Chapter 2
Logic without Assumptions (1947)

Karl R. Popper

Abstract This article is a corrected reprint of K. R. Popper (1947b). Logic without Assumptions. In: *Proceedings of the Aristotelian Society* 47, pp. 251–292.

Editorial notes: In the original publication it says "*Meeting of the Aristotelian Society at 21, Bedford Square, W.C.1, on May 5th, 1947 at 8 p.m.*" below the title. We omitted this here. Popper sent lists of corrections to Bernays and Quine. These have not been worked into this reprint. They are reproduced in § 21.8.1 and § 30.5.1 of this volume. Further remarks by Popper can be found in footnote 8 of the unpublished typescript "A Note on the Classical Conditional" (this volume, Chapter 17, p. 286). In this footnote Popper also criticizes the review by McKinsey (1948, this volume, § 13.14). The manuscript "A General Theory of Inference" (this volume, Chapter 15) seems to be an early version of this article. The term "minimum calculus" refers to Johansson's (1937) "Minimalkalkül".

251 In this paper I shall try to explain, with a minimum of technicalities, some results of investigations into the field of deductive logic. The main problem to be discussed is *the problem of deduction itself* – more precisely, the problem of giving a satisfactory definition of "valid deductive inference".

Our method will be as follows: after having introduced, in section (1), a few auxiliary technical terms, we shall propose a definition, criticize it, and replace it by a better one, and repeat this procedure. After a few steps we shall reach, in this way, our first result – a definition which is a generalization of one due to A. Tarski[1]; and the rest of our investigation will be directed towards improving this result with the aim of avoiding some crucial objections originally urged by Tarski against his own definition.

The improved definition reached in this way suffices for establishing the validity of propositional logic and of the lower functional logic, without any further assumption.

[1] Cp. A. Tarski's lecture (delivered in 1935) "Ueber den Begriff der logischen Folgerung", ⟨Tarski, 1936b⟩.

D. Binder et al. (eds.), *The Logical Writings of Karl Popper*,
Trends in Logic 58, https://doi.org/10.1007/978-3-030-94926-6_2

(1)

Three of the auxiliary technical terms which will be introduced and explained in this section, “interpretation”, “statement-preserving interpretation”, and “form-preserving interpretation”, are, like some other very general concepts, somewhat tedious to deal with. Their generality, and even triviality, presents an obstacle to their intuitive understanding. There is, as it were, so little in them that those who try to grasp them
252 are left with the unpleasant | feeling that there is nothing to grasp. Those who are not satisfied with concepts in which there is so little, or who cannot believe that there is not more in them, may be assured that these three auxiliary concepts are the only technical terms of such an unsatisfying generality which they will be expected to handle. The other terms, “logical form” and “logical skeleton”, introduced at the end of this section, speak for themselves.

We begin by considering a number of languages, Latin, Dutch, German, Russian, etc., and by considering *translations* from one of these languages into the others – say, from Latin into English. There will be good translations and bad translations. Let us suppose that we know all the languages under consideration well enough to be unfailing judges of the various translations offered; that is to say, we can say whether a translation renders the full meaning of the statements to be translated or not; also, whether perhaps a certain kind of expression (such as the nouns or the verbs) are properly translated, while other kinds of expressions are rendered without regard to their correct meaning, etc.

Now by an *interpretation of one language in another* we shall understand a kind of translation, good or bad. Our intention is to use the term “interpretation” in such a wide sense that even extremely bad translations can still be classed as interpretations. But obviously, it will be necessary to introduce some kind of limitation to the badness of a translation if we are not prepared to accept the claim, for example, that the first twelve words of “Pride and Prejudice” constitute an interpretation of, say, the whole original text of the “Golden Ass”. For certain purposes it might indeed be of some advantage to use the term “interpretation” in such a wide sense that even the example just given would be covered. But for our purposes, it turns out that a slightly narrower use of the term is useful – especially in view of the fact that we shall not be interested in such things as exhortations or exclamations, but only in *statements*.

253 | Imagine that an extremely bad translator knows just enough about a certain language – Latin, say – as to recognize what is, and what is not, a complete statement in that language. (He may, for example, know nothing beyond the fact that statements are separated by full stops.) He may then proceed by “translating” every Latin statement into an English one. We who know both, Latin and English, will, of course, realize that his mock-translation is shockingly bad, i.e., that most of the English statements do not render in any way the meaning of those Latin statements to which they correspond, and which they should translate. However, the mock-translation in question has perhaps *one* advantage. Since to each statement of the Latin original there corresponds exactly one statement of the alleged translation, it will be possible, if necessary, to check the claims of the translator step by step; in other words, there will be at least a definite claim to be checked, and a fairly definite method of checking it.

Now it turns out that it is convenient to confine our investigation to translations, including mock translations, which satisfy the minimum requirement that every complete statement of the original text is, however badly or arbitrarily, translated by – or, more precisely, co-ordinated with – one complete and meaningful statement of the translation. And a translation or mock-translation which satisfies this minimum requirement will be called here, from now on, an "*interpretation*".

It is important to realize that the concept thus defined is an extremely wide one. We may, for example, choose to co-ordinate the first, second, third . . . statement of the "Golden Ass" with the first, second, third . . . statement of "Pride and Prejudice": the result will be an interpretation in the sense here defined. But we may also choose to co-ordinate the first, third, fifth . . . statement of the "Golden Ass" with one statement, say the first of "Pride and Prejudice", and the second, fourth, sixth . . . statement of the "Golden Ass" with the statement "In Italy it rains more often than in Egypt"; and
254 the result of this utterly | arbitrary co-ordination will still be an interpretation of the "Golden Ass" in the English language, in the sense here defined.

It is perhaps difficult to imagine that such a wide concept as this concept of interpretation is of any use. However, it will help us to define a few slightly narrower concepts which turn out to be exceedingly useful; more particularly, the concept of a *statement-preserving interpretation* and, furthermore, that of a *form-preserving interpretation*.

It is possible that a certain statement of some text may, word for word, recur in some other places of that text. Let us assume, as before, that the translator or interpreter of the Latin text knows so little about this language that he only recognizes the places where full statements end. But let us also assume that he has such a splendid memory that he recognizes each Latin statement whenever it re-occurs completely, and that he decides to translate it, in all places of its complete re-occurrence, by the same statement of the English language by which he translated it when it occurred first. If he follows this method consistently, we shall say that his interpretation preserves recurrences of complete statements, or more briefly, that it is a *statement-preserving interpretation*.

The concept of a statement-preserving interpretation is still a very wide one, and one might be tempted, at first, to think that it is still too wide to be useful. It is so wide that it allows us, for example, to translate every statement of the Latin language by *one* and the same English statement – say, the statement "In Italy it rains more often than in Egypt." For in this case, our condition (viz., that every statement of the Latin text must be translated, whenever it recurs, by the same English statement by which it was translated when it first occurred) is clearly satisfied. In other words, if we wish to give a statement-preserving interpretation of some Latin text in the English language, it will be sufficient if we have at least *one* English statement at our disposal into which we may "translate" all the statements of the Latin text; although we may, of course,
255 use more | than one English statement – indeed, any number of different English statements up to the extreme case in which their number is equal to the number of the different Latin statements contained in the text to be interpreted. It is clear that, in general, it will be impossible to re-translate a statement-preserving interpretation, i.e., to re-construct the original text from the translation together with the rules of

co-ordination (e.g., a kind of statement dictionary), except in the extreme case where with every different statement of the original a different statement of the interpretation is co-ordinated. (In this extreme case we speak of a *strictly* statement-preserving interpretation.)

The conception of a statement-preserving interpretation has one outstanding advantage over the conception of an interpretation as it was first introduced, viz., that it can be easily extended so as to cover not only an interpretation of a given *text* in some language or other, but also the interpretation of a whole *language* L_1 (say, the Latin language) in another language L_2 (say, the English language). For this purpose all that is necessary is to assume that we have co-ordinated, by some method or other, to every statement of L_1 one statement of L_2. This might be done, for example, by giving a proper translation of every statement of L_1 in L_2. In this extreme case (in which the interpretation is strict) we have to use as many different statements of L_2 as there are in L_1. Or again, to use another extreme case as an example, we might co-ordinate with every statement of L_1 one and the same statement of L_2. Or we might choose, in intermediate cases, a group of 2 or 20 or 200 statements of L_2, and co-ordinate the various statements of L_1 with these, by some method or other (say, on the basis of some alphabetic similarity of the first letters). Once we have co-ordinated with every statement of L_1 a statement of L_2, we have, of course, laid down a method of giving for every text in L_1 a statement-preserving interpretation in L_2.

It may be remarked that the languages L_1 and L_2 may well coincide. We can
256 construct, for example, various | statement-preserving interpretations of the English language (L_1) in the English language (L_2), by selecting one or more – possibly a very great number – of English statements (L_2) into which all other English statements (L_1) are to be translated, in accordance with some dictionary or code.

I hope that the conception of a statement-preserving interpretation will be reasonably clear by now. That it can be a useful conception will still appear doubtful, and has to be shown later. But it will be realized that this conception comprises a great deal, not only interpretations which do not take any notice of the meaning of the statements of the language L_1 – the language to be interpreted – but also interpretations which preserve the meaning of these statements, i.e., *proper translations*. Of course, we cannot assert that the conception of a statement-preserving interpretation comprises all proper translations; there may be an excellent translation from Latin into English which is not statement-preserving but which, at times, may split up a Latin statement into more than one English statement, or which uses one English statement to render several Latin ones. But most proper translations which avoid such cases will fall into the class of statement-preserving interpretations.

It should be noted that every statement-preserving interpretation automatically preserves recurrences of groups or sequences of statements. If, for example, the group of the statements $a, b, c, \ldots$ of language L_1 occurs in a text more than once, it will be translated into L_2, by every statement-preserving interpretation, in the same manner whenever the group recurs. This, of course, is not necessarily the case with groups of *expressions shorter than statements*. Such expressions – for example, single words, or groups of words – may, by a statement-preserving interpretation, be rendered differently every time they recur in a certain text; provided, of course, that they do not recur as parts of a recurring complete statement.

Within the wide class of statement-preserving interpretations we have, of course, many sub-classes, and among them not only proper translations, but also interpretations
257 | which preserve the recurrence of certain groups of words. Among these, one class of interpretations is especially important for our purposes; we shall call this class the "*form-preserving interpretations*".

The intuitive idea of a form-preserving interpretation of a language L_1 in a language L_2 is that of an interpretation which is not only statement-preserving but which also preserves what is usually called the "logical form" of the statements which are to be interpreted. A definition which is adequate to this idea can be easily given if we assume that we can distinguish between *two kinds of signs* of the languages which we are considering, viz., the *formative signs* and the *descriptive signs*.

Formative signs (they have sometimes been called "logical signs"[2]) are such signs as, for example, the full stop, or, in the English language, such words as "and", "or", "if . . . then . . . ", "neither . . . nor . . . ", "all", "some", "there exists at least one", etc. All signs, and groups of signs, which are not classed as formative will be called "*descriptive signs*". Examples are words such as "kitten", "mountain", "strategist", "aluminium", or groups of words such as "Greek orators", "elephant bones", "elderly disgruntled newspaper reader", etc.; also – but only if "of" is not considered as formative – "Orators of Greece", "bones of an ancient ancestor of the elephants", etc. Other descriptive signs are adjectives such as "grey" or "soft", and proper names (names of individual persons or of other physical things). We shall assume, for the time being, that the distinction between formative and descriptive signs can be applied with ease and without ambiguity to all the languages in which we are interested, and more especially, to the language L_1 which is to be interpreted in some other language L_2. (This assumption will be challenged later, in section (4).)

258 | We can now define a *form-preserving interpretation* as an interpretation which (a) preserves the meaning of all the formative signs, i.e., gives a proper translation of all the formative signs, and which (b) preserves recurrences of those groups of non-formative (descriptive) expressions which, in a proper translation, would fill the spaces between the translated formative signs.

According to this definition, a proper translation (if it is statement-preserving) will in general be a form-preserving interpretation; but so will be a translation of "All men are mortal" into "All kittens are green", provided we decide to preserve, in case the descriptive sign "man" recurs in L_1, its rendering by "kitten", and, in the same way, the rendering of "mortal" by "green".

According to our definition, it is only necessary to preserve *recurrences* of descriptive signs; two or more different descriptive signs of L_1 may be rendered by the same descriptive signs of L_2, subject, of course, to the proviso that every statement

[2] The term "logical sign", introduced by Carnap together with the term "descriptive sign", has been used also by others, for example by Tarski. I prefer "formative sign" in order not to suggest that logical signs are something like logical technical terms, such as the terms "deducible from", "compatible with", "negation", etc.

is rendered by *a statement* – a true one or a false one, but in any case a meaningful statement.

Among the form-preserving interpretations, there will be some which not only preserve recurrences of descriptive signs but also differences between descriptive signs. These might be called “strictly form-preserving interpretations”.

In general – that is, except if they are strict – form-preserving interpretations cannot be re-translated or decoded. Even a knowledge of all the translation rules enables us to translate only in *one* direction – say from L_1 to L_2 – but does not enable us to re-translate our interpretation.

A necessary and sufficient condition for the possibility of decoding or retranslating a form-preserving interpretation L_2 of L_1 back into L_1 is that every statement a_1 of L_1 which is interpreted by a_2 of L_2 is, in turn, a form-preserving interpretation of a_2. This is, at the same time, a necessary and sufficient condition for a form-preserving interpretation to be strict. The situation with statement-preserving interpretations
259 is analogous. This enables us to *define the* | *strictness* of form-preserving and statement-preserving interpretations in the following way:

A form-preserving (or a statement-preserving) interpretation of L_1 in L_2 is strict if, and only if, there exists such a form-preserving (or statement-preserving) interpretation of L_2 in L_1 that each statement of L_2 is interpreted, in its turn, by the same statement of L_1 which it interprets.

So much about the ideas of a statement-preserving and of a form-preserving interpretation.[3]

With the help of the latter idea, it is now very easy to define the idea of the *logical form of a statement and of a sequence of statements* (e.g., of an argument):

Two statements a_1 and a_2, not necessarily belonging to the same language, have the same logical form if, and only if, there exist two form-preserving interpretations such that a_1 interprets a_2, and *vice versa*. (Instead of demanding two interpretations and saying “*vice versa*”, we may also insert the word “strictly” before “form-preserving”.)

Similarly, if A_1 and A_2 are groups or series of statements. (We assume that statements belonging to the same series also belong to the same language.)

Two series of statements, A_1 and A_2, not necessarily belonging to the same

[3] Apart from the two kinds of interpretations which have been explained – statement-preserving and form-preserving – there are others of some interest. I may mention the truth-preserving interpretations, i.e. interpretations which do not necessarily preserve the meaning of the statements involved but which render every true statement by a true statement. This may be achieved, in a trivial way, by translating *all* statements of L_1, whether true or false, into true statements (say, into arithmetical truisms) of L_2. If we wish to exclude such a trivial method, we can demand that the interpretation should be not only truth preserving but truth-value preserving, i.e. that it should preserve not only the truth but also the falsity of the statements of L_1. In this case, we may still render all statements of L_1 by merely two statements of L_2, viz. by one which is true and one which is false. But this interpretation will be, nevertheless, far less trivial, since we have to consider the truth or falsity of every statement of L_1 before we correlate it with one of the two statements of L_2. (In order to avoid misunderstandings, I may mention that my concept of interpretation does not coincide with Carnap’s concept, as used in his *Introduction to Semantics*, ⟨Carnap, 1942⟩, esp. p. 203 and pp. 212f. This can be seen from the fact that one of Carnap’s demands, expressed in our terminology, is that L_2 contains at least as many different statements as L_1. Also, what Carnap calls a “true interpretation” does not coincide with our “truth preserving interpretation”.)

260 language, have the same logical | form if, and only if, there exist two form-preserving interpretations such that the first (the second, etc.) statement of A_1 interprets the first (second, etc.) statement of A_2, and *vice versa*. (Again, we can simplify the definition by referring to strictness.)

The term "logical form" usually occurs in the context "the same logical form" or "different logical form", and for such contexts our definition suffices. If a separate definition is desired, we can define the *logical form of the statement* a_1 as the class of all statements (of any number of languages) which have *the same logical form as* a_1.

This is a very abstract idea; but a more concrete idea is available in case our two statements a_1 and a_2 belong to the same language. We then can say that they have not only the same logical form but, more concretely, the same "logical skeleton"; for example, the statements of the English language "All men are mortal" and "All kittens are green" have both the logical skeleton:

"All . . . . are ——"

The logical skeleton of a statement or a group of statements is obtained simply by eliminating all descriptive signs, indicating, at the same time, recurrences of descriptive signs, by some method or other. It is clear that all statements or arguments of a certain language with the same logical skeleton have the same logical form, and *vice versa*; and it is clear that the concept of logical skeleton is not only more concrete but also simpler, since it is possible to define it directly, with the help of the distinction between formative and descriptive signs, without introducing the idea of an interpretation. On the other hand, our idea of a logical form is more general, and gives us the means of constructing logic as a theory of language – or of languages – *without tying us down to any particular language*.

And this, indeed, is one of the main points of our use of interpretations: we operate with the ideas of a statement-preserving and of a form-preserving interpretation partly because this method allows us, as will be seen, to combine the modern view of logic
261 as a theory of language with the | old intuitive idea that *the validity of an inference does not depend on the language in which it is formulated*; or more precisely with the idea that, if an inference is valid in one language, then it remains valid in every proper (and form-preserving) translation.[4]

[4] I believe that, in this way, we can preserve whatever is tenable in those objections to the modern view which emphasize that logic does not deal with mere "sentences" but rather with "propositions" or perhaps with "objective thoughts" (or "thought contents") or "thoughts" or "judgements", etc. (I do not use the term "sentence" but "statement" in order to indicate that, if I speak of a statement, I do not abstract from the meaning expressed by it – whatever this may be. A further analysis of the idea of a "meaningful statement" is very desirable; but it is not, as far as I can see – cp. the end of section (6) – necessary for the understanding of logic and the problem of its foundations; nor do I believe that either psychology, behaviourism, operationalism, verificationism, phenomenalism, or perceptionalism, etc., can have anything to offer towards a solution of the problem I have in mind.)

(2)

Armed with the technical terms introduced in section (1), we now turn to the analysis of the idea of a valid (deductive) inference.

By an inference, valid or invalid, we shall understand, in this paper, a number of statements, at least two, of some language, for example English, of which one is marked out for a conclusion and the others for premises (for example, by writing the conclusion last, below a horizontal line, etc.).

Our problem will be to analyse, in the most general way possible, the conditions under which such an inference (or argument) is called "valid"; or using a slightly different terminology, the conditions under which the relationship of deducibility actually holds between some premises and a conclusion.

We shall take most of our examples, for the sake of simplicity, from syllogistic logic; but this should not create the impression that we are more concerned with syllogistic logic than with any of the more developed systems (including the so-called "alternative" or "non-Aristotelian" systems, and those which use modalities).

262 | We begin by considering two simple examples of inferences; (a) a valid one, and (b) an invalid one.

(a)	(b)
All kittens are green	*All men are mortal*
Joe is a kitten	*Socrates is mortal*
Joe is green	*Socrates is a man*

Our task is to explain in a general way what we mean by saying that (a) is valid while (b) is not – without appeal to any recognized system of rules of inference of which (a) may be an observance or application. How could we try to explain to somebody who has not studied such a system of logical rules that (a) is valid and (b) invalid?

We might try to argue on the following lines:

It is conceivable that the conclusion of (b) is false, *even if the premises are both true*, while this is not conceivable in the example (a). For, assume that all kittens are really green, and that Joe is really a kitten, and assume nothing else; then, clearly, Joe must be green. But assume that all men are really mortal, and that Socrates is really mortal, but do not assume anything else (more especially, do *not* assume that Socrates is a man any more than, say, Joe); then, clearly, it is conceivable that the premises of (b) are true, and the conclusion false, since Socrates may be mortal and, for example, a kitten. In other words, for (b) a state of affairs (in which "Socrates" is the name of a kitten) is possible which renders both premises true and the conclusion false, while with (a) every state of affairs which renders the premises true would also render the conclusion true.

These considerations lead to our first preliminary and tentative definition (D1):

(D1) *An inference is valid if, and only if, every possible state of affairs which renders all the premises true also renders the conclusion true.*

We may call a state of affairs which renders all the premises of an inference

true and, at the same time, the conclusion false, a counter-example of that inference.
263 (A counter-example of (b) is provided, for example, by a state | of affairs in which "Socrates" is the name of a kitten.) Using this term, we may re-formulate (D1) in this way:

(D1′) *An inference is valid if, and only if, no counter-example of it exists.*

The main objection to this first tentative definition is the vagueness of the term "state of affairs", and more especially "possible state of affairs". This latter term may even be suspected of introducing a vicious circularity. For we are discussing logical validity, i.e., logic. But "possible" may very well mean "logically possible", and thus presuppose what we wish to define. (The same may be said of words like "conceivable", etc.)

In order to avoid these question-begging terms, we shall make use of the technical terms introduced in section (1). We can either use the comparatively simple term "logical skeleton" or the more complicated term "logical form" (or, in its stead, the term "form-preserving interpretation"). Let us first use, tentatively, the following two logical skeletons of (a) and of (b):

(a+)	(b+)
All . . . . are ——	*All . . . . are* ——
, , , , is a	*, , , , is* ——
————————	————————
, , , , is ——	*, , , , is a . . . .*

Applying to (a+) and (b+) similar considerations as we applied to (a) and (b), we arrive at our second tentative definition:

(D2) *An inference is valid if, and only if, every inference with the same logical skeleton whose premises are all true has a true conclusion.*

We may now re-define our term "counter-example" as follows:

A counter-example of an inference is an inference with the same logical skeleton whose premises are all true and whose conclusion is false.

If we use the term "counter-example" in this second sense (if necessary, we can distinguish it by the attribute "skeleton-preserving"), then we can give an alternative
264 | formulation (D2′) of (D2), such that the wording of (D2′) is identical with that of (D1′), although its meaning of course, has changed with the meaning of "counter-example".

Another possibility is that we use the term "logical form" instead of "logical skeleton", leaving everything else unchanged. Owing to the fact, however, that "logical form" is, in its turn, defined with the help of the term "form-preserving interpretation", it turns out that it is preferable to use the latter term instead. We thus arrive at (D3). This is the definition described in the beginning of this paper as our first result, and as a generalization of Tarski's definition:

(D3) *An inference is valid if, and only if, every form-preserving interpretation of it whose premises are all true has a true conclusion.*

(D3′) is taken to have the same wording as (D1′) and (D2′); but “counter-example” is now defined in this way:

A counter-example (or, more fully, a form-preserving counter-example) of an inference is a form-preserving interpretation whose premises are all true and whose conclusion is false.

An immediate result of our definition (D3) is the following theorem (T1) which has the same wording as (D2), except that the term “logical form” takes the place of “logical skeleton”:

(T1) An inference is valid if, and only if, every inference of the same logical form whose premises are all true has a true conclusion.

(Similarly, if we re-define “counter-example” again, so that it means an inference of the same logical form with true premises and a false conclusion, then we obtain from (D3) a theorem (T1′) with the same wording as (D1′) etc., but, of course, a different meaning.)

In view of (T1), we can say that the main difference between (D2) and (D3), i.e. our second and third definition, is that (D2) refers to “logical skeleton” and (D3) – indirectly – to “logical form”.

265 | The considerable merits of these three definitions, and especially those of (D3), will be discussed in the next section; and in section (4), we shall present those objections against (D3) which will lead us to the construction of an improved definition.

(3)

Proceeding to a discussion of our three definitions (D1), (D2), and (D3), I shall first explain the points in which they are all strong; next the strong points of (D2) and (D3); and ultimately, those in which (D3) is superior to the others.

All three definitions make use of the fundamental idea of *transmission of truth from the premises to the conclusion*; that is to say, of the idea that, if the premises are true, the conclusion *must* be true also. And all the definitions except the first succeed in explaining, or rather avoiding, this “must” (which is one of those dangerous question-begging terms) by pointing out that the transmission of truth *depends* solely upon the logical skeleton, or logical form, of the argument; and the term “depends” (which is also dangerous since it may mean “Logically depends”) is avoided by the simple method of referring to *all* inferences of the same logical skeleton or form.

Both the reference to the transmission of truth and to the logical skeleton or form seem to me intuitively highly satisfactory. It is the main point of the practical usefulness of deduction that, if we know that the premises are true and the inference valid, we can rely on the conclusion being true. In this way, inference allows us to obtain from reliable primary information reliable secondary information; and it allows us, by using as premises primary information from different sources, to derive secondary information unknown to any one of the sources.

Moreover, the transmission of truth from the premises to the conclusion which in itself is pragmatically as well as intuitively such an important point, means that, whenever the conclusion of a valid inference is false, at least one of the premises
266 must be false; for otherwise we would have a | counter-example and the inference would be invalid. In other words, the transmission of truth from the premises to the conclusion means also the *re-transmission of falsity from the conclusion to (at least one of) the premises*. This is, from the pragmatic point of view, just as important an aspect of a valid deduction as the obtaining of reliable secondary information. It enables us to reject prejudices by falsifying their consequences; and it allows us to test a hypothesis by the method[5] of trying to refute some of the conclusions which follow from it; for, if one of these is not true, the hypothesis cannot be true either.

It may be objected here that the term "true" which plays such a role in our definitions as well as in these considerations is vague; also, that we often do not know whether a certain statement is true or not. While this latter point must be admitted, the vagueness of the idea of truth need not be admitted; on the contrary, most people know that the word "true" can be easily eliminated from any such context as (1) "The statement 'the snow is white' is true" or (2) "The statement 'the snow is red' is true". They know that the whole of (1) asserts precisely the same as the statement (1′) "The snow is white" and that (2) asserts precisely the same as (2′) "The snow is red". But such a knowledge about the way in which a term can be eliminated from a simple context is, precisely, a knowledge of its meaning.

Why then, it may be asked, do we not eliminate the word "true" from our definitions? The answer is that it is easily eliminated from such simple but fundamental contexts where it refers to certain single statements, but not from contexts in which, more generally, we speak *about kinds of statements of some language* – say, about all statements of a certain logical skeleton or form, etc. It is especially in such a context that the term "true" – meaning precisely the same as in the other context – becomes useful.[6]

267 | I am therefore not prepared to admit that the employment of the term "true" in our definition is objectionable. Nevertheless, it will be seen that, in our final and improved definition (see also (D5)), there is a possibility of avoiding this term. At the present moment, however, I wish to emphasize that it must be the main point of every definition of validity which is intuitively satisfactory that it provides for the transmission of truth (whether or not the idea of truth enters the definition); and our definitions ⟨do⟩[a] provide for it.

Apart from referring to the transmission of truth, our second and third definitions refer to the *logical skeleton or form* of the argument in question. This too can be shown to agree with our intuitive idea of a valid inference. One immediate consequence

[5] Cp. my *Logik der Forschung* ⟨Popper, 1935⟩, and "The Poverty of Historicism, III", ⟨Popper, 1945b⟩, esp. pp. 78ff.

[6] The theory of truth underlying these remarks has been developed by Tarski in his analysis on the concept of truth (in Polish, ⟨Tarski, 1933b⟩; in German, "*Der Wahrheitsbegriff in den formalisierten Sprachen*", ⟨Tarski, 1935a⟩).

[a] Originally: "to".

which is part of this intuitive idea is that, if a certain inference is valid, all other inferences of the same logical skeleton or form must be valid too. It is this fact which allows us to lay down *rules of inference* which give a description of the logical form of an argument. A rule of inference asserts that from premises of a certain kind, a conclusion of a certain kind can be deduced.

We shall call such a rule of inference "valid" if, and only if, every inference drawn in observance of the rule is valid.

The fact that, whenever an inference is valid, there will also be a valid rule of inference (describing either the skeleton or the form of the argument; e.g., "Barbara") is also an immediate consequence of our second and of our third definition.

Ultimately I may mention very briefly, as a further advantage common to our second and third definitions, the fact that they can be both easily extended so as to be applicable not only to statements but also to *statement functions*.

A statement function (e.g. "He is green") can be obtained by replacing a descriptive sign ("Joe") of a statement by an appropriate variable, e.g. a pronoun ("He"). Such a function is neither true nor false. Nevertheless, it is useful for certain developments
268 to treat, say, |

(a–)

All kittens are green
He is a kitten

He is green

as a valid inference.

In order to extend our second and third definition so as to cover such cases, their wording need not be changed. All that is necessary is to include statement functions wherever we have so far considered statements, or groups of statements, and to specify, if necessary, the real statements – i.e., those which alone can be true or false – by some adjective; we shall call them, say, "proper statements". Accordingly, a *true* or a *false* example of the same logical form or skeleton as (a–) can only be an example consisting of proper statements, since functions can be neither true not false; and the same will hold for a "form-preserving interpretation whose premises are all true", etc.: such an interpretation will have to consist of proper statements.

We turn now to the weak points of our definition (D2), and the main reason for preferring the third definition to the second. The main difference between the two definitions is that (D2) refers to other arguments *of the same logical skeleton* as the argument in question and thereby confines its reference to other arguments *belonging to the same language*; (D3), on the other hand, refers to all form-preserving interpretations and therefore to *an unspecified number of different languages*, viz., to all those into which the formative signs can be properly translated. We shall show that this difference has two consequences:

(1) Our second definition (D2) is satisfactory only if the language L to which the argument under consideration belongs is sufficiently rich in descriptive signs.
(2) Our third definition is, besides, superior to the second because it yields the immediate consequence that the validity or otherwise of an inference or rule of

inference is independent of the language in which it is formulated, in the sense
269 that if it is valid in one language, then every one of | its proper form-preserving translations into another language will be valid also.[7]

While point (2) does not need an explanation, point (1) does, and the rest of this section will be devoted to construct a simple example of a language *L* in which, owing to the poverty of its dictionary, an invalid inference would appear as valid from the point of view of (D2), simply because no counter-example exists within *L*.

Let us assume a language *L* with the usual formative signs, "all", "some", "is", "are", "not" etc. We may also include the words "and", "or", as well as the prefix "non-" (e.g. say, "non-tall") and the word "who" among the formative signs, so that we may construct out of the simple descriptive signs "tall", "fair", and "Greeks" such complex signs as "fair or tall", "non-tall and fair", "Greeks who are tall", etc., which are made up of both formative and descriptive signs.

As to the descriptive dictionary of *L*, we may admit proper names of persons,
270 such as "Socrates", etc., and | property-names or class names such as "Greeks", "Englishmen", "Frenchmen" etc., as well as, say, "tall", "short", "wise", "foolish", "fair" and "dark". The main point is that our descriptive dictionary must not contain any *simple* descriptive name synonymous with, say, "short and foolish" or "short and not-foolish" etc.; nor must it contain any simple descriptive name of a property which is common to all persons. (We may consider the class of all persons as our universe of discourse; and we may say that we must not have in *L* a *purely descriptive* name of the universal class.) The upshot of all this is that we achieve in this way that our language does not possess two different simple names of two classes of which the one is included in the other.

Consider now this example of an inference in *L*:

---

[7] Cp. the end of section (1) and note 4. It is even possible to avoid the words "form-preserving" here, by a further generalization of (D3). For this purpose, we may replace the word "valid" in (D3) – or, for that matter, in our final definition (D6) – by "*formally valid*", and then add the following definitions of "*informally (or materially) valid*" and of "valid"; "An inference is informally valid if, and only if, it is formally invalid and there exists a proper translation which is formally valid." – "An inference is valid if it is formally or informally valid." – An example of an informally valid inference is:

(a−)

Peter's speed surpasses Richard's
Richard's speed surpasses Spencer's
———
Peter's speed surpasses Spencer's

For a formal counter-example, replace "speed" by "dog" and "surpasses" by "passes" or "surprises". On the other hand, we may say that the argument is properly translated if we replace "surpasses" by "is greater than" and that in this translation it is formally valid. What is often called the "rational reconstruction" of an argument is usually an attempt to show – e.g. by giving translation rules such as definitions, etc. – that it is informally valid. Of course the question whether a certain argument is informally valid or not can hardly ever be answered without further assumption of a more or less questionable character, such as intuitive translation rules – and it is hardly possible to give an answer in the negative except on intuitive grounds. (Informal inference in our sense, it may be noted, has played a certain role in the discussions of the idea of "entailment".) In the remainder of the present paper, we shall confine ourselves to formal validity without explicit mention of the word "formal".

(b′)

All Greeks are wise
Socrates is wise

Socrates is a Greek

This is, clearly, of the same form as (b) and (b+) and should therefore be recognized as *invalid*. However, *no counter-example can be formulated within L*. For, a counter-example in L must have true premises of the same skeleton as (b′). But although true universal statements such as "All Greeks are tall or non-tall", "All Greeks are non-Frenchmen", or "All Greeks who are tall are non-short" etc., exist in L, no true statement with the simpler skeleton "All . . . are —" happens to exist. (There exist true statements with the skeleton "All . . . are . . ." in L, but only such as would make the conclusion true.) Thus, within L, no counter-example exists to (b′), and (b′) would have to be described as valid from the point of view of our second definition, contrary to our logical intuition, and contrary also to our third definition which in this point and, it seems, in all others, agrees with our logical intuition.[8] (Of course, if we enrich the descriptive vocabulary of L a little, for example, by introducing the new simple term "personalities", defined, say, by "tall or short or both", then we can at once give a counter-example within L.)

(4)

Our third definition is, I believe, perfectly satisfactory on intuitive grounds. Nevertheless, there are serious objections. These were first urged by Tarski against his own definition.[9]

The weak point of our definition is this. Our definition is based, in the last analysis, upon the fundamental *distinction between formative and descriptive signs*. (Of all the technical terms introduced by us, only "statement-preserving interpretation" – and, of course, "interpretation" – do not presuppose this distinction.) Now this distinction is an intuitive one, and accordingly vague and uncertain. For example, the words "greater" and "smaller" (and "equal") are usually classed (with "identical" and "different") as

---

[8] The fact that a definition such as (D2) cannot be satisfactorily applied to languages with a poor
271 dictionary is mentioned in Tarski's lecture quoted | here in note 1. Tarski's own definition (which was the starting point of my investigation) avoids this drawback by combining, as it were, features of (D1) and (D2). In our terminology, it might be perhaps rendered: "An inference is valid if, and only if, every state of affairs which satisfies the logical skeleton of the premises satisfies that of the conclusion." And by a "state of affairs which satisfies a logical skeleton" (Tarski would call it a "model" of the skeleton), we would have to think of real things, and their properties and relations (not of the *names* of things, or of properties, or of relations, for these may be wanting). Tarski's definition is free of the disadvantage (1) discussed in the text, but since it is bound to the logical skeleton of some (one) language L, it does not fully share with (D3) the advantages (2) and the possibilities discussed in the foregoing footnote – at least, not without something like a "rational reconstruction" which might assimilate it to our conceptual apparatus.

[9] For Tarski's definition, see notes 8 and 1.

formative, while the intuitively similar words “taller” and “shorter” are classed as descriptive. This may be all right, but no satisfactory principle is forthcoming by which the classification can be justified.

Besides, it is by no means the case that the distinction can be drawn in every language. Thus most verbs, such as “runs” combine formative and descriptive functions
272 (of “is” and “running”). But this is only another way | of saying that the distinction is not applicable to “runs”. Since we have languages in which statements occur (“Achilles runs”) to which the distinction is inapplicable, it is clear that we may have whole languages without any term which strikes us intuitively as either purely formative or purely descriptive.

That there is something more needed than a reliance on an intuitive classification of the sign of a language may be illustrated by this example:

(c)

Bob sits in the cinema in the same row as Pat, and to the left of Pat
Pat sits in the cinema in the same row as Tim, and to the left of Tim

---

Bob sits in the cinema in the same row as Tim, and to the left of Tim.

This seems, intuitively, valid; even after constructing something like a counter-example by substituting “surprise” for “left” do we feel, intuitively, that (c) may be in order; we feel that the apparent counter-example only sounds like one, and that “to the left of” and “to the surprise of” have really not the same logical form, because there is something formative in “left”, as it were. But it is not encouraging to find that we have no principle at our disposal that may help us to clear up the matter directly.

One thing we have to do is to operate with a somewhat restricted list of formative signs, and to make sure that there is no ambiguity in our use of these formative signs. Aristotle moved first in this direction, by restricting his investigation to the language of “categorical propositions”. He introduced thereby an artificial limitation and rigidity which is quite foreign to naturally grown languages, but, it seems, necessary if we wish to construct a theory of inference. A language with such an unnatural rigidity of its rules may be called an “artificial language” or “calculus”. Modern logicians have followed Aristotle in confining their discussions to one or the other artificial language or calculus, and the method, I believe, is practically unavoidable.

273 | But it does not solve our fundamental problem. The logician who constructs an artificial language usually lays down a list of formative signs whose meanings he explains, and he treats all signs which are not in this list, and not definable on the basis of the list, as descriptive. This method is very successful and need not be criticized. But the choice of his signs (and their meaning) appears very largely as arbitrary, or as justified mainly on unanalysable intuitive grounds.

On this choice, however, everything depends in logic. It is clear that, by transferring a sign – e.g., the word “left” in our last example – so far considered as descriptive, into the list of formative signs, or *vice versa* (without altering the intuitive meaning attached to it), we may alter the logical form of some, or even of all our arguments, with the result that arguments so far valid become invalid, and *vice versa*. And this can be done as long as we have no objective criterion of the formative or descriptive character of a sign.

Our objection may be summed up as follows:

The proposed definition of valid inference may permit us to reduce the intuitive problem whether a certain inference is valid or not to the problem whether a certain sign is formative or not. But this problem remains intuitive.

I do not intend to create the impression, by this formulation, that I consider our definition of validity as futile. On the contrary, it is very valuable. Nevertheless, the situation is serious. If there is no clear-cut distinction between valid and invalid inferences, then, it can be shown, there is also no clear-cut distinction between logical and factual (or empirical) statements, or, to use a more traditional terminology, between analytic or synthetic propositions. This is a question of major philosophical importance. One might even say that it involves the whole question of empiricism, and of scientific method.[10]

(5)

274 | Is our problem insoluble? I do not think so. One way to a possible solution follows here in outline. It is explained, more fully, in the subsequent sections.

There are inferences – admittedly, so trivial ones that they have hardly ever attracted much notice – which can be shown, on our definition of validity, to be valid *whatever the logical form of the statements involved*. Their validity is thus independent of that distinction between formative and descriptive signs which we have seen to be of such a problematic character. We shall say of these inferences that they are *absolutely valid*. Absolute validity can be defined in terms of statement-preserving interpretations, and therefore without referring to the distinction between formative and descriptive signs.

Next we observe that once we have a certain system of absolutely valid rules of inference at our disposal, it is possible to define the logical force or import of the various formative signs in terms of deducibility (i.e., in terms of the predicate: “from the statements a, b, c . . ., the statement *d can be deduced*”). We shall call a definition of a formative sign in terms of this relation an *inferential definition*.

With the help of the idea of an inferential definition, we can characterise the formative signs as signs which can be defined by an inferential definition. In this way we can meet the objections raised against the distinction between formative and descriptive signs, and re-establish (D3), as it were. Ultimately, we shall propose a definition (D6) which makes use only of the ideas of absolute validity and of an inferential definition, and which no longer refers to truth.

---

[10] These important problems, raised by our (or rather by Tarski’s) definition of valid inference, have been only little discussed, except by Tarski himself – who seems to consider the problem insoluble – and by Carnap, who takes a more hopeful view of the matter; cp. R. Carnap, *Introduction to Semantics* ⟨Carnap, 1942⟩, esp. p. vii, where “the distinction between logical and descriptive signs” is mentioned, and pp. 56ff.

(6)

There are inferences which are valid according to all our definitions, in spite of the fact that the logical form of the statements involved is irrelevant. They are very trivial inferences indeed – so trivial that some logicians refused to accept them. Accordingly, our first task will be to argue that they are valid.

275 | Consider the example:

(d)

$$\frac{\textit{Joe is a kitten}}{\textit{Joe is a kitten}}$$

From the point of view of all our definitions (and from the point of view of most modern systems of logic) this is definitely a valid inference – even though it is something like an extreme case, or a "zero-case", as it were. For it is clear that no counter-example of (d) can exist. Some logicians have objected to calling these cases "valid inferences" on grounds such as these: inference is a movement of the mind; but our mind does not move in such cases as (d); thus this cannot be an inference. But since even these logicians hardly suggest that (d) is actually invalid, not much harm can be done by calling it an inference. It seems that their objections are due to the fact that in such zero-cases our intuition is not a reliable guide. (Is zero a number, or is it not, rather, nothing? Is one a number? No, one thing is just one thing, and not a number of things! Even two is not a number, but a couple. A number of things are at least three, it seems.) But there are several reasons why we should call (d) a valid inference – quite apart from our definition.

One is that it would be awkward to forbid that the middle and major term of a syllogism may ever coincide, as in the example:

(e)

$$\frac{\begin{array}{l}\textit{All kittens are kittens}\\ \textit{Joe is a kitten}\end{array}}{\textit{Joe is a kitten}}$$

But if (e) is an inference (and, of course, valid), then we may clearly omit the first premise as redundant; and in this way we obtain (d).

But there are stronger reasons for accepting (d). There are very strong reasons for not excluding the possibility of cases of *mutual inference* (i.e., from a premise to a
276 conclusion | and also from this conclusion to its premise). A mutual inference can, of course, have only *one* premise (since an inference can have only one conclusion), but we can hardly forbid inferences from one premise, since we can combine several premises into one by merely joining them with the help of the word "and". If we forbid mutual inference, then we can never be sure whether an inference, however convincing, is valid; for if anybody should later discover an equally convincing argument running from the conclusion to the premise (or to the conjunction of the premises) then we would have to declare both inferences invalid.

But if mutual inferences are possible, then (d) must be possible too. For there is a principle of inference which has never been challenged – the transitivity principle which says that the conclusion c of a conclusion b of the premise a is itself a conclusion of the premise a; and this principle if applied to a mutual inference, yields:

(6.1) "From the statement a we may deduce a",

that is, a rule of inference of which (d) is an observance. In other words, whoever is not prepared to admit (d) and (6.1) must either forbid mutual inference, which has the most awkward consequences, or he must give up the following transitivity principle:

(6.2) If from the statement a we can deduce b, and if from b we can deduce c, then we can also deduce c from a.

We shall therefore rely on our definition of validity and accept (d) and (6.1) as valid.

But the validity of (d) and (6.1) does clearly not depend on the distinction between formative and descriptive signs. *The fact that no form-preserving counter-example exists is here a direct consequence of the fact that no statement-preserving counter-example exists* – together with the fact that, of course, all form-preserving interpretations must be statement-preserving.[11]

Thus there are valid inferences whose validity does not depend upon the distinction
277 between formative and | descriptive signs. We shall call these inferences "absolutely valid"; and we define, tentatively:

(D4) *An inference is absolutely valid if, and only if, every statement-preserving interpretation whose premises are true has a true conclusion.*

Or alternatively:

(D4′) *An inference is absolutely valid if no statement-preserving counter-example of it exists.*

The field of absolutely valid inferences, and of absolutely valid rules of inference[12], covered by this definition, is somewhat trivial but not at all unimportant. It can be shown[13] that all absolutely valid rules of inference which we shall need for the logic of statements – and there are many and even somewhat complicated rules among them – can be reduced to two – the one a generalization of (6.1) and the other ⟨a⟩ generalization of (6.2). By "reduced", I mean here: every inference which is an observance of some of the rules in question can be shown to be an observance of these two rules – obtained, possibly, by applying them many times in succession, which is permitted by the generalized transitivity rule (6.2g) itself.

In order to formulate these two rules more easily, we introduce the following abbreviation: we shall write

---

[11] The definition of "statement-preserving counter-example" (cp. "form-preserving counter-example") is obvious, in view of section (2).

[12] The definition is again obvious: a rule of inference is absolutely valid if, and only if, every inference drawn in observance of it is absolutely valid.

[13] Cp. my "New Foundations for Logic," forthcoming in *Mind*, 1947 ⟨as Popper, 1947c⟩.

$$\text{“}a, b, c \ldots / d\text{”}$$

instead of “from the premises $a, b, c \ldots$, the conclusion d can be deduced”; and we shall write

$$\text{“}a_1, a_2, \ldots, a_n / b\text{”}$$

instead of “from the premises $a_1, a_2, \ldots, a_n$, the conclusion b can be deduced”. (Note that our variables “a”, “b”, “a_1”, etc., are variable *names* of statements, i.e., that only *names* of statements may be substituted for them, but not the statements themselves; this distinguishes these variables from the so-called propositional variables of the calculus of propositions.)

278 | With the help of this notation, we can write our two generalized rules:

(6.1g) $$a_1, a_2, \ldots, a_n / a_i,$$

provided a_i is identical with one of the n statements $a_1, a_2, \ldots, a_n$ (in symbols: provided that $1 \leq i \leq n$).

(6.2g) If, from the n premises $a_1, a_2, \ldots, a_n$ taken together, each of the m statements $b_1, b_2, \ldots, b_m$ can be separately derived, then,

$$\text{if } b_1, b_2, \ldots, b_m / c \text{ then } a_1, a_2, \ldots, a_n / c.$$

These two rules, or rather the system consisting of these two rules, can be easily shown to be absolutely valid and therefore, *a fortiori*, valid. They are all the absolutely valid rules needed for the proof that our final definitions are adequate for the logic of proper statements. In order to extend our results to statement functions, we need another set of trivial rules which can easily be shown to be absolutely valid, and which allow us to operate with the idea of *the result of substituting, in some statement function one variable for another*. In order to indicate the character of these rules, I shall use the symbol “$a\binom{x}{y}$” as an abbreviation of “the result for substituting the variable y for the variable x in the statement function a”.[14] The trivial rules mentioned are all of the character of

(6.3) $$\text{If } x = y, \text{ then } a / a\tbinom{x}{y} \text{ and } a\tbinom{x}{y} / a.$$

The triviality of this rule will be realized by considering that, if $x = y$, the result of substituting y for x in a leaves a completely unchanged; or in other words, that

(6.4) $$a\tbinom{x}{x} / a.$$

Accordingly, $a\binom{x}{x}$ and a will be mutually deducible, whatever the logical form of
279 a may be; and this is what | (6.3) asserts. If we abbreviate “mutually deducible” by “//”, defining:

(6.5) $$a // b \text{ if, and only if, } a / b \text{ and } b / a,$$

then we can write (6.3) as follows

[14] The symbols “x” and “y” are (variable) *names* of variables.

(6.6) $$a\binom{x}{x} /\!/ a.$$

The triviality of (6.6) is, in view of (6.4), quite obvious.

The rules of inference (6.3) and (6.6) are clearly not only valid, but absolutely valid. For there can be no interpretation which preserves statements, including statement functions, and which renders a by a true statement, but not $a\binom{x}{x}$ which, in view of (6.4) is only another name for a.[15]

Our concept of absolute validity, we have seen, is independent of the problematic distinction between formative and descriptive signs. In addition, it is also possible to define it in such a manner that the much less problematic *concept of truth is also avoided*. For it turns out that a great many properties of statements can replace "true" in our definition – among them such trivial properties as "containing less than five words (or signs)", for which we shall briefly say "short". The properties which may replace truth can be called *freely interpretable properties*, or briefly, *free properties*. Intuitively speaking,[16] a property is called "free" if we are free to choose (by an
280 appropriate choice | of the translation rules) for any individual statement a_1 of L_1, whether or not it shall be interpreted by a statement which possesses the property in question. The strict definition of "free property" (given in note 16) makes use of the term "statement-preserving interpretation", but does not use any specific name of a free property, such as "short" or "true".

Now it can be shown[17] that if every statement-preserving interpretation transmits *one* free property in a certain argument, then it transmits *all* free properties in this argument. This allows us to replace (D4) by:

(D5) *An inference is absolutely valid if, and only if, there exists at least one free property such that every statement-preserving interpretation whose premises all possess this free property has a conclusion with the same free property.*

This definition avoids the term "truth"; but in the presence of the definition of "free property", and in view of the fact that "true" designates a free property, we can obtain (D4) from (D5).[18]

---

[15] Note that, in these considerations, we never assume anything about the logical form of a (not even that it is a function rather than a proper statement). For a list of similarly trivial rules which are needed for developing the logic of functions and which can all be shown to be absolutely valid, see rules (6.1) to (6.6) of my *New Foundations for Logic* ⟨Popper, 1947c⟩. Cp. note 13.

[16] The strict definition is: "A property of statements is a free property if, and only if, to every arbitrary division of the statements of L_1 into two exclusive classes A and B, there exists a statement-preserving interpretation which interprets all statements of A by statements which possess the property in question and all statements of B by statements which do not possess it." – As an illustration, I may also mention two examples of properties of a statement a_m which are *not* free: (1) "a_m is a statement of L_m." – (2) "a_m does not occur in text x after b_n."

[17] With the help of the fact that the relation designated by "statement-preserving interpretation" is transitive.

[18] The transition from (D4) to (D5), including its effect on (D6), corresponds roughly to the transition from Truth Tables to abstract Matrices, in the sense of Łukasiewicz and Tarski. In Carnap's terminology, it may be described as one from *General Semantics* to *General Syntax*, since "true" is a Semantical concept, while "short" is Syntactical; "free property" is also Syntactical, for it is defined merely with the help of "statement-preserving interpretation" which, in view of section

Absolute validity cannot replace validity; its main advantage is that its definition is free from the objections we raised (following Tarski's suggestions) against our definition of validity. But before going any further, we may ask whether there are not analogous objections left. Admittedly, absolute validity does no longer depend on the
281 distinction between formative and descriptive signs. But | does it not depend upon the distinction between statements and non-statements? And is this distinction not as intuitive as the other?

The answer to this last question must be "yes". But the situation is, nevertheless, not analogous to the one we have laboured so hard to avoid. Intuitive doubts about the distinction of formative and descriptive signs may arise even in connection with an artificial language. As opposed to this, any doubt whether or not a certain expression of an artificial language is a statement can hardly arise on intuitive grounds. (If at all, then it arises as a very definite problem of reconstruction, for example, in order to avoid certain paradoxes.) But the main difference between the two problems is this: the distinction between formative and descriptive languages affects our central problem – validity. A given inference may be valid or invalid, according to the way we draw the line. This is never the case with the other distinction. It cannot affect the decision as to the validity or invalidity of some given inference; it can only affect the question whether a certain sequence of expressions is an inference (valid or invalid) or no inference at all.

The problem of validity is always: given that this is an inference – is it valid? It is clear that the two questions – the one of characterizing the formative signs, the other of characterizing the statements – have a very different status relative to this problem. And while the solutions of the problems of validity and of formative signs are closely interdependent, I do not believe that the problem of validity and that of statements are likely to affect each other in any way.[19]

(7)

Absolute validity cannot, of course, replace validity. Its definition may be free from flaws, but this advantage is bought at a high price. Absolute validity, one is tempted to say, is a property of inferences which are absolutely trivial.

282 | We have therefore to return to the problem of validity, and with it, to that of formative signs. But we return to these problems with means of solving them. By our definition of absolute validity, we have acquired the right to operate freely with all

(1), appears to by a Syntactical term (as opposed to "form-preserving interpretation" into whose definition the Semantic idea of a proper translation enters). But, it may be asked, what about our assertion that truth is a free property? And what about the derivation of (D4) from (D5)? Altogether, I am very doubtful whether the distinction between Semantics and Syntax really fits our approach which is very largely meta-metalinguistic. (Our term "valid" belongs to the meta-metalanguage.) Yet we may perhaps describe the transition in question as a change of emphasis from the Semantics of Semantics to the Semantics of Syntax.

[19] Cp. note 4.

kinds of rules of inference, *as long as these do not refer to the logical form of the statements involved.*

With these means at our disposal, we can not yet define what we mean when we say:

"a is the negation of b"

or perhaps

"a is the disjunction of b and c".

But we do possess the means of defining what we mean when we say:

"a has the same (logical) force as a negation of b"

or perhaps

"a has the same (logical) force as a disjunction of b and c".

We can express the phrase "a has the same (logical) force as b" (or "a is equivalent to b") with the help of the symbol "//", by "$a//b$", which has been defined above (6.5) as mutual deducibility. There is an alternative definition of equivalence, viz.:

(7.1) $a//b$ if, and only if, for every c: a/c if, and only if, b/c.

Now just as this defines the phrase: "a has the same force as b, *whatever the logical form of a and b may be*", so we can also define the phrase: "a has the same force as the negation of b, *whatever the logical form of a and b may be*" by:[20]

(7.2) $a//$the negation of b if, and only if, for every c: $a, b/c$ and, if $a, c/b$ then c/b.

283 | Similarly we can define:

(7.3) $a//$the disjunction of b_1 and b_2 if, and only if,
for every c_1 and c_2: $a, c_1/c_2$ if, and only if, $b_1, c_1/c_2$ and $b_2, c_1/c_2$.

I shall not discuss the adequacy of these and the other definitions mentioned here, since this follows from material I have given elsewhere[21]. I may mention, however,

[20] (7.2) defines classical negation; if we replace the last phrase of (7.2) by "if $b, c/a$ then c/a" we obtain the negation belonging to Heyting's intuitionistic calculus. Both definitions can be proffered at the same time, but if in one language, a classical as well as an intuitionistic negation exists of every statement, then the latter becomes equivalent to the former, or in other words, classical negation then absorbs or assimilates its weaker kin. This remark follows from the observations made in my paper quoted in note 13 upon rule 4.2e, and modifies one made in that paper; but the negation belonging to Johansson's "Minimum Calculus", is not absorbed by classical negation; and the same holds for the following definition of a negation which seems to agree excellently with intuitionistic intentions, although it is not equivalent with that of Heyting's calculus: "$a//$the impossibility of b if, and only if, for every c: a/c or b/c, and if b/a, then c/a." (This definition is no longer "purely derivational"; this fact marks the transition to modal logic.)

[21] Cp. note 13. For propositional logic, one definition – apart from (7.1) – is sufficient: "$a//$the alternative denial of b_1 and b_2 if, and only if, for every c_1: c_1/b_1 and c_1/b_2 and $b_1, b_2/c_1$, if and only if, for every c_2: $a, c_1/c_2$ and if $a, c_2/c_1$ then c_2/c_1." Combined with (7.4) this suffices for functional logic; and combined with the definition at the end of note 20, for modal logic (whether propositional or functional).

that the two definitions (7.2) and (7.3), form a sufficient basis for propositional logic – say, for Russell's system, and that we can give similar definitions for all the known compounds of propositional logic.

In order to provide for functional logic, we define universal and existential quantification by a similar method. One of them is sufficient, in the presence of (7.2); we choose existential quantification:[22]

(7.4) Provided x is distinct from y,
$a\binom{y}{x}$ //the result of the existential quantification with respect to x of $b\binom{y}{x}$ if, and only if, for every c: $a\binom{y}{x}/c\binom{y}{x}$ if, and only if, $b\binom{x}{y}/c\binom{y}{x}$.

The definitions given here form a sufficient basis of propositional and (the lower) functional logic, in a sense which will be explained; but they are given only as
284 illustrations of our method. We can, by this method, give many | more definitions; not only of the logical force of certain formative signs, but also definitions of such matters as the exclusiveness (or contradictoriness) of a number of statements, and of their exhaustiveness (or "logical disjunctness", to use Carnap's expression). For simplicity's sake, we confine ourselves to two statements, a_1 and a_2:

(7.5) a_1 and a_2 are exclusive if, and only if,
for every b_1 and b_2: if b_1/a_1 then, if b_1/a_2 then b_1/b_2.

(7.6) a_1 and a_2 are exhaustive if, and only if,
for every b_1 and b_2: if a_1/b_2 then, if a_2/b_2 then b_1/b_2.

These concepts[23] are interesting since they provide us with a definition of complementarity:

(7.7) a//the complement of b if, and only if, a and b are exhaustive as well as exclusive.

Now on the basis of (7.2) and (7.7) it can be shown that, if a_1//the negation of b, and a_2//the complement of b, then $a_1//a_2$.

Had we known this before, we need not have defined "complement"; "negation" would have done just as well. Until, however, an equivalence between two definitions such as (7.2) and (7.7) is established, we must always be careful to use different names. *But this, indeed, is the main precaution necessary.* We need not make sure, in any other way, that our system of definitions is consistent. For example, we may define (introducing an arbitrarily chosen name "opponent"):

(7.8) a//the opponent of b if, and only if, for every c: b/a and a/c.

---

[22] It is understood that two rules are at our disposal (because of their absolute validity) which are described, in my "New Foundations," as "rules of substitution", and which read, in the notation of that paper:

"$(Eyb)\binom{y}{x}//Ex(b\binom{y}{x})$" and "If y/z then: $(Ezb)\binom{y}{x}//Ez(b\binom{y}{x})$".

Alternatively, a supplementary definition yielding these two rules may be added to (7.4).

[23] Their extension to n statements may be indicated by: "$a_1, a_2, \dots, a_n$ are exhaustive if, and only if, for every b_1 and b_2, if a_1/b_2 then, if a_2/b_2 then, ... if a_n/b_2, then b_1/b_2."

This definition has the consequence that every language which has a sign for "opponent of b" – analogous to the sign for "negation of b" – will be inconsistent (i.e. every one of its statements will be paradoxical). But this need not lead us to abandon (7.8); it only means that no consistent language will have a sign for "opponent of b".

285 | Each of our definitions gives rise to a number of rules of inference. For example, (7.1) gives rise to the rule:

(7.91) If $a//b$, then b/a.

Similarly (7.2) gives rise to

(7.92) If $a//$the negation of b, then if $a, c/b$, then c/b.

And (7.3) gives rise to

(7.93) If $a//$the disjunction of b_1 and b_2, then b_1/a and b_2/a.

To every of these rules we may add: "*whatever the logical form of the statements involved may be*".

All these rules are absolutely valid. (That is to say, if they are added to 6.1g and 6.2g, the resulting system is absolutely valid.) Why? Because no statement-preserving counter-example can be found. Consider (7.91): since we have laid down, in effect, that we can replace "$a//b$" by "a/b and b/a", we can in (7.91) replace "If $a//b$ then" by "If a/b and b/a then"; and it is clear that no statement-preserving counter-example can be found to any inference drawn in observance of the resulting rule. The same holds for all other rules of this kind.

Since we may consider the definitions themselves as rules, we may say, therefore, that they are all absolutely valid. Whether the concepts they define are adequately defined, is a different matter; this will depend upon our intentions. If, for example, we intend to define conjunction (or rather, its logical force) and use the right hand side of our definition of disjunction (7.3) for this purpose, then we shall have an inadequate definition. But the definition will, nevertheless, define something – namely just what we usually would call "disjunction" rather than "conjunction".

So much about these definitions. We now turn to the problem of formative signs, and of valid inference. We define:

(7.2$^{\mathrm{D}}$) The language L_1 contains a (preceding or succeeding) *sign of negation* if, and only if, it contains a sign which, if joined to any statement b of L_1 (placed before b, or placed after b) forms a new statement a of L_1, such that this resulting statement $a//$the negation of b, in the sense of (7.2).

286 | Note that, according to this definition, L_1 may contain more than one sign of negation.

(7.3$^{\mathrm{D}}$) The language L_1 contains a (preceding, or intervening, or succeeding) sign of disjunction if, and only if, it contains a sign which, if joined to any pair of statements b_1 and b_2 of L_1 (placed before b_1 followed by b_2, or placed between b_1 and b_2, or placed after b_2 preceded by b_1) forms a new statement

a of L_1, such that this resulting statement $a//$the disjunction of b_1 and b_2, in the sense of (7.3).

In the same way, we may define the other formative signs; we can define them as *signs which, in certain contexts, form statements of a defined logical force.*

Definitions such as (7.2^{D}) and (7.3^{D}) may be called inferential definitions. They are characterized by the fact that they define a formative sign by its logical force which is defined, in turn, by a definition in terms of inference (i.e. of "/").

It is now easy to define the term "formative sign".

A sign s of a language L_1 is a formative sign if, and only if, s can be defined by an inferential definition.

According to this definition of the term "formative sign", the question whether a certain sign s used in a certain way is formative remains an open question until someone either produces an inferential definition of s or shows, by some method or other, that such a definition does not exist (as can indeed be done in certain cases, for example, the negation sign of Johansson's minimum calculus). But the remark that this question may remain open does not constitute an objection against our definition; and especially the fact that inferential definitions of the intuitively recognized formative signs of propositional, functional, and even modal logic can be given (as has been shown) establishes, it seems, the adequacy of our definition of the term "formative sign".

I believe that this definition solves, fundamentally, our crucial problem. It provides a rational basis for the distinction between formative and descriptive signs; and with this, the objections of section (4) can now be met. Our definition achieves,
287 however, more than this result. It | also shows the rationality of the idea of a form-preserving interpretation, since it makes it clear that a form-preserving interpretation does not depend on the intuitive knowledge possessed by a translator – his intuitive knowledge of the meaning of the formative signs of a language. For we see now that precise translation rules for formative signs can be obtained from their definitions, because they are defined with the help of rules of inference which are absolutely valid, and because absolute validity, in its turn, depends merely upon the idea of statement-preserving interpretations, which does not presuppose a knowledge of languages.

We can therefore now adopt the *wording of* (D3) *as our final definition of valid inference.* Of course, the words of (D3) have, in view of our new results, acquired a somewhat different meaning, namely, a more precise one.

But it seems that we can, if we wish, go further; it seems that we can avoid, in view of (D5), even the reference to truth which occurs in the wording of (D3), by defining:

(D6) *An inference is valid if, and only if, it is either absolutely valid, or it can be shown, on the basis of the inferential definition of the formative signs, to have been drawn in observance of absolutely valid rules (including those which define the logical force of the statements involved).*

The wording of this definition should be capable of some improvement, but the idea, I think, is clear; and considering the actual techniques of establishing valid rules of inference (indicated in my paper quoted in note 13) it seems to be adequate. A new problem arises, however, in connection with this definition, viz. to prove, as can be done in the case of (D5) and (D4), that (D6) actually yields (D3), that is, guarantees the transmission of truth. But this problem must be left for another occasion.

(8)

An objection which may have troubled the reader for some time may now be briefly discussed. Is not our procedure rather circular? We assert that the validity of the
288 logical rules of inference *follows* from certain definitions. | This may be so; but does not this *derivation* assume the validity of some rules of inference – and probably just of those which are to be derived?

This objection cannot be discussed here in full, for that would mean a disquisition on the distinction between *languages* and *metalanguages*, of the general status of any metalinguistic investigation, and, more particularly, of any logical investigation. But a very brief answer may be attempted.

If we investigate such problems as the way in which two statements, *a* and *b*, are related – whether the one follows from the other, or contradicts the other, or is complementary to it, etc. – then we are investigating certain objects which are parts of some language, i.e. *linguistic objects*. If we wish to study linguistic objects, we must discuss them, and make statements about them, as with all other objects. If we use our language in order to discuss linguistic objects, then we say that we use our language as a *metalanguage*. The language which we study is called the *object language* (or language under investigation).

Now it is important to realize that we cannot at the same time *use* a word, or a statement, or an argument, and *study* it; and also that the study of words or statements presupposes the unhampered, although careful, use of some language in precisely the same way as the study of trees or of mental processes or of music.

It is for this reason that we should never attempt to reflect on our metalanguage *while* we are analysing an object language; and it is for this reason that we should not attempt to analyse the arguments we are using while engaged in analysing the rules of argumentation.

Indeed, if this were not so, then *all* logical investigation would be impossible. For if we wish to study something, we can, clearly, not begin by giving up the use of all argumentation; and logic is no exception to this rule.

In deriving rules of inference from definitions, we cannot, of course, avoid *using* inferential arguments, just as in studying plants or animals. *After* we have done our
289 job, we may | then analyse the inferential arguments used. This is a legitimate and interesting new problem, but it is quite different from the original problem – that of deriving certain rules from certain definitions.

If we actually analyse (using a meta-metalanguage) the procedure used before in the metalanguage, then we find that there is no circularity involved. Surely, we must *know how* to use our metalanguage when studying the rules of inference of certain object languages (as we did), and surely, we must know how to draw inferences in the metalanguage while engaged in these studies. But we do not formulate the rules we are using, and *these rules, therefore, never appear as premises in our derivations*. But only if they did could we speak of circularity.

This situation would not be affected if, for example, we might have to *use* negation (say, the word "not") while *discussing* negation. But we actually find, when we investigate our metalanguage, that it is possible to carry out our investigations in the metalanguage without using negation or rules of inference pertaining to negation; and this example, surely, reaffirms the view that our investigation is not circular.

(9)

We have, so far, only indicated how the well-known *rules of inference* of logic can be derived, not from assumptions such as primitive rules, but from definitions. In conclusion, I shall briefly mention how the *logical theorems* of a language (for example the well-known tautologies of the calculus of propositions, including, of course, the primitive propositions or axioms) can be obtained from our definitions without further assumption (such as the assumption of primitive propositions or axioms).

The procedure is based on a new definition – the definition of logical demonstrability, or more precisely, of "*a* is (logically) demonstrable":

(9.1) *a* is demonstrable if, and only if, *a* can be validly inferred from any statement *b* whatsoever.

290 | This definition may seem at first intuitively rather puzzling. But it only gives a precise form, in terms of valid inference, to the somewhat vague intuitive idea that a statement *a* is logically demonstrable if, and only if, it is true merely because of its logical form. That *a* is true merely because of its logical form can be expressed, more clearly, by saying that all statements of the same logical form must be true. But this means that no form-preserving interpretation of *a* can be false, and therefore, that no inference with the conclusion *a* can be invalid, whatever the premises may be; for if no form-preserving interpretation of *a* can be false, then it is clear that *no form-preserving counter-example can be found*.

There are other arguments to show that our definition (9.1) is in keeping with the intuitive idea of demonstration or proof, and that it brings out a distinction between derivation (i.e., non-demonstrative inference) and demonstration which has been often neglected by logicians.[24]

A derivation never establishes the truth of the conclusion. It only establishes the

[24] The distinction between derivation and proof has been emphasized by Carnap (in his *Logical Syntax*, ⟨Carnap, 1934a⟩ and ⟨Carnap, 1937⟩, as well as in his *Introduction to Semantics* ⟨Carnap, 1942⟩). Our treatment agrees fundamentally with his results but goes beyond them, in so far as our treatment of the theory of derivation is quite independent of the fact whether or not there are demonstrable statements in the language under consideration, while Carnap's treatment of derivation makes essential use of certain rules of derivation which are only re-formulations of axioms; cp. his *Introduction*,⟨Carnap, 1942⟩ p. 167, and his *Formalization of Logic*, ⟨Carnap, 1943,⟩ p. 9.

fact that the question whether or not the conclusion is true can be shifted back to the question whether or not the premises are all true. Thus, if we find that the conclusion is false, this does not show that the argument, the derivation is invalid – it only shows that at least one of the premises must be false.

With a proof it is different. If somebody tells us that he has proved a, and we find that a is false, then, clearly, something must have been wrong with the argument, with the proof. In other words, we use the word "proof" intuitively in such a way that a proof must be invalid if the proved statement is false. A counter-example establishing
291 | the invalidity of a derivation must consist of true premises and a false conclusion. For a counter-example establishing the invalidity of a proof, it suffices to show that the conclusion (or a form-preserving interpretation of it) is false.

We see that a proof, as opposed to a derivation, does not assert the truth of the conclusion *provided* the premises are true, but that it asserts the truth of the, conclusion *absolutely* – independently of the question whether any particular other statement is true. The dependance upon any particular premise is, as it were, thrown off. Our definition shows how this is possible. If a statement is provable whenever it is derivable from any premise whatsoever then, if we suspect the truth of one premise, we can always replace it by another. The conclusion does no longer depend on any particular premise or set of premises, or on any particular assumption – it is established under all possible assumptions, under all possible circumstances. Since it can be derived from any description of facts, it must be true whatever the facts are, or independently of all facts.

But the strongest reason for accepting (9.1) as a satisfactory definition of logical demonstrability is that, if we accept this definition, we can show, in our metalanguage, on the basis of our definitions and without any further assumption, that all the well-known primitive propositions or axioms of the various logical calculi, and with them, of course, the theorems, are demonstrable.

Besides we can show that those methods which have been usually called proofs (such as the indirect proof, or the *reductio ad absurdum*) are indeed demonstrations in the sense of our definition, that is to say, that they establish the demonstrability of their conclusions, as opposed, for example, to syllogisms, the *modus ponens*, etc., which are not demonstrations but merely derivations.

As an interesting minor result it turns out that the dilemma differs from most of the other classical arguments in being a method or figure of demonstration and not a figure of derivation, such as the syllogism – in spite of the fact that it has been
292 treated, under the name of "*syllogismus* | *cornutus*", as if it had the same status as the categorical or the hypothetical syllogism.

Philosophically of greater interest is perhaps the result that, while valid *demonstrations must not be circular* – they must throw off any dependence upon any specific premise –, *derivations are always circular*; indeed, our analysis shows that all derivations are based upon absolutely valid rules, which are clearly circular (cp. example (d) and rule (6.1) in section 6), and upon inferential definitions, which cannot be the means of establishing anything new. The point is interesting in view of Mill's much discussed criticism of the syllogism.[b] Syllogisms are, of course, circular; they must be so since they are derivations; and any suggestion that the circle is vicious only reveals that they are mistaken for demonstrations.

To sum up, it is possible to construct a general metalinguistic theory of logic, applicable to any ⟨objectlanguage⟩[c], without assuming a distinction between formative and descriptive signs, without assuming anything like the customary primitive rules of inference, or any primitive propositions or axioms; and this construction is based on definitions which, in the last analysis, go back to the somewhat trivial concept of a statement-preserving interpretation.

References

Carnap, R. (1934a). *Logische Syntax der Sprache*. Wien: Springer Verlag.
– (1937). *Logical Syntax of Language*. Kegan Paul, Trench, Trubner & Co Ltd.
– (1942). *Introduction to Semantics*. Harvard University Press.
– (1943). *Formalization of Logic*. Cambridge, Mass.: Harvard University Press.
Johansson, I. (1937). Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus. In: *Compositio Mathematica* 4, pp. 119–136.
McKinsey, J. C. C. (1948). Review of Logic without Assumptions and New Foundations for Logic. In: *Journal of Symbolic Logic* 13 (2), pp. 114–115. Reprinted as § 13.14 of this volume.
Mill, J. S. (1843). *A System of Logic*. London: John W. Parker.
Popper, K. R. (1935). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*. Wien: Julius Springer.
– (1944a). The Poverty of Historicism, I. In: *Economica* 11 (42), pp. 86–103. Cf. Popper (1957).
– (1944b). The Poverty of Historicism, II. In: *Economica* 11 (43), pp. 119–137. Cf. Popper (1957).
– (1945b). The Poverty of Historicism, III. In: *Economica* 12 (46), pp. 69–89. Cf. Popper (1957).
– (1947b). Logic without Assumptions. In: *Proceedings of the Aristotelian Society* 47, pp. 251–292.
– (1947c). New Foundations for Logic. In: *Mind* 56 (223), pp. 193–235. Reprinted as Chapter 3 of this volume.
– (1957). *The Poverty of Historicism*. London: Routledge & Kegan Paul.
Tarski, A. (1933b). Pojęcie prawdy w językach nauk dedukcyjnych (= The Concept of Truth in the Languages of Deductive Sciences). In: *Prace Towarzystwa Naukowego Warszawskiego, Wydział III Nauk Matematyczno-fizycznych (= Travaux de la Société des Sciences et des Lettres de Varsovie, Classe III Sciences Mathématiques et Physiques)* 34, vii + 116pp.

[b] Cf. Mill (1843, Vol. I, Book II, Ch. 3, "Of the Functions, and Logical Value, of the Syllogism").

[c] Correction according to § 30.5.1, this volume.

Tarski, A. (1935a). Der Wahrheitsbegriff in den formalisierten Sprachen. In: *Studia Philosophica* 1, pp. 261–405. English translation in Tarski (1956).
– (1936b). Über den Begriff der logischen Folgerung. In: *Actes du Congrès international de philosophie scientifique, Sorbonne, Paris 1935*. Fasc. VII, Logique. Paris: Hermann, pp. 1–11. English translation in Tarski (1956).
– (1956). *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Translated by J. H. Woodger. Oxford: Clarendon Press.