

# Notes on Zermelo's axioms for set theory

Randall Holmes

September 22, 2020

I'm going to organize these point by point using the section numbers in Zermelo's paper, with some intermissions.

1. He introduces a domain  $\mathbf{B}$  of all objects. He says that some of these objects are *sets* (primitive notion). He introduces equality as a primitive notion.
2. In section 2, he introduces membership  $a \in b$  as a primitive relation. He tells us that if  $a \in b$ ,  $b$  must be a set (and telegraphs that there can be a set with no elements, but there can be only one: this follows from a later axiom). In any event what he says here is enough to see that if there is any object which is not a set, it has no elements.

It is a feature of most modern treatments of foundations of mathematics based on set theory that all objects are sets. This isn't a natural assumption philosophically, though philosophers have gamely played along with it, and it is nice to see that in this source text this assumption is not made.

3. In section 3, he defines the subset relation:  $M \subseteq N$  iff  $M$  and  $N$  are sets (important) and for all  $x$ , if  $x$  is an element of  $M$  then  $x$  is an element of  $N$ . Notice that for any  $x$  which is not a set and any  $y$ , it is true that any  $z$  which belongs to  $x$  (there aren't any) belongs to  $y$ , but we do not say  $x \subseteq y$ . He also defines disjointness of sets.
4. In section 4, he introduces the idea of an assertion  $\phi$  being "definite", meaning that it can be decided whether it is true or false. This is important to him, and more in a philosophical than a mathematical way. A propositional function  $P(x)$  of an object  $x$  ranging over a

class  $K$  is definite if  $P(x)$  is definite for each element of the class  $K$ . Question: what does he mean by “class”? Is  $\mathbf{B}$  a class in this sense? He does know that  $\mathbf{B}$  is not a set (see section 10). He tells us that  $a \in b$  is always definite, and that  $M \subseteq N$  is always definite (because the propositional function  $x \in N$  of  $x$  is definite for each  $x \in M$ , I interpret him as saying).

He now introduces some axioms:

**Axiom I:** If  $M \subseteq N$  and  $N \subseteq M$  then  $M = N$ . Sets with the same elements are the same. This is called the axiom of extensionality.

**informal definition:** He introduces the notation  $\{a, b, c, \dots, r\}$  for a finite set whose elements are exactly  $a, b, c, \dots, r$ . A motto of mine: whenever a mathematician introduces those rows of dots, he is cheating.

**Axiom II:** There is a set  $0$  with no elements (why does he feel constrained to call it “fictitious”?); for any object  $a$  the set  $\{a\}$  which has  $a$  as an element and no other element exists; for any two objects  $a, b$ , the set  $\{a, b\}$  which has  $a$  as an element, has  $b$  as an element and has no other elements exists. This is called the axiom of elementary sets.

5. In section 5 he observes that the sets  $0$ ,  $\{a\}$  and  $\{a, b\}$  are uniquely determined by axiom I. He also notes that  $a = b$  is always “definite” because it is equivalent to  $a \in \{b\}$ . I was going to say in protest that it is definite because it is equivalent to  $a \subseteq b \wedge b \subseteq a$ , but that is actually inadequate: that only shows that equality of sets is definite. The question of why Zermelo wants to talk about definiteness of assertions is really interesting. Does he ever say anything that isn’t definite?

The notation  $\emptyset$  for the empty set is more usual now, but we will stick with  $0$ .

6. In section 6 he raises a really interesting issue in philosophy of set theory as it were in passing. He notes that  $0 \subseteq M$  and  $M \subseteq M$  are always true: he defines a “part” of a set  $M$  as a subset of  $M$  other than  $0$  or  $M$ . We might think of an object as a part of itself, and prefer to say “proper part” here. We can feel with Zermelo discomfort with the idea that all sets have a common part  $0$ . But in any event, the

philosophical point to make is that the elements of a set are not parts of the set (whatever a part of a set may be taken to be). The relation of part to whole, however it is defined, must be transitive, and  $a \in \{a, \}$ ,  $\{a, b\} \in \{\{a, b\}\}$  but  $a \notin \{\{a, b\}\}$ . A set does not have its elements as its parts but its subsets; it does have its singleton subsets as atomic parts correlated with its elements. David Lewis wrote a whole book about this, *Parts of Classes*, which I quite recommend as reading.

Now (still in section 6) we get

**Axiom III:** Whenever the propositional function  $P(x)$  is definite for each element of a set  $M$ , there is  $M_P$  such that  $M_P \subseteq M$  and for every  $x$ ,  $x \in M_P$  if and only if  $x \in M$  and  $P(x)$ . This is called the Axiom of Separation.

We introduce the notation  $\{x \in M : P(x)\}$  for the set  $M_P$ .

This axiom gives the kind of ability to define objects correlated with properties which Frege wanted in his Axiom V governing “courses of values”. But the restriction of this formation of objects from properties to properties of elements of a previously given set seems to make this workable without contradiction. Moreover, nothing is being given up in terms of actual mathematical practice: we do not construct sets of mathematical interests by considering properties of all objects taken indiscriminately, but by considering properties of objects of a particular sort.

Zermelo talks about the importance of ensuring that the property  $P(x)$  is “definite”. I am quite interested in what he thinks this notion is doing for him.

7. In section 7, Zermelo notes that for  $M_1 \subseteq M$  (and in fact for any  $M_1$ ) we can define  $M \setminus M_1$  as  $\{x \in M : x \notin M_1\}$ , Zermelo calls this the complement of  $M_1$ ; it would more usually be called the complement of  $M_1$  relative to  $M$ . This is provided by Axiom III.
8. In section 8, Zermelo introduces the intersection  $[M, N]$  of sets  $M$  and  $N$ , provided by axiom III as  $\{x \in M : x \in N\}$ . This is more usually written  $M \cap N$  now. The ability to write the assertion that  $M$  and  $N$  are disjoint as  $[M, N] = 0$  is worth noting.

9. In section 9 he extends intersections first to intersections  $[M, N, R, \dots]$  of a list of given sets and then to the notion  $\bigcap T$  of the intersection of all elements of a nonempty set  $T$ . His analysis of this is interesting. He says that by Axiom III for a given set  $T$  and each  $a$  we can define a subset  $T_a = \{t \in T : a \in t\}$  of all elements of  $T$  which contain  $a$ . It is then a definite question for each  $a$  whether  $T_a = T$  (that is, whether every element of  $T$  contains  $a$ ) and so by Axiom III there is a set  $\bigcap T = \{a \in t : T_a = T\}$ , where  $t$  is any element of  $T$  (it does not matter which one you use). A more usual definition would be  $\{a \in t : (\forall u \in T : a \in u)\}$ : but Zermelo in his formulation avoids using a quantifier.
10. In section 10 we see the form which the Russell paradox argument takes in this theory. It does not lead to paradox.

**Theorem:** For each set  $M$ , there is a set  $M_0$  such that  $M_0 \subseteq M$  and  $M_0 \notin M$ .

**Proof:** It is definite for each  $x \in M$  whether  $x \in x$  (and Zermelo takes pains to note that nothing in his system prevents  $x \in x$  from being true for some  $x$ ).

Let  $M_0 = \{x \in M : x \notin x\}$ . Clearly  $M_0 \subseteq M$ .

Now either  $M_0 \in M_0$  or not. If  $M_0 \in M_0$  then  $M_0$  contains an  $x$  such that  $x \in x$ , which is incompatible with its definition. So  $M_0 \notin M_0$ . But then  $M_0 \in M$  is impossible, as if we had  $M_0 \in M$  it would meet the conditions to belong to  $M_0$ , and we have already shown that it cannot meet these conditions. So  $M_0 \notin M$  as desired.

Zermelo then observes that it follows from this that we cannot have a set which contains every object, and so (interestingly) he observes that the domain **B** cannot be a set. It is philosophically very interesting that he makes this observation: one could further ask what sort of thing **B** is... and this is a natural opening to the later idea of proper classes.

Still in section 10, Zermelo introduces two more axioms.

**Axiom IV:** To every set  $A$  there corresponds a set  $\mathcal{P}(A)$ , the power set of  $A$ , whose elements are exactly the subsets of  $A$ . (Axiom of power set)

**Axiom V:** To every set  $A$ , there corresponds a set  $\bigcup A$ , the union of  $A$ , whose elements are exactly the elements of the elements of  $A$ .  
(Axiom of union)

**side comments of ours about these axioms:** It is very interesting that the union  $\bigcup A$  has to be provided by axiom where the related set  $\bigcap A$  is provided already by Axiom III.

An interesting point about the power set, which is related to the intellectual origins of Russell's paradox and the theorem of section 10, is the theorem of Cantor that the power set of  $A$ , even for  $A$  infinite, is larger in size than the set  $A$ .

We give an informal argument for this. We say that two sets  $A$  and  $B$  are the same size if there is a bijection from  $A$  to  $B$ , that is, a function  $F$  from  $A$  to  $B$  which is one to one and onto  $B$ . The set  $\mathcal{P}(A)$  is at least as large as  $A$  (each element  $a$  of  $A$  corresponds to  $\{a\} \in \mathcal{P}(A)$ ). Suppose that  $A$  is at least as large as  $\mathcal{P}(A)$ , that is, there is a one-to-one map  $f$  from  $\mathcal{P}(A)$  to a subset of  $A$ . This map has an inverse  $f^{-1}$  defined for some but not all elements of  $A$ . For the sake of argument we extend  $f^{-1}(a)$  to be 0 for  $a$  in  $A$  which is not in the range of  $f$ . Now define  $R$  as  $\{a \in A : a \notin f^{-1}(a)\}$ . We ask whether  $f(R)$ , a well-defined element of  $A$  by hypothesis, is an element of  $R$ : this will be true if and only if  $f(R) \in A$  and  $f(R) \notin f^{-1}(f(R))$ . Now  $f(R) \in A$  is true and  $f^{-1}(f(R)) = R$ , so we have  $f(R)$  an element of  $R$  if and only if  $f(R)$  is not an element of  $R$ , which is a contradiction: there can be no such  $f$ .

This is not a proof in Zermelo's system, because we do not yet know how to talk about functions like  $f$  in Zermelo's system, but shortly we will.

11. In section 11 we define binary unions of sets. We define  $M + N$  (usually now written  $M \cup N$ ) as  $\bigcup\{M, N\}$ . More generally we define  $M + N + R + \dots$  as  $\bigcup\{M, N, R, \dots\}$ . Some algebraic laws such as  $M + 0 = M + M = M$  are noted.
12. In section 12 it is noted that the commutative and associative laws  $M + N = M + N$  and  $(M + N) + R = M + (N + R)$  hold for union of sets. Further, distributive laws  $[M + N, R] = [M, R] + [N, R]$  and  $[M, N] + R = [M + R, N + R]$  hold. Zermelo notes that these theorems are proved by Axiom I and logic.

13. If  $M$  is a set different from  $0$  and  $a \in M$ , it is definite whether  $M = \{a\}$ . It is definite whether a set has one element or not. This is Zermelo's language: here is my own argument: for any set  $M$ . it is definite whether there is an element  $a$  of  $M$  such that every element  $x$  of  $M$  is equal to  $a$ . If there is such an element  $a$  of  $M$ , then  $M = \{a\}$ , and if there is not,  $M$  is not a singleton.

We say that a set  $T$  is a pairwise disjoint collection of sets or a partition if for each pair  $A, B$  of distinct elements of  $T$  we have that  $A$  and  $B$  are disjoint.

Let  $T$  be a pairwise disjoint collection of sets. We define  $\prod T$  as the collection of all subsets  $X$  of  $\bigcup T$  such that for each  $A \in T$ ,  $A \cap X$  has exactly one element. A product  $\prod\{M, N\}$  is written  $MN$  and a product  $\prod\{M, N, R, \dots\}$  is written  $MNR\dots$

Before stating the next axiom, it is worth noting the relationship between the addition and multiplication operations now defined on sets and addition and multiplication of numbers. If  $M$  has  $m$  elements and  $N$  has  $n$  elements and  $M, N$  are disjoint, then  $M + N = M \cup N$  has  $m + n$  elements and  $MN$  has  $mn$  elements. Similar statements are true for sums and products of finite collections of disjoint sets.

We would like it to be the case that a product of infinitely many nonempty sets has to be nonempty, and that is what the next axiom says. The final axiom provides us with something no previous axiom has done, an example of an infinite set.

**Axiom VI:** If  $T$  is a pairwise disjoint collection of sets and  $0 \notin T$ , then  $\prod T$  is nonempty. In other words, for any pairwise disjoint collection  $P$  of nonempty sets, we can find a set which contains exactly one element of  $P$ : we can choose one element from each set and collect them. This is called the Axiom of Choice.

**Axiom VII:** There is a set  $Z$  such that  $0 \in Z$  and for each element  $a$  of  $Z$ ,  $\{a\}$  is also an element of  $Z$ . (axiom of infinity).

Observe that Axiom II already gives us infinitely many distinct objects,  $0, \{0\}, \{\{0\}\}, \dots$ , but nothing before axiom VII allows us to construct an infinite set.

**what if everything is a finite set?** I actually pause to investigate this. I give an alternative collection of axioms:

- A:** same as Axiom I: sets with the same elements are the same.
- B:** The empty set  $0$  exists. For each object  $a$ ,  $\{a\}$  exists.
- C:** for each set  $A$  and object  $b$ ,  $A + \{b\} = A \cup \{b\}$  exists.
- D:** for each propositional function  $P(x)$  such that  $P(0)$ , for each object  $a$ ,  $P(a) \rightarrow P(\{a\})$ , and for each set  $A$  and object  $b$ ,  $P(A) \wedge P(b) \rightarrow P(A \cup \{b\})$  is true, we have that  $P(x)$  is true for all  $x$ .

Axioms A,B,C should be recognizable as consequences of Axioms I, II, and V. Axiom D should remind one of the principle of mathematical induction.

Though we do not have a formal definition of the notion of a finite set, you should notice that if the propositional function  $P(x)$  is “ $x$  is a finite set”, the conditions of axiom D apply, so we should expect that everything is actually a finite set in the system described by axioms A – D.

Now the punchline is that axioms I-VI (but not axiom VII) are all consequences of axioms A-D.

I is easy: it is the same as axiom A.

II: axiom B says that  $0$  and  $\{a\}$  exist; axioms B and C give us  $\{a\} + \{b\} = \{a, b\}$ .

III: Let  $Q(x)$  be a propositional function. We show using axiom D that the predicate  $P(x)$  asserting that  $\{y \in x : Q(y)\}$  exists holds of every  $x$ .

$P(0)$  is true because  $\{y \in 0 : Q(y)\} = 0$ .

$P(\{a\})$  is true (whether or not  $P(a)$  is true) because  $\{y \in \{a\} : Q(y)\}$  is either  $0$  or  $\{a\}$  depending on whether  $Q(a)$  is true, and both of these sets exist.

Suppose  $P(A)$  is true. Then whether or not  $P(b)$  is true  $\{y \in A + \{b\} : Q(y)\}$  exists because it is either  $\{y \in A : Q(y)\}$  (if  $\neg Q(b)$ ) which exists by hypothesis, or it is  $\{y \in A : Q(y)\} \cup \{b\}$ , if  $Q(b)$  is true, which exists by hypothesis and axiom C.

V:  $a \cup 0 = a$  exists.  $a \cup \{b\}$  exists by axiom C.  $a \cup (b \cup \{x\}) = (a \cup b) \cup \{x\}$  exists by axiom C if  $a \cup b$  exists. So by axiom D  $a \cup x$  exists for any set  $x$ . We use this to prove the existence of general unions of sets:  $\bigcup 0 = 0$ .  $\bigcup \{a\} = a$ .  $\bigcup (A \cup \{b\}) = (\bigcup A) \cup b$ , so by Axiom D  $\bigcup x$  exists for every  $x$ .

IV: We show the existence of the product of any set  $a$  and a singleton  $\{x\}$ :  $0\{x\} = 0$ ;  $\{a\}\{x\} = \{a, x\}$  which exists. If  $A\{a\}$  exists then  $(A \cup \{b\})\{x\} = A\{x\} \cup \{b, x\}$ .

$\mathcal{P}(0) = \{0\}$  which exists.  $\mathcal{P}(\{a\}) = \{\{a\}\} \cup \{0\}$ , which exists.  $\mathcal{P}(A \cup \{x\})$  is the union of  $\mathcal{P}(A)$  and the product of  $\{x\}$  and  $\mathcal{P}(A)$ , which exists if  $\mathcal{P}(A)$  exists.

VI: We want to prove that if  $P$  is a partition,  $P$  has a choice set. If  $P = 0$ ,  $0$  is a choice set for  $P$ . If  $P = \{a\}$ ,  $a$  is a nonempty set with an element  $x$  and  $\{x\}$  is a choice set for  $\{a\}$ . If  $P = A + \{b\}$  is a partition and  $A$  has a choice set  $C$ , then  $b$  is a nonempty set with an element  $x$  and  $C \cup \{x\}$  is a choice set for  $A + \{b\}$ .

So in fact all the axioms I-VI hold if we allow only the construction of finite sets by listing.

We do not want to restrict ourselves in this way, since we do want to consider things like the collection of natural numbers and the collection of real numbers, and their general subsets.

14. In section 14 we define the actual set of natural numbers, which we might call  $\mathbb{N}$  and which Zermelo calls  $Z_0$ .

We define an *inductive set* as a set which contains  $0$  and for each  $a$  contains  $\{a\}$  if it contains  $a$ . So Axiom VII simply asserts the existence of an inductive set.

That a set is inductive is a definite property.

There is not necessarily a set of all inductive sets, but by Axiom III there is a set  $I_Z$  of all inductive elements of  $\mathcal{P}(Z)$ . Note that  $Z \in I_Z$ , so  $I_Z$  is nonempty. Define  $Z_0$ , or  $\mathbb{N}$ , as  $\bigcap I_Z$ . It should then be observed that for any inductive set  $Z'$ , whether it is a subset of  $Z$  or not,  $Z \cap Z'$  is inductive and belongs to  $I_Z$ , and so  $Z_0 \subseteq Z' \cap Z \subseteq Z$ , from which we see that  $Z_0$  is the intersection of *all* inductive sets.

We informally define a function from natural numbers (whatever they may be) to their representatives in Zermelo's theory: define  $\#0$  as the set  $0$  and recursively define  $\#(n+1)$  as  $\{\#n\}$ . If being an image under  $\#$  is a definite predicate, it is clearly inductive. Further, by ordinary mathematical induction, it has no inductive subsets, so it actually is  $Z_0$ . The final step is to suggest that the answer to what the natural



numbers are might be that they *are* the Zermelo natural numbers (the elements of  $Z_0$ ).