Oxford University Press, Amen House, London E.C.4
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS KARACHI CAPE TOWN IBADAN
Geoffrey Cumberlege, Publisher to the University

LOGIC, SEMANTICS, METAMATHEMATICS

PAPERS FROM 1923 TO 1938

BY
ALFRED TARSKI

TRANSLATED BY
J. H. WOODGER

OXFORD
AT THE CLARENDON PRESS
1956

ıected with the investigations of Gödel. As is well known, :el has developed a method which makes it possible, in every ›ry which includes the arithmetic of natural numbers as a ;, to construct sentences which can be neither proved nor ›roved in this theory. But he has also pointed out that the ecidable sentences constructed by this method become de- ,ble if the theory under investigation is enriched by the ition of variables of higher type. The proof that the sen- ›es involved actually in this way become decidable again s on the definition of truth. Similarly—as I have shown means of the methods used in developing semantics—for given deductive theory it is possible to indicate concepts ch cannot be defined in this theory, although in their con- ; they belong to the theory, and become definable in it if theory is enriched by the introduction of higher types. ımarizing, we can say that the establishment of scientific antics, and in particular the definition of truth, enables us natch the negative results in the field of metamathematics ı corresponding positive ones, and in that way to fill to e extent the gaps which have been revealed in the deductive ,hod and in the very structure of deductive science.[1]

More detailed information about many of the problems discussed in this le can be found in VIII. Attention should also be called to my later ›r, Tarski, A. (82). While the first part of that paper is close in its content ıe present article, the second part contains polemical remarks regarding ›us objections which have been raised against my investigations in the of semantics. It also includes some observations about the applicability ›mantics to empirical sciences and their methodology.

XVI

ON THE CONCEPT OF LOGICAL CONSEQUENCE†

THE concept of *logical consequence* is one of those whose introduction into the field of strict formal investigation was not a matter of arbitrary decision on the part of this or that investigator; in defining this concept, efforts were made to adhere to the common usage of the language of everyday life. But these efforts have been confronted with the difficulties which usually present themselves in such cases. With respect to the clarity of its content the common concept of consequence is in no way superior to other concepts of everyday language. Its extension is not sharply bounded and its usage fluctuates. Any attempt to bring into harmony all possible vague, sometimes contradictory, tendencies which are connected with the use of this concept, is certainly doomed to failure. We must reconcile ourselves from the start to the fact that every precise definition of this concept will show arbitrary features to a greater or less degree.

Even until recently many logicians believed that they had succeeded, by means of a relatively meagre stock of concepts, in grasping almost exactly the content of the common concept of consequence, or rather in defining a new concept which coincided in extent with the common one. Such a belief could easily arise amidst the new achievements of the methodology of deductive science. Thanks to the progress of mathematical logic we have learnt, during the course of recent decades, how to present mathematical disciplines in the shape of formalized deductive theories. In these theories, as is well known, the

† BIBLIOGRAPHICAL NOTE. This is a summary of an address given at the International Congress of Scientific Philosophy in Paris, 1935. The article first appeared in print in Polish under the title 'O pojciu wynikania logicznego' in *Przegląd Filozoficzny*, vol. 39 (1936), pp. 58–68, and then in German under the title 'Über den Begriff der logischen Folgerung', *Actes du Congrès International de Philosophie Scientifique*, vol. 7 (Actualités Scientifiques et Industrielles, vol. 394), Paris, 1936, pp. 1–11.

proof of every theorem reduces to single or repeated application of some simple rules of inference—such as the rules of substitution and detachment. These rules tell us what transformations of a purely structural kind (i.e. transformations in which only the external structure of sentences is involved) are to be performed upon the axioms or theorems already proved in the theory, in order that the sentences obtained as a result of such transformations may themselves be regarded as proved. Logicians thought that these few rules of inference exhausted the content of the concept of consequence. Whenever a sentence follows from others, it can be obtained from them—so it was thought—in more or less complicated ways by means of the transformations prescribed by the rules. In order to defend this view against sceptics who doubted whether the concept of consequence when formalized in this way really coincided in extent with the common one, the logicians were able to bring forward a weighty argument: the fact that they had actually succeeded in reproducing in the shape of formalized proofs all the exact reasonings which had ever been carried out in mathematics.

Nevertheless we know today that the scepticism was quite justified and that the view sketched above cannot be maintained. Some years ago I gave a quite elementary example of a theory which shows the following peculiarity: among its theorems there occur such sentences as:

A_0. *0 possesses the given property P*,

A_1. *1 possesses the given property P*,

and, in general, all particular sentences of the form

A_n. *n possesses the given property P*,

where 'n' represents any symbol which denotes a natural number in a given (e.g. decimal) number system. On the other hand the universal sentence:

A. *Every natural number possesses the given property P*,

cannot be proved on the basis of the theory in question by means of the normal rules of inference.[1] This fact seems to me to speak

[1] For a detailed description of a theory with this peculiarity see IX; for the discussion of the closely related rule of infinite induction see VIII, pp. 258 ff.

for itself. It shows that the formalized concept of consequence, as it is generally used by mathematical logicians, by no means coincides with the common concept. Yet intuitively it seems certain that the universal sentence A follows in the usual sense from the totality of particular sentences $A_0, A_1, \ldots, A_n, \ldots$. Provided all these sentences are true, the sentence A must also be true.

In connexion with situations of the kind just described it has proved to be possible to formulate new rules of inference which do not differ from the old ones in their logical structure, are intuitively equally infallible, i.e. always lead from true sentences to true sentences, but cannot be reduced to the old rules. An example of such a rule is the so-called rule of infinite induction according to which the sentence A can be regarded as proved provided all the sentences $A_0, A_1, \ldots, A_n, \ldots$ have been proved (the symbols 'A_0', 'A_1', etc., being used in the same sense as previously). But this rule, on account of its infinitistic nature, is in essential respects different from the old rules. It can only be applied in the construction of a theory if we have first succeeded in proving infinitely many sentences of this theory—a state of affairs which is never realized in practice. But this defect can easily be overcome by means of a certain modification of the new rule. For this purpose we consider the sentence B which asserts that all the sentences $A_0, A_1, \ldots, A_n, \ldots$ are *provable* on the basis of the rules of inference hitherto used (not that they have actually been proved). We then set up the following rule: if the sentence B is proved, then the corresponding sentence A can be accepted as proved. But here it might still be objected that the sentence B is not at all a sentence of the theory under construction, but belongs to the so-called metatheory (i.e. the theory *of* the theory discussed) and that in consequence a practical application of the rule in question will always require a transition from the theory to the metatheory.[1] In order to avoid this objection we shall restrict

[1] For the concept of metatheory and the problem of the interpretation of a metatheory in the corresponding theory see article VIII, pp. 167 ff., 184, and 247 ff.

consideration to those deductive theories in which the arithmetic of natural numbers can be developed, and observe that in every such theory all the concepts and sentences of the corresponding metatheory can be interpreted (since a one-one correspondence can be established between expressions of a language and natural numbers).[1] We can replace in the rule discussed the sentence B by the sentence B', which is the arithmetical interpretation of B. In this way we reach a rule which does not deviate essentially from the rules of inference, either in the conditions of its applicability or in the nature of the concepts involved in its formulation or, finally, in its intuitive infallibility (although it is considerably more complicated).

Now it is possible to state other rules of like nature, and even as many of them as we please. Actually it suffices in fact to notice that the rule last formulated is essentially dependent upon the extension of the concept 'sentence provable on the basis of the rules hitherto used'. But when we adopt this rule we thereby widen the extension of this concept. Then, for the widened extension we can set up a new, analogous rule, and so on *ad infinitum*. It would be interesting to investigate whether there are any objective reasons for assigning a special position to the rules ordinarily used.

The conjecture now suggests itself that we can finally succeed in grasping the full intuitive content of the concept of consequence by the method sketched above, i.e. by supplementing the rules of inference used in the construction of deductive theories. By making use of the results of K. Gödel[2] we can show that this conjecture is untenable. In every deductive theory (apart from certain theories of a particularly elementary nature), however much we supplement the ordinary rules of inference by new purely structural rules, it is possible to construct sentences which follow, in the usual sense, from the theorems of this theory, but which nevertheless cannot be proved in this theory on the basis of the accepted rules of

[1] For the concept of metatheory and the problem of the interpretation of a metatheory in the corresponding theory see article VIII, pp. 167 ff., 184, and 247 ff.

[2] Cf. Gödel, K. (22), especially pp. 190 f.

inference.[1] In order to obtain the proper concept of consequence, which is close in essentials to the common concept, we must resort to quite different methods and apply quite different conceptual apparatus in defining it. It is perhaps not superfluous to point out in advance that—in comparison with the new—the old concept of consequence as commonly used by mathematical logicians in no way loses its importance. This concept will probably always have a decisive significance for the practical construction of deductive theories, as an instrument which allows us to prove or disprove particular sentences of these theories. It seems, however, that in considerations of a general theoretical nature the proper concept of consequence must be placed in the foreground.[2]

The first attempt to formulate a precise definition of the proper concept of consequence was that of R. Carnap.[3] But this

[1] In order to anticipate possible objections the range of application of the result just formulated should be determined more exactly and the logical nature of the rules of inference exhibited more clearly; in particular it should be exactly explained what is meant by the structural character of these rules.

[2] An opposition between the two concepts in question is clearly pointed out in article IX, pp. 293 ff. Nevertheless, in contrast to my present standpoint, I have there expressed myself in a decidedly negative manner about the possibility of setting up an exact formal definition for the proper concept of consequence. My position at that time is explained by the fact that, when I was writing the article mentioned, I wished to avoid any means of construction which went beyond the theory of logical types in any of its classical forms; but it can be shown that it is impossible to define the proper concept of consequence adequately whilst using exclusively the means admissible in the classical theory of types; unless we should thus limit our considerations solely to formalized languages of an elementary and fragmentary character (to be exact, to the so-called languages of finite order; cf. article VIII, especially pp. 268 ff.). In his extremely interesting book, Carnap, R. (10), the term (*logical*) *derivation* or *derivability* is applied to the old concept of consequence as commonly used in the construction of deductive theories, in order to distinguish it from the concept of *consequence* as the proper concept. The opposition between the two concepts is extended by Carnap to the most diverse derived concepts ('f-*concepts*' and '*a-concepts*', cf. pp. 88 ff., and 124 ff.); he also emphasizes—to my mind correctly—the importance of the proper concept of consequence and the concepts derived from it, for general theoretical discussions (cf. e.g. p. 128).

[3] Cf. Carnap, R. (10), pp. 88 f., and Carnap, R. (11) especially p. 181. In the first of these works there is yet another definition of consequence which is adapted to a formalized language of an elementary character. This definition is not considered here because it cannot be applied to languages of a more complicated logical structure. Carnap attempts to define the concept of logical consequence not only for special languages, but also within the framework of what he calls '*general syntax*'. We shall have more to say about this on p. 416, note 1.

attempt is connected rather closely with the particular properties of the formalized language which was chosen as the subject of investigation. The definition proposed by Carnap can be formulated as follows:

The sentence X follows logically *from the sentences of the class K if and only if the class consisting of all the sentences of K and of the negation of X is contradictory.*

The decisive element of the above definition obviously is the concept 'contradictory'. Carnap's definition of this concept is too complicated and special to be reproduced here without long and troublesome explanations.[1]

I should like to sketch here a general method which, it seems to me, enables us to construct an adequate definition of the concept of consequence for a comprehensive class of formalized languages. I emphasize, however, that the proposed treatment of the concept of consequence makes no very high claim to complete originality. The ideas involved in this treatment will certainly seem to be something well known, or even something of his own, to many a logician who has given close attention to the concept of consequence and has tried to characterize it more precisely. Nevertheless it seems to me that only the methods which have been developed in recent years for the establishment of scientific semantics, and the concepts defined with their aid, allow us to present these ideas in an exact form.[2]

Certain considerations of an intuitive nature will form our starting-point. Consider any class K of sentences and a sentence X which follows from the sentences of this class. From an intuitive standpoint it can never happen that both the class K consists only of true sentences and the sentence X is false. Moreover, since we are concerned here with the concept of logical, i.e. *formal*, consequence, and thus with a relation which is to be uniquely determined by the form of the sentences between which it holds, this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of the

[1] See footnote 3 on p. 413.

[2] The methods and concepts of semantics and especially the concepts of truth and satisfaction are discussed in detail in article VIII; see also article XV.

objects to which the sentence X or the sentences of the class K refer. The consequence relation cannot be affected by replacing the designations of the objects referred to in these sentences by the designations of any other objects. The two circumstances just indicated, which seem to be very characteristic and essential for the proper concept of consequence, may be jointly expressed in the following statement:

(F) *If, in the sentences of the class K and in the sentence X, the constants—apart from purely logical constants—are replaced by any other constants (like signs being everywhere replaced by like signs), and if we denote the class of sentences thus obtained from K by 'K′', and the sentence obtained from X by 'X′', then the sentence X′ must be true provided only that all sentences of the class K′ are true.*

[For the sake of simplifying the discussion certain incidental complications are disregarded, both here and in what follows. They are connected partly with the theory of logical types, and partly with the necessity of eliminating any defined signs which may possibly occur in the sentences concerned, i.e. of replacing them by primitive signs.]

In the statement (F) we have obtained a necessary condition for the sentence X to be a consequence of the class K. The question now arises whether this condition is also sufficient. If this question were to be answered in the affirmative, the problem of formulating an adequate definition of the concept of consequence would be solved affirmatively. The only difficulty would be connected with the term 'true' which occurs in the condition (F). But this term can be exactly and adequately defined in semantics.[1]

Unfortunately the situation is not so favourable. It may, and it does, happen—it is not difficult to show this by considering special formalized languages—that the sentence X does not follow in the ordinary sense from the sentences of the class K although the condition (F) is satisfied. This condition may in fact be satisfied only because the language with which we are

[1] See footnote 2 on p. 414.

dealing does not possess a sufficient stock of extra-logical constants. The condition (F) could be regarded as sufficient for the sentence X to follow from the class K only if the designations of all possible objects occurred in the language in question. This assumption, however, is fictitious and can never be realized.[1] We must therefore look for some means of expressing the intentions of the condition (F) which will be completely independent of that fictitious assumption.

Such a means is provided by semantics. Among the fundamental concepts of semantics we have the concept of the *satisfaction of a sentential function* by single objects or by a sequence of objects. It would be superfluous to give here a precise explanation of the content of this concept. The intuitive meaning of such phrases as: *John and Peter satisfy the condition 'X and Y are brothers'*, or *the triple of numbers* 2, 3, *and* 5 *satisfies the equation* '$x+y = z$', can give rise to no doubts. The concept of satisfaction—like other semantical concepts—must always be relativized to some particular language. The details of its precise definition depend on the structure of this language. Nevertheless, a general method can be developed which enables us to construct such definitions for a comprehensive class of formalized languages. Unfortunately, for technical reasons, it would be impossible to sketch this method here even in its general outlines.[2]

One of the concepts which can be defined in terms of the concept of satisfaction is the concept of *model*. Let us assume that in the language we are considering certain variables correspond to every extra-logical constant, and in such a way that every sentence becomes a sentential function if the constants in it are replaced by the corresponding variables. Let L be any class of sentences. We replace all extra-logical constants which

[1] These last remarks constitute a criticism of some earlier attempts to define the concept of formal consequence. They concern, in particular, Carnap's definitions of logical consequence and a series of derivative concepts (L-consequences and L-concepts, cf. Carnap, R. (10), pp. 137 ff.). These definitions, in so far as they are set up on the basis of 'general syntax', seem to me to be materially inadequate, just because the defined concepts depend essentially, in their extension, on the richness of the language investigated.

[2] See footnote 2 on p. 414.

occur in the sentences belonging to L by corresponding variables, like constants being replaced by like variables, and unlike by unlike. In this way we obtain a class L' of sentential functions. An arbitrary sequence of objects which satisfies every sentential function of the class L' will be called a *model* or *realization of the class L of sentences* (in just this sense one usually speaks of models of an axiom system of a deductive theory). If, in particular, the class L consists of a single sentence X, we shall also call the model of the class L the *model of the sentence X*.

In terms of these concepts we can define the concept of logical consequence as follows:

The sentence X follows logically *from the sentences of the class K if and only if every model of the class K is also a model of the sentence X*.†

It seems to me that everyone who understands the content of the above definition must admit that it agrees quite well with common usage. This becomes still clearer from its various consequences. In particular, it can be proved, on the basis of this definition, that every consequence of true sentences must be true, and also that the consequence relation which holds between given sentences is completely independent of the sense of the extra-logical constants which occur in these sentences. In brief, it can be shown that the condition (F) formulated above is necessary if the sentence X is to follow from the sentences of the class K. On the other hand, this condition is in general not sufficient, since the concept of consequence here defined (in agreement with the standpoint we have taken) is independent of the richness in concepts of the language being investigated.

Finally, it is not difficult to reconcile the proposed definition with that of Carnap. For we can agree to call a class of sentences

† After the original of this paper had appeared in print, H. Scholz in his article 'Die Wissenschaftslehre Bolzanos, Eine Jahrhundert-Betrachtung', *Abhandlungen der Fries'schen Schule*, new series, vol. 6, pp. 399–472 (see in particular p. 472, footnote 58) pointed out a far-reaching analogy between this definition of consequence and the one suggested by B. Bolzano about a hundred years earlier.

contradictory if it possesses no model. Analogously, a class of sentences can be called *analytical* if every sequence of objects is a model of it. Both of these concepts can be related not only to classes of sentences but also to single sentences. Let us assume further that, in the language with which we are dealing, for every sentence X there exists a negation of this sentence, i.e. a sentence Y which has as a model those and only those sequences of objects which are not models of the sentence X (this assumption is rather essential for Carnap's construction). On the basis of all these conventions and assumptions it is easy to prove the *equivalence of the two definitions*. We can also show —just as does Carnap—that those and only those sentences are analytical which follow from every class of sentences (in particular from the empty class), and those and only those are contradictory from which every sentence follows.[1]

I am not at all of the opinion that in the result of the above discussion the problem of a materially adequate definition of the concept of consequence has been completely solved. On the contrary, I still see several open questions, only one of which—perhaps the most important—I shall point out here.

Underlying our whole construction is the division of all terms of the language discussed into logical and extra-logical. This division is certainly not quite arbitrary. If, for example, we were to include among the extra-logical signs the implication sign, or the universal quantifier, then our definition of the concept of consequence would lead to results which obviously contradict ordinary usage. On the other hand, no objective grounds are known to me which permit us to draw a sharp

[1] Cf. Carnap, R. (10), pp. 135 ff., especially Ths. 52.7 and 52.8; Carnap, R. (11), p. 182, Ths. 10 and 11. Incidentally I should like to remark that the definition of the concept of consequence here proposed does not exceed the limits of syntax in Carnap's conception (cf. Carnap, R. (10), pp. 6 ff.). Admittedly the general concept of satisfaction (or of model) does not belong to syntax; but we use only a special case of this concept—the satisfaction of sentential functions which contain no extra-logical constants, and this special case can be characterized using only general logical and specific syntactical concepts. Between the general concept of satisfaction and the special case of this concept used here approximately the same relation holds as between the semantical concept of true sentence and the syntactical concept of analytical sentence.

boundary between the two groups of terms. It seems to be possible to include among logical terms some which are usually regarded by logicians as extra-logical without running into consequences which stand in sharp contrast to ordinary usage. In the extreme case we could regard all terms of the language as logical. The concept of *formal* consequence would then coincide with that of *material* consequence. The sentence X would in this case follow from the class K of sentences if either X were true or at least one sentence of the class K were false.[1]

In order to see the importance of this problem for certain general philosophical views it suffices to note that the division of terms into logical and extra-logical also plays an essential part in clarifying the concept 'analytical'. But according to many logicians this last concept is to be regarded as the exact formal correlate of the concept of *tautology* (i.e. of a statement

[1] It will perhaps be instructive to juxtapose the three concepts: 'derivability' (cf. p. 413, note 2), 'formal consequence', and 'material consequence', for the special case when the class K, from which the given sentence X follows, consists of only a finite number of sentences: $Y_1, Y_2, \ldots, Y_n$. Let us denote by the symbol 'Z' the conditional sentence (the implication) whose antecedent is the conjunction of the sentences $Y_1, Y_2, \ldots, Y_n$ and whose consequent is the sentence X. The following equivalences can then be established:

the sentence X is (logically) derivable from the sentences of the class K if and only if the sentence Z is logically provable (i.e. derivable from the axioms of logic);

the sentence X follows formally from the sentences of the class K if and only if the sentence Z is analytical;

the sentence X follows materially from the sentences of the class K if and only if the sentence Z is true.

Of the three equivalences only the first can arouse certain objections; cf. article XII, pp. 342–64, especially 346. In connexion with these equivalences cf. also Ajdukiewicz, K. (2), p. 19, and (4), pp. 14 and 42.

In view of the analogy indicated between the several variants of the concept of consequence, the question presents itself whether it would not be useful to introduce, in addition to the special concepts, a general concept of a relative character, and indeed the concept of *consequence with respect to a class L of sentences*. If we make use again of the previous notation (limiting ourselves to the case when the class K is finite), we can define this concept as follows:

the sentence X follows *from the sentences of the class K* with respect to the class L of sentences *if and only if the sentence Z belongs to the class L.*

On the basis of this definition, derivability would coincide with consequence with respect to the class of all logically provable sentences, formal consequences would be consequences with respect to the class of all analytical sentences, and material consequences those with respect to the class of all true sentences.

which 'says nothing about reality'), a concept which to me personally seems rather vague, but which has been of fundamental importance for the philosophical discussions of L. Wittgenstein and the whole Vienna Circle.[1]

Further research will doubtless greatly clarify the problem which interests us. Perhaps it will be possible to find important objective arguments which will enable us to justify the traditional boundary between logical and extra-logical expressions. But I also consider it to be quite possible that investigations will bring no positive results in this direction, so that we shall be compelled to regard such concepts as 'logical consequence', 'analytical statement', 'and 'tautology' as relative concepts which must, on each occasion, be related to a definite, although in greater or less degree arbitrary, division of terms into logical and extra-logical. The fluctuation in the common usage of the concept of consequence would—in part at least—be quite naturally reflected in such a compulsory situation.

[1] Cf. Wittgenstein, L. (91), Carnap, R. (10), pp. 37–40.

XVII

SENTENTIAL CALCULUS AND TOPOLOGY†

In this article I shall point out certain formal connexions between the sentential calculus and topology (as well as some other mathematical theories). I am concerned in the first place with a topological interpretation of two systems of the sentential calculus, namely the ordinary (two-valued) and the intuitionistic (Brouwer–Heyting) system. With every sentence $\mathfrak{A}$ of the sentential calculus we correlate, in one-one fashion, a sentence $\mathfrak{A}_1$ of topology in such a way that $\mathfrak{A}$ is provable in the two-valued calculus if and only if $\mathfrak{A}_1$ holds in every topological space. An analogous correlation is set up for the intuitionistic calculus. The present discussion seems to me to have a certain interest not only from the purely formal point of view; it also throws an interesting light on the content relations between the two systems of the sentential calculus and the intuitions underlying these systems.

In order to avoid possible misunderstandings I should like to emphasize that I have not attempted to adapt the methods of reasoning used in this article to the requirements of intuitionistic logic.[1] For valuable help in completing this work I am indebted to Professor A. Mostowski.

[1] Most results of this article were obtained in the year 1935. The connexion between the intuitionistic calculus and Boolean algebra (or the theory of deductive systems, see § 5) was discovered by me still earlier, namely in 1931. Some remarks to this effect can be found in article XII of the present book and in Tarski, A. (80). Only after completing this paper did I become acquainted with the work, then newly published, of Stone, M. H. (70). In spite of an entirely different view of Brouwerian logic there is certainly some connexion between particular results of the two works, as can easily be seen comparing Stone's Th. 7, p. 22, and my Th. 4.11. In their mathematical content these two theorems are closely related. But this does not at all apply to the two works as wholes. In particular, Th. 4.24, in which I see the kernel of this paper, tends in quite a different direction from Stone's considerations.

† Bibliographical Note. This article is the text of an address given by the author on 30 September 1937 to the Third Polish Mathematical Congress in Warsaw (see *Annales de la Société polonaise de mathématique*, vol. 16 (1937), p. 192). The article first appeared under the title 'Der Aussagenkalkül und die Topologie', in *Fundamenta Mathematicae*, vol. 31 (1938), pp. 103–34.