Chapter 3
New Foundations for Logic (1947)

Karl R. Popper

Abstract This article is a corrected reprint of K. R. Popper (1947c). New Foundations for Logic. In: *Mind* 56 (223), pp. 193–235. Errata and additional footnotes have been inserted from K. R. Popper (1948e). Corrections and Additions to "New Foundations for Logic". In: *Mind* 57 (225), pp. 69–70.

Editorial notes: We have changed (D *n*) to (D*n*) throughout. The footnotes are now numbered consecutively, whereas in the original footnotes were numbered per page. Popper may have discussed the content of this article with Bernays, considering the last passage of their joint work "On Systems of Rules of Inference", Chapter 14 of this volume, where one can also find an explanation of footnote 9. The term "minimum calculus" refers to Johansson's (1937) "Minimalkalkül".

1. Introduction[1]

193 Logic ought to be simple; and, in a way, even trivial. Complications in logic all arise from two sources. Reiterated applications of trivialities may result, in the end, in such complexity that the thread is lost, as it were, and that one has to resort to the laborious method of carefully checking every step in order to ensure that all is well. A minute analysis is thus often made necessary, and this creates the other source of complication – the need for a high degree of precision in the formulation even of trivialities.

[1] This paper is a report and a discussion of some results obtained by the author during the last ten years. Formal proofs are avoided here but they will be published elsewhere. A knowledge of symbolic logic is not assumed; on the contrary, it may be necessary to warn readers against the attempt to identify our theory of compound statements and quantification with the calculi of propositions or functions. Our theory is more general: we do not construct one calculus or language, but we investigate a wide range of languages, among them some that do not contain any distinct symbols corresponding to those symbols which play such a role in the calculus of propositions and in similar calculi.

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_3

It is assumed here that the central topic of logic is the *theory of formal or deductive inference* (or of derivability; or of deducibility; or of logical consequence: all these expressions are here taken to mean the same thing). We shall first attempt to determine this notion of logical derivation or derivability by laying down a few very simple primitive rules for it. This will be done in sections 2 and 3.

In spite of the triviality of these rules, the concept of derivability thus determined turns out to provide an exceedingly powerful instrument. It will be shown that with
194 its help, we can draw up systems of rules of inference which cover not only | the theory of compound statements (corresponding to the "calculus of propositions") but also the theory of quantification, that is to say, of universal and existential statements (corresponding to the "lower functional calculus").

But we shall go even further. Surveying our system of rules of inference, we shall find that it is possible to lay down definitions of all the logical concepts, if we take derivability as our sole undefined specifically logical concept. But this means that all these other concepts,[2] in principle even though not in practice, can be dispensed with, and that the primitive rules of sections 2 and 3, in spite of their triviality, somehow cover the whole of the theory of inference.

2. General Theory of Derivation

If we wish to discuss rules which tell us under what circumstances a certain statement – the conclusion – follows from certain other statements – the premises –, then we have first to come to an agreement how to write such rules.

Take "*a*"; "*b*"; "*c*"; . . . to be names of statements; then we can write[a]

$$\begin{array}{c} a \\ b \\ c \\ \hline d \end{array}$$

or something like it, in order to express the assertion: "From the premises, a, b, and c the conclusion d can be derived." This manner of writing a rule of derivation is known from traditional logic, where one often writes rules of inference (or "moods") in this way; for example, the *modus ponendo ponens* is often written:

$$\begin{array}{c} \textit{If A then B} \\ A \\ \hline B \end{array}$$

[2] Except the concept "the result of substituting x for y in the formula a" which is introduced, in section 6, by way of a few exceedingly trivial primitive rules of inference (and which may be said to be implicitly definable in terms of derivability).

[a] In Popper (1947c) the following two derivations are inside quotation marks.

which may be read: “From the premise ‘*If A then B*’ together with the premise *A*, the conclusion *B* can be derived.”[3]

195 | In order to save space, and to make our way of writing less clumsy, we shall use here a slightly different notation; the main difference being that we shall write everything in one line, using the symbol “/” instead of the horizontal line; we shall, furthermore, use commas in order to divide the various premises, instead of writing them in different lines. In other words, the notation:

$$“a, b, c/d”$$

will be used in order to express the assertion: “From the statements a, b, and c, the statement d can be derived.”

Very often we shall have to discuss inferences from many premises. Let n be the number of the premises we are considering. Then the expression:

$$“a_1, \dots, a_n/b”$$

will be used for conveying the assertion: “From the statements $a_1, a_2, a_3, \dots, a_n$, the statement b can be derived.”

The sign “/” designates the *specifically logical concept of deducibility* whose importance was emphasised in the *Introduction*. It should be noted that, although we may operate with as many premises as we like, we draw only *one* conclusion at a time; in other words, an expression such as “$a_1, \dots, a_n/b, c$” has no meaning for us, as we are using our notation. If we wish to assert that b as well as c follows from $a_1, \dots, a_n$, then we have to say: “$a_1, \dots, a_n/b$ and $a_1, \dots, a_n/c$”.

Having explained our notation, we proceed to the discussion of actual rules of derivation. The most trivial and at the same time the simplest rule of inference is undoubtedly

$$a/a \tag{2.1}$$

in words: “From the statement a, the statement a can be derived.”

This rule is so trivial that many classical logicians have hesitated to admit it. They have felt that nothing could be gained by admitting it, and that the dignity of the procedure of inference would be imperilled if this were called an inference. But these considerations are irrelevant. Since the rule is, obviously[4], not invalid, we should have to admit it even if it were useless. But, contrary to first impressions, it is far from being useless: within those more subtle analyses of more complicated threads to which we referred in the first section, it turns out to be exceedingly useful; and it is even characteristic, altogether, of the triviality of the fundamental rules from which the edifice of logic arises. We should, furthermore, keep in mind that triviality of the

[3] For a precise rendering, Quine’s signs for so-called “quasi quotation” should be used instead of our single quotation marks. But since we shall not make any use of this device in what follows, it seems preferable not to burden the reader with it.

[4] In the sense of every adequate definition of inference such as Tarski’s; *cp.* his and my papers mentioned in ⟨footnote 9⟩. ⟨Footnote added in the Errata.⟩

196 basic | assumptions is an advantage rather than a disadvantage, provided that what we obtain, at the end, is adequate for our purpose.

Another trivial valid rule is the following:

$$\text{If } a_1, \ldots, a_n/b \text{ then } a_1, \ldots, a_n, a_{n+1}/b \tag{2.2}$$

which can be translated into: "If an inference is valid, then adding a new premise does not make it invalid." Indeed, the addition of new premises can only strengthen the premises, – just as an omission, in general, weakens them. ("Strengthening" and "weakening" will be defined and discussed later in this section.)

A third triviality is that the *order* of the premises does not affect the validity of the inference. This leads us to lay down the rule

$$\text{If } a_1, \ldots, a_n/b \text{ then } a_n, \ldots, a_1/b. \tag{2.3}$$

This rule permits us to reverse the order of the premises; by itself, it does not allow us to put the premises into any desired order. But, combined with the previous two rules, and with one (the principle of transitivity) which we shall soon discuss, it does, in fact, allow us to achieve precisely this. (This fact can be established; but we omit all such proofs in this paper.)

Before proceeding to the discussion of a more complicated rule, the principle of the transitivity of derivability, it should be mentioned that the three rules so far explained, *i.e.*, 2.1, 2.2, and 2.3, can (in the presence of the transitivity principle) be replaced by one single rule, of about equal simplicity, *viz.*, the following:

"From the premises $a_1, \ldots, a_n$, we can derive, as a conclusion, any statement a_i which is one of the premises, *i.e.*, for which $1 \leq i \leq n$." This can be expressed in our notation by

$$a_1, \ldots, a_n/a_i \qquad (1 \leq i \leq n). \tag{2.4}$$

The notation might perhaps irritate some who are not used to this kind of thing; but it is only the notation which makes this rule appear less trivial than the others; the idea is simplicity itself, and the rule, obviously, is valid. For if (as we asserted) the order of the premises is irrelevant; if adding to them only strengthens them, and does not invalidate the inference; and if

$$a/a$$

is valid, then, obviously,

$$a_1, \ldots, a_n/a_1 \tag{2.41}$$

must be valid, and also

$$a_1, \ldots, a_n/a_2$$

197 | provided n is, at least, equal to two; and also

$$a_1, \ldots, a_n/a_3$$

provided n is at least equal to three; and generally we have 2.4, *i.e.*,

$$a_1, \ldots, a_n/a_i$$

provided n is at least equal to i, which is precisely what we express by the condition "$1 \leq i \leq n$". We shall call 2.4 the "generalised principle of reflexivity" and refer to it by "(Rg)", *i.e.*, generalised reflexivity principle.

We now turn to the principle of transitivity. In its simplest form, this rule can be expressed by

(2.5) If a/b, then: if b/c then a/c

or in words: "If b follows from a, then: if c follows from b, then c follows from a." We shall refer to it by "(Ts)", *i.e.*, "simplest Transitivity principle".

This rule (Ts), although perhaps slightly less trivial, is still very simple. In the next section, we shall indicate[5] a method by which it can be retained in the simple form in which we have introduced it here. This method, however, makes use of a new concept, that of the conjunction of two statements, – say, the statement a and the statement b. The introduction of this concept simplifies certain things very considerably – this is, indeed, the reason why it is introduced. But it goes hand in hand with a restriction of the generality of our approach. This is a serious step which should not be taken before we have seen how far we can get with our present method. Besides, only if we see that things get a little complicated can we appreciate the simplification achieved by the new method. We shall therefore proceed to explain the more complicated forms of the rule of transitivity which are needed in many inferences.

But we shall first introduce here some linguistic abbreviations. We shall write in our rules sometimes "→" instead of "if . . . then . . ."; for example, we shall write a rule like 2.2 sometimes:

(2.2+) $a_1, \ldots, a_n/b \rightarrow a_1, \ldots, a_n, a_{n+1}/b$

or the rule 2.3:

(2.3+) $a_1, \ldots, a_n/b \rightarrow a_n, \ldots, a_1/b.$

The introduction of the sign "→" for "if . . . then . . ." is a very different matter from the introduction of a new logical concept. It is *merely* an abbreviation; and in order to emphasise this fact, we shall quite frequently revert to our ordinary way of writing.
198 | It is important to keep this in mind, and not to think that we introduce a logical symbolism in an underhand manner. (A new logical sign must be introduced explicitly by a rule or by rules determining its use. The only specifically logical sign we have so far introduced is "/", and we are still busy explaining the rules, such as 2.1, 2.2, 2.3, etc., which determine its use.)

[5] *Cp.* ⟨footnote 13⟩. ⟨Footnote added in the Errata.⟩

If we use our new abbreviation, rule 2.5, that is to say (Ts), becomes:

(2.5+) $$a/b \rightarrow (b/c \rightarrow a/c).$$

The brackets correspond to the colon of the original formulation.

Apart from "→" we shall make use of "&", in the usual way, as abbreviation for "and"; again, we shall indicate the informal character of this arrangement by reserving the right of continuing to write "and" whenever we like.

We now proceed to the more complicated forms of the transitivity principle.

The simple form discussed so far allows us only to handle cases of inference from *one* premise. It can be extended to the case of more than one premise by writing, instead of 2.5+:

(2.5e) $$a_1, \ldots, a_n/b \rightarrow (b/c \rightarrow a_1, \ldots, a_n/c).$$

In words: "If the statement b follows from the premises $a_1, a_2, \ldots, a_n$, then, if c follows from b, c also follows from $a_1, a_2, \ldots, a_n$." We shall refer to this form of the principle by "(Te)", *i.e.*, "extended transitivity principle". It is a very important principle, and it will be used, in the next section, together with (Rg) and another rule, to form what we shall call our "basis II", which makes use of conjunction. But at present, we do not wish to use conjunction, and we need, therefore, a stronger and more complicated principle which we shall call the "generalised principle of transitivity" and which we shall denote by "(Tg)".

This generalised principle of transitivity is, indeed, not so simple as our other principles, and this is the reason why we shall attempt to do without it, that is to say, why we shall design a method for avoiding its acceptance as a primitive rule.

Why we need it here is due to the following reason:

We may be able to derive the statements b_1 as well as b_2 from the premises $a_1, \ldots, a_n$; *i.e.*, we may have

$$a_1, \ldots, a_n/b_1 \;\&\; a_1, \ldots, a_n/b_2.$$

At the same time, we may have

$$b_1, b_2/c.$$

199 | Clearly, we have then also

$$a_1, \ldots, a_n/c.$$

Or, putting it in one line, the following rule will be generally valid:

$$(a_1, \ldots, a_n/b_1 \;\&\; a_1, \ldots, a_n/b_2) \rightarrow (b_1, b_2/c \rightarrow a_1, \ldots, a_n/c).$$

Now this is perhaps not much more complicated than (Te). But we cannot stop here. It is possible that c does not follow from b_1 and b_2 together, but that it follows from the m premises $b_1, \ldots, b_m$; or in other words, it may be that

$$b_1, \ldots, b_m/c$$

holds; and it is quite possible that *each* of these m premises of c, *i.e.*, *each* of the statements, $b_1, b_2, \ldots, b_m$, follows in turn as a conclusion from the premises $a_1, a_2, \ldots, a_n$. In this case,

$$a_1, \ldots, a_n/c$$

will hold also.

But how shall we write down a rule stating this more complicated kind of transitivity? There is no very simple way of writing it down.

One way of writing it is this:

$$(2.5g) \qquad (a_1, \ldots, a_n/b_1 \;\&\; a_1, \ldots, a_n/b_2 \;\&\; \ldots \;\&\; a_1, \ldots, a_n/b_m) \rightarrow (b_1, \ldots, b_m/c \rightarrow a_1, \ldots, a_n/c).$$

Another way of writing it is this:

$$(2.5g^1) \qquad \left\{ \begin{array}{l} \quad a_1, \ldots, a_n/b_1 \\ \& \; a_1, \ldots, a_n/b_2 \\ \vdots \qquad\quad \vdots \\ \& \; a_1, \ldots, a_n/b_m \end{array} \right\} \rightarrow (b_1, \ldots, b_m/c \rightarrow a_1, \ldots, a_n/c).$$

Or we may introduce a new abbreviation for any expression like

$$a/b_1 \;\&\; a/b_2 \;\&\; a/b_3 \;\&\; \ldots \;\&\; a/b_m$$

(whatever the number of the premises may be) by using instead the brief expression:

$$\prod_{i=1}^{m} a/b_i$$

This will be found repulsive by many, and we shall not continue to use anything as
200 complicated as this in the remainder of the | paper. But we may mention that our rule 2.5g is in this way much easier to write:

$$(2.5g\dagger) \qquad \prod_{i=1}^{m} a_1, \ldots, a_n/b_i \rightarrow (b_1, \ldots, b_m/c \rightarrow a_1, \ldots, a_n/c).$$

Another way out would be to write:

(2.5g‡) If, for all values of i between 1 and m (*i.e.*, for $1 \leq i \leq m$), we have $a_1, \ldots, a_n/b_i$, then we also have: if $b_1, \ldots, b_m/c$ then $a_1, \ldots, a_n/c$.

But whatever way of writing we may adopt – there is no doubt that this general principle of transitivity is considerably less simple and convincing than, say, (Te), *i.e.*,

$$(2.5e) \qquad a_1, \ldots, a_n/b \rightarrow (b/c \rightarrow a_1, \ldots, a_n/c).$$

Before we proceed, in the next section, to develop a method which allows us to dispense with anything more complicated than the rules (Te) or even (Ts), I wish to make clear that the system consisting of the two rules (Rg) and (Tg), *i.e.*, 2.4 and 2.5g, or alternatively, the system consisting of the four rules, 2.1; 2.2; 2.3; and (Tg), is sufficient as a basis for the construction of propositional and functional logic, as undertaken in sections 3 to 7. (The precise sense in which these sets form a "basis" will become clear in the sequel.)

The two systems are of *equal effect* in the sense that it can be shown that every inference which is valid in the one is also valid in the other. Furthermore, each of the two systems is independent in the sense that it can be shown of any one of the rules so far mentioned that, if it is omitted, the system in question is no longer of equal effect with the other system. (They can also easily be shown to be consistent.)

With the help of each of the two systems, a large number of what we may call "secondary rules of inference" can be shown to be valid. A rule is called "secondary" to some other rules which are "primary" relative to it if it does not add anything new to these primary rules; that is to say, if every inference which is asserted as valid by the secondary rule could be drawn merely by force of the primary rules alone, for example, by reiterated application of one of the primary rules, or by applying one after the other. ("Primitive" rules are rules of a set of rules such that all the other rules considered are secondary to the rules of this set; and an independent set of primitive
201 rules is called a "basis".) Many of the secondary rules of our system | are, of course, trivial, but some are not. Among the more trivial ones is the following (which we already know):

(2.41) $$a_1, \ldots, a_n / a_1$$

This is clearly a valid rule whenever 2.4 is valid.

(2.6) If $a_1, \ldots, a_n / b$ then $a_{n+1}, a_1, \ldots, a_n / b$.

This rule is very similar to rule 2.2, but it does not follow from 2.2 alone. (To say that a rule "follows" from certain other rules is merely an elliptical way of saying that it is secondary to them.) On the contrary, the system consisting of rules 2.1 to 2.3 and (Tg) remains independent if 2.3 is replaced by 2.6 which shows that 2.6 is not only independent of 2.2 but also of 2.2 together with 2.1 and (Tg).[6]

Another secondary rule is one which we mentioned when introducing the rule 2.3 (which permits the reversal of the order of the premises). It is the rule that the premises may be rearranged in any desired order. This simple rule, if formulated in our usual way, looks a little complicated – again a case where the complication is not due to the contents of a rule but to the need of expressing it precisely, and in a handy and standardised way. The rule is:

---

[6] I owe to Professor Paul Bernays the suggestion of replacing 2.3 by 2.6. It can be shown, however, that our original set 2.1 to 2.3 and (Tg) is in the following sense weaker than the new system (with 2.6 instead of 2.3). If, in our original system, 2.1 is replaced by the stronger rule 2.41, then the system remains independent. In the new system, if 2.1 is so replaced, 2.2 becomes dependent. (We thus obtain a system of three primitive rules, *viz.*, 2.41; 2.6; 2.5g.)

(2.7) $a_1, \ldots, a_{n+m}/b \rightarrow a_n, \ldots, a_1, a_{n+1}, \ldots, a_{n+m}/b \qquad (0 \leq n \leq n+m).$

Rule 2.7, together with 2.5g and 2.41, may be taken for another basis of equal effect with the ones mentioned.

An obvious but very important secondary rule is the following:

(2.8) If in any valid inference, one or more of the premises are such that they follow from the others, then they can be omitted as redundant, *i.e.*, without invalidating the inference.

In view of

$$a/a,$$

this rule 2.8 allows us to omit from a set of premises any premise which occurs more than once.

It will have been noticed that we have rules of two kinds. Some, such as "a/a"
202 or "$a_1, \ldots, a_n/a_n$" may be called *absolute* | rules of derivation; others such as "$a/b \rightarrow a, c/b$" (a secondary rule to 2.2) may be called *conditional* rules of derivation. In a conditional rule, we may distinguish between the antecedent rule and the consequent rule. Thus in the conditional rule

$$a, b/c \rightarrow b, a/c,$$

the rule "$a, b/c$" is the antecedent rule, while "$b, a/c$" is the consequent rule.

Some important and no longer trivial principles can be shown to hold of all such conditional rules. In order to formulate them, we first explain the terms "strengthening" and "weakening" as applied to premises or conclusions.

If a conclusion can be validly derived from some premise(s), we say that it is therefore logically *weaker* than the premise(s) or at most equal (in strength) to them, and that the premise(s) are logically *stronger* than the conclusion or at least equal to it. We shall, for brevity, say simply "weaker" and "stronger", but we have to keep in mind that these terms are here intended to be used in such a way as *not to exclude equal strength*; for the validity of "a/a" reminds us that a conclusion may, very obviously, be as strong as the premise from which it is drawn.

We further say that we "strengthen" (or "weaken") the premise(s) or the conclusion of a rule if we replace some or all of them by others in such a way that as a result they become logically stronger (or weaker, as the case may be); equal strength, again, is not excluded.

Now we can formulate some of the secondary rules which can be shown to hold in our system as follows:

(2.9) In a valid conditional rule,
- (a) the premises of the antecedent rule may be weakened,
- (b) the conclusion of the antecedent rule may be strengthened,
- (c) the premises of the consequent rule may be strengthened,
- (d) the conclusion of the consequent rule may be weakened,

without affecting the validity of the conditional rule.

It is possible to show that any change of any valid conditional rule of our system, effected in accordance with rule 2.9, leads to a new valid rule which, too, is within our system.

Another interesting principle is connected with *mutual deducibility* (*or logical equivalence*) of statements. If a and b are mutually deducible, we write

$$a//b.$$

203 | "//" is a new concept of our system, and can be easily defined on the basis of "/", by the following obvious Definition:

(D//) $\quad a//b$ if, and only if, a/b & b/a.

This definition may be abbreviated by using "↔" for "if, and only if":

(D//+) $\quad a//b \leftrightarrow (a/b \;\&\; b/a).$

We shall use "↔" quite often, but, in the same way as "→" and "&", as an abbreviation only.

Now it can be shown of our system that, whenever $a//b$, we can, without impairing its validity, substitute a for b, in some or all places of the occurrence of b, in any valid inference. If we call two mutually deducible statements "logically equivalent", then we can call this very general rule the *substitutivity principle for logical equivalence*. And we can read "$a//b$" not only "a is logically equivalent to b", but also "a is substitutionally equal to b"; and we can call mutual deducibility also "*substitutional equality*". (The problem of the substitutivity of logical equivalence will be mentioned again in later sections.)

But is our system of rules valid?

We can define validity of inference (following in the main Tarski[7]) in the following way: An inference is valid if every interpretation (*i.e.*, for our present purpose: every actual set of statements which can be used as an example of one of our rules) which makes the premises true also makes the conclusion true.

If we call an interpretation which makes all the premises true and the conclusion false a *counterexample*, then we can also say: An inference is valid if no counterexample exists.[8]

Now it can be very easily shown of all our rules that they are valid in this sense. Take, for example, the rule 2.41:

---

[7] *Cp.* A. Tarski, "Ueber den Begriff der logischen Folgerung" ⟨Tarski, 1936b⟩. We shall not discuss here the concept of interpretation. (Tarski speaks of "models".) But the fact, established in this paper, that formative signs can be defined in terms of a concept of deducibility which does not assume any formative signs in turn, opens a way to applying Tarski's concept without difficulty. I have in mind the difficulty, mentioned by Tarski, of distinguishing between formative ("logical") and descriptive signs. This difficulty seems now to be removed. (*Cp.* my "Logic without Assumptions," forthcoming ⟨as Popper, 1947b⟩.)

[8] See also my "Why are the Calculuses of Logic and Mathematics Applicable to Reality?" ⟨Popper, 1946c⟩, esp. pp. 47ff.

$$a_1, \ldots, a_n / a_1.$$

204 | Assume that a counterexample exists. Then there must be a certain interpretation which makes all the premises true and, at the same time, the conclusion false; thus, in the light of this interpretation, the conclusion a_1 will be a false statement. But since a_1 occurs in the premises also, they cannot possibly be all true statements. Accordingly, no counterexample can exist.

In a similar way, all the other rules can easily be shown to be valid, and this explains their triviality: the degree of their triviality is, as it were, in proportion to the ease with which their validity can be shown.

3. Another Approach. Conjunction

We now proceed to discuss the replacement of the complicated rule (Tg) ⟨or⟩ 2.5g by the simpler rule (Te) or 2.5e. But in order to do this, we must somewhat restrict the generality of our approach.

Indeed, the rudiment of a theory of inference sketched in the foregoing section[9] is about as general as such a theory can possibly be. It can be applied to any language in which we can identify statements (or sentences, or propositions – there is no need for us here to enter into the problem of the possibility or necessity of distinguishing between these entities), that is to say, expressions of which we might reasonably say that they are true or that they are false. Nothing is presupposed of our $a, b, c, \ldots$ except that they are statements, and our theory shows, thereby, that there exists a rudimentary theory of inference for any language that contains statements, whatever their logical structure or lack of structure may be. (But this means, for any *human* language, that is to say, for any language which is not only expressive and evocative of response, as animal languages are, but also descriptive, *i.e.*, containing means for describing facts, or statements of facts, which may be true or false.)

We may mention here in passing that our symbols "*a*"; "*b*"; "*c*"; etc., are *variables*, similar to those used by mathematicians in algebra; and as the values of the mathematicians' variables are *numbers*, so the values of our variables are statements.
205 | The student of algebra studies a universe of discourse consisting of numbers (of some kind – say, natural numbers, or real numbers). We study a universe of discourse consisting of statements (of some kind – say, the statements of a certain language). The mathematician allows that certain *figures* (not numbers) – for example, the figures "1", "2", etc., may be substituted for his variables; that is to say, what may replace his variables is any *name* of a number. Similarly, we must assume that *names of*

[9] Professor Bernays has drawn my attention to the close relationship that exists between this theory and the system of five axioms for the theory of consequence developed by A. Tarski in his "Ueber einige fundamentale Begriffe der Metamathematik" ⟨Tarski, 1930b⟩. Bernays designed a method of translating Tarski's *prima facie* very different and more abstract approach into a more elementary one, very similar to the approach presented in this section. (He has also drawn my attention to a somewhat similar system; see ⟨Hertz, 1923⟩ and ⟨Hertz, 1931⟩.) With the help of this method it can be easily shown that our system is equivalent to the first four of Tarski's five axioms.

statements (not the statements themselves) may be substituted for our variables. This is what we mean when we say that the mathematicians' symbols "a"; "b"; etc., are variable names of numbers, *i.e.*, variables which designate unspecified numbers; in the same sense, our symbols "a"; "b"; etc., are variables which designate unspecified statements, and they may be described, in the sense indicated, as variable names of statements.

Now the universe of statements envisaged in the foregoing section may consist of the statements of any language, whatever its logical structure. But in the present section, a certain restriction is imposed on the admissible languages – they must contain what we shall call *conjunctions*.

We shall define conjunction in what follows. The intuitive meaning of our definition can be conveyed perhaps in this way: We say that a language L contains conjunctions (or that it contains the operation of conjunction) if we can join any two statements – the statement a, say, and the statement b – in such a way as to form a new statement c which is logically equivalent to a and b together. This is done, in English, by linking them together with the help of the word "and". But we need not suppose that any such word exists: the link may be effected in very different ways; moreover, the new statement need not even contain the old ones as recognisable separate parts (or "components"). We may therefore say:

A language L contains the operation of conjunction if, and only if, it contains, for any two statements, a and b, a third statement c which is logically equivalent to them (*i.e.*, a little more precisely, equivalent to the first two statements taken together).

Now the remark on logical equivalence definitely needs here some clarification. According to the way in which we have introduced "//", we cannot write "$a, b // c$"; rather, only *one* statement is allowed to occur on the left- as well as on the right-hand side of "//" only. For we must remember that our conclusions always consist of *one* statement only; and in the case of "//", conclusions (as well as premises) stand on both sides.

206 | Nevertheless, the statement c may be said to be equivalent to the two statements a, b if we have:

$$c/a \tag{1}$$

$$c/b \tag{2}$$

$$a, b/c \tag{3}$$

We are thus led to a definition along the following lines:

(D3.01) The statement c is a conjunction of the two statements a and b if, and only if, c/a, c/b, and $a, b/c$.

(D3.02) A language L contains the operation of conjunction if, and only if, it contains, with every pair of statements, a, and b, a third statement c which is the conjunction of a and b.

Definitions D3.01 and D3.02 are as such quite in order, and adequate. But in view of our aim – the simplification of the transitivity principle – we shall *not* adopt D3.01. Rather, we shall replace D3.01 by a similar but slightly stronger definition.

The situation is this: D3.01 makes, of course, very essential use of our specifically logical concept "/". Its force will depend, therefore, upon the force of this concept "/"; but this means upon the primitive rules of derivation assumed to hold. A closer analysis shows that D3.01 is perfectly adequate as long as we assume all those rules of inference which we are accustomed, intuitively, to use; which means, in effect, the systems of primitive rules discussed in section 2. But if we drop some of these rules, or if we replace 2.5g by the weaker form of the transitivity principle 2.5e, then our definition is no longer quite adequate. It turns out, in this case, that it does not entitle us to assert that *more than two* separate statements can be joined into a conjunction or, in other words, that there exists, for any three statements a, b, and c, an equivalent statement d. On the basis provided by section 2, this can be shown to hold, but not if 2.5g is replaced by 2.5e or 2.5. (It is in these somewhat subtle points that our trivial system of logic becomes complicated.)

But the main point of using conjunction is just that it should permit us to *link up any number of statements into one*. If it does this, then we can always replace the m conclusions of m different inferences from the same premises by one inference. But if we can do this, then (Te) or perhaps even (Ts) will be, clearly, sufficient. At the same time, many other simplifications would be possible. For example, we could use "//" in order to express equivalence between $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ – namely by first replacing each of the two sets by the conjunction of all statements belonging to it.

207 We see here that it is a practical need – the need for | simplification, abbreviation, greater lucidity of certain formulations, etc. – which leads us to introduce conjunction; and we may safely assume that it was precisely such a need which led, in the actual development of most human languages to the invention of a linguistic device, such as the use of the word "and", which guarantees that the language contains conjunction in the sense defined above.

Since our definition is, as we have said, only adequate on the basis of the whole system of section 2, we shall have to modify it in order to achieve our purpose. But before proceeding to do so, we shall make some comments upon another aspect of our definitions D3.01 and D3.02. These comments will go into some detail, but our labours will be rewarded, for our findings concerning conjunction apply equally to the hypothetical, the disjunction, the negation, the quantifiers, etc.

If we designate the conjunction of the statements a and b by the letter "c", it is necessary for us always to add, explicitly, whenever we mention c, that this c is a conjunction of the statements a and b. We can avoid this by introducing a new sign, such as "$\wedge(a, b)$" – which may be read: "the conjunction of the statement a and the statement b", – or, more briefly, "$a \wedge b$", – which may be read "(the statement) *a-and-b*", or simply "*a-and-b*". Using this method, D3.01 becomes:

(D3.01+) The statement $a \wedge b$ is a conjunction of a and b if, and only if, $a \wedge b/a$, $a \wedge b/b$, and $a, b/a \wedge b$.

But this can be split up into two parts: a definition D3.0 and three primitive rules of inference:

(D3.0) "$a \wedge b$" is the (variable) name of the conjunction of the two statements a and b (called the two components of the conjunction).

The three rules of inference for $a \wedge b$, are:

(3.1) $$a \wedge b/a.$$

(3.2) $$a \wedge b/b.$$

(3.3) $$a, b/a \wedge b.$$

It is clear that (D3.0) is nothing but the introduction of a new label, and that only the three rules give the new label a meaning, – by relating it to deducibility.

A warning should be given here: as we have introduced it, *the symbol* "$\wedge$" *has no separate meaning at all*. It would be a mistake to consider it as an abbreviation of the word "and", – perhaps even as an alternative abbreviation to "&". We do not even assume that the language we are discussing – the language to which our statements $a, b, c, \ldots$ belong – possesses a special sign for linking statements into conjunctions.
208 We can | easily imagine a language L, consisting, say, of simple subject-predicate statements, in which the end of a whole statement is always indicated by the word "STOP". In such a language, a conjunction of two or more statements may be expressed simply by omitting the word "STOP" between them, rather than by inserting a word corresponding to our "and". Furthermore, a language may dispose of different means for expressing conjunction. Our language may perhaps express conjunction not only by the omission of the word "STOP" but alternatively (taking a clue from Latin) by replacing the word "STOP" by the word "QUE"; and this may mean that the two preceding statements are intended to form a conjunction. (If three statements are to be linked, "QUEQUE" may conclude the third, etc.)

Any number of different methods of expressing conjunction may be invented, and may actually occur, even within one and the same language.

It is for this reason that we say that $a \wedge b$ is a conjunction of a and b, rather than that it is *the* conjunction of a and b. But however many conjunctions of a and b we may have in a language, they all must be equivalent if they satisfy our rules 3.1 to 3.3; for any two conjunctions of a and b which satisfy these rules (and those of section 2) can be shown to be mutually deducible.

The upshot of all this is that what we have defined is not so much the conjunction of a and b but the precise *logical force* (or the logical import) of any statement c that is equal in force to a conjunction of a and b. This may be made clearer, perhaps, by re-formulating our definition D3.01 in this way:

(D3.03) The statement c is equivalent (or substitutionally equal) to any conjunction of the statements a and b, if, and only if, c/a and c/b and $a, b/c$.

Or in an alternative way of writing:

(D3.03+) $$c//a \wedge b \leftrightarrow (c/a \;\&\; c/b \;\&\; a, b/c).$$

From this definition, we can obtain the three rules 3.1, 3.2 and 3.3 in the following way: if c and $a \wedge b$ are substitutionally equal, then: (1) $c/a \;\&\; c/b \;\&\; a, b/c$ (by force of D3.02), and (2) we may substitute "$a \wedge b$" for "c" (by force of substitutional equality). But the result of this substitution leads precisely to the rules 3.1 to 3.3.

It will now be reasonably clear, I hope, what we mean if we say that we can define conjunction on the basis of the system of section 2, with "/" as our only specific logical term. It may be mentioned, in addition, that if we wished, we could easily proceed further, and define, with the help of our definition of a conjunction, the *sign of*
209 *conjunction itself* (in English, the word "AND") | along the following lines. We could distinguish between preceding, intervening, and succeeding signs of conjunction (*i.e.*, signs preceding the two statements to be linked, or placed between them – like "AND" – or succeeding them – like "QUE" in our example above). Then we can define: A sign S of a language L is a preceding sign of conjunction in L if, and only if, for any pair of statements a and b of L, an expression consisting of S followed by a and by b is, in L a conjunction of a and b. Corresponding definitions would have to be laid down for intervening and succeeding signs of conjunction. Next, we could define: A language L contains a (preceding, etc.) sign of conjunction if, and only if, it contains a sign S such that S is a (preceding, etc.) sign of conjunction.

These definitions are intended to assume as little as possible about the particular method by which a particular language may express conjunction. But they are not applicable to *all* languages containing conjunction; they cannot achieve generality. It is always possible to design new methods of expressing conjunction which are not covered by any definition of this kind (for example, underlining in written language, intonation in spoken language, etc.); and I have mentioned this way of defining the term "sign of conjunction in L" merely in order to show that we do not, perhaps, pay for the generality of our approach by an incapacity of applying our theory to the usual types of language.

These comments upon the character of our definitions, and on the possibilities of extending them to the definition of certain signs of some language, apply, as indicated, not only conjunction, but equally to disjunction, negation, hypotheticals, the quantifiers, etc., that is, to all *formative signs*[10] of a language.

We now turn back to the particular problems of conjunction. First we may mention that our definitions can be made much neater; for it so happens that the rules 3.1 to 3.3 can be replaced, in the presence of the rules of section 2, by either of the two following rules:

$$a \wedge b/c \;\leftrightarrow\; a, b/c. \tag{3.4}$$

$$c/a \wedge b \;\leftrightarrow\; c/a \;\&\; c/b. \tag{3.5}$$

210 | Using 3.4, instead of using 3.1 to 3.3 (as we did in our definition D3.03+) we can replace D3.03+ by the following definition:

$$c//a \wedge b \;\leftrightarrow\; (\text{for any statement } d\text{: } c/d \;\leftrightarrow\; a, b/d). \tag{D3.04+}$$

Similarly, 3.5 would give rise to the following definition:

[10] I prefer to speak of the "formative signs" of an object language where Carnap and Tarski speak of "logical signs". The reason is that I consider logic as a purely metalinguistic affair; on the other hand, the signs in question (of the object language) characterise the formal structure of that language.

(D3.05+) $c//a \wedge b \leftrightarrow$ (for any statement d: $d/c \leftrightarrow d/a$ & d/b).

Each of these two definitions is adequate, on the basis of the system of section 2, but each of them presupposes (Tg), *i.e.* 2.5g, for adequacy.

But there is a method of generalising all these rules, by introducing many premises, $a_1, \ldots, a_n$, instead of one premise a. The various generalised rules look like this:

(3.1g) $a_1, \ldots, a_n, b \wedge c/b.$

(3.2g) $a_1, \ldots, a_n, b \wedge c/c.$

(3.3g) $a_1, \ldots, a_n, b, c/b \wedge c.$

(3.4g) $a_1, \ldots, a_n, b \wedge c/d \leftrightarrow a_1, \ldots, a_n, b, c/d.$

(3.5g) $a_1, \ldots, a_n/b \wedge c \leftrightarrow a_1, \ldots, a_n/b$ & $a_1, \ldots, a_n/c.$

On the basis provided by section 2, each of the last two rules,[11] 3.4g as well as 3.5g, turns out to be replaceable by the three rules 3.1g to 3.3g, and *vice versa*; but this result is in so far trivial as, on this basis, it also turns out that the generalisation is superfluous, that is to say, that we can everywhere omit the premises $a_1, \ldots, a_n$, which brings us back to our original forms 3.1 to 3.5. But this situation changes completely if we drop the generalised transitivity principle (Tg) and replace it by the simpler form (Te). In this case, all the rules 3.1 to 3.5 and 3.1g to 3.4g still follow from 3.5g;[b] but the opposite is not the case.

Thus rule 3.5g is more powerful[12] than the others; and indeed, it turns out that it allows us to dispense with (Tg); that is to say, it can be shown that, in the presence of 3.5g, (Tg) is secondary to (Te) and (Rg).

We shall briefly sketch the proof of this contention. The assumption of the generalised transitivity rule (Tg) is that the following m rules are valid:

$$a_1, \ldots, a_n/b_1 \;\&\; a_1, \ldots, a_n/b_2 \;\&\; a_1, \ldots, a_n/b_3 \;\&\; \ldots \;\&\; a_1, \ldots, a_n/b_m \qquad (1)$$

Now we obtain, from the first of these, *i.e.*, from

$$a_1, \ldots, a_n/b_1 \;\&\; a_1, \ldots, a_n/b_2 \qquad (2)$$

by rule 3.5g,

$$a_1, \ldots, a_n/b_1 \wedge b_2 \qquad (3)$$

211 | accordingly, we also have, by (1),

---

[11] Rule 3.4g has been suggested to me by Professor Bernays. This led me to adopt the simplified rule 3.4 (as rule 4.1 in section 4).

[12] It is more powerful relative to those transitivity principles which we are considering here (with a view to avoiding an unspecified number of antecedents). ⟨Footnote added in the Errata.⟩

[b] This statement is not correct; 3.4g does not follow from 3.5g. The error will lead to the failure of Basis II; cp. 4.6. Cp. also the remark by Bernays in his letter to Popper, 12 May 1948 (this volume, § 21.7).

$$a_1, \ldots, a_n / b_1 \wedge b_2 \;\&\; a_1, \ldots, a_n / b_3 \tag{4}$$

and, applying rule 3.5g again, we obtain

$$a_1, \ldots, a_n / (b_1 \wedge b_2) \wedge b_3 \tag{5}$$

and so on, until all the various conclusions $b_1, \ldots, b_n$ are incorporated into *one*, as components of a conjunction. Now this process, which makes use only of 3.5g, allows us to replace (1) by one single rule of inference. To this single rule we can apply (Te), the extended form of the transitivity rule.[c]

We may sum up our results as follows:

We have now two approaches for deducibility.

Approach I consists of the systems of section 2 – in its simplest form, a basis consisting of two primitive rules: rule (Rg), the generalised principle of reflexivity, and rule 2.5g the generalised principle of transitivity, denoted by (Tg). We repeat these two rules, which constitute what we call the basis I:

(Rg) $$a_1, \ldots, a_n / a_i \qquad (1 \leq i \leq n)$$

(Tg) $$\prod_{i=1}^{m} a_1, \ldots, a_n / b_i \rightarrow (a_1, \ldots, a_n / c \rightarrow b_1, \ldots, b_m / c)$$

This approach is applicable to any language.

Approach II is less generally applicable: it assumes that the language under consideration contains (in accordance with definition D3.02) the operation of conjunction. In its simplest form, called "basis II", it consists of a postulate (PC), expressing this assumption and replacing the definition D3.02, and of three primitive rules, *viz.* of (Rg) and (Te), the simple form of the principle of transitivity previously called 2.5e, and of 3.5g, which we shall call the generalised principle of conjunction. This is basis II:

(PC) If a is a statement and b is a statement, then $a \wedge b$ is a statement.

(Rg) $$a_1, \ldots, a_n / a_i \qquad (1 \leq i \leq n).$$

(Te) $$a_1, \ldots, a_n / b \rightarrow (b / c \rightarrow a_1, \ldots, a_n / c).$$

(Cg) $$a_1, \ldots, a_n / b \wedge c \leftrightarrow (a_1, \ldots, a_n / b \;\&\; a_1, \ldots, a_n / c).$$

These two approaches[13] are, of course, not equivalent; basis II is richer than basis I, since it contains, for example, all the rules 3.1 to 3.5, and 3.1g to 3.5g (which basis I
212 does not contain), and | many other rules which may be derived from it; but they become equivalent in force if we add to basis I postulate (PC) together with, for

[13] There are a number of alternative bases in each of the two approaches. I mention only basis I′, consisting of 2.1; 2.2; 2.3; either (Tg) or 2.8; basis II′ uses only (Ts), together with 2.1; 2.2; 2.3; Cg, and 3.4g.

[c] Popper's proof is not correct; $b_1, b_2/c \rightarrow b_1 \wedge b_2/c$ does not follow from the rules of Basis II. Cf. this volume, § 4.6.

example, the rules 3.1 to 3.3 (or with one of the rules 3.4 or 3.5, or else with one of the Definitions D3.03 to D3.05).

4. The Logic of Compound Statements

Either of the two bases described in the foregoing sections suffices for the construction of modern logic,[14] *i.e.*, of the logic of compound statements[15] and of the logic of quantification (or the propositional and the lower functional calculus), and therefore *a fortiori* for constructing classical logic, so far as it is valid.

The logic of compound statements comprises the rules of inference pertaining to conjunction, to the hypothetical or conditional statement, its converse, and the bi-conditional, to disjunction, negation, etc. The system of these rules can be constructed by precisely the same method as was used in the previous section for introducing conjunction:

We first assume postulates (such as PC in section 3), one for each of the compounds we wish to introduce, in order to assure, for every pair of statements, say a and b, the existence of the corresponding compound statement. In this way, we may postulate for every pair of statements a and b the existence of a statement $a > b$ (if we want our language to contain hypothetical compounds); of a statement $a < b$; (if we want converse hypotheticals); of a statement $a \mathrel{\widehat{=}} b$ (if we want bi-conditionals); and of a statement $a \vee b$ (if we want disjunctions). Furthermore, we postulate that for every statement a, a statement $\neg a$ exists (if we want negations).

These postulates have the function of confining our investigation to languages in which such compounds exist. They can be replaced by definitions such as D3.02; or alternatively we may formulate our theory hypothetically, by introducing an assertion about deducibility by a remark such as "If a language contains negation, then the following rules of inference hold: . . ." In other words, the postulates or definitions do not really form a part of our theory of inference – they are there solely to indicate
213 explicitly that the application of our theory is limited, if we | wish to operate with certain compounds, to languages which contain these compounds.

The meaning of the compounds will be determined by *primitive rules of derivation* (4.1 to 4.6), one primitive rule for each of the six compounds mentioned.

It may be asked whether any actually spoken language contains these compounds? The answer, I believe, is: they do not; but they contain them approximately. The actually spoken languages are a kind of natural growth, with rules of use which are not clearly defined, and which cannot be clearly defined; if we define them in a definite way, then we replace the actually spoken and growing language by something else –

[14] I exclude the theory of membership (as Quine calls it) as mathematical. Classical syllogistic (so far as valid) appears as a small part of the logic of statement functions with one variable.

[15] The term "compound" is in so far misleading as we do not assume here that the so-called "components" are actually parts of the statements (the conjunction, etc.) whose components they are; see section 3. That is, we do not assume statement-composition, although our theory is applicable to it.

by a kind of artificial reconstruction of it, or to put it bluntly, by a kind of artificial calculus.

This cannot be otherwise; for there is no reason why all people who speak a certain naturally grown language should conform, in their habits of speech, to a precise system of rules of use. On the contrary, they will, at best, develop that degree of precision which is requisite to the ordinary problems of the day (rather than to scientific and, more especially, mathematical and logical problems). A grammarian who attempts to legislate for them in fact attempts to impose an artificial calculus upon them; as a rule, he obtains his calculus by a process of idealising the speaking habits of what he considers to be well educated people. The logician also constructs an artificial model language or calculus, but as a rule (I hope) without attempting to impose it upon anybody.

Why does he do it? As I indicated above, the fundamental problem of logic is the theory of inference – in the most elementary form, the drawing up of rules of valid inference. In this task, he cannot succeed without introducing some artificial precision into the language in which these inferences are formulated. For what he aims at is, at the very least, a distinction between valid and invalid inferences. But whether or not a certain concrete instance of an inference is valid will depend, very often,[16] upon the meaning of certain words which may occur in the premises or in the conclusion; or more precisely, on the rules determining the use of these words. If, therefore, these rules are fluid, then the question whether or not a certain inference is valid cannot be answered.

214 | Thus the logician cannot fulfil his most elementary task without constructing something like an artificial model language; and every logician from Aristotle onward has done so – although few were quite clear about what they were doing.

The primitive rules of derivation with the help of which we are about to determine the meaning of our compounds, in a way go beyond this method of constructing an artificial model language – they characterise a very wide class of (artificial) languages as conforming to certain definite and precise rules – rules which make it possible for us to develop a theory of inference.

Still, when all is said and done, it so happens that to each of our artificial compounds, there exists a closely corresponding compound proposition in English. The correspondence between the conjunction $a \wedge b$ and the English statements constructed with "and", for example "It rains and it is wet", are particularly close. But even the conditional statement $a > b$ corresponds very closely indeed to some current English expressions. The correspondence with a phrase such as "If it rains, then it is wet" or "Provided it rains, it is wet" is quite good, but that with the more complex statement "It does not rain without its being wet" is better. Accordingly, we can pronounce the symbol "$a > b$" perhaps "if-a-then-b" or better "a-not-without-b", remembering, however, that it is *not* an abbreviation for the English phrase but a *variable name* of any statement which stands in a certain logical relationship to the two statements a and b.

[16] But not always; for the theory of section 2 (or of our approach I) is independent of the meaning of any particular *word*; all it presupposes is that we can identify *statements*.

Similarly, we can read "$a < b$" perhaps "*a-if-b*" or "*a-provided-b*", and "$a \mathrel{\widehat{=}} b$" perhaps "*a-if-and-only-if-b*".

The disjunction $a \vee b$ does not correspond badly to a phrase such as "It rains or it is wet or both". The correspondence with the briefer phrase "It rains or it is wet" is very bad indeed, but we may for the sake of brevity pronounce "$a \vee b$", nevertheless "*a-or-b*", having in mind that this is short for "*a-or-b-or-both*".

There is no very fitting English phrase corresponding to the negation $\neg a$. The negation of the statement "It rains", corresponds very well to "It does not rain" (rather than to the impossible phrase "Not it rains"), yet the negation of "All men are mortal" does not correspond to "All men are not mortal" but rather to "Not all men are mortal". A generally applicable phrase which is not too bad in English and which corresponds very well to $\neg a$ is "It is not the case that . . .". Abbreviating this, we can read "$\neg a$" perhaps as "*not-a*".

215 | We now state our primitive rules:

(4.1) $$a \wedge b/c \;\leftrightarrow\; a, b/c.$$

Or: "From the one premise *a-and-b* we can derive c if, and only if, from the two premises a and b we can derive c." (This is the same as 3.4. Note the trivial character of the rule. Nevertheless, it is important and very powerful in connexion with basis I; with basis II it is, of course, redundant.)

(4.2) $$a/b > c \;\leftrightarrow\; a, b/c.$$

Or: From the statement a we can derive the statement *b-not-without-c* if, and only if, we can derive from the two statements a and b the statement c. (Example: From "It rains incessantly every week-end" we can derive "Provided I frequently walk during week-ends, I frequently walk while it rains" if, and only if, we can derive from the two premises "It rains incessantly every week-end" and "I walk frequently during week-ends" the conclusion "I walk frequently while it rains".)

(4.3) $$a/b < c \;\leftrightarrow\; a, c/b.$$

(4.4) $$a/b \mathrel{\widehat{=}} c \;\leftrightarrow\; a, b/c \;\&\; a, c/b.$$

(4.5) $$a \vee b/c \;\leftrightarrow\; a/c \;\&\; b/c.$$

(4.6) $$\neg a, b/\neg c \;\leftrightarrow\; c, b/a.$$

Rules 4.1 to 4.5 constitute *the positive logic of compound statements*, that is to say, they suffice for that part of propositional logic which is independent of negation. If our rule 4.6 for negation is added, we obtain the whole of compound statements (or of propositional logic). If, instead, the two rules for quantification discussed in section 7 are added, we obtain the whole of positive logic.

Positive logic as defined by these rules does not yet contain all valid rules of inferences in which no use is made of negation: there is a further region which we may call the "extended positive logic" (or, if we exclude quantification, the "extended positive logic of compound statements"). This extension is achieved by adding to the system without 4.6 an additional rule for hypothetical statements, *viz.*:

(4.2e) $$a, b > c/b \leftrightarrow a/b.$$

This rule may be, loosely, characterised as asserting that, if we add to a premise another one which merely informs us that something depends on our conclusion, then we do not thereby strengthen in any way the power of our original premise to yield
216 that particular conclusion. An example is: From the two | premises “All men are mortal” and “If Plato is mortal then he cannot be a god”, we can derive the conclusion “Plato is mortal” if, and only if, we can derive this conclusion from the first premise alone.

If rule 4.2e is added as primary to our rules 4.1 to 4.5, then all valid rules of the logic of compounds in which negation does not occur become secondary rules. Also we do not need, in the presence of rule 4.2e, the whole force of our rule of negation 4.6, but can obtain this rule as a secondary rule from some weaker rules[17] (from the rules of the so-called intuitionist logic). On the other hand, if we have as primary the two rules 4.2 and 4.6, then 4.2e becomes a secondary rule.

Considering the triviality of these rules – especially of the “positive” rules 4.1 to 4.5 – their power is truly amazing. To give a few examples:

From 4.1 alone (of course, the basis is always assumed), we can obtain as a secondary rule the associative law for conjunction

(4.01) $$(a \wedge b) \wedge c // a \wedge (b \wedge c);$$

furthermore, on the basis I, all the rules of section 3.

From 4.2 alone, we obtain, on either of the two bases, first of all

(4.020) $$a, a > b/b,$$

that is to say, the *modus ponens*; furthermore, the important rules

(4.021) $$a/b > a.$$

(4.022) $$a > (a > b)/a > b.$$

On the basis II, or if we add either rule 4.1 or 4.6, we obtain furthermore

(4.023) $$a > b/(b > c) > (a > c).$$

(4.024) $$a, b/c > d \leftrightarrow a, b, c/d.$$

This last rule is a kind of generalisation of 4.2 which cannot, on the basis I, be obtained from 4.2 alone, but which can also be obtained with the help of 4.6 (instead of 4.1 or of basis II).

From the rule 4.024 alone, all the rules of positive logic containing no compound

[17] In fact, 4.6 assumes more than we need: we can replace in it “↔” by “→”, and still get 4.6 as secondary rule. It may be mentioned that our rule 4.6 amounts in effect only to a principle underlying the classical theory of “indirect reduction”.

except hypotheticals are derivable without any other help, – of course, always assuming
217 at least basis I. | Among these rules are

(4.025) $$a > (b > c)/b > (a > c).$$

(4.026) $$a > (b > c)/(a > b) > (a > c).$$

If 4.2e is added, we obtain *all* valid rules containing no compound other than hypotheticals. If we combine 4.1, 4.5 and 4.6, then we obtain, for example

(4.5) $$a \wedge (b \vee c)//(a \wedge b) \vee (b \wedge c).$$

From 4.6 alone we obtain, on either of the two bases, for example:

(4.061) $$\neg a, b_1/a \rightarrow \neg a, b_1/b_2 \ \&\ b_1, b_2/a,$$

a rule which, in its effect, is equal to 4.6; furthermore

(4.062) $$a//\neg\neg a.$$

(4.063) $$a, \neg a/b.$$

(4.064) $$a/b \ \&\ \neg a/b \rightarrow c/b.$$

(4.065) $$\neg c/b \rightarrow (a, b/c \leftrightarrow a/c).$$

These rules can hardly be called trivial; 4.063, for example, is a rule whose validity has been questioned in this journal[d] not very long ago; 4.065 is, in spite of its comparative simplicity, hardly self-evident.

These examples are here given mainly with the intention of illustrating our thesis that we should not despise trivial primitive rules: they may give rise to secondary rules which are far from being trivial.

Regarding the question of the triviality of our primitive rules, it may further be observed that the rules 4.1 to 4.5, that is, the rules of the positive logic of compounds, are not only very obvious, but have, besides, the character of definitions (so-called contextual definitions) of the compounds which they introduce. The newly introduced symbol (the definiendum) occurs only once, on the left-hand side, while on the other side our specific logical term "/" is used, as defining term, apart from certain phrases of ordinary English, or their abbreviations.

The position is different with the rule of negation, 4.6. Here *two* negations occur on the left at the same time: the one is somehow linked up with the other, and can, therefore, not always be eliminated alone. Rule 4.061 is, in this respect, perhaps slightly preferable. But it has a very similar defect. The variable "*a*" occurs twice, and the rule is therefore not necessarily an effective means of eliminating the symbol "¬" from any context. But a definition should be able to achieve this.

218 | The peculiar situation of negation in this respect is connected, I believe, with the fact that various systems of logic have been developed which use a weaker form of negation (notably Heyting's intuitionistic logic which formalises Brouwer's views on

[d] Popper here refers to Jeffreys (1942).

logic). It is, therefore, of considerable interest to search for an explicit definition for negation which might permit us to eliminate it from any context, and which might bring it into line with the other compounds.

5. Explicit Definitions of the Compounds

It can be easily shown that the introduction of the compounds cannot in any way impair the substitutivity of logical equivalence. That is to say, whenever a and b are mutually deducible, we may substitute a for b, or b for a, at some or all places of occurrence, in any rule of inference, however complicated, without affecting the validity or invalidity of the rule. If we, therefore, use the method already exemplified by definitions D3.03+ to D3.05+, then we cannot fail to obtain definitions which allow us to eliminate the defined sign without difficulty from any context.

All the definitions so obtained are rather obvious adaptations of the primitive rules of the last section, except the definition of negation which is based on 4.061 rather than on 4.6.

The Conjunction:

(D5.1) $$a//b \wedge c \leftrightarrow \text{(for any } c_1\text{: } a/c_1 \leftrightarrow b, c/c_1).$$

The Hypothetical:

(D5.2) $$a//b > c \leftrightarrow \text{(for any } a_1\text{: } a_1/a \leftrightarrow a_1, b/c).$$

The Converse Hypothetical:

(D5.3) $$a//b < c \leftrightarrow \text{(for any } a_1\text{: } a_1/a \leftrightarrow a_1, c/b).$$

The Bi-Conditional:

(D5.4) $$a//b \mathrel{\widehat{=}} c \leftrightarrow \text{(for any } a_1\text{: } a_1/a \leftrightarrow a_1, c/b \;\&\; a_1, b/c).$$

The Disjunction:

(D5.5) $$a//b \vee c \leftrightarrow \text{(for any } c_1\text{: } a/c_1 \leftrightarrow b/c_1 \;\&\; c/c_1).$$

The Negation[18]:

(D5.5) $$a//\neg b \leftrightarrow \text{(for any } a_1 \text{ and } b_1\text{: } a, a_1/b \rightarrow (a, a_1/b_1 \;\&\; a_1, b_1/b)).$$

[18] In this formula, the last occurrence of "b_1" may be omitted. It is added only in order to make a certain inherent symmetry (which is nothing but the symmetry between the laws of contradiction and excluded middle) more obvious. For an alternative definition, see ⟨footnote 20⟩.

While none of these definitions use a defined term in the definiens, we may add some
219 others which make use of defined terms: |

The Exclusive Disjunction ("*a-or-b-but not both*"):

(D5.7) $$a /\!/ a \neq b \;\leftrightarrow\; a /\!/ \neg(a \mathbin{\widehat{=}} b).$$

The Alternative Denial ("*not-a-or-not-b*"):

(D5.8) $$a /\!/ a \curlywedge b \;\leftrightarrow\; a /\!/ \neg(a \wedge b).$$

The Conjoint Denial ("*not-a-and-not-b*"):

(D5.9) $$a /\!/ a \downarrow b \;\leftrightarrow\; a /\!/ \neg(a \vee b).$$

The Tautology:

(D5.00) $$a /\!/ t(b) \;\leftrightarrow\; a /\!/ b \vee \neg b.$$

The Contradiction:

(D5.0) $$a /\!/ f(b) \;\leftrightarrow\; a /\!/ b \wedge \neg b.$$

For the last two, we can easily prove the rules:

(5.001) $$t(a) /\!/ t(b).$$

(5.01) $$f(a) /\!/ f(b).$$

That is to say, all tautologies are substitutionally equal, and the same holds for all contradictions. This shows that we may omit as irrelevant the reference to "a" or "b", and simply write "t" and "f" instead of "$t(a)$" and "$f(a)$".

In a way, all these definitions are quite unnecessary. True, they may replace the primitive rules of section 4, but we can work just as well or better with the much simpler primitive rules. Nothing is added to our system or taken away if we replace the rules of section 4 by the definitions of the present section.

The only reason why we nevertheless present these definitions here is that they establish beyond any doubt the fact that our primitive rules of section 5 have the character of definitions, even the rule for negation.

This fact is, no doubt, interesting. It may help to clear up the discussion of the so-called "alternative systems of logic". From our point of view, these systems are not alternative systems of logic, but alternative ways of using certain labels such as, for example, the label "negation". For it is plain that, by using different primitive rules, or by using different definitions, we can define different concepts. There is no problem involved here. If we introduce instead of 4.6 a similar primitive rule like

(5.06_{M}) $$a, b / \neg c \;\leftrightarrow\; c, b / \neg a,$$

then this rule defines a concept similar to our concept of negation, as defined by 4.6

(the "classical" concept of negation), but also a little different. This concept may be called "negation$_{\mathrm{M}}$" (it is, in the presence of the other rules except 4.6, identical with
220 the | negation of Johansson's "Minimum Calculus"). It so happens that negation$_{\mathrm{M}}$ is so closely related to our classical negation that all rules of inference based on negation$_{\mathrm{M}}$ are also valid for classical negation, but not *vice versa*. (In this sense negation$_{\mathrm{M}}$ can be said to be weaker than classical negation.) This is due to the simple fact that rule 5.06 is a valid secondary rule of our classical system. If we add to rule 5.06 another rule (which is not secondary to 5.06 but secondary to our classical system), *viz.*:

(5.06$_{\mathrm{I}}$) $$\neg a, \neg\neg a / a,$$

then these two rules together define another concept – we call it "negation$_{\mathrm{I}}$" – the intuitionist concept of negation.

All these concepts can co-exist, in one and the same model language, as long as they are distinguished.[19] Thus there is no quarrel between alternative systems. The rules of inference pertaining to the various concepts of negation are not identical, to be sure. But this is very satisfactorily explained by the simple fact that the various concepts have been given a meaning by rules which are not identical.[20]

The question whether ordinary spoken language uses the one or the other of these concepts does not, I believe, arise. Ordinary language is not sufficiently precise for us to ask such questions about it. As far as the somewhat more precise and therefore more artificial language of science is concerned, we are still asking too much, without doubt, if we ask such a question in connexion with it. But I suggest that there are good reasons why, for most purposes of science, the classical concept should be preferable to the others – simply because it is stronger, more explicit. This does not prevent us from using, for certain purposes, especially in parts of mathematics, the interesting concept negation$_{\mathrm{I}}$ side by side with the classical one.

The upshot of all these considerations is this: If we have an artificial model language with signs for conjunction, the conditional . . . etc. (we have called them "formative signs" of the language in question) then the meaning of these formative signs can be exhaustively determined by the rules of inference in which the signs

---

[19] It is, however, not always possible to distinguish them, or to prevent one from becoming absorbed or assimilated by another. If, for example, a language L contains classical negation and the conditional, then there exists, for every b and c, a statement $b > c$ such that 4.2e holds for every a (besides 4.2). But in the presence of 4.2e (and 4.2) all the rules which are valid for classical negation can be derived from those for intuitionist negation (as indicated on ⟨p. 133f⟩). This shows that if L contains definitions of both classical and intuitionist negation, the latter is unavoidably absorbed or assimilated by the former. ⟨Footnote added in the Errata.⟩

[20] We can use the following signs: "$\langle \neg_c\, a \rangle$" for the classical negation of a, "$\langle \neg_i\, a \rangle$" for the intuitionist negation of a. We can then introduce the following two explicit definitions, of which the first is a somewhat preferable alternative to (D5.6):

(D5.6c) $$a /\!/ \langle \neg_c\, b \rangle \leftrightarrow \text{(for any } b_1\text{: } a, b / b_1 \;\&\; (a, b_1 / b \rightarrow b_1 / b)).$$

(D5.6i) $$a /\!/ \langle \neg_i\, b \rangle \leftrightarrow \text{(for any } b_1\text{: } a, b / b_1 \;\&\; (b, b_1 / a \rightarrow b_1 / a)).$$

occur; this fact is established by defining our definitions of these formative signs
221 explicitly in terms of rules of inference. | There is good reason to believe that, in our ordinary language, we learn the intuitive meaning of formative signs in the same way; for no doubt, we learn their intuitive meaning by learning *how to use them*. In other words, we learn their meaning by learning the rules of their use, and these rules are mainly the rules of *their use in connexion with the drawing of inferences*.

Thus it is trivial that once we understand intuitively the meaning of formative signs, the proper handling of the rules of inference pertaining to them is part of this knowledge.

This is why rules of inference have that compelling character which mislead idealistic logicians to speak of laws of thought. Instead of taking this compulsion as the unanalysed and unanalysable *datum* of logic, we can indeed explain it quite easily. Once we know how to use the words "and", "provided", etc., that is to say, once we have learned the language and have got an intuitive grasp of its use, we feel compelled to admit that from a and $a > b$, b can be deduced; in exactly the same way as we feel compelled to admit that this animal here is a cow (provided it is a cow), once we have learned how to use the word "cow", and have got an intuitive grasp of its use.

6. Statement-Functions. Substitution as a Logical Operation

We now extend the scope of our inquiry. So far, we have considered only one kind of expressions, namely statements. Now we are going to consider, in addition to statements, so-called "statement-functions".

An example of a statement-function is: "He is a charming fellow". Whether this is true or false will depend on the question whom we mean by "he", or in other words, whose name we substitute for the pronoun[21] "he"; whether we substitute "Ernest", say, or "Edward". The expression "he is a charming fellow" is as such neither true nor false, it is *no statement*, but it can be converted into a statement by a fitting substitution – of a name, or of a descriptive phrase such as "Ernest's best friend". Other examples are: "This is an ugly fireplace"; "He is his younger brother"; "He owes him the legacy left to him by his father". The last two examples show the need for making distinctions: they become clearer if we write, say, "He_1 is the younger brother of he_2", and "He_1 owes he_2 the legacy left to he_3 by the father of he_4" – of course, something else may be meant, for example, "He_1 owes he_2 the legacy left to he_3 by the father of he_1", or "He_1 owes to he_2 the legacy left to he_2 by the father of he_3", etc.

222 | The need for a sufficiently long list of distinct pronouns is obvious, and we shall assume, of the languages under consideration, that they dispose of such a sufficient list. (See below, postulate PF3.)

Now consider again the last example. It is clear that we can *derive* from it the statement-function "He_1 owes to he_2 some legacy". Similarly, we can *derive* from the

[21] The excellent suggestion to treat name-variables as pronouns is, according to Quine ⟨Quine, 1940, p. 71⟩, due to Peano.

statement-function "He is not only a charming fellow but an excellent physician" the statement-function "He is an excellent physician". In other words, if our treatment of deducibility is to be as general as we intend it to be, then a treatment of the deducibility of statement-functions becomes necessary. At the same time, it will be necessary to treat combinations of statements and statement-functions.

We shall, in fact, assume that the theory of derivation developed in previous sections holds not only for statements but also for statement-functions.

Although statements and statement-functions should be clearly distinguished, they are, of course, rather similar things. A statement-function is, as it were, a statement with certain holes or openings in it, which can be filled up or closed up by inserting names. For this reason, statement-functions are often called "*open statements*", and statements proper are often called "*closed statements*". The terminology has great advantages, and we shall adopt it; not, however, without issuing a warning that the term "open statement" must not mislead anybody into believing that open statements are statements in the ordinary sense. They are not, for they are neither true nor false. "Open statement" is only an alternative, and often a more convenient, term for "statement-function". Only closed statements are statements proper.

Since we do not want to say that open statements are statements, we shall use another term – "formulae" – to describe the class of expressions which consists of open statements as well as of closed statements. Thus we shall say that every open statement is a formula, and that every closed statement is a formula.

But apart from formulae, we have another class of expressions to consider – the pronouns "he", "he$_1$", "he$_2$" ... "it", etc. We shall call these pronouns the *name-variables* under consideration, since names may be substituted for them.

It is obvious that these two classes of expressions, *formulae* and *pronouns or name-variables* have no member in common. This will be expressed in our postulate PF1.

223 | Thus our universe of discourse will consist, from now on, of two entirely distinct classes of expressions – formulae, F, and name-variables, N. We shall introduce the convention to use from now on "a"; "b"; "a_1"; "b_1" ... as variable names of formulae rather than as variable names of closed statements only (as before), and furthermore, to use "x"; "y"; "z"; "u"; "w"; etc., as variable names of pronouns. It so happens that "x"; "y"; etc., can be described as variable names of name-variables. This is perhaps a little confusing, for a moment or two, but quite in order.

The symbols "x"; "y"; etc., are variables which refer to certain symbols such as "he"; "he$_1$"; "it"; "it$_1$"; etc., of the languages under consideration. The symbols "he", "he$_1$", etc., refer in their turn to certain men – to Ernest, to Edward, etc. These *men* – not perhaps their *names* – constitute the "realm of individuals" or the "universe of individuals" to which the language L under consideration refers. This universe of the discourse of L is not the universe of *our* discourse; *we*, rather, discuss languages – L, for example – and especially certain classes of expressions.

The name-variables of L may occur in the formulae of L in two ways: either as *free variables* or as *bound variables*.

In the open statements which we used as examples, all the variables occurred freely. But we can *bind* a freely occurring variable by placing before the formula in which it occurs the phrase

(A) "Whosoever he_1 may be . . ."

or the phrase

(E) "There exists at least one he_1 such that . . ."

The phrase (A) is called the "universal quantifier (binding he_1)", the phrase (E) is called "the existential quantifier (binding he_1)".

The theory of quantifiers will be sketched in the next section; in the present section, we deal merely with formulae and variables in general. But it may be mentioned that, if all variables of a function are bound, the function is a closed statement, *i.e.*, a statement proper. Or in other words, an open statement may give rise to a closed one in either of the two following ways: by substituting names for the free variables, or by binding the free variables with the help of quantifiers. For example, the open statement "If he is a good physician, then he is a good man" may give rise to the statements:

(1) "If Ernest is a good physician then he is a good man."

224 (2) (A) "Whosoever he may be – if he is a good physician then | he is a good man." (This may be said to be the same as "All good physicians are good men": a universal statement.)

(2) (E) "There exists some he such that if he is a good physician then he is a good man." (This may be interpreted as: "There exists somebody who, if he is a good physician, is also a good man": an existential statement.)

We said before that the quantifiers may be placed before a function; but in another language (or in the same), they may be placed after the function, etc. In other words, the same problems arise as those we discussed in section 3 in connexion with conjunction, and the position of the sign "and". We shall, therefore, not assume that all the languages we wish to discuss contain distinguishable signs which we can call quantifiers. What we shall discuss, in the next section, is merely the way in which the formulae, denoted by

$$a$$

and by

$$Axa$$

are related, if Axa is equivalent to the result of applying universal quantification (with regard to the variable x) to the formula a.

The sign "Ax" has therefore no more independent meaning for us than the sign "$\wedge$"; nevertheless, we shall loosely speak of "Ax" as of the name of a universal quantifier binding x, and of "Ex" as of the name of an existential quantifier binding x.

Leaving the quantifiers and the rules of inference which are to determine their meaning to the next section, we now turn to the theory of another and even more fundamental logical operation, also comparable to conjunction, *viz.*, the *operation of substituting the variable y for the variable x in the formula a*. The intuitive idea of substituting the variable "he_1" for the variable "he_2" in the formula "he_1 loved he_2" or "he_1 killed he_2", and the difference such a substitution should make to the meaning of a formula, seems to be perfectly clear. In fact, this idea seems to be so obvious that it is usually accepted as a kind of descriptive term that is not in need (and perhaps not capable of) logical analysis – least of all, of an analysis in terms of derivability.

But the operation of substitution plays a very important part in quantification theory, and an attempt to reduce it to the concept of derivability seems highly desirable if we wish to build up a coherent theory.

For this purpose we first introduce the symbol "$a\binom{x}{y}$" as a convenient name for the formula which is the *result* of the substitution of the variable y for the variable
225 x in the formula a, in | all places of the occurrence of x. (If x does not occur in a, then $a\binom{x}{y}$ is the same as a.) As in the case of the compounds, we are not going to characterise this result itself, but rather its *logical force* – *i.e.*, any formula which is equivalent to the result.

To this end, we first postulate, as a preparation, PF1, PF2, and PF3.

(PF1) If L is one of the languages under consideration, then the formulae of L and the name-variables of L have no member in common.

(PF2) $a\binom{x}{y}$ is a formula (of L) if, and only if, a is a formula (of L), and x and y are name-variables (of L).

(PF3) If a is a formula then there exists at least one name-variable x such that $a//a\binom{x}{y}$ for any name-variable y.

The contents of the first two of these postulates are plain. The third has mainly the function of ensuring that, for every formula (especially for every open statement), there exists at least one variable which does not occur in it.[22] (We mentioned before that, if x does not occur in a, then $a\binom{x}{y}$ is the same as a.)

We now proceed to the primitive rules of derivation which lay down the precise meaning in which the term "$a\binom{x}{y}$" is used within the theory of inference. These rules look complicated, but this is only due to their notation: their content is trivial indeed.

(6.1) If, for every z, $a//a\binom{y}{z}$ and $b//b\binom{y}{z}$, then $a//b \rightarrow a\binom{x}{y}//b\binom{x}{y}$,

that is, in brief: equal substitutions in equal formulae have equal results. This rule assures the continued substitutivity of equivalence in the theory of statement-functions and of quantification.

(6.2) $$a\binom{x}{x}//a.$$

This is plain.

(6.3) If $x \neq y$, then $\left(a\binom{x}{y}\right)\binom{x}{z} //a\binom{x}{y}$,

that is: once we have substituted y for x, then x no longer occurs (provided $y \neq x$), and further substitutions for x have, therefore, no effect.

(6.4) $$\left(a\binom{x}{y}\right)\binom{y}{z}//\left(a\binom{x}{z}\right)\binom{y}{z},$$

226 | that is: if we substitute first y for x and then z for y, then the result is the same as if we had substituted at once z for x and, besides, z for y.

---

[22] Or, at least, upon which it does not depend. ⟨Footnote added in the Errata.⟩

(6.5) $$(a\tbinom{x}{y})\tbinom{z}{y} /\!/ (a\tbinom{z}{y})\tbinom{x}{y},$$

that is, if we substitute for two variables x and z, the same variable y, then it is irrelevant which of these two substitutions we undertake first.

(6.6) If $w \neq x$; $x \neq u$; and $u \neq y$, then $(a\tbinom{x}{y})\tbinom{u}{w} /\!/ (a\tbinom{u}{w})\tbinom{x}{y}$.

This says that, if we substitute for two *different* variables, x and u, two other variables, y and w, then it is irrelevant which of the two substitutions we undertake first – provided always that the variable to be substituted for x does not coincide with u, and *vice versa*.

These six primitive rules determine the meaning of the symbol "$a\tbinom{x}{y}$" in a way precisely analogous to the way in which, say, rules 3.1 to 3.3 determine the meaning of conjunction. And the meaning of "$a\tbinom{x}{y}$" is determined, precisely as was that of conjunction, with the help of the concept of derivability, "/".

To give a few examples of secondary rules: we can obtain first, from postulate PF3

(6.01) If, for any x and y, $a\tbinom{x}{y} /\!/ b\tbinom{x}{y}$, then $a /\!/ b$.

From this and 6.1, we obtain

(6.02) If, for every z, $a /\!/ a\tbinom{y}{z}$ and $b /\!/ b\tbinom{y}{z}$, then
$$a /\!/ b \leftrightarrow (\text{for every } x\text{: } a\tbinom{x}{y} /\!/ b\tbinom{x}{y}).$$

Another example, secondary to 6.2; 6.4; 6.5, is:

(6.03) $$(a\tbinom{x}{y})\tbinom{y}{x} /\!/ a\tbinom{y}{x}.$$

Quite important is the fact that, with the help of "$a\tbinom{x}{y}$", we can define a new concept – the non-occurrence of x in a, or, perhaps more precisely, the non-relevant occurrence of x in a (or the conception: "*a-does-not-depend-on-x*"). We denote this idea by "$a_{\grave{x}}$" (read: "*a-without-x*") and define it as follows:

(D6.1) $$a /\!/ a_{\grave{x}} \leftrightarrow (\text{for any } y\text{: } a /\!/ a\tbinom{x}{y}).$$

227 | That is: a is equivalent to $\langle a \rangle$*-without-x* if, and only if, a is equivalent to whatever may be the result of substituting in $\langle a \rangle$ for x.

If we neglect this concept then the result is, as a rule, that some formulae can be proved in the systems in question which, in effect, amount to the assertion that there exists only *one* individual, in the realm of individuals to which the name-variables of L refer, that is in the universe of discourse of the language under consideration. This result is highly undesirable, and in order to make explicit that such a result is frankly contradictory to our intentions, I suggest including in our list of postulates for the logic of functions the following:[e]

---

[e] Cp. the remark by Bernays in a letter to Popper (this volume, § 21.7, end of page 4) and Popper's reply (this volume, § 21.8).

(PF4) In every language under consideration, there exists at least one a, one x, and one y, such that[23]

$$a/a\binom{x}{y} \rightarrow t/f.$$

Since "t/f" can be shown to be contradictory, this postulate has the result that a theory in which the rule

$$a/a\binom{x}{y} \tag{1}$$

or even

$$a_{\grave{y}}/a_{\grave{y}}\binom{x}{y} \tag{2}$$

holds becomes frankly contradictory. But these rules (1) or (2) are just the rules asserting in effect that there exists only *one* individual. For they assert that *all* formulae such as "he_1 loves he_2" are equivalent to "he_1 loves he_1"; and this can be so only if "he_1 is not identical with he_2" is equivalent to "he_1 is not identical with he_1", which is only the case if there exists only one he.

This indicates that we may be able to express *identity and difference of individuals* with the help of the means at our disposal; and although we cannot, of course, define the sign of identity in L (there may be no such sign in L), we can, again, as in the case of conjunction, define the logical force of any statement-function expressing identity between the variables x and y. If we denote such a function by "$Idt(x, y)$" we can define:[24]

(D6.2) $a /\!/ Idt(x, y) \leftrightarrow$ (for every b and z: $((b /\!/ b_{\grave{x}}$ & $b /\!/ b_{\grave{y}}) \rightarrow a, b\binom{z}{x}/b\binom{z}{y})$ &
$(($for every c and u: $((c /\!/ c_{\grave{x}}$ & $c /\!/ c_{\grave{y}}) \rightarrow b, c\binom{u}{x}/c\binom{u}{y}))) \rightarrow b/a))$

Difference is easy to define:

(D6.3) $$a /\!/ Dff(x, y) \leftrightarrow a /\!/ \neg Idt(x, y).$$

This shows that these two concepts are logical concepts, and that the corresponding signs – if any – of the languages under consideration are formative signs.

---

[23] Here we could write, more simply, "b/c" instead of "t/f". ⟨Footnote added in the Errata.⟩

[24] To be exact, we have, of course, to add the following postulate as a preliminary to the definition. (A similar postulate would have to be added for difference.)

(P *Idt*) If x and y are name variables, then $Idt(x, y)$ is a formula.

Furthermore, our defining rules of substitution should be extended by introducing such rules as:

228 (A) $$(Idt(x, y))\binom{x}{z} /\!/ Idt(z, y).$$ |

(B) $$(Idt(x, y))\binom{y}{z} /\!/ Idt(x, z).$$

(C) If $x \neq u \neq y$, then $Idt(x, y)\binom{u}{z} /\!/ Idt(x, y)$.

7. Quantification

The theory of quantification is too subtle to be analysed in detail within the framework of this paper. All I shall attempt is to show that our two rules, 7.1 and 7.2, are not more complicated than previous ones, and that they can be considered as *definitions of universal and existential quantifications*.

Our two primitive rules are[25]

(7.1) If $b\binom{x}{y}$ is a formula, then: $a_{\hat{y}}/Axb_{\hat{y}} \leftrightarrow a_{\hat{y}}/b_{\hat{y}}\binom{x}{y}$

or in a different way of writing:

If a and $b\binom{x}{y}$ are formulae, then: $a/Axb \leftrightarrow a/b\binom{x}{y} \quad (a = a_{\hat{y}};\ b = b_{\hat{y}})$

(7.2) If $a\binom{x}{y}$ and b are formulae, then: $Exa_{\hat{y}}/b_{\hat{y}} \leftrightarrow a_{\hat{y}}\binom{x}{y}/b_{\hat{y}}$

or in alternative formulation:

$$Exa/b \leftrightarrow a\binom{x}{y}/b \qquad (a = a_{\hat{y}};\ b = b_{\hat{y}}).$$

If we wish, we can transform these rules into explicit definitions:

229 (D7.1) $a_{\hat{y}}/\!/Axb_{\hat{y}} \leftrightarrow$ (for every $c_{\hat{y}}$: $c_{\hat{y}}/a_{\hat{y}} \leftrightarrow c_{\hat{y}}/b_{\hat{y}}\binom{x}{y}$)

(D7.2) $a_{\hat{y}}/\!/Exb_{\hat{y}} \leftrightarrow$ (for every $c_{\hat{y}}$: $a_{\hat{y}}/c_{\hat{y}} \leftrightarrow b_{\hat{y}}\binom{x}{y}/c_{\hat{y}}$).

The definitions are, as in previous cases, no improvement – they only establish the fact that quantification can be defined directly on the basis provided by sections 2 or 3, and the primitive rules 6.1 to 6.6.

I do not claim that the two new primitive rules are trivial in the sense in which our basic rules of sections 2 and 3, or even the rules of section 4, are trivial. There is, without doubt, a certain subtlety involved in the way these rules make use of the concepts "$a\binom{x}{y}$" and "$a_{\hat{x}}$"[26] defined in section 6, – a subtlety which defies a simple

---

[25] Postulates and rules of substitution have also to be assumed, corresponding to those mentioned in the last footnote.

[26] The use of "$a_{\hat{x}}$", however, is avoidable in all the primitive rules and definitions; *i.e.*, we can write, more lengthily but more precisely and clearly ($x \neq y$ must always be assumed here):

(7.1) $a\binom{y}{x}/Ax(b\binom{y}{x}) \leftrightarrow a\binom{y}{x}/b\binom{x}{y}$

(7.2) $Ex(a\binom{y}{x})/b\binom{y}{x} \leftrightarrow a\binom{x}{y}/b\binom{y}{x}$

(D7.1) $a\binom{y}{x}/\!/Ax(b\binom{y}{x}) \leftrightarrow$ (for every c: $c\binom{y}{x}/a\binom{y}{x} \leftrightarrow c\binom{y}{x}/b\binom{x}{y}$)

(D7.2) $a\binom{y}{x}/\!/Ex(b\binom{y}{x}) \leftrightarrow$ (for every c: $a\binom{y}{x}/c\binom{y}{x} \leftrightarrow b\binom{x}{y}/c\binom{y}{x}$)

Of the "rules of substitution" referred to in the last note we may state explicitly:

$$(Axa)\binom{x}{y}/\!/Ay(a\binom{x}{y});$$

explanation. Yet anybody who is conversant with quantification theory will, I am sure, be astonished that rules of this comparative degree of simplicity – and, for those who are accustomed to this kind of thing, even triviality – are sufficient. For the hitherto known comparable systems of quantification theory are much more complicated. For example, the well-known system of Hilbert and Ackermann contains (in place of our two rules) two axioms; two rules rather similar to our own; four very complicated rules for the handling of variables; and at least one rule forbidding certain operations (which are permitted in our system). All these rules, except the last mentioned, can be shown to be secondary to our two rules. Or to take another system – the very interesting and suggestive one of Quine's *Mathematical Logic*. This system takes only the universal quantifier as primitive. (The existential quantifier is introduced by a well-known definition with the help of the universal quantifier and negation.) Quine's system corresponds therefore to our *one* rule 7.1. But Quine needs five axioms – partly rather complicated ones – which can, of course, all be shown to be secondary to our rule 7.1.

Among the secondary rules which can be shown to be valid in every adequate theory of quantification, there are extremely complicated ones – here logic definitely ceases to be trivial. But it is our contention that these complications are merely due to the wealth of possible iterations and combinations of essentially trivial rules. For even if our latest defining rules are no longer very trivial, they are only definitions, that is to say, nothing but abbreviations; and they all go back to one specific logical concept – the concept of derivability, "/", introduced by the undoubtedly trivial rules of sections 2 and 3.

8. Derivation and Demonstration

230 | If we omit from our system the definitions of tautology and contradiction, D5.0 and D5.00, then our whole system is one of *purely derivational logic*. By this we mean the following. Take any language in which ordinary descriptive statements occur, such as, in English, statements like "If it is raining then the streets are slippery" whose truth or falsity cannot be decided by merely logical considerations (in contra-distinction to the tautology "If it is raining then it is raining"). Call these descriptive statements "non-logical statements". Then we can define:

(D8.1) A system of primitive rules of derivation is called "purely derivational" if, and only if, we can give such examples for each primitive rule that all of the statements (whether components or compounds) which occur in the examples are non-logical.

The point of a purely derivational logic is this: it is a system intended from the

---

and:

$$\text{If } x \neq y \text{, then } (Axa)\binom{y}{z} /\!/ Ax(a\binom{y}{z}).$$

Corresponding rules of substitution hold for existential quantification. ⟨Footnote added in the Errata.⟩

start to be a theory of inference in the sense that it allows us to derive from certain informative (non-logical) statements other informative statements.

Most systems of modern logic are not purely derivational, and some (for example in the case of Hilbert Ackermann) are not derivational at all. They operate not so much with rules of inference as with axioms or with rules of proof (axiom schemata). That is to say, they take as primitive such assertions as “All statements designated by ‘$a > a$’ are true” (or “are provable” or “follow from the empty class of premises”, etc.; these are just so many ways of stating axiom schemata).

These procedures are in themselves unexceptionable. But they are liable to blur the distinction between derivation and proof; they are liable to create the impression that every logical proof is a derivation whose first premises are logical axioms.

But this impression is definitely wrong. It is true that a derivation whose premises are logical axioms is a proof. But this is so only (1) because every derivation whose premises are provable is a proof, and (2) because the so-called logical axioms are in fact provable.

Let us consider for a moment one of the intuitive ways which we use in proofs, – say in an indirect proof.

Assume that we succeed in deriving, from one premise a

$$a/b \tag{1}$$

$$a/\neg b \tag{2}$$

231 | Then we argue in this way: a must be absurd, for from it follows b as well as $\neg b$, and therefore $b \wedge \neg b$, which is certainly false. Thus a is *logically refuted*; but its negation, $\neg a$, must be true. Thus we have proved $\neg a$.

If we look at this argument, then we see that it refers to two derivations, (1) and (2); and that it argues from derivability to refutability and demonstrability.

In this way, *all proofs refer to derivations*; the derivations are the most conspicuous parts in proofs; but the derivation is not the proof.

Take another example.

If we find that the following derivations are valid

$$a/b \tag{1}$$

$$\neg a/b \tag{2}$$

then we say that we have proved b. For we say, one of the two premises, a or $\neg a$, must be true; and if both the derivations are valid, the truth of the true premise must be transmitted to the conclusion. Thus b must be true – we have proved b.

Now what these proof schemata have in common is this: They use derivation.

But while *derivation* only establishes that the conclusion is true *provided* that the premises are true, the proof seeks to establish the truth of its conclusion independently of the question whether the premises are true or false. It tries to establish the truth of the conclusion unconditionally.

But is such a thing possible? Yes, it is possible, although only for comparatively rather uninformative statements – for what we may call “logical truisms”.

We can now give a definition of a logical proof or demonstration:

(D8.2) The statement a is demonstrable $\leftrightarrow$ the statement a can be validly derived from any premise whatsoever.

This means, indeed, that the statement can be shown, by means of derivations, to be true unconditionally – independently of any particular premise. (If we derive it from b and doubt the truth of b, we can at once derive it from $\neg b$ also, or from c, etc.)

We shall write

$$\vdash a$$

for "the statement a is demonstrable". Then D8.2 becomes:

(D8.2+) $\vdash a \leftrightarrow$ (for any statement b: b/a);

we can, similarly, define refutability, writing

$$\dashv a$$

for "a is refutable":

(D8.3) $\dashv a \leftrightarrow$ (for any statement b: a/b).

232 Now on the basis of our purely derivational logic, it is possible to show strictly: |

(8.1) All the so-called axioms of the logic of propositions and propositional functions are provable; for example, those of *Principia Mathematica*, etc.

(8.2) Every statement that can be validly derived from a demonstrable statement is demonstrable, *i.e.*,

$$(\vdash a \; \& \; a/b) \vdash b.$$

(8.3) If we add demonstrable statements to the premises of any derivation, then the validity or invalidity of the derivation remains unaffected.

All these are principles which have been more or less intuitively used so far, but which can be strictly established as consequences of our definition of proof.

But there are further results. For example we can easily show

(8.4) $a/b \rightarrow (\vdash a \rightarrow \; \vdash b)$,

which is only another form of 8.2; *but at the same time, the converse is not true*, that is to say, we must be careful not to assert that

$$(\vdash a \rightarrow \; \vdash b) \rightarrow a/b.$$

(Take a to be non-logical and b to be contradictory: this is a counterexample.)

This warning is very important, for there are valid *rules of proof* of the theory of statement-functions such that

(8.5) $\vdash a_{\hat{y}} \leftrightarrow \; \vdash a_{\hat{y}}\binom{x}{y}$

for which the corresponding *rule of derivation*

$$a_{\grave{y}}/a_{\grave{y}}\binom{x}{y}$$

is invalid.[27] Among the valid rules of proof are the important rules

$$a_{\grave{x}}/b \rightarrow a_{\grave{x}}/Axb. \tag{8.9}$$

$$a_{\grave{y}}/b_{\grave{y}}\binom{x}{y} \rightarrow a_{\grave{y}}/Axb. \tag{8.91}$$

Now these – for example 8.9 – may be transformed in this way:

$$\vdash a_{\grave{x}} > b \rightarrow \ \vdash a_{\grave{x}} > Axb. \tag{8.92}$$

But they must not be transformed in this way:[28]

$$a_{\grave{x}} > b/a_{\grave{x}} > Axb,$$

233 | because otherwise

$$\vdash a > Axa$$

becomes provable, and with it,

$$a//a\binom{x}{y}$$

could be shown to be valid – which violates our postulate PF4, of section 6.

Now all the mistakes here warned against do actually vitiate some otherwise very excellent books on modern logic – an indication that the distinction between (conditional) rules of proof or rules of demonstration on the one side and rules of derivation on the other cannot be neglected without involving oneself in contradictions.

It may be mentioned, in passing, that derivational logic *must* operate with a theory of open sentence functions, while demonstrational logic may neglect this part of logic.

We may perhaps sum up by saying that a purely derivational logic can get over an awkward dualism which demonstrational logic cannot avoid. Demonstrational logic must start with axioms – at least one – and rules of inferences – at least something like the *modus ponens*. As opposed to this, derivational logic operates monistically, *i.e.*, with rules of inference alone. The so-called axioms of the various systems of the logic of demonstration can be proved without assuming one of them.

9. Metalanguage and Object Language

Why do we study logic? We all understand our language, which means that we all know how to draw inferences, as we have seen. We do not need logical theory for

---

[27] *Cp.*, as opposed to this, Carnap, *Formalization of Logic* ⟨Carnap, 1943⟩, p. 137, D. 28–2, rule 10.

[28] *Cp.* Carnap, *ibid.*, p. 138, rule 11; this is undoubtedly suggested by Hilbert-Bernays' way of writing these rules in the same style as the *modus ponens* (*cp.* pp. 105f.) which suggests that they are rules of derivation. Hilbert Ackermann, on the other hand, write p. 57 the *modus ponens* as if it were a conditional rule of proof.

being able to draw inferences; nor does the study of logical theory lead to a method of making all inferences fool-proof.

There are two obvious reasons for studying logic. One is curiosity, theoretical interest in languages and ⟨in⟩ their rules of use. This needs no apology. The other is practical. Our naturally grown languages have not been designed for the use to which we put them in scientific and mathematical investigations. They definitely do not always stand up too well to the strain of modern civilisation: paradoxes arise, spurious theories; the conclusiveness of some of our subtle arguments becomes doubtful. Such practical needs lead us to study the rules of language in order to design an instrument fit for use in science.

But for the study of language, we have to *use* language; and we have to use it freely and fully, without being afraid of using it. There is no study of any object whatever
234 without the full | use of language if we need it. As the zoologist uses language to describe the behaviour of reptiles or lions, so the student of language must use language in order to describe the behaviour of statements, formulae, variables, etc.

But is there not a danger that our studies may be vitiated if we use and study language at the same time? Undoubtedly there is. But there exists a simple device which seems to get over these difficulties beautifully. It is, simply, always to distinguish the language which is the object of our studies (the "object language") from the language we are using. The latter, if used for the discussion of other languages, is called the "metalanguage".

The fundamental idea of the distinction between object language and metalanguage is this: we simply cannot use *and* discuss one and the same language at the same time. Thus we decide to acknowledge the impossibility, and *we do not worry too much about the language we are using* – as long as we are using it. If we feel doubtful whether our way of using it was quite in order, then nothing will prevent us from proceeding, later, to an investigation of our metalanguage, using thereby what has been called a "meta-metalanguage" (etc., as long as we like to go on).

I mentioned that we should use the metalanguage freely. By this I mean that we must not be deterred, for example, from using conjunction in the metalanguage when defining conjunction in the object language. We *must* speak, if we wish to define. It is impossible to speak properly in our language of ordinary use without using conjunction, the conditional, and indeed all its formative signs. Here is no vicious circularity: we do not attempt to define the words we are using, but we use words in order to define very general concepts referring to other languages.

The demand that, in order to avoid circularity, we should avoid, in defining signs of the object language, the use of formative signs in the metalanguage, is no better than the impossible demand that we should avoid the use of statements in the metalanguage; for in most languages we can have no statements at all if we avoid formative signs. Incidentally we can, in our definitions, avoid even the appearance of circularity – we can for example define negation without making use of negation in the metalanguage, and we can define "$a \wedge b$" without using "&" in the metalanguage, and "$a \vee b$" without using "or".

As to our own investigations, we find, if we proceed to study our metalanguage in
235 the meta-metalanguage, that our meta-language | can be easily formalised. It turns

out that we do not need more than positive logic, including quantifiers (if we exclude the postulate PF4 which need not be assumed). Thus it happened that we defined negation, without using it in the meta-language (incidentally and without intending it). This fact may be interesting for some reason or other; but had we to make use of negation in the metalanguage, in order to define what we mean by negation in the various object languages we were then studying, there would have been no circularity involved.

References

Carnap, R. (1943). *Formalization of Logic*. Cambridge, Mass.: Harvard University Press.

Hertz, P. (1923). Über Axiomensysteme für beliebige Satzsysteme. Teil II. Sätze höheren Grades. In: *Mathematische Annalen* 89, pp. 76–102.

– (1931). Vom Wesen des Logischen, insbesondere der Bedeutung des modus barbara. In: *Erkenntnis* 2, pp. 369–392.

Jeffreys, H. (1942). Does a Contradiction Entail every Proposition? In: *Mind* 51 (201), pp. 90–91.

Johansson, I. (1937). Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus. In: *Compositio Mathematica* 4, pp. 119–136.

Popper, K. R. (1946c). Why Are the Calculuses of Logic and Arithmetic Applicable to Reality? In: *Logic and Reality*. Aristotelian Society, Supplementary Volume 20, pp. 40–60. Reprinted in Popper (2004).

– (1947b). Logic without Assumptions. In: *Proceedings of the Aristotelian Society* 47, pp. 251–292. Reprinted as Chapter 2 of this volume.

– (1947c). New Foundations for Logic. In: *Mind* 56 (223), pp. 193–235. Reprinted as Chapter 3 of this volume.

– (1948e). Corrections and Additions to “New Foundations for Logic”. In: *Mind* 57 (225), pp. 69–70.

– (2004). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.

Quine, W. V. O. (1940). *Mathematical Logic*. 1st ed. New York: W. W. Norton & Company.

Tarski, A. (1930a). Fundamentale Begriffe der Metholodogie der deduktiven Wissenschaften. I. In: *Monatshefte für Mathematik und Physik* 37, pp. 361–404. English translation in Tarski (1956).

– (1930b). Über einige fundamentale Begriffe der Metamathematik. In: *Comptes rendus des Séances de la Société des Sciences et des Lettes de Varsovie, Classe III*. Vol. 23, pp. 22–29. English translation in Tarski (1956).

– (1936a). Grundzüge des Systemenkalküls. Zweiter Teil. In: *Fundamenta Mathematicae* 26, pp. 283–301.

– (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.