

8. THE AXIOM OF CHOICE

In this section we will discuss an axiom of set theory, the Axiom of Choice.

Every human language has grammar and vocabulary, and people communicate by arranging the objects of the language in patterns. We imagine that our communications evoke similar, or at least related, mental states in others. We also use these patterns to elicit mental states in our “future selves,” as reminder of past imaginings so that we can start at a higher level in an ongoing project and not have to recreate each concept from scratch should we return to a task. It is apparent that our brains are built to do this.

But words are all defined in terms of each other. Ultimate meaning, if there is any to be found, is derived from pointing out the window at instances in the world, or from introspection. Very often ambiguity or multiple meaning of a phrase is the point of a given communication, and provides the richness and subtlety characteristic of poetry, for instance, or the beguiling power of political speech.

Set theory is a language mathematicians have invented to encode mathematics. But unlike most human languages, this language does everything possible to avoid blended meaning, to expose the logical structure of statements and keep the vocabulary of undefined terms to an absolute minimum. Many mathematicians believe what they do is “art.” But ambiguity and internal discord is not part of our particular esthetic ensemble.

Most mathematicians believe that, though set theory may be unfinished, it serves its purpose well. Virtually all mathematical structures can be successfully modeled in set theory, to the extent that most mathematicians never think of any other way of speaking or writing.

Together, the collection of axioms (which, along with logical conventions defines the language) normally used by most mathematicians is called the **Zermelo-Fraenkel Axioms**, or simply **ZF** and the set theory that arises from these axioms is called **Zermelo-Fraenkel Set Theory**. You saw explicit mention of two axioms from ZF, the Axiom of Infinity and the Axiom of the Empty Set, in Section 5. We have used others without mention on almost every page. For example we have formed power sets.

The Axiom of the Power Set For any set A there is a set $\mathbb{P}(A)$ consisting of all, and only, the subsets of A .

Asserting the existence of a set with this feature is a dramatic and “non-constructive” thing to do, *particularly* when the underlying set is infinite. We are not told how to create this set. We just have a means of recognizing if a set we have in hand is a member of this power set, or not.

And where, exactly, did that first infinite set come from? The Axiom of Infinity brings it into existence, out of nothing, simply because mathematician *want infinite sets* and this seems to be a logically consistent way to produce them.

There is another extremely useful—and arguably even less constructive—axiom which we discuss now.

We will present and presume to be true, wherever convenient, the four equivalent and useful statements below, one of which is called the Axiom of Choice. This axiom is frequently abbreviated to **AC**. The collection of the axioms of standard set theory plus this axiom is frequently denoted **ZFC**.

The discussions regarding equivalence of the Axiom of Choice and the other three statements, and the history associated with them, is a fascinating story which deserves study by every serious student of mathematics.

The Axiom of Choice: If J and X are sets and $A: J \rightarrow \mathbb{P}(X)$ is an indexed collection of nonempty sets then there is a function $f: J \rightarrow X$ such that $f(\beta) \in A_\beta \forall \beta \in J$. A function with this property is called a **choice function** for A .

Essentially, this axiom states that given any generic set \mathbb{S} of nonempty sets, there is a way of selecting one element from each member of \mathbb{S} . The other axioms do not imply that such a selection can be made, unless every member of \mathbb{S} has an element with some unique property, which would allow it to be singled out.

Zorn’s Lemma: If S is a set with a partial order and if every chain in S possesses an upper bound in S , then S has a maximal member.

Zermelo’s Theorem: Every nonempty set can be well ordered.

Kuratowski’s Lemma: Each chain in a partially ordered set S is contained in a **maximal chain** in S (that is, a chain in S not contained in any other chain in S .)

Kuratowski’s Lemma is also often called **The Hausdorff Maximal Principle**.

That Zorn’s Lemma implies Kuratowski’s Lemma is immediate. Suppose S is a set with a partial order and C is a chain in S . Let \mathbb{W} denote the set of all chains in S which contain the chain C , ordered by containment. Any chain in \mathbb{W} is

bounded above by the union of the chain, so Zorn's Lemma implies that \mathbb{W} contains a maximal member. That maximal member is a chain in S not properly contained in any other chain in S .

On the other hand, assuming Kuratowski's Lemma to be true, suppose S is a set with a partial order and that every chain in S possesses an upper bound in S . This time let \mathbb{W} denote the set of *all* chains in S . Let X denote a maximal member of \mathbb{W} . So X is a chain in S not contained in any other chain. Let M be any upper bound for X . By maximality of X , M must actually be in X and cannot be less than any other member of S : that is, M is maximal in S . So Zorn's Lemma is true.

In the last two paragraphs we have shown that Zorn's Lemma and Kuratowski's Lemma are equivalent statements.

We will now show that Zorn's Lemma implies AC. Suppose \mathbb{S} is any nonempty set of nonempty sets and X is the union of all the sets in \mathbb{S} . Let $B = \mathbb{S} \times X$. Now let Q denote the set of all subsets of $\mathbb{P}(B)$ which are choice functions on their domains: that is, $T \in Q$ exactly when T is nonempty and there is at most one ordered pair in T whose first component is any particular member of \mathbb{S} , and also $s \in A$ whenever $(A, s) \in T$. These are called "**partial choice functions.**" Order Q by containment. The union of any chain in Q is a member of Q so Zorn's Lemma implies that Q has a maximal member. This maximal member is a choice function on its domain, which must by maximality be all of \mathbb{S} .

The fact that Zermelo's Theorem implies AC is also straightforward: given any nonempty set \mathbb{S} of nonempty sets, well order the set $X = \bigcup_{S \in \mathbb{S}} S$. For each $S \in \mathbb{S}$ let $f(S)$ be the least element of S with respect to this ordering. f is the requisite choice function.

The opposite implication is a bit trickier. It involves using a choice function to create the well order.

Suppose set A has more than one element and $f: \mathbb{P}(A) - \{\emptyset\} \rightarrow A$ is a choice function: that is, $f(B) \in B$ whenever $\emptyset \neq B \subset A$.

Let \mathbb{B} denote the set of all nonempty containment-chains in $\mathbb{P}(A) - \{\emptyset\}$ which are well ordered and satisfy the condition:

Whenever I_K is an initial segment of one of these chains and if J is the union of all the sets in I_K then $J \neq A$ and $K = J \cup \{f(A - J)\}$.

\mathbb{B} is nonempty: for example, $\{\{f(A)\}, \{f(A), f(A - \{f(A)\})\}\}$ is in \mathbb{B} .

The condition above implies that each of the chains in \mathbb{B} must start with the set $\{f(A)\}$, and the successor to any set K in such a chain (if, of course, K is not the last set in the chain) has exactly one more member than does K . It also implies directly that if two different chains X and W of this kind have a common initial segment, so that $I_K \subset X$ and $I_G \subset W$ and $I_K = I_G$ then $K = G$. In other words, the least successor of an initial segment is determined by the sets in the initial segment, and *not* by the specific chain within which the initial segment sits.

Suppose that X is one of these chains. We will call S a "starting chunk" of X if $\emptyset \neq S \subset X$ and whenever $B, C \in X$ the condition $B \in S$ and $C \subset B$ implies $C \in S$. Now it might be that a starting chunk is as short as $\{f(A)\}$ or it could, possibly, be all of X . But if it is *not* all of X then because X is well ordered there

is a least member K of X not in S and so S contains all members of X less than K . That is, $S = I_K$ for some $K \in X$. So starting chunks are either initial segments or the entire chain.

Now suppose X and W are unequal chains, members of \mathbb{B} . Then one, say X , would contain a least set K not in the other. The initial segment I_K of X is contained in W . If there were a set in W not in I_K but less than some member of I_K then there would be a least member of W of this kind. Call that least member G . But then the initial segment I_G of W would be a starting chunk of X and by the above remark we would have $G \in X$, contrary to its definition.

So there are no missing members of W between members of I_K , which is therefore a starting chunk of W . Since $K \notin W$ we must have $I_K = W$, and conclude that W is an initial segment of X .

To reiterate: for each pair of members of \mathbb{B} , one is an initial segment of the other.

Now let S be the union of all the members of \mathbb{B} . Each set in S comes from a member of \mathbb{B} and since one of any pair of members of \mathbb{B} is an initial segment of the other we conclude that S itself is a chain, and well ordered too.

Let J denote the union of all the sets in S . If $J \neq A$ then we could extend S to $S \cup \{J \cup \{f(A - J)\}\}$ which satisfies the conditions for membership in \mathbb{B} but is strictly longer than its longest member, a contradiction. We conclude that $J = A$.

So we can use S to create an order on A . If a and b are members of A there is a least member S_a of S containing a and a least member S_b containing b . Declare $a \leq b$ precisely when $S_a \subset S_b$. If J is the union of the sets in the initial segment determined by S_a then $a \notin J$ so it must be that $a = f(A - J)$. So this relation makes A into a total order. Further, if $\emptyset \neq T \subset A$ then the collection of all of the S_t with $t \in T$ has a least member, which produces a least member of T . So the order on A is a well order.

We conclude that the existence of a choice function on $\mathbb{P}(A) - \{\emptyset\}$ implies that A can be well ordered. So AC implies Zermelo's Theorem.

Upon accepting the Axiom of Choice, as we will do throughout this book, well ordered sets are plentiful and can be used.

At this point we have shown the following implications among the conditions which we claim to be equivalent to the Axiom of Choice.

$$\begin{array}{ccc} \text{Zorn} & \iff & \text{Kuratowski} \\ \downarrow & & \\ \text{AC} & \iff & \text{Zermelo} \end{array}$$

The **Principles of Induction** and **Recursive Definition** are incredibly powerful and useful techniques, extending the idea of Induction on the Integers to many more well ordered sets and situations more varied than merely checking if an indexed set of propositions are all true. The methods are detailed in Section 14. It is important to note, and the reader should check, that the proof of the version of Recursive Definition we use here does *not* require AC.

We will now use Induction and Recursive Definition to show that Zermelo's Theorem implies Kuratowski's Lemma, thereby proving that any of the four conditions

listed above implies the others. The discussion below is a typical usage of this type of argument. It uses first a recursive definition to deduce that a certain function exists, and then induction to confirm various properties of that function.

We suppose we have a chain in a partially ordered set. We will line up the members of the set not already in the chain and test them one at a time. When it is an element's turn, if it can be added to yield a bigger chain than we have up to that point we select it. Otherwise we discard it. Then we go on to the next element and repeat until we have exhausted the possibilities. The product is a maximal chain. A rigorous justification can be produced after digesting the result in 14.3

Assume Zermelo's Theorem to be true, and suppose H is a nonempty chain in set K with partial order \preceq . Suppose $B = K - H$ is nonempty. There is a well order \leq for B . Since we have two orders in hand, we will use prefixes to describe which order is in use. We will let I_β stand for a \leq -initial segment for any $\beta \in B$.

Suppose y is a fixed element of H . For the \leq -first element α of B , let $P(\alpha)$ equal α if $H \cup \{\alpha\}$ is a \preceq -chain, and let $P(\alpha)$ be y otherwise.

Having found $P(\beta)$ for all $\beta \in B$ with $\alpha \leq \beta < \gamma$ for some $\gamma \in B$ define $P(\gamma)$ to be γ if $H \cup \{\gamma\} \cup P(I_\gamma)$ is a \preceq -chain, and let $P(\gamma)$ be y otherwise.

This serves to define $P(\gamma)$ for each $\gamma \in B$.

$H \cup P(B)$ must be a \preceq -chain: if not it must contain two \preceq -incomparable members s and t , which cannot both be in H . If one of the two, say s , is in H then there is a $\beta \in B$ with $P(\beta) = t = \beta$. But then $H \cup \{\beta\} \cup P(I_\beta)$ is not a chain, violating the defining condition for $P(\beta)$. A similar contradiction occurs if neither s nor t are in H , by examining the point at which the *second* of the two points would have been added. So in fact $H \cup P(B)$ must be a \preceq -chain.

No additional members of K can be added to $H \cup P(B)$ without causing the resulting set to fail to be a \preceq -chain: once again, letting γ be the \leq -least member of B which *could* be added, if any, yields a contradiction. That member *would* have been added at stage γ .

So $A \cup P(B)$ is a maximal \preceq -chain in K , and Kuratowski's Theorem holds.

8.1. Exercise. *Fill in the details of a direct proof using Induction and Recursive Definition that Zermelo's Theorem implies Zorn's Lemma. We assume that K is a set with partial order \preceq for which every chain has an upper bound. We assume also that K has a well order \leq with \leq -first member α .*

We would like to conclude that K has a \preceq -maximal element.

Let α denote the \leq -first member of K and define $G(\alpha) = \alpha$. Having defined G on \leq -initial segment I_β for $\beta > \alpha$ let $G(\beta) = \beta$ if β is a \preceq -upper bound for $G(I_\beta)$, and otherwise let $G(\beta) = \alpha$.

Show that $G(K)$ is a chain and that $G(K)$ has a \preceq -last member, which is \preceq -maximal in K .

8.2. Exercise. *(i) An axiom equivalent to our Axiom of Choice is produced if we add to that axiom the condition that $A_\alpha \cap A_\beta = \emptyset$ whenever $\alpha \neq \beta$.*

(ii) Consider the statement: “Whenever \mathbb{B} is a nonempty set of nonempty pairwise disjoint sets, there is a set S for which $S \cap x$ contains a single element for each $x \in \mathbb{B}$.” Show that this statement is equivalent to the Axiom of Choice.

(iii) Let \mathbb{B} be a (nonempty) set of sets. \mathbb{B} is said to have **finite character** provided that $A \in \mathbb{B}$ if and only if every finite subset of A is in \mathbb{B} . **Tukey’s Lemma** states that every set of sets of finite character has a maximal member: a set not contained in any other member. Show that Tukey’s Lemma is equivalent to the Axiom of Choice. (hint: To prove that Tukey’s Lemma implies the Axiom of Choice examine the set of partial choice functions and note that it satisfies the conditions of Tukey’s Lemma.)

The use of AC in the formation of mathematical arguments has historically been the subject of controversy centered around the nebulous nature of the objects whose existence is being asserted in each case. In applications the axioms of set theory are usually used to affirm the existence of one precisely defined set whose elements share an explicit property. That is less obviously the case when AC is invoked.

Applications which require less than the full strength of AC are common. In an effort to control, or at least record, how the axiom is being used, weaker variants have been created. Some mathematicians award “style points” to proofs using one of these, or which avoid AC altogether. We list two of these weaker versions of AC below.

The Axiom of Dependent Choice: If X is a nonempty set and $R \subset X \times X$ is a binary relation with domain all of X , then there is a sequence $r: \mathbb{N} \rightarrow X$ for which $(r_k, r_{k+1}) \in R \forall k \in \mathbb{N}$.

The Axiom of Countable Choice: If X is a nonempty set and $r: \mathbb{N} \rightarrow \mathbb{P}(X)$ is a sequence of nonempty subsets of X then there is a sequence $f: \mathbb{N} \rightarrow X$ such that $f(n) \in A_n \forall n \in \mathbb{N}$.

These axioms are frequently abbreviated to **DC** and **AC $_{\omega}$** , respectively.

8.3. Exercise. (i) Prove the implications $AC \Rightarrow DC \Rightarrow AC_{\omega}$.

(ii) Suppose X is infinite. For each $k \in \mathbb{N}$ let S_k denote the set of all subsets of X which have 2^k elements. ZF alone implies that S_k is nonempty for each k , and you may assume this. Let S denote the set of all the S_k . Use AC_{ω} twice to prove that there is a one-to-one function $f: Y \rightarrow \mathbb{N}$ for an infinite subset Y of X . Any set Y (infinite or not) for which there is a function of this kind is called **countable**, and the result here may be paraphrased as “Any infinite set has an infinite countable subset in ZF+ AC_{ω} .”

(iii) Sometimes the use of an axiom, particularly a variant of the Axiom of Choice, is hard to spot in an argument. It seems so reasonable, it is hard to see you are assuming anything. The theorem that “The union of a countable set of countable sets is countable.” is an example.

Suppose A is a countable set of countable sets, and let B denote the union of all the members of A . Because A is countable, there exists one-to-one $T: A \rightarrow \mathbb{N}$. Because each member of A is countable, for each nonempty set $S \in A$ there is a nonempty set F_S consisting of all one-to-one functions from S to \mathbb{N} . Using T , this collection of sets of functions is seen to be countable, so AC_{ω} guarantees that we

can pick a function from each. It is easy to overlook this step, and merely assert “Because each member of A is countable there exists one-to-one $G_S: S \rightarrow \mathbb{N}$ for each $S \in A$.” and get on with the discussion using these selected functions. But it is AC_ω which endorses this selection.

To finish the argument, for each $x \in B$ we let $A_x = \{S \in A \mid x \in S\}$ and define i_x to be the least integer in $T(A_x)$. We define W_x to be that member of A_x with $T(W_x) = i_x$. The function $H: B \rightarrow \mathbb{N}$ given by

$$H(x) = 2^{i_x} \cdot 3^{G_{W_x}(x)}$$

is one-to-one, so B is countable.

8.4. Exercise. (i) Any chain in a tree is contained in a branch.

(ii) Prove **König’s Tree Lemma**: If S is an infinite rooted tree but each $t \in S$ is the immediate predecessor of only finitely many members of S then S has an infinite branch. (hint: Let K denote those members of S with an infinite number of successors and for each $t \in K$ let $M_t = T_t \cap K - \{t\}$. Let f denote a choice function for these sets: $f(t) \in M_t \forall t \in K$. Use induction on \mathbb{N} to create an infinite chain.)

The next section contains another important consequence of the Axiom of Choice. Many more can be found scattered in appendices and chapters throughout this book.

Those who want a slightly more detailed look at the ZF axioms can find them listed in Sections 11 and 13. The discussions there are rudimentary but, I hope, a practical guide providing a taste of modern set theory.