

## ON THE THEORY OF TYPES<sup>1</sup>

W. V. QUINE

In this paper the theory of logical types will be examined, and certain departures from it will be suggested. Though the purpose of the paper is not primarily expository, an approach has been possible which presupposes no familiarity with special literature. Matters at variance with such an approach have been confined to appendices and footnotes.

In the early pages the logical paradoxes will be considered—an infinite series of them, of which Russell's paradox is the first. Then Russell's simple theory of types will be formulated, in adaptation to a minimal set of logical primitives: inclusion and abstraction. Two aspects of the theory will be distinguished: an *ontological doctrine* and a *formal restriction*. It will be found that by repudiating the former we can avoid certain unnatural effects of the type theory—notably the reduplication of logical constants from type to type, and the apparent dependence of finite arithmetic upon an axiom of infinity. But the formal restriction itself has unnatural effects, which survive, even in an aggravated form, after the type ontology has been dropped. A liberalization of the formal restriction will be proposed which removes the more irksome of these anomalies.

**1. Basic formal concepts.** Use will be made of the following logical notions and notations.

*Membership:* " $(x \in y)$ " means that  $x$  is a member of the class  $y$ . (In this and the ensuing notations, parentheses will be suppressed when there is no risk of ambiguity.)

*Inclusion:* " $(x \subset y)$ " means that the class  $x$  is included in the class  $y$ ; i.e., that every member of  $x$  is a member of  $y$ .

*Identity:* " $(x = y)$ " means that  $x$  and  $y$  are the same object.

*Conditional, Biconditional, Conjunction, Denial:* " $(\dots \supset \dots)$ ", " $(\dots \equiv \dots)$ ", " $(\dots \wedge \dots)$ ", and " $\sim \dots$ " mean respectively "If ... then ...", "... if and only if ...", "... and ...", and "It is not the case that ...", where the blanks are filled by any statements.

*Abstraction:* " $\hat{x} \dots$ " denotes the class whose members are just the objects  $x$  satisfying the condition "...". (The blank is filled by any statement, ordinarily one containing " $x$ ".)

*Universal quantification:* " $(x) \dots$ " means that the condition "... is satisfied by all values of " $x$ ".

*Universal class:*  $V$  is the class to which everything belongs.

*Null class:*  $\Lambda$  is the class to which nothing belongs.

*Unit class:*  $\iota x$  is the class whose sole member is  $x$ .

---

Received March 19, 1938.

<sup>1</sup> The main ideas of this paper were presented in an address before the mathematical fraternity Pi Mu Epsilon and the New York University Philosophical Society at their annual joint meeting in New York, February 24, 1938.

It is well known that these notions are sufficient for mathematical logic and indeed for mathematics generally. The further mathematical and logical notions are constructible from this basis by definition. The above list is in fact much longer than necessary, for various notions of the list are definable in terms of the remainder. Thus " $\forall$ ", " $\wedge$ ", and " $xy$ " are definable as abbreviations respectively of " $\hat{x}(x=x)$ ", " $\hat{x}\sim(x=x)$ ", and " $\hat{x}x\subset y$ "; and " $\hat{x}$ " is definable in turn as an abbreviation of " $\hat{y}(y=x)$ ". Further definitions are possible, until finally the twelve items of the list are reduced to just two: *inclusion* and *abstraction*.<sup>2</sup> Every mathematical or logical term (noun) or formula (statement) thus becomes a definitional abbreviation of a term or formula which is built up from variables merely by alternating the devices of inclusion and abstraction in the following fashion: Variables, which are our simplest *terms*, are joined by the inclusion notation to make a *formula*; from this formula a new *term* is made by abstraction, i.e., by prefixing a variable bearing a circumflex accent; two such terms, in turn, or one such term and one variable, are then joined by the inclusion notation to make a new *formula*; and so on.

"Term" and "formula", in this sense, are more rigorously describable with help of a rudimentary metamathematical symbolism. Let us use Greek letters (other than " $\iota$ ", " $\epsilon$ ") to denote any unspecified *expressions*. Then let us write " $\zeta I \eta$ " to denote the *inclusion compound* of  $\zeta$  and  $\eta$ . That is, " $\zeta I \eta$ " denotes the expression formed by putting  $\zeta$  and  $\eta$ , whatever they may be, in the respective blanks of " $( \subset )$ ". E.g., where  $\zeta$  is " $\hat{x}(x\subset x)$ " and  $\eta$  is " $y$ ",  $\zeta I \eta$  is " $(\hat{x}(x\subset x) \subset y)$ ". Finally, let us write " $\zeta_\alpha$ " to denote the *abstract* of  $\zeta$  with respect to  $\alpha$ . That is, " $\zeta_\alpha$ " denotes the expression formed by applying a circumflex accent to the expression  $\alpha$ , whatever it may be, and prefixing this to the expression  $\zeta$ . E.g., where  $\alpha$  is " $x$ " and  $\zeta$  is " $(x\subset y)$ ",  $\zeta_\alpha$  is " $\hat{x}(x\subset y)$ ". These notations involving Greek letters do not form part of the notation of logic, but aid us merely in talking about the notation of logic.

Now the *terms* are describable recursively thus:

(I) *Variables* (letters " $x$ ", " $y$ ", ...) are terms, and if  $\zeta$  and  $\eta$  are terms and  $\alpha$  is a variable then  $(\zeta I \eta)_\alpha$  is a term.

"Term" is of course to be construed in the narrowest way conformable to (I). In other words, the terms constitute the smallest class which embraces all variables and all expressions  $(\zeta I \eta)_\alpha$  such that  $\zeta$  and  $\eta$  are in the class and  $\alpha$  is a variable.

Finally, the *formulae* are directly describable thus:

(II) *Formulae* are the expressions  $\zeta I \eta$  such that  $\zeta$  and  $\eta$  are terms.

**2. The paradoxes.** It is clear from the explanation of abstraction that any statement of the form

$$(1) \quad \text{"}(\hat{x}\hat{x}\dots x\dots) \equiv \dots z\dots\text{"}$$

<sup>2</sup> See Quine [11], pp. 145-147. (Bracketed numerals refer to listings at the end of the present paper.) In what follows, use will be made of the fact that definitional reduction to inclusion and abstraction is possible; but familiarity with the actual definitions will not be presupposed, nor indeed any acquaintance with [11].

should be true, where "... $x$ ..." is any statement about  $x$  and "... $z$ ..." is the result of substituting " $z$ " for " $x$ " therein. But it is equally clear that any statement of the form

$$(2) \quad \text{“} \text{---} \equiv \sim \text{---} \text{”}$$

is a self-contradiction and must be rejected as false. Next let "K" be short for " $\hat{x}\sim(x\epsilon x)$ ". Now

$$(3) \quad \text{“}(K\epsilon K) \equiv \sim(K\epsilon K)\text{”}$$

is of the form (2), and should hence be false; yet it is the same as

$$\text{“}(K\epsilon \hat{x}\sim(x\epsilon x)) \equiv \sim(K\epsilon K)\text{”},$$

which is of the form (1) and should hence be true.

This difficulty, known as *Russell's paradox*, arises in precisely similar fashion when we confine ourselves to the primitive notation of inclusion and abstraction. For, the signs " $\equiv$ ", " $\sim$ ", and " $\epsilon$ " used in the above account are merely abbreviations of our primitive notation, according to the series of definitions alluded to in section 1. The paradox then shows that our simple grammatical rule, whereby inclusion and abstraction are applied alternately to produce terms and formulae, is too liberal; it enables us to get a freak combination such as (3), or rather the formula whereof (3) is a definitional abbreviation. This formula is a freak in that it illustrates a form which should be false in all cases, according to the meanings of our signs, and at the same time illustrates another form which should be true in all cases.

Our already very rudimentary equipment for generating terms and formulae must therefore be further restricted, so as to exclude such freak results. By way of such an added restriction, however, we cannot content ourselves with the direct stipulation that a formula is to be discarded as meaningless if, like (3), it is an instance simultaneously of a valid and a contradictory form. The fault of this stipulation is the lack of any finite test for the general case. We are trying to purify our language of idioms which might deceive us into contradicting ourselves; hence a restriction is of no avail which remains inapplicable until we have discovered ourselves in contradiction. Rather, we must isolate some immediately observable or at least finitely testable feature which (3) and similar freak cases have in common; then we may discard as meaningless all formulae exhibiting that feature.

Examination of (3) suggests that we could avoid Russell's paradox by rejecting as meaningless all formulae of *self-membership*—all formulae of the form " $x\epsilon x$ ". More accurately, since " $\epsilon$ " is not one of our primitive signs, the proposal would be to reject as meaningless all formulae of the kind which the definitions would *abbreviate* in the form " $x\epsilon x$ "; rejecting also, of course, all formulae having any such meaningless formulae as parts.

This would indeed dispose of (3), but it is easily seen that other paradoxes analogous to Russell's would still arise in spite of the suggested restriction. Namely, let  $K'$  be  $\hat{x}(y)\sim(x\epsilon y.y\epsilon x)$ ; i.e., let "K'" be used as an abbreviation of

the *term* (built up of inclusion and abstraction) which the series of definitions would abbreviate as " $\hat{x}(y) \sim (x \epsilon y \cdot y \epsilon x)$ ". The principle (1) tells us, then, that

$$"(z \epsilon K') \equiv (y) \sim (z \epsilon y \cdot y \epsilon z)"$$

should be true, and hence also, in particular,

$$(4) \quad "(\iota K' \epsilon K') \equiv (y) \sim (\iota K' \epsilon y \cdot y \epsilon \iota K')"$$

But familiar logical principles transform the right side of (4) successively into

$$(5) \quad "(y)((y \epsilon \iota K') \supset \sim (\iota K' \epsilon y))",$$

$$(6) \quad "(y)((y = K') \supset \sim (\iota K' \epsilon y))",$$

$$(7) \quad "\sim (\iota K' \epsilon K')".$$

We thus have the contradiction

$$(8) \quad "(\iota K' \epsilon K') \equiv \sim (\iota K' \epsilon K')".$$

Yet no self-membership was involved here: no expression of the form " $x \epsilon x$ ", either explicitly or implicitly through definitional abbreviations.

Russell's paradox and this one are merely the first two of a series. In the general case, we take  $K^{(n)}$  as

$$\hat{x}(y_1)(y_2) \cdots (y_n) \sim (x \epsilon y_1 \cdot y_1 \epsilon y_2 \cdot y_2 \epsilon y_3 \cdots y_n \epsilon x);$$

then, analogously to (4), we have

$$(9) \quad "(\iota^n K^{(n)} \epsilon K^{(n)}) \equiv (y_1)(y_2) \cdots (y_n) \sim (\iota^n K^{(n)} \epsilon y_1 \cdot y_1 \epsilon y_2 \cdots y_n \epsilon \iota^n K^{(n)})"$$

where " $\iota^n$ " represents  $n$  occurrences of " $\iota$ ". We next transform the right side of (9) into

$$"(y_1)(y_2) \cdots (y_n)((y_n \epsilon \iota^n K^{(n)}) \supset ((y_{n-1} \epsilon y_n) \supset \cdots ((y_1 \epsilon y_2) \supset \sim (\iota^n K^{(n)} \epsilon y_1)) \cdots))",$$

and finally we reduce this to " $\sim (\iota^n K^{(n)} \epsilon K^{(n)})$ " by  $n$  steps each of which is analogous to the step from (5) to (7).

It might appear that all the paradoxes would drop out if we strengthened our restriction to exclude all formulae which contain an *epsilon cycle*; i.e., all formulae which contain parts having the forms " $x \epsilon y_1$ ", " $y_1 \epsilon y_2$ ",  $\dots$  and " $y_n \epsilon x$ " (or the primitive expansions of these). Actually even this restriction is too mild, for there are formulae which contain no epsilon cycles but are logically equivalent to others which do. A simple example is

$$(10) \quad "(y)((x = y) \supset (x \epsilon y))"$$

(or its expansion into primitives); this contains no epsilon cycle, but it is logically equivalent to " $x \epsilon x$ " and hence leads to an equivalent of Russell's paradox.

**3. The theory of types.** A set of restrictions which does presumably avoid all paradoxes is provided by Russell's *theory of types*.<sup>3</sup> We must distinguish between the *metaphysical* or *ontological* aspect of this theory and the *metalogical* or *formal* aspect. In its ontological aspect the theory stipulates that if an individual is a member of a class  $x$ , then  $x$  must be composed exclusively of individuals; if an individual is a member of a member of a class  $x$ , then  $x$  must be composed exclusively of classes composed exclusively of individuals; and so on. Individuals are said to be of *type* 0, and classes of objects of type  $n$  are said to be of *type*  $n + 1$ ; and in these terms the theory of types amounts, in its ontological aspect, to demanding that all the members of a class be alike with respect to type.

The metalogical or formal aspect of the type theory is commonly set forth in terms which are not altogether formal, but involve reference also to the type ontology. Thus expressed, the formal aspect of the theory consists essentially of this stipulation: it is to be regarded as *meaningless*, rather than merely false, to indicate the relation of membership as holding between objects which are not of consecutive ascending types; likewise meaningless, rather than false, to indicate inclusion or identity as holding between objects which are not of the same type. Thus " $x \in y$ ", " $x \subset z$ ", " $x = z$ ", and all contexts thereof, become meaningless unless the values of " $x$ ", " $y$ ", and " $z$ " are thought of as restricted to the respective types  $n$ ,  $n+1$ , and  $n$  (for some  $n$ ).

The formal side of the type theory is studied most easily in application to our primitive notation of inclusion and abstraction, since the modes of notational combination to be scrutinized for meaningfulness are here reduced to a minimum. Thus applied, the stipulation is merely that " $\subset$ " be used only between terms which designate things of the same type.

This is not yet formal enough, because inspection of a term (variable or abstract) suggests no one appropriate type. A natural course, therefore, would be to modify our notation to the extent of attaching numerical indices to variables, thus indicating what type of objects each variable is to admit as values.<sup>4</sup> Then, since a class is of next higher type than its members, an abstract would denote a term of next higher type than the type indicated for its circumflexed variable. The formal aspect of the theory of types would thus reduce to this explicitly formal stipulation: The sign " $\subset$ " must occur only between terms

<sup>3</sup> What is relevant here is Russell's simple theory of class types. His theory took on a more complex form when applied to relations; and underlying his classial types and relational types there was his still more elaborate theory of types of so-called propositional functions. (See Whitehead and Russell [16], Vol. I, pp. 37-65.) But later work has made it apparent that both of these more complicated parts of the theory are superfluous. The complication regarding relations is eliminated through the reduction of relations to classes by Wiener [17] and Kuratowski [6]. (See also Gödel [5], p. 176; Tarski [15], pp. 363-364; Quine [10], pp. 123-124.) The superfluousness of the other complication, at the level of "propositional functions," was first suggested by Chwistek [3]. (See also Ramsey [12], pp. 20-29; Church [2], p. 169; Quine [8].)

<sup>4</sup> Such a notation has been used in some works, e.g. Tarski [14], pp. 97-103. To facilitate comparison with Tarski [14], it was used also in Quine [11].

with equal index numbers—the index number of an abstract being understood as the index number of the circumflexed variable plus one.

Russell's practice, however, which is more usual and more convenient, is to dispense with such indices; to leave the variables "typically ambiguous," in the sense of allowing them to denote objects of *any* types conformable to the context.<sup>5</sup> Applied to our primitive notation, this procedure of "typical ambiguity" consists in recognizing as meaningful any term or formula such that indices *could* be attached to all variables conformably with the described requirement on " $\subset$ ".

Thus, as applied to our primitive notation, the formal aspect of the theory of types comes to consist of the following stipulation:

(III) *A term or formula  $\zeta$  is to be retained as meaningful only if all terms occurring in  $\zeta$  can be assigned numbers in such a way that (a) " $\subset$ " connects only terms having like assignments, and (b) whatever number is assigned to an abstract  $\theta_\alpha$ , the next lower number is assigned to the variable  $\alpha$ .*

But this formulation is still not quite accurate. It proceeds on the understanding that all recurrences of a letter are recurrences of the same variable, and hence subject to the same numerical assignment. But actually such uniformity of assignment is in certain cases unnecessary. In " $\hat{x}(x \subset y) \subset x$ ", or " $\hat{x}(x \subset y) \subset \hat{x}(x \subset z)$ ", obviously the first two occurrences of " $x$ " have nothing to do with subsequent ones; the circumflexed variable of abstraction is relevant only to the abstract to which it belongs, and any recurrence of the same letter outside that abstract is merely an alphabetical coincidence. In the examples cited, the first two occurrences of " $x$ " could be rewritten as " $w$ " without any change in meaning. In general, the circumflexed prefix of an abstract  $\theta_\alpha$  affects only those occurrences of  $\alpha$  which are *free* in  $\theta$ : i.e., which are in  $\theta$  but are not in any abstract  $\eta_\alpha$  within  $\theta$ .<sup>6</sup>

In a refined formulation of (III), then, we would speak of assigning numbers to individual *term occurrences* rather than to terms. Instead of (b), we would say merely that, whatever number is assigned to an *occurrence* of an abstract  $\theta_\alpha$ , the next lower number is assigned to each *free occurrence* of  $\alpha$  in that occurrence of  $\theta$ . This compels like assignments to all free occurrences of  $\alpha$  in the occurrence of  $\theta$ , but imposes no uniformity on assignments to other occurrences of  $\alpha$ . Finally, we need an added condition to deal with occurrences of  $\alpha$  which lie outside all abstracts  $\theta_\alpha$ ; i.e., occurrences of  $\alpha$  which are free in the original term or formula  $\zeta$ . These free occurrences must of course still have like assignments among themselves. We thus arrive at this formulation of the formal aspect of type theory:

(IV) *A term or formula  $\zeta$  is to be retained as meaningful only if all term occurrences in  $\zeta$  can be assigned numbers in such a way that*

<sup>5</sup> To some extent, in the notation of Whitehead and Russell [16], type differences are reflected by styles of variables; but this is an inessential mnemonic device. See Quine [7], pp. 30–31.

<sup>6</sup> Note that the variable " $x$ " in a context of quantification " $(x) \dots$ " turns out to be a variable of abstraction, having the context " $\hat{x} \dots$ ", when the definitions of Quine [11] are applied.

- (a) " $\subset$ " connects only term occurrences having like assignments;
- (b) whatever number is assigned to an occurrence of an abstract  $\theta_\alpha$ , the next lower number is assigned to each free occurrence of  $\alpha$  in that occurrence of  $\theta$ ;
- (c) any two free occurrences in  $\zeta$  of the same variable have like assignments.

It is generally believed that the restrictions imposed by the formal aspect of the theory of types are sufficient to rescue logic and mathematics from the paradoxes. If the particular formulae (3)–(10) which led to the paradoxes of section 2 were written out in full primitive notation, it could easily be seen that they are all rejected as meaningless by (IV). The same is true of the terms " $K$ ", " $K'$ ", etc.

**4. Abandonment of the type ontology.** One especially unnatural and awkward effect of the type theory is the infinite reduplication of each logically definable class. There is no longer one universal class  $V$  to which everything belongs, for the theory of types demands that the members of a class be alike in type. We must thus content ourselves with a separate universal class for each type. The same reduplication affects all other classes definable in logical terms; even the numbers 0, 1, etc. lose their uniqueness, giving way to a duplicate for every type.

This reduplication is particularly strange in the case of the null class. One feels that classes should differ only with respect to their members, and this is obviously not true of the various null classes. A unique null class indeed still seems permissible, vacuously, if we think only of the requirement that members be alike in type. However, other requirements of type theory would be violated. For example, we want the null class to be included in each class; hence, inasmuch as it is regarded as meaningless to relate classes of unlike types by inclusion, we need a new null class to be included in each class of new type.

The constants " $v$ ", " $\Lambda$ ", " $0$ ", " $1$ ", etc. are thus "typically ambiguous," just as is the case with variables. Indeed, since our terms are built up of variables by means solely of inclusion and abstraction, *all* our terms are typically ambiguous; and the constants under consideration are merely definitional abbreviations of certain of these terms.

Another effect of the type theory appears in connection with a theorem of arithmetic, namely the theorem to the effect that  $n \neq n+1$  for finite  $n$ . The proof of this theorem depends on producing a class of at least  $n$  members; and this is accomplished as follows. We start with  $V$  and  $\Lambda$ , determined as of any one type. Then there are four classes having none, one, or both of  $V$  and  $\Lambda$  as members. Then there are 16 classes having none, one, two, three, or all of these four classes as members. After a finite number of steps of this kind we reach a level providing at least  $n$  classes. These, together, compose a class of at least  $n$  members as was required.

But observe that this process carries us higher and higher in the hierarchy of types. Consequently the proof establishes only that  $n \neq n+1$  when the numbers are construed as of sufficiently high types. Within lower types the theorem may still fail.

A remedy suggested by Whitehead and Russell is an axiom, valid for each

type, to the effect that there is an infinite class.<sup>7</sup> Some such axiom is in any case presumably needed for the theory of infinite numbers; but that it should be needed for proving finite inequalities is an anomalous effect merely of the theory of types.

But this is part of a broader problem, raised by the device of typical ambiguity. This device operates in such a way that Whitehead and Russell *could* have proved their theorem of inequality after all, in its full generality, without adding any axiom of infinity. That is, if they considered merely their logical formalism they could present in symbolic form precisely the proof outlined above. In some of the intermediate steps of the proof it would be contextually apparent, despite the typical ambiguity of the symbols, that the classes dealt with were of progressively higher types; but these contextual evidences would have dropped out by the time one reached the conclusion " $n \neq n+1$ ".

Such a proof, though admitted by the apparent formalism of *Principia mathematica* and related systems, seems to involve an abuse of typical ambiguity: a theorem is unconditionally asserted which, judged merely on its internal structure, admits determinations of type not covered by the proof. Hence Whitehead and Russell did not choose this easy way; indeed, to avoid being deceived into this fallacious sort of argument they even brought in a heuristic notation of suffixes for keeping track of the range of types covered by a proof.<sup>8</sup> No such precautions were explicit in the initial formalism of their system, and indeed it would be a matter of some complexity to incorporate them explicitly. Obviously the abuse of typical ambiguity would be much more convenient. Further, despite its apparent lack of cogency, this practice seems never to yield any intrinsically undesirable theorems.

The awkward situations thus far considered actually depend, not on the formal aspect of type theory, but only on the ontological aspect. Let us then try abandoning the ontological aspect altogether, retaining only the formal restrictions: for if the theory of types is adequate at all as a safeguard against contradictions, it must be adequate in its formal aspect alone.

The whole notion of type is now dropped. Some classes may now contain both individuals and classes as members, and some classes may even be members of themselves. Typical ambiguity of variables disappears; each variable may henceforward be thought of simply as having the unrestricted universe as its range. Typical ambiguity of constants similarly disappears; the sign " $\forall$ " now denotes just the unique universal class, to which absolutely everything belongs; the sign " $\Lambda$ " a unique null class; the sign " $0$ " a unique number 0; and so on.

Such expressions as " $\forall \epsilon V$ ", " $\sim(\Lambda \epsilon \Lambda)$ ", etc., can now be taken literally; the universal class is indeed a member of itself, the null class not. These expressions were also countenanced under the standard theory of types; but one took care to explain that in " $\forall \epsilon V$ " the typical ambiguity of the sign " $\forall$ " was to be resolved differently in the two occurrences. For the first occurrence the type

<sup>7</sup> See Whitehead and Russell [16], Vol. II, p. 203.

<sup>8</sup> See Whitehead and Russell [16], Vol. I, pp. 415-417; Vol. II, pp. vii-xxxi, 5-12, 285-290.



was to be lower by one than for the second. A similar remark would be applied to " $\sim(\Lambda\epsilon\Lambda)$ ", " $\Lambda\epsilon V$ ", " $\Lambda\epsilon 0$ ", " $\sim(0\epsilon 0)$ ", etc. But in discarding the type ontology we slough off this complication; we abandon typical ambiguity, restore the uniqueness of the logically definable classes, and cease to be offended in general by self-membership and other so-called confusions of type.

The effect is observable not only in the case of variables and constants, but also in the case of functions. For example the *negate*  $\bar{x}$ , defined as  $\dot{y}\sim(y\epsilon x)$ , is construed under the theory of types as comprising as members not *all* the non-members of  $x$ , but just those non-members of  $x$  which are of appropriate type for membership in  $x$ . Abandoning the type ontology, however, we restore  $\bar{x}$  to its common-sense status: the class of absolutely everything except the members of  $x$ .

Abandonment of the type ontology disposes also of such difficulties as the one about numerical inequalities. In effect, we now adopt without question the practice described above as abuse of typical ambiguity; but the procedure no longer turns upon typical ambiguity, nor involves any special assumption. Construction of *any* class of  $n$  or more members now provides a proof that *the* number  $n$  is distinct from *the* number  $n+1$ .

All this freedom is gained without altering the restriction (IV) on meaningful terms and formulae. We merely divorce this restriction from any connotations of type. The type ontology was at best only a graphic representation or metaphysical rationalization of the formal restrictions; and though some such rationalization may well be desired, it seems clear in particular that the type ontology afforded less help than hindrance.

**5. Relaxation of the formal restriction.** Removed from its background of types and viewed as an ultimate restriction, (IV) itself remains arbitrary and unnatural. We shall see that the unnatural features can in large part be eliminated by moderating (IV) in a certain way; but let us consider first what some of the unnatural features are.

Note that the meaninglessness of a given *term* is not, in general, difficult to conceive—quite apart from any theory of types. An abstract purports to denote a class whose members are all and only the objects  $x$  satisfying a given formula; but, for certain formulae, there may be no such corresponding class—every class may either miss some of those objects or else contain some others in addition. Russell's paradox shows, e.g., by *reductio ad absurdum*, that there can be no class corresponding to the formula " $\sim(x\epsilon x)$ ".

Moreover, if we concede the meaninglessness of a given term we are of course ready to concede also the meaninglessness of any formula containing that term. There remains, however, the case of a meaningless formula containing only meaningful terms. Every formula, in primitive notation, is an inclusion compound; so the case now under consideration is the case of two meaningful terms, say abstracts, joined by " $\subset$ " to make a meaningless formula. The whole is a *meaningless* statement of inclusion concerning two *genuine* classes. Having abandoned the type ontology, we can no longer excuse ourselves with the thought that the two classes are of different types; and hence it is hard to admit that it means nothing to say that the one class is included in the other.

Another somewhat unnatural effect of (IV) is that many terms and formulae such as " $x\epsilon y$ " (or its primitive expansion) are retained as meaningful while certain substitution instances thereof, such as " $x\epsilon x$ ", are rejected as meaningless. And there is a still more unnatural effect, which is the reverse: many formulae such as " $x\epsilon x$ " are rejected as meaningless, while substitution instances thereof, such as " $\forall\epsilon V$ ", are retained as meaningful. Intuitively it would seem, e.g., that " $x\epsilon x$ " can be meaningless only through meaninglessness in general of self membership; and that " $\forall\epsilon V$ " should then be meaningless by the same token. We no longer have the excuse originally provided by the theory of types, namely that because of typical ambiguity " $\forall\epsilon V$ " is really not a case of self-membership.

These anomalies of substitution were illustrated just now with definitionally abbreviated formulae " $x\epsilon y$ ", " $x\epsilon x$ ", " $\forall\epsilon V$ ". But simple examples in primitive notation are also easily found, e.g.

$$"\hat{x}(x\subset z)\subset w", \quad "\hat{x}(x\subset z)\subset z", \quad "\hat{x}(x\subset y(y\subset y))\subset y(y\subset y)".$$

Another anomaly is the fact that a conjunction, conditional, or other truth function composed of meaningful formulae may itself be meaningless; e.g., " $x=y$ " and " $x\epsilon y$ " are meaningful but " $x=y.x\epsilon y$ " is meaningless. From an intuitive standpoint it is hard to concede that two formulae can be meaningful separately, understood separately, and yet meaningless in conjunction. Nor can we appeal any longer to "confusion of type" as an excuse.

Now all those anomalies can be swept away by one simple change in the restriction (IV). We merely omit (c), obtaining the following:

(V)<sup>9</sup> *A term or formula  $\zeta$  is to be retained as meaningful only if all term occurrences in  $\zeta$  can be assigned numbers in such a way that*

- (a) " $\subset$ " connects only term occurrences having like assignments, and
- (b) whatever number is assigned to an occurrence of an abstract  $\theta_\alpha$ , the next lower number is assigned to each free occurrence of  $\alpha$  in that occurrence of  $\theta$ .

The difference between (IV) and (V) is illustrated by " $\hat{x}(x\subset y)\subset y$ ". This formula becomes meaningful; for, (a) and (b) are fulfilled by assigning 0 to " $x$ ", 0 to the first occurrence of " $y$ ", 1 to " $\hat{x}(x\subset y)$ ", and 1 to the second occurrence of " $y$ ". On the other hand such assignment of 0 and 1 to the occurrences of " $y$ " would have violated (c) of (IV). Other expressions which (IV) renders meaningful include the formulae " $x\epsilon x$ ", " $\sim(x\epsilon x)$ ", and " $x=y.x\epsilon y$ " (or their primitive expansions); also the formula (10) of section 2; also the terms " $\hat{x}(x=y.x\epsilon y)$ " and " $\hat{x}\sim(x=y.x\epsilon y)$ "; not, however, " $\hat{x}(x\epsilon x)$ ", nor any of the series " $K$ ", " $K'$ ", ... of section 2. In general, it is easily seen that the terms and formulae which are meaningful under (V) comprise just those which would

<sup>9</sup> If one prefers the kind of notation which attaches indices to the variables (see section 3), he will find that the moderated theory embodied in (V) is easily adapted also to that procedure. The indices would be viewed as belonging, not to the general notation of variables, but to the notation of abstraction; only variables of abstraction would bear them. A variable  $\alpha$  would bear an index at its initial circumflexed occurrence in  $\theta_\alpha$ , and the same index at all of its free occurrences in the formula  $\theta$ ; but no index outside such contexts. As previously, we would define the index number of an abstract as one plus the index number of its circumflexed variable; and we would forbid use of " $\subset$ " between terms with unequal index numbers. But unindexed variables would remain unaffected by the restriction.

be meaningful also according to (IV) if all free occurrences of variables were replaced by distinct letters.

Though (V) is a much milder restriction than (IV), no threat of paradox has appeared; and a scheme as liberal as (V), published earlier,<sup>10</sup> has already had expert scrutiny.<sup>11</sup> There is indeed no proof that paradoxes are excluded, but then neither is there such a proof for the original theory of types.<sup>12</sup>

**6. Disappearance of the anomalies.** Under (V) it ceases to be true that a formula can be meaningless and yet contain only meaningful terms. For, consider a formula  $\zeta I \eta$  such that  $\zeta$  and  $\eta$  are meaningful terms according to (V). Then numbers can be assigned to all term occurrences in  $\zeta$  conformably with (a) and (b); and similarly for  $\eta$ . Let  $S_1$  be such a system of assignments for  $\zeta$ , and let  $m$  be the number which it assigns to the occurrence of  $\zeta$  itself; and let  $S_2$  and  $n$  be similarly related to  $\eta$ . If  $S_2$  is changed by adding  $m-n$  to each of the assigned numbers, the result  $S_3$  will still satisfy (a) and (b); this is apparent from the purely relative nature of (a) and (b). Now  $S_1$  and  $S_3$  constitute together a system of assignments  $S_4$  for  $\zeta I \eta$  as a whole. Each occurrence of " $\subset$ " within  $\zeta$  or  $\eta$  connects term occurrences having like assignments under  $S_4$ , since  $S_1$  and  $S_3$  fulfill (a); and the remaining occurrence of " $\subset$ " in  $\zeta I \eta$  also connects term occurrences having like assignments under  $S_4$ , for it connects  $\zeta$  and  $\eta$ , both of which are assigned  $m$ . Thus  $S_4$  satisfies (a). Furthermore,  $S_4$  satisfies (b); for,  $S_1$  and  $S_3$  both satisfy (b), and there is no abstract  $\theta_a$  in  $\zeta I \eta$  which is not in  $\zeta$  or  $\eta$ . Hence  $\zeta I \eta$  is a formula according to (V).

Thus every inclusion compound of meaningful terms is now meaningful. Terms are the essential locus of meaninglessness; a formula can be meaningless only derivatively, though containing a meaningless term.

Equivalently, indeed, we might omit the reference to formulae in (V), omit also the original description (II) of formulae, and then simply describe a meaningful formula once and for all as an expression  $\zeta I \eta$  such that  $\zeta$  and  $\eta$  are meaningful terms.

Another anomaly which disappears is the possibility of a meaningless truth

<sup>10</sup> Quine [9], pp. 79-80. The primitives in [9] are different, and abstraction is not among them. But [9] is related to the present scheme in this way: if an abstract " $\hat{x} \dots$ " (lacking " $y$ ") is meaningful under (V), then we can prove in [9] that  $(\exists y)(x)((xey) = \dots)$ ; i.e., that there is a class such as " $\hat{x} \dots$ " purports to express. This is seen as follows. Suppose " $\hat{x} \dots$ " formed from " $\hat{x} \dots$ " by replacing all free occurrences of variables by new and distinct letters. As observed above, then, " $\hat{x} \dots$ " and hence also " $\dots$ " will be meaningful under (IV). But the formulae meaningful under (IV) are just those which are "stratified" in the sense of [9]—due allowance being made for the difference in primitive notation. Hence R3' of [9] yields " $(\exists y)(x)((xey) = \dots)$ ". From this, by substitution on the free variables (a form of inference allowed by the rules of [9]), we derive " $(\exists y)(x)((xey) = \dots)$ ".

Note, incidentally, these four corrections of [9]. (i) Of the two explanations of stratification on page 78, only the first is relevant; the second was included because it was erroneously supposed equivalent. Cf. Bernays [1]. (ii) The dated postscript at the end of [9] leads nowhere. Cf. Quine [10], note 4; or, indeed, section 2 above. (iii) R1 should end with " $\omega$ " instead of " $\psi$ ". (iv) R4 should end with " $\chi$ " instead of " $\psi$ ".

<sup>11</sup> See Rosser [13]; also Bernays [1], Curry [4].

<sup>12</sup> See Appendix B.

function of meaningful formulae. By examining the definitions whereby denial, conjunction, the conditional, etc. are introduced in terms of inclusion and abstraction, it could easily be seen that all truth functions of meaningful formulae are meaningful according to (V).

Finally, the anomalies of substitution also disappear. It becomes true that every substitution instance of a meaningful term or formula is meaningful, and that every term or formula having a meaningful substitution instance is meaningful. Preparatory to establishing this, we need an explicit formulation of *logical substitution*: A term of formula  $\zeta'$  is said to result from substituting a term  $\eta$  for a variable  $\beta$  in a term of formula  $\zeta$  if  $\zeta'$  is formed by putting  $\eta$  for all free occurrences of  $\beta$  in  $\zeta$ , and no free occurrence of  $\beta$  in  $\zeta$  stands within a term  $\theta_\alpha$  such that  $\alpha$  has a free occurrence in  $\eta$ .<sup>13</sup> Now what is to be proved is that, if  $\eta$  is a meaningful term according to (V), and  $\zeta$  and  $\zeta'$  are as just now described, then  $\zeta'$  is meaningful if and only if  $\zeta$  is.

Suppose first that  $\zeta$  is meaningful, and hence admits of a system  $S_1$  of assignments conforming to (a) and (b). Likewise  $\eta$  admits of such a system  $S_2$ . Let  $m_1, \dots, m_k$  be the numbers assigned by  $S_1$  to the respective free occurrences of  $\beta$  in  $\zeta$ , and let  $n$  be the number assigned by  $S_2$  to the occurrence of  $\eta$  in itself. Now let  $S_3$  be the following system of assignments to all term occurrences in  $\zeta'$ : throughout that occurrence of  $\eta$  which supplants the  $i$ th free occurrence of  $\beta$  in  $\zeta$ , we make assignments as in  $S_2$  but with  $m_i - n$  added to each assignment; and to all other term occurrences in  $\zeta'$  we make assignments just as in  $S_1$ . Since  $n + (m_i - n) = m_i$ , we see that  $S_3$  assigns the same numbers to the substitute occurrences of  $\eta$ , in  $\zeta'$ , which  $S_1$  assigned to the corresponding occurrences of  $\beta$  in  $\zeta$ . Outside such occurrences of  $\eta$ , further,  $S_3$  simply duplicates  $S_1$ . Then, since  $S_1$  conforms to (a), it follows that  $S_3$  also conforms to (a) insofar at least as concerns any occurrence of "C" outside the substitute occurrences of  $\eta$ .

Now consider an occurrence of an abstract  $\theta_\alpha$  in  $\zeta'$ . Even if this occurrence of  $\theta$  contains some of the substitute occurrences of  $\eta$ , we know from the definition of substitution that the free occurrences of  $\alpha$  in  $\theta$  will fall outside such occurrences of  $\eta$ . Hence, in all cases at least except where the occurrence of  $\theta_\alpha$  is wholly inside one of the substitute occurrences of  $\eta$ ,  $S_3$  will agree with  $S_1$  in its assignments to the free occurrences of  $\alpha$  in  $\theta$ . Then, since  $S_1$  conforms to (b), we see that  $S_3$  conforms to (b) insofar at least as concerns occurrences of  $\theta_\alpha$  not within the substitute occurrences of  $\eta$ .

But  $S_3$  also conforms to (a) and (b) insofar as concerns any occurrence of "C" or  $\theta_\alpha$  within the  $i$ th substitute occurrence of  $\eta$ ; for,  $S_2$  conformed to (a) and (b), and addition of a constant  $m_i - n$  to each assignment does not affect this property. Hence  $S_3$  conforms completely to (a) and (b). Hence  $\zeta'$  is meaningful according to (V).

It remains to prove, conversely, that  $\zeta$  is meaningful if  $\zeta'$  is. If  $\zeta'$  is meaningful, it has a system  $S_1$  of assignments conforming to (a) and (b). Now let  $S_2$  be the following set of assignments to all term occurrences in  $\zeta$ : the free occurrences of  $\beta$  in  $\zeta$  receive the same assignments which the corresponding occur-

<sup>13</sup> Cf. Tarski [14], p. 103, or Quine [11], p. 146.

rences of  $\eta$  received under  $S_1$ , and all other term occurrences receive the same assignments as in  $S_1$ . Now it is immediately apparent that  $S_2$ , like  $S_1$ , conforms to (a). Next consider any occurrence of an abstract  $\theta_\alpha$  in  $\zeta$ .! No occurrence of  $\alpha$  in that occurrence of  $\theta$  is simultaneously a free occurrence of  $\beta$  in  $\zeta$ , by the definition of freedom (§3). Hence all occurrences of  $\alpha$  in the occurrence of  $\theta$  are assigned numbers by  $S_2$  in accordance with  $S_1$ . Then, since  $S_1$  conforms to (b), so does  $S_2$ . Therefore  $\zeta$  is meaningful according to (V).

**Appendix A. Elimination of the retroactive feature.** As formulated, the theory of types performs a peculiar function of expurgation: a totality of terms and formulae is first specified as in (I)–(II), and afterward certain of these are weeded out by (IV). The case is similar under (V).

Both theories can be freed of this retroactive feature. The recursive description of term given in (I) can be supplanted by a narrower one which provides, from the very beginning, just those terms which would be left standing by (IV). The same is possible with regard to (V).

These formulations are more complicated than the formulations presented earlier, and they appear to be less convenient technically. It may be worth while, however, to record them. Proof of their equivalence with the previous formulations will not be undertaken.

The recursive description of *term* which would supplant (I) and (IV) is the following. It is at the same time a recursive description of an auxiliary notion of *rank*.

(VI) (a) *A variable is a term and has rank 0 with respect to itself.*

(b) *If  $\alpha$  is a variable,  $\zeta$  and  $\eta$  are terms, and for each variable  $\gamma$  there is at most one number  $r$  such that  $\zeta$  or  $\eta$  has rank  $r$  with respect to  $\gamma$ , then  $(\zeta\eta)_\alpha$  is a term; and if  $\zeta$  or  $\eta$  has rank  $m$  with respect to  $\alpha$ , and  $\zeta$  or  $\eta$  has rank  $n$  with respect to a variable  $\beta$  distinct from  $\alpha$ , then  $(\zeta\eta)_\alpha$  has rank  $n - m + 1$  with respect to  $\beta$ .*

Now the formulae are describable thus:

(VII) *Formulae are the expressions  $\zeta$  such that  $\zeta_\alpha$  is a term if  $\alpha$  is a variable.*

Granted that (VI) yields as terms just those terms in the sense (I) which are meaningful according to (IV), it is then obvious also that (VII) will yield as formulae just those formulae in the sense (I)–(II) which are meaningful according to (IV).

For a recursive definition of *term* supplanting (I) and (V), we change (VI) to just this extent: instead of “for each variable . . . with respect to  $\gamma$ ,” we put “there is at most one number  $m$  such that  $\zeta$  or  $\eta$  has rank  $m$  with respect to  $\alpha$ .” The formulae are then describable as in (VII), or, equivalently and more simply, as in (II).

**Appendix B. Two deductive systems.** The question of the consistency of (V) has no precise meaning until a deductive system is specified. Such a system will now be presented, comprising just two postulates and four rules of inference. The postulates, expressed with help of the abbreviations “ $\epsilon$ ” and “ $\supset$ ” (as defined in Quine [11]), are these:

$$(x \subset y) \supset ((y \subset z) \supset (x \subset z))$$

$$(\hat{w}(w\epsilon y \supset w\epsilon z) \subset \hat{w}(w\epsilon x \supset w\epsilon y)) \supset (x \subset y)$$

In stating the rules of inference, I shall write " $\zeta$  repl  $\eta$ " to mean that when  $\zeta$  is put for an occurrence of  $\eta$  in any theorem the result is a theorem. I shall write " $\zeta E\eta$ " to denote the expression formed by putting  $\zeta$  and  $\eta$  in the respective blanks of " $( \epsilon )$ " and expanding the result into primitives according to the definitions in [11]. "Term" and "formula" are to be understood in the sense of (I)–(II); and "meaningful" is to be understood in the sense of (V). The four rules, then, are these:

- (1) *If  $\zeta$  and  $\eta_\alpha$  are meaningful terms and  $\eta$  is a theorem,  $\zeta I\eta_\alpha$  is a theorem.*
- (2) *If  $\zeta$  and  $\zeta_\alpha I\eta_\beta$  are theorems, so is  $\eta$ .*
- (3) *If  $\zeta$  results from substituting the term  $\eta$  for the variable  $\alpha$  in the formula  $\theta$ , then  $\zeta$  repl  $\eta E\theta_\alpha$ .*
- (4) *If  $\zeta$  is a term containing no free occurrence of the variable  $\alpha$ , then  $(\alpha E\zeta)_\alpha$  repl  $\zeta$ .*

Rules (1)–(3) are essentially R3–5 of [11]. Rule (4) allows, indirectly, alphabetical change of a variable of abstraction; no such rule was included in [11], because the above two postulates were rendered in [11] as rules R1'–2' with unspecified variables.

The above rules are so framed that no formula can become a theorem if meaningless according to (V). It seems likely, however, that even this degree of restriction is unnecessary. Presumably the stipulation of the meaningfulness of  $\eta_\alpha$  can be dropped from (1) without contradiction. The thus liberalized system will yield theorems violating (V); hence the word "meaningful," in connection with (V), should be abandoned in favor of a more neutral word. Let us adopt rather the word "stratified" to denote conformity to (V). (This word now acquires a broader sense, of course, than in Quine [9].) The system then assumes this form:

*Postulates as before.*

*Rules: (1') If  $\zeta$  is a stratified term,  $\alpha$  is a variable, and  $\eta$  is a theorem, then  $\zeta I\eta_\alpha$  is a theorem.*

(2)–(4) as before.

By giving (1) the relaxed form (1'), we let down the bars to unstratified terms  $\eta_\alpha$  such that  $\eta$  is a theorem; such terms, e.g., as " $\hat{x}(xex \supset xex)$ " (or its primitive expansion). In effect, thus, we recognize such terms as meaningful. Stratification (conformity to (V)) becomes merely a sufficient condition for meaningfulness, not a necessary one. The question of a necessary condition for meaningfulness is abandoned.

This course is strongly recommended by intuitive considerations. The meaningfulness of an unstratified term  $\eta_\alpha$  is conveniently thought of in general as non-existence of the class which  $\eta_\alpha$  purports to describe; yet if  $\eta$  is a theorem, the term  $\eta_\alpha$  (e.g. " $\hat{x}(xex \supset xex)$ ") would still seem to describe a genuine enough class, namely V. More generally, on similar grounds, we should like to allow for the meaningfulness of an unstratified term  $\eta_\alpha$  whenever there is a meaningful term  $\zeta_\alpha$  such that  $\zeta$  and  $\eta$  are equivalent. The system under consideration permits all this.

Technically, also, the system under consideration is more convenient than

the system involving (1). Much less attention to (V) is required in the course of deductions. This is especially striking in the case of inference by substitution for free variables—a form of inference not listed among the above rules, but capable of justification on the basis of rules (1) (or (1')) and (3).<sup>14</sup> Under the version (1)–(4) we must, in effect, inspect not just the substituted term but the whole resulting formula for conformity to (V); under the version (1')–(4), on the other hand, we need inspect only the substituted term.

## REFERENCES

- [1] Paul Bernays. Review of [9]. *This JOURNAL*, vol. 2 (1937), pp. 86–87.
- [2] Alonzo Church. Review of Chwistek. *Ibid.*, pp. 168–170.
- [3] Leon Chwistek. *Antynomje logiki formalnej. Przegląd filozoficzny*, vol. 24 (1921), pp. 164–171.
- [4] H. B. Curry. Review of [9]. *Zentralblatt für mathematik*, vol. 16 (1937), p. 193.
- [5] Kurt Gödel. *Ueber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik*, vol. 38 (1931), pp. 173–198.
- [6] Casimir Kuratowski. *Sur la notion de l'ordre dans la théorie des ensembles. Fundamenta mathematicae*, vol. 2 (1920), pp. 161–171.
- [7] W. V. Quine. *A system of logistic*. Cambridge, Mass., 1934.
- [8] ———. *On the axiom of reducibility. Mind*, vol. 45 NS (1936), pp. 498–500.
- [9] ———. *New foundations for mathematical logic. The American mathematical monthly*, vol. 44 (1937), pp. 70–80.
- [10] ———. *On Cantor's theorem. This JOURNAL*, vol. 2 (1937), pp. 120–124.
- [11] ———. *Logic based on inclusion and abstraction. Ibid.*, pp. 145–152.
- [12] F. P. Ramsey. *The foundations of mathematics and other logical essays*. New York and London, 1931.
- [13] J. B. Rosser. *On the consistency of Quine's "New foundations for mathematical logic."* Abstract in the *Bulletin of the American Mathematical Society*, vol. 44 (1938), p. 43.
- [14] Alfred Tarski. *Einige Betrachtungen über die Begriffe der  $\omega$ -Widerspruchsfreiheit und der  $\omega$ -Vollständigkeit. Monatshefte für Mathematik und Physik*, vol. 40 (1933), pp. 97–112.
- [15] ———. *Der Wahrheitsbegriff in den formalisierten Sprachen. Studia philosophica*, vol. 1 (1935), pp. 261–405.
- [16] A. N. Whitehead and Bertrand Russell. *Principia Mathematica*. 2d edition. Cambridge, England, 1925–27.
- [17] Norbert Wiener. *A simplification of the logic of relations. Proceedings of the Cambridge Philosophical Society*, vol. 17 (1912–14), pp. 387–390.

HARVARD UNIVERSITY

---

<sup>14</sup> Cf. Quine [11], p. 152 (M27).