The Axiom of Choice: Equivalent Formulations and Surprising Consequences

Zhijie Chen

November 2, 2025

What is the mysterious axiom of choice? Why the hell do we need it even just to show that every infinite set has a countable subset? Why does every vector space have a basis? …

Contents

> To choose one sock from each of infinitely many pairs of socks requires the Axiom of Choice, but for shoes the Axiom is not needed.[1]
>
> — Bertrand Russell

> The Axiom of Choice is obviously true; the well-ordering principle is obviously false; and who can tell about Zorn's lemma?[2]
>
> — Jerry Bona

> Thought is subversive and revolutionary, destructive and terrible; thought is merciless to privilege, established institutions, and comfortable habit. Thought looks into the pit of hell and is not afraid. Thought is great and swift and free, the light of the world, and the chief glory of man.
>
> — Bertrand Russell

1 Introduction

Instead of throwing the definitions and theorems straight at your face, we first raise some seemingly trivial but somewhat interesting questions.

(1) Is it true that every infinite set has a countable (countably infinite) subset?

(2) Is it true that the Cartesian product of any nonempty collection of nonempty sets is nonempty?

For the first question, you may say that hey you are just too stupid. We can construct an infinite sequence $\{x_n\}_{n\in\mathbb{N}} \subseteq X$ by the following. Take $x_0 \in X$. Inductively, after $\{x_n\}_{n\leq m}$ is fixed, take $x_{m+1} \in X\backslash\{x_n\}_{n\leq m}$ because X is infinite. But you only proved by induction (recall what induction means) that for any $n \in \mathbb{N}$ there exists a (finite) injective sequence of length n in X, rather than an infinite one. A correct proof (of course the proposition is true) is given below using the axiom of choice.

For the second question, we first make precise the notion of the Cartesian product of an arbitrary collection of sets.

Definition 1. Suppose I is an index set and $\mathcal{A} = \{A_i : i \in I\}$ is a collection of sets. The Cartesian product of $\{A_i : i \in I\}$ is defined as

$$\prod_{i\in I} A_i = \left\{(f : I \to \bigcup \mathcal{A}) : f(i) \in A_i, \forall i \in I\right\}.$$

If $I = \emptyset$ or $\emptyset \in \mathcal{A}$ then the Cartesian product is certainly empty. Otherwise, that $\prod_i A_i \neq \emptyset$ is equivalent to the existence of an $f : I \to \bigcup \mathcal{A}$ with $f(i) \in A_i$ for each i, called a *choice function*. This is precisely what the axiom of choice says!

Definition 2. Let I be an index set and $\mathcal{A} = \{A_i : i \in I\}$ be a collection of sets indexed by I. A choice function on $\mathcal{A}$ is a function $f : I \to \bigcup \mathcal{A}$ such that $f(i) \in A_i$ for each $i \in I$. Note that a set X can always be indexed by itself: $X = \{x : x \in X\}$, so a choice function on X is a function $f : X \to \bigcup X$ such that $f(S) \in S$ for each $S \in X$.

The Axiom of Choice (AC). *Any set of nonempty sets has a choice function. Equivalently, the Cartesian product of any nonempty collection of nonempty sets is nonempty.*

Remark. The axiom of choice ("C"), together with Zermelo-Fraenkel ("ZF") set theory, constitutes the standard form[3] of axiomatic set theory (ZFC), the most common foundation of

[1] The axiom of choice is needed only when we cannot give an explicit way to make the selection.

[2] This is a joke. Although the three are mathematically equivalent, many mathematicians find the axiom of choice intuitive, the well-ordering principle counterintuitive, and Zorn's lemma too complex for any intuition. I have the same feeling. :)

[3] There are other models, some of which hair-raising. Set theorists are definitely no human.

mathematics. AC is independent of ZF in the sense that if ZF is consistent (this is unknown[4]), then so is ZFC; and if ZF is consistent, then so is ZF with the negation of AC.

Enough of those confusing words about axiomatic set theory. Next we give two equivalent formulations of AC (Zorn's lemma and the well-ordering theorem) and prove their equivalence. After that, we present some interesting and perhaps surprising consequences. For readers' convenience, we here list the results that will be covered, and state Zorn's lemma and the well-ordering theorem.

Remark. If you are not at all interested in the proofs of the equivalence of the three formulations, you can totally skip them all without hindering the understanding of the interesting and perhaps surprising results about the axiom of choice.

Results that will be covered (in the following order, without any interdependence):

(1) (Done) The Cartesian product of any nonempty collection of nonempty sets is nonempty.

(2) Every infinite set has a countable subset.

(3) Every vector space has a basis.

(4) There does not exist a perfect generalization of the length of an interval to all subsets of $\mathbb{R}$. (We will make that notion precise.)

(5) Every nontrivial finitely generated group possesses maximal subgroups.

Zorn's Lemma. *If P is a nonempty partially ordered set (poset) in which every chain has an upper bound, then P has a maximal element.*

The Well-Ordering Theorem. *Every set can be well-ordered, i.e., totally ordered such that every nonempty subset has a least element.*

2 Equivalence of the Three Formulations

In the following text, we shall frequently treat functions and relations as their underlying sets.

2.1 AC implies Zorn's lemma

For the sake of contradiction, assume that P is a nonempty poset such that

(1) Every chain in P has an upper bound;

(2) P has no maximal elements.

Let $\mathcal{T}$ be the set of all chains in P; it is a set because it is a subclass of the power set $\mathcal{P}(P)$. Define functions

$$F : \mathcal{T} \to \mathcal{P}(P)\backslash\{\emptyset\}, \quad G : P \to \mathcal{P}(P)\backslash\{\emptyset\}$$

by

$$F(T) = \{x \in P : c \leq x, \forall c \in T\}, \quad G(a) = \{x \in P : a \leq x, a \neq x\}.$$

So $F(T)$ is the set of all upper bounds of the chain T, and $G(a)$ is the set of elements strictly greater than a. That $F(T)$ and $G(a)$ are nonempty follows from the two conditions above.

Now consider $\{F(T) : T \in \mathcal{T}\}$, treated as a collection of nonempty subsets of P indexed by $\mathcal{T}$, and $\{G(a) : a \in P\}$, treated as a collection of nonempty subsets of P indexed by P. By AC, let f be a choice function on $\{F(T) : T \in \mathcal{T}\}$ and g a choice function on $\{G(a) : a \in P\}$, i.e.,

- $f : \mathcal{T} \to P$ satisfies $(c \leq f(T), \forall c \in T)$ for each $T \in \mathcal{T}$;

[4]It follows from Gödel's second incompleteness theorem that ZF cannot prove its own consistency unless it is actually inconsistent.

- $g : P \to P$ satisfies $x \leq g(x), x \neq g(x)$ for each $x \in P$.

So $f(T)$ selects an upper bound of the chain T, and $g(x)$ selects an element strictly greater than x.

Define $h : \mathcal{T} \to P$ by $h = g \circ f$. This function assigns to each chain an upper bound outside the chain.

Remark. The following steps require some knowledge of ordinals and transfinite recursion. One can refer to Chapter 2 of *Set Theory* by Thomas Jech.[5]

Define a sequence $\langle a_\alpha : \alpha \in \mathrm{Ord} \rangle$ by transfinite recursion:

$$a_\alpha = h(\operatorname{im} a \restriction_\alpha) = h(\{a_\xi : \xi < \alpha\})$$

for each $\alpha \in \mathrm{Ord}$. One should be convinced that the recursion rule can be formulated as a function on the class of all transfinite sequences. By the property of h as noted above, for any ordinal α, $a \restriction_\alpha$ is an injective transfinite sequence whose image is a chain in P, ensuring that the recursion is well-defined.

That immediately leads to a contradiction. Let $a^{-1} : \operatorname{im} a \to \mathrm{Ord}$ be the surjective inverse function. By the axiom schema of replacement, $\operatorname{im} a^{-1} = \mathrm{Ord}$ is a set, a contradiction. That completes the proof.

2.2 Zorn's lemma implies AC

This proof is a typical application of Zorn's lemma. Suppose $\emptyset \notin X \neq \emptyset$ is a nonempty collection of nonempty sets. Let $U = \bigcup X$.

Let

$$P = \left\{(f : X' \to U) : X' \subseteq X, f(S) \in S, \forall S \in X'\right\}$$

be a poset under set inclusion, i.e., function extension.

Because we can easily construct an f by taking X' to be a singleton, P is nonempty. Let

$$T = \{(f_i : X_i \to U) : i \in I\}$$

be a chain in P. The following steps lead to that $\bigcup T$ is an upper bound of T in P. The proof of each step is nothing more than verifying the definitions and hence is omitted.

(1) $\bigcup T = \bigcup_{i \in I} f_i : \left(\bigcup_{i \in I} X_i\right) \to U$.

(2) $\bigcup T \in P$.

(3) $\bigcup T$ is an upper bound of T.

By Zorn's lemma, P has a maximal element $\tilde{f} : \tilde{X} \to U$. If $\tilde{X} = X$ then $\tilde{f}$ is a choice function we seek. Otherwise take $S \in X \backslash \tilde{X}$ and $x \in S$, and $\left(\tilde{f} \cup \{(S, x)\}\right) \in P$ contradicts the maximality of $\tilde{f}$. The proof is completed.

2.3 Zorn's lemma implies the well-ordering theorem

This proof is essentially the same as the proof that Zorn's lemma implies AC. Suppose X is a nonempty set (if it is empty the desired result holds trivially).

Let

$$P = \left\{(X', \leq) : X' \subseteq X \text{ is well-ordered by } \leq\right\}$$

be partially ordered by

$$(X', \leq_{X'}) \leq (X'', \leq_{X''}) \iff X' \subseteq X'', (\leq_{X'}) \subseteq (\leq_{X''}).$$

[5] I personally think the chapter on ordinals is very difficult, but Qinxiang Cao told me it is very easy!

For a chain T in P, we can similarly show that $\bigcup T$ is an upper bound of T in P. By Zorn's lemma, P has a maximal element which is necessarily X with a well-ordering, completing the proof.

2.4 AC implies the well-ordering theorem

Suppose X is a nonempty set (if it is empty the desired result holds trivially). By AC, let f be a choice function on $\mathcal{P}(X)\backslash\{\emptyset\}$.

Define a sequence a by transfinite recursion:

$$a_\alpha = f(P\backslash \operatorname{im}\langle a_\xi : \xi < \alpha\rangle) = f(P\backslash\{a_\xi : \xi < \alpha\}).$$

If $\{a_\alpha : \alpha < \theta\} = P$ for some $\theta \in \mathrm{Ord}$, then the process terminates here and we obtain a θ-sequence $\langle a_\alpha : \alpha < \theta\rangle$. This is actually always the case because otherwise Ord would be a set by the injectivity of a (see the next paragraph).

Because f is a choice function, a is injective and hence a bijection between an ordinal θ and P. That gives the desired well-ordering: for $a_\alpha, a_\beta \in P$, $a_\alpha \le a_\beta$ if and only if $\alpha \le \beta$. That this is a well-ordering follows from that every ordinal, including θ, is well-ordered.

2.5 The well-ordering theorem implies AC

Suppose $\emptyset \notin X \neq \emptyset$ is a nonempty collection of nonempty sets. By the well-ordering theorem, let $\bigcup X$ be well-ordered by $\le$.

The idea is simple and straightforward. AC is needed only when no explicit rule of selection can be given (see the first quote at the beginning of the article and the associated footnote). The well-order gives us a canonical way to make the selection, so we do not need AC.

We make the above notion precise by explicitly constructing a choice function. Let $\varphi(x, y, p)$ be the first-order formula

$$\varphi(S, s, X) = (S \in X \wedge s \in S \wedge \forall x(x \in S \to s \le x)).$$

We wrote $\varphi(x, y, p)$ above to emphasize that φ is meant to be a function (with parameter p). It can be verified that φ is the function that maps each set in X to the least element of it, a choice function as desired.

If you prefer something more formal, we can make φ a function on the universe V and restrict this class function to obtain a set function (a set) *in* the set $X \times \bigcup X$.

$$\varphi(S, s, X) = ((S \in X \wedge s \in S \wedge \forall x(x \in S \to s \le x)) \vee (S \notin X \wedge s = \emptyset)).$$

Remark. We cannot give a well-order to every set in X because that requires many choices (one choice for each set in X), which per se requires the axiom of choice. Our approach requires only one choice, namely choosing a well-order on $\bigcup X$.

3 Consequences of the Axiom of Choice

After all those proofs (they actually can be skipped without hindering the understanding of the results below), we arrive at some interesting and perhaps surprising consequences of the axiom of choice.

3.1 Every infinite set has a countable subset

The problem with the naive proof in the introduction is that inductions (of natural numbers) can never truly reach the infinity. Transfinite recursion comes to the rescue.

Let X be an infinite set. By AC, let f be a choice function on $\mathcal{P}(X)\backslash\{\emptyset\}$. Define an ω-sequence $\langle a_n : n < \omega\rangle$ by transfinite recursion:

$$a_n = f(X\backslash \operatorname{im} a \restriction_n) = f(X\backslash\{a_0, \dots, a_{n-1}\}).$$

Here ω is the least nonzero limit ordinal, i.e., the set $\mathbb{N}$ of natural numbers in the usual sense, and $n < \omega$ simply means $n \in \mathbb{N}$. We use this notation to emphasize that we are performing a transfinite recursion (with no limit steps).

That solves the problem of infinity and well-defines an injective (infinite) sequence in X, proving the desired result.

3.2 Every vector space has a basis

In linear algebra courses we usually deal only with finite-dimensional vector spaces. Let V be a vector space (of arbitrary dimension) over a field K.

Remark. One might search the internet for the definition of a vector space over a field. But in this article the field can be simply taken to be $\mathbb{R}$.

- A linear combination of a (possibly infinite) subset $A \subseteq V$ is a *finite* sum of the form $c_1v_1 + \cdots + c_nv_n$, where $a_i \in K, v_i \in A$ for each i.
- A subset $A \subseteq V$ is said to be linearly independent if there do not exist $v, v_1, \dots, v_n \in A$ such that v is a linear combination of $v_1, \dots, v_n$.
- A subset $A \subseteq V$ is said to be spanning if every vector in V is a linear combination of A.
- A linearly independent subset $A \subseteq V$ is said to be a basis of V if it spans V.

In short, a linear combination has to be finite. This is fairly reasonable: we hope $\{1, x, x^2, \dots\}$ to be a basis of the vector space $\mathbb{R}[x]$ of polynomials in x with real coefficients, and we do not hope power series to mess things up.

Notice that a basis is a linearly independent subset that is maximal under set inclusion. That motivates the use of Zorn's lemma. Let P be the set of all linearly independent subsets of V, partially ordered by set inclusion. If $V = \{0\}$ then the result is trivial; hence assume V is nontrivial. Taking A to be the singleton of any nonzero vector leads to $P \neq \emptyset$.

Let T be a chain in P. Straightforward verifications of definitions lead to that $\bigcup T$ is an upper bound of T in P. By Zorn's lemma, P has a maximal element B. It has to span V, because otherwise let $v \in V$ be a vector not spanned by B, and $B \cup \{v\} \in P$ (as you can verify) contradicts the maximality of B. Now B is a linearly independent subset that spans V, i.e., a basis of V.

3.3 There does not exist a perfect generalization of the length of an interval to all subsets of $\mathbb{R}$

The Riemann integral is not good enough:

- It does not handle functions with too many discontinuities;
- It does not handle unbounded functions;
- It interacts badly with limits (in terms of interchanging limits and integrals).

For a better theory of integration, a notion of size (a *measure*) is needed for more general subsets of $\mathbb{R}$ than intervals.

Remark. Do not ask any further questions regarding the sentence above, unless you would like to run into the hell of measure theory, real analysis and functional analysis.

No worry, we do not talk about σ-algebras and measures here. We only prove the following disappointing fact that there does not exist a perfect notion of size that satisfies all desired properties.

Theorem 3 (Vitali)**.** *There does not exist a function $\mu : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ satisfying all the following properties.*

(1) *For every open interval $I = (a, b)$, $\mu(I) = \ell(I)$ where $\ell(I) = b - a$ is the length of the interval. Here $a < b$ and $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$.*

(2) *(Countable additivity) For any pairwise disjoint sequence $\{A_k\}_{k\in\mathbb{Z}^+} \subseteq \mathcal{P}(\mathbb{R})$,*

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

(3) *(Translation invariance) For every $A \subseteq \mathbb{R}$ and $t \in \mathbb{R}$, $\mu(t + A) = \mu(A)$, where $t + A = \{t + a : a \in A\}$.*

All conditions above seem natural: our notion of size should be a generalization of the length of intervals; countable additivity is needed in order to take limits (the core of calculus); a natural notion of size should be invariant under translation. However, we now show that such a perfect function does not exist, assuming the axiom of choice.

For the sake of contradiction, assume such a μ does exist. We first derive some handy properties of μ.

(1) $\mu(\emptyset) = 0$.

(2) If $A \subseteq B \subseteq \mathbb{R}$, then $\mu(A) \leq \mu(B)$.

(3) $\mu([a, b]) = b - a$ for $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$.

To prove (1), take $A_1 = (0, 1)$, $A_2 = A_3 = \cdots = \emptyset$ and apply countable additivity. For (2), take $A_1 = A$, $A_2 = B\backslash A$, $A_3 = A_4 = \cdots = \emptyset$ and apply countable additivity. For (3), apply (2) to $(a, b) \subseteq [a, b] \subseteq (a - \varepsilon, b + \varepsilon)$.

With all preparations made, we prove the theorem. Define an equivalence relation $\sim$ on $A = [-1, 1] \cap \mathbb{Q}$ by $a \sim b$ if $a - b \in \mathbb{Q}$. Let $P = \{[a] : a \in A\}$ be the set of equivalence classes under $\sim$. By AC, let f be a choice function on P and let $V = \operatorname{im} f$. So f assigns to each equivalence class a representative of it, and V contains a unique representative of every equivalence class.

Let $\{q_1, q_2, \dots\}$ be a pairwise distinct enumeration of all rationals in $[-2, 2]$, or more formally, an injective sequence with image $[-2, 2] \cap \mathbb{Q}$. We shall be working with $\bigcup_{k=1}^{\infty}(q_k + V)$.

First note that the $(q_k + V)$'s are pairwise disjoint. Assume that $(q_i + V) \cap (q_j + V) \neq \emptyset$, i.e., there exist $a, b \in [-1, 1]$ with $a \not\sim b$ (by definition of V) such that $q_i + a = q_j + b$. But q_i, q_j are rational, which implies that $a - b \in \mathbb{Q}$, a contradiction.

Observe that

$$[-1, 1] \subseteq \bigcup_{k=1}^{\infty}(q_k + V) \subseteq [-3, 3].$$

The second inclusion is obvious. We show the first inclusion. For $x \in [-1, 1]$, let a be the unique element of $V \cap [x]$. Then $x = (x - a) + a$ where $x - a \in [-2, 2] \cap \mathbb{Q}$ and $a \in V$. Hence $x \in (x - a) + V$.

The formula above implies that

$$2 \leq \mu\left(\bigcup_{k=1}^{\infty}(q_k + V)\right) = \sum_{k=1}^{\infty} \mu(q_k + V) = \sum_{k=1}^{\infty} \mu(V) \leq 6,$$

where the inequalities follow from property (2) and (3), and the equalities follow from countable additivity and translation invariance. But it is not possible for $\mu(V) \in [0, \infty]$ to satisfy $2 \leq \sum_{k=1}^{\infty} \mu(V) \leq 6$. We reach a contradiction and the proof is completed.

Remarks. A small minority of mathematicians objects to the use of the axiom of choice. Nonetheless, the proof above warns one away from trying to construct a notion of size that satisfies all desired properties, because any such construction will necessarily contradict the axiom of choice, which is consistent with ZF.

A similar proof (after deriving some properties of outer measure) leads to that the outer measure on $\mathbb{R}$ is not additive.

Remark. This disappointing fact leads to the definitions of σ-algebras and measurable spaces, measures, measurable functions, integration with respect to a measure...

3.4 Every nontrivial finitely generated group possesses maximal subgroups

This is a purely group-theretical question so no motivations will be given.

We first consider the (very easy) finite case. Let G be a finite group of order $n > 1$. Assume for the sake of contradiction that G has no maximal subgroups. We start from the trivial subgroup $1 \leq G$. 1 is not maximal; hence there exists some subgroup H_1 such that $1 < H_1 < G$. H_1 is not maximal; hence we have some $1 < H_1 < H_2 < G$. Repeat this process n times (n is finite) and we have $1 < H_1 < \cdots < H_n < G$. But $1 < H_1 < \cdots < H_n$ implies that $|H_n| > n = |G|$, a contradiction. The proof is completed.

For the infinite case such technique no longer works. We need that G is finitely generated. Let $G = \langle g_1, \ldots, g_n \rangle$ be a finitely generated group.

Let $P = \{H : H < G\}$ be a poset under $\leq$ (set inclusion). For a chain $T = \{H_i : i \in I\}$ in P, we show that $\bigcup T$ is an upper bound of T in P following the steps below.

- Show that $\bigcup T \leq G$.

  $\bigcup T$ is nonempty. For $x, y \in \bigcup T$, suppose $x \in H_i$ and $y \in H_j$. Because T is a chain, $H_i \leq H_j$ or $H_j \leq H_i$. Without loss of generality, suppose $H_i \leq H_j$. Then $x, y \in H_j$, which implies that $xy^{-1} \in H_j \subseteq \bigcup T$.

- Show that $\bigcup T \neq G$.

  Assume that $\bigcup T \in G$. Then $g_k \in H_{i_k}$ for each $1 \leq k \leq n$. Because T is a chain, $H_{i_1} \leq H_{i_2}$ or $H_{i_2} \leq H_{i_1}$. Then $g_1, g_2 \in H_{i_1}$ (W.L.O.G suppose this way) or $g_1, g_2 \in H_{i_2}$. Now $g_1, g_2 \in H_{i_1}$ and $g_3 \in H_{i_3}$. Because T is linearly ordered, g_1, g_2, g_3 all belong to one subgroup. Repeat this process n times (n is finite) and we have $g_1, \ldots, g_n \in H_{i_j}$ for some $i_j \in I$. Then $H_{i_j} = G$, a contradiction.

- Show that $\bigcup T$ is an upper bound of T. (Trivial.)

By Zorn's lemma, P has a maximal element, namely a maximal subgroup.

Afterword

It is quite a surprise that I wrote so many words in a single day. In fact, I started writing this blog at midnight on November 1, 2025, and finished the main part by 6 p.m. the same day, writing more than 3,000 words.

I have always found set theory quite an interesting subject (as has one of my good friends). Set theory and manifold theory are probably the two branches of mathematics I find the most appealing. However, both subjects are very abstract and obscure (at least to me), and it takes an unacceptably large amount of time (for a sophomore in AI) to study even the basics of them. Due to the absence of a textbook suitable for my level, and perhaps also to my lack of mathematical maturity, I sadly gave up studying set theory half a year ago. Now I am trying to read *An Introduction to Manifolds* by Loring W. Tu, a first-year graduate-level textbook on manifold theory that assumes only one year of real analysis and a semester of abstract algebra. It takes me a great amount of time — often spending half a day for less than 10 pages. I have no idea if this is a wrong choice, but wish me good luck.