THE AXIOM OF CHOICE

1. WHAT IS THE AXIOM OF CHOICE?

Why was this the axiom that broke mathematics? Here is its modern formulation (very close to the original one).

Axiom 1. *Let $(A_i)_{i\in I}$ be a collection of nonempty sets. There is a* choice function $f : \{A_i : i \in I\} \to \bigcup_{i\in I} A_i$ *such that for every $i \in I$, $f(A_i)$ is an element of A_i.*

This is a more formal way of saying there is a function that chooses exactly one element from each set A_i. The nontriviality of the axiom comes from the distinction between *successive* choices and *simultaenous* choices. Imagine, if you will, that the index set I is finite so that we only have finitely many sets $A_1, \cdots, A_n$. It is easy to define a choice function *successively*: first pick an element of A_1, then of $A_2, \ldots,$ then of A_n. This is a construction of a choice function by induction, and is a well-defined procedure to get a function.

Now, suppose we have an infinite collection of sets $(A_i)_{i\in I}$, but each A_i is a subset of the natural numbers. We can define a choice function *simultaneously*: define $f(A_i)$ to be the least element of A_i. This is distinct from the successive case because we don't need to go one-by-one here; we have a prescribed rule that we can apply to all sets simultaneously.

The problem is when we have neither a successive nor a simultaneous way of defining a choice function from a prescribed rule. What if the sets $(A_i)_{i\in I}$ form an infinite collection of arbitrary subsets of the real numbers? The successive construction only works for finitely many sets, because we cannot describe an "infinite" successive construction in finitely many steps. We have no idea what the sets A_i look like, so we cannot define a simultaneous rule that works for all of them.

This problem is at the heart of the Axiom of Choice. Mathematicians want to make this choice function anyway, because it must be possible, right? It turns out that our usual axioms of set theory only allow for choices that can be finitely defined, and arbitrary choice functions do not constitute such choices. This is why we need a whole new axiom to do this.

2. The history of the Axiom of Choice

Around 1870, Cantor defined a successive limit point operation that behaves like a derivative on sets: P' is the set of limit points of P. In general, $P^{(n)}$ is the set of limit points of $P^{(n-1)}$. What is P^∞? Well,

$$P^\infty = \bigcap_{n\in\mathbb{N}} P^{(n)},$$
$$P^{\infty+1} = \left(P^\infty\right)'$$

and so on!

These numbers $\infty, \infty+1, \cdots$ are *ordinals*. In 1883, he showed that the ordinals generalize the natural numbers because the order on them is a *well order*. He asserted that every set could be well-ordered. In 1904, Julius König disproved this assertion at the International Congress of Mathematicians, but Ernst Zermelo was skeptical! He formulated the Axiom of Choice in terms of something he called a *covering* of a set. He then deduced the Well-Ordering Principle (WOP) as a consequence of the Axiom of Choice. This caused a huge amount of controversy among mathematicians, many of whom thought the Well-Ordering Principle was "obviously false".

Here is a quick breakdown of the stages of choice-making that led mathematicians to relying on the Axiom of Choice.

(1) Mathematicians selected arbitrary elements from a *finite* collection of sets, often not caring too much about which specific elements they chose. This does not require the Axiom of Choice.

(2) Mathematicians (likely analysts and number theorists in the nineteenth century) needed to choose an infinite collection of elements, and started making infinitely many choices, often with a prescribed rule.

(3) Cantor chooses infinitely many elements from a collection of sets with no prescribed rule, for example in the proof of the Intermediate Value Theorem. He could have prescribed a rule!

(4) Chaos. Mathematicians start making infinitely many choices that have no possible prescribed rule. This probably began in 1871 when Heine proved that continuity at a point was equivalent to sequential continuity at that point.

Zermelo's proof of WOP let to scandal and outrage in the European mathematical community. Baire, Borel, and Lebesgue all shared constructivistic sympathies. Borel thought Zermelo's proof was no better than trying to order a set by picking a first element, then a second, then a third, and so on . . . . Hadamard wrote Borel a letter dissenting, which Borel forwarded to Baire, who was even more vehement about the proof's incorrectness. Borel also wrote a letter to Lebesgue, who was a little more cautious: "Can one prove the existence of a mathematical object without defining it?" Borel forwarded Baire's and Lebesgue's letters to Hadamard, who thought their objection was purely psychological. The question was “Is a set well-orderable?” and not “Can *one* well-order a set?” Hadamard wrote back about this to Borel, who responded by saying that even Cantor’s arguments had only heuristic value. Meanwhile, Lebesgue, Borel, and Baire were all publishing results that used the Axiom of Choice; they just didn’t realize it. In 1905, Vitali constructed the first non–Lebesgue measurable set. In 1907, Lebesgue finally broke and attacked Zermelo’s proof for the choice functions being undecidable. In 1908, Borel also broke and argued that it did not make sense to talk about the continuum, because only countably many real numbers could even be defined.

Hadamard was also publishing his thoughts anonymously in a journal. A man named Jules Richard wrote in with what is now known as *Richard’s Paradox*: list all two-letter combinations, three-letter combinations, and so on. Delete the ones that do not correspond to real numbers. The remaining words form a countable set. But, using the same trick as Cantor, you can find a real number that is not in this set. Richard’s intention was to cast aspersions on set theory because he was skeptical about the existence of uncountable sets. Poincaré also wrote in to this journal, and was the second French mathematician to publicly accept Zermelo’s proof.

Peano had an interesting take on Richard’s paradox. He thought the issue was not with mathematics, but with natural language, which lacked the precision necessary for symbolic logic. Nevertheless, he rejected Zermelo’s proof because a proof could not have an infinite number of steps.

The axiom was better accepted in Germany: Hamel used it to construct a discontinuous additive function on the reals, Hausdorff consciously used this to prove theorems about ordered sets. In 1905, Hilbert published in the *Annalen* articles by Borel, J. König, Jourdain, Bernstein, Schoenflies, and Hamel. Only Hamel accepted Zermelo’s proof.

In Hungary, Julius König initially turned to philosophical arguments to dispute Zermelo’s proof. His son, Dénes König, tried to use it to prove that $\alpha = \omega\alpha$ for every infinite cardinal α. They both came around eventually.

In 1908, in response to the criticism, Zermelo published a modified proof along with an axiomatization of set theory, which also did not sit well with other people. Fraenkel and Skolem later contributed to Zermelo’s list of axioms, which are now abbreviated as ZF (for Zermelo–Fraenkel). Together with Choice, the axioms become ZFC.

Even a decade after Zermelo’s proof, most mathematicians did not realize how much they depended on the axiom, and it was not yet widely accepted. Cipolla was the first to realize that the statement about sequential continuity relied on the axiom. When Hausdorff produced a paradoxical decomposition of the sphere in 1914, Borel used this to argue that the axiom could not possibly be true. It was only around 1918 when Sierpinski founded the Warsaw School of Mathematics that the axiom was systematically studied by a large group of mathematicians.

Sierpinski was not interested in arguing that the axiom must be true or false, but in learning which proofs in mathematics had used the axiom. He emphasized that every proof by Baire and Lebesgue used the axiom of countable choice. For example, the statement that Lebesgue measure is countably additive. He also observed that, so far, the use of the axiom had not led to any contradiction in mathematics. Set theory really popped off after this.

Zorn’s lemma: Around 1909, Hausdorff published a *Maximal Principle*, which was essentially what we now call Zorn’s lemma. Two other mathematicians, Kuratowski and Bochner, also discovered this indepdendently. The final form of Zorn's lemma was published in 1935 not as a theorem, but as an axiom. Zorn hoped it would supersede WOP in abstract algebra; indeed, Emil Artin realized it implied AC. Zorn himself used the lemma to prove several theorems in abstract algebra.

A lot of increasingly complicated things happened in set theory in the next 30 years. Mathematicians were interested in the consistency and independence of the Axiom of Choice with the other axioms. Around 1938, Gödel proved that AC is consistent with the axioms of ZF, which means that there is no contradiction in ZFC, unlike what several mathematicians used to believe. In 1963, Paul Cohen introduced a powerful technique called *forcing* (that I know nothing about) to prove that AC is independent of ZF, so it is not implied by any of the other axioms.

3. Statements related to the Axiom of Choice

Equivalent statements

1. *Zorn's lemma*: Let $\mathcal{P}$ be a partially ordered set so that any chain has a maximal element. Then, $\mathcal{P}$ has a maximal element.
2. *Well-ordering principle*: Every set A has an order $\leq$ on it such that every nonempty subset $S \subseteq A$ has a $\leq$-minimal element.
3. Every set has a cardinality.
4. The direct product of nonempty sets is nonempty.

Consequences

5. *Banach–Tarski*: The unit ball can be partitioned into finitely many pieces and rotated to form two disjoint unit balls.
6. Let $f : \mathbb{R} \to \mathbb{R}$ be a function that is sequentially continuous at a point $x \in \mathbb{R}$, i.e. for every sequence $(x_n)_{n\in\mathbb{N}}$ converging to x, $f(x_n) \to f(x)$. Then, f is continuous at x.
7. A countable union of countable sets is countable.
8. If there is a surjection from $A \to B$, there is an injection from $B \to A$.
9. If p is any statement about sets, then p is either true or false.

Statements consistent with the negation

10. Sets have a well-defined notion of volume.

11. There is an infinite set of real numbers that does not contain a countably infinite set.

12. The real numbers are a countable union of countable sets.

4. Other axioms

There are weaker forms of the Axiom of Choice that are sometimes useful. Two popular ones are the Axiom of *Dependent* Choice and the Axiom of *Countable* Choice.

Axiom 2 (Dependent Choice)**.** *Let X be a nonempty set and $\mathcal{R}$ any subset of $X \times X$ such that for all $x \in X$, there is some $y \in X$ with $(x, y) \in \mathcal{R}$. Then, there is a countably infinite sequence $(x_n)_{n\in\mathbb{N}} \subset X$ with $(x_n, x_{n+1}) \in \mathcal{R}$ for all $n \in \mathbb{N}$.*

Axiom 3 (Countable Choice)**.** *Let $(A_n)_{n\in\mathbb{N}}$ be a countably infinite collection of disjoint nonempty sets. Then, there is a choice function $f : \{A_n : n \in \mathbb{N}\} \to \bigcup_n A_n$ such that $f(A_n) \in A_n$ for all $n \in \mathbb{N}$.*

Both Dependent Choice and Countable Choice are sufficient to give us the nice results we want for countable sets: a countable union of countable sets is countable, every infinite set has a countable subset, and Lebesgue measure is countably additive. Dependent Choice is strictly stronger than Countable Choice in that it implies Countable Choice, but is not equivalent to it. For example, the Baire Category Theorem follows from Dependent Choice but not Countable Choice.

Theorem 1. *Dependent Choice implies Countable Choice.*

Proof. Let $(A_n)_{n\in\mathbb{N}}$ be a countable collection of sets. We want to construct a choice function using Dependent Choice. Set $A = \bigcup_{n\in\mathbb{N}} A_n$, and define the relation

$$\mathcal{R} = \{(x, y) : x \in A_n, y \in A_{n+1}\}.$$

By Dependent Choice, there is a sequence $(x_n)_{n\in\mathbb{N}}$ with $(x_n, x_{n+1}) \in \mathcal{R}$. Our choice function is now $f(A_n) = x_n$. □

The Principle of Mathematical Induction (PMI) follows from the standard axioms of set theory without Choice. Here is a "proof" of Countable Choice using PMI, handily broken down into steps.

1. Let $(A_n)_{n\in\mathbb{N}}$ be a countable collection of sets.

2. By the Principle of Mathematical Induction, for every $n \in \mathbb{N}$, there is a choice function C_n on $\{A_1, \cdots, A_n\}$.

3. In fact, the inductive construction shows that C_n and C_{n-1} agree on $\{A_1, \cdots, A_{n-1}\}$.

4. Define the choice function $C_\infty = \bigcup_{n\in\mathbb{N}} C_n$. (Taking a union of functions is formally well-defined without the Axiom of Choice.) That is, $C_\infty(A_n) = C_n(A_n)$.

5. This is a choice function on $(A_n)_{n\in\mathbb{N}}$, so Countable Choice is actually a theorem of set theory without Choice.

Can this really be true? Does the Axiom of Countable Choice follow from the other axioms of set theory?

Unfortunately, no.

There is a secret *choice* hiding between steps 3 and 4. Induction guarantees that for any finite n, we can find choice functions $C_1, \cdots, C_n$ that are compatible with each other. However, we need to find an infinite sequence of choice functions $(C_n)_{n\in\mathbb{N}}$ that are compatible with each other. The secret step is *choosing* such a sequence, which is, of course, an application of Countable Choice, making this 'proof' a tautology!

Here is a fun aside on the *Law of the Excluded Middle* (LEM). Informally, the LEM states that for every proposition P, either P or its negation is true. Constructive set theorists are unhappy with the LEM, because they want to *prove* that either P or its negation is true, instead of just *knowing* that one is true. Diaconescu's theorem says, sorry constructivists,

Theorem 2. *In constructive set theory, the Axiom of Choice implies the LEM.*

However, the Axiom of Countable Choice is not enough to imply the LEM!

The last axiom we'll talk about is, in some sense, the nemesis of the Axiom of Choice: the Axiom of Determinacy.

Axiom 4. *Every game that is nice has a winning strategy.*

If you're interested, ask me what a nice game is after class!

The world with the Axiom of Determinacy is nicer in many ways: every set of real numbers is Lebesge measurable, Banach–Tarski fails, the Well-Ordering Principle fails. However, we lose everything we know about cardinality.

Theorem 3. *Assuming ZF + Determinacy, the real numbers can be partitioned into nonempty disjoint sets* $(A_i)_{i\in I}$ *such that* $|I| > |\mathbb{R}|$.

The real numbers can be partitioned into more nonempty sets than the size of the real numbers. Imagine a hotel with many floors where every real number corresponds to a room, and a floor is a set of real numbers. This theorem says that the hotel has more floors than rooms, even though each floor has at least one room!

5. Problems

Here is a list of common results that use the Axiom of Choice in some way. You are under no obligation to work on any of them. You can also skip ones you don't understand or don't enjoy (for example, if you don't like ring theory).

5.1. The direct product of nonempty sets is nonempty

Given a collection of sets $(A_i)_{i\in I}$, their direct product is the set of all sequences indexed by I:

$$\prod_i A_i = \left\{(a_i)_{i\in I} : a_i \in A_i\right\}.$$

Problem 1. Show that the Axiom of Choice (in its choice function form) is equivalent to the direct product of nonempty sets being nonempty.

5.2. Every ring has a maximal ideal

The examples of rings you should have in mind are the integers $\mathbb{Z}$, polynomial rings like $\mathbb{Z}[X]$, and the matrix rings $M_n(\mathbb{R})$. The matrix ring $M_n(\mathbb{R})$ is an especially important example because multiplication is not commutative.

In general, a ring R is a set with addition and multiplication. The ring R has to form an abelian group under addition. The multiplication satisfies the following properties

(i) *Associativity*: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

(ii) *Multiplicative identity*: There is some $1 \in R$ such that $a \cdot 1 = 1 \cdot a = 1$ for all $a \in R$.

(iii) *Left and right distributivity*:

$$a \cdot (b + c) = ab + ac,$$
$$(b + c) \cdot a = ba + ca.$$

An element $u \in R$ is called a *unit* if it has a *multiplicative* inverse u^{-1}.

Definition 4. An *ideal* I in a ring R is a nonempty subset that is closed under addition, and closed under multiplication by elements of R. That is, if $a \in I$ and $r \in R$, then $ar \in I$ and $ra \in I$.

Being an ideal is stronger than just being a subring. For example, let $\mathbb{Z}[X, Y]$ be the polynomial ring in two variables over $\mathbb{Z}$. The set $\mathbb{Z}[X]$ is a subring because it is closed under addition and multiplication. However, it is not an ideal because $X \in \mathbb{Z}[X]$ but $YX \notin \mathbb{Z}[X]$. One notable property of an ideal I is that $I = R$ if and only if I contains unit of R. This is useful to check if an ideal is proper: find a unit that doesn't belong to it.

The most standard application of Zorn's lemma to ring theory is to show that every ring R has a maximal proper ideal. That is, there is some $M \lhd R$ such that $M \neq R$ and M is not contained in any other proper ideal of R.

Problem 2. Use Zorn's lemma to show that every ring has a maximal ideal.

5.3. Every vector space has a basis

The examples of vector spaces you should have in mind are $\mathbb{R}^n$, for $n \in \mathbb{N}$. A vector space V over $\mathbb{R}$ is an abelian group under addition that is also closed under scalar multiplication by elements of $\mathbb{R}$. That is,

(i) The space V has an operation $+$ that turns it into an abelian group.

(ii) For any $v \in V$ and $\lambda \in \mathbb{R}$, λv is a well-defined element of V.

(iii) Scalar multiplication has an identity: $1 \cdot v = v$.

(iv) The scalar multiplication is compatible with the real numbers: $\lambda_1(\lambda_2 v) = (\lambda_1\lambda_2)v$.

(v) *Distributivity*:

$$\lambda(v + w) = \lambda v + \lambda w$$
$$(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v.$$

A *basis* for a vector space V is a minimal generating set for it. For example, the vectors $(1, 0)$ and $(0, 1)$ form a basis of $\mathbb{R}^2$. Here is a more formal definition.

Given distinct vectors $v_1, \cdots, v_n \in V$, a *linear combination* of them is an expression of the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$, for $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$. The linear combination is *nontrivial* if at least one of $\lambda_1, \cdots, \lambda_n$ is nonzero. Linear combinations are always finite!

A set of vectors S *spans* V if every $v \in V$ can be expressed as a linear combination of elements from S. A set of vectors T is *linearly independent* if no nontrivial linear combination of elements from T is equal to zero.

For example, the set $\{(1, 1), (1, 0), (0, 1)\}$ spans V, but it is not linearly independent because $1 \cdot (1, 0) + 1 \cdot (0, 1) - 1 \cdot (1, 1) = (0, 0)$.

Definition 5. A *basis* for V is a set S that spans V and is linearly independent.

A basis is a maximal linearly independent set, or a minimal spanning set. However, the existence of a basis is equivalent to the Axiom of Choice!

Problem 3. Show that every vector space has a basis using the Well-Ordering Principle.

5.4. Sequential continuity implies continuity

Many introductory analysis textbooks conflate sequential continuity with continuity. This is not a crime: a function is sequentially continuous everywhere if and only if it is continuous everywhere, without the Axiom of Choice. However, this becomes false when we restrict our attention to sequential continuity at a point.

Here is the definition of sequential continuity.

Definition 6. A function $f : \mathbb{R} \to \mathbb{R}$ is *sequentially continuous* at a point $x \in \mathbb{R}$ if for every sequence $(x_n)_{n \in \mathbb{N}}$ converging to x, $f(x_n)$ converges to $f(x)$.

The definition of continuity for ε-enjoyers is

Definition 7. A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at a point $x \in \mathbb{R}$ if for every $\varepsilon > 0$, there is some $\delta > 0$ such that $|x - y| < \delta$ implies that $|f(x) - f(y)| < \varepsilon$.

If you aren't familiar with this definition, think of the ε as measuring arbitrarily small distances to $f(x)$. The function f is continuous at x if whenever you are within δ of x in the domain, you can get within ε of $f(x)$ in the range.

Problem 4. Show, using a proof by contradiction, that if a function f is sequentially continuous at x, then it is continuous at x. Where did you use the Axiom of Choice in your proof? What form of it did you use?