# PRINCIPIA MATHEMATICA

## TO ✱56

BY

ALFRED NORTH WHITEHEAD

AND

BERTRAND RUSSELL, F.R.S.

CAMBRIDGE
UNIVERSITY PRESS

# CHAPTER II

## THE THEORY OF LOGICAL TYPES

THE theory of logical types, to be explained in the present Chapter, recommended itself to us in the first instance by its ability to solve certain contradictions, of which the one best known to mathematicians is Burali-Forti's concerning the greatest ordinal. But the theory in question is not wholly dependent upon this indirect recommendation: it has also a certain consonance with common sense which makes it inherently credible. In what follows, we shall therefore first set forth the theory on its own account, and then apply it to the solution of the contradictions.

### I. *The Vicious-Circle Principle.*

An analysis of the paradoxes to be avoided shows that they all result from a certain kind of vicious circle*. The vicious circles in question arise from supposing that a collection of objects may contain members which can only be defined by means of the collection as a whole. Thus, for example, the collection of *propositions* will be supposed to contain a proposition stating that "all propositions are either true or false." It would seem, however, that such a statement could not be legitimate unless "all propositions" referred to some already definite collection, which it cannot do if new propositions are created by statements about "all propositions." We shall, therefore, have to say that statements about "all propositions" are meaningless. More generally, given any set of objects such that, if we suppose the set to have a total, it will contain members which presuppose this total, then such a set cannot have a total. By saying that a set has "no total," we mean, primarily, that no significant statement can be made about "all its members." Propositions, as the above illustration shows, must be a set having no total. The same is true, as we shall shortly see, of propositional functions, even when these are restricted to such as can significantly have as argument a given object $a$. In such cases, it is necessary to break up our set into smaller sets, each of which is capable of a total. This is what the theory of types aims at effecting.

The principle which enables us to avoid illegitimate totalities may be stated as follows: "Whatever involves *all* of a collection must not be one of the collection"; or, conversely: "If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total." We shall call this the "vicious-circle principle," because it enables us to avoid the vicious circles involved in the assumption of illegitimate totalities. Arguments which are condemned by the vicious-circle

* See the last section of the present Chapter. Cf. also H. Poincaré, "Les mathématiques et la logique," *Revue de Métaphysique et de Morale*, Mai 1906, p. 307.

principle will be called "vicious-circle fallacies." Such arguments, in certain circumstances, may lead to contradictions, but it often happens that the conclusions to which they lead are in fact true, though the arguments are fallacious. Take, for example, the law of excluded middle, in the form "all propositions are true or false." If from this law we argue that, because the law of excluded middle is a proposition, therefore the law of excluded middle is true or false, we incur a vicious-circle fallacy. "All propositions" must be in some way limited before it becomes a legitimate totality, and any limitation which makes it legitimate must make any statement about the totality fall outside the totality. Similarly, the imaginary sceptic, who asserts that he knows nothing, and is refuted by being asked if he knows that he knows nothing, has asserted nonsense, and has been fallaciously refuted by an argument which involves a vicious-circle fallacy. In order that the sceptic's assertion may become significant, it is necessary to place some limitation upon the things of which he is asserting his ignorance, because the things of which it is possible to be ignorant form an illegitimate totality. But as soon as a suitable limitation has been placed by him upon the collection of propositions of which he is asserting his ignorance, the proposition that he is ignorant of every member of this collection must not itself be one of the collection. Hence any significant scepticism is not open to the above form of refutation.

The paradoxes of symbolic logic concern various sorts of objects: propositions, classes, cardinal and ordinal numbers, etc. All these sorts of objects, as we shall show, represent illegitimate totalities, and are therefore capable of giving rise to vicious-circle fallacies. But by means of the theory (to be explained in Chapter III) which reduces statements that are verbally concerned with classes and relations to statements that are concerned with propositional functions, the paradoxes are reduced to such as are concerned with propositions and propositional functions. The paradoxes that concern propositions are only indirectly relevant to mathematics, while those that more nearly concern the mathematician are all concerned with *propositional functions*. We shall therefore proceed at once to the consideration of propositional functions.

### II. *The Nature of Propositional Functions.*

By a "propositional function" we mean something which contains a variable $x$, and expresses a *proposition* as soon as a value is assigned to $x$. That is to say, it differs from a proposition solely by the fact that it is ambiguous: it contains a variable of which the value is unassigned. It agrees with the ordinary functions of mathematics in the fact of containing an unassigned variable; where it differs is in the fact that the values of the function are propositions. Thus *e.g.* "$x$ is a man" or "$\sin x = 1$" is a propositional function. We shall find that it is possible to incur a vicious-circle it is by no means improbable that it should be found to be deducible from some other more fundamental and more evident axiom. It is possible that the use of the vicious-circle principle, as embodied in the above hierarchy of types, is more drastic than it need be, and that by a less drastic use the necessity for the axiom might be avoided. Such changes, however, would not render anything false which had been asserted on the basis of the principles explained above: they would merely provide easier proofs of the same theorems. There would seem, therefore, to be but the slenderest ground for fearing that the use of the axiom of reducibility may lead us into error.

## VIII. *The Contradictions.*

We are now in a position to show how the theory of types affects the solution of the contradictions which have beset mathematical logic. For this purpose, we shall begin by an enumeration of some of the more important and illustrative of these contradictions, and shall then show how they all embody vicious-circle fallacies, and are therefore all avoided by the theory of types. It will be noticed that these paradoxes do not relate exclusively to the ideas of number and quantity. Accordingly no solution can be adequate which seeks to explain them merely as the result of some illegitimate use of these ideas. The solution must be sought in some such scrutiny of fundamental logical ideas as has been attempted in the foregoing pages.

(1) The oldest contradiction of the kind in question is the *Epimenides*. Epimenides the Cretan said that all Cretans were liars, and all other statements made by Cretans were certainly lies. Was this a lie? The simplest form of this contradiction is afforded by the man who says "I am lying"; if he is lying, he is speaking the truth, and vice versa.

(2) Let $w$ be the class of all those classes which are not members of themselves. Then, whatever class $x$ may be, "$x$ is a $w$" is equivalent to "$x$ is not an $x$." Hence, giving to $x$ the value $w$, "$w$ is a $w$" is equivalent to "$w$ is not a $w$."

(3) Let $T$ be the relation which subsists between two relations $R$ and $S$ whenever $R$ does not have the relation $R$ to $S$. Then, whatever relations $R$ and $S$ may be, "$R$ has the relation $T$ to $S$" is equivalent to "$R$ does not have the relation $R$ to $S$." Hence, giving the value $T$ to both $R$ and $S$, "$T$ has the relation $T$ to $T$" is equivalent to "$T$ does not have the relation $T$ to $T$."

(4) Burali-Forti's contradiction* may be stated as follows: It can be shown that every well-ordered series has an ordinal number, that the series of ordinals up to and including any given ordinal exceeds the given ordinal by one, and (on certain very natural assumptions) that the series of all ordinals (in order of magnitude) is well-ordered. It follows that the series of.all

* "Una questione sui numeri transfiniti," *Rendiconti del circolo matematico di Palermo*, Vol. XI. (1897). See *256.

ordinals has an ordinal number, $\Omega$ say. But in that case the series of all ordinals including $\Omega$ has the ordinal number $\Omega + 1$, which must be greater than $\Omega$. Hence $\Omega$ is not the ordinal number of all ordinals.

(5) The number of syllables in the English names of finite integers tends to increase as the integers grow larger, and must gradually increase indefinitely, since only a finite number of names can be made with a given finite number of syllables. Hence the names of some integers must consist of at least nineteen syllables, and among these there must be a least. Hence "the least integer not nameable in fewer than nineteen syllables" must denote a definite integer; in fact, it denotes 111,777. But "the least integer not nameable in fewer than nineteen syllables" is itself a name consisting of eighteen syllables; hence the least integer not nameable in fewer than nineteen syllables can be named in eighteen syllables, which is a contradiction[*].

(6) Among transfinite ordinals some can be defined, while others can not; for the total number of possible definitions is $\aleph_0$[†], while the number of transfinite ordinals exceeds $\aleph_0$. Hence there must be indefinable ordinals, and among these there must be a least. But this is defined as "the least indefinable ordinal," which is a contradiction[‡].

(7) Richard's paradox[§] is akin to that of the least indefinable ordinal. It is as follows: Consider all decimals that can be defined by means of a finite number of words; let $E$ be the class of such decimals. Then $E$ has $\aleph_0$ terms; hence its members can be ordered as the 1st, 2nd, 3rd, .... Let $N$ be a number defined as follows: If the $n$th figure in the $n$th decimal is $p$, let the $n$th figure in $N$ be $p + 1$ (or 0, if $p = 9$). Then $N$ is different from all the members of $E$, since, whatever finite value $n$ may have, the $n$th figure in $N$ is different from the $n$th figure in the $n$th of the decimals composing $E$, and therefore $N$ is different from the $n$th decimal. Nevertheless we have defined $N$ in a finite number of words, and therefore $N$ ought to be a member of $E$. Thus $N$ both is and is not a member of $E$.

In all the above contradictions (which are merely selections from an indefinite number) there is a common characteristic, which we may describe as self-reference or reflexiveness. The remark of Epimenides must include itself in its own scope. If *all* classes, provided they are not members of themselves, are members of $w$, this must also apply to $w$; and similarly for the

[*] This contradiction was suggested to us by Mr G. G. Berry of the Bodleian Library.

[†] $\aleph_0$ is the number of finite integers. See *123.

[‡] Cf. König, "Ueber die Grundlagen der Mengenlehre und das Kontinuumproblem," *Math. Annalen*, Vol. LXI. (1905); A. C. Dixon, "On 'well-ordered' aggregates," *Proc. London Math. Soc.* Series 2, Vol. IV. Part I. (1906); and E. W. Hobson, "On the Arithmetic Continuum," *ibid.* The solution offered in the last of these papers depends upon the variation of the "apparatus of definition," and is thus in outline in agreement with the solution adopted here. But it does not invalidate the statement in the text, if "definition" is given a constant meaning.

[§] Cf. Poincaré, "Les mathématiques et la logique," *Revue de Métaphysique et de Morale*, Mai 1906, especially sections VII. and IX.; also Peano, *Revista de Mathematica*, Vol. VIII. No. 5 (1906), p. 149 ff.

analogous relational contradiction. In the cases of names and definitions, the paradoxes result from considering non-nameability and indefinability as elements in names and definitions. In the case of Burali-Forti's paradox, the series whose ordinal number causes the difficulty is the series of all ordinal numbers. In each contradiction something is said about *all* cases of some kind, and from what is said a new case seems to be generated, which both is and is not of the same kind as the cases of which *all* were concerned in what was said. But this is the characteristic of illegitimate totalities, as we defined them in stating the vicious-circle principle. Hence all our contradictions are illustrations of vicious-circle fallacies. It only remains to show, therefore, that the illegitimate totalities involved are excluded by the hierarchy of types which we have constructed.

(1) When a man says "I am lying," we may interpret his statement as: "There is a proposition which I am affirming and which is false." That is to say, he is asserting the truth of some value of the function "I assert $p$, and $p$ is false." But we saw that the word "false" is ambiguous, and that, in order to make it unambiguous, we must specify the order of falsehood, or, what comes to the same thing, the order of the proposition to which falsehood is ascribed. We saw also that, if $p$ is a proposition of the $n$th order, a proposition in which $p$ occurs as an apparent variable is not of the $n$th order, but of a higher order. Hence the kind of truth or falsehood which can belong to the statement "there is a proposition $p$ which I am affirming and which has falsehood of the $n$th order" is truth or falsehood of a higher order than the $n$th. Hence the statement of Epimenides does not fall within its own scope, and therefore no contradiction emerges.

If we regard the statement "I am lying" as a compact way of simultaneously making all the following statements: "I am asserting a false proposition of the first order," "I am asserting a false proposition of the second order," and so on, we find the following curious state of things: As no proposition of the first order is being asserted, the statement "I am asserting a false proposition of the first order" is false. This statement is of the second order, hence the statement "I am making a false statement of the second order" is true. This is a statement of the third order, and is the only statement of the third order which is being made. Hence the statement "I am making a false statement of the third order" is false. Thus we see that the statement "I am making a false statement of order $2n + 1$" is false, while the statement "I am making a false statement of order $2n$" is true. But in this state of things there is no contradiction.

(2) In order to solve the contradiction about the class of classes which are not members of themselves, we shall assume, what will be explained in the next Chapter, that a proposition about a class is always to be reduced to a statement about a function which defines the class, *i.e.* about a function which

is satisfied by the members of the class and by no other arguments. Thus a class is an object derived from a function and presupposing the function, just as, for example, $(x) . \phi x$ presupposes the function $\phi\hat{x}$. Hence a class cannot, by the vicious-circle principle, significantly be the argument to its defining function, that is to say, if we denote by "$\hat{z}(\phi z)$" the class defined by $\phi\hat{z}$, the symbol "$\phi\{\hat{z}(\phi z)\}$" must be meaningless. Hence a class neither satisfies nor does not satisfy its defining function, and therefore (as will appear more fully in Chapter III) is neither a member of itself nor not a member of itself. This is an immediate consequence of the limitation to the possible arguments to a function which was explained at the beginning of the present Chapter. Thus if $\alpha$ is a class, the statement "$\alpha$ is not a member of $\alpha$" is always meaningless, and there is therefore no sense in the phrase "the class of those classes which are not members of themselves." Hence the contradiction which results from supposing that there is such a class disappears.

(3) Exactly similar remarks apply to "the relation which holds between $R$ and $S$ whenever $R$ does not have the relation $R$ to $S$." Suppose the relation $R$ is defined by a function $\phi(x, y)$, *i.e.* $R$ holds between $x$ and $y$ whenever $\phi(x, y)$ is true, but not otherwise. Then in order to interpret "$R$ has the relation $R$ to $S$," we shall have to suppose that $R$ and $S$ can significantly be the arguments to $\phi$. But (assuming, as will appear in Chapter III, that $R$ presupposes its defining function) this would require that $\phi$ should be able to take as argument an object which is defined in terms of $\phi$, and this no function can do, as we saw at the beginning of this Chapter. Hence "$R$ has the relation $R$ to $S$" is meaningless, and the contradiction ceases.

(4) The solution of Burali-Forti's contradiction requires some further developments for its solution. At this stage, it must suffice to observe that a series is a relation, and an ordinal number is a class of series. (These statements are justified in the body of the work.) Hence a series of ordinal numbers is a relation between classes of relations, and is of higher type than any of the series which are members of the ordinal numbers in question. Burali-Forti's "ordinal number of all ordinals" must be the ordinal number of all ordinals of a given type, and must therefore be of higher type than any of these ordinals. Hence it is not one of these ordinals, and there is no contradiction in its being greater than any of them*.

(5) The paradox about "the least integer not nameable in fewer than nineteen syllables" embodies, as is at once obvious, a vicious-circle fallacy. For the word "nameable" refers to the totality of names, and yet is allowed to occur in what professes to be one among names. Hence there can be no such thing as a totality of names, in the sense in which the paradox speaks

* The solution of Burali-Forti's paradox by means of the theory of types is given in detail in *256.

of "names." It is easy to see that, in virtue of the hierarchy of functions, the theory of types renders a totality of "names" impossible. We may, in fact, distinguish names of different orders as follows: (*a*) Elementary names will be such as are true "proper names," *i.e.* conventional appellations not involving any description. (*b*) First-order names will be such as involve a description by means of a first-order function; that is to say, if $\phi ! \hat{x}$ is a first-order function, "the term which satisfies $\phi ! \hat{x}$" will be a first-order name, though there will not always be an object named by this name. (*c*) Second-order names will be such as involve a description by means of a second-order function; among such names will be those involving a reference to the totality of first-order names. And so we can proceed through a whole hierarchy. But at no stage can we give a meaning to the word "nameable" unless we specify the order of names to be employed; and any name in which the phrase "nameable by names of order $n$" occurs is necessarily of a higher order than the $n$th. Thus the paradox disappears.

The solutions of the paradox about the least indefinable ordinal and of Richard's paradox are closely analogous to the above. The notion of "definable," which occurs in both, is nearly the same as "nameable," which occurs in our fifth paradox: "definable" is what "nameable" becomes when elementary names are excluded, *i.e.* "definable" means "nameable by a name which is not elementary." But here there is the same ambiguity as to type as there was before, and the same need for the addition of words which specify the type to which the definition is to belong. And however the type may be specified, "the least ordinal not definable by definitions of this type" is a definition of a higher type; and in Richard's paradox, when we confine ourselves, as we must, to decimals that have a definition of a given type, the number $N$, which causes the paradox, is found to have a definition which belongs to a higher type, and thus not to come within the scope of our previous definitions.

An indefinite number of other contradictions, of similar nature to the above seven, can easily be manufactured. In all of them, the solution is of the same kind. In all of them, the appearance of contradiction is produced by the presence of some word which has systematic ambiguity of type, such as *truth*, *falsehood*, *function*, *property*, *class*, *relation*, *cardinal*, *ordinal*, *name*, *definition*. Any such word, if its typical ambiguity is overlooked, will apparently generate a totality containing members defined in terms of itself, and will thus give rise to vicious-circle fallacies. In most cases, the conclusions of arguments which involve vicious-circle fallacies will not be self-contradictory, but wherever we have an illegitimate totality, a little ingenuity will enable us to construct a vicious-circle fallacy leading to a contradiction, which disappears as soon as the typically ambiguous words are rendered typically definite, *i.e.* are determined as belonging to this or that type.

Thus the appearance of contradiction is always due to the presence of words embodying a concealed typical ambiguity, and the solution of the apparent contradiction lies in bringing the concealed ambiguity to light.

In spite of the contradictions which result from unnoticed typical ambiguity, it is not desirable to avoid words and symbols which have typical ambiguity. Such words and symbols embrace practically all the ideas with which mathematics and mathematical logic are concerned: the systematic ambiguity is the result of a systematic analogy. That is to say, in almost all the reasonings which constitute mathematics and mathematical logic, we are using ideas which may receive any one of an infinite number of different typical determinations, any one of which leaves the reasoning valid. Thus by employing typically ambiguous words and symbols, we are able to make one chain of reasoning applicable to any one of an infinite number of different cases, which would not be possible if we were to forego the use of typically ambiguous words and symbols.

Among propositions wholly expressed in terms of typically ambiguous notions practically the only ones which may differ, in respect of truth or falsehood, according to the typical determination which they receive, are existence-theorems. If we assume that the total number of individuals is $n$, then the total number of classes of individuals is $2^n$, the total number of classes of classes of individuals is $2^{2^n}$, and so on. Here $n$ may be either finite or infinite, and in either case $2^n > n$. Thus cardinals greater than $n$ but not greater than $2^n$ exist as applied to classes of classes, but not as applied to classes of individuals, so that whatever may be supposed to be the number of individuals, there will be existence-theorems which hold for higher types but not for lower types. Even here, however, so long as the number of individuals is not asserted, but is merely assumed hypothetically, we may replace the type of individuals by any other type, provided we make a corresponding change in all the other types occurring in the same context. That is, we may give the name "relative individuals" to the members of an arbitrarily chosen type $\tau$, and the name "relative classes of individuals" to classes of "relative individuals," and so on. Thus so long as only hypotheticals are concerned, in which existence-theorems for one type are shown to be implied by existence-theorems for another, only *relative* types are relevant even in existence-theorems. This applies also to cases where the hypothesis (and therefore the conclusion) is *asserted*, provided the assertion holds for any type, however chosen. For example, any type has at least one member; hence any type which consists of classes, of whatever order, has at least two members. But the further pursuit of these topics must be left to the body of the work.