The Good & Evil of the Axiom of Choice

Kenneth Chu

Department of Mathematics

University of Utah

March 19, 2002

Abstract

The Axiom of Choice (AC) is undeniably the most (well, the only) controversial axiom of Set Theory, the foundation of modern mathematics. In this talk, there are three points I would like to make:

1. AC is intuitively appealing (a lot of people will disagree).

2. AC is everywhere; in particular, a great deal of powerful results in functional analysis and PDEs rely on consequences of AC.

3. AC has certain "absurd" consequences for which one might almost want to reject AC as an axiom.

Personally, I find it mildly interesting to be "AC-alert," yet I am starting to feel that I am being perceived as "AC-lunatic" instead. I am planning on using this talk to vindicate myself (it might very well do the exact opposite). Come see for yourself; it is going to be fun.

The Axiom of Choice (AC)

Let $X = \{X_\alpha\}_{\alpha \in A}$ be a collection of non-empty sets, indexed by the set A. Then there exists a function $c : A \to \bigcup_\alpha X_\alpha$ such that for each $\alpha \in A$, we have $c(\alpha) \in X_\alpha$.

Such a function is called a *choice function*.

In plain English ...
Given a collection of nonempty sets, one can *choose* one element from each set in the collection.

The Zermelo-Fraenkel-Choice (ZFC) Set Axioms

1. (Empty Set Axiom) There exists a set $\varnothing$ which contains no elements.
2. (Axiom of Infinity) There exists a set a such that $\varnothing \in a$ and $\{x\} \in a$ whenever $x \in a$.

---

3. (Extension Axiom) Two sets are (said to be) identical if they have the same elements.
4. (Pair Axiom) If x, y are sets, then there is a set $\{x, y\}$ whose elements are x and y.
5. (Union Axiom) If x is a set (of sets), then there exists a set $\bigcup x$ whose elements are the elements of the elements of x.
6. (Power Set Axiom) If x is a set, then there exists a set $\mathcal{P}(x)$ whose elements are the subsets of x.
7. (Selection Axiom) Let a be a set and let ϕ be a "statement" involving one free variable x. Then there exists a set b whose elements are the elements x of a such that $\phi(x)$ is true, i.e. $b = \{x \in a \mid \phi(x) \text{ is true}\}$
8. (Replacement Axiom) Let a be a set and $\psi(x, y)$ be a statment involving two free variables such that for any x, there exists at most one y that satisfies $\psi(x, y)$. Then, $b := \{y \mid \psi(x, y) \text{ holds for some } x \in a\}$ is a set.

---

9. (Foundation Axiom) For any nonempty set x, there exists $y \in x$ such that $x \cap y = \varnothing$. (This axiom prevents Russel's paradox.)
10. The Axiom of Choice (AC).

Remarks on ZFC

- If we delete AC from ZFC, we obtain the axioms of what is called the Zermelo-Fraenkel (ZF) set theory.

- It has been proved AC is independent of ZF: There are (consistent) alternative set theories in which AC fails.

- Not all axioms in ZF (as presented here) are independent.

Theorem 1: AC is intuitively appealing!

Proof To convince you of this, I would argue that the negation of AC is "absurd."

Choice functions & Cartesian products

Recall that the standard set-theoretic way of defining the Cartesian product $X \times Y$ of two nonempty sets X and Y is as follows:

$$X \times Y := \{f : \{1,2\} \to X \cup Y \mid f(1) \in X, f(2) \in Y\}.$$

Hence if $f \in X \times Y$, then f is simply a choice function on $\{X_i\}_{i=1}^2$, where $X_1 = X$ and $X_2 = Y$.

Similarly, for a finite collection $\{X_i\}_{i=1}^n$ of nonempty sets,

$$X_1 \times \cdots \times X_n := \left\{ f : \{1, \ldots, n\} \to \bigcup_{i=1}^n X_i \;\middle|\; f(i) \in X_i, \forall\, i \right\}.$$

In other words,

$$\begin{matrix}\text{Cartesian product} \\ \text{of } X_1, \ldots, X_n\end{matrix} = \left\{ \begin{matrix}\text{choice functions} \\ \text{on } \{X_i\}_{i=1}^n\end{matrix} \right\}$$

An induction argument shows that the nonemptiness of each X_i implies that choice functions always exist for a finite collection of nonempty sets; in terms of Cartesian products, this says that $X_1 \times \cdots \times X_n$ is nonempty.

Observe also that if you add a finite number of nonempty sets to a finite collection of nonempty sets, the resulting Cartesian product becomes larger!

Now, if AC is ever to fail, then there must exist some infinite collection $\{X_\alpha\}_{\alpha \in A}$ of nonempty sets whose Cartesian product $\Pi_{\alpha \in A} X_\alpha$ is empty, but yet each of its finite sub-product, supposedly contained in the infinite product $\Pi_{\alpha \in A} X_\alpha$, is nonempty! ■

Theorem 2: AC is everywhere!

Proof Clearly, it suffices to show that AC is used even in a completely innocent-looking 3rd year level real analysis proof. Hence, consider:

(Cantor) The union of a countable family $\{A_i\}_{i=1}^{\infty}$ of countable sets is countable.

Proof

$$\text{Countability of } A_i \iff \mathcal{F}_i := \{f : \mathbb{N} \to A_i \mid f \text{ bijective}\} \neq \varnothing.$$

By AC, we may pick, for each A_i, an element $f_i \in \mathcal{F}_i$. We now see that the map $(i, j) \longmapsto f_i(j)$ is a bijection from $\mathbb{N} \times \mathbb{N}$ onto $\cup A_i$. And, $\mathbb{N} \times \mathbb{N} \approx \mathbb{Q}$ is itself denumerable. ∎

Remark AC is necessary for the validity of the above in following sense: There are models of set theory (with the ZF axioms) in which the above statement actually fails!

Other VERY familiar results whose dependence on AC might have eluded you as well as some of the greatest mathematicians of all times:

- Suppose $f : \mathbb{R} \to \mathbb{R}$ is bounded above (hence $\sup_{\mathbb{R}} f \in \mathbb{R}$ exists.) Then, for any upper bound $\alpha \in \mathbb{R}$ of f, $\alpha = \sup_{\mathbb{R}} f$ if and only if there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ such that $f(x_n) \to \alpha$ as $n \to \infty$.

- (Cantor, Borel, Russell) Every infinite set has a denumerable subset.

- (Cantor) $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if it is sequentially continuous.

Some STELLAR Consequences of AC

- The Bolzano-Weierstrass Theorem: For metric spaces (or first countable topological spaces), compactness is equivalent to sequential compactness.

- The Hahn-Banach Extension Theorem: Every bounded linear functional on a subspace of a normed space has an extension to the whole space of the same norm.

- Tychonoff's Theorem: The product of a collection of compact topological spaces with the product topology is itself a compact topological space. ($\Longleftrightarrow$ AC in ZF)

- Every proper ideal of a ring with identity is contained in a maximal ideal. ($\Longleftrightarrow$ AC in ZF)

- The Weak Hilbert Nullstellensatz: If $\mathbb{F}$ is an algebraically closed field, then an ideal $I \subseteq \mathbb{F}[x_1, \ldots, x_n]$ is proper if and only if its zero set $Z(I) \subseteq \mathbb{F}^n$ is non-empty.

- The Banach-Alaoglu Theorem: The norm closed unit ball of the norm dual of a Banach space is weak-$*$ compact.

- (Consequence of Banach-Alaoglu) A Banach space is reflexive if and only if its norm closed unit ball is weakly compact.

A Crash Course in Functional Analytic Methods in PDEs

We seek solutions to the following boundary value problem (BVP):

$$\begin{cases} -\Delta u &= f(x,u(x)), & x \in \Omega \\ u &= 0, & x \in \partial\Omega \end{cases}$$

Weak Formulation

$$\int_\Omega \nabla u(x) \cdot \nabla \varphi(x)\, dx - \int_\Omega f(x,u(x))\varphi(x)\, dx = 0,$$

$$\text{for all } \varphi \in C_0^\infty(\Omega) \subseteq H_0^1(\Omega) =: X.$$

Energy Functional

Define $\Phi : H_0^1(\Omega) \longrightarrow \mathbb{R}$ by

$$\Phi(u) = \frac{1}{2}\int_\Omega |\nabla u(x)|^2\, dx - \int_\Omega \int_0^{u(x)} f(x,\tau)\, d\tau\, dx.$$

Then, Φ is *differentiable* and $\Phi' : X \longrightarrow X^*$ can be computed:

$$\langle\, \Phi'(u), \varphi \,\rangle = \int_\Omega \{\nabla u(x) \cdot \nabla \varphi(x) - f(x,u(x))\varphi(x)\}\, dx$$

What have we done?

$$\text{Weak formulation} \Longleftrightarrow \Phi'(u) = 0$$

We have turned our original BVP into a critical point problem.

A Crash Course in Functional Analytic Methods in PDEs (cont'd)

Game Plan
We have turned our original BVP into a critical point problem. One way to proceed is to determine conditions which ensure that Φ has a minimum, in particular a critical point.

Lower Half of the Extreme Value Theorem
A lower semi-continuous $\mathbb{R}$-valued function (preimages of open upper half rays are open) defined on a compact set is bounded below and attains its infimum.

There are conditions that ensure that Φ is weakly lower semi-continuous (usually invokes Hahn-Banach) and coercive ($\Phi(u) \to \infty$ as $||u|| \to \infty$).

Punchline
Coercivity implies that minimization is essentially done on a closed ball $B \subseteq X = H_0^1(\Omega)$. Banach-Alaoglu says that the closed ball B is weakly compact. Hence weak lower semi-continuity implies that Φ admits a minimum on B by the Lower Half of Extreme Value Theorem.

So Far, (AC seems) So Good

- AC is intuitively appealing ...
- AC is everywhere: Even 3rd year math students need to use it for Real Analysis homework ...
- Results whose proofs directly invoke AC are numerous and appear in all branches of mathematics. (Not to mention the ones that only invoke it implicitly.)
- It keeps the algebraic geometers, PDE people, etc. employed ...

Where Is the Evil?

Let the Evil Begin!!

Remember?

Remember Emina?

Our beloved Chairwoman Alibegovic masterfully presented a most memorable GSAC colloquium roughly the same time last year on the curious phenomenon called the Banach-Tarski Paradox.

We express our thanks and adoration for her!!

(Applaud.)

The Banach-Tarski Paradox
Let B be the unit ball in $\mathbb{R}^n, n \geq 3$. Then there exists a decomposition of B into finitely many disjoint subsets which can be reassembled, via rigid motions in $\mathbb{R}^n$, into two disjoint unit balls.

The proof of the Paradox requires the use of AC. (Ask the Chairwoman Herself for a proof.)

Some authors seem to present the Paradox as a novelty result of AC.

My take on it:
The Paradox, among other "evil" consequences of AC, are *warnings* that there *might* be something wrong with AC, because it has consequences that *seem* absurd, e.g. the Banach-Tarski Paradox.

So, Here is the Heart of the Controversy Surrounding AC (in my opinion)

It seems an innocently natural property to assume *sets* to have.

Its negation seems absurd.

It is required in many garden-variety arguments in mathematics.

Countable unions of countable sets are countable.

It implies a large number of very important results, supporting entire fields of mathematical research.

Hahn-Banach, Weak Hilbert Nullstellensatz, Banach-Alaoglu, etc.

And yet, it also has "absurd" consequences for which one almost wants to reject AC as an axiom!!

The Banach-Tarski Paradox

A Much Closer-to-Home Evil of AC

Who is the culprit of you having to go through the nightmare* of measure theory?

*Exaggerated language used here for dramatic effects, no offence to Analysis lovers ...

Definition

A function $\mu : \mathcal{A} \to [0, \infty]$ on a σ-algebra $\mathcal{A}$ of subsets of $\mathbb{R}$ is called an *Lebesgue-like measure* on $\mathbb{R}$ if

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

2. μ is σ-additive.

3. $\mu(I) =$ length of I, for every finite interval $I \subseteq \mathbb{R}$.

4. μ is translation-invariant.

If the domain of a Lebesgue-like measure μ is all of $\mathcal{P}(\mathbb{R})$, then μ is called an *ideal measure* on $\mathbb{R}$.

Theorem 3

In ZFC, there exists no ideal measure on $\mathbb{R}$.

The standard proof of this fact uses AC, and that usage of AC is essentially the same as that in the proof of the Banach-Tarski Paradox.

Proof of Theorem 3

Define an equivalence relation $\sim$ on $[0,1]$ by:

$$x \sim y \iff x - y \in \mathbb{Q}.$$

Each equivalence class of $\sim$ is non-empty, and by AC, we may thus construct a set E which intersects each equivalence class at exactly one point.

Claim: $\mu(E)$ must be undefined for any Lebesgue-like measure μ.

Let $r_1, r_2, \ldots$ be an enumeration of $\mathbb{Q} \cap [-1,1]$, and for each $n \in \mathbb{N}$, let $E_n := r_n + E$.

Then (a) E_n is pairwise disjoint, and (b) $[0,1] \subseteq \bigcup_{n=1}^{\infty} E_n \subseteq [-1,2]$.

Suppose on the contrary that $\mu(E)$ is defined. Then $\mu(E_n)$ is defined for each n by translation-invariance of μ.

Hence,

$$3 = \mu([-1,2]) \geq \mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n) = \sum_{n=1}^{\infty} \mu(E) \Longrightarrow \mu\left(\bigcup_{n=1}^{\infty} E_n\right) = 0.$$

But,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \geq \mu([0,1]) = 1 > 0.$$

We have reached a contradiction. $\blacksquare$

Corollary

In ZFC, there exist non-Lebesgue measurable subsets in $\mathbb{R}$.

So, if we admit AC in our system of axioms, then a measure theory on a proper sub-σ-algebra of $\mathcal{P}(\mathbb{R})$ is really the best we can hope for.

And, Measure Theory is largely concerned with selecting a proper sub-σ-algebra of $\mathcal{P}(\mathbb{R})$ that is large enough for practical use and on which a measure can be defined.

Last Nail to the Coffin

We have seen that AC implies the non-existence of ideal measures on $\mathbb{R}$, or equivalently, existence of non-Lebesgue measurable subsets of $\mathbb{R}$.

A natural question: Is AC necessary? That is, can a non-Lebesgue measurable set be constructed in ZF (i.e. without using the AC)?

Answer: No!! Robert M. Solovay[†] proved that the AC is also a necessary condition in the sense that there are "alternative" set theories (with ZF but not AC) such that every set in $\mathcal{P}(\mathbb{R})$ is Lebesgue measurable.

[†] *A model of set-theory in which every set of reals is Lebesgue measurable*, 1970

The Greatest Evil (or Good?) of AC: Alternative Set Theories

First, think of Non-Euclidean Geometries

One of the great intellectual shocks of the 19th century was the discovery of non-Euclidean geometries:

It shattered two millenia of certainty that Euclid had described the ideal geometry to which our surveying and measuring were approximations. Theorems of Euclidean Geometry were no longer "true" but merely consequences of its axioms.

The situation with Set Theory is similar ...

Alternative Set Theories

AC has prompted logicians and mathematicians to study which results depend on AC and to what extent.

In order to do so, they construct alternative set theories by adding and deleting axioms in ZFC, check consistency of the new set theory and examine what happens to the original result in the new set theory.

For example, one can replace AC in ZFC with the following

Axiom of Determinacy (AD) In any two-player game in which draws are not possible, one player has a winning strategy.

One can prove that ZF+AD is consistent and in it, every subset of $\mathbb{R}$ is Lebesgue measurable.

Alternative Set Theories, so what?

Dieudonné [1976]:
... there is an infinity of different possible mathematics, and for the time being no definitive reason compels us to choose one of them rather than another.

Mostowski [1967]:
... axiomatic set theory is hopelessly incomplete ... Of course, if there are a multitude of set theories, then none of them can claim the central place in mathematics. Only their common part could claim such a position; but it is debatable whether this common part will contain all the axioms needed for a reduction of mathematics to set theory.

For example, the existence of solutions result for the BVP I outlined for you earlier is not really "true" — it is merely a consequence of the axioms we have adopted, namely ZFC.

But ZFC is not God-given. Why do we choose it?

Because it is convenient? Because it gives us all these high-power results? But we know darn well ZFC isn't even that convenient: in ZFC, you can cut up a ball, put the pieces together and get two balls back!

Maybe there will come a day when math biologists will use a certain set theory to model immunology, another set theory to model blood flow, yet another the brain ...

********* *The End* *********