THE JOURNAL OF PHILOSOPHY

VOLUME LXVIII, NO. 8, APRIL 22, 1971

THE ITERATIVE CONCEPTION OF SET

A SET, according to Cantor, is "any collection...into a whole of definite, well-distinguished objects...of our intuition or thought."[1] Cantor alo sdefined a set as a "many, which can be thought of as one, i.e., a totality of definite elements that can be combined into a whole by a law."[2] One might object to the first definition on the grounds that it uses the concepts of *collection* and *whole*, which are notions no better understood than that of *set*, that there ought to be sets of objects that are not objects of our thought, that 'intuition' is a term laden with a theory of knowledge that no one should believe, that *any* object is "definite," that there should be sets of ill-distinguished objects, such as waves and trains, etc., etc. And one might object to the second on the grounds that 'a many' is ungrammatical, that if something is "a many" it should hardly be thought of as one, that *totality* is as obscure as *set*, that it is far from clear how laws can combine anything into a whole, that there ought to be other combinations into a whole than those effected by "laws," etc., etc. But it cannot be denied that Cantor's definitions could be used by a person to identify and gain some understanding of the sort of object of which Cantor wished to treat. Moreover, they do suggest—although, it must be conceded, only very faintly—two important characteristics of sets: that a set is "determined" by its elements in the sense that sets with exactly the same elements are

[1] "Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die 'Elemente' von M genannt werden) zu einem Ganzen." Georg Cantor, *Gesammelte Abhandlungen*, Ernst Zermelo, ed. (Berlin, 1932), p. 282.

[2] "...jedes Viele, welches sich als Eines denken lässt, d.h. jeden Inbegriff bestimmter Elemente, welcher durch ein Gesetz zu einem Ganzen verbunden werden kann" (*ibid.*, p. 204).

identical, and that, in a sense, the clarification of which is one of the principal objects of the theory whose rationale we shall give, the elements of a set are "prior to" it.

It is not to be presumed that the concepts of *set* and *member of* can be explained or defined by means of notions that are simpler or conceptually more basic. However, as a theory about sets might itself provide the sort of elucidation about sets and membership that good definitions might be hoped to offer, there is no reason for such a theory to begin with, or even contain, a definition of 'set'. That we are unable to given informative definitions of *not* or *for some* does not and should not prevent the development of quantificational logic, which provides us with significant information about these concepts.

I. NAIVE SET THEORY

Here is an idea about sets that might occur to us quite naturally, and is perhaps suggested by Cantor's definition of a set as a totality of definite elements that can be combined into a whole *by a law.*

By the law of excluded middle, any (one-place) predicate in any language either applies to a given object or does not. So, it would seem, to any predicate there correspond two sorts of thing: the sort of thing to which the predicate applies (of which it is true) and the sort of thing to which it does not apply. So, it would seem, for any predicate there is a set of all and only those things to which it applies (as well as a set of just those things to which it does not apply). Any set whose members are exactly the things to which the predicate applies—by the axiom of extensionality, there cannot be *two* such sets—is called the *extension* of the predicate. Our thought might therefore be put: "Any predicate has an extension." We shall call this proposition, together with the argument for it, the *naive conception of set.*

The argument has great force. How could there *not* be a collection, or set, of just those things to which any given predicate applied? Isn't anything to which a predicate applies similar to all other things to which it applies in precisely the respect that it applies to them; and how could there fail to be a set of all things similar to one another in this respect? Wouldn't it be extremely implausible to say, of any particular predicate one might consider, that there weren't two kinds of thing it determined, namely, a kind of thing of which it is true, and a kind of thing of which it is not true? And why should one not take these kinds of things to be sets? Aren't kinds sets? If not, what is the difference?

Let us denote by '$\mathcal{K}$' a certain standardly formalized first-order language, whose variables range over all sets and individuals (= nonsets), and whose nonlogical constants are a one-place predicate letter 'S' abbreviating 'is a set', and a two-place predicate letter 'ϵ', abbreviating 'is a member of'. Which sentences of this language, together with their consequences, do we believe state truths about sets? Otherwise put, which formulas of $\mathcal{K}$ should we take as axioms of a set theory on the strength of our beliefs about sets?

If the naive conception of set is correct, there should (at least) be a set of just those things to which ϕ applies, if ϕ is a formula of $\mathcal{K}$. So (the universal closure of) $\ulcorner(\exists y)(Sy \mathbin{\&} (x)(x \epsilon y \leftrightarrow \phi)))\urcorner$ should express a truth about sets (if no occurrence of 'y' in ϕ is free).

We call the theory whose axioms are the axiom of extensionality (to which we later recur), i.e., the sentence

$$(x)(y)(Sx \mathbin{\&} Sy \mathbin{\&} (z)(z \epsilon x \leftrightarrow x \epsilon y) \rightarrow x = y)$$

and all formulas $\ulcorner(\exists y)(Sy \mathbin{\&} (x)(x \epsilon y \leftrightarrow \phi))\urcorner$ (where 'y' does not occur free in ϕ) *naive set theory.*

Some of the axioms of naive set theory are the formulas

$$(\exists y)(Sy \mathbin{\&} (x)(x \epsilon y \leftrightarrow x \neq x))$$
$$(\exists y)(Sy \mathbin{\&} (x)(x \epsilon y \leftrightarrow (x = z \vee x = w)))$$
$$(\exists y)(Sy \mathbin{\&} (x)(x \epsilon y \leftrightarrow (\exists w)(x \epsilon w \mathbin{\&} w \epsilon z)))$$
$$(\exists y)(Sy \mathbin{\&} (x)(x \epsilon y \leftrightarrow (Sx \mathbin{\&} x = x)))$$

The first of these formulas states that there is a set that contains no members. By the axiom of extensionality, there can be at most one such set. The second states that there is a set whose sole members are z and w; the third, that there is a set whose members are just the members of members of z.

The last, which states that there is a set that contains all sets whatsoever, is rather anomalous; for if there is a set that contains all sets, a universal set, that set contains itself, and perhaps the mind ought to boggle at the idea of something's *containing* itself. Nevertheless, naive set theory is simple to state, elegant, initially quite credible, and natural in that it articulates a view about sets that might occur to one quite naturally.

Alas, it is inconsistent.

Proof of the Inconsistency of Naive Set Theory (Russell's paradox)

No set can contain all and only those sets which do not contain themselves. For if any such set existed, if it contained itself, then, as it contains *only* those sets which do not contain themselves, it would not contain itself; but if it did not contain itself, then, as it contains *all* those sets which do not contain themselves, it would contain itself. Thus any such set would have to contain itself if and only if it did not contain itself. Consequently, there is no set that contains all and only those sets which do not contain themselves.

This argument, which uses no axioms of naive set theory, or any other set theory, shows that the sentence

$$\sim (\exists y)(Sy \ \& \ (x)(x \epsilon y \leftrightarrow (Sx \ \& \sim x \epsilon x)))$$

is *logically valid* and, hence, is a theorem of any theory that is expressed in $\mathcal{K}$. But one of the axioms and, hence, one of the theorems, of naive set theory is the sentence

$$(\exists y)(Sy \ \& \ (x)(x \epsilon y \leftrightarrow (Sx \ \& \sim x \epsilon x)))$$

Naive set theory is therefore inconsistent.

II. THE ITERATIVE CONCEPTION OF SET

Faced with the inconsistency of naive set theory, one might come to believe that any decision to adopt a system of axioms about sets would be *arbitrary* in that no explanation could be given why the particular system adopted had any greater claim to describe what we conceive sets and the membership relation to be like than some other system, perhaps incompatible with the one chosen. One might think that no answer could be given to the question: why adopt *this* particular system rather than that or this other one? One might suppose that any apparently consistent theory of sets would have to be unnatural in some way or fragmentary, and that, if consistent, its consistency would be due to certain provisions that were laid down for the express purpose of avoiding the paradoxes that show naive set theory inconsistent, but that lack any independent motivation.

One might imagine all this; but there is another view of sets: the *iterative conception of set*, as it is sometimes called, which often strikes people as entirely natural, free from artificiality, not at all ad hoc, and one they might perhaps have formulated themselves.

It is, perhaps, no more natural a conception than the naive conception, and certainly not quite so simple to describe. On the other hand, it is, as far as we know, consistent: not only are the sets whose existence would lead to contradiction not assumed to exist in the axioms of the theories that express the iterative conception, but the more than fifty years of experience that practicing set theorists have had with this conception have yielded a good understanding of what can and what cannot be proved in these theories, and at present there just is no suspicion at all that they are inconsistent.[3]

[3] The conception is well known among logicians; a rather different version of it is sketched in chapter IX of Joseph R. Shoenfield's *Mathematical Logic* (Reading, Mass.: Addison-Wesley, 1967). I learned of it principally from Putnam, Kripke, and Donald Martin. Authors of set-theory texts either omit it or relegate it to back pages; philosophers, in the main, seem to be unaware of it, or of the preeminence of ZF, which may be said to embody it. It is due primarily to Zermelo and Russell.

The standard, first-order theory that expresses the iterative conception of set as fully as a first-order theory in the language $\mathcal{L}$ of set theory[4] can, is known as *Zermelo-Fraenkel set theory*, or 'ZF' for short. There are other theories besides ZF that embody the iterative conception: one of them, Zermelo set theory, or "Z", which will occupy us shortly, is a *subsystem* of ZF in the sense that any theorem of Z is also a theorem of ZF; two others, von-Neumann-Bernays-Gödel set theory and Morse-Kelley set theory, are supersystems (or extensions) of ZF, but they are most commonly formulated as second-order theories.

Other theories of sets, incompatible with ZF, have been proposed.[5] These theories appear to lack a motivation that is independent of the paradoxes in the following sense: they are not, as Russell has written, "such as even the cleverest logician would have thought of if he had not known of the contradictions."[6] A final and satisfying resolution to the set-theoretical paradoxes cannot be embodied in a theory that blocks their derivation by artificial technical restrictions on the set of axioms that are imposed *only because* paradox would otherwise ensue; these other theories survive only through such artificial devices. ZF alone (together with its extensions and subsystems) is not only a consistent (apparently) but also an independently motivated theory of sets: there is, so to speak, a "thought behind it" about the nature of sets which might have been put forth even if, impossibly, naive set theory had been consistent. The thought, moreover, can be described in a rough, but informative way without first stating the theory the thought is behind.

In order to see why a conception of set other than the naive conception might be desired even if the naive conception were consistent, let us take another look at naive set theory and the anomalousness of its axiom, '$(\exists y)(Sy \,\&\, (x)(x \epsilon y \leftrightarrow (Sx \,\&\, x = x)))$'.

According to this axiom there is a set that contains all sets, and therefore there is a set that contains itself. It is important to realize how odd the idea of something's containing itself is. Of course a set can and must *include* itself (as a subset). But *contain* itself? Whatever tenuous hold on the concepts of *set* and *member* were given one by Cantor's definitions of 'set' and one's ordinary understanding of 'element', 'set', 'collection', etc. is altogether lost if one is to suppose that some sets are members of themselves. The idea is paradoxical not in the sense that it is contradictory to suppose that some set is

[4] $\mathcal{L}$ contains (countably many) variables, ranging over (pure) sets, '$=$', and 'ϵ', which is its sole nonlogical constant.

[5] For example, Quine's systems NF and ML.

[6] Russell, *My Philosophical Development* (New York: Simon & Schuster, 1959), p. 80.

a member of itself, for, after all, '$(\exists x)(Sx \,\&\, x\epsilon x)$' is obviously consistent, but that if one understands 'ϵ' as meaning 'is a member of', it is very, very peculiar to suppose it true. For when one is told that a set is a collection into a whole of definite elements of our thought, one thinks: Here are some things. Now we bind them up into a whole.[7] *Now* we have a set. We don't suppose that what we come up with after combining some elements into a whole could have been one of the very things we combined (not, at least, if we are combining two or more elements).

If $(\exists x)(Sx \,\&\, x\epsilon x)$, then $(\exists x)(\exists y)(Sx \,\&\, Sy \,\&\, x\epsilon y \,\&\, y\epsilon x)$. The supposition that there are sets x and y each of which belongs to the other is almost as strange as the supposition that some set is a self-member. There is of course an infinite sequence of such cyclical pathologies: $(\exists x)(\exists y)(\exists z)(Sx \,\&\, Sy \,\&\, Sz \,\&\, x\epsilon y \,\&\, y\epsilon z \,\&\, z\epsilon x)$, etc. Only slightly less pathological are the suppositions that there is an ungrounded set,[8] or that there is an infinite sequence of sets $x_0, x_1, x_2, \ldots$, each term of which belongs to the previous one.

There does not seem to be any argument that is guaranteed to persuade someone who really does not see the peculiarity of a set's belonging to itself, or to one of its members, etc., that these states of affairs are peculiar. But it is in part the sense of their oddity that has led set-theorists to favor conceptions of set, such as the iterative conception, according to which what they find odd does not occur.

We describe this conception now. Our description will have three parts. The first is a rough statement of the idea. It contains such expressions as 'stage', 'is formed at', 'earlier than', 'keep on going', which must be exorcised from any formal theory of sets. From the rough description it sounds as if sets were continually being created, which is not the case. In the second part, we present an axiomatic theory which partially formalizes the idea roughly stated in the first part. For reference, let us call this theory the *stage theory*. The third part consists in a derivation from the stage theory of the axioms of a theory of sets. These axioms are formulas of $\mathcal{L}$, the language of set theory, and contain none of the metaphorical expressions which are employed in the rough statement and of which abbreviations are found in the language in which the stage theory is expressed.

Here is the idea, roughly stated:

A set is any collection that is formed at some stage of the following process: Begin with individuals (if there are any). An individual is

[7] We put a "lasso" around them, in a figure of Kripke's.

[8] x is *ungrounded* if x belongs to some set z, each of whose members has a member in common with z.

an object that is not a set; individuals do not contain members. At stage zero (we count from zero instead of one) form all possible collections of individuals. If there are no individuals, only one collection, the null set, which contains no members, is formed at this 0th stage. If there is only one individual, two sets are formed: the null set and the set containing just that one individual. If there are two individuals, four sets are formed; and in general, if there are n individuals, 2^n sets are formed. Perhaps there are infinitely many individuals. Still, we assume that one of the collections formed at stage zero is the collection of all individuals, however many of them there may be.

At stage one, form all possible collections of individuals and sets formed at stage zero. If there are any individuals, at stage one some sets are formed that contain both individuals and sets formed at stage zero. Of course some sets are formed that contain only sets formed at stage zero. At stage two, form all possible collections of individuals, sets formed at stage zero, and sets formed at stage one. At stage three, form all possible collections of individuals and sets formed at stages zero, one, and two. At stage four, form all possible collections of individuals and sets formed at stages zero, one, two, and three. Keep on going in this way, at each stage forming all possible collections of individuals and sets formed at earlier stages.

Immediately after all of stages zero, one, two, three, . . . , there is a stage; call it stage omega. At stage omega, form all possible collections of individuals formed at stages zero, one, two,. . . . One of these collections will be the set of *all* sets formed at stages zero, one, two, ...

After stage omega there is a stage omega plus one. At stage omega plus one form all possible collections of individuals and sets formed at stages zero, one, two . . . , and omega. At stage omega plus two form all possible collections of individuals and sets formed at stages zero, one, two,. . . , omega, and omega plus one. At stage omega plus three form all possible collections of individuals and sets formed at earlier stages. Keep on going in this way.

Immediately after all of stages zero, one, two,. . . , omega, omega plus one, omega plus two,. . . , there is a stage, call it stage omega plus omega (or omega times two). At stage omega plus omega form all possible collections of individuals and sets formed at earlier stages. At stage omega plus omega plus one. . . . . . .

. . . omega plus omega plus omega (or omega times three) . . .

. . . (omega times four) . . .

. . . omega times omega . . . . . .

Keep on going in this way. . . .

According to this description, sets are formed over and over again:

in fact, according to it, a set is formed at every stage later than that at which it is first formed. We could continue to say this if we liked; instead we shall say that a set is formed only once, namely, at the earliest stage at which, on our old way of speaking, it would have been said to be formed.

That is a rough statement of the iterative conception of set. According to this conception, no set belongs to itself, and hence there is no set of all sets; for every set is formed at some earliest stage, and has as members only individuals or sets formed at still earlier stages. Moreover, there are not two sets x and y, each of which belongs to the other. For if y belonged to x, y would have had to be formed at an earlier stage than the earliest stage at which x was formed, and if x belonged to y, x would have had to be formed at an earlier stage than the earliest stage at which y was formed. So x would have had to be formed at an earlier stage than the earliest stage at which it was formed, which is impossible. Similarly, there are no sets x, y, and z such that x belongs to y, y to z, and z to x. And in general, there are no sets $x_0, x_1, x_2, \ldots, x_n$ such that x_0 belongs to x_1, x_1 to $x_2, \ldots$, x_{n-1} to x_n, and x_n to x_0. Furthermore it would appear that there is no sequence of sets $x_0, x_1, x_2, x_3, \ldots$ such that x_1 belongs to x_0, x_2 belongs to x_1, x_3 belongs to x_2, and so forth. Thus, if sets are as the iterative conception has them, the anomalous situations do not arise in which sets belong to themselves or to others that in turn belong to them.

The sets of which ZF in its usual formulation speaks ("quantifies over") are not all the sets there are, if we assume that there are some individuals, but only those which are formed at some stage under the assumption that there are no individuals. These sets are called *pure* sets. All members of a pure set are pure sets, and any set, all of whose members are pure, is itself pure. It may not be obvious that any pure sets are ever formed, but the set Λ, which contains no members at all, is pure, and is formed at stage 0. $\{\Lambda\}$ and $\{\{\Lambda\}\}$ are also both pure and are formed at stages 1 and 2, respectively. There are many others. From now on, we shall use the word 'set' to mean 'pure set'.

Let us now try to state a theory, the stage theory, that precisely expresses much, but not all, of the content of the iterative conception. We shall use a language, $\mathcal{G}$, in which there are two sorts of variables: variables 'x', 'y', 'z', 'w',..., which range over sets, and variables 'r', 's', 't', which range over stages. In addition to the predicate letters 'ϵ' and '$=$' of $\mathcal{L}$, $\mathcal{G}$ also contains two new two-place predicate letters 'E', read 'is earlier than', and 'F', read 'is formed at'. The rules of formation of $\mathcal{G}$ are perfectly standard.

Here are some axioms governing the sequence of stages:

(I) $(s) \sim s\mathrm{E}s$ (No stage is earlier than itself.)
(II) $(r)(s)(t)((r\mathrm{E}s \; \& \; s\mathrm{E}t) \rightarrow r\mathrm{E}t)$ (*Earlier than* is transitive.)
(III) $(s)(t)(s\mathrm{E}t \vee s=t \vee t\mathrm{E}s)$ (*Earlier than* is connected.)
(IV) $(\exists s)(t)(t \neq s \rightarrow s\mathrm{E}t)$ (There is an earliest stage.)
(V) $(s)(\exists t)(s\mathrm{E}t \; \& \; (r)(r\mathrm{E}t \rightarrow (r\mathrm{E}s \vee r=s)))$ (*Immediately* after any stage there is another.)

Here are some axioms describing when sets and their members are formed:

(VI) $(\exists s)((\exists t)t\mathrm{E}s \; \& \; (t)(t\mathrm{E}s \rightarrow (\exists r)(t\mathrm{E}r \; \& \; r\mathrm{E}s)))$ (There is a stage, not the earliest one, which is not immediately after any one stage. In the rough description, stage omega was such a stage.)
(VII) $(x)(\exists s)(x\mathrm{F}s \; \& \; (t)(x\mathrm{F}t \rightarrow t=s))$ (E v ery set is formed at some unique stage.)
(VIII) $(x)(y)(s)(t)((y \epsilon x \; \& \; x\mathrm{F}s \; \& \; y\mathrm{F}t) \rightarrow t\mathrm{E}s)$ (Every member of a set is formed *before*, i.e., at an earlier stage than, the set.)
(IX) $(x)(s)(t)(x\mathrm{F}s \; \& \; t\mathrm{E}s \rightarrow (\exists y)(\exists r)(y \epsilon x \; \& \; y\mathrm{F}r \; \& \; (t=r \vee t\mathrm{E}r)))$ (If a set is formed at a stage, then, at or after any earlier stage, at least one of its members is formed. So it never happens that all members of a set are formed before some stage, but the set is not formed at that stage, if it has not been formed already.)

We may capture part of the content of the idea that at any stage every *possible* collection (or set) of sets formed at earlier stages is formed (if it has not yet been formed) by taking as axioms all formulas $\ulcorner(s)(\exists y)(x)(x \epsilon y \leftrightarrow (\chi \; \& \; (\exists t)(t\mathrm{E}s \; \& \; x\mathrm{F}t)))\urcorner$, where χ is a formula of the language $\mathcal{J}$ in which no occurrence of 'y' is free. Any such axiom will say that for any stage there is a set of just those sets to which χ applies that are formed before that stage. Let us call these axioms *specification axioms.*

There is still one important feature contained in our rough description that has not yet been expressed in the stage theory: the analogy between the way sets are *inductively generated* by the procedure described in the rough statement and the way the natural numbers 0, 1, 2,. . .are inductively generated from 0 by the repeated application of the successor operation. One way to characterize this feature is to assert a suitable induction principle concerning sets and stages; for, as Frege, Dedekind, Peano, and others have enabled us to see, the content of the idea that objects of a certain kind are inductively generated in a certain way is just the proposition than an appropriate induction principle holds of those objects.

The principle of mathematical induction, the induction principle governing the natural numbers, has two forms, which are interderivable on certain assumptions about the natural numbers. The first version of the principle is the statement

$$(P)[(P0 \;\&\; (n)[Pn \rightarrow PSn]) \rightarrow (n)Pn]$$

which may be read, 'If 0 has a property and if whenever a natural number has the property its successor does, then every natural number has the property'. The second version is the statement

$$(P)[(n)((m)[m < n \rightarrow Pm] \rightarrow Pn) \rightarrow (n)Pn]$$

It may be read, '*If each natural number has a property provided that all smaller natural numbers have it, then every natural number has the property*'.

The induction principle about sets and stages that we should like to assert is modeled after the second form of the principle of mathematical induction. Let us say that a stage s *is covered by* a predicate if the predicate applies to every set formed at s. Our analogue for sets and stages of the second form of mathematical induction says that *if each stage is covered by a predicate provided that all earlier stages are covered by it, then every stage is covered by the predicate.* The full force of this assertion can be expressed only with a second-order quantifier. However, we can capture some of its content by taking as axioms all formulas

$$\ulcorner (s)((t)[t\mathrm{E}s \rightarrow (x)(x\mathrm{F}t \rightarrow \theta) \rightarrow (x)(x\mathrm{F}s \rightarrow \chi)) \geq (s)(x)(x\mathrm{F}s \rightarrow \chi)\urcorner$$

where χ is a formula of $\mathcal{J}$ that contains no occurrences of 't' and θ is just like χ except for containing a free occurrence of 't' wherever χ contains a free occurrence of 's'. [Observe that '$(x)(x\mathrm{F}s \rightarrow \chi)$' says that χ applies to every set formed at stage s and, hence, that s is covered by χ.] We call these axioms *induction axioms*.

III. ZERMELO SET THEORY

We complete the description of the iterative conception of set by showing how to derive the axioms of a theory of sets from the stage theory. The axioms we derive speak only about sets and membership: they are formulas of $\mathcal{L}$.

The axiom of the null set: $(\exists y)(x) \sim x\epsilon y$. (There is a set with no members.)

Derivation. Let χ = '$x = x$'. Then

$$(s)(\exists y)(x)(x\epsilon y \leftrightarrow (x = x \;\&\; (\exists t)(t\mathrm{E}s \;\&\; x\mathrm{F}t)))$$

is a specification axiom, according to which, for any stage, there is a set of all sets formed at earlier stages. As there is an earliest stage, stage 0, before which no sets are formed, there is a set that contains no members.

Note that, by axiom (IX) of the stage theory, any set with no members is formed at stage 0; for if it were formed later, it would have to have a member (that was formed at or after stage 0).

The axiom of pairs: $(z)(w)(\exists y)(x)(x\epsilon y \leftrightarrow (x=z \vee x=w))$. (For any sets z and w, not necessarily distinct, there is a set whose sole members are z and w.)
Derivation. Let χ = '$(x=z \vee x=w)$'. Then

$$(s)(\exists y)(x)(x\epsilon y \leftrightarrow ((x=z \vee x=w) \ \&\ (\exists t)(t\mathrm{E}s \ \&\ x\mathrm{F}t)))$$

is a specification axiom, according to which, for any stage, there is a set of all sets formed at earlier stages that are identical with either z or w. Any set is formed at some stage. Let r be the stage at which z is formed; s, the stage at which w is formed. Let t be a stage later than both r and s. Then there is a set of all sets formed at stages earlier than t that are identical with z or w. So there is a set containing just z and w.

The axiom of unions: $(z)(\exists y)(x)(x\epsilon y \leftrightarrow (\exists w)(x\epsilon w \ \&\ w\epsilon z))$. (For any set z, there is a set whose members are just the members of members of z.)
Derivation. '$(s)(\exists y)(x)(x\epsilon y \leftrightarrow ((\exists w)(x\epsilon w \ \&\ w\epsilon z) \ \&\ (\exists t)(t\mathrm{E}s \ \&\ x\mathrm{F}t)))$' is a specification axiom, according to which, for any stage, there is a set of all members of members of z formed at earlier stages. Let s be the stage at which z is formed. Every member of z is formed before s, and hence every member of a member of z is also formed before s. Thus there is a set of all members of members of z.

The power-set axiom: $(z)(\exists y)(x)(x\epsilon y \leftrightarrow (w)(w\epsilon x \rightarrow w\epsilon z))$. (For any set z, there is a set whose members are just the subsets of z.)
Derivation. '$(s)(\exists y)(x)(x\epsilon y \leftrightarrow ((w)(w\epsilon x \rightarrow w\epsilon z) \ \&\ (\exists t)(t\mathrm{E}s \ \&\ x\mathrm{F}t)))$' is a specification axiom, according to which, for any stage, there is a set of all subsets of z formed at earlier stages. Let t be the stage at which z is formed and let s be the next later stage. If x is a subset of z, then x is formed before s. For otherwise, by axiom (IX), there would be a member of x that was formed at or after t and, hence, that was not a member of z. So there is a set of all subsets of z formed before s, and hence a set of all subsets of z.

The axiom of infinity:

$$(\exists y)((\exists x)(x\epsilon y \ \&\ (z) \sim z\epsilon x) \\ \&\ (x)(x\epsilon y \rightarrow (\exists z)(z\epsilon y \ \&\ (w)(w\epsilon z \leftrightarrow (w\epsilon x \vee w=x)))))$$

(Call a set null if it has no members. Call z a successor of x if the members of z are just those of x and x itself. Then there is a set which contains a null set and which contains a successor of any set it contains.)
Derivation. Let us first observe that every set x has a successor. For let y be a set containing just x and x (axiom of pairs), and let w be a set containing just x and y (axiom of pairs again), and let z contain just the members of members of w (axiom of unions). Then z is a successor of x,

for its members are just x and x's members. Next, note that if z is a successor of x, x is formed at r, and t is the next stage after r, then z is formed at t. For every member of z is formed before t. So z is formed at or before t, by axiom (IX). But x, which is in z, is formed at r. So z cannot be formed at or before r. So z cannot be formed before t. Now, by axiom (VI), there is a stage s, not the earliest one, which is not immediately after any stage. '$(s)(\exists y)(x)(x\epsilon y \leftrightarrow (x = x \ \& \ (\exists t)(t\mathrm{E}s \ \& \ x\mathrm{F}t)))$' is a specification axiom, according to which, for any stage, there is a set of all sets formed at earlier stages. So there is a set y of all sets formed before s. y thus contains all sets formed at stage 0, and hence contains a null set. And if y contains x, y contains all successors of x (and there are some), for all these are formed at stages immediately after stages before s and, hence, at stages themselves before s.

Axioms of separation (Aussonderungsaxioms): All formulas

$$\ulcorner (z)(\exists y)(x)(x\epsilon y \leftrightarrow (x\epsilon z \ \& \ \phi))\urcorner$$

where ϕ is a formula of $\mathfrak{L}$ in which no occurrence of 'y' is free.
Derivation. If ϕ is a formula of $\mathfrak{L}$ in which no occurrence of 'y' is free, then $\ulcorner (s)(\exists y)(x)(x\epsilon y \leftrightarrow ((x\epsilon z \ \& \phi) \ \& \ (\exists t)(t\mathrm{E}s \ \& \ x\mathrm{F}t)))\urcorner$ is a specification axiom, which we may read, 'For any stage s, there is a set of all sets formed at earlier stages, which belong to z and to which ϕ applies. Let s be the stage at which z is formed. All members of z are formed before s. So, for any z, there is a set of just those members of z to which ϕ applies, which we would write, $\ulcorner (z)(\exists y)(x)(x\dot{\epsilon} y \leftrightarrow (x\epsilon z \ \& \ \phi))\urcorner$. A formal derivation of an Aussonderungsaxiom would use the specification axiom described and axioms (VII) and (VIII) of the stage theory.

Axioms of regularity: All formulas $\ulcorner (\exists x)\phi \rightarrow (\exists x)(\phi \ \& \ (y)(y\epsilon x \rightarrow -\psi))\urcorner$, where ϕ does not contain 'y' at all and ψ is just like ϕ except for containing an occurrence of 'y' wherever ϕ contains a free occurrence of 'x'.
Derivation. The idea: Suppose ϕ applies to some set x'. x' is formed at some stage. That stage is therefore not covered by $\ulcorner \sim \phi\urcorner$. By an induction axiom, there is then a stage s not covered by $\ulcorner \sim \phi\urcorner$, although all stages earlier than s are covered by $\ulcorner \sim\phi\urcorner$. Since s is not covered by $\ulcorner \sim\phi\urcorner$, there is an x, formed at s, to which $\ulcorner \sim\phi\urcorner$ does not apply, i.e., to which ϕ applies. If y is in x, however, y is formed before s, and hence the stage at which is is formed is covered by $\ulcorner \sim\phi\urcorner$. So $\ulcorner \sim \phi\urcorner$ applies to y (which is what $\ulcorner \sim \psi\urcorner$ says).

For a formal derivation, contrapose, reletter, and simplify the induction axiom

$$\ulcorner (s)((t)(t\mathrm{E}s \rightarrow (x)(x\mathrm{F}t \rightarrow \sim \phi)) \rightarrow (x)(x\mathrm{F}s \rightarrow \sim \phi)) \rightarrow (s)(x)(x\mathrm{F}s \rightarrow \sim \phi)\urcorner$$

so as to obtain

$$\ulcorner (\exists s)(\exists x)(x\mathrm{F}s \ \& \ \phi) \rightarrow (\exists s)(\exists x)(x\mathrm{F}s \ \& \ \phi \ \& \ (y)(t)(t\mathrm{E}s \ \& \ y\mathrm{F}t \rightarrow \sim \psi))\urcorner$$

Assume $\ulcorner (\exists x)\phi\urcorner$. Use axiom (VII) and modus ponens to obtain

$$\ulcorner (\exists s)(\exists x)(x\mathrm{F}x \ \& \ \phi \ \& \ (y)(t)(t\mathrm{E}s \ \& \ y\mathrm{F}t \rightarrow \sim \psi))\urcorner$$

Use axioms (VII) and (VIII) to obtain $\ulcorner(\exists x)(\phi \mathbin{\&} (y)(y \epsilon x \rightarrow \sim \psi))\urcorner$ from this.

The axioms of regularity (partially) express the analogue for sets of the version of mathematical induction called the *least-number principle:* if there is a number that has a property, then there is a least number with that property. The analogue itself has been called the *principle of set-theoretical induction.*[9] Here is an application of set-theoretical induction.

Theorem: No set belongs to itself.
Proof. Suppose that some set belongs to itself, i.e., that $(\exists x)x \epsilon x$.

$$(\exists x)x \epsilon x \rightarrow (\exists x)(x \epsilon x \mathbin{\&} (y)(y \epsilon x \rightarrow \sim y \epsilon y))$$

is an axiom of regularity. By modus ponens, then, some set x belongs to itself though no member of x (not even x) belongs to itself. This is a contradiction.

The axioms whose derivations we have given are those statements which are often taken as axioms of ZF and which are deducible from all (sufficiently strong[10]) theories that can fairly be called formalizations of the iterative conception, as roughly described. (The axiom of extensionality has a special status, which we discuss below.) Other axioms than those we have given could have been taken as axioms of the stage theory. For example, we could have fairly taken as an axiom a statement asserting the existence of a stage, not immediately later than any stage, but later than some stage that is itself neither the earliest stage nor immediately later than any stage. Such an axiom would have enabled us to deduce a stronger axiom of infinity than the one whose derivation we have given, but this stronger statement is not commonly taken as an axiom of ZF. We could also have derived other statements from the stage theory, such as the statement that no set belongs to any of its members, but this statement is never taken as an axiom of ZF. We do not believe that the axioms of replacement or choice can be inferred from the iterative conception.

One of the axioms of regularity,

$$(\exists x)x \epsilon z \rightarrow (\exists x)(x \epsilon z \mathbin{\&} (y)(y \epsilon x \rightarrow \sim y \epsilon z)$$

is sometimes called *the* axiom of regularity; in the presence of other axioms of ZF, all the other axioms of regularity follow from it. The name 'Zermelo set theory' is perhaps most commonly given to the

[9] By Tarski, among others.
[10] 'Sufficiently strong' may here be taken to mean "at least as strong as the stage theory."

theory whose axioms are '$(x)(y)((z)(z\epsilon x \leftrightarrow x\epsilon y) \rightarrow x=y)$', i.e., the axiom of extensionality, and the axioms of the null set, pairs, and unions, the power-set axiom, the axiom of infinity, all the Aussonderungsaxioms, and the axiom of regularity.[11] With the exception of the axiom of extensionality, then, all the axioms of Zermelo set theory follow from the stage theory.

IV. ZERMELO-FRAENKEL SET THEORY

The axioms of replacement. ZF is the theory whose axioms are those of Zermelo set theory and all axioms of replacement.[12] A formula of $\mathfrak{L}$ is an axiom of replacement if it is the translation into $\mathfrak{L}$ of the result "substituting" a formula of $\mathfrak{L}$ for 'F' in

$$F \text{ is a function} \rightarrow (z)(\exists y)(x)(x\epsilon y \leftrightarrow (\exists w)(w\epsilon z \ \& \ F(w)=x))$$

There is an extension of the stage theory from which the axioms of replacement could have been derived. We could have taken as axioms all instances (that can be expressed in $\mathfrak{J}$) of a principle which may be put, 'If each set is correlated with at least one stage (no matter how), then for any set z there is a stage s such that for each member w of z, s is later than some stage with which w is correlated'. This *bounding* or *cofinality* principle is an attractive further thought about the interrelation of sets and stages, but it does seem to us to be a *further* thought, and not one that can be said to have been meant in the rough description of the iterative conception. For that there are exactly ω_1 stages does not seem to be excluded by anything said in the rough description; it would seem that $R\omega_1$ (see below) is a model for any statement of $\mathfrak{L}$ that can (fairly) be said to have been implied by the rough description, and not all of the axioms of replacement hold in $R\omega_1$.[13] Thus the axioms of replacement do not seem to us to follow from the iterative conception.

Adding the axioms of replacement to those of Zermelo set theory enables us to define a sequence of sets, $\{R_\alpha\}$, with which the stages of the stage theory may be identified. Suppose we put R_0 = the null set; $R_{\alpha+1} = R_\alpha \cup$ the power set of R_α, and $R_\lambda = \bigcup_{\beta<\lambda} R_\beta$ (λ a limit ordi-

[11] In his 1908 paper, "Investigations in the Foundations of Set Theory I" ["Untersuchungen über die Grundlagen der Mergenlehre," *Mathematische Annaler*, LX (1908): 261–281; reprinted in Jean van Heijenoort, ed., *From Frege to Godel* (Cambridge, Mass.: Harvard, 1967), pp. 119–215], Zermelo took as axioms versions of the axioms of extensionality, the null set, pairs (and unit set), unions, the power-set axiom, the axiom of infinity, the Aussonderungsaxioms, and the axiom of choice.

[12] Sometimes the axiom of choice is also considered one of the axioms of ZF.

[13] Worse yet, R_{δ_1} would also seem to be such a model. (δ_1 is the first nonrecursive ordinal.)

nal)—axioms of replacement ensure that the operation R is well-defined—and say that s is a stage if $(\exists\alpha)s = R_\alpha$, that x is formed at s if x is subset but not a member of s, and that s is earlier than t if, for some α, β, $s = R_\alpha$, $t = R_\beta$, and $\alpha < \beta$. Then we can prove as theorems of ZF not only the translations into the language of set theory of the axioms of the stage theory, but also those of all those stronger axioms asserting the existence of stages further and further "out" that might have been suggested by the rough description (and those of the instances of the bounding principle which are expressible in $\mathcal{G}$ as well). ZF thus enables us to describe and assert the full first-order content of the iterative conception within the language of set theory.

Although they are not derived from the iterative conception, the reason for adopting the axioms of replacement is quite simple: they have many desirable consequences and (apparently) no undesirable ones. In addition to theorems about the iterative conception, the consequences of replacement include a satisfactory if not ideal[14] theory of infinite numbers, and a highly desirable result that justifies inductive definitions on well-founded relations.

The axiom of extensionality. The axiom of extensionality enjoys a special epistemological status shared by none of the other axioms of ZF. Were someone to deny another of the axioms of ZF, we would be rather more inclined to suppose, on the basis of his denial alone, that he believed that axiom false than we would if he denied the axiom of extensionality. Although 'there are unmarried bachelors' and 'there are no bachelors' are equally preposterous things to say, if someone were to say the former, he would far more invite the suspicion that he did not mean what he said than someone who said the latter. Similarly, if someone were to say, "there are distinct sets with the same members," he would thereby justify us in thinking his usage nonstandard far more than someone who asserted the denial of some other axiom. Because of this difference, one might be tempted to call the axiom of extensionality "analytic," true by virtue of the meanings of the words contained in it, but not to consider the other axioms analytic.

It has been persuasively argued, by Quine and others, however, that until we have an acceptable explanation of how a sentence (or what it says) can be true in virtue of meanings, we should refrain from calling *anything* analytic. It seems probable, nevertheless, that whatever justification for accepting the axiom of extensionality there may be, it is more likely to resemble the justification for ac-

[14] An ideal theory would decide the continuum hypothesis, at least.

cepting most of the classical examples of analytic sentences, such as 'all bachelors are unmarried' or 'siblings have siblings' than is the justification for accepting the other axioms of set theory. That the concepts of *set* and *being a member of* obey the axiom of extensionality is a far more central feature of our use of them than is the fact that they obey any other axiom. A theory that denied, or even failed to affirm, some of the other axioms of ZF might still be called a set theory, albeit a deviant or fragmentary one. But a theory that did not affirm that the objects with which it dealt were identical if they had the same members would only by charity be called a theory of *sets* alone.

The axiom of choice. One form of the axiom of choice, sometimes called the "multiplicative axiom," is the statement, 'For any x, if x is a set of nonempty disjoint sets (two sets are disjoint if nothing is a member of both), then there is a set, called a *choice set* for x, that contains exactly one member of each of the members of x'.

It seems that, unfortunately, the iterative conception is neutral with respect to the axiom of choice. It is easy to show that, since, as is now known, neither the axiom of choice nor its negation is a theorem of ZF, neither the axiom nor its negation can be derived from the stage theory. Of course the stage theory, which is supposed to formalize the rough description, could be extended so as to decide the axiom. But it seems that no additional axiom, which would decide choice, can be inferred from the rough description without the assumption of the axiom of choice itself, or some equally uncertain principle, in the inference. The difficulty with the axiom of choice is that the decision whether to regard the rough description as implying a principle about sets and stages from which the axiom could be derived is as difficult a decision, because essentially the same decision, as the decision whether to accept the axiom.

Suppose that we tried to derive the axiom by arguing in this manner: Let x be a set of nonempty disjoint sets. x is formed at some stage s. The members of members of x are formed at earlier stages than s. Hence, at s, if not earlier, there is a set formed that contains exactly one member of each member of x. But to assert this is to beg the question. How do we know that such a choice set *is* formed? If a choice set is formed, it is indeed formed at or before s. But how do we know that one is formed at all? To argue that at s we can choose one member from each member of x and so form a choice set for x is also to beg the question: "we *can't* choose" one member from each member of x if there is no choice set for x.

To say this is not to say that the axiom of choice is not both obvious and indispensable. It is only to say that the justification for its acceptance is not to be found in the iterative conception of set.

GEORGE BOOLOS
Massachusetts Institute of Technology

A PLEA FOR SUBSTITUTIONAL QUANTIFICATION *

IN this note I shall discuss the relevance to ontology of what is called the *substitutional* interpretation of quantifiers. According to this interpretation, a sentence of the form '$(\exists x)Fx$' is true if and only if there is some closed term 't' of the language such that 'Ft' is true. This is opposed to the *objectual* interpretation according to which '$(\exists x)Fx$' is true if and only if there is some object x in the universe of discourse such that 'F' is *true of* that object.[1]

Ontology is not the only connection in which substitutional quantification has been discussed in recent years. It has been advocated as a justification for restrictions on the substitutivity of identity in intensional languages[2] or has been found necessary to

* I owe to Sidney Morgenbesser and W. V. Quine the stimulus to write this paper. I am also indebted to Hao Wang for a valuable discussion, and to Quine for illuminating comments on an earlier version.

[1] It is noteworthy that the substitutional interpretation allows truth of, say, first-order quantified sentences to be given a direct inductive definition, while in the objectual interpretation the fundamental notion is truth *of* (satisfaction), and truth is defined in terms of it. Davidson's version of the correspondence theory of truth would not be applicable to a substitutional language. See "True to the Facts," this JOURNAL, LXVI, 21 (Nov. 6, 1969): 748–764.

Quine has pointed out to me that in a language with infinitely many singular terms, a problem arises about defining truth for *atomic* formulas. Objectual quantification can be nontrivial for a language with no singular terms at all, but, if there are such, the problem can be resolved by an auxiliary definition which assigns them denotations, relative to a sequence that codes an assignment to the free variables. In the substitutional case, we need to define the truth of an atomic formula $Pt_1 \ldots t_m$ directly for closed terms $t_1 \ldots t_m$, for example, inductively by reducing the case for more complex terms to that for less complex terms. In the usual language of elementary number theory, closed terms are constructed from '0' by applications of various function symbols which (except for the successor symbol, say 'S') have associated with them defining equations that can be regarded as rules for reducing closed terms to canonical form, namely, as numerals (a numeral is 'S' applied finitely many times to '0'), so that an atomic formula $Pt_1 \ldots t_m$ is true if and only if $Pn_1 \ldots n_m$ is true, where n_i is the numeral corresponding to t_i. The truth of predicates applied to numerals is defined either trivially (as in the case of '$=$') or in a similar recursive manner.

Quine discusses this issue in §6 of an unpublished paper, "Truth and Disquotation."

[2] Ruth Barcan Marcus, "Modalities and Intensional Languages," *Synthese*, XIII,