The Axiom of Choice and its Implications

by Jonathan Ben-Artzi

November 14, 2007

Abstract

Having been formulated only in 1904, for many hundreds of years of modern mathematics the axiom of choice was (unknowingly) assumed. In this talk I will discuss the axiom of choice itself and what it means, I will show some important equivalent statements (including Zorn's lemma), and I will present a few theorems (both in algebra and in analysis) that fail without assuming the axiom of choice.

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.

Bertrand Russell, 1919

1 Introduction and History

The basic axioms that we use in mathematical logic are known as the **Zermelo-Fraenkel axioms**, ZF in short. Together with the **axiom of choice** AC, they are called ZFC. The axiom of choice was formulated in 1904 by Ernst Zermelo, who, in 1908 formulated a first version of the axioms that will later be known as ZF. Abraham Fraenkel and Thoralf Skolem modified that first version in 1922, to produce ZF.

Kurt Gödel proved in 1939 that if ZF are consistent, they do not disprove AC. That is, ZFC is a consistent set of axioms. In 1963 Paul Cohen proved that if ZF are consistent, they do not prove AC. That is, ZF¬C is a consistent set of axioms.

The axiom of choice is usually singled out, and theorems that rely on it are pointed out as such, since it is not constructive (as we will later see) - it assumes the existence of a set, without telling us exactly what that set is, or how it is constructed.

The history regarding **Zorn's lemma** is not completely clear. It is clear that it is actually not Max Zorn who first proved it (in 1935)! One account is that it was first proved by Moore and Kuratowski (independently) in 1922. Another account is that it was proved by Hausdorff in 1914, and the reason it is named "Zorn's lemma" is since Zorn was known to use it a lot to prove theorems in algebra in the 1930s, and thus demonstrated how useful this lemma is.

Crucial theorems that require the axiom of choice include the Hahn-Banach theorem in functional analysis, proved in the 1920s by Hans Hahn and Stefan Banach independently, and the algebraic fact that any vector space has a basis.

While the thought of a world without the axiom of choice seems like "abstract nonsense", there are some interesting paradoxes that arise in the world *with* the axiom of choice. One of these paradoxes, the Banach-Tarski paradox, is an actual real-life physical paradox.

2 Definitions

Definition 1. *A* **partial ordering** *on a set X is a relation $\leqslant$ on X such that:*

1. *(reflexive)* $a \leqslant a$ *for all* $a \in X$
2. *(antisymmetric)* $a \leqslant b$ *and* $b \leqslant a$ *then* $a = b$
3. *(transitive)* $a \leqslant b$ *and* $b \leqslant c$ *then* $a \leqslant c$

A set together with a partial ordering is called a **partial ordered set** *or a* **poset**.

Definition 2. *A* **lattice** *is a poset such that every two element subset has an lub and a glb. It is a* **complete lattice** *if every subset has an lub and a glb.*

Definition 3. *A totally ordered set is said to be* **well ordered** *if every non-empty subset has a least element.*

Definition 4. *Let A be an index set, and let $\{S_\alpha \mid \alpha \in A\}$ be an indexed family of nonempty sets S_α. A function $f\colon A \longrightarrow \bigcup_\alpha S_\alpha$ such that $f(\alpha) \in S_\alpha$ for all $\alpha \in A$ is called a* **choice function**.

3 Formulation and Proof of Equivalences

Lemma 5. *Let X be a poset such that every well ordered subset has an* lub *in X. If $f\colon X \longrightarrow X$ is such that $f(x) \geqslant x$ for all $x \in X$, then f has a fixed point.*

Theorem 6. *The following statements are equivalent:*

- *A) For each set X, there is a function $f\colon \mathfrak{P}_0(X) \longrightarrow X$ such that $f(S) \in S$ for all $\phi \neq S \subseteq X$.*
- *B) If X is a poset such that every well ordered subset has an* lub *in X then X contains a maximal element, i.e., and element $a \in X$ such that $a' \geqslant a \Longrightarrow a' = a$.*
- *C)* ***(Hausdorff's Restricted Maximal Chain Theorem)*** *If X is a poset then X contains a maximal chain, i.e., a chain not properly contained in any other chain in X.*
- *D)* ***(Zorn's lemma)*** *If X is a poset such that every chain in X has an upper bound, then X has a maximal element.*
- *E)* ***(Zermelo's Well Ordering Theorem)*** *Every set can be well ordered.*
- *F) If $f\colon X \longrightarrow Y$ is onto, then there is a section $g\colon Y \longrightarrow X$ of f, i.e., an injection $g\colon Y \longrightarrow X$ such that $f \circ g = \mathbb{1}_Y$.*
- *G)* ***(Axiom of Choice)*** *If $\{S_\alpha \mid \alpha \in A\}$ is an indexed family of nonempty sets S_α then there exists a choice function on $\{S_\alpha \mid \alpha \in A\}$, i.e., a function $f\colon A \longrightarrow \bigcup_\alpha S_\alpha$ such that $f(\alpha) \in S_\alpha$ for all $\alpha \in A$.*

Proof.

I. $A \Longrightarrow B$: By contradiction. We let $X_a = \{x \in X \mid x > a\}$. Then by the contradiction assumption, $X_a \neq \phi$ for all a. We let $g\colon \mathfrak{P}_0(X) \longrightarrow X$ be a choice function as in (A). Define $f\colon X \longrightarrow X$ by $f(a) = g(X_a) > a$. Then $f(x) > x$ for all $x \in X$ contrary to Lemma 5.

II. $B \Longrightarrow C$: Let S be the collection of all chains in X ordered by $\subseteq$. If $C \subseteq S$ is a chain of chains then $\bigcup_{Y \in C} Y$ is a chain. Thus, every chain in S has an lub. By (B), S has a maximal element.

III. $C \Longrightarrow D$: We pick a maximal chain C and note that an upper bound for it will be maximal.

IV. $D \Longrightarrow E$: We let W be the collection of elements of the form $(U, \ll)$, where $U \subseteq X$ and $\ll$ is a well ordering on U. We order these element by $(U, \ll) \leqslant (V, \ll') \iff$ they are equal, or $(U, \ll)$ is an initial segment of $(V, \ll')$ and $\ll$ is a restriction of $\ll'$ to $U \times U$.

Every chain in W has a (least) upper bound, namely the union of its elements. By (D), there exists a maximal (w.r.t. $\leqslant$) well ordering, say $(U, \ll)$.

We claim that $U = X$. Otherwise, we let $x \in X \backslash U$, and define $(U \cup \{x\}, \ll')$ to be an extension of $(U, \ll)$, defining x to be the larger than any $u \in U$. This contradicts maximality of U.

V. $E \Longrightarrow F$: Well order X, and let $g(y)$ be the first element of $f^{-1}(y)$. Then $(f \circ g)(y) = y$.

VI. $F \Longrightarrow G$: We let $S = \bigcup_\alpha S_\alpha$ and $X = \{\langle s, \alpha\rangle \in S \times A \mid s \in S_\alpha\}$. We let p_S and p_A be the projections onto the first and second coordinates. p_A is onto since for all α, $S_\alpha \neq \phi$. By (F) there is a section g: $A \longrightarrow X$ for p_A, i.e. $g(\alpha) = \langle s, \alpha\rangle$ for some $s \in S_\alpha$.

Let $f = p_S \circ g$: $A \longrightarrow S$. Then f is a choice function since $f(\alpha) = p_S(g(\alpha)) = p_S(\langle s, \alpha\rangle) = s$ for some $s \in S_\alpha$ and for all $\alpha \in A$.

VII. $G \Longrightarrow A$: For $T \in \mathfrak{P}_0(X)$, we define $S_T = T$. Then $\mathfrak{P}_0(X) = \{S_T \mid T \in \mathfrak{P}_0(X)\}$ is an indexed family of nonempty sets. Also, $\bigcup_T S_T = X$ since for any $x \in X$, $x \in \{x\} = S_{\{x\}}$. So, by (G) there is a function $f\colon \mathfrak{P}_0(X) \longrightarrow \bigcup_T S_T$ such that $f(T) \in S_T = T$ for any $\phi \neq T \subseteq X$. $\square$

4 Results of the Axiom of Choice (AC)

4.1 Equivalent Results

4.1.1 Algebra

- Every vector space has a basis.
- Every ring with a unit that is not trivial contains a maximal ideal.

4.1.2 Topology

- Tychonoff's Theorem: *Any product of compact spaces is compact.*

4.2 Weaker Results (that require AC)

4.2.1 Set theory

- Any infinite set has a countable subset.
- A countable union of countable sets is countable.

4.2.2 Algebra

- Every field has an algebraic closure.

4.2.3 Measure theory

- Vitali Theorem: There exists $V \subseteq \mathbb{R}$ that is not Lebesgue measurable.
- The Banach-Tarski paradox.

4.2.4 Functional analysis

- The Hahn-Banach Theorem: *If $Y \subseteq X$ is a vector subspace of a normed linear space X, and if ψ is a continuous linear functional on Y, then there exists a continuous linear functional φ on X that extends ψ, and with $\|\varphi\| = \|\psi\|$.*
- The Baire Category Theorem: *If $(X,\, d)$ is a complete metric space, then (a) the intersection of countably many open dense sets is nonempty, and (b) X is not the union of countably many closed nowhere dense sets.*
- The Banach-Alaoglu Theorem: *Let X be a normed space with dual X^*. Then the closed unit ball in X^* is compact in the weak-* topology.*
- Every Hilbert space has an orthonormal basis.

5 Results of the Negation of the Axiom of Choice (¬AC)

- There is a model of ZF¬C in which the real numbers are a countable union of countable sets.
- There is a model of ZF¬C with a field with no algebraic closure.
- There is a model of ZF¬C with a vector space with no basis.
- There is a model of ZF¬C with a vector space with two bases of different cardinalities.

- There is a model of ZF¬C in which any $V \subseteq \mathbb{R}$ is Lebesgue measurable.
- In *all models* of ZF¬C, the generalized continuum hypothesis does not hold, i.e., it is not true that $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for all ordinals α.

6 Examples

6.1 Why the axiom of choice is not constructive

Suppose we want to classify all subsets of $\mathbb{N}$ as either "small" sets or not "small". Define the word "small" as follows:

1. Any set with zero or one members is small.
2. Any union of two small sets is small.
3. A set is small iff its complement isn't small.

It is not hard to give examples that satisfy any two of these three conditions (small may mean "finite"; small may mean "does not contain 1"; small may mean "contains at most one of the numbers 1, 2, 3").

However, it is impossible to give an example which satisfies all three conditions, although such an example is guaranteed to exist by the axiom of choice (the proof of this is nontrivial).

6.2 Ugly Functions

> *Logic sometimes breeds monsters. For half a century there has been springing up a host of weird functions which seems to strive to have as little resemblance as possible to honest functions that are of some use . . . They are invented on purpose to show our ancestor's reasonings at fault, and we shall never get anything more out of them.*

Henri Poincaré, 1906

Definition 7. *The equation $f(x+y) = f(x) + f(y)$ is called the* **Cauchy-equation**.

A function $f: \mathbb{R} \longrightarrow \mathbb{R}$ that satisfies the Cauchy-equation satisfies:

i. $f(r \cdot x) = r \cdot f(x)$ for all rational r and real x.
ii. An immediate result is $f(r) = f(1) \cdot r$ for all rational r, and, if f is continuous we get
iii. $f(x) = f(1) \cdot x$ for all $x \in \mathbb{R}$.

Definition 8. *Non-continuous solutions of the Cauchy-equation are called* **ugly**.

There are no solutions of the Cauchy-equation that are not continuous in ZF, but assuming AC, not only do we get ugly solutions, but there are far more such ($2^{\left(2^{\aleph_0}\right)}$ to be exact) than there are continuous solutions ($2^{\aleph_0}$ to be exact).

Furthermore, we have the following results: (for proofs see [2]).

Theorem 9. *If $f: \mathbb{R} \longrightarrow \mathbb{R}$ is ugly, then its graph is dense in $\mathbb{R}^2$.*

Theorem 10. *Ugly functions are non-measurable.*

6.3 The Vitali Monsters

Named after Giuseppe Vitali who showed this in 1905.

Theorem 11. *There exists $V \subseteq \mathbb{R}$ that is not Lebesgue measurable.*

Proof. (Any operation we do here is understood to be mod 1). Consider the interval $I = [0, 1)$. For any $x \in I$, define the equivalence relation $x \sim y$ iff $x - y$ is a rational number. Let $[x]$ be the equivalence class of x. By the axiom of choice, we can choose a set $V \subseteq I$ that has exactly one representative from each equivalence class. Such V is called a *Vitali Monster*. We will now show that V is not measurable.

First, we enumerate the rationals in I by p_k, and define $V_k = V + p_k$. If V is measurable, then V_k is measurable for any k, and $m(V_k) = m(V_j)$ for all j, k, since the Lebesgue measure is translation invariant. Further, we notice that $V_k \cap V_j = \phi$ for $j \neq k$, and that $\bigcup_k V_k = I$. Thus we should have that:

$$1 = m(I) = m\left(\bigcup_k V_k\right) = \sum_k m(V_k) = \sum_k m(V),$$

which is impossible. □

Remark 12. Regarding V through Zorn's lemma, one might see that it is, in fact, a maximal set contained in I such that for any $x, y \in V$, $x - y$ is irrational.

Remark 13. Since there are $2^{\aleph_0}$ equivalence classes, and each equivalence class is countable, there are exactly $\aleph_0^{\left(2^{\aleph_0}\right)} = 2^{\left(2^{\aleph_0}\right)}$ Vitali Monsters.

6.4 The Banach-Tarski Paradox

This paradox was shown by Banach and Tarski in 1924.

Consider the space $\mathbb{R}^n$ with $n \geqslant 3$. Let G be the group of isometries on $\mathbb{R}^n$.

Definition 14. *Let $A, B \subseteq \mathbb{R}^n$, $n \geqslant 3$, and suppose $A = \bigcup_{i=1}^k A_i$, $B = \bigcup_{i=1}^k B_i$ are disjoint nontrivial decompositions of A and B into k subsets. If there exist $g_1, \ldots, g_k \in G$ such that $g_i(A_i) = B_i$, we say that A and B are **G-equidecomposable** using k pieces, and that A_i and B_i are **congruent**, and we write $A_i \approx B_i$.*

Definition 15. *A set $E = A \cup B$ that is the disjoint union of two sets A and B (all in $\mathbb{R}^n$) such that E and A, as well as E and B are G-equidecomposable is called **paradoxical**. A set E that is decomposable into such sets A and B is said to have a **paradoxical decomposition**.*

Theorem 16. *A three-dimensional Euclidean ball is equidecomposable with two copies of itself, hence making it paradoxical.*

In other words, this means that we can decompose the ball in $\mathbb{R}^3$ into finitely many pieces, re-arrange them without distorting them, and produce two *identical* balls, hence contradicting preservation of mass.

Proof. I will give a rough sketch of the proof. A detailed treatment is given in [2] and [3].

Step 1. A paradoxical decomposition of the free group in two generators. Recall that the free group F_2 on two generators a and b is the group of all finite words in a, a^{-1}, b, b^{-1} appearing in minimal form (i.e. $a a^{-1}$, $a^{-1} a$, $b b^{-1}$, $b^{-1} b$ all vanish). The neutral element of this group is the empty string e. Now, we let $S(a)$ be all strings that begin with an a, and similarly for $S(a^{-1}), S(b), S(b^{-1})$. So we have:

$$F_2 = \{e\} \cup S(a) \cup S(a^{-1}) \cup S(b) \cup S(b^{-1}). \tag{1}$$

But we may alternatively write:

$$F_2 = a\, S(a^{-1}) \cup S(a) \tag{2}$$

and

$$F_2 = b\, S(b^{-1}) \cup S(b). \tag{3}$$

We now cut F_2 into four pieces according to the decomposition of (1), ignoring e. By taking $S(a)$, and multiplying $S\left(a^{-1}\right)$ on the left by a, we get back F_2 as in (2). Similarly we construct another copy of F_2 using (3).

Step 2. A group of rotations in $\mathbb{R}^3$ that is isomorphic to F_2. We consider two orthogonal axes in $\mathbb{R}^3$ and let A and B be rotations of some irrational multiple of π around each of these axes. A and B form a free group H (on two generators). So $H \cong F_2$.

Step 3. Construct a paradoxical decomposition of $S^2 = \partial B_1(0) \subseteq \mathbb{R}^3$. For any $x \in S^2$, let $[x]$ be the orbit of x under H. Notice that $\mathfrak{C} := \left\{[x] \mid x \in S^2\right\}$ is dense in S^2. Furthermore (I omit this part of the proof), $S^2 \backslash \mathfrak{C}$ has measure 0.

By the axiom of choice, we may choose $x \in [x]$ for any orbit $[x]$. Let M be the set of all chosen x. Thus almost every point $s \in S^2$ can be reached in exactly one way, by applying the proper rotation from H to the proper element of M.

Thus, the paradoxical decomposition of H yields a paradoxical decomposition of S^2.

Step 4. Extend the decomposition of S^2 to a decomposition of $B_1(0)$. Connect any $s \in S^2$ with the origin $0 \in \mathbb{R}^3$ and consider the segment connecting them, open at the origin and closed at s. The paradoxical decomposition that we have for S^2 thus extends to $B_1(0) \backslash \{0\}$, which can be further extended to $B_1(0)$. □

Remark 17. The Banach-Tarski Paradox is slightly more general: *Any two bounded subsets A and B of $\mathbb{R}^3$, that contain some ball each, are equidecomposable.*

Remark 18. The main omitted parts in the above proof are:

- In **step 1**, taking into account the empty string makes things slightly more complicated.
- In **step 3**, it is actually a theorem (not so hard) to show that a paradoxical decomposition of F_2 induces a paradoxical decomposition on a space X on which F_2 acts with *no fixed points*.
- Since H actually does have fixed points in its action on $\mathbb{R}^3$, adjustments are made to the above.
- The origin requires some treatment.

These issues are resolved since the "problematic" sets are countable and are thus relatively easy to deal with.

Remark 19. The fact that there is a need to deal with countably many "problematic" points results in having to decompose into more parts.

While we have shown that we can cut F_2 into four parts and then reassemble two copies of it, Raphael Robinson proved that there is no paradoxical decomposition of the unit ball into four parts. He managed to show that it is possible with five parts, which is therefore a sharp result.

Bibliography

[1] Bredon, *Topology and Geometry*, Springer, 1993
[2] Herrlich, *Axiom of Choice*, Springer, 2006
[3] Jech, *The Axiom of Choice*, Elsevier, 1973