A PRACTICAL FORM OF TYPE THEORY I

A. M. Turing

It is usual for mathematicians to pay lip-service to the theory of types, but they will not usually make any attempt to bring their mathematics into line with it. An occasional paradox may perhaps be attributed to neglect of types, but no suggestions are made for the avoidance of these paradoxes short of the expression of all mathematics in the formalism ~~mf~~ of Principia Mathematica (say).[1] In the present paper a system ~~xixx~~ will be described which takes account of type theory, but at the same time follows very closely the normal mathematical outlook. The type theory intrudes itself on the system only very slightly, and its effect may be summed up in the form of one or two simple and natural cautions, which are easily carried over to unformalised mathematics: this should, I hope, enable all serious mathematics such ~~terkxxbrxxghxtxintoxlinexxith~~ as ~~xxixt~~ is supposedly based on the theory of types to be brought genuinely into line with it, at ~~xxxx~~ the cost of very little additional trouble tothe ~~writer~~ mathematician.

The paper will appear in ~~three~~ two parts. ~~In~~ The first part is written chiefly for the mathematician who wishes to increase the rigour of his proofs along the lines indicated in the previous paragraph, rather than for the logician. The emphasis will be on notation and meaning rather than on axioms and rules of procedure; these will however be given for the sake of completeness. ~~The notation will be illustrated by examples fromthe theory of real analysis etc.~~ The second part will be devoted to a little axiomatic development, and the justification of th e system in the case of the 'finite universe' i.e. the case where there is only a finite number of individuals. ~~Th e third part~~ II will establish a very complete connection between this system and th at of Church[2] ~~(Church 1)~~. This connection seems to be valuable because Church's system has greater theoretical simplicity th an the xxxxxxx proposed 'practical system', but is less convenient for th e formalisation of proofs. Consequently it will be natural to express proofs in the practical system, but metamathematical results in terms of Church's system.

The author wishes to repudiate any implication that may be suggested by this paper to the effect that he believes the Russell philosophy of mathematics[1] to be the truest xxxxxxx. He does believe however that it is the one which is most easily understood, and also that it describes xxxxxxxxxxxx most closely xxxxx th e accepted xxxxxxxxxxxxxxxx form of present day mathematical thinking. This paper is concerned with giving accurate expression to that thinking. When xxxxx that is done xxxxxxxxxxxx it will be easier to see the limitations of the outlook which goes with this form of thinking.

6. <u>Metamathematical results</u>

Two theorems are quoted here of which proofs will be published in part II.

<u>Theorem A</u>. With an interpretation of formulae essentially as described in § 1 (in terms of tables etc.) and with a finite universe of N individuals, the necessary and sufficinet condition for a propositional formula to be interpretable as true is that it be provable by the use of rules I to IX and X_N.

Before we can state theorem B, we shall have to define what we mean by the equivalence of two logical systems.

<u>Definition</u>. A logical system 1 will be said to be <u>equivalent</u> to the logical system 2 if to each proposition-like formula $\underline{A}$ of 1 we can make correspond a proposition like formula $\underline{A}^{(1,2)}$ of 2, and conversely to each proposition-like formula $\underline{P}$ of 2 we can make correspond a proposition-like formula $\underline{P}^{(2,1)}$ of 1, in such a way that

~~ii~~ i) If $\underline{A}$ is provable in 1 then $\underline{A}^{(1,2)}$ is provable in 2.

ii) If $\underline{P}$ is provable in 2 then $\underline{P}^{(2,1)}$ is provable in 1.

iii) If $\underline{A}$ is a proposition-like formula of 1 then $(\underline{A}^{(1,2)})^{(2,1)} \equiv \underline{A}$ is provable in 1.

~~iii~~ iv) If $\underline{P}$ is a proposition-like formula of 2 then $(\underline{P}^{(2,1)})^{(1,2)} \equiv \underline{P}$ is provable in 2.

v) If $\underline{A}$ and $\underline{B}$ are proposition-like formulae of 1 then we can prove $(\underline{A} \equiv \underline{B})^{(1,2)} \equiv (\underline{A}^{(1,2)} \equiv \underline{B}^{(1,2)})$ in 2.

vi) If $\underline{P}$ and $\underline{Q}$ are proposition-like formulae of 2 then we can prove $(\underline{P} \equiv \underline{Q})^{(2,1)} \equiv (\underline{P}^{(2,1)} \equiv \underline{Q}^{(2,1)})$ in 1.

~~Refer xxxxxxxxxxxxx~~ It is understood that for each system there is a special kind of formulae 'proposition-like formulae' defined, that every provable formula is necessarily proposition-like, but that it is a comparatively trivial matter to determine

1. Th e "practical system" for a finite universe.

In this section the notation of the practical system will be explain ed. The explanation will be in terms of the 'finite universe', i.e. we start with a finite number of objects or 'individuals' and build up other entities from th ese. We can then formulate certain rules which give valid results in this case and hope that they will apply in the infinite case also. We cannot of course hope that all such rules will work. We have to imagine that many rules of this kind have been tried, found wanting and rejected, and that others are still in use. This rather unsatisfactory-sounding process is as good an account as the author feels can be given of the way in which current mathematical procedure has grown up. But ~~however all~~ whatever the truth of this may be the finite universe provides a first class ground on whoih to describe the practical system, and we proceed accordingly.

Our finite universe has initially as its members the 'individuals' $U_1 \ldots U_N$. Although these ~~are~~ include all the individuals they need not exhaust our stock-in-trade, for we can also bring in functions taking the individuals as arguments and having them also as values. With our increased range of commodities we can then go into business again and produce a still greater variety of objects, and repeat without limit. There obviously arises a ~~xxxx~~ great variety of different kinds of function, which may need to be distinguished, but for the present system we need only trouble ourselves with th e very broadest divisions, which will be called ~~categories~~ types~~, or, when differently arranged, types~~. These divisions are described below.

whether a formula is proposition like. Specifically we may say that the statement " $\underline{A}$ is a proposition-like formula" should be equivalent to " $\varphi(n) = 0$ " where n is the Gödel representation of $\underline{A}$ and φ is some primitive recursive function. It is also understood that both systems "include the propositional calculus": this is required in connection with the xxxxx logical equivalence signs in ~~v) and~~ iii) to vi).

We are justified in describing this relation as the equivalence of the two systems, for the relation is transitive, symmetric, and reflexive, as I shall now show. The symmetry of the relation follows at once from the fact that interchange of systesm 1 and 2 simply interchanges conditions i) and ii), iii) and iv), v) and vi). Reflexiveness is proved by taking $\underline{A}^{(1,1)}$ to be $\underline{A}$. Transitivity is not quite so easy. We shall have to bring in a third system 3. We will define $\underline{A}^{(1,3)}$ to be $(\underline{A}^{(1,2)})^{(2,3)}$ and $\underline{A}^{(3,1)}$ to be $(\underline{A}^{(3,2)})^{(2,1)}$ We assume conditions i) to vi) to hold for the pairs 1,2 and 2,3 and attempt to prove them for the pair 1,3. Because of the symmetry it is sufficient to prove i), iii), v). To prove i) we must prove $(\underline{A}^{(1,2)})^{(2,3)}$ in 3 assuming $\underline{A}$ provable in 1. Now by i) for the pair 1,2 we see that $\underline{A}^{(1,2)}$ is provable in 2, and then by i) for the pair 2,3 we get $(\underline{A}^{(1,2)})^{(2,3)}$ in 3. To prove iii) we must prove $(((\underline{A}^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv \underline{A}$ in ~~2~~ 1. Using iii) for the pair 2,3 gives us $((\underline{A}^{(1,2)})^{(2,3)})^{(3,2)} \equiv \underline{A}^{(1,2)}$ (in 2), whence by ii) for the pair 1,2 we have $(((\underline{A}^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv (A^{(1,2)})^{(2,1)}$ Also by vi) for the pair 1,2 we have

$$(((A^{(1,2)})^{(2,3)})^{(3,2)} \equiv A^{(1,2)})^{(2,1)} \equiv ((((A^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv (A^{(1,2)})^{(2,1)})$$

and by ~~ii~~ iii) for the pair 1,2 we have $(A^{(1,2)})^{(2,1)} \equiv \underline{A}$. Combining these last three results by the rules of the propositional calculus we obtain $(((A^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv \underline{A}$ as required.

1. Th e "practical system" for a finite universe.

In this section the notation of the practical system will be explain ed. The explanation will be in terms of the 'finite universe', i.e. we start with a finite number of objects or 'individuals' and build up other entities from th ese. We can then formulate certain rules which give valid results in this case and hope that they will apply in the infinite case also. We cannot of course hope that all such rules will work. We have to imagine that many rules of this kind have been tried, found wanting and rejected, and that others are still in use. This rather unsatisfactory-sounding process is as good an account as the author feels can be given of the way in which current mathematical procedure has grown up. But ~~however all~~ whatever the truth of this may be the finite universe provides a first class ground on whoih to describe the practical system, and we proceed accordingly.

Our finite universe has initially as its members the 'individuals' $U_1 \ldots U_N$. Although these ~~are~~ include all the individuals they need not exhaust our stock-in-trade, for we can also bring in functions taking the individuals as arguments and having them also as values. With our increased range of commodities we can then go into business again and produce a still greater variety of objects, and repeat without limit. There obviously arises a ~~very~~ great variety of different kinds of function, which may need to be distinguished, but for the present system we need only trouble ourselves with th e very broadest divisions, which will be called types categories, ~~or, when differently arranged,~~ types. These divisions are described below.

To prove v) for the pair 1,3 we must prove

$$((A \equiv B)^{(1,2)})^{(2,3)} \equiv ((\underline{A}^{(1,2)})^{(2,3)} \equiv (\underline{B}^{(1,2)})^{(2,3)})$$

By an application of v) to the pair 1,2 followed by an application od i) to the pair 2,3 we get $((\underline{A} \equiv \underline{B})^{(1,2)} \equiv (\underline{A}^{(1,2)} \equiv \underline{B}^{(1,2)}))^{(2,3)}$ and by an application of v) to the pair 2,3 we have

$$((\underline{A} \equiv \underline{B})^{(1,2)} \equiv (A^{(1,2)} \equiv \underline{B}^{(1,2)}))^{(2,3)} \equiv (((\underline{A} \equiv \underline{B})^{(1,2)})^{(2,3)} \equiv (A^{(1,2)} \equiv B^{(1,2)})^{(2,3)})$$

Combining these by the propositional calculus gives

$((\underline{A} \equiv \underline{B})^{(1,2)})^{(2,3)} \equiv (\underline{A}^{(1,2)} \equiv \underline{B}^{(1,2)})^{(2,3)}$. Condition v) applied to 2,3 also gives $(\underline{A}^{(1,2)} \equiv \underline{B}^{(1,2)})^{(2,3)} \equiv ((\underline{A}^{(1,2)})^{(2,3)} \equiv (\underline{B}^{(1,2)})^{(2,3)})$ from which we now obtain the required result.

We are now in a position to enunciate theorem B.

<u>Theorem B</u>. The practical system is equivalent to the simplified type theory of Church, if the ~~proposition-like~~ propositional formulae of the practical system are taken to be the formulae without free variables, and the proposition-like formulae of Church's system are taken to be the formulae of type o without free variables.

The proofs of both theorems A and B are long and tedious, and not very difficult. Their length is due chiefly to the necessity of separate consideration of every method of forming formulae, every axiom and every rule of procedure, and in the case of theorem B the fact that there are six separate conditions to be proved.

The individuals $U_1, \ldots, U_N$ form type 0.

The functions of individuals, taking individuals as values, together with the individuals, form type 1.

The functions of arguments in type 1, with values also in type 1, together with the individuals, form type 2.

. . . .

The functions of arguments in type n, with values also in type n, together with the individuals, form type $n+1$.

. . . .

It is convenient to require functions to be defined throughout the appropriate type, i. e. not to permit such definitions as " $f(0) = 0$, but, if x is different from 0, $f(x)$ is undefined". In order to cover such cases we shall set apart from the outset a particular individual U_1, which we shall rename C, to be the value of a function in all cases where it would normally be regarded as undefined. This does not require any formal axiom: it is merely an explanation of the part which is actually played by C. We make the following conventions concerning this element C.

a) So far as possible C is on a level with all other individuals, at any rate as far as the most fundamental part of the theory is concerned. The chief violation of this principle is in c) below.

b) The function which is everywhere undefined (value C) is a normal member of type 1.

c) The value of a function is C if the argument belongs to every type to which the function belongs. In particular ~~$U(V)$~~ the value is always C if ~~V~~ the 'function' is an individual.

Certain other conventions which might appear desirable are incompatible with the above. It might for instance be considered desirable

Footnotes

1. A. N. Whitehead and Bertrand Russell, Principia Mathematica, Cambridge, England, 1925.

2. Alonzo Church, A formulation of the simple theory of types, this ~~journal~~ JOURNAL, vol. 5 (1940), pp. 56-68.

3. We shall use heavy type letters throughout to represent variable or undetermined formulae of tables. This is in the metamathematical discussions. In short we mean all our statements to be true whatever substitutions of formulae (or tables) are made for the heavy type letters.

4. There is no very simple corresponding form for $(\lambda x)\mathbf{A}$. This may have some connection with the fact that it has not been customary until recently to have a notation for abstraction.

to have a convention by which if any part of a formula was undefined then the whole is undefined. This would mean essentially that the value of any function is C if the argument is C ~~$(U(C)=C$~~ which would be a serious violation of convention a). In practise this other convention may be largely obeyed by restricting our interest mainly to functions which satisfy this condition ~~$(U(C)=C$~~ . Another suggestion might be to class the function which is everywhere undefined as an individual, on the grounds (undistributed middle!) that all individuals have this property. We have specifically excluded this.

The functions and individuals together will be known as <u>terms</u>. With our finite universe it is convenient to think of the functions as given by tables, consisting of two columns, in the first of which would appear all the necessary arguments, and opposite them in the second column the appropriate values. Thus with $N=4$ a typical member of type 1 would be represented by the table

U_2	U_3
U_1	U_1
U_3	U_1
U_4	U_4

(1)

Illegibilities (third sheet)

notation

Likewise for $(\exists x)\underline{A}$, $(\iota x)\underline{A}$ and $(\lambda x)\underline{A}$, the formulae which have to be proved in these cases being respectively ...

~~The proof,~~ It is not difficult, and is omitted.

it is hardly necessary to do so

ordinary

'class of all classes which are not members of themselves'

branch

~~there~~ outcome

there is at least one way in which

~~proof~~

Material which is written in ink, unless obviously forming words, is not to be typed, but a generous space left. The same applies to those typed letters which have been ringed in pencil. The space left in the latter case should be considerably more than actually occupied, as they will be replaced by manuscript letters.

use ~~single~~ double spacing

It would be a convenience to have the table rearranged with the first column in natural order. In ~~this~~ the case of the above table (1) we should simply have to interchange the first two rows. Such a table may be said to be in normal form. We can do this for all tables of type 1, and when we have done this we are in a position to define a natural order for the members of type 1. With both tables in normal form, the earlier table is to be the one with the earlier value in the ~~first~~ last row ~~for~~ in which the two tables differ. Thus the table (1) above precedes

U_1	U_1
U_2	U_4
U_3	U_3
U_4	U_4

(2)

since when (1) is put into normal form they differ ~~first~~ last in the ~~second~~ third row, and there (1) has the value U[illegible], but (2) has the value U_3. ~~xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx~~ We shall also adopt the convention that the individuals in type 1 precede the tables. We may now continue the numbering of terms so as to include all type 1, simply numbering them in the natural order just defined. The numbers will extend from 1 to $N + N^N$. It may be verified that the ~~numbers of the~~ above tables (1) and (2) are ~~40~~ 205 and ~~64~~ 241 respectively. A similar process may now be carried out for type 2 and then for type 3. In general when we are dealing with type n we have already numbered the members of type n-1. It is easily verified that those tables which have already appeared as members of type n-1 have the order which they had in that type, and precede all the new tables. The order of ~~the new~~ any two (new or old) tables is ~~determined by the value~~ that of the last [illegible] ~~first~~ values in which they differ.

Illegibilities , continued

It will be noticed that we have said "$(\underline{A}\underline{B})$ where $\underline{A}$ and $\underline{B}$ are term formulae" rather than " $(\underline{U}\,\underline{V})$ where $\underline{U}$ and $\underline{V}$ are tables". The distinction between term formulae and tables is not altogether a metaphysical one. One rather concrete way in which they differ is that term formulae can have free variables, but tables cannot.

If we are regarding the tables and the formulae as different sorts of object, (and it seems convenient to do so from some points of view) we should avoid using U_r both to denote a table and a formula. We shall use it as table only and call the corresponding formula U_r^H .

meaningful

, or sometimes 'Put'.

appropriate

whole

appears to have

only

i.e., more strictly, under these circumstances we may interchange $(x, r)\ \underline{P}$ and $D^r x \supset \underline{P}$ as conclusions of an inference.

It may be remarked that the 'definition of equality'... omit four lines... may be proved from the other axioms and rules. The proof is left tothe reader, with the suggestion that he should put for X ..

It may be verified that this last condition, which we have only laid down for the new tables, actually applies for any pair of tables.

Let us now introduce the notation (UV) to denote the result of looking up V in the table U: in slightly different words it is the entry against V in the table U. In other words again it is the value of the function U for the argument V, and might therefore, in agreement with current mathematical practice have been denoted by U(V). Our conventions require (UV) to be C in cases where the table gives no information: these are, just the cases where the category of U does not exceed that of V. We may also introduce the notation U= V to denote the identity of the terms U and V. It should be noticed that so long as U and V are tables known to belong to some particular type n, we can establish their identity by showing that they have the same values throughout type n-1. (This is known as the principle of extensionality and gives rise to the 'axiom of extensionality'). However the principle fails for individuals, for if U and V are individuals then (UX) is always identical with (VX), both being C, and yet U and V may well be different. The principle also fails when the types of the terms are unknown, for we can never then be sure that we have examined sufficient arguments for the functions: there may be some argument in a higher type than we have yet considered for which the two functions differ.

The expression (UV) described above always denotes a term. The expression U= V however denotes a "proposition". We might perhaps permit it the status of a term, by considering truth and falsity as particular individuals, but we shall not in fact adopt this procedure. The two kinds of expression, terms and propositions, will be described collectively as formulae.

1+ 2 carbons

Illegibilities

The following are the translations of the handwritten parts:-

Last Draft

mathematician

two

It

whatever the truth of

include

types

~~individual~~

the value
the 'function'

the value of any function is C
if ~~the~~ argument is C

which satisfy this condition

the case of the above table

last

last

third

respectively

It is easily verified that

The order of any two tables (new or old) is that of the last pair
of values in which they differ

original

constants for the variables, i.e.

of type r

of type γ

by substituting

that
of a kind that
Both bound and free variables are of common occurrence in mathematics

the variable x

that of

If P is a proposition then $(\sim P)$ denotes the negation of P, i.e. $(\sim P)$ is true is P is false and is false if P is true. If P and Q are both propositions then $(P \supset Q)$ or ~~x~~ 'P implies Q' is a proposition which is true unless P is true and Q false, in which case naturally it is false.

There are two other ways of forming propositions and one other way of forming terms. If U is a term and r is ~~ax~~ nonnegative integer, then $\mathfrak{D}^r U$ represents the proposition which states that U is of type r. ~~if~~ This is the only way of forming formulae that can really be said to be peculiar tothis system. ~~If P is a proposition then we can form the formulae of which the first is~~

Before we describe the two remaining ways of forming formule we must explain the use of formulae with variables. ~~We shall permit as formulae any expressions which~~ Let us first agree that any italic letters other than p,q,r,s,t, with any number of primes is a term variable, and that p,q,r,s,t with any number of primes are proposition varibles. Then we will permit as formulae any expressions which ~~has~~ always become formulae in the original ~~first~~ sense on substituting [illegible] for variables, etc. tables or individuals for the term variables, and propositions not containing variables for the proposition variables. Our other two ways of forming formulae apply principally to these formulae with variables. If P is a proposition formla then $(x, v)\, P$ and $(x, v)\, P$ are formulae of which the first is a propostion meaning ~~'~~ 'P is true for all x' and the second is a term with the meaning 'the x for which P is true'. These expressions really need tohave their meaining defined more carefully. We do not attempt to define a meaning for $(x, v)\, P$ itself, but rather we define the meaning of every expression that can be obtained from it by substituting constants for the variables. It is easily seen that to do this it will suffice to define the meaning of $(x,r)\ P$ in all cases where P contains no variables except ~~probably~~ possibly x (or rather probably) x itself. It is defined tobe true if substitution of any member of type r for x in P ~~gives~~ always gives a true proposition, and to be false if there is a member of type r making P false. Likewise we only need define $(\imath x,r)\ P$ in cases where P contains no variables other than x. It is defined to be C in all cases except where there is one and only one member of type r, U say, such that substituting U for x in P yields a true proposition ~~g~~ : if there is such a unique U then $(\imath x,r)\ P$ is that U.

The variables ~~in the formulae~~ which we first introduced into the formulae were all ~~what~~ of a kind that we call 'free' variables. This really means that it is permissible to substitute for them. However we may not substitute for the occurrences of x in such expressions as $(x,r)\ P$. ~~These are examples of variables are called~~ Variables occurring in this sort of way are called 'bound' variables. P.T.O. A typical case of the occurrence of a bound variable in analysis is the occurrence of x in the equation $\int_0^1 x\,dx = \frac{1}{2}$; a case of a free variable is that of x in the equation $x(x-1) = x^2 - x$. We can distinguish the two cases by trying the substitution of 1 for x.

~~In the formation of~~ We have now seven ways of forming formulae from other formulae typified by $(\cancel{UV})$ (AB), $(\imath x,r)\ P$, $\cancel{(U=V)}$ $(A=B)$, $\mathrm{D}^r A$, $(\sim P)$, $(P \supset Q)$, $(x,r)\ P$. Inthese we allow A, B to be any term formulae and P, Q to be any proposition formulae. The first two are terms and the remainder propositions. We do not restrict A, B, P, Q, in any way, e.g, we do not require them to have no bound variables. P.T.O.

These means of producing formulae with the help of other formulae get us nowhere if we have not also direct means for producing them. We have two such means. There are the tables and individuals themselves and the variables. P.T.O. Inpractice we do not use the individuals themselves much. We only give special names to three of the individuals , viz U_1^H is C, U_2^H is T and U_3^H is F. Those tables of higher type than the individuals can be referred to usually by means of some appropriate formula, e.g. [illegible]. Except inthe case of the finite universe this is practically the only means available. We cannot express by means of a table the value of a function for an infinity of arguments.

One further method of constructing formulae may be mentioned, although it is possible to do without it, defining it in terms of already explained. This is the idea of 'abstraction'. If $\underline{A}$ is a term formula, then $(\lambda x, r-1)\,\underline{A}$ is a term formula of type r. Its meaning is that it is the function whose value for the argument x is $\underline{A}$, provided that both argument and value are of type r-1: however if there should be a single argument of type r-1 giving a value not of type r-1 then the whole is undefined i.e. C. The definition of abstraction in previqusly described terms is

$$(\lambda x, r-1)\,\underline{A} \rightarrow (\iota y, r)\big((\sim D^{\circ} y) \vee ((x, r-1)(y\,x) = \underline{A})\big)$$

2. Abbreviations

In a logical system it is normally necessary to have means for abbreviating formulae or else they very soon become exceedingly cumbrous. In the present paper we shall confine ourselves to rather few forms of abbreviation, since we shall not be developing the system very far and we do not wish to become involved in ~~such~~ unnecessary technical discussions. Our forms of abbreviation are then limited to the following

a) The use of the arrow, whereby ~~[illegible]~~ we write the abbreviated form followed by an arrow and then by the full form of the formula, e.g.

$$Suc \rightarrow (\lambda n, \iota)(\lambda f, \iota)(\lambda x, o)(f(n(f)x))$$

means that Suc will be used as an abbreviation for $(\lambda n, \iota)(\lambda f, \iota)(\lambda x, o)(f(n(f)x))$. We may also use more complex expressions on the left of the arrow, ~~including~~ involving heavy type formulae: In these cases (as in b) below) it is understood that the formula on the left is an abbreviation of that on the right for any substitutions of formulae for the heavy type letters,

b) We put

$$\underline{A} \,\&\, \underline{B} \rightarrow \sim(\underline{A} \supset \sim \underline{B})$$
$$\underline{A} \vee \underline{B} \rightarrow \sim \underline{A} \supset \underline{B}$$
$$\underline{A} \equiv \underline{B} \rightarrow (\underline{A} \supset \underline{B}) \,\&\, (\underline{B} \supset \underline{A})$$
$$(\exists x, \iota)\, \underline{A} \rightarrow \sim (x, \iota) \sim \underline{A})$$

c) Formulae of form $\underline{A} \,\&\, \underline{B} \,\&\, \ldots \,\&\, \underline{P}$ we consider do not need any more brackets, since they have the same meaning whatever way the brackets are put in. Likewise for the formulae $\underline{A} \vee \underline{B} \vee \ldots \vee \underline{P}$.

3. Remarks on the deduction theorem

In symbolic logic it is usual to consider proofs to consist of series of formulae of which any one is deducible, in virtue of the rules of procedure from one or more (usually not more than two) previous formulae in the sequence, or else is an axiom. This outlook has its advantages, since it gives the most simple general form for proofs. On the other hand it has its disadvantages: actual proofs in common or garden mathematics take a slightly different form. It might be expected that these actual proofs would take a form so variable that it was profitless to try to fit them to any standard form. Fortunately this is only true to a limited extent, and there is a quite definite pattern in which they can all be fitted with only rather little idealization. According to this pattern the proof consists of a number of steps which may be classified as follows

a) Writing down an axiom

b) Writing down a hypothesis

c) Deducing a result from previous results which are to be axioms, hypotheses or results obtained by more complicated means.

d) Elimination of a hypothesis, i.e. if we have obtained a formula $\mathfrak{F}$ by the use of various hypotheses of which H is one we may consider $H \supset \mathfrak{F}$ to be a formula obtained by the use of the these hypotheses with H omitted.

The introduction of a hypothesis has the effect of converting the variables which appear free in it into constants. For instance if we make the hypothesis 'Let a be the radius of the earth', the letter 'a', which heretofore was a perfectly good variable (as good as 'x'!) must hereafter be treated as a constant and cannot be substituted for. Likewise if we make the hypothesis "Suppose $a = b^2$" we cannot substitute unrestrainedly for a, b, but we can substitute 5 for b and 25 for a because $25 = 5^2$. Actually we should make such a substitution in one step, but we may consider it as two or three, thus if the conclusion is "$a-1=(b-1)(b+1)$ the argument would be

$(a=b^2)\supset(a-1=(b-1)(b+1))$, by elimination of the hypothesis

$(25=5^2)\supset 25-1=(5-1)(5+1)$ by substitution

But $25=5^2$

Hence $25-1=(5-1)(5+1)$"

Hypotheses are very often introduced by the words 'let' or 'suppose' as in the above examples, or sometimes 'put'. Proof by reductio ad absurdum provides a very clear case of the use of a hypothesis. Such a proof goes as follows. "$\underline{P}$, for suppose not, then ... which gives a contradiction, so that $\underline{P}$ must be true". This we may interpret as follows

Take $\sim\underline{P}$ as a hypothesis

We deduce $\underline{P}'$

Eliminating the hypothesis gives $\sim\underline{P}\supset\underline{P}'$, from which $\underline{P}$ follows by the general result $(\sim\underline{P}\supset\underline{P}')\supset\underline{P}$.

Hypotheses also appear frequently when branches of mathematics are treated axiomatically. In this case all the 'axioms' are to be regarded as hypotheses which are to be eliminated only when definite applications are involved. If for instance we are treating geometry axiomatically we introduce the appropriate axioms as hypotheses, with the understanding that words such as 'point', 'line', 'intersect' (or rather, perhaps, 'the property of being a point' etc) are variables. These variables will appear free in the 'axioms' are therefore are to be treated as constants, as indeed they certainly will be.

In symbolic logic it is customary to prove a result which would justify the above technique of 'hypotheses'. This result is called the 'deduction theorem'. In the present system we have taken the bull by the horns and adopted the deduction theorem ~~xxxx~~ and the whole technique of hypotheses as fundamental ~~xx xxxxxxxxxx~~ Actually in the formal ~~xxx~~ system the advantage of this appears to have ~~xxx~~ been largely thrown away by working in terms of 'inferences' which appear to require a statement of the hypotheses in force at every line of the proof. This is ~~just~~ only a convenience for the formal ~~xxxxxxxxx~~ description of the system however. In an actual proof one just introduces and eliminates hypotheses and hopes that the reader will keep track. Sometimes this keeping track is difficult, but there need not be any difficulty in principle: occasional recapitulation of references to the hypotheses still in force should be quite adequate.

4. Formal account of the practical system.

We now describe the practical system in the usual formal manner, specifying what series of symbols are to be regarded as term-formulae, proposition formulae, variables, provable formulae etc. We do not follow this aspect very far in the present paper, believing that ~~the important thing from the~~ ~~strictly correct rules~~ ~~practical point of view is to have a sound notation, rather than sound rules of procedure~~ mathematics is suffering more from lack of sound notation than from lack of rules of procedure.

There is one small difference from the most customary form of description of such systems. We regard provable formulae as a special case of a formula deducible from hypotheses, and this latter idea is taken as fundamental. This is really only done for convenience.

Term variables. The symbols a,b,...,n,o,u,v,w,x,y,z,a',b',... are term variables.

Proposition variables. The symbols p,q,r,s,t,p',q',... are proposition variables.

Term formulae, proposition formulae, and formulae. Term variables are term-formulae. Terms $(U_1^{\mu}, U_2^{\mu}, \ldots)$ are term formulae. Proposition variables and T and F are proposition formulae. If $\underline{A}$ and $\underline{B}$ are term formulae and $\underline{P}$ and $\underline{Q}$ are proposition formulae and $\underline{x}$ is a variable and r a non-negative integer, then $(\underline{A}\ \underline{B})$ and $(\iota \underline{x}, r)\underline{P}$ are term formulae and

$\underline{A} = \underline{B}$, $(\underline{x}, r)\underline{P}$, $\underline{P} \supset \underline{Q}$, $\sim \underline{P}$, $D^{(r)}\underline{A}$

are proposition formulae. Term formulae and proposition formulae are formulae. No expression is a term variable, term formula, proposition variable, proposition formula or formula unless compelled to be so by the foregoing.

Free and bound variables. All proposition variables are free variables of the formulae in which they occur: they are never bound variables. Term variables have themselves as free variables and have no bound variables. Terms have no free or bound variables. The free and bound variables of $(\underline{A}\,\underline{B})$, $(?x, r)\underline{P}$, $(x, r)\underline{P}$, $(\underline{A} \sim \underline{B})$, $(\underline{P} \supset \underline{Q})$, $(\sim \underline{P})$, $\mathrm{U}^{x,r}\underline{A}$ are the free and bound variables of whichever of $\underline{A}$, $\underline{B}$, $\underline{P}$, $\underline{Q}$ are shown shown in them explicitly, except that in $(?x, r)\underline{P}$, and $(x, r)\underline{P}$ the variable $\underline{x}$ is bound and not free.

It may be observed that all four possible combinations are possible as regards whether a variable is bound or free in a formula. Examples are y, x, $(?x, o)\,x = x$, $x = (?x, o)\,x = x$

Deduction from hypotheses. Inferences. A sequence of symbols of the form ~~$H_1, H_2, \ldots, H_n \vdash \underline{P}$, where~~

$\underline{H}_1, \underline{H}_2, \ldots, \underline{H}_n \vdash \underline{P}$ where $\underline{H}_1, \ldots \underline{H}_n, \underline{P}$ are proposition formulae, is called an 'inference'. The formulae $\underline{H}_1, \ldots \underline{H}_n$ are called the hypotheses of the inference and $\underline{P}$ the conclusion. The set of hypotheses may be void. The rules of procedure which follow will be in the form of a definition of the 'permissible inferences'. The permissible inferences will in fact be those for which the conclusion is deducible from the hypotheses.

The rules of procedure

Rule I. If a part of a formula $\underline{A}$ is of form $(x, r)\underline{B}$ or $(\exists x, r)\underline{B}$ then we may replace that part by $(y, r) S^{x}_{y} \underline{B}|$ or $(\exists y, r) S^{x}_{y} \underline{B}|$ provided that $\underline{y}$ is not a free variable of $\underline{B}$. More strictly we should say that if $\underline{F}$ is the resulting formula and $\underline{H}_1, \dots \underline{H}_n \vdash \underline{A}$ is permissible then $\underline{H}_1, \dots \underline{H}_n \vdash \underline{F}$ is permissible.

Rule II. If a ~~any~~ substitution of proposition formulae for free proposition variables or of term formulae ~~variables~~ for term free ~~formal~~ variables is made in the conclusion of a permissible inference, a permissible inference results, provided that the variables in question do not occur free in any of the hypotheses.

Rule III. If $\underline{P}$ is the conclusion of a permissible inference I_1 and $\underline{P} \supset \underline{Q}$ is the conclusion of a permissible inference I_2, then $\underline{Q}$ is the conclusion of a permissible inference whose hypotheses are those of I_1 and of I_2 combined in any order.

Rule IV. From $D^r \underline{x} \supset \underline{P}$ we may deduce $(x, r)\underline{P}$ if $\underline{x}$ is not free in any hypothesis, and conversely from $(x, r)\underline{P}$ we may infer $D^r \underline{x} \supset \underline{P}$.

Rule V. We may use the deduction theorem, i.e. if $\underline{H}_1, \underline{H}_2, \dots \underline{H}_n \vdash \underline{P}$ is permissible then so is $\underline{H}_1 \dots \underline{H}_{n-1} \vdash \underline{H}_n \supset \underline{P}$

Rule VI. $\vdash \underline{P}$ is permissible if $\underline{P}$ is a provable formula of the propositional calculus, i.e. if it is a propositional formula built up entirely from propositional variables with the aid of brackets and $\supset$ and $\sim$, and such that whenever substitutions of T or F are made for each variable, and the result evaluated by the rules that $T \supset F$ is F etc we are finally left with the value T.

<u>Rule VII</u>. The following are permissible inferences for any propositional formula $\underline{P}$ and non-negative integer r.

$\vdash D^r x \sim \underline{P} \sim (y,r)\{S^x_y \underline{P} \mid \supset y \sim x\} . \supset . (\iota x, r)\underline{P} = x$

$\vdash (x,r) \sim \underline{P} \vee (\exists x)(\exists y). x \neq y \sim \underline{P} \sim S^x_y \underline{P} \mid . \supset (\iota x, r)\underline{P} = C$

provided y not free in $\underline{P}$.

<u>Rule VIII</u>. The following is a permissible inference for any ~~propositional~~ term formula $\underline{A}$ and non-negative integer r.

$\vdash (x,r) D^r \underline{A} . \supset (\exists y, r+1). (x,r)\, yx \sim \underline{A} \sim \sim D^0 y$

We may make one or two remarks about these rules. ~~Therefore~~ By means of VII we ~~are~~ have $\vdash \underline{P} \supset \underline{P}$, and this with III enables us to change the order of brackets. The second assertion under Rule VII might have been omitted. If we did so we should have to define a new description operator in terms of the original one in such a way that this second assertion applied for the new operator. It may be wondered why VII and VIII do not appear under the axioms instead of the rules, writing $yx \sim \overline{T}$ and yx for $\underline{P}$ and $\underline{A}$. In general this would be possible, if there were any further rules of this nature involved, but rules VII, VIII are themselves involved when we try to apply such axioms to get the effect of the corresponding rule.

Rule II. For any non-negative integers r, p ~~[illegible]~~ the following formulas ~~[illegible]~~ A1–C3, known as axioms, are ~~the conclusions of admissible [illegible] formulas without hypothesis (in short, we can just write the form)~~ provable.

A1. $\vdash C \neq T' \neq C \neq F' \neq T' \neq F' \neq D^0 C \neq D^0 T \neq D^0 F$

A2. $\vdash (x, r) D^r x$

A3.$^{(r)}$ $\vdash \sim D^{r-1} x \neq D^r y \supset yx = C$ A3$^{(0)}$ $D^0 y \supset yx = C$

A4. $\vdash D^r x \supset D^{r+1} x$

A5. $\vdash D^r x \supset D^{r-1} xy$

A6. $\vdash (y, r-1)\, D^{r-1} xy . \neq D^p x \neq (y, p-1) . \sim D^{r-1} y \supset xy = C : \supset D^r x$

B1. $\vdash x = x$

B2. $\vdash x = y \supset y = x$

B3. $\vdash x = y \neq y = z \supset x = z$

B4. $\vdash x = y \supset . zx = zy \neq xz = yz$

C1. $\vdash (f, r+1)(g, r+1) : (x, r)\, fx = gx . \supset : f = g \vee D^0 f \vee D^0 g$

(Axiom of Extensionality)

~~C2. $\vdash (f, r)(g, r) : (x, r+1)\, xf = xg . \supset f = g \vee f = C \vee g = C$~~

~~(Definition of equality)~~

C2 ~~C3~~. $\vdash (\exists i, r+2)(f, r+1) : (\exists x, r)\, fx = T' . \supset f . if = T'$

Rule IX. The following, the axiom of infinity, is provable

C3 ~~C4~~. $\vdash (\exists t, 1)(\exists 0, 0)(x, 0)(y, 0) : tx = ty \supset x = y . \neq 0 \neq tx$

When we have a finite universe instead of an infinite one we use

Rule I_{II}. The following D1-D6 are permissible.

D1. $\vdash \mathfrak{D}^r x \supset f = \mathfrak{D} x j \quad x = U_n^H$
U_n in Type r

D2. $\vdash \sim (U_n^H = U_m^H)$, if $n \neq m$

D3. $\vdash (U_m^H\, U_n^H) = U_k^H$, if the result of looking up U_m in U_n is U_k

D4. $\vdash \mathfrak{D}^r U_m^H$, if U_m is in Type r

D5. $c = c . \equiv T^H : \equiv : \sim T^H \equiv F^H$

It may be remarked that the 'definition of equality'

$$(f, r)(g, r)\left[((x, r\cdot 1)(xf = xg)) \supset (f = g)\right]$$

may be proved from the other axioms and rules. The proof is left to the reader, with the suggestion of putting for x

$$(\lambda k, r)(k, 0)(y, r)\left[(ky = gy \supset k = T) \vee (ky = gy \supset k = F)\right]$$

We can use T' and F' to represent the true and the false proposition, e.g. by putting

$T' \rightarrow c = c$
$F' \rightarrow \sim T'$

5. Relaxation of type ~~theory~~ notation

The form of type theory which we have described is one in which the types themselves do not intrude very much. Even so they do intrude still to an undesirable extent, and it would be desirable to see how much further they can be relegated to the background. A possible way of doing this will be described in this section.

We have agreed that expressions of the forms $(\underline{x})\,\underline{A}$, $(\exists\underline{x})\,\underline{A}$, $(\iota\underline{x})\,\underline{A}$, $(\lambda\underline{x})\,\underline{A}$ are not normally to be regarded as meaningful. However in a very large class of cases we can assign meanings to them in a satisfactory manner. These cases and the appropriate interpretation may be summed up in the following rule:-

If the formulae $(\underline{x},r)\,\underline{A} \equiv (\underline{x},r+1)\,\underline{A}$ are all provable ~~xxxxx~~ when $r \geqslant r_0$ for some r_0, then $(\underline{x})\underline{A}$ may be interpreted as meaning any one of $(\underline{x},r)\,\underline{A}$ with $r \geqslant r_0$.

P.T.O.

It might appear that this rule does not lead to a unique interpretation owing to the doubt in the value of r. Any doubt of this kind concerning the interpretation is however rendered ineffective by the following theorem:-

Let the variable $\underline{u}$ occur only once in $\underline{F}$ and there be free, then for any formulae $\underline{A}$, $\underline{B}$ we can prove $(\underline{A} == \underline{B}) \supset (S^{\underline{u}}_{\underline{A}}\,\underline{F}| == S^{\underline{u}}_{\underline{B}}\,\underline{F}|)$. Here $\underline{A} == \underline{B}$ means $\underline{A} = \underline{B}$ if $\underline{A}$ and $\underline{B}$ are term formulae and $\underline{A} \equiv \underline{B}$ if they are proposition formulae.

~~The proof of this~~ is [illegible] and [illegible]

The proof of this is by induction over the length of $\underline{F}$.
~~It is straightforward and can be left to the reader.~~

The most natural cases where we can apply the above rule are those of $(x).\underline{A} \supset \underline{B}$, $(\exists x).\underline{A} \cdot \underline{B}$, $(\iota x).\underline{A} \cdot \underline{B}$, where $\underline{A} \supset D^{r_0} x$ is provable for some r_0 . It is ~~xxxxxxxx~~ fairly easy to remember which are the most important expressions $\underline{A}$ of this kind: it is not so easy to remember the appropriate numbers r_0 but ~~this~~ is hardly necessary ~~xxxxxxxxxx~~ anyway if the ~~xxxxxxxxxxx~~ ~~xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx~~ notations $(x)\underline{A}$ etc. are adhered to throughout. When $\underline{A} \supset D^{r_0} x$ I shall call the class of x for which $\underline{A}$ is true a 'noun-class'. There is a very close connection between the rôle that these formulae play and that nouns (in ~~any~~ ordinary language), so much so that one ~~xxxx~~ tends to feel that Russell's type theory was largely anticipated by prehistoric man. This connection may be seen by translating $(x).\underline{A} \supset \underline{B}$, $(\exists x).\underline{A} \cdot \underline{B}$, $(\iota x).\underline{A} \cdot \underline{B}$ roughly as 'All $\underline{A}$ ~~xxx~~ satisfy $\underline{B}$', 'There exists an $\underline{A}$ satisfying $\underline{B}$' and 'The $\underline{A}$ which satisfies $\underline{B}$'. In each case $\underline{A}$ is translated in the form of a noun. It seems that the necessity to use nouns ~~xxxxx~~ automatically prevents us from committing type fallacies in common speech. We can probably only break down ~~xxxxxxxx~~ this 'safety device' by using nouns ~~such~~ as 'thing' or 'object', ~~xxxx~~ with the intended meaning 'anything whatever'. In the case of the Russell paradox [illegible] we use the word 'class' in very much that way. We use it to mean 'class of anythings whatever'.

~~The outcome of the matter at the present stage then is this. In order to make mathematics conform with type theory it is necessary to be able to recognise 'noun-classes' quickly~~

Using our new interpretation rule let us look again at the rules of procedure and the axioms. Can we reformulate them in such a way as to avoid the references to types? We can. The following alterations should be made

~~xxx~~ Rule I. Replace $(y,r)\ldots$ by $(y)\ldots$ etc.

Rule II. The formulae substituted must be interpretable.

Rule IV. To read: from $\underline{P}$ we may infer $(x)\underline{P}$ and conversely, provided $(x)\underline{P}$ is interpretable.

Rule VII. The formulae are to read

$$\vdash \underline{P} + (y)\left[S^y_x \underline{P} / \supset y = x\right] . \supset . (\imath x)\underline{P} = x$$

$$\vdash (x) \sim \underline{P} \vee (\exists x)(\exists y)\, \underline{P} + S^x_y \underline{P} / \sim x \neq y \; . \supset (\imath x)\underline{P} = C$$

and the reference to the integer r is to be omitted. The formulae have to be interpretable.

Rule VIII. As with Rule VII, and the formula is to read

$$\underline{B}\underline{A} = \overline{T} \supset . (\exists y) . \sim \mathfrak{D}^n_y + (x)\left[\underline{B}x = \overline{T} \supset yx = \underline{A}\right]$$

It may be observed that the 'noun-classes' are those classes of the form 'the class of those x for which $\underline{A}x = \overline{T}$'. This fact does not ~~xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx~~ immediately enable us to dispense with the fundamental noun-classes $\mathfrak{D}^r$, but we can do so in this way. Let us put

$$(\mathrm{Pot}\,\underline{A}\,\underline{B}) \rightarrow (\imath y)(x) . yx = \overline{T} \vee yx = C + . yx = \overline{T} \equiv (z) . \underline{B}z = \overline{T} \supset \underline{A}xz = \overline{T}$$

then we can prove that the description in $\mathrm{Pot}\,\underline{A}\,\underline{B}$ is always interpretable. ~~xxxxxxxxxxxxxxxxxxx~~ Having once proved this fact we can ~~xxxxxxxx~~ make use of it to construct a great variety of interpretable formulae. Likewise we can show that $\mathrm{Sum}\,\underline{A}\,\underline{B}$ is also an interpretable where $\mathrm{Sum}\,\underline{A}\,\underline{B} \rightarrow (\imath y)(x) . yx = \overline{T} \vee yx = C + . yx = \overline{T} \equiv . \underline{A}x = \overline{T} \vee \underline{B}x = \overline{T}$

These are adequate for our purposes since we can effectively define D^r in terms of D^0, Pot, Sum. Let us put

$$E^r \rightarrow (\iota_1, r+1)(x) : D^r x \supset y\, x = T \,.\vee.\, \sim D^r x \supset y\, x = C$$

then

$$E^{r+1} = \mathrm{Sum}(E^0, \mathrm{Pot}\, E^r E^r)$$

If we choose to adopt this attitude the axioms A may be omitted, except in so far as they relate to D^0, i.e. we retain A1, A2^0. We also retain the axioms B. Axiom C1 becomes

~~[illegible]~~

$$\mathrm{Pot}\, \rho y\, f = T \,.\equiv.\, \mathrm{Pot}\, \rho y\, g = T \,.\equiv.\, (x).\, g x = T \supset f x = g x \,:\supset f = g$$

~~(Axiom C3 is not of this ... separable from the ...)~~. Axiom C2 ~~[illegible]~~ may take the form

$$(y)(\exists z) : f \neq C \supset y f = T \,.\equiv.\, (f).\, y f = T \supset (x).\, f x = T \supset f\, x = T$$

Axiom C4 may take the form

$$(\exists r)\, E^r b = T \,.\equiv.\, (\exists n)\, E^n b = T \,.\equiv.\, (x)(y)\, E^n x = T \,.\, E^n y = T$$

$$\supset : b x = b y \supset x = y \,.\vee.\, a \neq b x$$

The above suggestion provides one possible way in which mathematics might be brought strictly into line with type theory without any very drastic alterations from present practice: in particular it would be unnecessary to keep track of what types the various formulae have. However in the case of the integers for instance, we should probably not wish to commit ourselves as to just what the integers were, whether they were classes of similar classes, or functions describing how often another function is to be iterated, or what. Our procedure in such a case would be to take the Peano axioms as a hypothesis. This hypothesis would take the form of a formula involving a variable ι which was afterwards to be interpreted as defining the integers, these being those x for which $\iota x = T$. In this way the hypothesis would effectively include the assumption that the class of integers was a 'noun class'.

~~xxxxxxxx~~ This particular suggestion ~~xxxx~~ is of course by no means the only feasible one, and since mathematics is unlikely to convert itself into a branch ~~form~~ of symbolic logic it is to be expected that a combination of techniques is the most likely outcome. However ~~the author considers it to be~~ it is of importance that it is possible to obey type theory without serious inconvenience. branch

there is at least one way in which

It would be natural at this point to illustrate the ideas involved by examples of well-known mathematical results expressed in the manner recommended above. There are two reasons why this is not being done. In the first place, as has been mentioned, it is not considered desirable to convert mathematics into a branch of symbolic logic: ~~xxxxxxxxx~~ examples of the kind desired would therefore consist of long passages of very conventional nature, with occasional slight variations from the conventional pattern. Secondly, it is the author's intention shortly to make a very comprehensive review of current mathematical notation, which will involve the discussion of such examples. In these discussions there will be many other points of interest, but the type aspect will also be included: the present moment is therefore thought premature for the consideration of examples.

with the author's compliments.

A. E. Ingham

B 3.

THE JOURNAL OF SYMBOLIC LOGIC
Volume 13, Number 2, June 1948

PRACTICAL FORMS OF TYPE THEORY

A. M. TURING

Russell's theory of types,[1] though probably not providing the soundest possible foundation for mathematics, follows closely the outlook of most mathematicians. The present paper is an attempt to present the theory of types in forms in which the types themselves only play a rather small part, as they do in ordinary mathematical argument. Two logical systems are described (called the "nested-type" and "concealed-type" systems). It is hoped that the ideas involved in these systems may help mathematicians to observe type theory in proofs as well as in doctrine. It will not be necessary to adopt a formal logical notation to do so.

1. The nested-type system for a finite universe. In this section the notation of the nested-type system will be explained. The explanation will be in terms of the 'finite universe,' i.e. we start with a finite number of objects or 'individuals' and build up other entities from these. We can then formulate certain rules which give valid results in this case and hope that they will apply in the infinite case also. We cannot of course hope that all such rules will work. We have to imagine that many rules of this kind have been tried, found wanting and rejected, and that others are still in use. This rather unsatisfactory-sounding process is as good an account as the author feels can be given of the way in which current mathematical procedure has grown up. But whatever the truth of this may be the finite universe provides a first class ground on which to describe the nested-type system, and we proceed accordingly.

Our finite universe has initially as its members the 'individuals' $U_1, \cdots, U_N$. Although these include all the individuals, they need not exhaust our stock-in-trade, for we can also bring in functions taking the individuals as arguments and having them also as values. With our increased range of commodities we can then go into business again and produce a still greater variety of objects, and repeat without limit. There obviously arises a great variety of different kinds of functions which may need to be distinguished, but for the present system we need only trouble ourselves with the very broadest divisions, which will be called types. These divisions are described below.

The individuals $U_1, \cdots, U_N$ form type 0.

The functions of individuals, taking individuals as values, together with the individuals themselves, form type 1.

The functions of arguments in type 1, taking values also in type 1, together with the members of type 1, form type 2.

. . . .

The functions of arguments in type n, taking values also in type n, together with the members of type n, form type $n + 1$.

. . . .

Received January 6, 1947.

[1] A. N. Whitehead and Bertrand Russell, ***Principia mathematica***, Cambridge, England, 1925.

It must be understood that by a "function" we mean the function itself and not merely one of its values. To illustrate the point by analogy with functions of a real variable, we should say that "sin" denotes a function, but that "sin 0.3" and "sin x" do not, although the latter is often used (incorrectly in the author's opinion) as if synonymous with "sin".

It is convenient to require functions to be defined throughout the appropriate type, i.e. not to permit such definitions as "$f(0) = 0$, but if x is different from 0 then $f(x)$ is undefined." In order to cover such cases we shall set apart from the outset a particular individual U_1, which we shall rename "C", to be the value of a function in all cases where it would normally be regarded as undefined. So far as possible we try to keep C on a par with the other individuals. We deviate from this principle by adopting the convention that the value of a function is always C unless the function is of higher type than the argument. (More strictly, if the function belongs to every type to which the argument belongs.) We respect the principle by refraining from considering every expression containing "C" to have the value C.

The functions and individuals together will be known as *terms*. With our finite universe it is convenient to think of the functions as given by tables, consisting of two columns, in the first of which appear all the necessary arguments, and opposite them in the second column the appropriate values. Thus with $N = 4$ a typical member of type 1 would be represented by the table

$$\text{(1)} \qquad \boxed{\begin{array}{cc} U_2 & U_3 \\ U_1 & U_1 \\ U_3 & U_1 \\ U_4 & U_4 \end{array}}$$

It would be a convenience to have the table rearranged with the first column in natural order. In the case of the above table (1) we should simply have to interchange the first two rows. Such a table may be said to be in normal form. We can do this for all tables of type 1, and when we have done so we are in a position to define a natural order for the members of type 1. With both tables in normal form, the earlier table is to be the one which has the earlier value in the last row in which the two tables differ. Thus the table (1) above precedes

$$\text{(2)} \qquad \boxed{\begin{array}{cc} U_1 & U_1 \\ U_2 & U_4 \\ U_3 & U_3 \\ U_4 & U_4 \end{array}}$$

since when (1) is put into normal form the two tables differ last in the third row, and there (1) has the value U_1 but (2) has the value U_3. We shall also adopt the convention that the individuals in type 1 precede the tables. We may now continue the numbering of terms so as to include all type 1, simply numbering them in the natural order just defined. The numbers will extend from 1 to $N + N^N$. It may be verified that the above tables (1) and (2) are U_{205} and U_{241} respectively. A similar process may now be carried out for type 2 and then for type 3. In general when we are dealing with type n we have already numbered the members of type $n-1$. It is easily verified that those tables which have already appeared as members of type $n-1$ have the order which they had in that type, and precede all the new tables. The order of any two tables (new or old) is that of the last pair of values in which they differ.

Let us now introduce the notation $(\boldsymbol{UV})$ to denote the result of looking up $\boldsymbol{V}$ in the table $\boldsymbol{U}$; in slightly different words it is the entry against $\boldsymbol{V}$ in the table $\boldsymbol{U}$.[2] In other words again it is the value of the function $\boldsymbol{U}$ for the argument $\boldsymbol{V}$, and might therefore, in agreement with current mathematical practice have been denoted by $\boldsymbol{U}(\boldsymbol{V})$. Our conventions require $(\boldsymbol{UV})$ to be C in cases where the table gives no information: these are just the cases where the lowest type to which $\boldsymbol{U}$ belongs does not exceed the lowest for $\boldsymbol{V}$. We may also introduce the notation $\boldsymbol{U} = \boldsymbol{V}$ to denote the identity of the terms $\boldsymbol{U}$ and $\boldsymbol{V}$. It should be noticed that so long as $\boldsymbol{U}$ and $\boldsymbol{V}$ are tables known to belong to some particular type n we can establish their identity by showing that they have the same values throughout type $n-1$ (this is known as the principle of extensionality and gives rise to the "axiom of extensionality"). The principle fails for individuals, for if $\boldsymbol{U}$ and $\boldsymbol{V}$ are individuals then $(\boldsymbol{UX})$ is always identical with $(\boldsymbol{VX})$, both being C, and yet $\boldsymbol{U}$ and $\boldsymbol{V}$ may well be different. The principle also fails when the types of the terms are unknown, for we can never then be sure that we have examined sufficient arguments for the functions. There may be some argument in a higher type than we have yet considered for which the two functions differ.

The expression $\boldsymbol{U} = \boldsymbol{V}$ which we have just introduced denotes a *proposition*, unlike $(\boldsymbol{UV})$ which was a term. Propositions may be thought of as having a value which is either truth (T) or falsity (F). By taking T and F to be individuals we could have arranged for the propositions to be included amongst the terms, but we have not in fact done so.

There are several other ways of forming propositions. If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are propositions then $(\sim\boldsymbol{P})$ is a proposition whose value is opposite to that of $\boldsymbol{P}$ and $(\boldsymbol{P} \supset \boldsymbol{Q})$ is one whose value is F if and only if $\boldsymbol{P}$ is T and $\boldsymbol{Q}$ is F. We may read $(\sim\boldsymbol{P})$ as "not $\boldsymbol{P}$" and $(\boldsymbol{P} \supset \boldsymbol{Q})$ as "$\boldsymbol{P}$ implies $\boldsymbol{Q}$." If $\boldsymbol{U}$ is a term then $D^r\,\boldsymbol{U}$ represents the proposition that $\boldsymbol{U}$ is in type r, i.e. it is T if and only if $\boldsymbol{U}$ is in type r.

We could of course introduce a great variety of further means for forming terms and propositions. We could for instance define $(\boldsymbol{P}\ \&\ \boldsymbol{Q})$ as a proposition whose value is T if and only if both $\boldsymbol{P}$ and $\boldsymbol{Q}$ are T. We shall be content however with comparatively few, namely those we have already introduced, together with one further way of forming propositions and one of forming terms. These cannot be described without bringing in the ideas of "variable" and "formula with variables." Variables are of little importance except as parts of formulas. All we need say about them is that as a matter of notation small italic letters with any number of primes will be used as variables. *The letters p, q, r, s, t,*

[2] We shall use heavy type letters throughout to represent variables or undetermined formulas or tables. They occur only in metamathematical discussions. All our statements are understood to be true whatever substitutions of formulas (or tables, as the case may be) are made for the heavy type capital letters, and whatever substitutions of variables are made for the small heavy type letters.

(possibly with primes) will be proposition variables and the others term variables. Small heavy type letters may be used to stand for any variable, with an obvious convention concerning the kind of variable. An example of a "formula with variables" is the expression $x = U_5$. On substituting a term, e.g. U_{10} for the the term variable x it becomes a proposition. Similarly $(U_{405}\,x)$ is a formula with variables: in this case substitution yields a term. In general a formula with variables or more briefly a *formula* is an *expression which yields a term or proposition on substituting terms and propositions for the (free) term and proposition variables respectively*. The formulas may be called *term formulas* or *proposition formulas* according as they give rise to terms or propositions on substitution. The word *free* in the definition should be ignored for the present.

We can now describe our remaining ways of forming terms and propositions. If $\boldsymbol{P}$ is a proposition formula with only the one free term variable $\boldsymbol{x}$ and no proposition variables then $(\iota\boldsymbol{x}, r)\ \boldsymbol{P}$ is a term and $(\boldsymbol{x}, r)\ \boldsymbol{P}$ is a proposition. Of these the term $(\iota\boldsymbol{x}, r)\ \boldsymbol{P}$ has the value C unless there is one and only one term $\boldsymbol{U}$ in type r for which the result $S^{\boldsymbol{x}}_{\boldsymbol{U}}\boldsymbol{P}\,|$ of substituting $\boldsymbol{U}$ for $\boldsymbol{x}$ in $\boldsymbol{P}$ is T: if there is a unique $\boldsymbol{U}$ with this property then the value of $(\iota\boldsymbol{x}, r)\ \boldsymbol{P}$ is that $\boldsymbol{U}$. The value of the proposition $(\boldsymbol{x}, r)\ \boldsymbol{P}$ is T if and only if all the results of substitution, $S^{\boldsymbol{x}}_{\boldsymbol{U}}\boldsymbol{P}\,|$, with $\boldsymbol{U}$ in type r, have the value T. We may read $(\iota\boldsymbol{x}, r)\ \boldsymbol{P}$ as "the $\boldsymbol{x}$ in type r such that $\boldsymbol{P}$" and $(\boldsymbol{x}, r)\ \boldsymbol{P}$ as "$\boldsymbol{P}$, for all $\boldsymbol{x}$ in type r."

Now consider the expression $(x, 3)(x = y)$. In it there occur the two variables x and y. If we substitute a term, e.g. U_6, for y we shall obtain a proposition, but if at the same time we substitute U_9 for x we shall obtain nonsense. We would like to excuse ourselves from making this second substitution and admit $(x, 3)(x = y)$ to membership of the class of formulas. Our excuse is that substitution should only be made for the *free* occurrences of a variable, and that the occurrences of x in $(x, 3)(x = y)$ are not free but bound. We say that a variable $\boldsymbol{u}$ occurs *bound* in a formula if the occurrence in question is in a part of form $(\iota\boldsymbol{u}, r)\boldsymbol{P}$ or $(\boldsymbol{u}, r)\boldsymbol{P}$. Thus the first occurrence of x in $(y, 1)\ [x = (\iota x, 0)\ (x = x)]$ is free and the others are bound. This expression is a proposition formula according to our definition. To verify this, first note that $x = x$ is a proposition formula with no free variables other than x and that $(\iota x, 0)\ (x = x)$ is therefore a term. Consequently $\boldsymbol{U} = (\iota x, 0)\ (x = x)$ is a proposition, and *a fortiori* a proposition formula, for any term $\boldsymbol{U}$. It has no free variables orher than y (indeed it has none at all), and therefore $(y, 1)\ [\boldsymbol{U} = (\iota x, 0)\ (x = x)]$ must be a proposition for any term $\boldsymbol{U}$, i.e. $(y, 1)\ [x = (\iota x, 0)\ (x = x)]$ is a proposition formula.

It will now be seen that terms and propositions are just term formulas and proposition formulas without free variables.

Free and bound variables are familiar in mathematics though they are seldom consciously recognized. A typical example of a bound variable is that of x in the integral $\int_0^1 x\,dx$; x occurs free in the equation $x(x-1) = 0$. A convenient method of distinguishing between bound and free variables is to make a substitution of a constant (of the appropriate kind) for the variable in question. If nonsense results the variable is certainly bound: if sense results it is most probably free. Sense may perhaps result from substitution for a bound variable if the result of the substitution and the original expression are interpreted according to different conventions. The double suffix summation convention of tensor theory provides an example of this. Using this convention the variable j in the expression $a_{ij}b_{jk}$ is bound, but we can substitute 1 for j and obtain a perfectly sensible expression; it is sensible because it is interpreted without applying the double suffix convention.

The outcome of our definition of "formulas" is that they will include terms, propositions, and variables. Also if $\mathbf{A}$ and $\mathbf{B}$ are term formulas, $\mathbf{P}$ and $\mathbf{Q}$ proposition formulas, $\mathbf{x}$ a term variable, and r a numeral representing a non-negative integer, then $(\mathbf{A}\ \mathbf{B})$ and $(\iota\mathbf{x}, r)\mathbf{P}$ are also term formulas and $(\mathbf{A} = \mathbf{B})$, $D^r\ \mathbf{A}$, $(\sim\mathbf{P})$, $(\mathbf{P} \supset \mathbf{Q})$, and $(\mathbf{x}, r)\mathbf{P}$ are proposition formulas. Our use of the letter "r" in these cases must not of course be confused with its use as a proposition variable. One further method of constructing formulas is worth mentioning although it is possible to do without it, and define it in terms already explained. This is "abstraction." If $\mathbf{A}$ is a term formula then $(\lambda\mathbf{x}, r)\,\mathbf{A}$ is a term formula of type $r + 1$. It stands for the function whose value for the argument $\mathbf{U}$ in type r is $S^{\mathbf{x}}_{\mathbf{U}}\mathbf{A}\,|$, provided that $S^{\mathbf{x}}_{\mathbf{U}}\mathbf{A}\,|$ is in type r for every $\mathbf{U}$ in type r: if however there is a single argument $\mathbf{U}$ in type r for which $S^{\mathbf{x}}_{\mathbf{U}}\mathbf{A}\,|$ is not in type r then $(\lambda\mathbf{x}, r)\,\mathbf{A}$ is C. We can define $(\lambda\mathbf{x}, r)\,\mathbf{A}$ in previously explained terms as

$$(\iota\mathbf{y}, r + 1)\ (\sim[(\mathbf{x}, r)\ (\mathbf{yx} = \mathbf{A}) \supset D^0\mathbf{y}])$$

where $\mathbf{y}$ is any variable not occurring free in $\mathbf{A}$.

In the case of a finite universe the individuals $U_1, \ldots, U_N$ form a part of the system. When dealing with an infinite universe this does not seem to be necessary, but it is convenient to retain symbols for three of them; these are U_1 which is called C and which we have already mentioned, U_2 which is called T' and U_3 which is called F'. These last two may be regarded as unofficial representatives of truth and falsity, looking after their interests amongst the terms: their official representatives are T and F which are propositions. The chief use of T' and F' is in connection with propositional functions. If we wish to express 'x is mortal' we form a function M which is defined for individuals (supposed to include mammals) and has the value T' for mortal arguments, F' for immortal arguments. Then "x is mortal" is written as $Mx = T'$.

At this point we should pause and consider what we have done. We have defined a class of expressions which we have called term-formulas and proposition-formulas, and which roughly correspond to the terms and propositions of mathematics. These formulas are given interpretations in the finite universe in terms of individuals and tables. Each term formula without free variables has an interpretation as a particular individual or table, and each proposition formula has an interpretation which is truth or falsity. We are able to determine whether a proposition formula without free variables is true by working out its interpretation, although this will be a very lengthy business unless the formula is very simple and N very small. The work involved in establishing the truth of formulas can be greatly reduced by the use of various rules, e.g. that if two formulas $\mathbf{P}, \mathbf{Q}$ are true then $\sim(\mathbf{P} \supset \sim\mathbf{Q})$ is true. A process of application of such rules may be allowed to oust the process of working out the interpretation.

Since the majority of the rules involved do not make any reference to the number N it is easy to forget the finite universe, and to allow the various rules to become reflex action. Eventually we break off almost all connection with the finite universe picture: in particular we repudiate such propositions as

$$(x, r)(y, r)((x \neq y) \supset ((fx) \neq (fy))) \supset (x, r)\ (\exists y, r)((fy) = x)$$

which are especially connected with such a picture. Finally we even repudiate the picture more violently by adopting an "axiom of infinity."

This, in my opinion, is a very idealised but essentially correct account of how the present mathematical argument-form has grown up. The last step or two may appear very lame, but I think this cannot be helped: I think that these last steps are not really sound.

One set of rules which can replace the finite universe picture is given below in §2 (rules I–X, XI_n).

Abbreviations. At this point we are obliged to introduce a few conventions which permit us to abbreviate our formulas. The unabbreviated formulas would be disagreeably cumbrous.

(a) We may introduce abbreviations by means of the arrow: a formula standing to the left of an arrow is understood to be an abbreviation of that on the right of it. If heavy type letters appear in these expressions it is understood that the formula on the left is an abbreviation of that on the right for any meaningful substitutions of formulas for the heavy-type letters. With these conventions we introduce the abbreviations:

$$(\mathbf{P}\ \&\ \mathbf{Q}) \rightarrow (\sim(\mathbf{P} \supset (\sim\mathbf{Q})))$$

$$(\mathbf{P} \vee \mathbf{Q}) \rightarrow ((\sim\mathbf{P}) \supset \mathbf{Q})$$

$$(\mathbf{P} \equiv \mathbf{Q}) \rightarrow ((\mathbf{P} \supset \mathbf{Q})\ \&\ (\mathbf{Q} \supset \mathbf{P}))$$

$$(\exists\mathbf{x}, r)\mathbf{P} \rightarrow (\sim((\mathbf{x}, r)(\sim\mathbf{P})))$$

$$(\exists!\mathbf{x}, r)\mathbf{P} \rightarrow ((\exists\mathbf{x}, r)\mathbf{P}\ \&\ (\mathbf{x}, r)(\mathbf{y}, r)(\mathbf{P} \neq S^{\mathbf{x}}_{\mathbf{y}}\mathbf{P}\,| \supset \mathbf{x} = \mathbf{y}))$$

$$(\mathbf{A} \neq \mathbf{B}) \rightarrow (\sim(\mathbf{A} = \mathbf{B}))$$

$$T \rightarrow (x, 0)x = x$$

$$F \rightarrow (\sim T)$$

The variable $\mathbf{y}$ must not be free in $\mathbf{P}$.

(b) Formulas of form $\mathbf{A}\ \&\ \mathbf{B}\ \&\ \ldots\ \&\ \mathbf{P}$ we consider not to need any more brackets, since they have the same meaning in whatever manner the brackets are put in. Strictly speaking this equivalence only applies in virtue of rule IV below, and the reader may prefer to adopt some definite convention of his own as to the way the missing brackets are to be supplied. Similar considerations apply to formulas of form $\mathbf{A} \vee \mathbf{B} \vee \ldots \vee \mathbf{P}$.

(c) We shall often leave brackets out in cases where it is quite obvious how they should be replaced. Excessive bracketing often makes the formulas difficult to read. It is not thought worth while to introduce definite conventions in the present paper: we rely on common sense instead. Likewise we permit alterations in the form of a pair of brackets. These common sense conventions have already been applied to some extent.

2. Formal account of the nested-type system. We now describe the practical system in the usual formal manner, specifying what series of symbols are to be regarded as term-formulas, proposition formulas, variables, provable formulas, etc. We do not follow this aspect very far in the present paper, believing that mathematics is suffering more from lack of sound notation than from lack of rules of procedure.

Term variables. The symbols $a, b, \ldots, n, o, u, v, w, x, y, z, a', b', \ldots$ are term variables.

Proposition variables. The symbols $p, q, r, s, t, p', q', \ldots$ are proposition variables.

Term formulas, proposition formulas, and formulas. Term variables are term-formulas. Terms $(U_1^H, U_2^H, \ldots)$ are term formulas. Prosposition variables are proposition formulas. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are term formulas and $\boldsymbol{P}$ and $\boldsymbol{Q}$ are proposition formulas and $\boldsymbol{x}$ is a term-variable and r a numeral representing a non-negative integer, then $(\boldsymbol{AB})$ and $(\iota\boldsymbol{x}, r)\boldsymbol{P}$ are term formulas and $(\boldsymbol{A} = \boldsymbol{B})$, $(\sim\boldsymbol{P})$, $(\boldsymbol{P} \supset \boldsymbol{Q})$, $D^r\boldsymbol{A}$, $(\boldsymbol{x}, r)\boldsymbol{P}$ are proposition formulas. Term formulas and proposition formulas are formulas. No expression is a term variable, term formula, proposition variable, proposition formula, or formula unless compelled to be so by the foregoing.

Free and bound occurrences of variables. Each occurrence of a variable in a formula is either a bound or a free occurrence, but cannot be both. Occurrences of proposition variables are always free. The occurrence of the term variable $\boldsymbol{X}$ in the formula $\boldsymbol{X}$ is free. In the formulas $(\boldsymbol{AB})$, $(\iota\boldsymbol{X}, r)\boldsymbol{P}$, $(\boldsymbol{A} = \boldsymbol{B})$, $(\sim\boldsymbol{P})$, $(\boldsymbol{P} \supset \boldsymbol{Q})$, $D^r\boldsymbol{A}$, $(\boldsymbol{X}, r)\boldsymbol{P}$ the occurrences of the various variables are free or bound according as they were free or bound in their corresponding occurrences in $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{P}$, or $\boldsymbol{Q}$ except that the occurrences of $\boldsymbol{X}$ in $(\boldsymbol{X}, r)\boldsymbol{P}$, $(\iota\boldsymbol{X}, r)\boldsymbol{P}$ are bound.

It may be observed that all four possible combinations concerning the presence or absence of a variable bound or free in a formula can occur. Examples are T', x, $(\iota x, 0)(x = x)$, $x = (\iota x, 0)(x = x)$.

Formulas and tautological formulas of the propositional calculus. The formulas of the propositional calculus are defined to be the least class of formulas containing the propositional variables, and containing $(\boldsymbol{P} \supset \boldsymbol{Q})$ and $(\sim\boldsymbol{P})$ whenever it contains $\boldsymbol{P}$ and $\boldsymbol{Q}$. Tautological formulas of the propositional calculus are those which always give the value T if a substitution of values T or F is made for the variables, and the result then evaluated as follows: T $\supset$ T is F, T $\supset$ F is F, F $\supset$ T is T, F $\supset$ F is T, $\sim$T is F, $\sim$F is T.

The rules of procedure (provable formulas). We word our rules of procedure in the form of a definition of the "provable formulas". Throughout, r is any numeral representing a non-negative integer.

Rule I (Change of bound variables). The formulas

$$(\boldsymbol{x}, r)\boldsymbol{P} \equiv (\boldsymbol{y}, r)\ S^{\boldsymbol{x}}_{\boldsymbol{y}}\boldsymbol{P}\,|$$
$$(\iota\boldsymbol{X}, r)\boldsymbol{P} \equiv (\iota\boldsymbol{y}, r)\ S^{\boldsymbol{x}}_{\boldsymbol{y}}\boldsymbol{P}\,|$$

are provable if $\boldsymbol{P}$ is a proposition formula in which $\boldsymbol{y}$ does not occur free, and $\boldsymbol{x}$ is not free at a place where $\boldsymbol{y}$ would be bound.

Rule II (Substitution). If $\boldsymbol{P}$ is provable, then $S^{\boldsymbol{x}}_{\boldsymbol{A}}\boldsymbol{P}\,|$ and $S^{\boldsymbol{q}}_{\boldsymbol{Q}}\boldsymbol{P}\,|$ are provable, where $\boldsymbol{A}$ and $\boldsymbol{Q}$ are respectively term and proposition formulas, and the bound variables of $\boldsymbol{P}$ are distinct both from $\boldsymbol{x}$ and $\boldsymbol{q}$ and from the free variables of $\boldsymbol{A}$ and of $\boldsymbol{Q}$.

Rule III (Quantifiers). If either of the two formulas $\boldsymbol{H} \supset (D^r\boldsymbol{x} \supset \boldsymbol{P})$, $\boldsymbol{H} \supset (\boldsymbol{x}, r)\boldsymbol{P}$ is provable, and $\boldsymbol{x}$ is not free in $\boldsymbol{H}$, then the other is also provable.

Rule IV (Propositional calculus). Any tautologous formula of the propositional calculus is provable.

Rule V (Modus ponens). If the formulas $\boldsymbol{P} \supset \boldsymbol{Q}$ and $\boldsymbol{P}$ are both provable then $\boldsymbol{Q}$ is provable.

Rule VI (Descriptions). If $\boldsymbol{P}$ is a proposition formula in which x does not occur bound, then the formulas

$$(\exists !x, r)\boldsymbol{P} \supset S^{x}_{(\iota x, r)\boldsymbol{P}}\boldsymbol{P}\,|$$
$$\sim(\exists !x, r)\boldsymbol{P} \supset (\iota x, r)\boldsymbol{P} = C$$
$$D^r(\iota x, r)\boldsymbol{P}$$

are provable.

Rule VII. The formula

$$(x, r)D^r\ \boldsymbol{A} \supset (\exists y, r+1)(\sim D^0 y\ \&\ (x, r)\ yx = \boldsymbol{A})$$

is provable provided y does not appear free in the term formula $\boldsymbol{A}$.

Rule VIII (Axioms). For any numeral r representing a non-negative integer the following formulas numbered A1 to C2 are provable:

A1. $C \neq T'\ \&\ C \neq F'\ \&\ T' \neq F'$
A2. $D^0C\ \&\ D^0T'\ \&\ D^0F'$
A3. $[D^0x \vee (D^{r+1}x\ \&\ \sim D^r y)] \supset xy = C$
A4. $D^r x \supset D^{r+1}x$
A5. $D^{r+1}x \supset D^r xy$
B1. $x = x$
B2. $(y = x\ \&\ y = z) \supset x = z$
B3. $x = y \supset (zx = zy\ \&\ xz = yz)$
C1. $(x, r)fx = gx \supset [f = g \vee D^0 f \vee D^0 g \vee \sim D^{r+1} f \vee \sim D^{r+1} g]$
(Axiom of extensionality.)
C2. $(\exists i, r+2)(f, r+1)((\exists x, r)fx = T') \supset f(if) = T']$
(Axiom of choice.)

Rule IX (Axiom of infinity). The following formula is provable:

C3. $(\exists h, 1)(\exists v, 0)(x, 0)(y, 0)[(hx = hy \supset x = y)\ \&\ v \neq hx]$

If we have a finite universe with N individuals instead of an infinite one we must replace rule IX by:

Rule IX_N. The following, D1 and D2, are provable:

D1. $D^0x \equiv (x = U_1^H \vee \ldots \vee U_N^H)$

D2. $U_n^H \neq U_m^H$

where m and n are different and not greater than N.

We may make a number of remarks about these axioms and rules:

(1) Axioms D1, D2 are rather stronger than is really necessary. Instead we could use the one axiom

$$D^0x \supset (x = U_1^H \vee \ldots \vee x = U_N^H)$$

which would be more nearly analogous to C3, but would admit the possibility of there being fewer than N individuals.

(2) The second formula under rule VI might have been omitted. If this had been done it would have been necessary to define a new description operator in terms of the old one in such a way that the second formula would apply for the new operator.

(3) It may be wondered why rules VI and VII do not appear under the axioms, $yx = T'$ being written for $\boldsymbol{P}$ and yx for $\boldsymbol{A}$. If there had been any more rules of this kind they could have been replaced by axioms, by making similar substitutions, but these axioms would only be equivalent to the corresponding rule in the presence of rules VI, VII. It will now be clear why rules VI, VII cannot themselves be written as axioms.

(4) A term U_m and its corresponding formula U_m^H are not regarded as identical as they were in §1. We have introduced a distinction rather similar to the distinction between the real and complex numbers π. This distinction will be of value in any attempt to provide a formal justification of the system in terms of tables: it would then be very embarrassing to have the same notation both for a formula and its interpretation. The author has carried through such a justification in detail, together with a proof that the system is complete for the finite universe. This provides a good check that no essential axioms have been omitted. The theorem mentioned in the next section provides a similar check.

(5) Although rule III does not permit $\boldsymbol{H} \supset (D^r\ x \supset \boldsymbol{P})$ to be deduced directly from $\boldsymbol{H} \supset (x, r)\boldsymbol{P}$ if x is free in $\boldsymbol{H}$, the deduction may be made indirectly.

(6) The axiom of choice is optional, i.e. we may drop this axiom and still retain a system adequate for the greater part of mathematics.

(7) We shall not carry out any proofs in this paper, but the following provable formulas are of interest:

$$x = y \supset (D^rx \supset D^ry)$$
$$(x, r)(\boldsymbol{P} \equiv \boldsymbol{Q}) \supset (\iota x, r)\ \boldsymbol{P} = (\iota x, r)\boldsymbol{Q}$$
$$(x, r)\boldsymbol{A} = \boldsymbol{B} \supset (\lambda x, r)\boldsymbol{A} = (\lambda x, r)\boldsymbol{B}$$
$$(x, r)D^r\boldsymbol{A} \supset (x, r)[((\lambda x, r)\boldsymbol{A})x = \boldsymbol{A}]$$
$$D^{r+1}\ (\lambda x, r)\boldsymbol{A}$$
$$(f, r)(g, r)[(x, r+1)(xf = xg) \supset f = g]$$
$$D^{r+1}x \equiv [(y, r+1)\{D^rxy\ \&\ (D^ry \vee xy = C)\}\ \&\ D^{r+2}x]$$

3. Equivalence with Church's system. The nested-type system described above may be proved equivalent, in a certain sense, to Church's simplified theory of types.[3] The proof is long and tedious, and would not justify publication, but it may be of interest to give an exact statement of the equivalence theorem. The form of "equivalence" used has a certain interest in itself.

Definition. A logical system 1 will be said to be *equivalent* to the logical system 2 if to each proposition-like formula $\boldsymbol{A}$ of 1 we can make correspond a proposition-like formula $\boldsymbol{A}^{(1,2)}$ of 2, and conversely to each proposition-like formula $\boldsymbol{P}$ of 2 we can make correspond a proposition-like formula $\boldsymbol{P}^{(2,1)}$ of 1, in such a way that

(i) If $\boldsymbol{A}$ is provable in 1 then $\boldsymbol{A}^{(1,2)}$ is provable in 2.

(ii) If $\boldsymbol{P}$ is provable in 2 then $\boldsymbol{P}^{(2,1)}$ is provable in 1.

(iii) If $\boldsymbol{A}$ is a proposition-like formula of 1 then $(\boldsymbol{A}^{(1,2)})^{(2,1)} \equiv \boldsymbol{A}$ is provable in 1.

(iv) If $\boldsymbol{P}$ is a proposition-like formula of 2 then $(\boldsymbol{P}^{(2,1)})^{(1,2)} \equiv \boldsymbol{P}$ is provable in 2.

(v) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are proposition-like formulas of 1 then we can prove $(\boldsymbol{A} \equiv \boldsymbol{B})^{(1,2)} \equiv (\boldsymbol{A}^{(1,2)} \equiv \boldsymbol{B}^{(1,2)})$ in 2.

(vi) If $\boldsymbol{P}$ and $\boldsymbol{Q}$ are proposition-like formulas of 2 then we can prove $(\boldsymbol{P} \equiv \boldsymbol{Q})^{(2,1)} \equiv (\boldsymbol{P}^{(2,1)} \equiv \boldsymbol{Q}^{(2,1)})$ in 1.

The formula $\boldsymbol{A}^{(1,2)}$ must be an effectively calculable function of $\boldsymbol{A}$ and $\boldsymbol{P}^{(2,1)}$ of $\boldsymbol{P}$.

It is understood that for each system there is defined a special kind of formulas called 'proposition-like formulas'; that every provable formula is necessarily proposition-like, and that it is a comparatively trivial matter to determine whether a formula is proposition-like or not. Specifically we may say that the statement "$\boldsymbol{A}$ is a proposition-like formula" should be equivalent to some statement of the form "$\varphi(n) = 0$" where n is the Gödel representation of $\boldsymbol{A}$ and φ is some primitive recursive function. It is also understood that both systems "include the propositional calculus": this is required in connection with the logical equivalence signs in (iii) to (vi).

We are justified in describing this relation as the equivalence of the two systems, for the relation is transitive, symmetric, and reflexive, as I shall now show. The symmetry of the relation follows at once from the fact that interchange of systems 1 and 2 simply interchanges conditions (i) and (ii), (iii) and (iv), (v) and (vi). Reflexiveness is proved by taking $\boldsymbol{A}^{(1,1)}$ to be $\boldsymbol{A}$. Transitivity is not quite so easy. We shall have to bring in a third system 3. We will define $\boldsymbol{A}^{(1,3)}$ to be $(\boldsymbol{A}^{(1,2)})^{(2,3)}$ and $\boldsymbol{A}^{(3,1)}$ to be $(\boldsymbol{A}^{(3,2)})^{(2,1)}$. We assume conditions (i) to (vi) to hold for the pairs 1,2 and 2,3 and attempt to prove them for the pair 1,3. Because of the symmetry it is sufficient to prove (i), (iii), (v). To prove (i) we must prove $(\boldsymbol{A}^{(1,2)})^{(2,3)}$ in 3 assuming $\boldsymbol{A}$ provable in 1. Now by (i) for the pair 1,2 we see that $\boldsymbol{A}^{(1,2)}$ is provable in 2, and then by (i) for the pair 2,3 we get $(\boldsymbol{A}^{(1,2)})^{(2,3)}$ in 3. To prove (iii) we must prove $(((\boldsymbol{A}^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv \boldsymbol{A}$ in 1.

[3] Alonzo Church, *A formulation of the simple theory of types*, this Journal, vol. 5 (1940), pp. 56–68.

Using (iii) for the pair 2,3 gives us $((A^{(1,2)})^{(2,3)})^{(3,2)} \equiv A^{(1,2)}$ (in 2), whence by (ii) for the pair 1,2 we have

$$(((A^{(1,2)})^{(2,3)})^{(3,2)} \equiv A^{(1,2)})^{(2,1)}$$

Also by (vi) for the pair 1,2 we have

$$(((A^{(1,2)})^{(2,3)})^{(3,2)} \equiv A^{(1,2)})^{(2,1)} \equiv ((((A^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv (A^{(1,2)})^{(2,1)})$$

and by (iii) for the pair 1,2 we have

$$(A^{(1,2)})^{(2,1)} \equiv A$$

Combining these last three results by the rules of the propositional calculus we obtain

$$(((A^{(1,2)})^{(2,3)})^{(3,2)})^{(2,1)} \equiv A$$

as required.
To prove (v) for the pair 1,3 we must prove

$$((A \equiv B)^{(1,2)})^{(2,3)} \equiv ((A^{(1,2)})^{(2,3)} \equiv (B^{(1,2)})^{(2,3)})$$

By an application of (v) to the pair 1,2 followed by an application of (i) to the pair 2,3 we get

$$((A \equiv B)^{(1,2)} \equiv (A^{(1,2)} \equiv B^{(1,2)}))^{(2,3)}$$

and by an application of (v) to the pair 2,3 we have

$$((A \equiv B)^{(1,2)} \equiv (A^{(1,2)} \equiv B^{(1,2)}))^{(2,3)} \equiv (((A \equiv B)^{(1,2)})^{(2,3)} \equiv (A^{(1,2)} \equiv B^{(1,2)})^{(2,3)})$$

Combining these by the propositional calculus gives

$$((A \equiv B)^{(1,2)})^{(2,3)} \equiv (A^{(1,2)} \equiv B^{(1,2)})^{(2,3)}$$

Condition (v) applied to 2,3 also gives

$$(A^{(1,2)} \equiv B^{(1,2)})^{(2,3)} \equiv ((A^{(1,2)})^{(2,3)} \equiv (B^{(1,2)})^{(2,3)})$$

from which we now obtain the required result.

Our definition of the equivalence of two systems could be summed up by saying that they are equivalent if we can translate from either system to the other in such a way that provable propositions translate into provable propositions again, and so that a double translation gives rise to a proposition equivalent to the original. This explanation ignores the last two conditions (v) and (vi), which are rather too tenuous for such rough handling.

The equivalence theorem then states that the nested-type system is equivalent to Church's system, if the proposition-like formulas of the nested-type system are taken to be the proposition formulas without free variables, and the proposition-like formulas of Church's system are those of type o without free variables.

4. Relaxation of type notation. The form of type theory which we have described is one in which the types themselves do not intrude very much. Even so they do still intrude to an appreciable extent, and it would be desirable to see how much further they can be relegated to the background. A possible way of doing so will be described in this section.

We could sum up the effect of type theory as it appears in this system by saying that we give no meaning to the expressions 'for all x, A,' 'there exists an x, such that A,' 'the x, such that A,' 'the function whose value for argument x is A' (usually expressed symbolically as $(x)A$, $(\exists x)A$, $(\iota x)A$, $(\lambda x)A$, respectively). Instead we give meaning to the expressions $(x, r)A$, $(\exists x, r)A$, $(\iota x, r)A$, $(\lambda x, r)A$. Nevertheless in a large class of cases we *can* assign meanings to $(x)A$, $(\exists x)A$, $(\iota x)A$, $(\lambda x)A$ in a satisfactory manner. A typical case is that of a formula of the form $(\iota x)P$ where P is such that we can prove $P \supset D^{10}x$, say. In this case for any integers $r, s \geqq 10$ we can prove $(\iota x, r)P = (\iota x, s)P$ and it is therefore natural to stipulate that $(\iota x)P$ shall stand for the common value of $(\iota x, 10)P$, $(\iota x, 11)P, \cdots$. We may say more generally that if $(\iota x, r_0)P = (\iota x, r)P$ is provable for all $r \geqq r_0$ then $(\iota x)P$ shall be said to be interpretable and to have the interpretation $(\iota x, r_0)P$. This is of course still only the beginning of a definition of "the interpretation of a formula with some type bounds omitted." In order to give the complete definition we must deal properly with formulas having free variables: results such as $P \supset D^{10}x$ (quoted above) are not normally provable if P has free variables other than x. On this account we introduce the idea of "interpretability under hypotheses"; the hypotheses involved are usually of the form $D^r x$. The complete definition is as follows:

All variables and C, T', F' provide their own interpretations under any hypotheses.

If A, B, P, Q have interpretations A', B', P', Q' under certain hypotheses, then (AB), $(A = B)$, $D^r A$, $(P \supset Q)$, $(\sim P)$ have the interpretations $(A'B')$, $(A' = B')$, $D^r A'$, $(P' \supset Q')$, $(\sim P')$ respectively under the same hypotheses.

If, for each $r \geqq r_0$, P has the interpretation P_r under hypothesis $H \,\&\, D^r x$ where H does not contain x free and we can prove

$$\text{(A)} \qquad H \supset (\iota x, r_0)P_{r_0} = (\iota x, r)P_r$$

then $(\iota x)P$ has the interpretation $(\iota x, r_0)P_{r_0}$ under hypothesis H. If instead of (A) we can prove

$$H \supset [(x, r_0)P_{r_0} \equiv (x, r)P_r]$$

then $(x)P$ has the interpretation $(x, r_0)P$ under H.

No formula has any interpretation unless compelled to by the foregoing.

It may be observed that every formula of the nested-type system is interpretable and provides its own interpretation. Also that if $H \supset K$ is provable and a formula has a certain interpretation under K then it has the same interpretation under H.

If P has the interpretation P_r under $H \,\&\, D^r x$ and we wish to show either that $(\iota x)P$ has the interpretation $(\iota x, r_0)P_{r_0}$, or that $(\exists x)P$ has the interpretation $(\exists x, r_0)P_{r_0}$ under H, it is sufficient to prove $P_r \supset D^{r_0}x$ $(r \geqq r_0)$.

It will be seen that this definition does not provide an effective means of determining whether or not an expression is interpretable. This need not be considered a serious drawback, as we seldom need to establish that an expression is not interpretable.

The most natural cases where we can apply the above definitions are those of $(x)(A \supset B)$, $(\exists x)(A \,\&\, B)$, $(\iota x)(A \,\&\, B)$ where $A \supset D^{r_0}x$ is provable for some r_0. It is fairly easy to remember which are the most important expressions A of this kind: e.g. in almost any formalisation we shall have " 'x is a real number' $\supset D^{r_0}x$" with $r_0 = 10$ say; this fact would be remembered in the form "the class of real numbers is all right." It is not so easy to remember the appropriate numbers r_0, but it is hardly necessary to do so if the notations $(x)A$ etc. are adhered to throughout. When A is such that for some r_0 we can prove $A \supset D^{r_0}x$ I shall call the class of x for which A is true a "noun-class." There is a very close connection between the part played by the formulas A in our system and nouns in ordinary language; so much so that one might say that type theory had been instinctively obeyed for thousands of years before its discovery by Russell. This connection may be seen by translating $(x)(A \supset B)$, $(\exists x)(A \,\&\, B)$, $(\iota x)(A \,\&\, B)$ roughly as "All A satisfy B," "There exists an A satisfying B" and "The A which satisfies B." In each case A is translated in the form of a noun. It seems that the necessity to use nouns prevents us automatically from committing type fallacies in common speech. We can probably only break down this 'safety device' by using nouns such as 'thing' or 'object' with the intended meaning 'anything whatever.' In the case of the Russell paradox ('class of all classes which are not members of themselves') we use the word 'class' in very much that way. We use it to mean 'class of anythings whatever.'

There are various ways in which we might make use of the idea of interpretable formulas to transform what we have called the 'nested type system' into something rather more closely analogous to common mathematical practice. One possibility is simply to regard the formulas without types as abbreviations of the appropriate formulas of the nested-type system, such formulas only being used when the appropriate metamathematical result justifying the interpretation has been established. This does not seem to be really satisfactory because of the frequent need to prove such metamathematical results. Alternatively we may set up some new symbolic system in which the formulas form a considerably wider class than those of the nested-type system, and are all interpretable as defined above. The author has investigated two such systems. In one of them the expression $(x, A)P$ had the meaning which we have assigned to $(x)(Ax = T' \supset P)$. This is always interpretable if A is interpretable and without free variables. This scheme leads to rather heavy formulas in the elementary stages, though it may have advantages when more advanced branches of mathematics are reached. The second system appears rather more hopeful, and will now be described briefly. It may be called the "concealed type" system.

The formulas in the concealed type system will be described as "admissible formulas" to distinguish them from the formulas of the nested-type system. The admissible formulas will in fact be included amongst the interpretable formulas associated with the nested-type system. There will be admissible term formulas (ATF) and admissible proposition formulas (APF). We define APF, ATF, and provable formula by a simultaneous induction. Consequently there is no rule for determining whether an expression is an admissible formula or not: this is not usual in logical systems, but there seems to be no good reason for a positive taboo on such an arrangement. We now give the inductive definitions.

Every term variable is an ATF and every proposition variable is an APF.

The symbols E, C, T', F' are ATF.

If A, B, F are ATF and P, Q, R, S are APF, and $P \supset Ax = T'$, $\sim Q \supset Ax = T'$ are provable formulas then $(\exists x)P$, $(x)Q$, $(B = F)$, $(R \supset S)$, $\sim R$ are APF, and $(\iota x)P$, (BF) are ATF. The variable x must not occur free in A.

Free and bound occurrences of variables are defined as in the nested-type system.

The symbol E corresponds to $(\lambda x, 0)T'$ of the nested type system. Its main purpose is to take the place of D^0 and indirectly to replace the other D^r. For any formula A we can prove $((\lambda x, 0)T')A \equiv D^0 A$ in the nested-type system.

If A and B are ATF not containing x, y, or z free then the two expressions below are ATF, viz.

$$(\iota y)(x)[(yx \vee yx = C) \;\&\; \sim Ey \;\&\; (yx \equiv (Ax \vee Bx))]$$

$$(\iota y)(x)[(yx \vee yx = C) \;\&\; \sim Ey \;\&\; (yx \equiv (z)\{(Bz \supset A(xz) \;\&\; (\sim Bz \supset xz = C)\}]$$

They may be abbreviated respectively to *Sum* $A\,B$ and *Pot* $A\,B$. In these formulas we have adopted the useful convention that a formula of form $A = T'$ may be abbreviated to A. The context will always enable one to determine when this abbreviation has been applied. We shall continue to use this convention.

Strictly speaking the definitions of *Sum* AB and *Pot* AB are invalid because the bound variables x, y, z were not specified. This technical difficulty may be resolved by requiring x, y, z to be the three earliest variables not appearing free in A, B.

The remainder of the definition consists of the axioms and rules of procedure. It may be remembered that these took the form of a definition in the nested-type system also

Rules of procedure (concealed-type system).

Rule I. The formulas

$$(x)P \equiv (y)S^x_y P|$$

$$(\iota x)P = (\iota y)S^x_y P|$$

are provable if $(x)P$ is an APF in which x is not bound in P, y does not occur free, x does not occur at a place where y would be bound, and $(\iota x)P$ is an ATF.

Rule II. If P is provable, then $S^x_A P|$ and $S^q_Q P|$ are provable, where A and P are respectively an ATF and an APF, and the bound variables of P are distinct both from x and q and from the free variables of A and of Q.

Rule III. If $H \supset P$ and $H \supset (x)P$ are both APF and one of them is provable then the other is provable also.

Rule IV. Any tautologous formula of the propositional calculus is provable.

Rule V. If the formulas $P \supset Q$ and P are both provable then Q is provable.

Rule VI. If P is an APF in which x does not occur bound, then the formulas

$$(\exists!x)P \supset S^{x}_{(\imath x)P}P|$$
$$\sim(\exists!x)P \supset (\imath x)P = C$$

are provable provided they are APF.

Rule VII. If A is an APF in which x, y, z, u do not occur free, then

$$(x)(ux \supset zA) \supset (\exists y)[(Pot\, zu)y \ \& \ (x)(ux \supset yx = A)]$$

is provable.

In rule VI the definition

$$(\exists!x)P \rightarrow (\exists x)P \ \& \ (x)(y)(P \ \& \ S^{x}_{y}P| \supset x = y)$$

is understood, y standing for a variable not occurring free in P.

The axioms are:

A1 $C \neq T' \ \& C \neq F' \ \& T' \neq F'$
A2 $EC \ \& \ ET' \ \& \ EF'$
A3 $Ex \supset xy = C$
B1 $x = x$
B2 $(y = x \ \& y = z) \supset x = z$
B3 $(x = y) \supset (zx = zy \ \& \ xz = yz)$
C1 $[(Pot\, yu)f \ \& \ (Pot\, yu)g \ \& \ (x)(ux \supset fx = gx)] \supset f = g$
C2 $(\exists i)[(Pot\, u(Pot\, Eu))i \ \& \ (f)\{[(Pot\, Eu)f \ \& \ (\exists x)fx] \supset f(if)\}]$
C3 $(\exists t)[(Pot\, E\, E)t \ \& \ (\exists v)[Ev \ \& \ (x)(y)\{(Ex \ \& \ Ey)$
$\supset ((tx = ty \supset x = y) \ \& \ v \neq tx)\}]]$

To complete our inductive definition we need only add that no expression is an ATF, APF, or provable formula unless compelled to be so by the foregoing.

We may say that roughly speaking type theory appears in the concealed type system only through the condition that $P \supset Ax = T'$ must be provable if $(\imath x)P$ is to be an ATF, and a similar condition for $(x)P$. The system is related to the nested-type system by the following metamathematical results:

(1) If we substitute $(\lambda x, 0)T'$ for E throughout an admissible formula without free variables we obtain an interpretable formula.

(2) If in a provable formula of the concealed-type system without free variables we make the substitution mentioned in (1) and then form an interpretation of the resulting formula we obtain a provable formula of the nested-type system.

(3) Every provable formula of the nested-type system is obtainable as in (2).

A valuable aid in the proof of these is the following result which concerns the nested-type system only:

(4) If A is a term formula containing only the variables $x_1, x_2, \cdots, x_n$ free, and $m_1, m_2, \cdots, m_n$ are non-negative integers, then there is an integer k such that $D^{m_1}x_1 \ \& \cdots \ \& D^{m_n}x_n \supset D^k A$ is provable.

NATIONAL PHYSICAL LABORATORY, TEDDINGTON