, etc. are replaced by corresponding symbols 'f_k', 'f_l', etc.

We shall use a recursive method in order to formulate a general definition of satisfaction of a sentential function by a sequence of classes, which will include all partial definitions of this notion as special cases which are obtained from the given scheme in the way described above. For this purpose it will suffice, bearing in mind the definition of sentential function, to indicate which sequences satisfy the inclusions $\iota_{k,l}$ and then to specify how the notion we are defining behaves when the three fundamental operations of negation, disjunction, and universal quantification are performed on sentential functions.

The operation of universal quantification calls for special consideration. Let x be any sentential function, and assume that we already know which sequences satisfy the function x. Considering the meaning of the operation of universal quantification, we shall say that the sequence f satisfies the function $\bigcap_k x$ (where k is a particular natural number) only if this sequence itself satisfies the function x and does not cease to satisfy it even when the kth term of this sequence varies in any way; in other words, if every sequence which differs from the given sequence in at most the kth place also satisfies the function. For example, the function $\bigcap_2 \iota_{1,2}$ is satisfied by those, and only those, sequences f for which the formula $f_1 \subseteq f_2$ holds without regard to the way in which the second term of this sequence is allowed to vary (as is easily seen, this is only possible when the first term is the null class).

After these explanations the understanding of the following definition should not be difficult.

DEFINITION 22. *The sequence f* satisfies *the sentential function x if and only 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 natural numbers k and l such that $x = \iota_{k,l}$ and $f_k \subseteq f_l$;* (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 natural number k and a sentential function y such that $x = \bigcap_k y$ and every infinite sequence of classes which differs from f in at most the k-th place satisfies the function y.*[1]

The following are examples of the application of the above definition to concrete sentential functions: the infinite sequence f satisfies the inclusion $\iota_{1,2}$ if and only if $f_1 \subseteq f_2$, and the function $\overline{\iota_{2,3}} + \overline{\iota_{3,2}}$ if and only if $f_2 \neq f_3$; the functions $\bigcap_2 \iota_{1,2}$ and $\bigcap_2 \iota_{2,3}$ are satisfied by those, and only those, sequences f in which f_1 is the null class and f_3 the universal class (i.e. the class of all individuals) respectively; finally, every infinite sequence of classes satisfies the function $\iota_{1,1}$ and no such sequence satisfies the function $\iota_{1,2} \cdot \overline{\iota_{1,2}}$.

The concept just defined is of the greatest importance for investigations into the semantics of language. With its help the meaning of a whole series of concepts in this field can easily be

[1] The normal definition, which is equivalent to the above recursive one, is as follows (cf. pp. 70, 90, and 93):

The sequence f satisfies the sentential function x if and only if we have fRx for every relation R which satisfies the following condition:

For any g and y, in order that gRy it is necessary and sufficient that g is an infinite sequence of classes, y is a sentential function and either (α) *there are natural numbers k and l such that $y = \iota_{k,l}$ and $g_k \subseteq g_l$ or* (β) *there is a sentential function z such that $y = \bar{z}$ and the formula gRz does not hold; or* (γ) *there are sentential functions z and t such that $y = z+t$ and gRz or gRt; or finally* (δ) *there is a natural number k and a sentential function z such that $y = \bigcap_k z$ and hRz for every infinite sequence h of classes which is distinct from g at the k-th place at most.*

defined, e.g. the concepts of denotation, definability,[1] and truth, with the last of which we are especially concerned here.

The concept of truth is reached in the following way. On the basis of Def. 22 and the intuitive considerations which preceded it, it is easy to realize that whether or not a given sequence satisfies a given sentential function depends only on those terms of the sequence which correspond (in their indices) with the free variables of the function. Thus in the extreme case, when the function is a sentence, and so contains no free variable (which is in no way excluded by Def. 22), the satisfaction of a function by a sequence does not depend on the properties of the terms of the sequence at all. Only two possibilities then remain: either every infinite sequence of classes satisfies a given sentence, or no sequence satisfies it (cf. the Lemmas A and B given below). The sentences of the first kind, e.g. $\bigcup_1 \iota_{1,1}$, are the *true sentences*; those of the second kind, e.g. $\bigcap_1 \iota_{1,1}$, can correspondingly be called the *false sentences*.†

[1] To say that the name x denotes a given object a is the same as to stipulate that the object a (or every sequence of which a is the corresponding term) satisfies a sentential function of a particular type. In colloquial language it would be a function which consists of three parts in the following order: a variable, the word 'is' and the given name x. As regards the concept of definability, I shall try to explain its content only in a particular case. If we consider which properties of classes we regard as definable (in reference to the system of the calculus of classes discussed here), we reach the following formulations:

We say that the sentential function x defines the property P of classes if and only if for a natural number k (α) x contains v_k as its only free variable, and (β) in order that an infinite sequence f of classes should satisfy x, it is necessary and sufficient that f_k should have the property P; we say that the property P of classes is definable if and only if there is a sentential function x which defines P.

On the basis of these stipulations it can be shown, for example, that such properties of classes as emptiness, of containing only one, two, three, etc., elements are definable. On the other hand the property of containing infinitely many elements is not definable (cf. the remarks given below in connexion with Ths. 14–16). It will also be seen that with this interpretation the concept of definability does not depend at all on whether the formalization of the science investigated admits of the possibility of constructing definitions. More exact discussions of definability will be found in articles VI and X of the present volume.

† A method of defining truth which is essentially equivalent to the method developed in this work, but is based upon a different idea, has recently been suggested by J. C. C. McKinsey in his paper 'A new definition of truth', *Synthèse*, vol. 7 (1948–9), pp. 428–33.

DEFINITION 23. *x is a* true sentence*—in symbols $x \in Tr$—if and only if $x \in S$ and every infinite sequence of classes satisfies x.*[1]

The question now arises whether this definition, about the formal correctness of which there is no doubt, is also materially correct—at least in the sense previously laid down in the convention **T**. It can be shown that the answer to this question is affirmative: *Def. 23 is an adequate definition of truth in the sense of convention* **T**, since its consequences include all those required by this convention. Nevertheless it can be seen without difficulty (from the fact that the number of these consequences is infinite) that the exact and general establishment of this fact has no place within the limits of the considerations so far brought forward. The proof would require the setting up of an entirely new apparatus: in fact it involves the transition to a level one step higher—to the meta-metatheory, which would have to be preceded by the formalization of the metatheory which forms the foundation of our investigations.[2] If we do not wish to depart from the level of our previous discussions, only one

[1] In the whole of the above construction we could operate with finite sequences with a variable number of terms instead of with infinite sequences. It would then be convenient to generalize the concept of finite sequence. In the usual interpretation of this term a sequence which has an nth term must also have all terms with indices less than n—we must now relinquish this postulate and regard any many-one relation as a finite sequence if its counter domain consists of a finite number of natural numbers distinct from 0. The modification of the construction would consist in eliminating from the sequences which satisfy the given sentential function all 'superfluous' terms, which have no influence on the satisfaction of the function. Thus if v_k, v_l, etc., occur as free variables in the function (of course in finite number), only those terms with the indices k, l, etc., would remain in the sequence which satisfies this function. For example, those, and only those, sequences f of classes would satisfy the function $\iota_{2,4}$ which consist of only two terms f_2 and f_4 verifying the formula $f_2 \subseteq f_4$. The value of such a modification from the standpoint of naturalness and conformity with the usual procedure is clear, but when we come to carry it out exactly certain defects of a logical nature show themselves: Def. 22 then takes on a more complicated form. Regarding the concept of truth, it is to be noted that—according to the above treatment—only *one* sequence, namely the 'empty' sequence which has no member at all, can satisfy a sentence, i.e. a function without free variables; we should then have to call those sentences true which are actually satisfied by the 'empty' sequence. A certain artificiality attaching to this definition will doubtless displease all those who are not sufficiently familiar with the specific procedures which are commonly used in mathematical constructions.

[2] See p. 188, footnote.

method, the empirical method, remains—the verification of the properties of Def. 23 in a series of concrete examples.

Consider, for example, the sentence $\bigcap_1 \bigcup_2 \iota_{1,2}$, i.e. '$\prod x_1 N \prod x_2 N I x_1 x_2$'. According to Def. 22 the sentential function $\iota_{1,2}$ is satisfied by those, and only those, sequences f of classes for which $f_1 \subseteq f_2$ holds, but its negation, i.e. the function $\overline{\iota_{1,2}}$, only by those sequences for which $f_1 \not\subseteq f_2$ holds. Consequently a sequence f satisfies the function $\bigcap_2 \overline{\iota_{1,2}}$, if every sequence g which differs from f in at most the 2nd place satisfies the function $\overline{\iota_{1,2}}$ and thus verifies the formula $g_1 \not\subseteq g_2$. Since $g_1 = f_1$ and the class g_2 may be quite arbitrary, only those sequences f satisfy the function $\bigcap_2 \overline{\iota_{1,2}}$, which are such that $f_1 \not\subseteq b$ for any class b. If we proceed in an analogous way, we reach the result that the sequence f satisfies the function $\bigcup_2 \iota_{1,2}$, i.e. the negation of the function $\bigcap_2 \overline{\iota_{1,2}}$, only if there is a class b for which $f_1 \subseteq b$ holds. Moreover, the sentence $\bigcap_1 \bigcup_2 \iota_{1,2}$ is only satisfied (by an arbitrary sequence f) if there is for an arbitrary class a, a class b for which $a \subseteq b$. Finally by applying Def. 23 we at once obtain one of the theorems which were described in the condition (α) of the convention T:

$\bigcap_1 \bigcup_2 \iota_{1,2} \in Tr$ *if and only if for all classes* a *there is a class* b *such that* $a \subseteq b$.

From this we infer without difficulty, by using the known theorems of the calculus of classes, that $\bigcap_1 \bigcup_2 \iota_{1,2}$ is a true sentence.

We can proceed in an exactly analogous way with every other sentence of the language we are considering. If for such a sentence we construct a corresponding assertion described in the condition (α) and then apply the mode of inference used above, we can prove without the least difficulty that this assertion is a consequence of the definition of truth which we have adopted. In many cases, with the help of only the simplest laws of logic (from the domain of the sentential calculus and the calculus of classes), we can draw definitive conclusions from theorems obtained in this way about the truth or falsity of the sentences in

question. Thus, for example, $\bigcap_1 \bigcup_2 (\iota_{1,2} + \overline{\iota_{1,2}})$ is shown to be a true and $\bigcap_1 \bigcap_2 \overline{\iota_{1,2}}$ a false sentence. With respect to other sentences, e.g. the sentence $\bigcap_1 \bigcap_2 \bigcap_3 (\iota_{1,2} + \iota_{2,3} + \iota_{3,1})$ or its negation, the analogous question cannot be decided (at least so long as we do not have recourse to the special existential assumptions of the metatheory, cf. p. 174): Def. 23 alone gives no general criterion for the truth of a sentence.[1] Nevertheless, through the theorems obtained, the meaning of the corresponding expressions of the type '$x \in Tr$' becomes intelligible and unambiguous. It should also be noted that the theorem expressed in the condition (β) of the convention T is also an obvious consequence of our definition.

With these discussions the reader will doubtless have reached the subjective conviction that Def. 23 actually possesses the property which it is intended to have: it satisfies all the conditions of convention T. In order to fix the conviction of the material correctness of the definition which has been reached in this way, it is worth while studying some characteristic general theorems that can be derived from it. With a view to avoiding encumbering this work with purely deductive matter, I shall give these theorems without exact proofs.[2]

THEOREM 1 (The principle of contradiction). *For all sentences* x, *either* $x \overline{\in} Tr$ *or* $\bar{x} \overline{\in} Tr$.

This is an almost immediate consequence of Defs. 22 and 23.

THEOREM 2 (The principle of excluded middle). *For all sentences* x, *either* $x \in Tr$ *or* $\bar{x} \in Tr$.

[1] At least when it is regarded from the methodological viewpoint this is not a defect of the definition in question; in this respect it does not differ at all from the greater part of the definitions which occur in the deductive sciences.

[2] The proofs are based on the general laws of logic, the specific axioms of the metascience and the definitions of the concepts occurring in the theorems. In some cases the application of the general properties of the concepts of consequence, of deductive system, etc., which are given in article V of the present volume is indicated. We are able to use the results obtained there because it can easily be shown that the concepts of sentence and consequence introduced here satisfy all the axioms upon which the above-mentioned work was based.

In the proof the following lemma, which follows from Defs. 11 and 22, plays an essential part:

LEMMA A. *If the sequence f satisfies the sentential function x, and the infinite sequence g of classes satisfies the following condition: for every k, $f_k = g_k$ if v_k is a free variable of the function x; then the sequence g also satisfies the function x.*

As an immediate consequence of this lemma and Def. 12 we obtain Lemma B which, in combination with Defs. 22 and 23 easily leads to Th. 1:

LEMMA B. *If $x \in S$ and at least one infinite sequence of classes satisfies the sentence x, then every infinite sequence of classes satisfies x.*

THEOREM 3. *If $X \subseteq Tr$ then $Cn(X) \subseteq Tr$; thus in particular $Cn(Tr) \subseteq Tr$.*

This theorem is proved by strong induction based chiefly on Defs. 15, 16, 22, and 23; the following simple lemma is also useful in this connexion:

LEMMA C. *If y is a universal quantification of the sentential function x, then in order that every infinite sequence of classes should satisfy x, it is necessary and sufficient that every infinite sequence of classes satisfies y.*

The results contained in Ths. 1–3 may be summarized in the following (obtained with the help of Defs. 18–20):

THEOREM 4. *The class Tr is a consistent and complete deductive system.*

THEOREM 5. *Every provable sentence is a true sentence, in other words, $Pr \subseteq Tr$.*

This theorem follows immediately from Def. 17, from Th. 3, and from Lemma D, the proof of which (on the basis of Def. 13 and Lemma C among others) presents no difficulty.

LEMMA D. *Every axiom is a true sentence.*

Th. 5 cannot be inverted:

THEOREM 6. *There exist true sentences which are not provable, in other words, $Tr \nsubseteq Pr$.*

This is an immediate consequence of Th. 2 and the following lemma, the exact proof of which is not quite easy:

LEMMA E. *Both $\bigcap_1 \bigcap_2 \iota_{1,2} \overline{\in} Pr$ and $\overline{\bigcap_1 \bigcap_2 \iota_{1,2}} \overline{\in} Pr$.*[1]

As a corollary from Ths. 1, 5, and 6, I give finally the following theorem:

THEOREM 7. *The class Pr is a consistent, but not a complete deductive system.*

In the investigations which are in progress at the present day in the methodology of the deductive sciences (in particular in the work of the Göttingen school grouped around Hilbert) another concept of a relative character plays a much greater part than the absolute concept of truth and includes it as a special case. This is the concept of *correct or true sentence in an individual domain a*.[2,3] By this is meant (quite generally and roughly speaking) every sentence which is true in the usual sense if we restrict the extension of the individuals considered to a given class a, or—somewhat more precisely—when we agree to interpret the terms 'individual', 'class of individuals', etc., as 'element of the class a', 'subclass of the class a', etc., respectively. Where we are dealing with the concrete case of sentences from the calculus of classes we must interpret expressions of the type '$\prod xp$' as '*for every subclass x of the*

[1] If we wish to include the sentence $\overline{\bigcap_1 \bigcap_2 \iota_{1,2}}$ among the acceptable sentences (as is often the case, cf. p. 186, footnote) we could use here, instead of Lemma E, the following Lemma E′:

Both $\bigcap_1 \bigcap_2 (\iota_{1,2} + \iota_{2,1}) \overline{\in} Pr$ and $\overline{\bigcap_1 \bigcap_2 (\iota_{1,2} + \iota_{2,1})} \overline{\in} Pr$.

The idea of the proof of both of these lemmas is the same as that of the proofs of the consistency and incompleteness of the lower functional calculus which is found in Hilbert, D., and Ackermann, W. (30), pp. 65–68.

[2] The discussion of these relativized notions is not essential for the understanding of the main theme of this work and can be omitted by those readers who are not interested in special studies in the domain of the methodology of the deductive sciences (only the discussions on pp. 208–9 are in closer connexion with our main thesis).

[3] In this connexion see Hilbert, D., and Ackermann, W. (30), especially pp. 72–81, and Bernays, P., and Schönfinkel, M. (5 *a*). But it should be emphasized that the authors mentioned relate this concept not to sentences but to sentential functions with free variables (because in the language of the lower functional calculus which they use there are no sentences in the strict sense of the word) and, connected with this, they use the term 'generally valid' instead of the term 'correct' or 'true'; cf. the second of the works cited above, pp. 347–8.

class a we have p', and expressions of the type '*Ixy*' as '*the subclass x of the class a is contained in the subclass y of the class a*'. We obtain a precise definition of this concept by means of a modification of Defs. 22 and 23. As derived concepts we introduce the notion of a *correct sentence in an individual domain with k elements* and the notion of a *correct sentence in every individual domain*. It is worthy of note that—in spite of the great importance of these terms for metamathematical investigations—they have hitherto been used in a purely intuitive sense without any attempt to define their meaning more closely.[1]

DEFINITION 24. *The sequence f* satisfies *the sentential function x* in the individual domain *a if and only if a is a class of individuals, f an infinite sequence of subclasses of the class a and x a sentential function satisfying one of the following four conditions*: (α) *there exist natural numbers k and l such that* $x = \iota_{k,l}$ *and* $f_k \subseteq f_l$; (β) *there is a sentential function y such that* $x = \bar{y}$ *and the sequence f does not satisfy y in the individual domain a*; (γ) *there are sentential functions y and z such that* $x = y + z$ *and f satisfies either y or z in the individual domain a*; (δ) *there is a natural number k and a sentential function y such that* $x = \prod_k y$ *and every infinite sequence g of subclasses of the class a which differs from f in at most the k-th place satisfies y in the individual domain a.*

DEFINITION 25. *x is a* correct (true) sentence in the individual domain *a if and only if* $x \in S$ *and every infinite sequence of subclasses of the class a satisfies the sentence x in the individual domain a.*

DEFINITION 26. *x is a* correct (true) sentence in an individual domain with *k* elements—*in symbols* $x \in Ct_k$—*if and only if there exists a class a such that k is the cardinal number of the class a and x is a correct sentence in the individual domain a.*

[1] An exception is furnished by Herbrand, J. (26) in which the author defines the concept of true sentence in a finite domain (pp. 108–12). A comparison of Herbrand's definition with Defs. 25 and 26 given in the text will lead the reader at once to the conclusion that we have to do here with like-sounding terms rather than with a relationship of content. Nevertheless, it is possible that with respect to certain concrete deductive sciences, and under special assumptions for the corresponding metatheory, Herbrand's concept has the same extension (and also the same importance for metamathematical investigations) as a certain special case of the concept introduced in Def. 25.

DEFINITION 27. *x is a* correct (true) sentence in every individual domain—*in symbols* $x \in Ct$—*if and only if for every class a, x is a correct sentence in the individual domain a.*

If we drop the formula '$x \in S$' from Def. 25, and thereby modify the content of Defs. 26 and 27, we reach concepts of a more general nature which apply not only to sentences but also to arbitrary sentential functions.

Examples of the application to concrete sentences of the concepts defined will be given below. In the interest of more convenient formulation of various properties of these concepts, I introduce some further symbolical abbreviations.

DEFINITION 28. $x = \epsilon_k$ *if and only if*

$$x = \overline{\prod_{k+1} \iota_{k,k+1}} \cdot \prod_{k+1} (\prod_{k+2} \iota_{k+1,k+2} + \overline{\iota_{k+1,k}} + \iota_{k,k+1}).$$

DEFINITION 29. $x = \alpha$ *if and only if*

$$x = \prod_1 (\prod_2 \iota_{1,2} + \bigcup_2 (\iota_{2,1} \cdot \epsilon_2)).$$

As is easily seen, the sentential function ϵ_k states that the class denoted by the variable v_k consists of only *one* element; the sentence α_1 which plays a great part in subsequent investigations, states that every non-null class includes a one-element class as a part.

DEFINITION 30. $x = \beta_n$ *if and only if either* $n = 0$ *and* $x = \prod_1 \overline{\epsilon_1}$, *or* $n \neq 0$ *and* $x = \prod_k^{k \leqslant n+1} \left(\sum_k^{n+1} \overline{\epsilon_k} + \sum_l^n \sum_k^l (\iota_{k,l+1} \cdot \iota_{l+1,k}) \right)$.

DEFINITION 31. $x = \gamma_n$ *if and only if either* $n = 0$ *and* $x = \beta_0$, *or* $n \neq 0$ *and* $x = \overline{\beta_{n-1}} \cdot \beta_n$.

It follows from these definitions that the sentences β_n and γ_n (where n is any natural number) establish that there are at most n, and exactly n, distinct one-element classes respectively, or, what amounts to the same thing, that there are n distinct individuals.

DEFINITION 32. *x is a* quantitative sentence (*or a* sentence about the number of individuals) *if and only if there exists a finite sequence p of n natural numbers such that either* $x = \sum_k^n \gamma_{p_k}$ *or* $x = \overline{\sum_k^n \gamma_{p_k}}$.

I shall now give a series of characteristic properties of the defined concepts and the more important connexions which relate them with notions already introduced. This is the place for some results of a more special nature which are connected with the particular properties of the calculus of classes and cannot be extended to other disciplines of related logical structure (e.g. Ths. 11–13, 24, and 28).

THEOREM 8. *If a is a class of individuals and k the cardinal number of this class, then in order that x should be a correct sentence in the individual domain a it is necessary and sufficient that $x \in Ct_k$.*

The proof is based on the following lemma (among other things) which follows from Def. 24:

LEMMA F. *Let a and b be two classes of individuals and R a relation which satisfies the following conditions*: (α) *for any f' and g', if $f'Rg'$ then f' is an infinite sequence of subclasses of a, and g' of subclasses of b*; (β) *if f' is any infinite sequence of subclasses of a, then there is a sequence g' such that $f'Rg'$*; (γ) *if g' is any infinite sequence of subclasses of b, then there is a sequence f' such that $f'Rg'$*; (δ) *for all f', g', f'', g'', k and l, if $f'Rg'$, $f''Rg''$, and k and l are natural numbers distinct from 0, then $f'_k \subseteq f''_l$ if and only if $g'_k \subseteq g''_l$. If fRg and the sequence f satisfies the sentential function x in the individual domain a, then the sequence g also satisfies this function in the individual domain b.*

From this lemma, with the help of Def. 25, we easily obtain Lemma G which, together with Def. 26, at once gives Th. 8:

LEMMA G. *If the classes a and b of individuals have the same cardinal number, and x is a correct sentence in the individual domain a, then x is also a correct sentence in the individual domain b.*

According to Th. 8 (or Lemma G) the extension of the concept 'a sentence which is correct in the individual domain a' depends entirely on one property of the class a, namely on its cardinal number. This enables us to neglect in the sequel all results concerning this concept, because they can be derived immediately from the corresponding theorems relating to the classes Ct_k.

With the help of Defs. 24 and 25 the Ths. 1–6 and Lemmas A–D can be generalized by replacing the expressions '*infinite sequence of classes*', '*the sequence . . . satisfies the sentential function . . .*', '*true sentence*', and so on, by '*infinite sequence of subclasses of the class a*', '*the sequence . . . satisfies the sentential function . . . in the individual domain a*', '*correct sentence in the individual domain a*', and so on, respectively. As a consequence of Th. 8 the results so obtained can be extended to sentences which belong to the classes Ct_k. In this way we reach, among other things, the following generalizations of Ths. 4–6:

THEOREM 9. *For every cardinal number k the class Ct_k is a consistent and complete deductive system.*

THEOREM 10. *For every cardinal number k we have $Pr \subseteq Ct_k$, but $Ct_k \nsubseteq Pr$.*

In reference to Th. 10 the following problem presents itself: how is the list of axioms in Def. 13 to be completed, so that the class of all consequences of this extended class of axioms may coincide with the given class Ct_k? Ths. 11 and 12 which follow immediately below contain the solution of this problem and also prove that—with respect to the language of the calculus of classes —the definition of a correct sentence in a domain with k elements (Def. 26) can be replaced by another equivalent one which is analogous to the definition of provable sentence (Def. 17) and therefore has a structural character.

THEOREM 11. *If k is a natural number, and X the class consisting of all the axioms together with the sentences α and γ_k, then $Ct_k = Cn(X)$.*

THEOREM 12. *If k is an infinite cardinal number, and X the class consisting of all the axioms together with the sentence α and all sentences $\bar{\gamma}_l$ (where l is any natural number), then $Ct_k = Cn(X)$.*

The proof of these theorems is based chiefly on Ths. 9 and 10 and the three following lemmas:

LEMMA H. *For every cardinal number k we have $\alpha \in Ct_k$.*

LEMMA I. *If k is a natural number and l a cardinal number distinct from k, then $\gamma_k \overline{\in} Ct_k$ and $\gamma_k \in Ct_l$, but $\overline{\gamma_k} \overline{\in} Ct_k$ and $\overline{\gamma_k} \in Ct_l$.*

LEMMA K. *If $x \in S$ and X is the class consisting of all the axioms together with the sentence α, then there is a sentence y which is equivalent to the sentence x with respect to the class X and such that either y is a quantitative sentence, or $y \in Pr$ or $\bar{y} \in Pr$.*

Lemmas H and I are almost immediately evident, but the proof of the very important and interesting Lemma K is rather difficult.[1]

By means of Th. 9 and Lemma I it is possible from Th. 12 to derive the following consequence which combined with Th. 11 brings out the essential differences existing in the logical structure of the classes Ct_k according to whether the cardinal number k is finite or infinite:

THEOREM 13. *If k is an infinite cardinal number, then there is no class X which contains only a finite number of sentences which are not axioms, and also satisfies the formula*

$$Ct_k = Cn(X).$$[2]

From Lemma I and Ths. 11 and 12 we easily obtain the following consequences:

THEOREM 14. *If k is a natural number and l a cardinal number distinct from k, then $Ct_k \nsubseteq Ct_l$ and $Ct_l \nsubseteq Ct_k$.*

THEOREM 15. *If k and l are infinite cardinal numbers, then $Ct_k = Ct_l$.*

THEOREM 16. *If k is an infinite cardinal number and $x \in Ct_k$, then there is a natural number l such that $x \in Ct_l$ (in other words, the class Ct_k is included in the sum of all the classes Ct_l).*

According to Ths. 14–16 (or Lemma I) there exists for every natural number k a sentence which is correct in every domain

[1] In its essentials this lemma is contained in the results to be found in Skolem, Th. (64), pp. 29–37.

[2] The idea of the proof of this theorem is the same as that of the proofs of Ths. 24 and 25 in article V of the present volume, pp. 78–9. If we take over from the latter Def. 3, p .76, and at the same time extend our present concept of consequence by adding the words '*or x is an axiom*' to the condition (α) of Def. 15, then we could derive the following consequence from Ths. 11 and 13:

In order that the class Ct_k should be an axiomatizable deductive system, it is necessary and sufficient that k be a natural number.

with k elements and in no domain with any other cardinal number. On the other hand, every sentence which is correct in one infinite domain is also correct in every other infinite domain (without reference to its cardinal number) as well as in certain finite domains. From this we infer that the language in question allows us to express such a property of classes of individuals as their being composed of exactly k elements, where k is any natural number; but we find in this language no means by which we can distinguish a special kind of infinity (e.g. denumerability), and we are unable, either with the help of a single or of a finite number of sentences, to distinguish two such properties of classes as finiteness and infinity.[1]

By means of Ths. 9, 11, and 12 we can prove

THEOREM 17. *If X is a consistent class of sentences which contains all the axioms together with the sentence α, then there is a cardinal number k such that $X \subseteq Ct_k$; if X is a complete deductive system, then $X = Ct_k$.*

If we combine this theorem with Ths. 11 and 12, we obtain a structural description of all complete deductive systems which contain all the axioms and the sentence α. It should be noted that the presence of the sentence α is essential here; the multiplicity of the systems which do not contain this sentence is significantly greater and their exhaustive description would not at all be a simple matter.[2]

The remaining considerations concern sentences which are correct in every individual domain, i.e. belong to the class Ct.

[1] These results, as well as Th. 19 given below, we owe to Löwenheim; cf. Löwenheim, L. (49) (especially Th. 4, p. 459) and Skolem, Th. (64).

[2] I have occupied myself in the years 1926–8 with problems of this type, i.e. with the structural description of all complete systems of a given science, in application to various elementary deductive sciences (algebra of logic, arithmetic of real numbers, geometry of straight lines, theory of order, theory of groups); on the results of these investigations, reports were made in the seminar exercises in the methodology of the deductive sciences which I conducted in Warsaw University in the years 1927/8 and 1928/9. Cf. Presburger, M. (61) (especially note 4 on p. 95), and XII, § 5. For a detailed discussion of certain closely related problems (as well as for further bibliographical references) see also the recent publications of the author, Tarski, A. (84) and (84 *a*).

THEOREM 18. *In order that $x \in Ct$ it is necessary and sufficient that, for every cardinal number k, $x \in Ct_k$ (in other words, the class Ct is the product of all the classes Ct_k).*

This theorem, which is an immediate consequence of Def. 27 and Th. 8, can be essentially sharpened by means of Ths. 9 and 16:

THEOREM 19. *In order that $x \in Ct$ it is necessary and sufficient that, for every natural number k, $x \in Ct_k$.*

The correctness of a sentence in all finite domains thus entails its correctness in every individual domain.

The following two corollaries are derivable from Ths. 9, 14, and 18:

THEOREM 20. *For every cardinal number k we have $Ct \subseteq Ct_k$, but $Ct_k \nsubseteq Ct$.*

THEOREM 21. *The class Ct is a consistent but not a complete deductive system.*

THEOREM 22. *$Pr \subseteq Ct$, but $Ct \nsubseteq Pr$.*

This theorem follows from Ths. 10 and 18 and Lemma L:

LEMMA L. *$\alpha \in Ct$ but $\alpha \bar{\in} Pr$.*

That $\alpha \in Ct$ follows at once from Lemma H and Th. 18. The exact proof of the second part of the lemma is considerably more difficult.

THEOREM 23. *If x is a quantitative sentence then $x \bar{\in} Ct$.*

The proof, which is based on Lemma I, Th. 18, and Def. 32, offers no difficulties.

THEOREM 24. *If X is the class consisting of all the axioms together with the sentence α, then $Ct = Cn(X)$.*

This theorem is most easily proved with the help of Ths. 11, 12, and 18. By using Lemma K we obtain from it at once:

THEOREM 25. *If $x \in S$, $x \bar{\in} Ct$ and $\bar{x} \bar{\in} Ct$, then there is a quantitative sentence y, which is equivalent to the sentence x with respect to the class Ct.*

By reference to Lemma L and Th. 24 we notice that we have the following situation: the concept of a sentence which is correct in every individual domain has a larger extension than the concept of provable sentence, since the sentence α belongs to the extension of the first concept but not to that of the second. But if we increase the system of axioms by just this single sentence α, the two concepts become identical in extension. Because it seems to me desirable that, with respect to the calculus of classes, the concepts of theorem and of correct sentence in every individual domain should not be distinct in extension,[1] I would advocate the inclusion of the sentence α among the axioms of this science.

The problem still remains of clarifying the relation of the absolute concept of truth defined in Def. 23 to the concepts we have just investigated.

If we compare Defs. 22 and 23 with Defs. 24 and 25 and apply Th. 8, we easily obtain the following result:

THEOREM 26. *If a is the class of all individuals then $x \in Tr$ if and only if x is a correct sentence in the domain a; thus if k is the cardinal number of the class a, then $Tr = Ct_k$.*

As an immediate consequence of Ths. 20 and 26 we have:

THEOREM 27. *$Ct \subseteq Tr$, but $Tr \nsubseteq Ct$.*

If we bring together Ths. 26 and 14 or Ths. 11 and 12, we reach the conclusion that those assumptions of the metatheory which determine the cardinal number of the class of all individuals (and which do not intervene in the proof of Th. 26 itself) exert an essential influence on the extension of the term 'true sentence'. The extension of this term is different according to whether that class is finite or infinite. In the first case the extension even depends on how big the cardinal number of this class is.

[1] This tendency will be discussed in the next paragraph. It should be mentioned that Schröder, although beginning with other ideas, has made the suggestion of completing the system of hypotheses of the calculus of classes with the sentence α (and even with still other sentences which, however, as can easily be shown, follow in a simple way from the sentence α); cf. Schröder, E. (62), vol. 2, Part 1, pp. 318–49. In this connexion I may remark that it seems to me that the inclusion of the sentence α in the 'formal' system of the algebra of logic (of which the calculus of classes is an interpretation) would not be useful, for many interpretations of this system are known in which the sentence in question is not satisfied.

Because we can show, on the basis of the system of assumptions here adopted, that the class of all individuals is infinite, Th. 26 in combination with Th. 12 makes a structural characterization of true sentences possible:

THEOREM 28. *In order that $x \in Tr$, it is necessary and sufficient that x is a consequence of the class which consists of all the axioms together with the sentence α and all sentences $\bar{\gamma}_l$, where l is any natural number.*

This sentence could, in its form, obviously be regarded as a definition of true sentence. It would then be a purely structural definition, completely analogous to Def. 17 of provable theorem. But it must be strongly emphasized that the possibility of constructing a definition of such a kind is purely accidental. We owe it to the specific peculiarities of the science in question (to those peculiarities which, among others, have been expressed in Lemma K, which is the most essential premiss in the proof of Ths. 12 and 28) as well as—in some degree—to the strong existential assumptions adopted in the metatheory. On the other hand—in contrast to the original definition—we have here no general method of construction which could be applied to other deductive sciences.

It is worth noticing that by analysing the proof of Th. 28 and of the lemmas from which this theorem follows, we can obtain a general structural criterion of truth for all sentences of the language investigated. From Th. 28 such a criterion for quantitative sentences is easily derivable, and the proof of Lemma K allows us effectively to correlate with every sentence of the language a sentence which is equivalent to it and which, if it is not quantitative, is manifestly true or manifestly false. An analogous remark holds for the concept of correctness in a given, or in every, individual domain.

Summarizing the most important results obtained in this section we can say:

We have succeeded in doing for the language of the calculus of classes what we tried in vain to do for colloquial language: namely to construct a formally correct and materially adequate semantical definition of the expression 'true sentence'.

Moreover, by making use of the special peculiarities of the calculus of classes, we have been able to transform this definition into an equivalent structural definition which even yields a general criterion of truth for the sentences of the language of this calculus.

§ 4. THE CONCEPT OF TRUE SENTENCE IN LANGUAGES OF FINITE ORDER

The methods of construction which I have used in the previous section for the investigation of the language of the calculus of classes can be applied, without very important changes, to many other formalized languages, even to those with a considerably more complicated logical structure. In the following pages the generality of these methods will be emphasized, the limits of their applicability will be determined, and the modifications which they undergo in their various concrete applications will be briefly described.

It is by no means my intention, in these investigations, to consider all languages that can conceivably be imagined, or which any one at any time could or might wish to construct; such an attempt would be condemned to failure from the start. In what I shall say here I shall consider exclusively languages of the same structure as those which are known to us at the present day (in the perhaps unfounded conviction that they will form in the future, as they have done hitherto, a sufficient basis for the foundation of the whole of deductive knowledge). And even these languages show such great differences in their construction that their investigation in a perfectly general, but at the same time precise, way must encounter serious difficulties. These differences are, of course, rather of a 'calligraphical' nature. In some languages, for example, only constants and variables occur, in others it is not possible to avoid the use of so-called technical signs (brackets, points, and so on). In some languages symbols of an exactly specified form are used as variables, so that the form of the variables depends on the part they play

and their significance. In others quite arbitrary symbols may be used as variables, so long as they are distinguished by their form from the constants. In some languages every expression is a system of linearly ordered signs, i.e. signs following one another in a line, but in others the signs may lie at different levels, not only alongside but also below one another. This calligraphy of the language nevertheless exerts a fairly strong influence on the form of the constructions in the domain of the metalanguage, as will doubtless be seen from a brief survey of the preceding paragraphs.[1] For those reasons alone the following exposition will have the nature of a sketch; wherever it takes a more precise form, it is dealing with concretely described languages which are constructed in the same way as the language of the calculus of classes (i.e. languages without technical signs, with variables of an exactly specified form, with linear arrangement of the signs in every expression and so on).[2]

Before we approach our principal task—the construction of the definition of true sentence—we must undertake, in every concrete case, the construction of a corresponding metalanguage and the establishment of the metatheory which forms the proper field of investigation. A metalanguage which meets our requirements must contain three groups of primitive expressions: (1) expressions of a general logical kind; (2) expressions having the same meaning as all the constants of the language to be discussed or which suffice for the definition of such expressions (taking as a basis the rules of definition adopted in the meta-

[1] Cf., for example, p. 191, footnote.

[2] In order to give the following exposition a completely precise, concrete, and also sufficiently general form, it would suffice if we chose, as the object of investigation, the language of some one complete system of mathematical logic. Such a language can be regarded as a universal language in the sense that all other formalized languages—apart from 'calligraphical' differences—are either fragments of it, or can be obtained from it or from its fragments by adding certain constants, provided that the semantical categories of these constants (cf. below, pp. 215 ff.) are already represented by certain expressions of the given language. The presence or absence of such constants exerts, as we shall show, only a minimal influence on the solution of the problem in which we are interested. As such a language we could choose the language of the general theory of sets which will be discussed in § 5, and which might be enriched by means of variables representing the names of two- and of many-termed relations (of arbitrary semantical categories).

theory); (3) expressions of the structural-descriptive type which denote single signs and expressions of the language considered, whole classes and sequences of such expressions or, finally, the relations existing between them. That the expressions of the first group are indispensable is evident. The expressions of the second group enable us to translate every concrete sentence or, more generally, every meaningful expression of the language into the metalanguage, and those of the third group provide for the assignment of an individual name to every such expression. These last two circumstances taken together play an essential part in the final formulation of the desired definition. Corresponding to the three groups of primitive expressions, the full axiom system of the metatheory includes three groups of sentences: (1) axioms of a general logical kind; (2) axioms which have the same meaning as the axioms of the science under investigation or are logically stronger than them, but which in any case suffice (on the basis of the rules of inference adopted) for the establishment of all sentences having the same meaning as the theorems of the science investigated;[1] finally, (3) axioms which determine the fundamental properties of the primitive concepts of a structural-descriptive type. The primitive expressions and axioms of the first group (as well as the rules of definition and inference) may be taken from any sufficiently developed system of mathematical logic; the expressions and axioms of the second group naturally depend on the special peculiarities of the language investigated; for the third group suitable examples are provided in the presentation of § 2. It is to be noted that the two first groups of primitive

[1] It has already been mentioned (p. 167) that we are here interested exclusively in those deductive sciences which are not 'formal' in a quite special meaning of this word. I have, moreover, brought forward various conditions—of an intuitive not a formal nature—which are satisfied by the sciences here investigated: a strictly determinate and understandable meaning of the constants, the certainty of the axioms, the reliability of the rules of inference. An external characteristic of this standpoint is just the fact that, among the primitive expressions and the axioms of the metatheory the expressions and axioms of the second group occur. For as soon as we regard certain expressions as intelligible, or believe in the truth of certain sentences, no obstacle exists to using them as the need arises. This applies also to the rules of inference which we may, if need be, transfer from the theory to the metatheory. In the sequel we shall convince ourselves that this need actually exists in the cases given.

expressions and axioms partly overlap one another, and in those cases in which mathematical logic, or a fragment of it, is the object of investigation (as is the case with the calculus of classes), they even combine to form *one* group.

The establishment of the metatheory having been completed, our next task is to distinguish from the totality of expressions of the language the especially important category of *sentential functions* and in particular of *sentences*. The expressions of the language investigated consist of *constants* and *variables*. Among the constants, which are usually finite in number, we find, as a rule, certain signs belonging to the sentential calculus and the predicate calculus: for example the signs of negation, logical sum, logical product, implication, equivalence, as well as the universal and existential quantifiers, which we have already met in § 2. In addition to these we sometimes find other signs which are connected with the individual peculiarities of the language and denote concrete individuals, classes, or relations; such, for example, as the inclusion sign of the language of the calculus of classes, which denotes a particular relation between classes of individuals. Usually there are infinitely many variables. According to their form, and the interpretation of the language, they represent names of individuals, classes, or relations (sometimes there are also variables which represent sentences, i.e. the so-called sentential variables).[1] Among the expressions which are formed from the signs of both kinds, we distinguish first of all the *primitive sentential functions*, corresponding to the inclusions $\iota_{k,l}$ of the calculus of classes. The exact description of the form of the sentential functions and the specification of their intuitive sense will depend upon the special peculiarities of the language in question. In any case they are certain complexes of constants which are names of individuals, classes, or relations, and of variables which represent these

[1] In many languages various other categories of constants and variables occur, e.g. name-forming functors which, in combination with variables, form composite expressions by which names of individuals, classes, and relations are represented (e.g. the word 'father' in colloquial language, or the sign of complementation in the complete language of the calculus of classes—cf. p. 161, note 1, and p. 168, note 3 The languages considered in the present article contain no signs and expressions of this kind.

names. The first sign of such a complex is always the name of a class or a relation or a corresponding variable, and is called a (*sentence forming*) *functor of the given primitive sentential function*;[1] the remaining signs are called *arguments*, namely 1st, 2nd,..., kth argument—according to the place they occupy. For every constant and variable of the language studied—with the exception of the constants of the sentential calculus and the universal and existential quantifiers—a primitive function can be constructed which contains this sign (the sentential variables, even when they appear in the language, do not occur in the primitive functions as functors or arguments, but each is regarded as an independent primitive function). Next we introduce the *fundamental operations on expressions* by means of which composite expressions are formed from simpler ones. In addition to the operations of negation, logical addition and universal quantification, which we have met with in § 2 (Defs. 2, 3, and 6), we consider here other analogously defined operations, such as logical multiplication, formation of implications and equivalences, as well as existential quantification. Each of these operations consists in putting in front of the expression considered, or in front of two successive expressions (according to the kind of operation), either one of the constants of the sentential calculus which belongs to the language, or one of the two quantifiers together with the variables immediately following it. The expressions which we obtain from the primitive functions by applying to them any number of times and in any order any of

[1] Thus sentence-forming functors which have names as arguments are here identified with the names of classes or relations (in fact the one-argument functors with names of classes and the rest with names of two- or many-termed relations). This interpretation seems artificial with the interpretation of the term 'functor' which was given by some examples on p. 161, note 1; in any case it certainly does not agree with the spirit and formal structure of the language of everyday life. Without going into details, it seems to me for various reasons to be neither necessary nor useful to distinguish between these two categories of expressions (i.e. sentence-forming functors and names of classes or relations). Moreover, the whole question is rather of a terminological nature and is without influence on subsequent developments. We may either regard the definition of functor given in the text as purely formal and disregard the current interpretation of the term, or so extend the interpretation of terms like 'name of a class', 'name of a relation' that we include expressions which are not names in the usual sense.

the fundamental operations, we call sentential functions. Among the variables which occur in a given sentential function we can distinguish—by means of recursive definitions—*free* and *bound* variables. Sentential functions without free variables are called sentences (cf. Defs. 10–12 in § 2).

Next we define yet other concepts which are closely connected with the deductive character of the science under investigation, namely the concepts of *axiom*, *consequence*, and *theorem*. Among the axioms we include as a rule certain logical sentences which are constructed in a manner similar to that used for the first kind of axioms of the calculus of classes (cf. § 2, Def. 13). Moreover the definition of axiom depends wholly on the individual peculiarities of the science investigated, sometimes even on accidental factors which are connected with its historical development. In the definition of the concept of consequence we follow—*mutatis mutandis*—the pattern of § 2. The operations by means of which we form the consequences of a given class of sentences differ in no essential points from the operations which were given in Def. 15. The consequences of the axioms are called *provable sentences* or *theorems*.

After this preliminary work we turn now to our principal task —the construction of a correct definition of *true sentence*. As we saw in § 3, the method of construction available to us presupposes first a definition of another concept of a more general kind which is of fundamental importance for investigations in the semantics of language. I mean the notion of the *satisfaction of a sentential function by a sequence of objects*. In the same section I have attempted to clarify the customary meaning of this expression in its ordinary usage. I have pointed out that in drawing up a correct definition of the concept of satisfaction use can be made of recursive definition. For this purpose it suffices—recalling the recursive definition of sentential function and bearing in mind the intuitive sense of the primitive sentential functions and the fundamental operations on expressions—to establish two facts: (1) which sequences satisfy the fundamental functions, and (2) how the concept of satisfaction behaves under the application of any of the fundamental operations (or to put

it more exactly: which sequences satisfy the sentential functions which are obtained from given sentential functions by means of one of the fundamental operations, assuming that it has already been established which sequences satisfy the sentential functions to which the operation is applied). As soon as we have succeeded in making precise the sense of this concept of satisfaction, the definition of truth presents no further difficulty: the true sentences may be defined as those sentences which are satisfied by an arbitrary sequence of objects.

In carrying out the plan just sketched in connexion with various concrete languages we nevertheless meet with obstacles of a fundamental kind; in fact, just at the point where we try finally to formulate the correct definition of the concept of satisfaction. In order to make clear the nature of these difficulties a concept must first be discussed which we have not hitherto had an opportunity of introducing, namely the concept of *semantical category*.

This concept, which we owe to E. Husserl, was introduced into investigations on the foundations of the deductive sciences by Leśniewski. From the formal point of view this concept plays a part in the construction of a science which is analogous to that played by the notion of type in the system *Principia Mathematica* of Whitehead and Russell. But, so far as its origin and content are concerned, it corresponds (approximately) rather to the well-known concept of part of speech from the grammar of colloquial language. Whilst the theory of types was thought of chiefly as a kind of prophylactic to guard the deductive sciences against possible antinomies, the theory of semantical categories penetrates so deeply into our fundamental intuitions regarding the meaningfulness of expressions, that it is scarcely possible to imagine a scientific language in which the sentences have a clear intuitive meaning but the structure of which cannot be brought into harmony with the above theory.[1]

[1] Cf. Leśniewski, S. (46), especially pp. 14 and 68; Ajdukiewicz, K. (3), pp. 9 and 148. From the formal point of view the theory of semantical categories is rather remote from the original theory of types of Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 37 ff.; it differs less from the simplified theory of types (cf. Chwistek, L. (12), pp. 12–14; Carnap, R. (8), pp. 19–22) and is an extension of the latter. Regarding the views expressed in the last paragraph of the text, compare the Postscript to this article (p. 268).

For reasons mentioned at the beginning of this section we cannot offer here a precise structural definition of semantical category and will content ourselves with the following approximate formulation: two expressions *belong to the same semantical category* if (1) there is a sentential function which contains one of these expressions, and if (2) no sentential function which contains one of these expressions ceases to be a sentential function if this expression is replaced in it by the other. It follows from this that the relation of belonging to the same category is reflexive, symmetrical, and transitive. By applying the principle of abstraction,[1] all the expressions of the language which are parts of sentential functions can be divided into mutually exclusive classes, for two expressions are put into one and the same class if and only if they belong to the same semantical category, and each of these classes is called a semantical category. Among the simplest examples of semantical categories it suffices to mention the category of the sentential functions, further the categories which include respectively the names of individuals, of classes of individuals, of two-termed relations between individuals, and so on. Variables (or expressions with variables) which represent names of the given categories likewise belong to the same category.

In connexion with the definition of semantical category the following question arises: in order to establish the fact that two given expressions belong to one and the same semantical category, is it necessary to consider all possible sentential functions which contain one of the given expressions and to investigate their behaviour when one of these expressions is replaced by the other, or does it suffice to make this observation in some or even in only *one* case? From the standpoint of the ordinary usage of language the second possibility seems much more natural; in order that two expressions shall belong to the same semantical category, it suffices if there exists *one* function which contains one of these expressions and which remains a function when this expression is replaced by the other. This principle, which can be called the *first principle of the theory of semantical categories*, is

[1] Cf. Carnap, R. (8), pp. 48–50.

taken strictly as a basis for the construction of the formalized languages here investigated.[1] It is especially taken into account in the definition of the concept of sentential function. It also exerts an essential influence on the definition of the operation of substitution, i.e. one of those operations with the help of which we form the consequences of a class of sentences. For if we wish that this operation, when carried out on any sentence, should always give a new sentence as a result, we must restrict ourselves to substituting for the variables only those expressions which belong to the same semantical category as the corresponding variables.[2] Closely connected with this principle is a general law concerning the semantical categories of sentence-forming functors: the functors of two primitive sentential functions belong to the same category if and only if the number of arguments in the two functions is the same, and if any two arguments which occupy corresponding places in the two functions also belong to the same category. From this it follows that, in particular, no sign can be simultaneously a functor of two functions which possess a different number of arguments, or of two such functions (even if they possess the same number of arguments)

[1] When applied to concrete languages the formulations given in the text—both the definition of semantical category and the above-mentioned principle—require various corrections and supplementations. They are in any case too general, for they also include expressions to which we do not usually ascribe independent meaning, and which we often include in the same semantical categories to which meaningful expressions belong (for example, in the language of the calculus of classes, the expressions 'N', '$\Pi\, x_r$', and '$AIx_r x_s$' would belong to the same semantical category); in the case of these meaningless expressions, it can easily be shown that even the first principle of semantical categories loses its validity. This fact is of no essential importance for our investigations, for we shall apply the concept of semantical category, not to composite expressions, but exclusively to variables. On the other hand, the examples which we shall encounter in the sequel show that the above formulations admit of very far-reaching simplifications in concrete cases. Thanks to a suitable choice of the signs used in the construction of the expressions of the language, the mere shape of the sign (and even of the composite expression) decides to which category it belongs. Consequently it is possible that in methodological and semantical investigations concerning a concrete language, the concept of semantical category does not explicitly occur at all.

[2] In the language of the calculus of classes, and in the languages which I shall describe in more detail in the sequel, such expressions can only be other variables; this explains the formulation of Def. 14 in § 2.

in which two arguments which occupy corresponding places belong to different categories.

We require a classification of the semantical categories; to every category a particular natural number is assigned called the *order of the category*. This order is also assigned to all expressions which belong to this category.[1] The meaning of this term can be determined recursively. For this purpose we adopt the following convention (in which we have in mind only those languages which we shall deal with here and we take account only of the semantical categories of the variables): (1) the 1st order is assigned only to the names of individuals and to the variables representing them; (2) among expressions of the $n+1$th order, where n is any natural number, we include the functors of all those primitive functions all of whose arguments are of at most the nth order, where at least one of them must be of exactly the nth order. Thanks to the above convention all expressions which belong to a given semantical category have the same order assigned to them, which is therefore called the order of that category.[2] On the other hand the category

[1] Cf. Carnap, R. (8), pp. 31–32.

[2] This classification by no means includes all semantical categories which are to be found in formalized languages. For example, it does not include sentential variables and functors with sentences as arguments—i.e. signs which occur in the sentential calculus—neither does it include functors which, together with the corresponding arguments, form expressions which belong to one of the categories distinct from sentential functions, such as the name-forming functors mentioned on p. 213, footnote.

In view of this, the definition of order given in the text could be widened in the following way: (1) to the 1st order belong sentences, names of individuals and expressions representing them; (2) among expressions of the $n+1$th order we include those functors with an arbitrary number of arguments of order $\leqslant n$, which together with these arguments form expressions of order $\leqslant n$, but are not themselves expressions of the nth order. Even this definition does not yet cover all meaningful expressions which occur in the deductive sciences. No signs which 'bind' variables fall under this definition (thus such signs as the universal and existential quantifiers, the signs 'Σ' and 'Π' of the theory of sets and analysis or the sign of integration), signs which—in contrast to the functors—can be called *operators*. (von Neumann speaks of *abstractions* in this connexion, see Neumann, J. v. (54).) On the other hand the latter classification is completely adapted to the system invented by Leśniewski and sketched by him in Leśniewski (46) and (47). This system contains no operators except the universal quantifier which belongs to no semantical category. I may add that, in my view, the lack of operators in Leśniewski's system constitutes a deficiency which restricts its 'universal' character (in the sense of p. 210, note 2) to a certain degree.

is by no means specified by the order: every natural number which is greater than 1 can be the order of many different categories. Thus, for example, both the names of classes of individuals and the names of two-, three-, and many-termed relations between individuals are expressions of the 2nd order.

It is desirable to classify the sentential functions of the language according to the semantical categories of the free variables occurring in them. We shall say of two functions that they *possess the same semantical type* if the number of free variables of every semantical category in the two functions is the same (or, in other words, if the free variables of the one function can be put into one-one correspondence with the free variables of the other in such a way that to every variable a variable of the same category corresponds). The class of all sentential functions which possess the same type as a given function we can call a *semantical type*.

We sometimes use the term 'semantical category' in a derivative sense, by applying it, not to the expressions of the language, but to the objects which they denote. Such 'hypostatizations' are not quite correct from a logical standpoint, but they simplify the formulation of many ideas. We say, for example, that all individuals belong to the same semantical category, but that no classes or relations belong to this category. From the general law stated above concerning sentence-forming functors we conclude that two classes belong to the same category if and only if all their elements belong to one and the same category. Two two-termed relations belong to the same category if and only if their domains belong to the same category and their counter domains belong to the same category. In particular, two sequences belong to the same category if and only if all their terms belong to the same category. A class and a relation, or two relations having different numbers of terms never belong to the same category. It also follows that there can be no class whose elements belong to two or more semantical categories; in an analogous way there can be no sequence whose terms belong to distinct semantical categories. Individuals are sometimes called objects of the 1st order,

classes of individuals and relations between individuals objects of the 2nd order, and so on.

The language of a complete system of logic should contain—actually or potentially—all possible semantical categories which occur in the languages of the deductive sciences. Just this fact gives to the language mentioned a certain 'universal' character, and it is one of the factors to which logic owes its fundamental importance for the whole of deductive knowledge. In various fragmentary systems of logic, as well as in other deductive sciences, the multiplicity of the semantical categories may undergo a significant restriction in both their number and their order. As we shall see, the degree of difficulty which we have to overcome in the construction of a correct definition of truth for a given concrete language, depends in the first place on this multiplicity of the semantical categories appearing in the language, or, more exactly, on whether the expressions and especially the variables of the language belong to a finite or an infinite number of categories, and in the latter case on whether the orders of all these categories are bounded above or not. From this point of view we can distinguish four kinds of languages: (1) languages in which all the variables belong to one and the same semantical category; (2) languages in which the number of categories in which the variables are included is greater than 1 but finite; (3) languages in which the variables belong to infinitely many different categories but the order of these variables does not exceed a previously given natural number n; and finally (4) languages which contain variables of arbitrarily high order. We shall call the languages of the first three kinds *languages of finite order*, in contrast to languages of the fourth kind, the *languages of infinite order*. The languages of finite order could be further divided into languages of the 1st, 2nd order, and so on, according to the highest order of the variables occurring in the language. By way of supplementation of the sketch given at the beginning of this section of the construction of a metatheory, it must be noted here that the metalanguage, on the basis of which the investigation is conducted, is to be furnished with at least all the

semantical categories which are represented in the language studied. This is necessary if it is to be possible to translate any expression of the language into the metalanguage.[1]

From the point of view of their logical structure the languages of the 1st kind are obviously the simplest. The language of the calculus of classes is a typical example. We have seen in § 3 that for this language the definition of the satisfaction of a sentential function by a sequence of objects, and hence the definition of true sentence, presents no great difficulties. The method of construction sketched there can be applied as a whole to other languages of the 1st order. It is clear that in doing this certain small deviations in detail may occur. Among other things it may be necessary to operate not with sequences of classes but with sequences of other kinds, e.g. with sequences of individuals or relations, according to the intended interpretation and the semantical categories of the variables occurring in the language.[2]

A particularly simple example of a language of the 1st kind which is worthy of attention is the language of the ordinary sentential calculus enlarged by the introduction of the universal and existential quantifiers. The simplicity of this language lies, among other things, in the fact that the concept of variable coincides with that of primitive sentential function. In the metatheory of the sentential calculus two different definitions can be given of provable theorem, the equivalence of which is in no way evident: the one is based on the concept of consequence and is analogous to Defs. 15–17 of § 2, the second is connected with the concept of the two-valued matrix. By virtue of this second definition we can easily determine whether any sentence is provable provided its structure is known.[3] If we now construct for this language a definition of true sentence strictly according

[1] Here—*mutatis mutandis*—the remarks of p. 211, footnote, also apply.

[2] Certain complications, which I shall not discuss here, arise if in addition to variables, composite expressions of the same semantical category also occur in the language investigated; the complete language of the calculus of classes which was mentioned on p. 168, note 3, will serve as an example, or the language of a system of arithmetic investigated in Presburger, M. (61) (cf. also p. 212, footnote).

[3] Cf. Hilbert, D., and Ackermann, W. (30), pp. 84–85; Łukasiewicz, J. (51), pp. 154 ff.; IV, § 4.

to the pattern given in § 3, we can easily convince ourselves that it represents a simple transformation of the second of these definitions of provable sentence, and thus the two terms 'provable theorem' and 'true sentence' in this case have the same extension. This fact provides us, among other things, with a general structural criterion for the truth of the sentences of this language. The method of construction laid down in the present work could thus be regarded, in a certain sense, as a generalization of the matrix method familiar in investigations on the sentential calculus.

Serious difficulties only arise when we consider languages of more complicated structure, e.g. languages of the 2nd, 3rd, and 4th kinds. We must now analyse these difficulties and describe the methods which enable us at least partially to overcome them. In order to make the exposition as clear and precise as possible I shall discuss in somewhat greater detail some concrete formalized languages, one of each kind. I shall try to choose examples which are as simple as possible, are free from all less essential, subordinate complications, and are at the same time sufficiently typical to exhibit the difficulties mentioned to the fullest extent and in the most striking form.

The language of the *logic of two-termed relations* will serve as an example of a language the 2nd order.[1] The only constants of this language are: the sign of negation 'N', the sign of logical sum 'A' and the universal quantifier '$\prod$'. As variables we can use the signs '$x_{'}$', '$x_{''}$', '$x_{'''}$', ... and '$X_{'}$', '$X_{''}$', '$X_{'''}$', ... The sign composed of the symbol 'x' and of k small additional strokes is called the *k-th variable of the 1st order*, and is denoted by the symbol 'v_k'. The sign analogously constructed with the symbol 'X' is called the *k-th variable of the 2nd order*, symbolically 'V_k'. The variables of the 1st order represent names of individuals, those of the 2nd order names of two-termed relations between individuals. From the material and also—in agreement with the further description of the language—from the formal

[1] This is a fragment of the language of the algebra of relations, the foundations of which are given in Schröder, E. (62), vol. 3—a fragment which nevertheless suffices to express every idea which can be formulated in this language.

point of view, the signs 'v_k' and 'V_k' belong to two distinct semantical categories. Expressions of the form 'Xyz' are regarded as primitive sentential functions, where in the place of 'X' any variable of the 2nd order, and in the place of 'y' and 'z' any variables of the 1st order may appear. These expressions are read: 'the individual y stands in the relation X to the individual z' and they are denoted—according to the form of the variables—by the symbols '$\rho_{k,l,m}$'. By the use of the sign '$\frown$' from § 2 we specify that $\rho_{k,l,m} = (V_k \frown v_l) \frown v_m$. The definitions of the fundamental operations on expressions, as well as those of sentential function, sentence, consequence, provable sentence, and so on, are all quite analogous to the definitions of § 2. But it must always be borne in mind that in this language two distinct categories of variables appear and that the expressions $\rho_{k,l,m}$ play the part of the inclusions $\iota_{k,l}$. In connexion with the first of these facts we have to consider not *one* operation of quantification (Defs. 6 and 9) but *two* analogous operations: with respect to a variable of the 1st order as well as with respect to a variable of the 2nd order, the results of which are denoted by the symbols '$\bigcap'_k x$', and '$\bigcap''_k x$' or '$\bigcup'_k x$' and '$\bigcup''_k x$' respectively. Correspondingly there will be two operations of substitution. Among the axioms of the logic of relations we include the sentences which satisfy the condition (α) of Def. 13, i.e. substitutions of the axioms of the sentential calculus, and universal quantifications of these substitutions, and also all sentences which are universal quantifications of expressions of the type

$$\bigcup''_k \bigcap'_l \bigcap'_m (\rho_{k,l,m} \cdot y + \overline{\rho_{k,l,m}} \cdot \bar{y}),$$

where k, l, and m are any natural numbers ($l \neq m$) and y any sentential function in which the free variable V_k does not occur. Considering their intuitive meaning the axioms of the last category may be called *pseudodefinitions*.[1]

[1] This term we owe to Leśniewski, who has drawn attention to the necessity of including pseudodefinitions among the axioms of the deductive sciences in those cases in which the formalization of the science does not admit the possibility of constructing suitable definitions (cf. p. 166, footnote). Pseudodefinitions can be regarded as a substitute for the *axiom of reducibility* of Whitehead, A. N., and Russell, B. A. W. (90), vol. 1, pp. 55 ff. It would not be difficult to show the connexion between these sentences and a group of axioms adopted in Neumann, J. v. (54), p. 18.

To obtain a correct definition of satisfaction in connexion with the language we are considering we must first extend our knowledge of this concept. In the first stage of operating with it we spoke of the satisfaction of a sentential function by one, two, three objects, and so on, according to the number of free variables occurring in the given function (cf. pp. 189 ff.). From the semantical standpoint the concept of satisfaction had there a strongly ambiguous character; it included relations in which the number of terms was diverse, relations whose last domain was a class of sentential functions, whilst the other domains—in the case of the language of the calculus of classes—consisted of objects of one and the same category, namely classes of individuals. Strictly speaking we were dealing not with *one* concept, but with an infinite number of analogous concepts, belonging to different semantical categories. If we had formalized the metalanguage it would have been necessary to use infinitely many distinct terms instead of the *one* term 'satisfies'. The semantical ambiguity of this concept increases still more when we pass to languages of more complicated logical structure. If we continue the intuitive considerations of § 3, analyse the examples given there and construct new ones after the same pattern, it soon becomes clear that a strict semantical correlation exists between the free variables of the sentential function and the objects which satisfy these functions: every free variable belongs to the same semantical category as the name of the object corresponding to it. If, therefore, at least two different categories occur among the variables of the language—as in the case we are investigating—it does not suffice to restrict consideration to only a single category of objects in dealing with the concept of satisfaction. The domains of the single relations which are covered by the term 'satisfaction', thus cease to be semantically unambiguous (only the last domain consists as before exclusively of sentential functions). But since the semantical category of a relation not only depends on the number of domains, i.e. the number of terms standing in the relation to one another, but also on the categories of these domains, the category of the concept of satisfaction, or rather the category of each single one of these

concepts, also depends on two circumstances. It depends on the number and also on the categories of the free variables which appear in the sentential functions to which the concept of satisfaction relates. In brief, it depends on what we have called the semantical type of the sentential function. To functions which belong to two distinct types two semantically distinct concepts of satisfaction always correspond.[1] Some examples will make this clear. We shall say that the objects R, a, and b satisfy the function $\rho_{1,2,3}$ if and only if R is a relation and a and b are individuals and we have aRb (i.e. a stands in the relation R to b). The function $\rho_{1,2,2} \cdot \rho_{3,2,2}$ is satisfied by the objects R, a, and S if and only if R and S are relations, a is an individual and we have both aRa and aSa. The function $\bigcap'_2 \bigcap'_3 (\overline{\rho_{1,2,3}} + \rho_{1,3,2})$ is satisfied by symmetrical relations and only by them, i.e. by relations such that, for all individuals a and b, if we have aRb we also always have bRa. The function $\bigcap'_1 (\overline{\rho_{1,2,3}} + \rho_{1,3,2})$ is satisfied by those and only those individuals a and b which satisfy the following condition: for every relation R, if aRb, then bRa, i.e. individuals which are identical. In the above examples we have sentential functions belonging to four different semantical types, and we are, therefore, dealing with four different relations of satisfaction, in spite of the fact that the number of free variables and also the number of terms in the relations is the same in the first two examples.

The semantical ambiguity attaching to the concept of satisfaction in its original conception renders an exact characterization of this concept in a single sentence, or even in a finite number of sentences, impossible, and so denies us the use of the only method so far known to us of constructing a definition of a true sentence. In order to avoid this ambiguity, in dealing with the calculus of classes we had recourse to an artifice which is used by logicians and mathematicians in similar situations. Instead of using infinitely many concepts of satisfaction of a sentential

[1] Moreover, functions of *one* semantical type can correspond to several semantically distinct concepts of satisfaction, provided the free variables of these functions belong to at least two distinct semantical categories; in addition to the number and the categories of the variables their arrangement must also be taken into consideration.

function by single objects, we tried to operate with the semantically uniform, if somewhat artificial, concept of the satisfaction of a function by a sequence of objects. It happened that this concept is sufficiently more general than the previous one to include it—intuitively speaking—as a special case (to define the logical nature of this inclusion would, however, be a little difficult). It will easily be seen that this method cannot be applied to the present problem without further difficulty. Satisfaction in its new form is a two-termed relation, whose domain consists of sequences and counter domain of sentential functions. As before, there exists between the free variables of a sentential function and the corresponding terms of the sequences which satisfy it, a strict semantical correlation. Thus if the language of the logic of relations contains variables of two different semantical categories, we must likewise use two categories of sequences in our investigations. For example, the function $\bigcap_2' \bigcap_3'(\overline{\rho_{1,2,3}}+\rho_{1,3,2})$ is satisfied exclusively by sequences of two-termed relations between individuals (namely by those and only those sequences F whose first term F_1 is a symmetrical relation). But the function $\bigcap_1''(\overline{\rho_{1,2,3}}+\rho_{1,3,2})$ is satisfied exclusively by sequences of individuals (i.e. by sequences f for which $f_2 = f_3$ holds). The domain of the relation of satisfaction and *eo ipso* the relation itself thus again becomes semantically ambiguous. Again we are dealing not with *one*, but with at least two different concepts of satisfaction. But still worse, a closer analysis shows that the new interpretation of the concept of satisfaction can no longer as a whole be maintained. For one and the same sentential function often contains free variables of two different categories. To deal with such functions we must operate with sequences whose terms likewise belong to two categories. The first term, for example, of the sequence which satisfies the function $\rho_{1,2,3}$ must be a relation, but the two following ones must be individuals. But it is known that the theory of semantical categories does not permit the existence of such heterogeneous sequences. Consequently the whole conception collapses. Thus changing the original interpretation of the concept of satisfaction has removed only *one* subsidiary cause of its ambiguity, namely the

diversity in the number of terms in the relations which are the object of the concept; another far more important factor, the semantical diversity of the terms of the relations, has lost none of its force.

Nevertheless the methods used in § 3 can be applied to the language now being investigated, although with certain modifications. In this case also it is possible to find an interpretation of the concept of satisfaction in which this notion loses its semantical ambiguity and at the same time becomes so general that it includes all special cases of the original concept. In fact, two different methods are available; I shall call them the *method of many-rowed sequences* and the *method of semantical unification of the variables*.

The first method requires that we should treat satisfaction not as a two-termed, but as a three-termed relation which holds between sequences of individuals, between sequences of two-termed relations and between sentential functions. We use the following mode of expression: 'the sequence f of individuals and the sequence F of relations together satisfy the sentential function x'. The content of this phrase can easily be visualized by means of concrete examples. For example, the sequence f of individuals and the sequence F of relations together satisfy the function $\rho_{1,2,3}$ if and only if the individual f_2 stands in the relation F_1 to the individual f_3. In order to formulate a general definition we proceed exactly in the manner of Def. 22 in § 3, care being taken to remember that, in the language we are considering, the expressions $\rho_{k,l,m}$ play the part of primitive sentential functions and that instead of *one* operation of universal quantification *two* related operations occur. The definition of true sentence is completely analogous to Def. 23.

This method can now be modified to some extent by treating satisfaction as a two-termed relation between so-called two-rowed sequences and sentential functions. Every ordered pair which consists of two sequences f and F is called a *two-rowed sequence* (or *two-rowed matrix*), where the kth term of the sequence f or of the sequence F is called the kth term of the first or second row respectively of the two-rowed sequence. In the present

case we have to deal with ordered pairs which consist of a sequence of individuals and a sequence of relations. It is easily seen that this modification is a purely formal one and has no essential effect on the construction as a whole. It is to this modification of the method that the term 'method of many-rowed sequences' is adapted.

To understand the method of semantical unification of the variables we begin with certain considerations which are not immediately connected with the language we are at present investigating. It is known that with every individual a a definite two-termed relation a^* can be correlated in such a way that to distinct individuals distinct relations correspond. For this purpose it suffices to take as a^* an ordered pair whose terms are identical with a, i.e. the relation R which holds between any two individuals b and c if and only if $b = a$ and $c = a$. On the basis of this correlation we can now correlate in a one-one fashion with every class of individuals a class of relations, with every many-termed relation between individuals a corresponding relation between relations, and so on. For example, to any class A of individuals there corresponds a class A^* of all those relations a^* which are correlated with the elements a of the class A. In this way every sentence about individuals can be transformed into an equivalent sentence about relations.

Bearing these facts in mind we return to the language of the logic of relations and change the intuitive interpretation of the expressions of this language without in any way touching their formal structure. All constants will retain their previous meaning, whilst all variables both of the 1st and 2nd order are from now on to represent names of two-termed relations. To the primitive sentential functions of the type 'Xyz', where instead of 'X' some variable V_k and instead of 'y' and 'z' any two variables v_l and v_m occur, we assign the following meaning: 'there exist individuals a and b such that a stands in the relation X to $b, y = a^*$, and $z = b^*$.' In this way the meaning of the composite sentential functions will likewise be modified. It is almost immediately evident that every true or false sentence in the earlier interpretation will remain true or false respectively in

the new one. By virtue of this new interpretation all the variables of the language now belong to one and the same semantical category, not indeed from the formal but from the intuitive point of view; they represent words of the same 'part of speech'. Consequently the language we are considering can be investigated by exactly the same methods as all languages of the 1st kind; in particular, satisfaction can be treated as a two-termed relation between sequences of relations and sentential functions. At the same time a complication of a technical nature—although an unimportant one—presents itself. Since two free variables of different orders but the same indices, e.g. v_l and V_l, may occur in the same sentential function, it is not clear without supplementary stipulations which terms of the sequence are to correspond to the variables of the 1st, and which to those of the 2nd order. To overcome this difficulty we shall stipulate that to every variable v_k a term of the sequence with an uneven index $2.k-1$ corresponds, and to every variable V_k a term with even index $2.k$ corresponds. For example, the sequence F of relations satisfies the function $\rho_{k,l,m}$ if and only if there are individuals a and b such that a stands in the relation $F_{2.k}$ to b, $F_{2.l-1} = a^*$, and $F_{2.m-1} = b^*$. Apart from this detail the definitions of satisfaction and of true sentence differ in no essential point from the definitions given in § 3.

The two methods described can be applied to all languages of the 2nd kind.[1] If the variables of the language studied belong to n different semantical categories, we regard satisfaction—under the method of many-rowed sequences—as an $n+1$-termed relation holding between n sequences of the corresponding categories and the semantical functions, or as a two-termed relation whose domain consists of n-rowed sequences (i.e. ordered

[1] This holds even for languages in which variables occur which are not included in the classification on p. 218 (cf. p. 218, note 2). I shall not deal with certain (not particularly important) difficulties which may occur here. But I take this opportunity of mentioning that sentential variables, even if they occur in the language, do not complicate the construction at all, and that, in particular, it would not be worth while to include them in the process of semantical unification. Sentences which contain such variables can be excluded by correlating with each of them, in one-many fashion, an equivalent sentence which does not contain sentential variables (cf. Hilbert, D., and Ackermann, W. (30), pp. 84–85).

n-tuples of ordinary sequences) and whose counter domain consists of sentential functions. Constructions based on this method form the most natural generalization of the constructions in § 3 and their material correctness appears to leave no doubts.

In applying the method of semantical unification of the variables, the choice of the *unifying category* plays an essential part, i.e. that semantical category in which all the variables of the language studied can be interpreted. Only one thing is required of the unifying category: that with all objects of every semantical category which is represented by the variables of the given language, effective objects of the chosen category can be correlated in a one-one fashion (i.e. so that to distinct objects, distinct objects correspond). Nevertheless, the choice of the unifying category is not always so simple as in the example discussed above in connexion with the language of the logic of relations; this choice cannot always be made from the categories which occur in the language. If, for example, the variables of the language represent names of two-termed relations between individuals and names of classes which consist of classes of individuals, then the simplest unifying category seems to be the category of two-termed relations between classes of individuals. I do not propose to enter into a further analysis of this problem (it would presuppose a knowledge of certain facts belonging to set theory). I add only the following remarks: (1) the unifying category cannot be of lower order than any one category among those occurring in the language; (2) for every language of the 2nd kind a unifying category can be found, even infinitely many such categories and in fact among categories of the nth order, where n is the highest order of the variables occurring in the language. As soon as the unifying category is specified, and the primitive sentential functions correspondingly interpreted, the further course of the work does not differ at all from the methods of construction used for languages of the 1st kind.

In contrast to the method of many-rowed sequences, there is no doubt that the second method is somewhat artificial. Nevertheless the definitions constructed by this method prove, on

closer analysis, to be intuitively evident to a scarcely less degree than the constructions based on the first method. At the same time they have the advantage of greater logical simplicity. In particular, when we are dealing with the definition of true sentence the proof of the equivalence of its two formulations presents no difficulty in any concrete case. The essential advantages of the method of unification of the variables only become clear, however, in the investigation of languages of the 3rd kind, since the method of many-rowed sequences here proves to be quite useless.

As a typical example of a language of the 3rd kind we choose the language of the *logic of many-termed relations*.[1] In this science we deal with the same constants 'N', 'A', and '$\prod$' and with the same variables of the 1st order v_k, as in the logic of two-termed relations. But we also find here variables of the 2nd order in greater multiplicity than before. As variables of this kind we shall use such signs as 'X'_r', 'X'_n', 'X'_m',..., 'X''_r', 'X''_n', 'X''_m',..., 'X'''_r', 'X'''_n', 'X'''_m',... and so on. The composite symbol constructed from the sign 'X' with k small strokes below and l such strokes above will be called the *kth variable functor with l arguments*, and denoted by 'V^l_k'. Intuitively interpreted, the variables v_k represent, as before, names of individuals, whilst the variables V^l_k represent names of l-termed relations between individuals, in particular for $l = 1$ names of one-termed relations, i.e. names of classes. Both from the intuitive and the formal points of view the signs v_k, V^1_k, V^2_k,... belong to infinitely many distinct semantical categories of the 1st and 2nd orders respectively. The fundamental sentential functions are expressions of the type '$Xxy...z$', where in place of 'X' any variable functor with l arguments and in place of 'x', 'y',..., 'z' variables of the 1st order, l in number, occur. These expressions are read as follows: 'the l-termed relation X holds between the l individuals $x, y,..., z$.' According to the number and form of the variables we denote the primitive functions by the symbols '$\rho_{k,m}$', '$\rho_{k,m,n}$',...,

[1] This is a language which resembles the language of the lower predicate calculus of Hilbert, D., and Ackermann, W. (30), pp. 43 ff., but is richer than the latter because variable functors can occur in it both as free and as bound variables.

putting $\rho_{k,m} = V^1_k \frown v_m$, $\rho_{k,m,n} = (V^2_k \frown v_m) \frown v_n$, and so on. In order to obtain a unified symbolism, which is independent of the number of variables, we shall use symbols of the type '$\rho^l_{k,p}$' (where 'p' represents the name of a finite sequence of natural numbers), the meaning of which is determined by the formula $\rho^l_{k,p} = (((V^l_k \frown v_{p_1}) \frown v_{p_2}) \frown \ldots) \frown v_{p_l}$.[1] The further definitions of the metatheory do not differ at all from the analogous definitions relating to the logic of two-termed relations and even to the calculus of classes. As operations of quantification we introduce quantification with respect to the variables v_k and the variables V^l_k and denote the result of the operations by the symbols '$\bigcap_k x$' and '$\bigcap^l_k x$' respectively. The list of axioms includes those which satisfy the condition (α) of Def. 13 of § 2, and pseudodefinitions which form a natural generalization of the pseudodefinitions from the logic of two-termed relations. Their more detailed description seems to be unnecessary.

We turn now to the problem of how the concept of satisfaction is to be conceived and the definition of truth to be constructed for the language we are now considering. Any attempt to apply the method of many-rowed sequences in this case fails completely. In this method the term 'satisfaction'—in whatever form—expresses the relation of dependence between n sequences of various categories and the sentential functions, where n is exactly equal to the number of semantical categories represented by the variables of the given language. In the case we are investigating the number n is indefinitely large and the metalanguage we are using—like all other actually existing formalized languages—provides no means for dealing with the relation of mutual dependence between objects which belong to infinitely many distinct semantical categories.[2]

[1] Strictly speaking the meaning of the symbol '$\rho^l_{k,p}$' should be defined recursively.

[2] In those cases in which, in logical and mathematical constructions, we deal with the mutual dependence between an arbitrary, not previously determined number of objects of one and the same semantical category, we mostly use ordinary sequences. For objects which belong to a finite number of distinct categories many-rowed sequences fulfil the analogous function. But on the basis of the known languages we find nothing like 'sequences with infinitely many rows' (of distinct semantical categories).

The method of semantical unification of the variables can, however, be applied to this language with complete success. To see this it suffices to note that we can correlate in a one-one fashion, with every n-termed relation R between individuals, a class R^* which consists of n-termed sequences of individuals, namely the class of all sequences f which satisfy the following condition: the relation R holds between the individuals $f_1, f_2, \ldots, f_n$. For example, the class of all sequences f with two terms f_1 and f_2 such that $f_1 R f_2$ corresponds to the two-termed relation R. Consequently every sentence concerning many-termed relations can be transformed into an equivalent sentence which asserts something about classes of sequences. It will be remembered that by sequences of individuals we mean two-termed relations between individuals and natural numbers. Accordingly all sequences of individuals, whatever the number of their terms, belong to one and the same semantical category and therefore the classes of these sequences, in contrast to many-termed relations, likewise belong to one and the same category.

On the basis of these considerations we now partially unify the semantical categories of the variables in the following way. To the variables v_k we give—at least provisionally—the same significance as before. But the variables V^l_k now represent the names of any classes which consist of finite sequences of individuals or of other objects of the same category (i.e. the names of objects of at least the 3rd order, according to the order which we assign to the natural numbers).[1] The primitive functions of the form '$Xxy\ldots z$', which begin with a functor with l arguments and hence contain l variables of the 1st order, are interpreted by phrases of the type: 'the sequence of individuals the first term of which is x, the second y,... and the lth (the last) is z, belongs to

[1] In systems of mathematical logic, e.g. in Whitehead, A. N., and Russell, B. A. W. (90), vol. 2, pp. 4 ff., the cardinal numbers and in particular the natural numbers are usually treated as classes consisting of classes of individuals (or other objects), namely as the classes of all those classes which are similar (in the *Principia Mathematica* sense) to a given class. For example, the number 1 is defined as the class of all those classes which have exactly *one* element. With this conception the natural numbers are thus objects of (at least) the 3rd, sequences of individuals of the 4th, and classes of these sequences of the 5th order.

the class X which consists of l-termed sequences'. From the intuitive, although not from the formal, standpoint, the variables from now on still belong to only two distinct semantical categories; in view of this circumstance we can use, in the further course of our work, the same methods as we employed in investigating languages of the 2nd kind.

By means of the phrase: 'the sequence f of individuals and the sequence F, whose terms form classes of finite sequences of individuals, together satisfy the given sentential function', we can bring into service the method of many-rowed sequences. To use this concept consistently we must first set up a one-one correlation between the variables V_k^l and the terms of the sequence F in such a way that terms with different indices correspond to different variables. This is most easily done by putting every variable V_k^l in correspondence with a term having the index $(2.k-1).2^{l-1}$. For example, the terms F_1, F_3, F_5, F_2, F_6, F_4,... correspond to the variables V_1^1, V_2^1, V_3^1, V_1^2, V_2^2, V_1^3,... .[1] With this convention the establishment of the meaning of the above phrase in its application to any concrete sentential function, and even the construction of a general definition of the concept in question, presents no further difficulties. Thus concerning the primitive functions, those and only those sequences f and F (of the categories given above) will together satisfy the function $\rho_{k,m}$ which satisfy the following condition: the sequence g of individuals, whose single term g_1 is identical with f_n, belongs to the class $F_{2.k-1}$. In an analogous way, those functions f and F will together satisfy the function $\rho_{k,m,n}$ which satisfy the following condition: the sequence g of individuals with two terms, where $g_1 = f_m$ and $g_2 = f_n$, belongs to the class $F_{(2.k-1).2}$. In general, in order that the sequences f and F should together satisfy the function $\rho_{k,p}^l$, it is necessary and sufficient that the sequence g of individuals with l terms, where $g_1 = f_{p_1}$, $g_2 = f_{p_2}$,..., $g_l = f_{p_l}$, should belong to the class $F_{(2.k-1).2^{l-1}}$ (which consists of sequences with the same number of terms).

[1] Instead of the function $f(k, l) = (2.k-1).2^{l-1}$ we could use any other function $f(k, l)$ which correlates the natural numbers in one-one fashion with ordered pairs of natural numbers. Set theory offers many examples of such correlations; cf. Fraenkel, A. (16), pp. 30 ff. and 96 ff.

If we wish to apply the method of unification of the variables we again make use of the fact that a one-one correlation can be set up between any individuals and certain classes of finite sequences, and in such a way that to every individual a there corresponds the class a^* containing as its only element a sequence whose only member is just the given individual. Beginning in this way we next modify the interpretation of the variables of 1st order in exactly the same direction in which we formerly modified the interpretation of the variables of the 2nd order. The primitive functions of the form '$Xxy...z$', containing $l+1$ signs, we now regard as having the same meaning as expressions of the type 'the l-termed sequence g of individuals which satisfies the conditions: $g_1^* = x$, $g_2^* = y$,..., $g_l^* = z$, belongs to the class X, which consists of sequences with l terms'. With this intuitive interpretation all variables now belong to the same semantical category. The further construction contains no essentially new features and the reader will encounter no serious difficulties in carrying it out.

The method of semantical unification of the variables can be applied with equal success to the investigation of any language of the 3rd kind. Determining the unifying category may sometimes be more difficult. As in the case of languages of the 2nd kind it is here impossible to restrict consideration to categories occurring in the language studied. In contrast to those languages it is never possible to make the choice from among the categories of one of the orders represented in the language. This difficulty is not, however, essential and exclusively concerns languages of the lowest order. For it is possible to prove that for those languages in which the order of the variables does not exceed a given number n, where $n > 3$, any category of the nth order can serve as the unifying category.

In this way the various methods at our disposal enable us to define the concept of satisfaction and with it to construct a correct definition of truth for any language of finite order. We shall see in the next section that these methods do not extend further; the totality of languages of finite order exhausts the domain of applicability of our methods. This is therefore the place in which

to summarize the most important consequences which follow from the definitions we have constructed.†

First, *the definition of true sentence is a correct definition of truth in the sense of convention* T of § 3. It embraces, as special cases, all partial definitions which were described in condition (α) of this convention and which elucidate in a more precise and materially correct way the sense of expressions of the type 'x is a true sentence'. Although this definition alone provides no general criterion of truth, the partial definitions mentioned do permit us definitely to decide in many cases the question of the truth or falsity of the sentences investigated.

In particular, it can be proved—on the basis of the axioms of the second group adopted in the metatheory (cf. p. 211)—that *all axioms of the science under investigation are true sentences.* In a similar manner we can prove, making essential use of the fact that the rules of inference employed in the metatheory are not logically weaker than the corresponding rules of the science itself, *that all consequences of true sentences are true.* These two facts together enable us to assert that *the class of true sentences contains all provable sentences of the science investigated* (cf. Lemma D and Ths. 3 and 5 of § 3).

Among the most important consequences of a general nature which follow from the definition of truth must be reckoned the *principle of contradiction* and the *principle of the excluded middle.*

These two theorems, together with the theorem on the consequences of true sentences already mentioned, show that *the class of all true sentences forms a consistent and complete deductive system* (Ths. 1, 2, and 4).

As an immediate, although a somewhat subsidiary, consequence of these facts we obtain the theorem that *the class of all provable sentences likewise forms a consistent* (although not necessarily complete) *deductive system.* In this way we are able to produce a proof of the consistency of every science for which we can construct the definition of truth. The proof carried out by

† Some further consequences of this type are discussed in the article of the author 'On undecidable statements in enlarged systems of logic and the concept of truth', *Journal of Symbolic Logic*, vol. 4 (1939), pp. 105–12; cf. in particular sect. 9, p. 111.

means of this method does not, of course, add much to our knowledge, since it is based upon premisses which are at least as strong as the assumptions of the science under investigation.[1] Nevertheless it seems to be worthy of note that such a general method of proof exists, which is applicable to an extensive range of deductive sciences. It will be seen that from the deductive standpoint this method is not entirely trivial, and in many cases no simpler, and in fact no other, method is known.†

In those cases in which the class of provable sentences is not only consistent but also complete, it is easy to show that it coincides with the class of true sentences. If, therefore, we identify the two concepts—that of true sentence and that of provable sentence—we reach a new definition of truth of a purely structural nature and essentially different from the original semantical definition of this notion.[2] Even when the provable sentences

[1] As Ajdukiewicz has rightly pointed out in a somewhat different connexion (cf. Ajdukiewicz, K. (2), pp. 39–40) it does not at all follow from this that this proof is not correct from the methodological standpoint—that it contains in some form a *petitio principii*. The assertion which we prove, i.e. the consistency of the science, does not occur in any way among the hypotheses of the proof.

[2] In the course of this work I have several times contrasted semantical definitions of true sentence with structural definitions. But this does not mean that I intend to specify the distinction between the two kinds of definitions in an exact way. From the intuitive standpoint these differences seem to be tolerably clear. Def. 23 in § 3—as well as other definitions constructed in the same way—I regard as a semantical definition because in a certain sense

† In connexion with the problem discussed in the last three paragraphs see the recent publications: Mostowski, A. (53 *e*) as well as Wang, H. (87 *c*). From the results of these authors it is seen that in some cases, having succeeded in constructing an adequate definition of truth for a theory T in its metatheory, we may still be unable to show that all the provable sentences of T are true in the sense of this definition, and hence we may also be unable to carry out the consistency proof for T in M. This phenomenon can roughly be explained as follows: in the proof that all provable sentences of T are true a certain form of mathematical induction is essentially involved, and the formalism of M may be insufficiently powerful to secure the validity of this inductive argument. Hence a certain clarification of the assumptions (on pp. 174 ff.) concerning foundations of the metatheory may be desirable. In particular the phrase 'from any sufficiently developed system of mathematical logic' (p. 170) should be understood in a way which does not deprive the metatheory of any normally applied modes of inference. If the theory T is of finite order our purpose will be fully achieved if we decide to provide the metatheory M with a logical basis as strong as the general theory of classes discussed in the following section.

do not form a complete system the question of the construction of a structural definition is not *a priori* hopeless. Sometimes it is possible, by adding certain structurally described sentences, to extend the axiom system of the science in a suitable way so that it becomes a system in which the class of all its consequences coincides with the class of all true sentences. But there can be no question of a general method of construction. I suspect that the attempt to construct a structural definition, even in relatively simple cases—e.g. in connexion with the logic of two-termed relations studied in the preceding section—would encounter serious difficulties. These difficulties would certainly become much greater when it came to the question of giving a general structural criterion of truth, although we have already dealt with two languages, that of the calculus of classes and that of

(which would be difficult to define) it represents a 'natural generalization', so to speak an 'infinite logical product', of those partial definitions which were described in convention T and which establish a direct correlation between the sentences of the language and the names of these sentences. Among the structural definitions, on the other hand, I include those which are constructed according to the following scheme: a class of sentences or other expressions is described in such a way that from the form of every expression it is possible to know whether it belongs to the given class or not. Further operations on expressions are given of such a kind that if certain expressions in finite number are given and if the form of an arbitrary other expression is given, then we can decide whether it can be obtained from the given expressions by means of the given operations. Finally the true sentences are defined as those which are obtained by applying the given operation to the expressions of the given class any number of times (it is to be noted that such a structural definition still in no way provides a general criterion of truth). Certain differences of a formal nature can be recognized between these two kinds of definitions. The semantical definition requires the use of terms of higher order than all variables of the language investigated, e.g. the use of the term 'satisfies'; but for the formulation of a structural definition the terms of perhaps two or three of the lowest orders suffice. In the construction of a semantical definition we use—explicitly or implicitly—those expressions of the metalanguage which are of like meaning with the expressions of the language investigated, whilst they play no part in the construction of a structural definition; it is easy to see that this distinction vanishes when the language studied is a fragment of logic. Moreover, the distinction as a whole is not very clear and sharp, as is shown by the fact that with respect to the sentential calculus the semantical definition can be regarded as a formal transformation of the structural definition based on the matrix method. At the same time it must be remembered that the construction of semantical definitions, based on the methods at present known to us, is essentially dependent upon the structural definitions of sentence and sentential function.

the sentential calculus, for which this problem could be relatively easily solved.[1]

In all cases in which we are able to define satisfaction and the notion of true statement, we can—by means of a modification of these definitions—also define two still more general concepts of a relative kind, namely the concepts of *satisfaction* and *correct sentence*—both *with respect to a given individual domain a*.[2] This modification depends on a suitable restriction of the domain of objects considered. Instead of operating with arbitrary individuals, classes of individuals, relations between individuals, and so on, we deal exclusively with the elements of a given class a of individuals, subclasses of this class, relations between elements of this class, and so on. It is obvious that in the special case when a is the class of all individuals, the new concepts coincide with the former ones (cf. Defs. 24 and 25, and Th. 26). As I have already emphasized in § 3 the general concept of correct sentence in a given domain plays a great part in present day methodological researches. But it must be added that this only concerns researches whose object is mathematical logic and its parts. In connexion with the special sciences we are interested in correct sentences in a quite specific individual domain for which the general concept loses its importance. Likewise it is only in connexion with sciences which are parts of logic that some general properties of these concepts, which were proved in § 3 for the language of the calculus of classes, preserve their validity. For example, it happens that in these sciences the extension of the term 'correct sentence in the individual domain a' depends exclusively on the cardinal number of the class a. Thus in these investigations we can replace this term by the more convenient term 'correct sentence in a domain with k elements' (Def. 26, Th. 8). The theorems previously discussed concerning the concept of truth, such as the principles of contradiction and the excluded middle can be extended to the concept of correct sentence in a given domain. The concept of correct sentence in every

[1] Cf. the remarks on pp. 207 f. and 221; I shall return to this problem in § 5 (cf. p. 254, footnote).

[2] See p. 199, note 2.

individual domain (Def. 27) deserves special consideration. In its extension it stands midway between the concept of provable sentence and that of true sentence; the class of correct sentences in every domain contains all theorems and consists exclusively of true sentences (Ths. 22 and 27). This class is therefore in general narrower than the class of all true sentences; it contains, for example, no sentences whose validity depends on the magnitude of the number of all individuals (Th. 23). If it is desired to transform the system of the provable sentences of every science into a complete one, it is necessary at the outset to add sentences to the system which decide the question how many individuals exist. But for various reasons another point of view seems to be better established, namely the view that the decision regarding such problems should be left to the specific deductive sciences, whilst in logic and its parts we should try to ensure only that the extension of the concept of provable sentence coincides with that of correct sentence in every individual domain. For a supporter of this standpoint the question whether the extension of these two concepts is actually identical is of great importance. In the case of a negative answer the problem arises of completing the axiom system of the science studied in such a way that the class of provable sentences thus extended now coincides with the class of sentences which are correct in every domain. This problem, which properly is equivalent to the question of structurally characterizing the latter concept, can be positively decided only in a few cases (cf. Th. 24).[1] Generally speaking the difficulties presented by this question are no less essential than those connected with the analogous problem of a structural definition of true sentence. We meet with similar difficulties when we attempt to define structurally the concept of correct sentence in a domain with k elements. Only in the case where k is a finite number is it easy to give a general method, modelled on the method of matrices from investigations on the extended sentential calculus, which makes a structural definition

[1] In the case of the lower functional calculus this problem, which is raised in Hilbert, D., and Ackermann, W. (30), p. 68, has recently been decided by Gödel, see Gödel, K. (20).

of this concept possible. In this way we even obtain a general criterion which enables us to decide from the form of any sentence whether it is correct in a domain with a previously given finite number of elements.[1]

I do not wish to enter here into a more detailed discussion of special investigations on the concepts just considered. Some results which are relevant here, relating to the calculus of classes, have already been given as examples in § 3. I will only mention that in recent years numerous results have been obtained which enable us to infer from the correctness of certain sentences in special individual domains or from their structural properties their correctness in every domain and thus their truth.[2] It is evident that all these results only receive a clear content and can only then be exactly proved, if a concrete and precisely formulated definition of correct sentence is accepted as a basis for the investigation.

§ 5. The Concept of True Sentence in Languages of Infinite Order

We come now to languages of the 4th kind, hence to those of infinite order and so lying beyond the scope of the methods of construction sketched in the preceding section. The language of the *general theory of classes* will serve as an example. This language is noteworthy because, in spite of its elementary structure and its poverty in grammatical forms, it suffices for

[1] Cf. Bernays, P., and Schönfinkel, M. (5 *a*), p. 352.

[2] According to the well-known theorems of Löwenheim and Skolem, certain categories of sentences are correct in every domain provided they are correct in all finite and denumerable domains. These sentences include, for example, all sentences of the logic of two- or many-termed relations, described in this section, which are generalizations of sentential functions in which variables of the 2nd order occur exclusively as free variables. In the case of the sentences of the calculus of classes this result—as is shown in Ths. 15 and 19 of § 3—can be essentially sharpened. Certain results of Bernays, Schönfinkel, and Ackermann have a narrower domain of application. They allow us to correlate a particular natural number k with sentences of a special structure in such a way that from the correctness of a given sentence in the domain with k elements (thus—as we already know—from purely structural properties of the sentence) its correctness in every domain follows. Cf. Ackermann, W. (1), Bernays, P., and Schönfinkel, M. (5 *a*), Herbrand, J. (26), Löwenheim, L. (49), Skolem, Th. (64), (65), and (66). For a systematic presentation of the results in this direction including more recent ones, see Church, A. (11 *a*).

the formulation of every idea which can be expressed in the whole language of mathematical logic. It is difficult to imagine a simpler language which can do this.[1]

In the general theory of classes the same constants occur as in the previously investigated sciences, i.e. the signs of negation and of logical sum, as well as the universal quantifier. As variables we use such symbols as 'X'', 'X''', 'X'''', and so on, i.e. signs composed of the symbol 'X' and a number of small strokes above and below. The sign having n strokes above and k below is called the *k-th variable of the n-th order* and is denoted by the symbol 'V_k^n'. The variables $V_k^1, V_k^2, V_k^3, \ldots$ represent respectively names of individuals, objects of the 1st order; classes of individuals, objects of the 2nd order; classes of such classes, objects of the 3rd order, and so on. These variables obviously belong to infinitely many semantical categories. As primitive sentential functions we have expressions of the type 'XY' where in the place of 'X' any variable of the $n+1$th order, and instead of 'Y' a variable of the nth order occurs. This expression is

[1] The language of the general theory of classes is much inferior to the language of Whitehead, A. N., and Russell, B. A. W. (90) in its stock of semantical categories, and still more inferior in this respect to the language used by Leśniewski in his system (cf. p. 210, note 2; p. 218, note 2). In particular, in this language no sentential variables and neither names of two- or many-termed relations, nor variables representing these names, occur. The dispensability of sentential variables depends on the fact mentioned on p. 229, footnote: to every sentence which contains sentential variables there is a logically equivalent sentence which does not contain such variables. The results of § 2, especially Defs. 13–17, suffice to show how such variables are to be avoided in setting up lists of axioms and in the derivation of theorems; cf. also Neumann, J. v. (54) (especially note 9, p. 38). The possibility of eliminating two-termed relations results from the following consideration. With every relation R we can correlate, in one-one fashion, a class of ordered pairs, namely, the class of all ordered pairs whose terms x and y satisfy the formula, xRy. If the relation is homogeneous, i.e. if the domain and counter domain of this relation belong to the same semantical category, then the ordered pair can be interpreted otherwise than we have done on p. 171, namely as classes having two classes as elements: the class whose only element is x and the class consisting of the two elements x and y. In order to apply an analogous method to inhomogeneous relations we must first correlate homogeneous relations with them in one-one fashion, and this presents no great difficulty. We proceed in an analogous way with many-termed relations. In this way every statement about two- or many-termed relations of arbitrary category can be transformed into an equivalent statement about individuals, classes of individuals, classes of such classes, and so on. Cf. Kuratowski, C. (38), p. 171, and Chwistek, L. (13), especially p. 722.

read: 'the class X (of $n+1$th order) has as an element the object Y (of nth order)', or 'the object Y has the property X'. For the designation of the primitive functions we employ the symbol '$\epsilon_{k,l}^n$', setting $\epsilon_{k,l}^n = V_k^{n+1} \frown V_l^n$. The further development of the science differs in no essential way from that of the logic of two- or many-termed relations. The quantifications of the sentential functions x with respect to the variable V_k^n are denoted by the symbols '$\bigcap_k^n x$' and '$\bigcup_k^n x$'. The axioms consist of (1) sentences which satisfy the condition (α) of Def. 13 of § 2, which are thus derived from the axioms of the sentential calculus by substitution, sometimes also followed by generalization; (2) pseudodefinitions, i.e. statements which are quantifications of sentential functions of the type

$$\bigcup_k^{n+1} \bigcap_l^n (\epsilon_{k,l}^n \cdot y + \overline{\epsilon_{k,l}^n} \cdot \bar{y}),$$

where y is any sentential function which does not contain the free variable V_k^{n+1}; (3) the *laws of extensionality*, i.e. sentences of the form

$$\bigcap_k^{p+2} \bigcap_l^{p+1} \bigcap_m^{p+1} (\bigcup_n^p (\epsilon_{l,n}^p \cdot \overline{\epsilon_{m,n}^p} + \overline{\epsilon_{l,n}^p} \cdot \epsilon_{m,n}^p) + \overline{\epsilon_{k,l}^{p+1}} + \epsilon_{k,m}^{p+1}),$$

which state that two classes which do not differ in their elements do not differ in any of their properties and are thus identical. In order to obtain in this science a sufficient basis for the establishment of various parts of mathematics and in particular of the whole of theoretical arithmetic, we must add to the above still one more axiom: (4) the *axiom of infinity*, i.e. the sentence

$$\bigcup_1^3 (\bigcup_1^2 \epsilon_{1,1}^2 \cdot \bigcap_1^2 (\overline{\epsilon_{1,1}^2} + \bigcup_2^2 (\epsilon_{1,2}^2 \cdot \bigcap_1^1 (\epsilon_{1,1}^1 + \overline{\epsilon_{2,1}^1}) \cdot \bigcup_1^1 (\epsilon_{1,1}^1 \cdot \overline{\epsilon_{2,1}^1})))),$$

which guarantees the existence of infinitely many individuals.[1] In the derivation of consequences from the axioms we apply the operations of substitution, detachment, and the introduction and removal of the universal quantifier, analogous to the operations described in conditions (γ)–(ζ) of Def. 15 in § 3.

When we try to define the concept of satisfaction in connexion with the present language we encounter difficulties which we cannot overcome. In the face of the infinite diversity of seman-

[1] In adopting the axiom of infinity we admittedly give up the postulate according to which only the sentences which are correct in every individual domain are to be provable sentences of logic (cf. p. 240).

tical categories which are represented in the language, the use of the method of many-rowed sequences is excluded from the beginning, just as it was in the case of the logic of many-termed relations. But the situation here is still worse, because the method of semantical unification of the variables also fails us. As we learnt in § 4, the unifying category cannot be of lower order than any one of the variables of the language studied. Sequences whose terms belong to this category, and still more the relation of satisfaction, which holds between such sequences and the corresponding sentential functions, must thus be of higher order than all those variables. In the language with which we are now dealing variables of arbitrarily high (finite) order occur: consequently in applying the method of unification it would be necessary to operate with expressions of 'infinite order'. Yet neither the metalanguage which forms the basis of the present investigations, nor any other of the existing languages, contains such expressions. It is in fact not at all clear what intuitive meaning could be given to such expressions.

These considerations seem to show that it is impossible to construct a general, semantically unambiguous concept of satisfaction for the language we are studying which will be applicable to all sentential functions without regard to their semantical type. On the other hand there appear to be no difficulties which would render impossible in principle a consistent application of the concept of satisfaction in its original formulation, or rather—in view of the semantical ambiguity of that formulation—of an infinite number of such concepts. Each of these concepts is, from the semantical standpoint, already specified and would relate exclusively to functions of a specific semantical type (e.g. to functions which contain a variable of the 1st order as the only free variable). Actually—independently of the logical structure of the language—the intuitive sense of none of these expressions raises any doubt. For every particular sentential function we can in fact define this meaning exactly by constructing for every phrase of the type 'the objects a, b, c,... satisfy the given sentential function' an intuitively equivalent phrase which is expressed wholly in terms

of the metalanguage. Nevertheless the problem of the construction of a correct definition for each of these concepts again presents us with difficulties of an essential nature. On the basis of the languages which we have previously studied it was easy to obtain each special concept of satisfaction by a certain specialization of the general concept; in the present case this way is clearly not open to us. A brief reflection shows that the idea of using the recursive method analogously to the definition of sentential function proves, in spite of its naturalness, to be unsuitable. It is easily seen that the composite functions of a particular semantical type cannot always be formed from simpler functions of the same type. On the contrary, if we are to be able to construct arbitrary functions of a given type, we must use for that purpose all possible semantical types.[1] It would, therefore, be necessary, in the recursive definition of any one of the special concepts of satisfaction, to cover, in one and the same recursive process, infinitely many analogous concepts, and this is beyond the possibilities of the language.

The central problem of our work, the construction of the definition of truth, is closely connected with these considerations. If we were successful in defining, if not the general, at least any one of the special concepts of satisfaction, then this problem would not offer the least difficulty.[2] On the other

[1] An external expression of this state of affairs is that in the definition of satisfaction not only is it essential to take free variables into account but also all the bound variables of the function in question, although these variables have no influence on the semantical type of the function; and whether the relation of satisfaction holds or not does not depend in any way on the terms of the sequence which correspond to these variables (cf. Def. 22 of § 3, condition (δ)). It is to be remembered that analogous difficulties to those mentioned in the text appeared earlier in the attempt to construct a recursive definition of truth by a direct route (cf. p. 189).

[2] For example, let us imagine that we have succeeded in some way in defining the concept of satisfaction in the case of sentential functions which contain a variable of 1st order as the only free variable. We could then operate freely with phrases of the type 'the individual a satisfies the sentential function y'. If we now consider some one concrete sentential function, e.g. $\bigcup_1^2 \epsilon_{1,1}^1$, which is satisfied by every arbitrary individual, we obtain at once the following definition of true sentence: *x is a true sentence if and only if every individual a satisfies the function $x.\bigcup_1^2 \epsilon_{1,1}^1$ (i.e. the conjunction of the sentence x and the function $\bigcup_1^2 \epsilon_{1,1}^1$)*. In an exactly analogous way we can pass from every other specific concept of satisfaction to the concept of truth.

hand we know of no method of construction which would not—directly or indirectly—presuppose a previous definition of the concept of satisfaction. Therefore we can say—considering the failure of previous attempts—that at present we can construct no correct and materially adequate definition of truth for the language under investigation.†

In the face of this state of affairs the question arises whether our failure is accidental and in some way connected with defects in the methods actually used, or whether obstacles of a fundamental kind play a part which are connected with the nature of the concepts we wish to define, or of those with the help of which we have tried to construct the required definitions. If the second supposition is the correct one all efforts intended to improve the methods of construction would clearly be fruitless. If we are to answer this question we must first give it a rather less indefinite form. It will be remembered that in the convention T of § 3 the conditions which decide the material correctness of any definition of true sentence are exactly stipulated. The construction of a definition which satisfies these conditions forms in fact the principal object of our investigation. From this standpoint the problem we are now considering takes on a precise form: it is a question of *whether on the basis of the metatheory of the language we are considering the construction of a correct definition of truth in the sense of convention* T *is in principle possible*. As we shall see, the problem in this form can be definitely solved, but in a *negative sense*.

It is not difficult to see that this problem exceeds the bounds of our previous discussion. It belongs to the field of the metametatheory. Its definitive solution, even its correct formulation, would require new equipment for investigation and especially the formalization of the metalanguage and the metatheory which uses it. But without going so far, and still avoiding

† The problem of the possibility of defining satisfaction and truth for the language under investigation will be considerably clarified by the discussion in the Postscript. It should be mentioned that the method of defining truth recently suggested in McKinsey, J. C. C. (53 *b*) is not based on a preliminary definition of satisfaction. Instead, McKinsey has to consider formalized languages with non-denumerably many constants and has to use a metalanguage which is provided with a very strong set-theoretical apparatus.

various technical complications, I believe I am able to give a fairly clear account of everything of a positive nature that can at present be established in connexion with the above problem.

In operating with the metalanguage we shall adhere to the symbolism given in §§ 2 and 3. To simplify the further developments and avoid possible misunderstandings we shall suppose the metalanguage to be so constructed that the language we are studying forms a fragment of it; every expression of the language is at the same time an expression of the metalanguage, but not vice versa. This enables us in certain cases (e.g. in the formulation of condition (α) of convention T) to speak simply of the expressions of the language itself, instead of expressions of the metalanguage which have the same meaning.

After these reservations and conventions we turn to the formulation and proof of the fundamental result.

THEOREM I. (α) *In whatever way the symbol 'Tr', denoting a class of expressions, is defined in the metatheory, it will be possible to derive from it the negation of one of the sentences which were described in the condition* (α) *of the convention* T;

(β) *assuming that the class of all provable sentences of the metatheory is consistent, it is impossible to construct an adequate definition of truth in the sense of convention* T *on the basis of the metatheory.*

The idea of the proof of this theorem can be expressed in the following words:[1] (1) a particular interpretation of the meta-

[1] We owe the method used here to Gödel, who has employed it for other purposes in his recently published work, Gödel, K. (22), cf. especially pp. 174–5 or 187–90 (proof of Th. VI). This exceedingly important and interesting article is not directly connected with the theme of our work—it deals with strictly methodological problems: the consistency and completeness of deductive systems; nevertheless we shall be able to use the methods and in part also the results of Gödel's investigations for our purpose.

I take this opportunity of mentioning that Th. I and the sketch of its proof was only added to the present work after it had already gone to press. At the time the work was presented at the Warsaw Society of Sciences (21 March 1931), Gödel's article—so far as I know—had not yet appeared. In this place therefore I had originally expressed, instead of positive results, only certain suppositions in the same direction, which were based partly on my own investigations and partly on the short report, Gödel, K. (21), which had been published some months previously.

After I had become acquainted with the above mentioned article I convinced myself, among other things, that the deductive theory which Gödel

language is established in the language itself and in this way with every sentence of the metalanguage there is correlated, in one-many fashion, a sentence of the language which is equivalent to it (with reference to the axiom system adopted in the metatheory); in this way the metalanguage contains as well as every particular sentence, an individual name, if not for that sentence at least for the sentence which is correlated with it and equivalent to it. (2) Should we succeed in constructing in the metalanguage a correct definition of truth, then the metalanguage—with reference to the above interpretation—would acquire that universal character which was the primary source of the semantical antinomies in colloquial language (cf. p. 164). It would then be possible to reconstruct the antinomy of the liar in the metalanguage, by forming in the language itself a sentence x such that the sentence of the metalanguage which is correlated with x asserts that x is not a true sentence. In doing this it would be possible, by applying the diagonal procedure[1] from the theory of sets, to avoid all terms which do not belong to the metalanguage, as well as all premisses of an empirical nature which have played a part in the previous formulations of the antinomy of the liar.[2]

had chosen as the object of his studies, which he called the 'system P', was strikingly similar to the general theory of classes considered in the present section. Apart from certain differences of a 'calligraphical' nature, the only distinction lies in the fact that in the system P, in addition to three logical constants, certain constants belonging to the arithmetic of the natural numbers also occur (a far-reaching analogy also exists between the system P and the system of arithmetic sketched in VI (see pp. 113–16)). Consequently the results obtained for the system P can easily be carried over to the present discussion. Moreover, the abstract character of the methods used by Gödel renders the validity of his results independent to a high degree of the specific peculiarities of the science investigated.

[1] Cf. Fraenkel, A. (16), pp. 48 ff.

[2] If we analyse the sketch of the proof given below we easily note that an analogous reconstruction could be carried out even on the basis of colloquial language, and that in consequence of this reconstruction the antinomy of the liar actually approximates to the antinomy of the expression 'heterological'. For a rather simple reconstruction of the antinomy of the liar in this direction see Tarski, A. (82), note 11, p. 371. It seems interesting that in this reconstruction all the technical devices are avoided which are used in the proof of Th. 1 (such as interpretation of the metalanguage in arithmetic or the diagonal procedure). In connexion with the last paragraph of the text cf. the concluding remarks of § 1, pp. 164 f., and in particular p. 165, note 1.

We shall sketch the proof a little more exactly.[1]

Let us agree for the moment to use the symbol 'n' instead of 'X''''. The existential quantifiation of the sentential function y with respect to the variable 'n' will be denoted by the symbol '$\bigcup_1^3 y$' as before. The variable 'n' thus represents names of classes the elements of which are classes of individuals. Among these classes we find, among other things, the natural numbers and generally speaking the cardinal numbers.[2]

I have already mentioned that all facts belonging to the arithmetic of the natural numbers can be expressed in the language of the general theory of classes. In particular, if a natural number k is given, a sentential function ι_k is, easily constructed in this language containing the symbol 'n' as the only free variable and which asserts that the class whose name is represented by this symbol is identical with the number k (and thus consists of just those classes of individuals which have exactly k elements).[2] For example:

$$\iota_1 = \bigcap_1^2(\epsilon_{1,1}^2 \cdot \bigcup_1^1 \bigcap_2^1 \bigcap_2^2(\epsilon_{1,1}^1 \cdot (\overline{\epsilon_{1,2}^1} + \overline{\epsilon_{2,1}^1} + \epsilon_{2,2}^1)) + \\ + \overline{\epsilon_{1,1}^2} \cdot \bigcap_1^1 \bigcup_2^1 \bigcup_2^2(\overline{\epsilon_{1,1}^1} + \epsilon_{1,2}^1 \cdot \epsilon_{2,1}^1 \cdot \overline{\epsilon_{2,2}^1})).$$

A general recursive definition of the sequence of functions ι_k within the metalanguage presents no great difficulty.

As I have already pointed out in § 2 (p. 184) a one-one correspondence can be set up without difficulty between the expressions of the language and the natural numbers; we can define in the metalanguage an infinite sequence ϕ of expressions in which every expression of the language occurs once and only once. With the help of this correlation we can correlate with every operation on expressions an operation on natural numbers (which possesses the same formal properties), with every class of expressions a class of natural numbers, and so on. In this way the metalanguage receives an interpretation in the arithmetic of the

[1] For the sake of simplicity we shall in many places express ourselves as though the demonstration which follows belonged to the metatheory and not to the meta-metatheory; in particular, instead of saying that a given sentence is provable in the metatheory, we shall simply assert the sentence itself. In any case it must not be forgotten that only a sketch of the proof is given here and one which is far from complete.

[2] See p. 233, note 1.

natural numbers and indirectly in the language of the general theory of classes.

Let us suppose that we have defined the class Tr of sentences in the metalanguage. There would then correspond to this class a class of natural numbers which is defined exclusively in the terms of arithmetic. Consider the expression '$\bigcup_1^3(\iota_n . \phi_n) \overline{\in} Tr$'. This is a sentential function of the metalanguage which contains 'n' as the only free variable. From the previous remarks it follows that with this function we can correlate another function which is equivalent to it for any value of 'n', but which is expressed completely in terms of arithmetic. We shall write this new function in the schematic form '$\psi(n)$'. Thus we have:

(1) *for any* n, $\bigcup_1^3(\iota_n . \phi_n) \overline{\in} Tr$ *if and only if* $\psi(n)$.

Since the language of the general theory of classes suffices for the foundation of the arithmetic of the natural numbers, we can assume that '$\psi(n)$' is one of the functions of this language. The function '$\psi(n)$' will thus be a term of the sequence ϕ, e.g. the term with the index k, '$\psi(n)$' $= \phi_k$. If we substitute 'k' for 'n' in the sentence (1) we obtain:

(2) $\bigcup_1^3(\iota_k . \phi_k) \overline{\in} Tr$ *if and only if* $\psi(k)$.

The symbol '$\bigcup_1^3(\iota_k . \phi_k)$' denotes, of course, a sentence of the language under investigation. By applying to this sentence condition (α) of the convention **T** we obtain a sentence of the form '$x \overline{\in} Tr$ *if and only if* p', where 'x' is to be replaced by a structural-descriptive or any other individual name of the statement $\bigcup_1^3(\iota_k . \phi_k)$, but '$p$' by this statement itself or by any statement which is equivalent to it. In particular we can substitute '$\bigcup_1^3(\iota_k . \phi_k)$' for '$x$' and for '$p$'—in view of the meaning of the symbol 'ι_k'—the statement 'there is an n such that $n = k$ and $\psi(n)$' or, simply '$\psi(k)$'. In this way we obtain the following formulation:

(3) $\bigcup_1^3(\iota_k . \phi_k) \in Tr$ *if and only if* $\psi(k)$.

The sentences (2) and (3) stand in palpable contradiction to one another; the sentence (2) is in fact directly equivalent to the negation of (3). In this way we have proved the first part of

the theorem. We have proved that among the consequences of the definition of the symbol 'Tr' the negation of one of the sentences mentioned in the condition (α) of the convention **T** must appear. From this the second part of the theorem immediately follows.

The assumption of consistency appearing in the part (β) of this theorem is essential. If the class of all provable sentences of the metatheory contained a contradiction, then every definition in the metatheory would have among its consequences all possible sentences (since they all would be provable in the metatheory), in particular those described in the convention **T**. On the other hand, as we now know,[1] there is no prospect of proving the consistency of the metatheory which we are working with, on the basis of the meta-metatheory. It is to be noted that, in view of the existence of an interpretation of the metatheory in the science itself (a fact which has played such an essential part in the proof sketched above), the assumption of the second part of Th. I is equivalent to the assumption of the consistency of the science investigated itself and from the intuitive standpoint is just as evident.

The result reached in Th. I seems perhaps at first sight uncommonly paradoxical. This impression will doubtless be weakened as soon as we recall the fundamental distinction between the content of the concept to be defined and the nature of those concepts which are at our disposal for the construction of the definition.

The metalanguage in which we carry out the investigation contains exclusively structural-descriptive terms, such as names of expressions of the language, structural properties of these expressions, structural relations between expressions, and so on, as well as expressions of a logical kind among which (in the present case) we find all the expressions of the language studied. What we call metatheory is, fundamentally, the *morphology of language*—a science of the form of expressions—a correlate of such parts of traditional grammar as morphology, etymology, and syntax.

[1] Cf. Gödel, K. (22), p. 196 (Th. XI).

The fact that the language studied and the deductive science carried out in this language are formalized has brought about an interesting phenomenon; it has been possible to reduce to structural-descriptive concepts certain other notions of a totally different kind, which are distinguished from the former both in their origin and in their usual meaning, namely the concept of consequence together with a series of related notions.[1] It has been possible to establish what may be called the *logic of the given science* as a part of morphology.

Encouraged by this success we have attempted to go further and to construct in the metalanguage definitions of certain concepts belonging to another domain, namely that called the *semantics of language*—i.e. such concepts as satisfaction, denoting, truth, definability, and so on. A characteristic feature of the semantical concepts is that they give expression to certain relations between the expressions of language and the objects about which these expressions speak, or that by means of such relations they characterize certain classes of expressions or other objects. We could also say (making use of the *suppositio materialis*) that these concepts serve to set up the correlation between the names of expressions and the expressions themselves.

For a long time the semantical concepts have had an evil reputation among specialists in the study of language. They have resisted all attempts to define their meaning exactly, and the properties of these concepts, apparently so clear in their content, have led to paradoxes and antinomies. For that reason the tendency to reduce these concepts to structural-descriptive ones must seem quite natural and well-founded. The following fact seemed to favour the possibility of realizing this tendency: it has always been possible to replace every phrase which contains these semantical terms, and which concerns particular

[1] The reduction of the concept of consequence to concepts belonging to the morphology of language is a result of the deductive method in its latest stages of development. When, in everyday life, we say that a sentence follows from other sentences we no doubt mean something quite different from the existence of certain structural relations between these sentences. In the light of the latest results of Gödel it seems doubtful whether this reduction has been effected without remainder.

structurally described expressions of the language, by a phrase which is equivalent in content and is free from such terms. In other words it is possible to formulate infinitely many partial definitions for every semantical concept, which in their totality exhaust all cases of the application of the concept to concrete expressions and of which the sentences adduced in condition (α) of convention **T** are examples. It was with just this end in view that, as a rule, we included in the metalanguage, with regard to the content of the semantical concepts, not only the names of expressions but all expressions of the language itself or expressions having the same meaning (even when these expressions were not of a logical kind, cf. pp. 210 f.), although such an enrichment of the metalanguage has no advantages for the pursuit of the 'pure' morphology of language.

In the abstract the fact mentioned has no decisive importance; it offers no path by which an automatic transition from the partial definitions to a general definition is possible, which embraces them all as special cases and would form their infinite logical product.[1] Only thanks to the special methods of construction which we developed in §§ 3 and 4 have we succeeded in carrying out the required reduction of the semantical concepts, and then only for a specified group of languages which are poor in grammatical forms and have a restricted equipment of semantical categories—namely the languages of finite order. Let it be remembered that the methods there applied required the use in the metalanguage of categories of higher order than all categories of the language studied and are for that reason fundamentally different from all grammatical forms of this language. The analysis of the proof of Th. I sketched above shows that this

[1] In the course of our investigation we have repeatedly encountered similar phenomena: the impossibility of grasping the simultaneous dependence between objects which belong to infinitely many semantical categories; the lack of terms of 'infinite order'; the impossibility of including, in *one* process of definition, infinitely many concepts, and so on (pp. 188 f., 232 f., 243, 245). I do not believe that these phenomena can be viewed as a symptom of the formal incompleteness of the actually existing languages—their cause is to be sought rather in the nature of language itself: language, which is a product of human activity, necessarily possesses a 'finitistic' character, and cannot serve as an adequate tool for the investigation of facts, or for the construction of concepts, of an eminently 'infinitistic' character.

circumstance is not an accidental one. Under certain general assumptions, it proves to be impossible to construct a correct definition of truth if only such categories are used which appear in the language under consideration.[1] For that reason the situation had fundamentally changed when we passed to the 'rich' languages of infinite order. The methods used earlier proved to be inapplicable; all concepts and all grammatical forms of the metalanguage found an interpretation in the language and hence we were able to show conclusively that the semantics of the language could not be established as a part of its morphology. The significance of the results reached reduces just to this.

But, apart from this, Th. I has important consequences of a methodological nature. It shows that it is impossible to define in the metatheory a class of sentences of the language studied which consists exclusively of materially true sentences and is at the same time complete (in the sense of Def. 20 in § 3). In particular, if we enlarge the class of provable sentences of the science investigated in any way—whether by supplementing the list of axioms or by sharpening the rules of inference—then we either add false sentences to this class or we obtain an incomplete system. This is all the more interesting inasmuch as the enlargement of the class of provable sentences to form a complete and consistent system in itself presents no difficulties.[1]

[1] From this, or immediately from certain results contained in Gödel, K. (22) (pp. 187–91), it can easily be inferred that a structural definition of truth—in the sense discussed on pp. 236 ff., especially on p. 237, note 2—cannot be constructed even for languages of finite order which are in some degree richer. From other investigations of this author (op. cit., p. 193; Th. IX) it follows that in certain elementary cases in which we can construct such a definition, it is nevertheless impossible to give a general structural criterion of the truth of a sentence. The first of these results applies, for instance, to the logic of two-termed and many-termed relations discussed in § 4. The second result applies, for example, to the lower predicate calculus ('engere Funktionenkalkül') of Hilbert-Ackermann (30), pp. 43 ff.; in this case, however, the result is applied, not to the notion of a true sentence, but to the related notion of a universally valid ('allgemeingültig') sentential function.

At this point we should like to call attention to the close connexion between the notions of 'structural definition of truth', and of 'general structural criterion of truth' discussed in this work, and the notions of recursive enumerability and general recursiveness known from the recent literature (see, for example, Mostowski, A. (53*f*), chap. 5). In fact, by saying that there is a 'structural definition of truth' for a given formalized theory we essentially mean that the set of all true sentences of this theory is recursively enumerable; when we say that there is a 'general structural criterion of truth' we mean that the set of all true sentences is general recursive.

An interpretation of Th. I which went beyond the limits given would not be justified. In particular it would be incorrect to infer the impossibility of operating consistently and in agreement with intuition with semantical concepts and especially with the concept of truth. But since one of the possible ways of constructing the scientific foundations of semantics is closed we must look for other methods. The idea naturally suggests itself of setting up semantics as a special deductive science with a system of morphology as its logical substructure. For this purpose it would be necessary to introduce into morphology a given semantical notion as an undefined concept and to establish its fundamental properties by means of axioms. The experience gained from the study of semantical concepts in connexion with colloquial language, warns us of the great dangers that may accompany the use of this method. For that reason the question of how we can be certain that the axiomatic method will not in this case lead to complications and antinomies becomes especially important.

In discussing this question I shall restrict myself to the theory of truth, and in the first place I shall establish the following theorem, which is a consequence of the discussion in the preceding section:

THEOREM II. *For an arbitrary, previously given natural number k, it is possible to construct a definition of the symbol 'Tr' on the basis of the metatheory, which has among its consequences all those sentences from the condition* (α) *of the convention* T *in which in the place of the symbol 'p' sentences with variables of at most the k-th order occur (and moreover, the sentence adduced in the condition* (β) *of this convention).*

By way of proof it suffices to remark that this theorem no longer concerns the language studied in its whole extent but only a fragment of it which embraces all those expressions which contain no variable of higher order than the *k*th. This fragment

[1] Cf. V, Th. 56, a result of Lindenbaum's (see p. 98 of the present volume).

is clearly a language of finite order and in fact a language of the 2nd kind. We can therefore easily construct the required definition by applying one of the two methods described in § 4. It is to be noted that the definition obtained in this way (together with the consequences given in Th. II) yields a series of theorems of a general nature, like the Ths. 1–5 in § 3, for example, if the formulations of these theorems are suitably weakened by restricting the domain of their applicability to sentences with variables of at most the kth order.

Hence it will be seen that, in contrast to the theory of truth in its totality, the single fragments of this theory (the objects of investigation of which are sentences which contain only variables whose order is bounded above) can be established as fragments of the metatheory. If, therefore, the metatheory is consistent we shall not find a contradiction in these fragments. This last result can be extended in a certain sense to the whole theory of truth, as the following theorem shows:

THEOREM III. *If the class of all provable sentences of the metatheory is consistent and if we add to the metatheory the symbol 'Tr' as a new primitive sign, and all theorems which are described in conditions* (α) *and* (β) *of the convention* T *as new axioms, then the class of provable sentences in the metatheory enlarged in this way will also be consistent.*

To prove this theorem we note that the condition (α) contains infinitely many sentences which are taken as axioms of the theory of truth. A finite number of these axioms—even in union with the single axiom from condition (β)—cannot lead to a contradiction (so long as there is no contradiction already in the metatheory). Actually in the finite number of axioms obtained from (α) only a finite number of sentences of the language studied appears and in these sentences we find a finite number of variables. There must, therefore, be a natural number k such that the order of none of these variables exceeds k. From this it follows, by Th. II, that a definition of the symbol 'Tr' can be constructed in the metatheory such that the axioms in question become consequences of this definition. In other words: these axioms, with

a suitable interpretation of the symbol 'Tr', become provable sentences of the metatheory (this fact can also be established directly, i.e. independently of Th. II). If any class of sentences contains a contradiction, it is easy to show that the contradiction must appear in a finite part of this class.[1] Since, however, no finite part of the axiom system described in Th. III contains a contradiction, the whole system is consistent, which was to be proved.

The value of the result obtained is considerably diminished by the fact that the axioms mentioned in Th. III have a very restricted deductive power. A theory of truth founded on them would be a highly incomplete system, which would lack the most important and most fruitful general theorems. Let us show this in more detail by a concrete example. Consider the sentential function '$x \bar{\in} Tr$ *or* $\bar{x} \bar{\in} Tr$'. If in this function we substitute for the variable 'x' structural-descriptive names of sentences, we obtain an infinite number of theorems, the proof of which on the basis of the axioms obtained from the convention T presents not the slightest difficulty. But the situation changes fundamentally as soon as we pass to the generalization of this sentential function, i.e. to the general principle of contradiction. From the intuitive standpoint the truth of all those theorems is itself already a proof of the general principle; this principle represents, so to speak, an 'infinite logical product' of those special theorems. But this does not at all mean that we can actually derive the principle of contradiction from the axioms or theorems mentioned by means of the normal modes of inference usually employed. On the contrary, by a slight modification in the proof of Th. III it can be shown that the principle of contradiction is not a consequence (at least in the existing sense of the word) of the axiom system described.

We could, of course, now enlarge the above axiom system by adding to it a series of general sentences which are independent of this system. We could take as new axioms the principles of contradiction and excluded middle, as well as those sentences which assert that the consequences of true sentences are true,

[1] Cf. V, Th. 48, p. 91 of the present volume.

and also that all primitive sentences of the science investigated belong to the class of true sentences. Th. III could be extended to the axiom system enlarged in this way.[1] But we attach little importance to this procedure. For it seems that every such enlargement of the axiom system has an accidental character, depending on rather inessential factors such, for example, as the actual state of knowledge in this field. In any case, various objective criteria which we should wish to apply in the choice of further axioms prove to be quite inapplicable. Thus it seems natural to require that the axioms of the theory of truth, together with the original axioms of the metatheory, should constitute a categorical system.[2] It can be shown that this postulate coincides in the present case with another postulate, according to which the axiom system of the theory of truth should unambiguously determine the extension of the symbol '*Tr*' which occurs in it, and in the following sense: if we introduce into the metatheory, alongside this symbol, another primitive sign, e.g. the symbol '*Tr'*' and set up analogous axioms for it, then the statement '$Tr = Tr'$' must be provable. But this postulate cannot be satisfied. For it is not difficult to prove that in the contrary case the concept of truth could be defined exclusively by means of terms belonging to the morphology of language, which would be in palpable contradiction to Th. I. For other reasons of a more general nature there can be no question of an axiom system that would be complete and would consequently suffice for the solution of every problem from the domain of the theory under consideration. This is an immediate methodological consequence of Th. I applied not to the language of the general theory of classes but to the richer language of the metatheory and the theory of truth (cf. the remarks on p. 254).

There is, however, quite a different way in which the foundations of the theory of truth may be essentially strengthened.

[1] For this purpose we must nevertheless to some extent sharpen the premisses of the theorem by assuming that the class of all provable sentences of the metatheory is not only consistent, but also ω-consistent in the sense of Gödel, K. (22), p. 187, or in other words, that this class remains consistent after a single application of the rule of infinite induction, which will be discussed below.

[2] See p. 174, note 1.

The fact that we cannot infer from the correctness of all substitutions of such a sentential function as '$x \in Tr$ *or* $\bar{x} \in Tr$' the correctness of the sentence which is the generalization of this function, can be regarded as a symptom of a certain imperfection in the rules of inference hitherto used in the deductive sciences. In order to make good this defect we could adopt a new rule, the so-called *rule of infinite induction*, which in its application to the metatheory may be formulated somewhat as follows: if a given sentential function contains the symbol 'x', which belongs to the same semantical category as the names of expressions, as its only free variable, and if every sentence, which arises from the given function by substituting the structural-descriptive name of any expression of the language investigated for the variable 'x', is a provable theorem of the metatheory, then the sentence which we obtain from the phrase '*for every x, if x is an expression then p*' by substituting the given function for the symbol 'p', may also be added to the theorems of the metatheory. This rule can also be given another formulation which differs from the foregoing only by the fact that in it, instead of speaking about expressions, we speak of natural numbers; and instead of structural-descriptive names of expressions, the so-called specific symbols of natural numbers are dealt with, i.e. such symbols as '0', '1', '1+1', '1+1+1', and so on. In this form the rule of infinite induction recalls the principle of complete induction, which it exceeds considerably in logical power. Since it is possible to set up effectively a one-one correspondence between expressions and the natural numbers (cf. the proof of Th. I) it is easy to see that the two formulations are equivalent on the basis of the metatheory. But in the second formulation no specific concepts of the metalanguage occur at all, and for this reason it is applicable to many other deductive sciences. In the case where we are dealing with a science in the language of which there are no specific symbols for the natural numbers this formulation requires certain external modifications. For example, in order to formulate the rule for the general theory of classes, instead of substitutions of a given sentential function we must operate with expressions of the type '$\bigcup_1^3(\iota_k.p)$', where, in the place of

'p' the function in question occurs and the symbol 'ι_k' has the same meaning as in the proof of Th. I.[1]

On account of its non-finitist nature the rule of infinite induction differs fundamentally from the normal rules of inference. On each occasion of its use infinitely many sentences must be taken into consideration, although at no moment in the development of a science is such a number of previously proved theorems effectively given. It may well be doubted whether there is any place for the use of such a rule within the limits of the existing conception of the deductive method. The question whether this rule does not lead to contradictions presents no less serious difficulties than the analogous problem regarding the existing rules, even if we assume the consistency of the existing rules and permit the use of the new rule not only in the theory but also in the corresponding metatheory and in particular in any attempted proof of consistency. Nevertheless from the intuitive standpoint the rule of infinite induction seems to be as reliable as the rules normally applied: it always leads from true sentences to true sentences. In connexion with languages of finite order this fact can be strictly proved by means of the definition of truth constructed for these languages. The fact that this rule enables many problems to be solved which are not solvable on the basis of the old rules is in favour of the acceptance of the new rule, not only in the theory but also in the metatheory. By the introduction of this rule the class of provable sentences is enlarged by a much greater extent than by any supplementation of the list of axioms.[2] In the case of certain elementary deductive sciences, this enlargement is so great that the class of theorems becomes a complete system and coincides

[1] I have previously pointed out the importance of the rule of infinite induction in the year 1926. In a report to the Second Polish Philosophical Congress, in 1927, I have given, among other things, a simple example of a consistent deductive system which after a single application of this rule ceases to be consistent, and is therefore not ω-consistent (cf. p. 258, note 1; see also IX, p. 282, note 2). Some remarks on this rule are to be found in Hilbert, D. (29), pp. 491–2.

[2] Thus, for example, if we adopt this rule in the metalanguage without including it in the language, we can prove that the class of provable sentences of the science is consistent, which previously was not possible. In connexion with this problem cf. Gödel, K. (22), pp. 187–91 and 196.

with the class of true sentences. Elementary number theory provides an example, namely, the science in which all variables represent names of natural or whole numbers and the constants are the signs from the sentential and predicate calculi, the signs of zero, one, equality, sum, product and possibly other signs defined with their help.†

If it is decided to adopt the rule of infinite induction in the metatheory, then the system of axioms to which Th. III refers already forms a sufficient foundation for the development of the theory of truth. The proof of any of the known theorems in this field will then present no difficulty, in particular the Ths. 1–6 in § 3 and the theorem according to which the rule of infinite induction when applied to true sentences always yields true sentences. More important still, these axioms, together with the general axioms of the metatheory, form a categorical (although not a complete) system, and determine unambiguously the extension of the symbol 'Tr' which occurs in them.

Under these circumstances the question whether the theory erected on these foundations contains no inner contradiction acquires a special importance. Unfortunately this question cannot be finally decided at present. Th. I retains its full validity: in spite of the strengthening of the foundations of the metatheory the theory of truth cannot be constructed as a part of the morphology of language. On the other hand for the present we cannot prove Th. III for the enlarged metalanguage. The premiss which has played the most essential part in the original proof, i.e. the reduction of the consistency of the infinite axiom system to the consistency of every finite part of this system, now completely loses its validity—as is easily seen—on account of the content of the newly adopted rule. The possibility that the question cannot be decided in any direction is not excluded (at least on the basis of a 'normal' system of the meta-metatheory, which is constructed

† This last remark enables us to construct a rather simple definition of truth for elementary number theory without using our general method. The definition thus constructed can be further simplified. In fact we can first structurally describe all true sentences which contain no variables (or quantifiers), and then define an arbitrary sentence to be true if and only if it can be obtained from those elementary true sentences by applying the rule of infinite induction arbitrarily many times.

according to the principles given at the beginning of § 4 and does not contain the semantics of the metalanguage). On the other hand the possibility of showing Th. III to be false in its new interpretation seems to be unlikely from the intuitive viewpoint. One thing seems clear: the antinomy of the liar cannot be directly reconstructed either in the formulation met with in § 1 or in the form in which it appeared in the proof of Th. I. For here the axioms adopted in the theory of truth clearly possess, in contrast to colloquial language, the character of partial definitions. Through the introduction of the symbol '*Tr*' the metalanguage does not in any way become semantically universal, it does not coincide with the language itself and cannot be interpreted in that language (cf. pp. 158 and 248).[1]

No serious obstacles stand in the way of the application of the results obtained to other languages of infinite order. This is especially true of the most important of these results—Th. I. The languages of infinite order, thanks to the variety of meaningful expressions contained in them, provide sufficient means for the formulation of every sentence belonging to the arithmetic

[1] This last problem is equivalent to a seemingly more general problem of a methodological nature which can be formulated as follows. We presuppose the consistency of the metatheory supplemented by the rule of infinite induction. We consider an infinite sequence t of sentences of the metatheory; further we take into the metatheory a new primitive sign 'N', and add as axioms those and only those sentences which are obtained from the scheme '$n \in N$ *if and only if* p' by substituting for the sign 'n' the kth specific symbol of the natural numbers (i.e. the expression composed of k signs '1' separated from one another by the signs '+') and for the sign 'p' the kth term of the sequence t (k being here an arbitrary natural number). The question now arises whether the class of provable sentences of the metatheory, when enlarged in this way, remains consistent. This problem may be called the *problem of infinite recursive definitions*. The axiom system described in it can—from the intuitive standpoint—be regarded as a definition *sui generis* of the symbol 'N', which is distinguished from normal definitions only by the fact that it is formulated in infinitely many sentences. In view of this character of the axioms the possibility of a negative solution of the problem does not seem very probable. From Th. II and the interpretation of the metatheory in the theory itself, it is not difficult to infer that this problem can be solved in a positive sense in those cases in which the order of all variables which occur in the sentences of the sequence t is bounded above. It is then even possible to construct a definition of the symbol 'N' in the metatheory such that all the axioms mentioned follow from it. This problem obviously does not depend on the specific properties of the metatheory as such; it can also be presented in the same or in a somewhat modified form for other deductive sciences, e.g. for the general theory of classes.

of natural numbers and consequently enable the metalanguage to be interpreted in the language itself. It is thanks to just this circumstance that Th. I retains its validity for all languages of this kind.[1]

Some remarks may be added about those cases in which not single languages but whole classes of languages are investigated. As I have already emphasized in the Introduction, the concept of truth essentially depends, as regards both extension and content, upon the language to which it is applied. We can only meaningfully say of an expression that it is true or not if we treat this expression as a part of a concrete language. As soon as the discussion is about a large number of languages the expression 'true sentence' ceases to be unambiguous. If we are to avoid this ambiguity we must replace it by the relative term 'a true sentence with respect to the given language'. In order to make the sense of this term precise we apply to it essentially the same procedure as before: we construct a common metalanguage for all the languages of the given class; within the metalanguage we try to define the expression in question with the help of the methods developed in §§ 3 and 4. If we are not successful we add this term to the fundamental expressions of the metalanguage and by the axiomatic method determine its meaning according to the instructions of Th. III of this section. On account of the relativization of this term we should nevertheless expect *a priori* that in carrying out the plan sketched above the earlier difficulties would be significantly increased and quite new complications might arise (connected for example with the necessity of

[1] A reservation is necessary here: if we choose as our starting-point the classification of semantical categories sketched on p. 218, note 2, then we again encounter languages of infinite order for which Th. I loses its validity. A typical example is furnished by the language of Leśniewski's *Prototothetic* (cf. Leśniewski, S. (46)). In consequence of the 'finitistic' character of all the semantical categories of this language, it is easy to construct, in the metalanguage, a correct definition of truth, by choosing as model the matrix method from the extended sentential calculus. Moreover, such a definition can be obtained in other ways: as Leśniewski has shown, the class of provable sentences of the prototothetic is complete, and therefore the concept of provable sentence coincides in its extension with that of true sentence. Th. I on the other hand applies without restriction to all languages in which the order of the semantical categories from the domain of Leśniewski's *Ontology* (cf. Leśniewski, S. (47)) is not bounded above.

defining the word 'language'). I do not propose to discuss the problem touched upon in more detail in this place. The prospects for such investigations at the present time seem to be rather limited. In particular it would be incorrect to suppose that the relativization of the concept of truth—in the direction mentioned above—would open the way to some general theory of this concept which would embrace all possible or at least all formalized languages. The class of languages which is chosen as the object of simultaneous study must not be too wide. If, for example, we include in this class the metalanguage, which forms the field of the investigations and already contains the concept of truth, we automatically create the conditions which enable the antinomy of the liar to be reconstructed. The language of the general theory of truth would then contain a contradiction for exactly the same reason as does colloquial language.

In conclusion it may be mentioned that the results obtained can be extended to other semantical concepts, e.g. to the concept of satisfaction. For each of these concepts a system of postulates can be set up which (1) contains partial definitions analogous to the statements described in condition (α) of the convention **T** which determine the meaning of the given concept with respect to all concrete, structurally described expressions of a given class (e.g. with respect to sentences or sentential functions of a specific semantical type), and (2) contains a further postulate which corresponds to the sentence from the condition (β) of the same convention and stipulates that the concept in question may be applied only to expressions of the given class. We should be prepared to regard such a definition of the concept studied as a materially adequate one if its consequences included all the postulates of the above system. Methods which are similar to those described in §§ 3 and 4 enable the required definition to be constructed in all cases where we are dealing with languages of finite order, or, more generally, in which the semantical concept studied concerns exclusively linguistic expressions in which the order of the variables is bounded above (cf. Th. II). In the remaining cases it can be shown—after the pattern of the proof of Th. I—that no definition with the properties mentioned can be

formulated in the metalanguage.[1] In order to construct the theory of the concept studied in these cases also, it must be included in the system of primitive concepts, and the postulate described above must be included in the axiom system of the metatheory. A procedure analogous to the proof of Th. III proves that the system of the metalanguage supplemented in this way remains internally consistent. But the deductive power of the added postulates is very restricted. They do not suffice for the proof of the most important general theorems concerning the concept in question. They do not determine its extension unambiguously and the system obtained is not complete, nor is it even categorical. To remove this defect we must strengthen the foundations of the metatheory itself by adding the rule of infinite induction to its rules of inference. But then the proof of consistency would present great difficulties which we are not able at present to overcome.

§ 6. Summary

The principal results of this article may be summarized in the following theses:

A. *For every formalized language of finite order a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage, making use only of expressions of a general logical kind, expressions of the language itself as well as terms belonging to the morphology of language, i.e. names of linguistic expressions and of the structural relations existing between them.*

B. *For formalized languages of infinite order the construction of such a definition is impossible.*

[1] This especially concerns the concept of definability (although in this case both the formulation of the problem itself, as well as the method of solution, require certain modifications in comparison with the scheme put forward in the text). In VI, I have expressed the conjecture that it is impossible to define this concept in its full extent on the basis of the metalanguage. I can now prove this conjecture exactly. This fact is all the more noteworthy in that it is possible—as I have shown in the article mentioned—to construct the definitions of the particular cases of the concept of definability which apply, not to the whole language, but to any of its fragments of finite order, not only in the metalanguage but also in the language itself.

C. *On the other hand, even with respect to formalized languages of infinite order, the consistent and correct use of the concept of truth is rendered possible by including this concept in the system of primitive concepts of the metalanguage and determining its fundamental properties by means of the axiomatic method* (the question whether the theory of truth established in this way contains no contradiction remains for the present undecided).

Since the results obtained can easily be extended to other semantical concepts the above theses can be given a more general form:

A'. *The semantics of any formalized language of finite order can be built up as a part of the morphology of language, based on correspondingly constructed definitions.*

B'. *It is impossible to establish the semantics of the formalized languages of infinite order in this way.*

C'. *But the semantics of any formalized language of infinite order can be established as an independent science based upon its own primitive concepts and its own axioms, possessing as its logical foundation a system of the morphology of language* (although a full guarantee that the semantics constructed by this method contains no inner contradiction is at present lacking).

From the formal point of view the foregoing investigations have been carried out within the boundaries of the methodology of the deductive sciences. Some so to speak incidental results will perhaps be of interest to specialists in this field. I would draw attention to the fact that with the definition of true sentence for deductive sciences of finite order a general method has been obtained for proving their consistency (a method which, however, does not add greatly to our knowledge). I would point out also that it has been possible to define, for languages of finite order, the concepts of correct sentence in a given and in an arbitrary individual domain—concepts which play a great part in recent methodological studies.

But in its essential parts the present work deviates from the main stream of methodological investigations. Its central

problem—the construction of the definition of true sentence and establishing the scientific foundations of the theory of truth—belongs to the theory of knowledge and forms one of the chief problems of this branch of philosophy. I therefore hope that this work will interest the student of the theory of knowledge above all and that he will be able to analyse the results contained in it critically and to judge their value for further researches in this field, without allowing himself to be discouraged by the apparatus of concepts and methods used here, which in places have been difficult and have not hitherto been used in the field in which he works.

One word in conclusion. Philosophers who are not accustomed to use deductive methods in their daily work are inclined to regard all formalized languages with a certain disparagement, because they contrast these 'artificial' constructions with the one natural language—the colloquial language. For that reason the fact that the results obtained concern the formalized languages almost exclusively will greatly diminish the value of the foregoing investigations in the opinion of many readers. It would be difficult for me to share this view. In my opinion the considerations of § 1 prove emphatically that the concept of truth (as well as other semantical concepts) when applied to colloquial language in conjunction with the normal laws of logic leads inevitably to confusions and contradictions. Whoever wishes, in spite of all difficulties, to pursue the semantics of colloquial language with the help of exact methods will be driven first to undertake the thankless task of a reform of this language. He will find it necessary to define its structure, to overcome the ambiguity of the terms which occur in it, and finally to split the language into a series of languages of greater and greater extent, each of which stands in the same relation to the next in which a formalized language stands to its metalanguage. It may, however, be doubted whether the language of everyday life, after being 'rationalized' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages.

§ 7. Postscript

In writing the present article I had in mind only formalized languages possessing a structure which is in harmony with the theory of semantical categories and especially with its basic principles. This fact has exercised an essential influence on the construction of the whole work and on the formulation of its final results. It seemed to me then that 'the theory of the semantical categories penetrates so deeply into our fundamental intuitions regarding the meaningfulness of expressions, that it is hardly possible to imagine a scientific language whose sentences possess a clear intuitive meaning but whose structure cannot be brought into harmony with the theory in question in one of its formulations' (cf. p. 215). Today I can no longer defend decisively the view I then took of this question. In connexion with this it now seems to me interesting and important to inquire what the consequences would be for the basic problems of the present work if we included in the field under consideration formalized languages for which the fundamental principles of the theory of semantical categories no longer hold. In what follows I will briefly consider this question.

Although in this way the field to be covered is essentially enlarged, I do not intend—any more than previously—to consider all possible languages which someone might at some time construct. On the contrary I shall restrict myself exclusively to languages which—apart from differences connected with the theory of semantical categories—exhibit in their structure the greatest possible analogy with the languages previously studied. In particular, for the sake of simplicity, I shall consider only those languages in which occur, in addition to the universal and existential quantifiers and the constants of the sentential calculus, only individual names and the variables representing them, as well as constant and variable sentence-forming functors with arbitrary numbers of arguments. After the manner of the procedure in §§ 2 and 4 we try to specify for each of these languages the concepts of primitive sentential function, fundamental operations on expressions, sentential function in general,

axiom, consequence, and provable theorem. Thus, for example, we include as a rule among the axioms—just as in the language of the general theory of classes in § 5—the substitutions of the axioms of the sentential calculus, the pseudo-definitions and the law of extensionality (perhaps also other sentences, according to the specific peculiarities of the language). In determining the concept of consequence we take as our model Def. 15 in § 2.

The concept introduced in § 4 of the order of an expression plays a part which is no less essential than before in the construction of the language we are now considering. It is advisable to assign to names of individuals and to the variables representing them the order 0 (and not as before the order 1). The order of a sentence-forming functor of an arbitrary (primitive) sentential function is no longer unambiguously determined by the orders of all arguments of this function. Since the principles of the theory of the semantical categories no longer hold, it may happen that one and the same sign plays the part of a functor in two or more sentential functions in which arguments occupying respectively the same places nevertheless belong to different orders. Thus in order to fix the order of any sign we must take into account the orders of all arguments in all sentential functions in which this sign is a sentence-forming functor. If the order of all these arguments is smaller than a particular natural number n, and if there occurs in at least *one* sentential function an argument which is exactly of order $n-1$, then we assign to the symbol in question the order n. All such sentence-forming functors—as well as the names of individuals and the variables representing them—are included among the signs of finite order. But account must also be taken of the possibility that yet other sentence-forming functors may occur in the language to which an infinite order must be assigned. If, for example, a sign is a sentence-forming functor of only those sentential functions which have all their arguments of finite order, where, however, these orders are not bounded above by any natural number, then this sign will be of infinite order.

In order to classify the signs of infinite order we make use of the notion of *ordinal number*, taken from the theory of sets, which

is a generalization of the usual concept of natural number.[1] As is well known, the natural numbers are the smallest ordinal numbers. Since, for every infinite sequence of ordinal numbers, there are numbers greater than every term of the sequence, there are, in particular, numbers which are greater than all natural numbers. We call them *transfinite ordinal numbers*. It is known that in every non-empty class of ordinal numbers there is a smallest number. In particular there is a smallest transfinite number which is denoted by the symbol 'ω'. The next largest number is $\omega+1$, then follow the numbers $\omega+2$, $\omega+3$,..., $\omega.2$, $\omega.2+1$, $\omega.2+2$,..., $\omega.3$,..., and so on. To those signs of infinite order which are functors of sentential functions containing exclusively arguments of finite order we assign the number ω as their order. A sign which is a functor in only those sentential functions in which the arguments are either of finite order or of order ω (and in which at least *one* argument of a function is actually of order ω), is of the order $\omega+1$. The general recursive definition of order is as follows: the order of a particular sign is the smallest ordinal number which is greater than the orders of all arguments in all sentential functions in which the given sign occurs as a sentence-forming functor.[2]

Just as in § 4, we can distinguish languages of finite and infinite order. We can in fact assign to every language a quite specific ordinal number as its order, namely the smallest ordinal number which exceeds the orders of all variables occurring in this language (the former languages of the nth order—as can easily be shown—retain their former order under this convention because the order of the names of individuals has been diminished. The language of the general theory of classes has the order ω).

It does not at all follow from these stipulations that every variable in the languages in question is of a definite order. On the contrary it seems to me (by reason of trials and other considerations) almost certain that we cannot restrict ourselves to the use of variables of definite order if we are to obtain languages

[1] Cf. Fraenkel, A. (16), pp. 185 ff.

[2] Cf. the introduction of the system of levels in Carnap, R. (10), pp. 139 ff. (p. 186 in English translation).

which are actually superior to the previous languages in the abundance of the concepts which are expressible by their means, and the study of which could throw new light on the problems in which we are here interested. We must introduce into the languages variables of indefinite order which, so to speak, 'run through' all possible orders, which can occur as functors or arguments in sentential functions without regard to the order of the remaining signs occurring in these functions, and which at the same time may be both functors and arguments in the same sentential functions. With such variables we must proceed with the greatest caution if we are not to become entangled in antinomies like the famous antinomy of the class of all classes which are not members of themselves. Special care must be taken in formulating the rule of substitution for languages which contain such variables and in describing the axioms which we have called pseudodefinitions. But we cannot go into details here.[1]

There is obviously no obstacle to the introduction of variables of transfinite order not only into the language which is the object investigated, but also into the metalanguage in which the investigation is carried out. In particular it is always possible to construct the metalanguage in such a way that it contains variables

[1] From the languages just considered it is but a step to languages of another kind which constitute a much more convenient and actually much more frequently applied apparatus for the development of logic and mathematics. In these new languages all the variables are of indefinite order. From the formal point of view these are languages of a very simple structure; according to the terminology laid down in § 4 they must be counted among the languages of the first kind, since all their variables belong to one and the same semantical category. Nevertheless, as is shown by the investigations of E. Zermelo and his successors (cf. Skolem, Th. (66), pp. 1–12), with a suitable choice of axioms it is possible to construct the theory of sets and the whole of classical mathematics on the basis provided by this language. In it we can express so to speak every idea which can be formulated in the previously studied languages of finite and infinite order. For the languages here discussed the concept of order by no means loses its importance; it no longer applies, however, to the expressions of the language, but either to the objects denoted by them or to the language as a whole. Individuals, i.e. objects which are not sets, we call objects of order 0; the order of an arbitrary set is the smallest ordinal number which is greater than the orders of all elements of this set; the order of the language is the smallest ordinal number which exceeds the order of all sets whose existence follows from the axioms adopted in the language. Our further exposition also applies without restriction to the languages which have just been discussed.

of higher order than all the variables of the language studied. The metalanguage then becomes a language of higher order and thus one which is essentially richer in grammatical forms than the language we are investigating. This is a fact of the greatest importance from the point of view of the problems in which we are interested. For with this the distinction between languages of finite and infinite orders disappears—a distinction which was so prominent in §§ 4 and 5 and was strongly expressed in the theses A and B formulated in the Summary. In fact, the setting up of a correct definition of truth for languages of infinite order would in principle be possible provided we had at our disposal in the metalanguage expressions of higher order than all variables of the language investigated. The absence of such expressions in the metalanguage has rendered the extension of these methods of construction to languages of infinite order impossible. But now we are in a position to define the concept of truth for any language of finite or infinite order, provided we take as the basis for our investigations a metalanguage of an order which is at least greater by 1 than that of the language studied (an essential part is played here by the presence of variables of indefinite order in the metalanguage). It is perhaps interesting to emphasize that the construction of the definition is then to a certain degree simplified. We can adhere strictly to the method outlined in § 3 without applying the artifice which we were compelled to use in § 4 in the study of languages of the 2nd and 3rd kinds. We need neither apply many-rowed sequences nor carry out the semantical unification of the variables, for having abandoned the principles of the theory of semantical categories we can freely operate with sequences whose terms are of different orders. On the other hand the considerations brought forward in § 5 in connexion with Th. I lose none of their importance and can be extended to languages of any order. It is impossible to give an adequate definition of truth for a language in which the arithmetic of the natural numbers can be constructed, if the order of the metalanguage in which the investigations are carried out does not exceed the order of the language investigated (cf. the relevant remarks on p. 253).

Finally, the foregoing considerations show the necessity of revising, to a rather important extent, the Theses A and B given in the conclusions of this work and containing a summary of its chief results:

A. *For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in the metalanguage with the help only of general logical expressions, of expressions of the language itself, and of terms from the morphology of language—but under the condition that the metalanguage possesses a higher order than the language which is the object of investigation.*

B. *If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed.*

From a comparison of the new formulation of the two theses with the earlier one it will be seen that the range of the results obtained has been essentially enlarged, and at the same time the conditions for their application have been made more precise.

In view of the new formulation of Thesis A the former Thesis C loses its importance. It possesses a certain value only when the investigations are carried out in a metalanguage which has the same order as the language studied and when, having abandoned the construction of a definition of truth, the attempt is made to build up the theory of truth by the axiomatic method. It is easy to see that a theory of truth built up in this way cannot contain an inner contradiction, provided there is freedom from contradiction in the metalanguage of higher order on the basis of which an adequate definition of truth can be set up and in which those theorems which are adopted in the theory of truth as axioms can be derived.[1]

Just as in the conclusion of this work, the Theses A and B can be given a more general formulation by extending them to other semantical concepts:

A′. *The semantics of any formalized language can be established as a part of the morphology of language based on suitably constructed*

[1] In particular, the question broached on p. 261 has a positive answer. The same also holds for the problem of infinite inductive definitions mentioned on p. 262, footnote.

definitions, provided, however, that the language in which the morphology is carried out has a higher order than the language whose morphology it is.

B′. *It is impossible to establish the semantics of a language in this way if the order of the language of its morphology is at most equal to that of the language itself.*

The Thesis A in its new generalized form is of no little importance for the methodology of the deductive sciences. Its consequences run parallel with the important results which Gödel has reported in this field in recent years. The definition of truth allows the consistency of a deductive science to be proved on the basis of a metatheory which is of higher order than the theory itself (cf. pp. 199 and 236). On the other hand, it follows from Gödel's investigations that it is in general impossible to prove the consistency of a theory if the proof is sought on the basis of a metatheory of equal or lower order.[1] Moreover Gödel has given a method for constructing sentences which—assuming the theory concerned to be consistent—cannot be decided in any direction in this theory. All sentences constructed according to Gödel's method possess the property that it can be established whether they are true or false on the basis of the metatheory of higher order having a correct definition of truth. Consequently it is possible to reach a decision regarding these sentences, i.e. they can be either proved or disproved. Moreover a decision can be reached within the science itself, without making use of the concepts and assumptions of the metatheory—provided, of course, that we have previously enriched the language and the logical foundations of the theory in question by the introduction of variables of higher order.[2]

Let us try to explain this somewhat more exactly. Consider an arbitrary deductive science in which the arithmetic of natural numbers can be constructed, and provisionally begin the investigation on the basis of a metatheory of the same order as the theory itself. Gödel's method of constructing undecidable sentences has been outlined implicitly in the proof of Th. I in

[1] Cf. Gödel, K. (22), p. 196 (Th. XI).

[2] Cf. Gödel, K. (22), pp. 187 ff., and in particular, p. 191, note 48 *a*.

§ 5 (p. 249 ff.). Everywhere, both in the formulation of the theorem and in its proof, we replace the symbol '*Tr*' by the symbol '*Pr*' which denotes the class of all provable sentences of the theory under consideration and can be defined in the metatheory (cf. e.g. Def. 17 in § 2). In accordance with the first part of Th. I we can obtain the negation of one of the sentences in condition (α) of convention T of § 3 as a consequence of the definition of the symbol '*Pr*' (provided we replace '*Tr*' in this convention by '*Pr*'). In other words we can construct a sentence x of the science in question which satisfies the following condition:

it is not true that $x \in Pr$ *if and only if* p

or in equivalent formulation:

(1) $$x \bar{\in} Pr \text{ if and only if } p$$

where the symbol 'p' represents the whole sentence x (in fact we may choose the sentence $\bigcup_1^3(\iota_k . \phi_k)$ constructed in the proof of Th. I as x).

We shall show that the sentence x is actually undecidable and at the same time true. For this purpose we shall pass to a metatheory of higher order; Th. I then obviously remains valid. According to Thesis A we can construct, on the basis of the enriched metatheory, a correct definition of truth concerning all the sentences of the theory studied. If we denote the class of all true sentences by the symbol '*Tr*' then—in accordance with convention T—the sentence x which we have constructed will satisfy the following condition:

(2) $$x \in Tr \text{ if and only if } p;$$

from (1) and (2) we obtain immediately

(3) $$x \bar{\in} Pr \text{ if and only if } x \in Tr.$$

Moreover, if we denote the negation of the sentence x by the symbol '$\bar{x}$' we can derive the following theorems from the definition of truth (cf. Ths. 1 and 5 in § 3):

(4) $$\text{either } x \bar{\in} Tr \text{ or } \bar{x} \bar{\in} Tr;$$

(5) $$\text{if } x \in Pr, \text{ then } x \in Tr;$$

(6) $$\text{if } \bar{x} \in Pr, \text{ then } \bar{x} \in Tr;$$

From (3) and (5) we infer without difficulty that

(7) $$x \in Tr$$

and that

(8) $$x \overline{\in}^{1} Pr.$$

In view of (4) and (7) we have $\bar{x} \overline{\in} Tr$, which together with (6) gives the formula

(9) $$\bar{x} \overline{\in} Pr.$$

The formulas (8) and (9) together express the fact that x is an undecidable sentence; moreover from (7) it follows that x is a true sentence.

By establishing the truth of the sentence x we have *eo ipso* —by reason of (2)—also proved x itself in the metatheory. Since, moreover, the metatheory can be interpreted in the theory enriched by variables of higher order (cf. p. 184) and since in this interpretation the sentence x, which contains no specific term of the metatheory, is its own correlate, the proof of the sentence x given in the metatheory can automatically be carried over into the theory itself: the sentence x which is undecidable in the original theory becomes a decidable sentence in the enriched theory.

I should like to draw attention here to an analogous result. For every deductive science in which arithmetic is contained it is possible to specify arithmetical notions which, so to speak, belong intuitively to this science, but which cannot be defined on the basis of this science. With the help of methods which are completely analogous to those used in the construction of the definition of truth, it is nevertheless possible to show that these concepts can be so defined provided the science is enriched by the introduction of variables of higher order.[1]

In conclusion it can be affirmed that the definition of truth and, more generally, the establishment of semantics enables us to match some important negative results which have been obtained

[1] Cf. my review, 'Über definierbare Mengen reeller Zahlen,' *Annales de la Société Polonaise de Mathématique*, t. ix, année 1930, Kraków, 1931, pp. 206–7 (report on a lecture given on 16 December 1930 at the Lemberg Section of the Polish Mathematical Society); the ideas there sketched were in part developed later in VI.

in the methodology of the deductive sciences with parallel positive results, and thus to fill up in some measure the gaps thereby revealed in the deductive method and in the edifice of deductive knowledge itself.

HISTORICAL NOTES. In the course of the years 1929 to 1935, in which I reached the final definition of the concept of truth and most of the remaining results described here, and in the last year of which the whole work appeared for the first time in a universal language, the questions here discussed have been treated several times. In the German language, in addition to my summary, Tarski, A. (76), works by Carnap have appeared in which quite similar ideas were developed (cf. R. Carnap, 'Die Antinomien und die Unvollständigkeit der Mathematik', *Monatshefte f. Math. u. Phys.* vol. 41 (Leipzig, 1934), pp. 263–84 and 'Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik, ibid., vol. 42, pp. 163–90; both articles being supplementations of R. Carnap (10)). The two articles have been incorporated in the English edition of Carnap's book, entitled *Logical Syntax of Language* (London, 1937).

It was to be expected that, in consequence of this lapse of six years, and of the nature of the problem and perhaps also of the language of the original text of my work, errors regarding the historical connexions might occur. And in fact Carnap writes in the second of the above-mentioned articles regarding my investigations that they have been carried out '. . . in connexion with those of Gödel . . .'. It will therefore not be superfluous if I make some remarks in this place about the dependence or independence of my studies.

I may say quite generally that all my methods and results, with the exception of those at places where I have expressly emphasized this—cf. footnotes, pp. 154 and 247—were obtained by me quite independently. The dates given in footnote, p. 154, provide, I believe, sufficient basis for testing this assertion. I may point out further that my article which appeared in French (VI), about which I had already reported in December 1930 (cf. the report in German in A. Tarski (74)) contains precisely those methods of construction which were used there for other purposes but in the present work for the construction of the definition of truth.

I should like to emphasize the independence of my investigations regarding the following points of detail: (1) the general formulation of the problem of defining truth, cf. especially pp. 187–8; (2) the positive solution of the problem, i.e. the definition of the concept of truth for the case where the means available in the metalanguage are sufficiently rich (for logical languages this definition becomes that of the term 'analytical' used by Carnap). Cf. pp. 194 and 236; (3) the method of proving consistency on the basis of the definition of truth, cf. pp. 199 and 236; (4) the axiomatic construction of the metasystem, cf. pp. 173 ff., and in connexion with this (5) the discussions on pp. 184 f. on the interpreta-

tion of the metasystem in arithmetic, which already contain the so-called 'method of arithmetizing the metalanguage' which was developed far more completely and quite independently by Gödel. Moreover, I should like to draw attention to results not relating to the concept of truth but to another semantical concept, that of definability reported on p. 276.

In the one place in which my work is connected with the ideas of Gödel—in the negative solution of the problem of the definition of truth for the case where the metalanguage is not richer than the language investigated—I have naturally expressly emphasized this fact (cf. p. 247, footnote); it may be mentioned that the result so reached, which very much completed my work, was the only one subsequently added to the otherwise already finished investigation.

IX

SOME OBSERVATIONS ON THE CONCEPTS OF ω-CONSISTENCY AND ω-COMPLETENESS†

In an extremely interesting article Gödel[1] introduces the concept of ω-consistency, and constructs an example of a deductive system which is consistent in the usual sense, but is not ω-consistent. In the present article I propose to give another simple example of such a system, together with some general remarks on the concept mentioned as well as on the corresponding concept of ω-completeness.[2]

The symbolical language in which I shall construct this system is closely related to the language of the system P used by Gödel. It is also the result of an exact formalization and simplification, as far as possible, of the language in which the system of *Principia Mathematica* of Whitehead and Russell[3] is constructed. In spite of its great simplicity this language suffices for the expression of every idea which can be formulated in *Principia Mathematica*.[4]

[1] See Gödel, K. (22).

[2] Already, in the year 1927, at the Second Conference of the Polish Philosophical Society in Warsaw in the lecture 'Remarks on some notions of the methodology of the deductive sciences' (listed by title in *Ruch Filozoficzny*, vol. 10 (1926–7), p. 96), I had pointed out the importance of these two concepts, and the rule of transfinite induction which is closely related to them and about which more is said in the text, but I had not suggested special names for these concepts. I also communicated the example of a consistent and yet not ω-consistent system which I give in the present article in a slightly altered form. Naturally it is not hereby claimed that I already knew then the results later obtained by Gödel or had even foreseen them. On the contrary, I had personally felt that the publication of the work of Gödel cited above was a most exciting scientific event.

[3] Whitehead, A. N., and Russell, B. A. W. (90).

[4] Cf. articles VI and VIII of the present work, where I have used the same or a very similar language.

† Bibliographical Note. This article first appeared under the title 'Einige Betrachtungen über die Begriffe ω-Widerspruchsfreiheit und der ω-Vollständigkeit', *Monatshefte für Mathematik und Physik*, vol. 40 (1933), pp. 97–112. For the historical information about the results of this article see footnote 2 above.