The Journal of Symbolic Logic
Volume 7, Number 4, December 1942

A SYSTEM OF AXIOMATIC SET THEORY

PART IV. GENERAL SET THEORY.[38]

PAUL BERNAYS

11. Elementary one-to-one correspondences, fundamental theorems on power. Our task in the treatment of general set theory will be to give a survey for the purpose of characterizing the different stages and the principal theorems with respect to their axiomatic requirements from the point of view of our system of axioms. The delimitation of "general set theory" which we have in view differs from that of Fraenkel's general set theory, and also from that of "standard logic" as understood by most logicians. It is adapted rather to the tendency of von Neumann's system of set theory—the von Neumann system having been the first in which the possibility appeared of separating the assumptions which are required for the conceptual formations from those which lead to the Cantor hierarchy of powers. Thus our intention is to obtain general set theory without use of the axioms V d, V c, VI.

It will also be desirable to separate those proofs which can be made without the axiom of choice, and in doing this we shall have to use the axiom V*—i.e., the theorem of replacement taken as an axiom. From V*, as we saw in §4, we can immediately derive V a and V b as theorems, and also the theorem that a function whose domain is represented by a set is itself represented by a functional set; and on the other hand V* was found to be derivable from V a and V b in combination with the axiom of choice.[39] (These statements on deducibility are of course all on the basis of the axioms I–III.)

In the development of general set theory we shall always have V a at our disposal, in some contexts as an axiom, in others as a theorem. But, as we have seen, V a in connection with the axioms I–III enables us to obtain the fundamental theorems on ordinals, and also the iteration theorem and number theory. Hence we shall be able to dispense with the axiom VII throughout our treatment of general set theory.

For this first more elementary part of general set theory we introduce an axiom which we shall call the *pair class axiom*, and which asserts that the pair class $A \times B$ is represented by a set if the classes A and B are represented by sets. This was obtained in §4 as a consequence of V b, c, and also as a consequence of V a, c, d.[40] Later we shall show that it is also derivable from IV, V a, V b, and hence, on the basis of the axiom of choice, can be dispensed with as an axiom for general set theory.[41]

Received April 25, 1941.

[38] Parts I, II, III appeared in this Journal, vol. 2 (1937), pp. 65–77, vol. 6 (1941) pp. 1–17, and vol. 7 (1942), pp. 65–89. Part V, continuing the treatment of general set theory, will appear in a later number of this Journal.

[39] Part II, pp. 2–3, consequence 3, consequence 4 and remark.

[40] Part II, pp. 4–5, consequence 8.

[41] Here and in similar cases, the basic axioms I–III are presupposed as a means of deduction without being expressly mentioned.

From the pair class axiom we are able to deduce by means of V a, b—hence also by means of V*—the theorem, stated in §4 as a consequence of V c,[42] that the sum of the classes A and B is represented by a set if A and B are represented by sets. In fact we may first reduce this theorem to the case that A and B have no common element. For if B_1 is the class of elements of B which do not belong to A, then the sum of A and B is the same as that of A and B_1; the classes A and B_1 have no common element, and if A and B are represented by sets, so likewise are A and B_1 (by V a). Moreover the assertion of the theorem is obvious if one of the classes A, B has no element or has only one element. Thus we may assume that A and B have no common element and that there are elements p, q of A and elements r, s of B such that $p \neq q$ and $r \neq s$. But then the class of triplets $\langle\langle a, b\rangle, c\rangle$ such that either $a \eta A$ & $b = r$ & $c = a$, or $a = p$ & $b \eta B$ & $b \neq r$ & $c = b$, or $a = q$ & $b = s$ & $c = r$ is a one-to-one correspondence between a subclass of $A \times B$ and the sum of A and B. By the pair class axiom the class $A \times B$ is represented by a set, and hence by V a and V b the sum of A and B is represented by a set.

We introduce at once some notations and definitions.

By the *set-sum* of the sets a and b we mean the set representing the sum of the classes represented by a and b, or in other words the set whose elements are those sets which are in a or in b. (Its existence, by the proof just given, follows from V a, b and the pair class axiom.) We shall denote the sum of A and B by $A + B$, and likewise the set-sum of a and b by $a + b$.

By the *difference* of the classes A, B we mean the intersection of A with the complementary class of B; and by the *set-difference* of the sets a, b we mean the set representing the difference of the classes represented by a, b (the existence of such a set follows from V a). We shall denote the difference of A and B by $A \div B$, and the set-difference of a and b by $a \div b$.

The *pair set* of a and b, which we shall denote by $a \times b$, is the set representing the pair class of the classes represented by a and b, or in other words the set of pairs $\langle c, d\rangle$ such that $c \epsilon a$ and $d \epsilon b$. (Its existence follows from the pair class axiom.)

The *class of mappings of the set c into the class A*, which we denote by $A^{[c]}$, is the class of functional sets which have c as domain and whose values belong to A.

Going on now to our survey, we first observe that there is an introductory discipline of class theory, namely the Boolean algebra (in the elementary sense) dealing with sum and intersection of classes, and the complementary class of a class, considered with respect to equalities and the subclass relation. In order to obtain this Boolean algebra, our axioms I (2), II (1), III a (1), (2), (3) suffice.[42a]

We have also a number of formal laws, of similar character, which concern the pair class $A \times B$ and the class of mappings $A^{[c]}$.

---

[42] Part II, p. 4, consequence 7.

[42a] This kind of a derivation of Boolean algebra of course is not an independent foundation, but the fundamental operations and relations of Boolean algebra are reduced, by means of the logical concepts, to the relation of a thing (set) belonging to a class. (Added September 28, 1942.)

The four laws concerning the pair class which assert that, for any classes A, B, C,

$$A \times B \sim B \times A, \qquad A \times (B \times C) \sim (A \times B) \times C,$$

$$A \times (B + C) = (A \times B) + (A \times C), \quad A \times (B \div C) = (A \times B) \div (A \times C),$$

are deducible by means of the axioms I–III.[43] (Notice that the first two laws assert merely a one-to-one correspondence, but the other two, an identity.)

The formal laws which concern $A^{[c]}$ are less elementary, since they require the axiom V* and the pair class axiom, or else axioms (such as IV, V a, V b) from which these two axioms can be derived as theorems. These laws for $A^{[c]}$ are as follows (all of them assert one-to-one correspondences):

For any classes A, B and any set c,

$$(A \times B)^{[c]} \sim A^{[c]} \times B^{[c]}.$$

For any class A, and for any sets b, c having no common element,

$$A^{[b+c]} \sim A^{[b]} \times A^{[c]}.$$

For any class A and any sets b, c,

$$A^{[b \times c]} \sim (A^{[b]})^{[c]}.$$

For any set a there is a one-to-one correspondence between the class of subsets of a and the class of mappings of a into the class represented by 2.

Since the translation into our system of the usual method of proving these laws concerning $A^{[c]}$ is rather direct, it will suffice to carry out, as an example, the proof of the first of them. To do this we must, given classes A, B and a set c, exhibit a one-to-one correspondence between the functional sets which assign to each element of c an element of $A \times B$, and the pairs of functional sets $\langle f, g \rangle$ such that f assigns to each element of c an element of A, and g assigns to each element of c an element of B. Such a one-to-one correspondence is the class of triplets $\langle h, \langle f, g \rangle\rangle$ such that $f \eta A^{[c]}$, and $g \eta B^{[c]}$, and h represents the class of triplets $\langle r, \langle s, t \rangle\rangle$ for which $\langle r, s \rangle \epsilon f$ and $\langle r, t \rangle \epsilon g$. In fact this class of triplets $\langle r, \langle s, t \rangle\rangle$ is, for any element f of $A^{[c]}$ and any element g of $B^{[c]}$, a function whose domain is represented by c and which therefore, by axiom V*, is itself represented by a set.

The three remaining laws are to be proved in an analogous direct way.

To the use of these laws on one-to-one correspondences—which are generalizations of laws for finite classes—there is adjoined the method—peculiar to the treatment of infinite classes—of obtaining one-to-one correspondences by means of the *Bernstein theorem* (or *Schröder-Bernstein theorem*), the statement of which is as follows:

If A, B, C are classes, and $A \subset B$, and $B \subset C$, and $A \sim C$, then $B \sim C$.

---

[43] See in particular Part I, p. 76, assertions 8 and 9.

Felix Bernstein's proof of this theorem[44] can be carried out within our system in the following way. By hypothesis there exists a one-to-one correspondence F between C and A. By the class theorem there exists the class S of those elements e of C for which there is a finite ordinal n and a functional set f such that n' is the domain of f, and $f(0)\ \eta\ C \div B$, and $f(n) = e$, and for every element k of n, $\langle f(k), f(k')\rangle \eta F$.[44a] Let H be the class of those elements $\langle e, a\rangle$ of F for which $e \eta S$. Then it is easily shown that H is a one-to-one correspondence between S and $S \div (C \div B)$. Therefore, since $B = (S \div (C \div B)) + (C \div S)$, and $C = S + (C \div S)$, the sum of H and the class of pairs $\langle b, b\rangle$ such that $b\ \eta\ C \div S$ is a one-to-one correspondence between C and B.

For this proof, as can be seen, there are required besides the axioms I–III only the theory of finite ordinals. Thus the Bernstein theorem can be proved on the basis of the axioms I–III and V a (or also with VII replacing V a).

Remark. By an application of the class theorem analogous to that at the beginning of the preceding proof we infer the existence, for any class A, of the class of all those sets b for which there exists a finite ordinal n and a functional set f such that n' is the domain of f, and $f(0)\ \eta\ A$, and $f(n) = b$, and for every element k of n, $f(k') \epsilon f(k)$. This class we shall call the *transitive closure* of A. It is easily seen to be a transitive class, and a subclass of any transitive class of which A is a subclass. Moreover by the *transitive closure of a set a* we understand the transitive closure of the class represented by a.[45] It follows further by the class theorem that the class of pairs $\langle a, b\rangle$ exists such that b belongs to the transitive closure of a.

The Bernstein theorem, as is known, has its chief importance in connection with the *foundations of the theory of powers* (*Mächtigkeiten*).

The first steps toward the comparison of powers can be made merely on the basis of the axioms I–III, V a. In Cantor's terminology, the classes A and B are said to be *of equal power* if $A \sim B$; and A is said to be *of lower power* than B, and B *of higher power* than A, if A is not of equal power with B but is of equal power with some subclass of B. We also use the expressions "of equal power," "of lower (higher) power" for sets, in the sense that the relation named holds for the classes represented by the sets; also similarly for a class and a set. That the class A is of lower power than the class B will be symbolized by $A \prec B$; and similarly the notations $a \prec b$, $a \prec B$, $A \prec b$ will be used.

[44] Concerning the proofs which have been given of this theorem see E. Borel, ***Leçons sur la théorie des fonctions,*** first edition, Paris 1898, Note I, pp. 102–107, and A. Korselt, *Über einen Beweis des Äquivalenzsatzes*, ***Mathematische Annalen,*** vol. 70 (1911), pp. 294–296. In the proofs of Korselt, Zermelo, and Peano, the concept of finite number is eliminated by the Dedekind method of operating with intersections. However, this method of proof is in some respects less elementary, and this has the effect that it is applicable in our system only to the case that the class C (in our above formulation of the Bernstein theorem) is represented by a set.

[44a] The class S can be defined in other words as the class of those elements of C which, for at least one element a of $C \div B$, belong to the converse domain of the iterator of F on a—see Part II, §6, p. 12. (Added September 28, 1942.)

[45] The concept of the transitive closure of a set amounts to the same thing as Finsler's concept of the system *der in einer Menge wesentlichen Mengen*, though the latter is defined in a somewhat different way. P. Finsler, *Über die Grundlegung der Mengenlehre*, ***Mathematische Zeitschrift,*** vol. 25 (1926), see §7, pp. 693–694.

From the definition of the relation $\prec$ the formal laws characterizing it as an ordering relation can be deduced in the usual way. That $\overline{A \prec A}$ for every class A follows immediately; that, for arbitrary classes A, B, C,

$$A \prec B \;\&\; A \backsim C \rightarrow C \prec B,$$

$$A \prec B \;\&\; B \backsim C \rightarrow A \prec C$$

follows by means of the composition lemma; and that, for arbitrary classes A, B, C

$$A \prec B \;\&\; B \prec C \rightarrow A \prec C$$

is to be proved by means of the Bernstein theorem and the composition lemma.

Among the general theorems on comparison of power which are provable by means of the axioms I–III and V a there belongs also the famous *Cantor theorem* concerning the subsets of a set, which can be formulated as follows: The class of the elements of a set a is of lower power than the class of the subsets of a.

The proof proceeds in the well-known way. First there is a one-to-one correspondence between the class A of the elements of a and the class of those non-empty subsets of a which have only one element. On the other hand, if there were a one-to-one correspondence C between A and the class of the subsets of a, there would exist the class S of those elements b of A not in $C(b)$. By V a the class S would be represented by a subset s of a. This set s would be assigned by C to an element r of a. And, by the definition of S, the contradiction would result that $r \epsilon s$ if and only if $\overline{r \epsilon s}$.

In this proof the use of V a is essential. This appears from the impossibility of proving, by the method of the preceding reasoning (the Cantor "diagonal procedure"), the assertion that every class is of lower power than the class of its subsets. And in fact this assertion can be immediately seen to be false, since the class of all sets is identical with the class of its subsets.

However, the Cantor argument applies not only to subsets of a set but also to subclasses of a class; and although in our system we have neither classes of classes nor functions assigning classes to sets, nevertheless we can carry out in it the application of the Cantor diagonal procedure to the subclasses of a class. This possibility arises from the following circumstance. To an assignment (in the usual sense) of classes to sets which are the elements of a class A, there corresponds the relation between a set a belonging to A and a set b belonging to the class assigned to a. If this relation can be formulated by means of a constitutive expression, then the class C exists of pairs $\langle a, b\rangle$ such that the relation in question holds between a and b; and the class assigned to an element a of A is the class of sets b for which $\langle a, b\rangle \eta C$.

We shall say that a class B *is assigned to a set a by means of the class of pairs C* if B is the class of sets b such that $\langle a, b\rangle \eta C$.

Now we can prove that, for any class A, (1) there is a class of pairs by means of which to each element of A a subclass of A is assigned in a one-to-one way, and (2) there is no class of pairs by means of which every subclass of A is assigned to an element of A, hence *a fortiori* no class of pairs by means of which such an assignment is made in a one-to-one way. In fact, (1) by means of the class of pairs $\langle c, c\rangle$ such that $c\eta A$ there is assigned to each element a of A the class whose only element is a, and this of course is a one-to-one assignment; and (2) if there were a class of pairs C by means of which every subclass of A were assigned to an element of A, then the class S of elements d of A such that $\langle d, d\rangle$ did not belong to C would be assigned by means of C to an element r of A; but then we should have the contradiction that $\langle r, r\rangle\eta C$ if and only if $\overline{\langle r, r\rangle\eta C}$.

In this way the Cantor theorem can be expressed and proved within the axiomatic frame under consideration, not only for the subsets of a set but also for the subclasses of a class—however without the consequence arising that for every class there is another class of higher power, which of course would lead to a contradiction. We are not even able in general set theory, as we have defined it, to infer from the forms of the Cantor theorem proved that for every set there is a set of higher power. This last follows only by application of the axiom V d (a fact which motivates the designation of this axiom as "power axiom").

However we can infer from the Cantor theorem that for every class represented by a set there exists a class of higher power, or, what amounts to the same thing, that the class of all sets is of higher power than any class which is represented by a set.

This result can also be obtained in another way, by use of the axiom V b. Since every class is a subclass of the class of all sets, every class is either of lower power or of equal power to the class of all sets. Thus if there were a class C represented by a set and not of lower power than the class of all sets, there would exist a one-to-one correspondence between C and the class of all sets; then by V b the class of all sets would be represented by a set, and therefore by V a every class would be represented by a set. But we have seen that the class of all ordinals is not represented by a set.

From these considerations concerning the Cantor theorem and the class of all sets it becomes clear in particular how the set-theoretic paradoxes are avoided in our system by the distinction of classes and sets.

12. Numeration and well-ordering, cardinal numbers. A principal point in the theory of power, as is well known, is the question of *comparability* and, in connection with it, the relation between powers and ordinals.

A class A is called *comparable* with the class B if the disjunction holds that

$$A \backsim B \vee A \prec B \vee B \prec A.$$

The comparability of sets is defined in the same way.

By the theorems on finiteness, any finite classes A, B are comparable (and likewise any finite sets). In fact it follows easily from the theorems on finiteness that if A and B are finite classes we have $A \backsim B$ or $A \prec B$ or $B \prec A$ according as the number attributable to A is the same as or lower than or higher than the number attributable to B. We also can infer that every finite class is of lower power than any infinite class.

For proof of the comparability of arbitrary classes A, B rather strong axiomatic assumptions seem to be required. In general set theory we shall prove that any two sets are comparable. Afterwards for the proof that any two classes are comparable we shall have to add the axioms V c, d, and VII (or VII*).

In both these proofs the axiom of choice has an essential rôle. This is not surprising. In fact it is quite natural to use the axiom of choice for the theory of powers, as may be seen from the following considerations.

Cantor's concept of power refers to one-to-one correspondence. But a comparison of powers could just as well be based on the general concept of a function. Indeed we could first introduce a concept, *A is of at least as high a power as B*—the notation could be, say, "$A \gtrsim B$"—defining it to mean that either B is empty or there exists a function with the domain A and the converse domain B. Then we could define the classes A, B to be *of equal power* if at the same time $A \gtrsim B$ and $B \gtrsim A$; A to be *of higher power than* B if $A \gtrsim B$ but not $B \gtrsim A$; and A to be *of lower power than* B if B is of higher power than A. This method of defining equal power, higher power, and lower power is intuitively as well motivated as the usual definition due to Cantor. (For finite classes it comes to the same thing.) And as long as the equivalence of the two definitions is not established, we have two competing concepts of power. Now this equivalence can be expressed by the statement that, for arbitrary classes A and B, $A \gtrsim B$ if and only if $A \sim B \vee B \prec A$ (note that we are using the symbols $\sim$, $\prec$ in their originally defined signification); and the implication

$$A \sim B \vee B \prec A \rightarrow A \gtrsim B$$

follows directly by axioms I–III. So what is in question is the general validity of the implication

$$A \gtrsim B \rightarrow A \sim B \vee B \prec A.$$

This, however, follows rather directly from the theorem that every function has an inverse function, which we found to be equivalent to the axiom of choice on the basis of the axioms I–III.[46]

Remark. Apparently the theorem that, if A and B are any classes,

$$A \gtrsim B \rightarrow A \sim B \vee B \prec A$$

cannot be deduced from the axioms I–III, V–VII (i.e., from our complete set with exception of the axiom of choice). But a proof of impossibility has not been given. The same thing is to be said of the theorem stating the general validity of the implication

$$A \gtrsim B \;\&\; B \gtrsim A \rightarrow A \sim B,$$

which of course on the assumption of the axiom of choice is an immediate consequence of the Bernstein theorem.

The inference from $A \gtrsim B$ to $A \sim B \vee B \prec A$ occurs in particular in Zermelo's

[46] See Part II, §4, p. 1, consequence 2 and remark.

proof of the *generalized Julius König theorem*,[47] which can be formulated as follows: If f and g are functional sets with the same domain d and if, for every element a of d, $f(a) \prec g(a)$, then the sum S of the elements of the converse domain of f is of lower power than the class P of functional sets with the domain d which assign to each element a of d an element of $g(a)$.

Let us briefly consider the proof. Two things are to be shown, first that there exists a one-to-one correspondence between S and a subclass of P, secondly that there is no one-to-one correspondence between S and P.

The first of these results in the following way. By our assumptions on f, g, the theorem of replacement, and the axiom of choice, there exist functional sets g_1, h, t with the domain d such that, for every element a of d, $g_1(a)$ is a proper ubset of $g(a)$, $h(a)$ is a one-to-one correspondence between $f(a)$ and $g_1(a)$, and $t(a) \epsilon g(a) \div g_1(a)$. Let Q be the class of pairs $\langle s, p \rangle$ such that $s \eta S$, $p \eta P$, and, for every element a of d, $p(a) = (h(a))(s)$ if $s \epsilon f(a)$, and $p(a) = t(a)$ otherwise. This class is obviously a one-to-one correspondence between S and a subclass of P.

The second part of the proof consists in showing that, for every one-to-one correspondence C between S and a subclass of P, there exists an element of P not belonging to the converse domain of C. For this purpose we notice that, for each element a of d, there exists the class of pairs $\langle s, b \rangle$ of an element s of $f(a)$ and an element b of $g(a)$ such that $(C(s))(a) = b$. This class (which depends on a) is a function; its domain M is represented by $f(a)$ and its converse domain N is a subclass of $g(a)$. We have $M \gtrsim N$ and therefore, by the above-mentioned consequence of the axiom of choice, $M \backsim N \vee N \prec M$. Since $f(a) \prec g(a)$, it follows (by the formal properties of the relation "of lower power") that N cannot represent $g(a)$. Thus for every element a of d there exists an element of $g(a)$ which, for every element s of $f(a)$, is different from $(C(s))(a)$. Moreover, by the class theorem, the axiom of choice, and the theorem of replacement, there exists a functional set q with the domain d, satisfying the conditions that, for every element a of d, $q(a) \epsilon g(a)$ and $(x)(x \epsilon f(a) \rightarrow q(a) \neq (C(x))(a))$. And q belongs to P, but, as follows immediately, not to the converse domain of C.

Thus we prove the generalized König theorem within general set theory.

Going on now to prove, within general set theory, by means of the axioms I–IV, V a, b, that any two sets are comparable, we follow the analogy of the case of finite sets. The comparability of finite sets results from the existence, for any finite set, of a one-to-one correspondence to an ordinal. Thus we have to prove for an arbitrary set the existence of a one-to-one correspondence between this set and an ordinal. As follows from the axiom of replacement, such a one-to-one correspondence must be represented by a functional set. We shall call a functional set representing a one-to-one correspondence between a

[47] This theorem was proved by Zermelo in 1904 as a generalization of a theorem presented by Julius König at the Heidelberg Congress of 1904. Cf. J. König, *Zum Kontinuum-Problem*, ***Mathematische Annalen***, vol. 60 (1905), pp. 177–180, and E. Zermelo, *Untersuchungen über die Grundlagen der Mengenlehre*, ***Mathematische Annalen***, vol. 65 (1908), see theorem 33 and footnote, pp. 277–279. In the following statement and its proof, as will be seen, a restricting premiss of both the König and the Zermelo theorem, which was adapted to the theory of powers, is eliminated.

set and an ordinal a *numeration* of the set. Thus the theorem to be proved can be formulated as *the numeration theorem*: Every set has a numeration.

For the proof of this we use the method of Zermelo's first proof of the well-ordering theorem.[48]

This method yields at the same time a more general theorem whose demonstration is independent of the axiom of choice. For its formulation we first make the following definition. A numeration h of a set c will be called *adapted to* F, where F is a function assigning to every proper subset p of c an element of $c \div p$, if for every ordinal n in the domain of h the value $h(n)$ is identical with the value assigned by F to the proper subset of c which represents the class of those elements of c assigned by h to the ordinals lower than n.

Then we state the following *theorem of adapted numeration*: For any set c and any function F which assigns to every proper subset p of c an element of $c \div p$, there exists a numeration of c that is adapted to F.

Proof. It is easily seen that the condition that a set be a numeration of a subset of c, adapted to F, can be formulated by a constitutive expression. Therefore the class L exists of those numerations of subsets of c which are adapted to F. Every element of L is a functional set whose domain is an ordinal and which represents a one-to-one correspondence. If h and k are elements of L, and m is the domain of h and n the domain of k, and m is a subset of n, then h is a subset of k. For otherwise there would be a lowest ordinal l among those ordinals which are first member of an ordered pair that is in h but not in k, and, since h and k are numerations adapted to F, we should have $h(l) = k(l)$, in contradiction to the characterizing property of l. Consequently, of any two elements of L, one is a subset of the other, and so the sum S of the elements of L is a one-to-one correspondence The domain of S is a transitive class of ordinals, and the converse domain of S is the class of those elements of c which are in the converse domain of a numeration of a subset of c adapted to F. By V a, since the converse domain of S is a subclass of c, it is represented by a set; and by V b, since the converse class of S is (like S) a one-to-one correspondence, the domain of S and S itself are each represented by a set. The set m representing the domain of S is a transitive set of ordinals and so is itself an ordinal. Therefore the set s representing S is a numeration of a subset d of c, and it is easily shown that this numeration is adapted to F. Moreover d must be identical with c. For otherwise the set t whose elements are the elements of s together with the pair $\langle m, F(d)\rangle$ would be a numeration of a subset of c, adapted to F. And then t would be a subset of s; but this is immediately seen to be impossible. Therefore s is a numeration of c that is adapted to F.—For this proof, as will be seen, the axioms I–III, V a, b have sufficed.

Now the numeration theorem is easily proved as follows. Let c be a set. By the class theorem the class of pairs $\langle a, b\rangle$ exists such that $a \subset c$ and $b \epsilon c \div a$. Therefore by the axiom of choice there exists a function F assigning to every proper subset p of c an element of $c \div p$. And consequently by the theorem of adapted numeration, there exists a numeration of c that is adapted to F.

[48] Cf. E. Zermelo, *Beweis, dass jede Menge wohlgeordnet werden kann*, ***Mathematische Annalen***, vol. 59 (1904), pp. 514–516.

From the numeration theorem and the theorem of adapted numeration several rather direct consequences are to be drawn.

First we have the result in connection with the theory of powers that *any two sets are comparable*. In fact, since of any two ordinals one is a subset of the other, it follows by the numeration theorem that of any two sets one is of equal power with a subset of the other.

The situation is made more explicit by introduction of the concept of a *cardinal*. By a *cardinal* (or a *cardinal number*) we understand an ordinal for which there exists no lower ordinal of equal power. For any set there exists a unique cardinal to which it has a one-to-one correspondence. For since the class of numerations of a set c is not empty, there is a lowest among the ordinals which occur as domains of numerations belonging to that class—or in other words there is a lowest of the ordinals m such that $m \backsim c$. This lowest ordinal is obviously a cardinal, and there cannot be another cardinal of equal power with c.

The uniquely determined cardinal that is of equal power with a set c we shall call *the cardinal number of* c, and also *the cardinal number of* C if C is the class represented by c.

Clearly a set a is of equal power with a set b, of lower power than b, or of higher power than b according as the cardinal number of a is the same as, lower than, or higher than that of b. And the corresponding thing holds for classes which are represented by sets.

By the theorem (γ) of §6,[49] every finite ordinal is a cardinal, and the cardinal number of a finite set or class is the number attributable to it.

Another direct consequence of the numeration theorem is the well-ordering theorem.

In our system the easiest way of introducing the concept of *order* is that of defining ordered classes and ordered sets as classes of pairs. As is well known, this can be done as follows.

Definition. A class M of pairs is called *an order of a class* C if:

1. $\langle a, b\rangle \eta M \rightarrow a\eta C \,\&\, b\eta C$.
2. $a\eta C \,\&\, b\eta C \rightarrow \langle a, b\rangle \eta M \vee \langle b, a\rangle \eta M$.
3. $\langle a, b\rangle \eta M \,\&\, \langle b, a\rangle \eta M \rightarrow a = b$.
4. $\langle a, b\rangle \eta M \,\&\, \langle b, c\rangle \eta M \rightarrow \langle a, c\rangle \eta M$.

We then also say that the class C is *ordered by the class* M, and we also speak of *the class* C *in the order* M. *An order of a set* c is to be defined in the same way; or the definition can be given by saying that an order of a set c is the same as an order of the class represented by c.

There can be at most one class C which is ordered by a given class of pairs M. For if there is such a class C, it must be the same as the class of sets a such that $\langle a, a\rangle \eta M$.

If a and b are two distinct elements of a class or a set which is ordered by a class M, then, by conditions 2 and 3 in the definition of order, one and only one of the pairs $\langle a, b\rangle$, $\langle b, a\rangle$ belongs to M. We shall say that a *precedes* b *in the order* M if $a \neq b$ and $\langle a, b\rangle \eta M$.

[49] See Part II, §6, p. 16.

Remark. It might appear more natural to define an ordering class in such a way that no pair $\langle a, a\rangle$ belongs to it; but this would have the disadvantage that a class or a set with only one element could not be regarded as ordered, or else that the ordered set (a) could not be distinguished from the ordered set (b) or from the ordered null set.

An order M of a class C (or of a set c) is called a *well-ordering* if every non-empty subclass of C (or every non-empty subset of c) has a *first element* in the order M—i.e., an element which precedes every other element in the order M.

A well-ordering of a set c can be obtained immediately from a numeration of c. Indeed every numeration h of a set c determines an order M of c by the condition that, a and b being elements of c, the pair $\langle a, b\rangle$ belongs to M if and only if the ordinal to which a is assigned by h is not higher than the ordinal to which b is assigned. Now this ordering of c—let us call it *the order associated with* h—is a well-ordering. For if s is a non-empty subset of c, the class of those ordinals to which an element of s is assigned by h has a lowest element; and the value of h for this lowest element is obviously the first element of s in the order in question. Thus from the numeration theorem we can infer that *every set has a well-ordering.* This is the *well-ordering theorem.*

On the other hand from the theorem of adapted numeration it follows that every well-ordering of a set c is associated with a numeration of c. For let c be well-ordered by M and (making use of the class theorem) let F be the function whose domain is the class of proper subsets of c and whose value for a proper subset p of c is the first element of $c \div p$ in the order M. Then by the theorem of adapted numeration there exists a numeration of c that is adapted to F; and it is readily seen that M is the well-ordering associated with this numeration.

Moreover it is easily shown that a well-ordering of a set is associated with only one numeration of the set.

The foregoing applies in particular to the *natural order* of a set of ordinals. By the *natural order of a class, or a set, of ordinals* we mean the class of pairs $\langle a, b\rangle$ such that a and b are elements of the class, or set, and a is not higher than b. This order is obviously a well-ordering. Thus of any set of ordinals there is one and only one numeration with which the natural order is associated; we shall call it *the numeration in the natural order.*

We insert here some discussion of the proof of the well-ordering theorem. A natural question is whether we could not follow more closely Zermelo's method of proving this theorem, applying the treatment we have made of numerations directly to well-orderings instead. This indeed would be possible, but would not be advantageous for our purposes, since in any case we want to have the numeration theorem, and the passage from it to the well-ordering theorem is more immediate than the inverse passage. Moreover for the treatment of well-orderings, as we have defined them, we should need the pair class axiom in order to show that any well-ordering of a set is represented by a set—whereas our intention is to derive the pair class axiom by means of the numeration theorem.

This last inconvenience can, however, be avoided by using another way of introducing the notion of order, due to Sierpiński.[50] Thus we come to a proof of the well-ordering theorem which is intermediate between the first proof of Zermelo and his second[51] one. Let us briefly indicate this method.

We start from the following definition. An *ordering class for a set c* is a class A of subsets of c having the three following properties:

1. Of any two elements of A, one is a subset of the other.

2. In every element s of A there is one and only one element which is not in any proper subset of s belonging to A. This element will be called the *terminal* element of s (with respect to A).

3. For every element b of c there is an element s of A such that b is in s but in no proper subset of s belonging to A. (In view of 2, this amounts to postulating that every element b of c is the terminal element of some element of A.)

From this definition it follows immediately that there is a one-to-one correspondence between any set c and any ordering class for it. Hence it follows by the axiom V b that every ordering class for a set is represented by a set. We shall call a set representing an ordering class for a set c an *ordering set for c*.

By the *order generated by an ordering set t* for c we understand the class of pairs $\langle a, b\rangle$ such that $a \epsilon c$ and $b \epsilon c$ and a is in every element of t in which b is. This class of pairs is obviously an order. And it is easily seen that the order generated by an ordering set t for c is a well-ordering of c if and only if every non-empty subset s of t has an element which is a subset of every element of s. If this condition is satisfied by an ordering set t for c we call it a *well-ordering set for c*. (In particular, under this definition, the null set is a well-ordering set for itself.)

Moreover a well-ordering set r for a subset of c will be called *adapted to F*, where F is a function which assigns to every proper subset p of c an element of $c \div p$, if, for every element b of r, the terminal element of b is the value of F for the set of the remaining elements of b.

Now let c be any set. We are to prove that there exists a well-ordering set for c. As before, we conclude that, in virtue of the axiom of choice, there exists a function assigning to every proper subset p of c an element of $c \div p$. Let F be such a function.

First we show that, if h and k are well-ordering sets for subsets of c and are both adapted to F, and if k is not a subset of h, then there is an element d of k such that h is the set of those elements of k which are proper subsets of d. Let D be the class of those elements of k which have a subset that is in one of the sets h, k but not in both (notice that the subset in question is not required to be

[50] Cf. W. Sierpiński, *Une remarque sur la notion de l'ordre*, ***Fundamenta mathematicae,*** vol. 2 (1921), pp. 199–200. The Sierpiński concept of order is a modification of that introduced by Kuratowski in his paper, *Sur la notion de l'ordre dans la théorie des ensembles*, ***Fundamenta mathematicae,*** vol. 2 (1921), pp. 161–171. The idea of representing any ordering of a set by a class of subsets such that any two distinct subsets a and b belonging to the class satisfy the condition $a \subset b \vee b \subset a$ goes back to Hessenberg. Cf. G. Hessenberg, *Grundbegriffe der Mengenlehre*, ***Abhandlungen der Fries'schen Schule*** (Göttingen), n. s. vol. 1 (1906), in particular pp. 674–685 ("Vollständig ordnende Systeme").

[51] E. Zermelo, *Neuer Beweis für die Möglichkeit einer Wohlordnung*, ***Mathematische Annalen,*** vol. 65 (1908), see §1, pp. 107–111.

a proper subset). Then D is not empty, since k has an element not in h. And since k is a well-ordering set, the subset of k which represents D has an element d which is a subset of every element of D. Let t be the set of sets b such that $b \epsilon k$ and $b \subset d$, and let s be the set representing the sum of the elements of t (which sum is a subclass of c). Then $d = s + (F(s))$. An element b of t cannot belong to D, and consequently every subset of b (including b itself) is either in both of the sets h, k or in neither. From this it follows that t is a subset of h, and that every subset of an element of t which is in h is also in t. And this entails $t = h$; for otherwise, since h is a well-ordering set adapted to F, there would be an element of h which was on the one hand not in t and on the other hand a subset of every element of $h \div t$, and this element of h would be identical with $s + (F(s))$; so we should have $d \epsilon h$, $d \epsilon k$, and at the same time d would be a subset of every element of D and of every element of $h \div t$, and this leads to a contradiction with $d \eta D$. Thus h is the set of those elements of k which are proper subsets of d.

Then we consider the class H of those well-ordering sets for subsets of c which are adapted to F (the existence of H following from the class theorem). From the foregoing it follows that the sum of the elements of H is represented by some one of its elements, say g. But then g must be a well-ordering set for c itself—since otherwise the set $g + (F(g))$ would be an element of H and thus a subset of g, which is impossible. Therefore there exists a well-ordering set for c. And the order generated by this well-ordering set is a well-ordering of c.

In this way we have also proved at the same time that, for any function F which assigns to every proper subset p of a set c an element of $c \div p$, there exists a well-ordering set for c that is adapted to F. And in the proof of this the axiom of choice has not been used.

In regard to this demonstration of the well-ordering theorem, it will be observed that the possibility of imitating within our system of general set theory (with use merely of the axioms I–IV, V a, b) Zermelo's second proof of the well-ordering theorem is due to an essential modification of that proof by which it becomes more elementary. In fact, in order to translate the second proof of Zermelo directly into our system, we should have to apply the axiom V d. Likewise this axiom would be required for translating into our system Hartogs's proof (without use of the axiom of choice) that there exists no set which is of higher power than every well-ordered set.[52]

ZURICH

[52] Cf. F. Hartogs, *Über das Problem der Wohlordnung*, ***Mathematische Annalen***, vol. 76 (1915), pp. 438–443.