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

fallacy at the very outset, by admitting as possible arguments to a propositional function terms which presuppose the function. This form of the fallacy is very instructive, and its avoidance leads, as we shall see, to the hierarchy of types.

The question as to the nature of a function* is by no means an easy one. It would seem, however, that the essential characteristic of a function is *ambiguity*. Take, for example, the law of identity in the form "A is A," which is the form in which it is usually enunciated. It is plain that, regarded psychologically, we have here a single judgment. But what are we to say of the object of the judgment? We are not judging that Socrates is Socrates, nor that Plato is Plato, nor any other of the definite judgments that are instances of the law of identity. Yet each of these judgments is, in a sense, within the scope of our judgment. We are in fact judging an ambiguous instance of the propositional function "A is A." We appear to have a single thought which does not have a definite object, but has as its object an undetermined one of the values of the function "A is A." It is this kind of ambiguity that constitutes the essence of a function. When we speak of "ϕx," where x is not specified, we mean one value of the function, but not a definite one. We may express this by saying that "ϕx" *ambiguously denotes* ϕa, ϕb, ϕc, etc., where ϕa, ϕb, ϕc, etc., are the various values of "ϕx."

When we say that "ϕx" ambiguously denotes ϕa, ϕb, ϕc, etc., we mean that "ϕx" means one of the objects ϕa, ϕb, ϕc, etc., though not a definite one, but an undetermined one. It follows that "ϕx" only has a well-defined meaning (well-defined, that is to say, except in so far as it is of its essence to be ambiguous) if the objects ϕa, ϕb, ϕc, etc., are well-defined. That is to say, a function is not a well-defined function unless all its values are already well-defined. It follows from this that no function can have among its values anything which presupposes the function, for if it had, we could not regard the objects ambiguously denoted by the function as definite until the function was definite, while conversely, as we have just seen, the function cannot be definite until its values are definite. This is a particular case, but perhaps the most fundamental case, of the vicious-circle principle. A function is what ambiguously denotes some one of a certain totality, namely the values of the function; hence this totality cannot contain any members which involve the function, since, if it did, it would contain members involving the totality, which, by the vicious-circle principle, no totality can do.

It will be seen that, according to the above account, the values of a function are presupposed by the function, not vice versa. It is sufficiently obvious, in any particular case, that a value of a function does not presuppose the function. Thus for example the proposition "Socrates is human" can be perfectly apprehended without regarding it as a value of the function "x is human." It is true that, conversely, a function can be apprehended without

---

* When the word "function" is used in the sequel, "propositional function" is always meant. Other functions will not be in question in the present Chapter.

its being necessary to apprehend its values severally and individually. If this were not the case, no function could be apprehended at all, since the number of values (true and false) of a function is necessarily infinite and there are necessarily possible arguments with which we are unacquainted. What is necessary is not that the values should be given individually and extensionally, but that the totality of the values should be given intensionally, so that, concerning any assigned object, it is at least theoretically determinate whether or not the said object is a value of the function.

It is necessary practically to distinguish the function itself from an undetermined value of the function. We may regard the function itself as that which ambiguously denotes, while an undetermined value of the function is that which is ambiguously denoted. If the undetermined value is written "ϕx," we will write the function itself "$\phi \hat{x}$." (Any other letter may be used in place of x.) Thus we should say "ϕx is a proposition," but "$\phi \hat{x}$ is a propositional function." When we say "ϕx is a proposition," we mean to state something which is true for every possible value of x, though we do not decide what value x is to have. We are making an ambiguous statement about any value of the function. But when we say "$\phi \hat{x}$ is a function," we are not making an ambiguous statement. It would be more correct to say that we are making a statement about an ambiguity, taking the view that a function is an ambiguity. The function itself, $\phi \hat{x}$, is the single thing which ambiguously denotes its many values; while ϕx, where x is not specified, is one of the denoted objects, with the ambiguity belonging to the manner of denoting.

We have seen that, in accordance with the vicious-circle principle, the values of a function cannot contain terms only definable in terms of the function. Now given a function $\phi \hat{x}$, the values for the function* are all propositions of the form ϕx. It follows that there must be no propositions, of the form ϕx, in which x has a value which involves $\phi \hat{x}$. (If this were the case, the values of the function would not all be determinate until the function was determinate, whereas we found that the function is not determinate unless its values are previously determinate.) Hence there must be no such thing as the value for $\phi \hat{x}$ with the argument $\phi \hat{x}$, or with any argument which involves $\phi \hat{x}$. That is to say, the symbol "$\phi (\phi \hat{x})$" must not express a proposition, as "ϕa" does if ϕa is a value for $\phi \hat{x}$. In fact "$\phi (\phi \hat{x})$" must be a symbol which does not express anything: we may therefore say that it is not significant. Thus given any function $\phi \hat{x}$, there are arguments with which the function has no value, as well as arguments with which it has a value. We will call the arguments with which $\phi \hat{x}$ has a value "possible values of x." We will say that $\phi \hat{x}$ is "significant with the argument x" when $\phi \hat{x}$ has a value with the argument x.

* We shall speak in this Chapter of "values *for* $\phi \hat{x}$" and of "values *of* ϕx," meaning in each case the same thing, namely ϕa, ϕb, ϕc, etc. The distinction of phraseology serves to avoid ambiguity where several variables are concerned, especially when one of them is a function.

When it is said that *e.g.* "$\phi (\phi \hat{z})$" is meaningless, and therefore neither true nor false, it is necessary to avoid a misunderstanding. If "$\phi (\phi \hat{z})$" were interpreted as meaning "the value for $\phi \hat{z}$ with the argument $\phi \hat{z}$ is true," that would be not meaningless, but false. It is false for the same reason for which "the King of France is bald" is false, namely because there is no such thing as "the value for $\phi \hat{z}$ with the argument $\phi \hat{z}$." But when, with some argument a, we assert ϕa, we are not meaning to assert "the value for $\phi \hat{x}$ with the argument a is true"; we are meaning to assert the actual proposition which *is* the value for $\phi \hat{x}$ with the argument a. Thus for example if $\phi \hat{x}$ is "$\hat{x}$ is a man," ϕ (Socrates) will be "Socrates is a man," *not* "the value for the function '$\hat{x}$ is a man,' with the argument Socrates, is true." Thus in accordance with our principle that "$\phi (\phi \hat{z})$" is meaningless, we cannot legitimately deny "the function '$\hat{x}$ is a man' is a man," because this is nonsense, but we can legitimately deny "the value for the function '$\hat{x}$ is a man' with the argument '$\hat{x}$ is a man' is true," not on the ground that the value in question is false, but on the ground that there is no such value for the function.

We will denote by the symbol "$(x) . \phi x$" the proposition "ϕx always*," *i.e.* the proposition which asserts *all* the values for $\phi \hat{x}$. This proposition involves the function $\phi \hat{x}$, not merely an ambiguous value of the function. The assertion of ϕx, where x is unspecified, is a different assertion from the one which asserts all values for $\phi \hat{x}$, for the former is an ambiguous assertion, whereas the latter is in no sense ambiguous. It will be observed that "$(x) . \phi x$" does not assert "ϕx with all values of x," because, as we have seen, there must be values of x with which "ϕx" is meaningless. What is asserted by "$(x) . \phi x$" is all propositions which are values for $\phi \hat{x}$; hence it is only with such values of x as make "ϕx" significant, *i.e.* with all *possible* arguments, that ϕx is asserted when we assert "$(x) . \phi x$." Thus a convenient way to read "$(x) . \phi x$" is "ϕx is true with all possible values of x." This is, however, a less accurate reading than "ϕx always," because the notion of *truth* is not part of the content of what is judged. When we judge "all men are mortal," we judge truly, but the notion of truth is not necessarily in our minds, any more than it need be when we judge "Socrates is mortal."

III. *Definition and Systematic Ambiguity of Truth and Falsehood.*

Since "$(x) . \phi x$" involves the function $\phi \hat{x}$, it must, according to our principle, be impossible as an argument to ϕ. That is to say, the symbol "$\phi \{(x) . \phi x\}$" must be meaningless. This principle would seem, at first sight, to have certain exceptions. Take, for example, the function "$\hat{p}$ is false," and consider the proposition "$(p) . p$ is false." This should be a proposition asserting all propositions of the form "p is false." Such a proposition, we

* We use "always" as meaning "in all cases," not "at all times." Similarly "sometimes" will mean "in some cases."

should be inclined to say, must be false, because "p is false" is not always true. Hence we should be led to the proposition

$$\{(p) . p \text{ is false}\} \text{ is false,}$$

i.e. we should be led to a proposition in which "$(p) . p$ is false" is the argument to the function "$\hat{p}$ is false," which we had declared to be impossible. Now it will be seen that "$(p) . p$ is false," in the above, purports to be a proposition about all propositions, and that, by the general form of the vicious-circle principle, there must be no propositions about *all* propositions. Nevertheless, it seems plain that, given any function, there is a proposition (true or false) asserting all its values. Hence we are led to the conclusion that "p is false" and "q is false" must not always be the values, with the arguments p and q, for a single function "$\hat{p}$ is false." This, however, is only possible if the word "false" really has many different meanings, appropriate to propositions of different kinds.

That the words "true" and "false" have many different meanings, according to the kind of proposition to which they are applied, is not difficult to see. Let us take any function $\phi\hat{x}$, and let ϕa be one of its values. Let us call the sort of truth which is applicable to ϕa "*first* truth." (This is not to assume that this would be first truth in another context: it is merely to indicate that it is the first sort of truth in our context.) Consider now the proposition $(x) . \phi x$. If this has truth of the sort appropriate to it, that will mean that every value ϕx has "first truth." Thus if we call the sort of truth that is appropriate to $(x) . \phi x$ "*second* truth," we may define "$\{(x) . \phi x\}$ has second truth" as meaning "every value for $\phi\hat{x}$ has first truth," *i.e.* "$(x) . (\phi x$ has first truth)." Similarly, if we denote by "$(\exists x) . \phi x$" the proposition "ϕx sometimes," *i.e.* as we may less accurately express it, "ϕx with some value of x," we find that $(\exists x) . \phi x$ has second truth if there is an x with which ϕx has first truth; thus we may define "$\{(\exists x) . \phi x\}$ has second truth" as meaning "some value for $\phi\hat{x}$ has first truth," *i.e.* "$(\exists x) . (\phi x$ has first truth)." Similar remarks apply to falsehood. Thus "$\{(x) . \phi x\}$ has second falsehood" will mean "some value for $\phi\hat{x}$ has first falsehood," *i.e.* "$(\exists x) . (\phi x$ has first falsehood)," while "$\{(\exists x) . \phi x\}$ has second falsehood" will mean "all values for $\phi\hat{x}$ have first falsehood," *i.e.* "$(x) . (\phi x$ has first falsehood)." Thus the sort of falsehood that can belong to a general proposition is different from the sort that can belong to a particular proposition.

Applying these considerations to the proposition "$(p) . p$ is false," we see that the kind of falsehood in question must be specified. If, for example, first falsehood is meant, the function "$\hat{p}$ has first falsehood" is only significant when p is the sort of proposition which has first falsehood or first truth. Hence "$(p) . p$ is false" will be replaced by a statement which is equivalent to "all propositions having either first truth or first falsehood have first falsehood." This proposition has *second* falsehood, and is not

a possible argument to the function "$\hat{p}$ has *first* falsehood." Thus the apparent exception to the principle that "$\phi \{(x) . \phi x\}$" must be meaningless disappears.

Similar considerations will enable us to deal with "not-p" and with "p or q." It might seem as if these were functions in which *any* proposition might appear as argument. But this is due to a systematic ambiguity in the meanings of "not" and "or," by which they adapt themselves to propositions of any order. To explain fully how this occurs, it will be well to begin with a definition of the simplest kind of *truth* and *falsehood*.

The universe consists of objects having various qualities and standing in various relations. Some of the objects which occur in the universe are complex. When an object is complex, it consists of interrelated parts. Let us consider a complex object composed of two parts a and b standing to each other in the relation R. The complex object "a-in-the-relation-R-to-b" may be capable of being *perceived*; when perceived, it is perceived as one object. Attention may show that it is complex; we then *judge* that a and b stand in the relation R. Such a judgment, being derived from perception by mere attention, may be called a "judgment of perception." This judgment of perception, considered as an actual occurrence, is a relation of four terms, namely a and b and R and the percipient. The perception, on the contrary, is a relation of two terms, namely "a-in-the-relation-R-to-b," and the percipient. Since an object of perception cannot be nothing, we cannot perceive "a-in-the-relation-R-to-b" unless a is in the relation R to b. Hence a judgment of perception, according to the above definition, must be true. This does not mean that, in a judgment which *appears* to us to be one of perception, we are sure of not being in error, since we may err in thinking that our judgment has really been derived merely by analysis of what was perceived. But if our judgment has been so derived, it must be true. In fact, we may define *truth*, where such judgments are concerned, as consisting in the fact that there is a complex *corresponding* to the discursive thought which is the judgment. That is, when we judge "a has the relation R to b," our judgment is said to be *true* when there is a complex "a-in-the-relation-R-to-b," and is said to be *false* when this is not the case. This is a definition of truth and falsehood in relation to judgments of this kind.

It will be seen that, according to the above account, a judgment does not have a single object, namely the proposition, but has several interrelated objects. That is to say, the relation which constitutes judgment is not a relation of two terms, namely the judging mind and the proposition, but is a relation of several terms, namely the mind and what are called the constituents of the proposition. That is, when we judge (say) "this is red," what occurs is a relation of three terms, the mind, and "this," and red. On the other hand, when we *perceive* "the redness of this," there is a relation of two terms, namely

the mind and the complex object "the redness of this." When a judgment occurs, there is a certain complex entity, composed of the mind and the various objects of the judgment. When the judgment is *true*, in the case of the kind of judgments we have been considering, there is a corresponding complex of the *objects* of the judgment alone. Falsehood, in regard to our present class of judgments, consists in the absence of a corresponding complex composed of the objects alone. It follows from the above theory that a "proposition," in the sense in which a proposition is supposed to be *the* object of a judgment, is a false abstraction, because a judgment has several objects, not one. It is the severalness of the objects in judgment (as opposed to perception) which has led people to speak of thought as "discursive," though they do not appear to have realized clearly what was meant by this epithet.

Owing to the plurality of the objects of a single judgment, it follows that what we call a "proposition" (in the sense in which this is distinguished from the phrase expressing it) is not a single entity at all. That is to say, the phrase which expresses a proposition is what we call an "incomplete" symbol*; it does not have meaning in itself, but requires some supplementation in order to acquire a complete meaning. This fact is somewhat concealed by the circumstance that judgment in itself supplies a sufficient supplement, and that judgment in itself makes no *verbal* addition to the proposition. Thus "the proposition 'Socrates is human'" uses "Socrates is human" in a way which requires a supplement of some kind before it acquires a complete meaning; but when I judge "Socrates is human," the meaning is completed by the act of judging, and we no longer have an incomplete symbol. The fact that propositions are "incomplete symbols" is important philosophically, and is relevant at certain points in symbolic logic.

The judgments we have been dealing with hitherto are such as are of the same form as judgments of perception, *i.e.* their subjects are always particular and definite. But there are many judgments which are not of this form. Such are "all men are mortal," "I met a man," "some men are Greeks." Before dealing with such judgments, we will introduce some technical terms.

We will give the name of "a *complex*" to any such object as "*a* in the relation R to b" or "*a* having the quality q," or "*a* and *b* and *c* standing in the relation S." Broadly speaking, a *complex* is anything which occurs in the universe and is not simple. We will call a judgment *elementary* when it merely asserts such things as "*a* has the relation R to b," "*a* has the quality q" or "*a* and *b* and *c* stand in the relation S." Then an *elementary* judgment is true when there is a corresponding complex, and false when there is no corresponding complex.

But take now such a proposition as "all men are mortal." Here the judgment does not correspond to *one* complex, but to many, namely "Socrates

* See Chapter III.

is mortal," "Plato is mortal," "Aristotle is mortal," etc. (For the moment, it is unnecessary to inquire whether each of these does not require further treatment before we reach the ultimate complexes involved. For purposes of illustration, "Socrates is mortal" is here treated as an elementary judgment, though it is in fact not one, as will be explained later. Truly elementary judgments are not very easily found.) We do not mean to deny that there may be some relation of the concept *man* to the concept *mortal* which may be *equivalent* to "all men are mortal," but in any case this relation is not the same thing as what we affirm when we say that all men are mortal. Our judgment that all men are mortal collects together a number of elementary judgments. It is not, however, composed of these, since (*e.g.*) the fact that Socrates is mortal is no part of what we assert, as may be seen by considering the fact that our assertion can be understood by a person who has never heard of Socrates. In order to understand the judgment "all men are mortal," it is not necessary to know what men there are. We must admit, therefore, as a radically new kind of judgment, such general assertions as "all men are mortal." We assert that, given that x is human, x is always mortal. That is, we assert "x is mortal" of *every* x which is human. Thus we are able to judge (whether truly or falsely) that *all* the objects which have some assigned property also have some other assigned property. That is, given any propositional functions $\phi\hat{x}$ and $\psi\hat{x}$, there is a judgment asserting ψx with every x for which we have ϕx. Such judgments we will call *general judgments*.

It is evident (as explained above) that the definition of *truth* is different in the case of general judgments from what it was in the case of elementary judgments. Let us call the meaning of *truth* which we gave for elementary judgments "elementary truth." Then when we assert that it is true that all men are mortal, we shall mean that all judgments of the form "x is mortal," where x is a man, have elementary truth. We may define this as "truth of the second order" or "second-order truth." Then if we express the proposition "all men are mortal" in the form

"$(x) . x$ is mortal, where x is a man,"

and call this judgment p, then "p is true" must be taken to mean "p has second-order truth," which in turn means

"$(x) . $ 'x is mortal' has elementary truth, where x is a man."

In order to avoid the necessity for stating explicitly the limitation to which our variable is subject, it is convenient to replace the above interpretation of "all men are mortal" by a slightly different interpretation. The proposition "all men are mortal" is equivalent to "'x is a man' implies 'x is mortal,' with all possible values of x." Here x is not restricted to such values as are men, but may have any value with which "'x is a man' implies 'x is mortal'" is *significant*, *i.e.* either true or false. Such a proposition is called a "formal implication." The advantage of this form is that the values which the variable may take are given by the function to which it is the argument: the

values which the variable may take are all those with which the function is significant.

We use the symbol "$(x) . \phi x$" to express the general judgment which asserts all judgments of the form "ϕx." Then the judgment "all men are mortal" is equivalent to

"(x).'x is a man' implies 'x is a mortal,'"

i.e. (in virtue of the definition of implication) to

"$(x) . x$ is not a man or x is mortal."

As we have just seen, the meaning of *truth* which is applicable to this proposition is not the same as the meaning of *truth* which is applicable to "x is a man" or to "x is mortal." And generally, in any judgment $(x) . \phi x$, the sense in which this judgment is or may be true is not the same as that in which ϕx is or may be true. If ϕx is an elementary judgment, it is true when it *points to* a corresponding complex. But $(x) . \phi x$ does not point to a single corresponding complex: the corresponding complexes are as numerous as the possible values of x.

It follows from the above that such a proposition as "all the judgments made by Epimenides are true" will only be prima facie capable of truth if all his judgments are of the same order. If they are of varying orders, of which the nth is the highest, we may make n assertions of the form "all the judgments of order m made by Epimenides are true," where m has all values up to n. But no such judgment can include itself in its own scope, since such a judgment is always of higher order than the judgments to which it refers.

Let us consider next what is meant by the negation of a proposition of the form "$(x) . \phi x$." We observe, to begin with, that "ϕx in some cases," or "ϕx sometimes," is a judgment which is on a par with "ϕx in all cases," or "ϕx always." The judgment "ϕx sometimes" is true if one or more values of x exist for which ϕx is true. We will express the proposition "ϕx sometimes" by the notation "$(\exists x) . \phi x$," where "$\exists$" stands for "there exists," and the whole symbol may be read "there exists an x such that ϕx." We take the two kinds of judgment expressed by "$(x) . \phi x$" and "$(\exists x) . \phi x$" as primitive ideas. We also take as a primitive idea the negation of an *elementary* proposition. We can then define the negations of $(x) . \phi x$ and $(\exists x) . \phi x$. The negation of any proposition p will be denoted by the symbol "$\sim p$." Then the negation of $(x) . \phi x$ will be *defined* as meaning

"$(\exists x) . \sim \phi x$,"

and the negation of $(\exists x) . \phi x$ will be *defined* as meaning "$(x) . \sim \phi x$." Thus, in the traditional language of formal logic, the negation of a universal affirmative is to be defined as the particular negative, and the negation of the particular affirmative is to be defined as the universal negative. Hence the meaning of negation for such propositions is different from the meaning of negation for elementary propositions.

An analogous explanation will apply to disjunction. Consider the statement "either p, or ϕx always." We will denote the disjunction of two propositions p, q by "$p \vee q$." Then our statement is "$p . \vee . (x) . \phi x$." We will suppose that p is an elementary proposition, and that ϕx is always an elementary proposition. We take the disjunction of two elementary propositions as a primitive idea, and we wish to *define* the disjunction

"$p . \vee . (x) . \phi x$."

This may be defined as "$(x) . p \vee \phi x$," *i.e.* "either p is true, or ϕx is always true" is to mean "'p or ϕx' is always true." Similarly we will define

"$p . \vee . (\exists x) . \phi x$"

as meaning "$(\exists x) . p \vee \phi x$," *i.e.* we define "either p is true or there is an x for which ϕx is true" as meaning "there is an x for which either p or ϕx is true." Similarly we can define a disjunction of two universal propositions: "$(x) . \phi x . \vee . (y) . \psi y$" will be defined as meaning "$(x, y) . \phi x \vee \psi y$," *i.e.* "either ϕx is always true or ψy is always true" is to mean "'ϕx or ψy' is always true." By this method we obtain definitions of disjunctions containing propositions of the form $(x) . \phi x$ or $(\exists x) . \phi x$ in terms of disjunctions of elementary propositions; but the meaning of "disjunction" is not the same for propositions of the forms $(x) . \phi x, (\exists x) . \phi x$, as it was for elementary propositions.

Similar explanations could be given for implication and conjunction, but this is unnecessary, since these can be defined in terms of negation and disjunction.

IV. *Why a Given Function requires Arguments of a Certain Type.*

The considerations so far adduced in favour of the view that a function cannot significantly have as argument anything defined in terms of the function itself have been more or less indirect. But a direct consideration of the kinds of functions which have functions as arguments and the kinds of functions which have arguments other than functions will show, if we are not mistaken, that not only is it impossible for a function $\phi \hat{z}$ to have itself or anything derived from it as argument, but that, if $\psi \hat{z}$ is another function such that there are arguments a with which both "ϕa" and "ψa" are significant, then $\psi \hat{z}$ and anything derived from it cannot significantly be argument to $\phi \hat{z}$. This arises from the fact that a function is essentially an ambiguity, and that, if it is to occur in a definite proposition, it must occur in such a way that the ambiguity has disappeared, and a wholly unambiguous statement has resulted. A few illustrations will make this clear. Thus "$(x) . \phi x$," which we have already considered, is a function of $\phi \hat{x}$; as soon as $\phi \hat{x}$ is assigned, we have a definite proposition, wholly free from ambiguity. But it is obvious that we cannot substitute for the function something which is not a function: "$(x) . \phi x$" means "ϕx in all cases," and depends for its significance upon the fact that there are "cases" of ϕx, *i.e.* upon the

ambiguity which is characteristic of a function. This instance illustrates the fact that, when a function can occur significantly as argument, something which is not a function cannot occur significantly as argument. But conversely, when something which is not a function can occur significantly as argument, a function cannot occur significantly. Take, *e.g.* "*x* is a man," and consider "$\phi\hat{x}$ is a man." Here there is nothing to eliminate the ambiguity which constitutes $\phi\hat{x}$; there is thus nothing definite which is said to be a man. A function, in fact, is not a definite object, which could be or not be a man; it is a mere ambiguity awaiting determination, and in order that it may occur significantly it must receive the necessary determination, which it obviously does not receive if it is merely substituted for something determinate in a proposition*. This argument does not, however, apply directly as against such a statement as "$\{(x) . \phi x\}$ is a man." Common sense would pronounce such a statement to be meaningless, but it cannot be condemned on the ground of ambiguity in its subject. We need here a new objection, namely the following: A proposition is not a single entity, but a relation of several; hence a statement in which a proposition appears as subject will only be significant if it can be reduced to a statement about the terms which appear in the proposition. A proposition, like such phrases as "the so-and-so," where grammatically it appears as subject, must be broken up into its constituents if we are to find the true subject or subjects†. But in such a statement as "*p* is a man," where *p* is a proposition, this is not possible. Hence "$\{(x) . \phi x\}$ is a man" is meaningless.

V. *The Hierarchy of Functions and Propositions.*

We are thus led to the conclusion, both from the vicious-circle principle and from direct inspection, that the functions to which a given object *a* can be an argument are incapable of being arguments to each other, and that they have no term in common with the functions to which they can be arguments. We are thus led to construct a hierarchy. Beginning with *a* and the other terms which can be arguments to the same functions to which *a* can be argument, we come next to functions to which *a* is a possible argument, and then to functions to which such functions are possible arguments, and so on. But the hierarchy which has to be constructed is not so simple as might at first appear. The functions which can take *a* as argument form an illegitimate totality, and themselves require division into a hierarchy of functions. This is easily seen as follows. Let $f(\phi\hat{z}, x)$ be a function of the two variables $\phi\hat{z}$ and *x*. Then if, keeping *x* fixed for the moment, we assert this with all possible values of ϕ, we obtain a proposition:

$$(\phi) . f(\phi\hat{z}, x).$$

* Note that statements concerning the significance of a phrase containing "$\phi\hat{z}$" concern the *symbol* "$\phi\hat{z}$," and therefore do not fall under the rule that the elimination of the functional ambiguity is necessary to significance. Significance is a property of signs. Cf. pp. 40, 41.

† Cf. Chapter III.

Here, if *x* is variable, we have a function of *x*; but as this function involves a totality of values of $\phi\hat{z}$*, it cannot itself be one of the values included in the totality, by the vicious-circle principle. It follows that the totality of values of $\phi\hat{z}$ concerned in $(\phi) . f(\phi\hat{z}, x)$ is not the totality of all functions in which *x* can occur as argument, and that there is no such totality as that of all functions in which *x* can occur as argument.

It follows from the above that a function in which $\phi\hat{z}$ appears as argument requires that "$\phi\hat{z}$" should not stand for *any* function which is capable of a given argument, but must be restricted in such a way that none of the functions which are possible values of "$\phi\hat{z}$" should involve any reference to the totality of such functions. Let us take as an illustration the definition of identity. We might attempt to define "*x* is identical with *y*" as meaning "whatever is true of *x* is true of *y*," *i.e.* "ϕx always implies ϕy." But here, since we are concerned to assert all values of "ϕx implies ϕy" regarded as a function of ϕ, we shall be compelled to impose upon ϕ some limitation which will prevent us from including among values of ϕ values in which "all possible values of ϕ" are referred to. Thus for example "*x* is identical with *a*" is a function of *x*; hence, if it is a legitimate value of ϕ in "ϕx always implies ϕy," we shall be able to infer, by means of the above definition, that if *x* is identical with *a*, and *x* is identical with *y*, then *y* is identical with *a*. Although the conclusion is sound, the reasoning embodies a vicious-circle fallacy, since we have taken "$(\phi) . \phi x$ implies ϕa" as a possible value of ϕx, which it cannot be. If, however, we impose any limitation upon ϕ, it may happen, so far as appears at present, that with other values of ϕ we might have ϕx true and ϕy false, so that our proposed definition of identity would plainly be wrong. This difficulty is avoided by the "axiom of reducibility," to be explained later. For the present, it is only mentioned in order to illustrate the necessity and the relevance of the hierarchy of functions of a given argument.

Let us give the name "*a*-functions" to functions that are significant for a given argument *a*. Then suppose we take any selection of *a*-functions, and consider the proposition "*a* satisfies all the functions belonging to the selection in question." If we here replace *a* by a variable, we obtain an *a*-function; but by the vicious-circle principle this *a*-function cannot be a member of our selection, since it refers to the whole of the selection. Let the selection consist of all those functions which satisfy $f(\phi\hat{z})$. Then our new function is

$$(\phi) . \{f(\phi\hat{z}) \text{ implies } \phi x\},$$

where *x* is the argument. It thus appears that, whatever selection of *a*-functions we may make, there will be other *a*-functions that lie outside our

* When we speak of "values *of* $\phi\hat{z}$" it is ϕ, not *z*, that is to be assigned. This follows from the explanation in the note on p. 40. When the function itself is the variable, it is possible and simpler to write ϕ rather than $\phi\hat{z}$, except in positions where it is necessary to emphasize that an argument must be supplied to secure significance.

selection. Such a-functions, as the above instance illustrates, will always arise through taking a function of two arguments, $\phi\hat{z}$ and x, and asserting all or some of the values resulting from varying ϕ. What is necessary, therefore, in order to avoid vicious-circle fallacies, is to divide our a-functions into "types," each of which contains no functions which refer to the whole of that type.

When something is asserted or denied about all possible values or about some (undetermined) possible values of a variable, that variable is called *apparent*, after Peano. The presence of the words *all* or *some* in a proposition indicates the presence of an apparent variable; but often an apparent variable is really present where language does not at once indicate its presence. Thus for example "A is mortal" means "there is a time at which A will die." Thus a variable time occurs as apparent variable.

The clearest instances of propositions not containing apparent variables are such as express immediate judgments of perception, such as "this is red" or "this is painful," where "this" is something immediately given. In other judgments, even where at first sight no variable appears to be present, it often happens that there really is one. Take (say) "Socrates is human." To Socrates himself, the word "Socrates" no doubt stood for an object of which he was immediately aware, and the judgment "Socrates is human" contained no apparent variable. But to us, who only know Socrates by description, the word "Socrates" cannot mean what it meant to him; it means rather "the person having such-and-such properties," (say) "the Athenian philosopher who drank the hemlock." Now in all propositions about "the so-and-so" there is an apparent variable, as will be shown in Chapter III. Thus in what *we* have in mind when we say "Socrates is human" there is an apparent variable, though there was no apparent variable in the corresponding judgment as made by Socrates, provided we assume that there is such a thing as immediate awareness of oneself.

Whatever may be the instances of propositions not containing apparent variables, it is obvious that propositional functions whose values do not contain apparent variables are the source of propositions containing apparent variables, in the sense in which the function $\phi\hat{x}$ is the source of the proposition $(x) . \phi x$. For the values for $\phi\hat{x}$ do not contain the apparent variable x, which appears in $(x) . \phi x$; if they contain an apparent variable y, this can be similarly eliminated, and so on. This process must come to an end, since no proposition which we can apprehend can contain more than a finite number of apparent variables, on the ground that whatever we can apprehend must be of finite complexity. Thus we must arrive at last at a function of as many variables as there have been stages in reaching it from our original proposition, and this function will be such that its values contain no apparent variables. We may call this function the *matrix* of our original proposition and of any other

propositions and functions to be obtained by turning some of the arguments to the function into apparent variables. Thus for example, if we have a matrix-function whose values are $\phi(x, y)$, we shall derive from it

$(y) . \phi(x, y)$, which is a function of x,

$(x) . \phi(x, y)$, which is a function of y,

$(x, y) . \phi(x, y)$, meaning "$\phi(x, y)$ is true with all possible values of x and y."

This last is a proposition containing no *real* variable, *i.e.* no variable except apparent variables.

It is thus plain that all possible propositions and functions are obtainable from matrices by the process of turning the arguments to the matrices into apparent variables. In order to divide our propositions and functions into types, we shall, therefore, start from matrices, and consider how they are to be divided with a view to the avoidance of vicious-circle fallacies in the definitions of the functions concerned. For this purpose, we will use such letters as a, b, c, x, y, z, w, to denote objects which are neither propositions nor functions. Such objects we shall call *individuals*. Such objects will be constituents of propositions or functions, and will be *genuine* constituents, in the sense that they do not disappear on analysis, as (for example) classes do, or phrases of the form "the so-and-so."

The first matrices that occur are those whose values are of the forms

$$\phi x, \ \psi(x, y), \ \chi(x, y, z \ldots),$$

i.e. where the arguments, however many there may be, are all individuals. The functions $\phi, \psi, \chi \ldots$, since (by definition) they contain no apparent variables, and have no arguments except individuals, do not presuppose any totality of functions. From the functions $\psi, \chi \ldots$ we may proceed to form other functions of x, such as $(y) . \psi(x, y), (\exists y) . \psi(x, y), (y, z) . \chi(x, y, z)$, $(y) : (\exists z) . \chi(x, y, z)$, and so on. All these presuppose no totality except that of individuals. We thus arrive at a certain collection of functions of x, characterized by the fact that they involve no variables except individuals. Such functions we will call "*first-order* functions."

We may now introduce a notation to express "any first-order function." We will denote any first-order function by "$\phi ! \hat{x}$" and any value for such a function by "$\phi ! x$." Thus "$\phi ! x$" stands for any value for any function which involves no variables except individuals. It will be seen that "$\phi ! x$" is itself a function of *two* variables, namely $\phi ! \hat{z}$ and x. Thus $\phi ! x$ involves a variable which is not an individual, namely $\phi ! \hat{z}$. Similarly "$(x) . \phi ! x$" is a function of the variable $\phi ! \hat{z}$, and thus involves a variable other than an individual. Again, if a is a given individual,

"$\phi ! x$ implies $\phi ! a$ with all possible values of ϕ"

is a function of x, but it is not a function of the form $\phi ! x$, because it involves an (apparent) variable ϕ which is not an individual. Let us give the name "predicate" to any first-order function $\phi ! \hat{x}$. (This use of the word "predicate"

is only proposed for the purposes of the present discussion.) Then the statement "$\phi \,!\, x$ implies $\phi \,!\, a$ with all possible values of ϕ" may be read "all the predicates of x are predicates of a." This makes a statement about x, but does not attribute to x a *predicate* in the special sense just defined.

Owing to the introduction of the variable first-order function $\phi \,!\, \hat{z}$, we now have a new set of matrices. Thus "$\phi \,!\, x$" is a function which contains no apparent variables, but contains the two real variables $\phi \,!\, \hat{z}$ and x. (It should be observed that when ϕ is assigned, we may obtain a function whose values do involve individuals as apparent variables, for example if $\phi \,!\, x$ is $(y) \,.\, \psi \,(x, y)$. But so long as ϕ is variable, $\phi \,!\, x$ contains no apparent variables.) Again, if a is a definite individual, $\phi \,!\, a$ is a function of the one variable $\phi \,!\, \hat{z}$. If a and b are definite individuals, "$\phi \,!\, a$ implies $\psi \,!\, b$" is a function of the two variables $\phi \,!\, \hat{z}$, $\psi \,!\, \hat{z}$, and so on. We are thus led to a whole set of new matrices,

$$f\,(\phi \,!\, \hat{z}), \; g\,(\phi \,!\, \hat{z}, \psi \,!\, \hat{z}), \; F\,(\phi \,!\, \hat{z}, x), \; \text{and so on.}$$

These matrices contain individuals and first-order functions as arguments, but (like all matrices) they contain no apparent variables. Any such matrix, if it contains more than one variable, gives rise to new functions of one variable by turning all its arguments except one into apparent variables. Thus we obtain the functions

$(\phi) \,.\, g\,(\phi \,!\, \hat{z}, \psi \,!\, \hat{z})$, which is a function of $\psi \,!\, \hat{z}$.

$(x) \,.\, F\,(\phi \,!\, \hat{z}, x)$, which is a function of $\phi \,!\, \hat{z}$.

$(\phi) \,.\, F\,(\phi \,!\, \hat{z}, x)$, which is a function of x.

We will give the name of *second-order matrices* to such matrices as have first-order functions among their arguments, and have no arguments except first-order functions and individuals. (It is not *necessary* that they should have individuals among their arguments.) We will give the name of *second-order functions* to such as either are second-order matrices or are derived from such matrices by turning some of the arguments into apparent variables. It will be seen that either an individual or a first-order function may appear as argument to a second-order function. Second-order functions are such as contain variables which are first-order functions, but contain no other variables except (possibly) individuals.

We now have various new classes of functions at our command. In the first place, we have second-order functions which have one argument which is a first-order function. We will denote a variable function of this kind by the notation $f\,!\,(\hat{\phi} \,!\, \hat{z})$, and any value of such a function by $f\,!\,(\phi \,!\, \hat{z})$. Like $\phi \,!\, x$, $f\,!\,(\phi \,!\, \hat{z})$ is a function of two variables, namely $f\,!\,(\hat{\phi} \,!\, \hat{z})$ and $\phi \,!\, \hat{z}$. Among possible values of $f\,!\,(\phi \,!\, \hat{z})$ will be $\phi \,!\, a$ (where a is constant), $(x) \,.\, \phi \,!\, x$, $(\exists x) \,.\, \phi \,!\, x$, and so on. (These result from assigning a value to f, leaving ϕ to be assigned.) We will call such functions "predicative functions of first-order functions."

In the second place, we have second-order functions of two arguments, one of which is a first-order function while the other is an individual. Let us denote undetermined values of such functions by the notation

$$f\,!\,(\phi \,!\, \hat{z}, x).$$

As soon as x is assigned, we shall have a predicative function of $\phi \,!\, \hat{z}$. If our function contains no first-order function as apparent variable, we shall obtain a predicative function of x if we assign a value to $\phi \,!\, \hat{z}$. Thus, to take the simplest possible case, if $f\,!\,(\phi \,!\, \hat{z}, x)$ is $\phi \,!\, x$, the assignment of a value to ϕ gives us a predicative function of x, in virtue of the definition of "$\phi \,!\, x$." But if $f\,!\,(\phi \,!\, \hat{z}, x)$ contains a first-order function as apparent variable, the assignment of a value to $\phi \,!\, \hat{z}$ gives us a second-order function of x.

In the third place, we have second-order functions of individuals. These will all be derived from functions of the form $f\,!\,(\phi \,!\, \hat{z}, x)$ by turning ϕ into an apparent variable. We do not, therefore, need a new notation for them.

We have also second-order functions of two first-order functions, or of two such functions and an individual, and so on.

We may now proceed in exactly the same way to third-order matrices, which will be functions containing second-order functions as arguments, and containing no apparent variables, and no arguments except individuals and first-order functions and second-order functions. Thence we shall proceed, as before, to third-order functions; and so we can proceed indefinitely. If the highest order of variable occurring in a function, whether as argument or as apparent variable, is a function of the nth order, then the function in which it occurs is of the $n + 1$th order. We do not arrive at functions of an infinite order, because the number of arguments and of apparent variables in a function must be finite, and therefore every function must be of a finite order. Since the orders of functions are only defined step by step, there can be no process of "proceeding to the limit," and functions of an infinite order cannot occur.

We will define a function of one variable as *predicative* when it is of the next order above that of its argument, *i.e.* of the lowest order compatible with its having that argument. If a function has several arguments, and the highest order of function occurring among the arguments is the nth, we call the function predicative if it is of the $n + 1$th order, *i.e.* again, if it is of the lowest order compatible with its having the arguments it has. A function of several arguments is predicative if there is one of its arguments such that, when the other arguments have values assigned to them, we obtain a predicative function of the one undetermined argument.

It is important to observe that all possible functions in the above hierarchy can be obtained by means of predicative functions and apparent variables. Thus, as we saw, second-order functions of an individual x are of the form

$(\phi) \,.\, f\,!\,(\phi \,!\, \hat{z}, x)$ or $(\exists \phi) \,.\, f\,!\,(\phi \,!\, \hat{z}, x)$ or $(\phi, \psi) \,.\, f\,!\,(\phi \,!\, \hat{z}, \psi \,!\, \hat{z}, x)$ or etc.,

where f is a second-order predicative function. And speaking generally, a

non-predicative function of the nth order is obtained from a predicative function of the nth order by turning all the arguments of the $n-1$th order into apparent variables. (Other arguments also may be turned into apparent variables.) Thus we need not introduce as variables any functions except predicative functions. Moreover, to obtain any function of one variable x, we need not go beyond predicative functions of *two* variables. For the function $(\psi).f!(\phi!\hat{z},\psi!\hat{z},x)$, where f is given, is a function of $\phi!\hat{z}$ and x, and is predicative. Thus it is of the form $F!(\phi!\hat{z},x)$, and therefore $(\phi,\psi).f!(\phi!\hat{z},\psi!\hat{z},x)$ is of the form $(\phi).F!(\phi!\hat{z},x)$. Thus speaking generally, by a succession of steps we find that, if $\phi!\hat{u}$ is a predicative function of a sufficiently high order, any assigned non-predicative function of x will be of one of the two forms

$$(\phi).F!(\phi!\hat{u},x),\quad(\exists\phi).F!(\phi!\hat{u},x),$$

where F is a predicative function of $\phi!\hat{u}$ and x.

The nature of the above hierarchy of functions may be restated as follows. A function, as we saw at an earlier stage, presupposes as part of its meaning the totality of its values, or, what comes to the same thing, the totality of its possible arguments. The arguments to a function may be functions or propositions or individuals. (It will be remembered that individuals were defined as whatever is neither a proposition nor a function.) For the present we neglect the case in which the argument to a function is a proposition. Consider a function whose argument is an individual. This function presupposes the totality of individuals; but unless it contains functions as apparent variables, it does not presuppose any totality of functions. If, however, it does contain a function as apparent variable, then it cannot be defined until some totality of functions has been defined. It follows that we must first define the totality of those functions that have individuals as arguments and contain no functions as apparent variables. These are the *predicative* functions of individuals. Generally, a predicative function of a variable argument is one which involves no totality except that of the possible values of the argument, and those that are presupposed by any one of the possible arguments. Thus a predicative function of a variable argument is any function which can be specified without introducing new kinds of variables not necessarily presupposed by the variable which is the argument.

A closely analogous treatment can be developed for propositions. Propositions which contain no functions and no apparent variables may be called *elementary propositions*. Propositions which are not elementary, which contain no functions, and no apparent variables except individuals, may be called *first-order propositions*. (It should be observed that no variables except *apparent* variables can occur in a proposition, since whatever contains a *real* variable is a function, not a proposition.) Thus elementary and first-order propositions will be values of first-order functions. (It should be remembered

that a function is not a constituent in one of its values: thus for example the function "$\hat{x}$ is human" is not a constituent of the proposition "Socrates is human.") Elementary and first-order propositions presuppose no totality except (at most) the totality of individuals. They are of one or other of the three forms

$$\phi!x;\quad(x).\phi!x;\quad(\exists x).\phi!x,$$

where $\phi!x$ is a predicative function of an individual. If follows that, if p represents a variable elementary proposition or a variable first-order proposition, a function fp is either $f(\phi!x)$ or $f\{(x).\phi!x\}$ or $f\{(\exists x).\phi!x\}$. Thus a function of an elementary or a first-order proposition may always be reduced to a function of a first-order function. It follows that a proposition involving the totality of first-order propositions may be reduced to one involving the totality of first-order functions; and this obviously applies equally to higher orders. The propositional hierarchy can, therefore, be derived from the functional hierarchy, and we may define a proposition of the nth order as one which involves an apparent variable of the $n-1$th order in the functional hierarchy. The propositional hierarchy is never required in practice, and is only relevant for the solution of paradoxes; hence it is unnecessary to go into further detail as to the types of propositions.

VI. *The Axiom of Reducibility.*

It remains to consider the "axiom of reducibility." It will be seen that, according to the above hierarchy, no statement can be made significantly about "all a-functions," where a is some given object. Thus such a notion as "all properties of a," meaning "all functions which are true with the argument a," will be illegitimate. We shall have to distinguish the order of function concerned. We can speak of "all predicative properties of a," "all second-order properties of a," and so on. (If a is not an individual, but an object of order n, "second-order properties of a" will mean "functions of order $n+2$ satisfied by a.") But we cannot speak of "all properties of a." In some cases, we can see that some statement will hold of "all nth-order properties of a," whatever value n may have. In such cases, no practical harm results from regarding the statement as being about "all properties of a," provided we remember that it is really a number of statements, and not a single statement which could be regarded as assigning another property to a, over and above all properties. Such cases will always involve some systematic ambiguity, such as that involved in the meaning of the word "truth," as explained above. Owing to this systematic ambiguity, it will be possible, sometimes, to combine into a single verbal statement what are really a number of different statements, corresponding to different orders in the hierarchy. This is illustrated in the case of the liar, where the statement "all A's statements are false" should be broken up into different statements referring to his statements of various orders, and attributing to each the appropriate kind of falsehood.

The axiom of reducibility is introduced in order to legitimate a great mass of reasoning, in which, prima facie, we are concerned with such notions as "all properties of a" or "all a-functions," and in which, nevertheless, it seems scarcely possible to suspect any substantial error. In order to state the axiom, we must first define what is meant by "formal equivalence." Two functions $\phi\hat{x}, \psi\hat{x}$ are said to be "formally equivalent" when, with every possible argument x, ϕx is equivalent to ψx, i.e. ϕx and ψx are either both true or both false. Thus two functions are formally equivalent when they are satisfied by the same set of arguments. The axiom of reducibility is the assumption that, given any function $\phi\hat{x}$, there is a formally equivalent *predicative* function, i.e. there is a predicative function which is true when ϕx is true and false when ϕx is false. In symbols, the axiom is:

$$\vdash : (\exists\psi) : \phi x . \equiv_x . \psi \,!\, x.$$

For two variables, we require a similar axiom, namely: Given any function $\phi(\hat{x}, \hat{y})$, there is a formally equivalent *predicative* function, i.e.

$$\vdash : (\exists\psi) : \phi(x, y) . \equiv_{x,y} . \psi\,!\,(x, y).$$

In order to explain the purposes of the axiom of reducibility, and the nature of the grounds for supposing it true, we shall first illustrate it by applying it to some particular cases.

If we call a *predicate* of an object a predicative function which is true of that object, then the predicates of an object are only some among its properties. Take for example such a proposition as "Napoleon had all the qualities that make a great general." We may interpret this as meaning "Napoleon had all the predicates that make a great general." Here there is a predicate which is an apparent variable. If we put "$f(\phi\,!\,\hat{z})$" for "$\phi\,!\,\hat{z}$ is a predicate required in a great general," our proposition is

$$(\phi) : f(\phi\,!\,\hat{z}) \text{ implies } \phi\,!\,(\text{Napoleon}).$$

Since this refers to a totality of predicates, it is not itself a predicate of Napoleon. It by no means follows, however, that there is not some one predicate common and peculiar to great generals. In fact, it is certain that there is such a predicate. For the number of great generals is finite, and each of them certainly possessed some predicate not possessed by any other human being —for example, the exact instant of his birth. The disjunction of such predicates will constitute a predicate common and peculiar to great generals*. If we call this predicate $\psi\,!\,\hat{z}$, the statement we made about Napoleon was equivalent to $\psi\,!\,(\text{Napoleon})$. And this equivalence holds equally if we substitute any other individual for Napoleon. Thus we have arrived at a predicate which is always equivalent to the property we ascribed to Napoleon, i.e. it belongs to those objects which have this property, and to no others. The axiom of reducibility states that such a predicate always exists, i.e. that any property

---

* When a (finite) set of predicates is given by actual enumeration, their disjunction is a predicate, because no predicate occurs as apparent variable in the disjunction.

of an object belongs to the same collection of objects as those that possess some predicate.

We may next illustrate our principle by its application to *identity*. In this connection, it has a certain affinity with Leibniz's identity of indiscernibles. It is plain that, if x and y are identical, and ϕx is true, then ϕy is true. Here it cannot matter what sort of function $\phi\hat{x}$ may be: the statement must hold for *any* function. But we cannot say, conversely: "If, with all values of ϕ, ϕx implies ϕy, then x and y are identical"; because "all values of ϕ" is inadmissible. If we wish to speak of "all values of ϕ," we must confine ourselves to functions of one order. We may confine ϕ to predicates, or to second-order functions, or to functions of any order we please. But we must necessarily leave out functions of all but one order. Thus we shall obtain, so to speak, a hierarchy of different degrees of identity. We may say "all the predicates of x belong to y," "all second-order properties of x belong to y," and so on. Each of these statements implies all its predecessors: for example, if all second-order properties of x belong to y, then all predicates of x belong to y, for to have all the predicates of x is a second-order property, and this property belongs to x. But we cannot, without the help of an axiom, argue conversely that if all the predicates of x belong to y, all the second-order properties of x must also belong to y. Thus we cannot, without the help of an axiom, be sure that x and y are identical if they have the same predicates. Leibniz's identity of indiscernibles supplied this axiom. It should be observed that by "indiscernibles" he cannot have meant two objects which agree as to *all* their properties, for one of the properties of x is to be identical with x, and therefore this property would necessarily belong to y if x and y agreed in *all* their properties. Some limitation of the common properties necessary to make things indiscernible is therefore implied by the necessity of an axiom. For purposes of illustration (not of interpreting Leibniz) we may suppose the common properties required for indiscernibility to be limited to predicates. Then the identity of indiscernibles will state that if x and y agree as to all their predicates, they are identical. This can be proved if we assume the axiom of reducibility. For, in that case, every property belongs to the same collection of objects as is defined by some predicate. Hence there is some predicate common and peculiar to the objects which are identical with x. This predicate belongs to x, since x is identical with itself; hence it belongs to y, since y has all the predicates of x; hence y is identical with x. It follows that we may *define* x and y as identical when all the predicates of x belong to y, i.e. when $(\phi) : \phi\,!\,x . \supset . \phi\,!\,y$. We therefore adopt the following definition of identity*:

$$x = y . = : (\phi) : \phi\,!\,x . \supset . \phi\,!\,y \quad \text{Df.}$$

---

* Note that in this definition the second sign of equality is to be regarded as combining with "Df" to form one symbol; what is defined is the sign of equality *not* followed by the letters "Df."

But apart from the axiom of reducibility, or some axiom equivalent in this connection, we should be compelled to regard identity as indefinable, and to admit (what seems impossible) that two objects may agree in all their predicates without being identical.

The axiom of reducibility is even more essential in the theory of classes. It should be observed, in the first place, that if we assume the existence of classes, the axiom of reducibility can be proved. For in that case, given any function $\phi\hat{z}$ of whatever order, there is a class α consisting of just those objects which satisfy $\phi\hat{z}$. Hence "ϕx" is equivalent to "x belongs to α." But "x belongs to α" is a statement containing no apparent variable, and is therefore a predicative function of x. Hence if we assume the existence of classes, the axiom of reducibility becomes unnecessary. The assumption of the axiom of reducibility is therefore a smaller assumption than the assumption that there are classes. This latter assumption has hitherto been made unhesitatingly. However, both on the ground of the contradictions, which require a more complicated treatment if classes are assumed, and on the ground that it is always well to make the smallest assumption required for proving our theorems, we prefer to assume the axiom of reducibility rather than the existence of classes. But in order to explain the use of the axiom in dealing with classes, it is necessary first to explain the theory of classes, which is a topic belonging to Chapter III. We therefore postpone to that Chapter the explanation of the use of our axiom in dealing with classes.

It is worth while to note that all the purposes served by the axiom of reducibility are equally well served if we assume that there is always a function of the nth order (where n is fixed) which is formally equivalent to $\phi\hat{x}$, whatever may be the order of $\phi\hat{x}$. Here we shall mean by "a function of the nth order" a function of the nth order relative to the arguments to $\phi\hat{x}$; thus if these arguments are absolutely of the mth order, we assume the existence of a function formally equivalent to $\phi\hat{x}$ whose absolute order is the $m+n$th. The axiom of reducibility in the form assumed above takes $n=1$, but this is not necessary to the use of the axiom. It is also unnecessary that n should be the same for different values of m; what is necessary is that n should be constant so long as m is constant. What is needed is that, where extensional functions of functions are concerned, we should be able to deal with any a-function by means of some formally equivalent function of a given type, so as to be able to obtain results which would otherwise require the illegitimate notion of "all a-functions"; but it does not matter what the given type is. It does not appear, however, that the axiom of reducibility is rendered appreciably more plausible by being put in the above more general but more complicated form.

The axiom of reducibility is equivalent to the assumption that "any

combination or disjunction of predicates* is equivalent to a single predicate," i.e. to the assumption that, if we assert that x has all the predicates that satisfy a function $f(\phi!\hat{z})$, there is some one predicate which x will have whenever our assertion is true, and will not have whenever it is false, and similarly if we assert that x has some one of the predicates that satisfy a function $f(\phi!\hat{z})$. For by means of this assumption, the order of a non-predicative function can be lowered by one; hence, after some finite number of steps, we shall be able to get from any non-predicative function to a formally equivalent predicative function. It does not seem probable that the above assumption could be substituted for the axiom of reducibility in symbolic deductions, since its use would require the explicit introduction of the further assumption that by a finite number of downward steps we can pass from any function to a predicative function, and this assumption could not well be made without developments that are scarcely possible at an early stage. But on the above grounds it seems plain that in fact, if the above alternative axiom is true, so is the axiom of reducibility. The converse, which completes the proof of equivalence, is of course evident.

VII. *Reasons for Accepting the Axiom of Reducibility.*

That the axiom of reducibility is self-evident is a proposition which can hardly be maintained. But in fact self-evidence is never more than a part of the reason for accepting an axiom, and is never indispensable. The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it. If the axiom is apparently self-evident, that only means, practically, that it is nearly indubitable; for things have been thought to be self-evident and have yet turned out to be false. And if the axiom itself is nearly indubitable, that merely adds to the inductive evidence derived from the fact that its consequences are nearly indubitable: it does not provide new evidence of a radically different kind. Infallibility is never attainable, and therefore some element of doubt should always attach to every axiom and to all its consequences. In formal logic, the element of doubt is less than in most sciences, but it is not absent, as appears from the fact that the paradoxes followed from premises which were not previously known to require limitations. In the case of the axiom of reducibility, the inductive evidence in its favour is very strong, since the reasonings which it permits and the results to which it leads are all such as appear valid. But although it seems very improbable that the axiom should turn out to be false,

---

* Here the combination or disjunction is supposed to be given intensionally. If given extensionally (i.e. by enumeration), no assumption is required; but in this case the number of predicates concerned must be finite.

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, Ω say. But in that case the series of all ordinals including Ω has the ordinal number $\Omega + 1$, which must be greater than Ω. Hence Ω 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 nth figure in the nth decimal is p, let the nth 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 nth figure in N is different from the nth figure in the nth of the decimals composing E, and therefore N is different from the nth 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 nth order, a proposition in which p occurs as an apparent variable is not of the nth 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 nth order" is truth or falsehood of a higher order than the nth. 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 α is a class, the statement "α is not a member of α" 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 ϕ. But (assuming, as will appear in Chapter III, that R presupposes its defining function) this would require that ϕ should be able to take as argument an object which is defined in terms of ϕ, 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 nth. 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 τ, 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.