The Axiom of Choice and Its Consequences

Prishita Dharampal

June 2022

Abstract

The Axiom of Choice is responsible for causing a rift in the mathematical community; those who believe it are in awe of the beautiful results it yields, and the rest are so baffled that they refuse to accept it. This paper introduces the reader to elementary concepts of set theory and topology, familiarizes them with the notion of the Axiom of Choice, discusses it's equivalences, and finally concludes by talking about the consequences of both, acceptance and negation of the Axiom.

Contents

1 Introduction

If we were to take a complete graph (every pair of vertices is connected by an edge) with an uncountably infinite number of vertices, and color each of those vertices such that no two vertices sharing an edge have the same color, how many colors do you think the graph would need?

Say, the graph was the unit distance graph; that is a graph who's vertices are points on the real plane (i.e. ordered pairs of real numbers), and two vertices are connected by an edge if and only if their distance in the plane is exactly one. Then what would be the answer?

Seven. Well, not exactly. We know it cannot be four (try it out using different unit distance graphs), so the answer is somewhere between five and seven.[1] We know this because of a theorem in graph theory, namely the De Bruijn–Erdős theorem, which (informally) states that an infinite graph, can be divided into smaller, finite graphs (also known as finite subgraphs), and if the maximum colors these graphs requires is q, then the chromatic number of the infinite graph is exactly q.

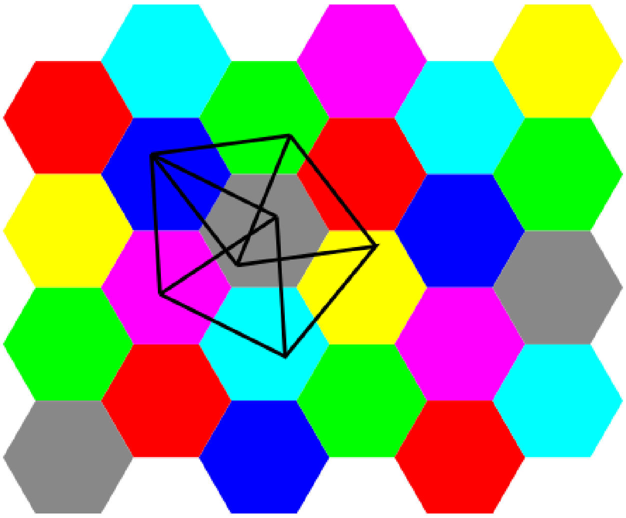


Figure 1. Image courtesy of Wikiwand.

We will talk about, and prove this theorem, when we near the end of the paper and have developed the thought process required.

This colorful, dainty theorem is one amongst many equivalences of Tychonoff's Theorem[2], which in turn is an elegant equivalent of the Axiom of Choice.

This paper starts off by giving the reader some preliminary information about the math it uses. Then in section 3 we explain and prove some basic topological facts, introduce the topic of spaces, and eventually prove Tychonoff's theorem. In section 4 the reader is introduced to the topics of infinity, the Zermelo-Frankeal Set Theory and the star of our paper: The Axiom of Choice; section 5 is a very short section on Zorn's Lemma. Section 6 discusses the consequences of the Axiom of Choice, and we end by talking about the effects of the negation of the Axiom in the mathematical world in section 7.

[1]This graph is a particularly interesting one; and all we know about it as of today is the fact that the chromatic number lies between 5 and 7.

[2]Tikhonov's surname is also transliterated as "Tychonoff", supposedly since he originally published in the German language. The first transcribed and detailed proof for the general case of Tychonoff's theorem seems to be by Eduard Čech.

Acknowledgements

The author of this paper would like to thank Chris Lambie-Hanson, George Simmons, Jay Daigle for their helpful explanatory texts, Simon Rubinstein-Salzedo for the opportunity and guidance, and Maxim Gilula for being an absolutely awesome TA.

2 Preliminary Information

"It is sometimes said that mathematics is the study of sets and functions. Of course, this oversimplifies matters; but it does come as close to the truth as an aphorism can."[3]

A lot of the math we have done till now is based fundamentally on sets. Simplistically, sets are collections of objects. (One must be careful here to avoid paradox; ignore these subtleties for now.)

Definition 2.1. *Sets* are represented as a collection of well-defined entities or elements. They are denoted using capital letters.

For example,

$A = \{1, 2, 3, \ldots, 19\}$

Here, A is the set of natural numbers less than 20.

Definition 2.2. A *function* is a relation between an input set (the domain) and an output set (the co-domain) that assigns to each element of its domain exactly one element of its co-domain.

Definition 2.3. The *order of a set* defines the number of elements a set is having. It describes the size of a set. The order of set is also known as its *Cardinality.*

For example,

Set A has a Cardinality of 19. It is denoted like so, $n(A) = 19$.

Definition 2.4. A *class* is a collection of sets that can be unambiguously defined by a property that all its members share.[4] A class is usually denoted using $\{A_i\}$, where A is the name of the class and i is the index number.

For example the class of all groups or the class of all vector spaces.

Definition 2.5. Generally, *open sets* can be defined as the members of a given collection of subsets of a given set, a collection that has the property of containing every union of its members, every finite intersection of its members, the empty set, and the whole set itself.

However, open sets are usually used to provide a notion of nearness that is similar to that of metric spaces, without having a notion of distance defined.

Conversely, a *closed set* is a set whose complement is an open set.

Definition 2.6. A *metric space* is a set together with a metric on the set. The metric is a function that defines a concept of distance between any two members of the set, which are usually called points.

Definition 2.7. A *sequence* is an enumerated collection of objects in which repetitions are allowed and order matters.

For instance $\{x_n\} = \{x_1, x_2, x_3, \ldots, x_n\}$.

Definition 2.8. A *partition of a set* A is a disjoint class $\{A_i\}$ of non-empty subsets of A whose union is A itself. The sets in $\{A_i\}$ are called *partition sets.*

Definition 2.9. When we associate $\{A_i\}$ with a binary relation in A, which is defined such that the elements are reflexive, symmetric, and transitive we get an *equivalence relation.*

It is denoted using the symbol: $\sim$; for example, $a \sim b$ if a and b belong to the same partition set.

Definition 2.10. A *choice function* is a function that is defined on a collection of nonempty sets and assigns a element of each set A in that collection to A by $f(A)$. That is $f(A)$ maps A to some element of A.

Definition 2.11. A binary relation on a set such that, for certain pairs of elements in the set, one of the elements precedes the other in the ordering, is known as a *partial ordering* of the set.

A partially ordered set is also often referred to as a *poset.*

Definition 2.12. A *chain* is a set all of whose elements are comparable to each other. Note that a subset of chain is also a chain. For example :

The set $\mathbb{Z}$, $\mathbb{Z}+$, and other similar sets are chains, because a well-order can be defined on them.

[3] [Sim63]

[4] Since we're working with the ZF set theory, the notion of class is informal

Definition 2.13. A *filter* on a set X is a collection $\mathcal{F}$ of subsets of $X(i.e. \mathcal{F} \subseteq \mathcal{P}(X))$ satisfying the following criteria:

1. $X \in \mathcal{F}$ and $\emptyset \notin \mathcal{F}$.
2. If $Z \subseteq Y \subseteq X$ and $Z \in \mathcal{F}$, then $Y \in \mathcal{F}$.
3. If $Y, Z \in \mathcal{F}, then Y \cap Z \in \mathcal{F}$.

Definition 2.14. An *ultrafilter* on X is a filter on X that contains as many sets as possible. *Note that if $\mathcal{F}$ is a filter on X and $Y \subseteq X$, it cannot be the case that both Y and $X \setminus Y$ are in $\mathcal{F}$.*

This paper also makes use of several abbreviations, here is a table for your reference.

Symbol	**Usage**
$\mathbb{N}$	Natural Numbers
$\mathbb{Z}$	Integers
$\mathbb{Q}$	Rational Numbers
$\mathbb{R}$	Real Numbers
$\mathbb{C}$	Complex Numbers
iff	if and only if
AC	Axiom of Choice
ZF	The Zermelo Frankeal Set Theory
ZFC	The Zermelo Frankeal Set Theory With Choice
$ZF\neg C$	The Zermelo Frankeal Set Theory Without Choice
$\varnothing$	An empty set
A'	Complement of A (any given set)

Additionally, we use several logical symbols in the paper; below attached is a table for your convenience.

Symbol	**Verbose name**	**Usage**	**Remarks**
$\neg$	$negation\, symbol$	$\neg A$	To denote the opposite of a given statement. English: not A.
$\wedge$	$conjunction\, symbol$	$A \wedge B$	To say A and B. English: and.
$\vee$	$disjunction\, symbol$	$A \vee B$	To say A or B. English: or.
$\rightarrow$	$conditional\, symbol$	$A \rightarrow B$	It is used to express a condition (if A, then B) or to denote a function's domain and co-domain.
$\leftrightarrow$	$bi-conditional\, symbol$	$A \leftrightarrow B$	English: if and only if.
$\mapsto$	$`maps\, to'\, symbol$	$A \mapsto B$	Used to describe the input and output of a function.
$\setminus$	$set-minus\, symbol$	$A \setminus B$	To denote that the set the set consists of the elements of A which are not elements of B.
$\cup$	$union\, symbol$	$A \cup B$	Refers to elements belonging to both A and B.
$\cap$	$intersection\, symbol$	$A \cap B$	Refers to elements common to A and B.
$\subseteq$	$subset\, symbol$	$A \subseteq b$	Used to denote that B is a subset of A.
$\subset$	$subset]\, symbol$	$A \subset b$	Used to denote that B is a proper subset of A.
$\in$	$`element\, of'\, symbol$	$a \in A$	Used to denote that a belongs to A.
$\notin$	$`not\, an\, element\, of'\, symbol$	$a \notin A$	Used to denote that a does not belong to A.
$\times$	$Cartesian\, product$	$A \times B$	Refers to the set of all ordered pairs from A and B.

3 Tychonoff's Theorem

[Tör15]

Theorem 3.1 (Tychonoff's Theorem). *The product of any non-empty class of compact spaces is compact in the product topology.*[5]

3.1 Introduction to Topology

We are used to having an intuitive concept of metric as the distance between two entities (or elements). When we transition to the world of topology, we have to abandon such a concept. This section provides an introduction to elementary topology, including basic definitions, concept of a *space*, and some theorems. We begin by defining what topology is.

For the extent of this section, let A be a non-empty set, and T be a topology on A.

Definition 3.2. A class T is a *topology* on A if it meets the following requisites :

(i). A union of every class in T is a set in T. $\bigcup T_i \in T$.

(ii). The intersection of every finite class of sets in T is a set in T.
$(\forall X \in T)(X_1 \cap X_2 \ldots \cap X_n)$, where X is a finite class of sets, and n is the number of such classes.

(iii). A and $\varnothing$ are in T.

Thus, a topology on A is closed under the formation of arbitrary unions and finite intersections.

Definition 3.3. The topology T on set A and set A constitute a *topological space*. It is formally denoted by (A, T), but it is common practice to refer to it as just A.

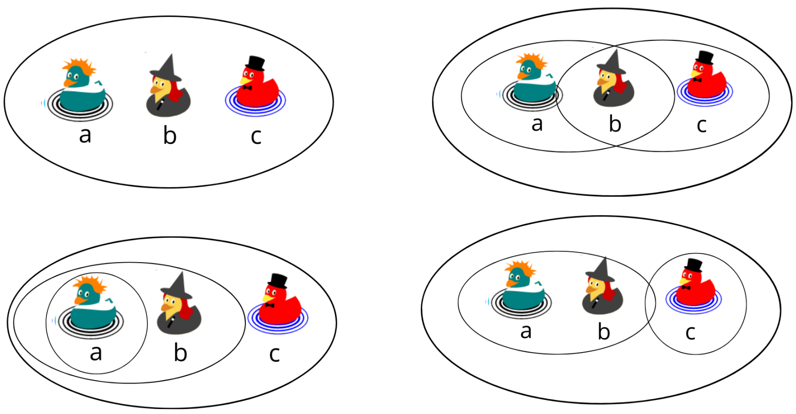


Figure 2. Different topologies for the set A of ducks, where A = {a, b, c}.

Definition 3.4. The sets in T are called *open sets* of the topological space A.

Definition 3.5. A *closed set* in a topological space is one whose complement is open. Here, all sets in T' are closed.

Definition 3.6. When the topology on A is defined as class of all subsets of A, it is known as a *discrete topology*. And a topological space whose topology is a discrete topology is called a *discrete space*.

[5] This must not be confused with box topology, here is an interesting thread to distinguish between the two.

Definition 3.7. Let P and Q be two topological spaces, and f a mapping from P to Q. Then f is *continuous* if $f^{-1}(R)$ is open in P whenever R is open in Q.

Definition 3.8. *Relative topology* on T is defined as the class of all intersections of T with open sets in A, where A is a topological space and T is a non-empty subset of A.

T with its relative topology is known as a subspace of A.

Theorem 3.9. *If T is a topological space, then the following is always true :*

(i). An intersection of closed sets in T is closed.

(ii). A finite union of closed sets in T is closed.

Proof. Let $\{B_i\}$ be a class of closed sets in A. From 3.5 we know that it is equivalent to show that $(\cup_{i/in\mathbb{Z}} B_i)'$, assuming that B' is open.

$$(\bigcap_{i\in\mathbb{Z}} B_i)' = \bigcup_{i\in\mathbb{Z}} B_i'$$
$$\text{if } \bigcup_{i\in\mathbb{Z}} B_i' \text{ is open then } \bigcap_{i\in\mathbb{Z}} B_i \text{ is closed.}$$

Hence we prove *(i)*.

Point *(ii)* can be proven the same way, by using $(\bigcup_{i=1}^{n} B_i)' = \bigcap_{i=1}^{n} B_i'$.

Now, onto defining bases.

Definition 3.10. A *base* for the topology T of a topological space A is a family $\mathfrak{b}$ of open subsets of A such that every open set of the topology is equal to the union of some sub-family of $\mathfrak{b}$.

A base is often also known as an *open base*. The sets in an open base is referred to as *basic open sets*.

Example. The set of all open intervals in the real number line $\mathbb{R}$ is a basis for the topology[6] on $\mathbb{R}$ because every open interval is an open set, and also every open subset of $\mathbb{R}$ can be written as a union of some family of open intervals.

Definition 3.11. An *open subbase* is a class of open subsets of A whose finite intersections form an open base. This open base is called the open base generated by the open subbase, and the sets in an open subbase are called *subbasic open sets*.

Remark 3.12. If X is a non-empty set, and S is an arbitrary class of subsets of X. Then S can serve as an open subbase for a topology on X, in the sense that the class of all unions of finite intersections of sets in S is a topology.

Definition 3.13. A class of closed sets in A is considered as a *closed base* if the class of the complements of aforementioned sets is an open base. Closed subbases can be defined similarly.

Definition 3.14. A class B_i of open subsets of A is said to be an *open cover* of A if $A = \bigcup_i B_i$. If a subclass of an open cover is itself an open cover it is called a *subcover*.

If all sets of an open cover are in a given open base, it is called a *basic open cover*, and if they all lie in some given open subbase it is called a *subbasic open cover*.

Below is a figure from Robin Törnkvist's essay on Tychonoff's theorem and its equivalence with the Axiom of Choice, to easily recap most of the definitions we just discussed.[7]

[6] Euclidean topology to be precise

[7] The notations may differ slightly, but the essence is maintained.

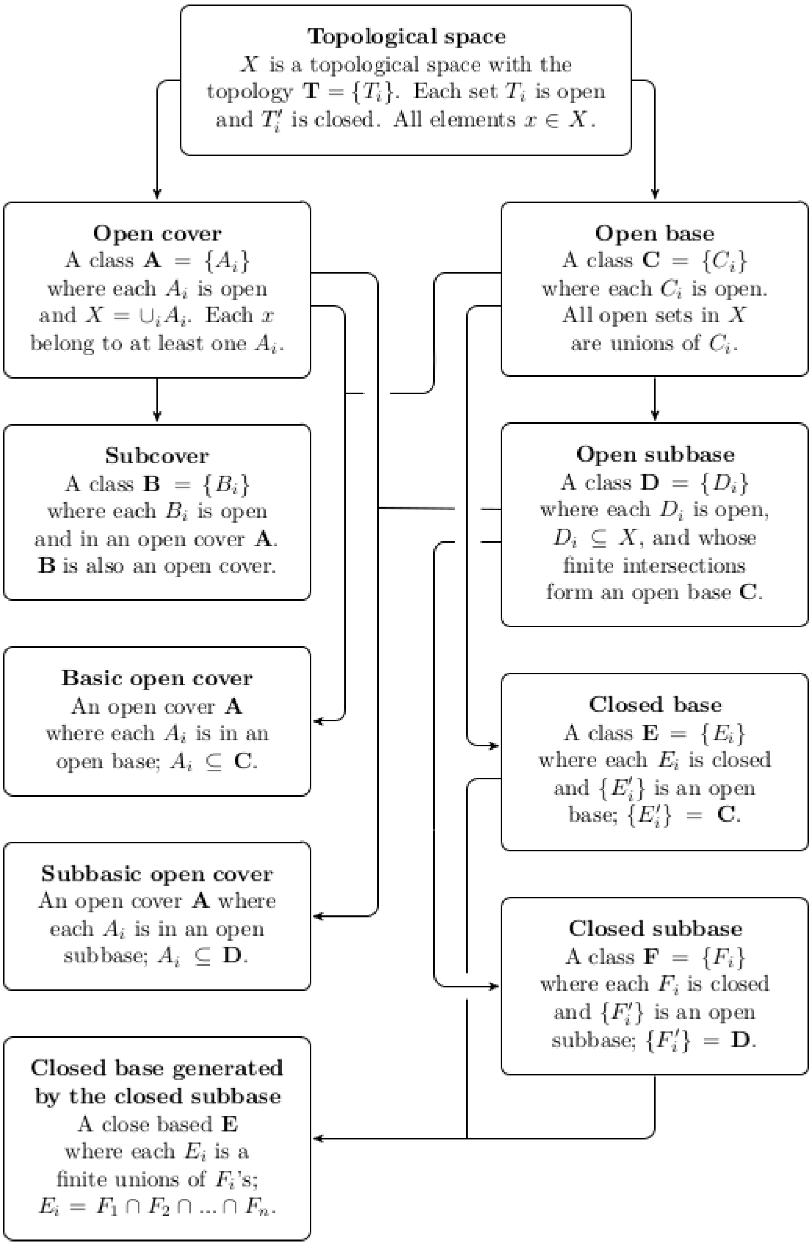


Figure 3. A mind map of basic topological concepts.

3.2 Product Spaces and Compact Spaces

Essentially, this section discusses taking the Cartesian product of topological spaces and then moves on to the concept of compactness.

Definition 3.15. For a class of non-empty sets A_i their *product* is defined to be the set of all ordered pairs $(a_1, a_2, \ldots, a_n)$ where a_i *in* A_i for each $i = \{1, 2, \ldots, n\}$.

Definition 3.16. If $a = (a_1, a_2, \ldots, a_n)$ is a point in the product form then the *mapping* m_i of the product onto its i'th co-ordinate is $m_i(a) = a_i$ and is called the *projection* of a onto the i'th co-ordinate.

Unfortunately this definition isn't generalised enough to be extended to arbitrary classes.

Remark 3.17. The product $(a_1, a_2, \ldots, a_n)$ is essentially a function a with the domain as the index set I, and the restriction as $a(i) = a_i$, an element of the set A_i for each i. Therefore,

Definition 3.18. If each element i in $\{A_i\}$ belongs to I, the product of sets is, $A = M_i A_i = \prod_i A_i$.

Definition 3.19. The product on A_i is,

$$A = M_i A_i$$

The *product topology* on A is defined as the class of all inverse images of open subsets in A_i,

$$R = m_i^{-1}(S_i)$$

where R is the class of all inverse images of open subsets in A_i, i is any index in I, and G_i is any open subset of A_i.

Remark 3.20. It is worth noting that from the previous definition we can understand that the projections m_i are continuous. Additionally, R can be defined as the class of all products of the form $R = M_i S_i$, except when $i = 1$.

Definition 3.21. The class of all finite intersections of R is called *defining open base*.

The class R is called the *defining open subbase* for the product topology, and the class of all complements of sets in R is called the *defining closed subbase*.

The defining closed subbase can also be defined as a product of $M_i U_i$, except when $i = 1$. Here U_i any closed subset of A_i.

Definition 3.22. A topological space A is compact iff every collection of closed subsets of A with the $\mathcal{FIP}$ has a non-empty common intersection.[8]

Informally, it is equivalent to considering A a compact space if every open cover of A has a finite subcover.

Theorem 3.23. *Every closed subspace of a compact space is compact.*

Proof. Let A be a compact space, B be a closed subset of A, G_i be an open cover of B, and H_i be a class of open subsets in A.

Each G_i, which is open in the relative topology of B, is the intersection with B of some H_i. Since B is closed we know that B' is open. by definition, this means that the class composed of B' and all H_i is an open cover of A. Also, since A is compact this open cover has a finite subcover. If B' occurs in this subcover we discard it, and what remains is a finite class of H_i whose union form A. The corresponding G_i then form a finite subcover of the original open cover of B. Thus, every closed subspace of a compact space is compact.

Theorem 3.24. *Any continuous image of a compact space is compact.*

Proof. Let $f : A \mapsto Y$ be a continuous mapping of a compact space A into an arbitrary topological space Y, and $\{G_i\}$ an open cover of $f(A)$. Hence, each G_i is an intersection with $f(A)$ of an open subset H_i of A. The class $\{f^{-1}(H_i)\}$ is an open cover of A since f is continuous and each H_i is open.

Since A is compact this class has a finite subcover. The union of the finite class of H_i of which these are the inverse images must contain $f(A)$, so the class of corresponding G_i is a finite subcover of the original open cover of $f(A)$. This implies that $f(A)$ is compact.

The following are different ways to show that a topological space A, is compact. The scope of this paper does not permit proofs for them, but they are quite intuitive and the reader may look them up on Google.

[8] A collection of subsets of a topological space A is said to have the finite intersection property ($\mathcal{FIP}$) if the intersection of any finite subcollection is non-empty.

1. A topological space A is compact if, and only if, every class of closed sets in A with empty intersections has a finite subclass with empty intersections.

2. A topological space A is compact if every basic open cover of A has a finite subcover.

3. A topological space A is compact if every subbasic open cover of A has a finite subcover, or equivalently, if every class of subbasic closed sets in A with the finite intersection property has non-empty intersections.

3.3 Proof of Tychonoff's Theorem

It is strongly recommended that the readers first familiarize themselves with the notions of the Axiom of Choice (section 4), Zorn's Lemma (section 5), and the definitions of compact spaces before proceeding with the proof.

Restating Theorem 3.1(Tychonoff's Theorem). *The product of any non-empty class of compact spaces is compact in the product topology.*

Proof. Let $\{A_i\}$ be a non-empty class of compact spaces, and form the product $A = M_i A_i$, $\{B_j\}$ be a non-empty subclass of the defining closed subbase for the product topology on A, hence each $\{B_j\}$ is a product of the form $B_j = M_i F_{ij}$, except when $i = 1$. We assume that the class $\{B_j\}$ has $\mathcal{FIP}$.

For a fixed i the class B_{ij} is a class of closed subsets of A_i with $\mathcal{FIP}$, and by the compactness of A_i and by definition 3.22,

$$(\exists a_i \in A_i)(a_i \in \bigcap B_{ij})$$

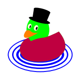
Upon doing this for each i, we obtain a point $a = \{a_i\}$ in A, such that,

$$a \in \bigcap_j B_j.$$

Hence proving,

$$\bigcap_j B_j \neq \varnothing.$$

And by definition of compact spaces[9] we have proved the theorem.



We had started off by talking about De Bruijn–Erdős theorem—a rather picturesque consequence of Tychonoff's Theorem. Tychonoff's theorem has been used to prove many other mathematical theorems; particularly interesting ones include, the Arzelà–Ascoli theorem, and the Curtis–Hedlund–Lyndon theorem.

But we mustn't forget that Tychonoff's theorem is just one of the consequences of Zorn's lemma, which in turn is an equivalence of the Axiom of Choice which is *somewhat* restricted to the mathematical realms of topology and abstract algebra. *However, the star of this paper the Axiom of Choice— a simple and intuitive notion, which has caused ~~friendly~~ math-wars, and bequeathed to us mind-boggling, counter-intuitive results.* So let's try to understand what it is.

[9]The definition refered to here is *'A topological space A is compact if every subbasic open cover of A has a finite subcover, or equivalently, if every class of subbasic closed sets in A with the finite intersection property has non-empty intersections'*, which requires Zorn's Lemma to be proved. Since we have not covered Zorn's lemma yet, we have not proved it.

4 The Axiom of Choice

4.1 Infinity

We have been dealing with infinity since middle school, maybe even earlier; but what exactly is infinity? It is often misconstrued as the largest number that could exist. It is not a number, rather it is a size. It is incomprehensibly big, and is forever expanding. Before delving into the nitty-gritties of the theorem, let's clarify some fundamental concepts about infinity.

- Infinity is a notion that denotes the unknown, unbounded, largest possible value.
- Not all infinities are equal: there are countable infinities and uncountable infinities.
- Countably infinite sets are those which can be put in a one-to-one correspondence with $\mathbb{N}$. Hence, if one can establish a bijection between an infinite set S and $\mathbb{N}$, then S is countably infinite.
- Uncountably Infinite Sets are those whose elements cannot be 'counted' using the same method. The set of all Real Numbers $\mathbb{R}$ is uncountably infinite. This is proven by Georg Cantor's Diagonal Argument.

4.2 Cantor's Diagonalisation Argument

It is impossible to put elements of a real interval in a one-to-one correspondence with $\mathbb{N}$.

Let $f : \mathbb{N} \to [1, 2]$ (whose Cardinality is same as that of $\mathbb{R}$).

Make a table of values of f, where the nth row contains the decimal expansion of $f(n)$. Perhaps $f(1) = \pi/2, f(2) = 37/22, f(3) = 11/7, f(4) = \sqrt{6}/2$, so that the table starts out like this.

n	$f(n)$
1	1.5707963...
2	1.6818181...
3	1.5714285...
4	1.7320508...
.	.
.	.

Even if this table goes on for infinity (the countable one) we cannot possibly list every number in [1,2]. For example, if we take the digits in the main diagonal of the table and add one to each of them, we come up with a new number that isn't in the table.

n	$f(n)$
1	1.**5**707963...
2	1.6**8**18181...
3	1.57**1**4285...
4	1.732**0**508...
.	.
.	.

The highlighted digits are 1.5810... on adding one to each of them we get 1.6921...

This number isn't in the table. In fact, there are many such numbers that aren't in the table. As long as we highlight at least one digit in each row and at most one digit in each column, we can change each the digits to get another number not in the table.

This proves that f cannot possibly be biject-ed to $\mathbb{N}$.

4.3 Zermelo–Fraenkel Set Theory

0. Axiom of Existence

   It tells us that the universe ξ is nonempty. (This axiom is superfluous, and is added only for completeness). It is a logical axiom—something; which is always true.

1. Axiom of Extensionality

   It states that two sets are equal iff they contain the same elements.

$$(\forall x)(\forall y)[x = y \leftrightarrow (\forall z)(z \epsilon x \leftrightarrow z \epsilon y)]$$

2. Axiom of Regularity[10]

   It guarantees that $x \in x, x \in y \in x, x \in y_0 \in \ldots \in y_n \in x$ or even $x_0 \ni x_1 \ni \ldots \ni x_n$ do not happen.

   We postulate that every nonempty set has an $\in$- minimal element. So if A is a set and $a \in A$, then a is an $\in$-minimal element of A iff no elements of a are in A. i.e. $a \cap A = \varnothing$.

$$(\forall x)[(\exists y)(y \in x \wedge (\forall z)(z \in y \rightarrow z \notin x)$$

3. Axiom Schema of Specification[11]

   If a is a set and $P(v)$ is a property expressible in our language, then there is a set $\{x \in a \mid P(a)\}$.

   If $\varphi(x, y_1, \ldots, y_n)$ is a formula $a, a_1, \ldots, a_n$ are sets then $\{z \in a \mid \varphi(z, y_1, \ldots, y_n)\}$ is a set. The schema consists of all formulae of the form:

$$(\forall x)(\forall y_1) \ldots (\forall y_n)(\exists u)(\forall z)[z \in u \leftrightarrow (z \in x \wedge \varphi(z, y_1, \ldots, y_n))]$$

   Note that the schema can only construct subsets, and does not allow the construction of entities of the more general form: $\{x \mid \varnothing(x)\}$.This restriction is necessary to avoid Russell's paradox[12] and its variants.

4. Axiom of Pairing

   For any two sets x, y there is a set whose elements are just x, y, denoted by $\{x, y\}$.

$$(\forall x)(\forall y)(\exists z)(\forall u)[(u \in z \leftrightarrow (u = x \vee u = y)]$$

5. Axiom of Union

   If x is a collection of sets then there is a set S whose elements are precisely all elements of sets in x. This is the union of x, denoted by $\cup_x$.

$$(\forall x)(\exists y)(\forall z)[(z \subseteq y \leftrightarrow (\exists u)(u \in x \wedge z \in u)]$$

6. Axiom Schema of Replacement

   If $F : a \rightarrow V$ is a function defined by a formula and a is a set, then there is a set b such that all values of F are in b. $(\exists! x)\varphi(x)$ is an abbreviation of

$$(\exists! x)\, \varphi\, (x) \wedge (\forall y)(\forall z)(\varphi(y) \wedge \varphi(z) \rightarrow y = z).^{13}$$

[10] Also known as Axiom of Foundation

[11] Also called the Axiom Schema of Separation or of Restricted Comprehension

[12] Here's a short video on it.

[13] $\exists! x$ means "there is exactly one x."

7. Axiom of Infinity

   Let S be a operation such that $S(x) = x \cup \{x\}$.

$$(\exists x)[\varnothing \in x \cap (\forall z)(z \in x \to S(z) \in x)]$$

8. Axiom of Power Set

   For each set x, the collection of all subsets of x is contained in a set. $z \subseteq x$ is an abbreviation for

$$(\forall x)(\exists y)(\forall z)(z \subseteq x \to z \in y)$$

4.4 Well-Ordering Theorem aka Zermelo's Theorem

The well-ordering theorem is the most commonly cited equivalence to the Axiom of Choice, quickly followed by Zorn's lemma.

Theorem 4.1. *Given any set S, there exists a well-order $\prec$ on S .*

A well-order on set S is a relation that allows us to interpret S as an ordinal number α and $\prec$ as the relation on ordinal numbers less than α . $\prec$ on a set S is:

- transitive, extensional, and well-founded.
- a precise linear order with no infinite descending sequence $x_0 \prec x_1 \prec x_2 \prec \dots$
- a total order with the property that every inhabited, non-empty subset U of S has a least element. $(\forall x \in U)(\bot_U \preceq x)$

Note: A well ordered set is also known as a ***woset***.

For the set $\mathbb{N}$ the well-order $\prec$ is the same relation as $<$.

However, the set of integers with our usual ordering on it is not well-ordered, neither is the set of rational numbers, nor the set of all positive rational numbers.[14]

$\mathbb{Z}$ can be well-ordered by the following relation.

For $a, b \in \mathbb{Z}$, $a \prec b$ whenever $|a| < |b|$ or if $|a| = |b|$, wherein a is a negative number and b is a positive number.

According to the well-ordering theorem, every single set admits some $\prec$ satisfying trichotomy and transitivity that makes the set well-ordered.

Cantor was first to idealise the well-ordering theorem. Ernst Zermelo gave a much clearer exposition of his ideas, and eventually went on to show that the well-ordering theorem had many beautiful consequences in mathematics.

4.5 Axiom of Choice: AC

According to AC if A is a collection of nonempty mutually disjoint sets, then we can find a set C that has exactly one element in common with every set from A. So this gives us a choice function: $A \to \bigcup A$ if $f(A) \in A$. This does not have many major implications for finite sets, as the process of manually choosing elements, although tedious, is feasible. Whereas for infinite sets, AC allows us to simultaneously choose an element from infinitely many sets in a collection, *without specifying a definite way to select them.*

It is may also be defined as: for any indexed collection of sets, there exists a choice function.

$$f : X \to \bigcup_{i \in X} S_i$$

[14] A well-order on $\mathbb{R}$ has not yet been established. Moreover, it has also been proven that it is impossible to write down an explicit well-ordering for the set of real numbers.

such that $f(i) \in S_i$ for all $i \in X$.

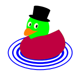
Essentially, this implies the existence of the following sets.

1. Let $A = P(\mathbb{N}) \setminus \varnothing$ then the function $f(A) = min(A)$ is a choice function for A.
2. Let $C = (\forall x)(\forall y)([x, y] \in \mathbb{R}^+)$ such that each interval has a finite length. Then we can define the choice function $f(C)$ to be the midpoint of the interval $[x, y]$.
3. And the most common example, in Bertrand Russell's words "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." That is to say, one can easily choose the left shoe every time, but since there is no rule to distinguish between the members of a pair of socks we need AC.

4.6 Equivalence of the Well-Ordering Theorem to the Axiom of Choice

Proof. To make a choice function for a collection of non-empty sets A let X be the union of sets in A, $X = \bigcup A$. Let R be a well-ordering of X (We know there exists one because we are assuming the well-ordering theorem).

$$R(S) \to A = min(S)$$

where S is a set in A. 

Remark 4.2. R involves a single arbitrary choice; applying the well-ordering theorem to each S of A separately would not work, since the theorem only asserts the existence of a well-ordering, and choosing for each S a well-ordering would require just as many choices as simply choosing an element from each S. Particularly, if A contains uncountably infinite sets, making all uncountably many choices is not allowed under ZF without AC.

On the other hand, the well-ordering theorem states that every set may be equipped with a well-order. This theorem follows from the Axiom of Choice, and is equivalent to it in the presence of excluded middle.

The Well-Ordering Theorem illustrates the non-constructive nature of AC very well; (even familiar sets like $\mathbb{R}$ often do not have an established well-order). This has resulted in people (only a minuscule community of mathematicians to be honest) rejecting it as an axiom; some even treat is as a "powerful-drug—only to be consumed when necessary." As if we provide "consumer-labels" with theorems, such that people can avoid certain math if it isn't delectable for them.

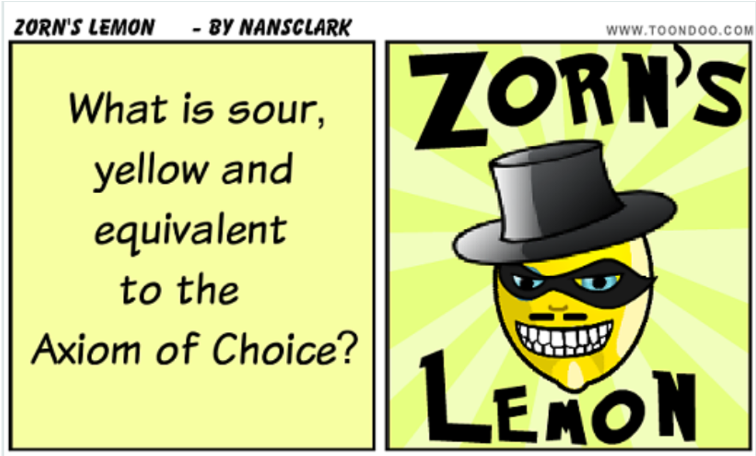
However, not only does the Axiom have interesting, quite counter-intuitive and incredibly beautiful consequences and equivalences, but it's negation has ~~blasphemous~~ ludicrous results.

5 Zorn's Lemma — Another Equivalent Statement

Zorn's lemma is probably the most common equivalent of AC, for that very reason we would not be discussing it in a lot of detail. This is how it is formally defined:

Lemma 5.1 (Zorn's Lemma)**.** *Every non-empty partially ordered set in which every totally ordered subset has an upper bound contains at least one maximal element.*

Definition 5.2. An element x is called maximal if in the set there is no larger element. It is called maximum if it is larger than all the other elements.



5.1 The Axiom of Choice implies Zorn's Lemma

Proof. Let A be a set partially ordered by $<$ such that each chain has an upper bound. Define $p(a) = \{b \in A \mid a < b\} \in P(A)$. Let $p(A) = \{p(a) \mid a \in A\}$. If $p(a) = \varnothing$ then it follows that a is maximal.

Assume $p(a) \neq \varnothing$

Then (AC) there is a choice function f on $p(A)$, and since for each $p(a)$ we have $f(p(a)) \in p(a)$ it follows that $a < f(p(a))$.

We define $f\alpha(p(a))$ for all ordinals α by transfinite induction[15]. And for a limit ordinal α, let $f\alpha(p(a))$ be an upper bound of $f_i(p(a))$ for $i < \alpha$.

This construction will go on forever, for any ordinal. Then we can easily construct an injective function from the ordinal to A as $g(\alpha) = f\alpha(p(a))$ for an arbitrary $a \in A$. This must be injective, since $\alpha < \beta$ implies $g(\alpha) < g(\beta)$. But that requires that A be a proper class,i.e. contradicting the fact that it is a set. Hence there can be no such choice function, and there must be a maximal element of A.

Hence, the Axiom of Choice implies the existence of Zorn's Lemma.

[15] Formally, transfinite induction is an extension of mathematical induction to well-ordered sets, for example to sets of ordinal numbers or cardinal numbers. Its correctness is a theorem of ZFC

6 The Consequences of AC

6.1 The Infinite Hat Paradox

[Dai21]

The Question:

There is an infinite line of people. All of them are wearing hats, the hats have random real numbers on them you could have 3 or 7, or 5.234 or π^e. Each person can see an infinite collection of people in front of them (assume there is a first person in the line, so the line is infinite only in one direction), and cannot hear the answers of the people preceding them. Starting at the back of the line, the host will ask each person to guess the number on their hats. Before the game starts, you all get a few minutes to talk and plan out your strategy. What should you do to get as many correct guesses as possible?
Right now, this puzzle seems plainly impossible. Let's discuss a common and simpler version of the same problem.[16]

6.1.1 A Simpler Version

The Simpler Question:

You and nine other individuals have been captured by super-intelligent alien overlords. The aliens think humans look quite tasty, but their civilization forbids eating highly logical and cooperative beings. Unfortunately, they're not sure whether you qualify, so they decide to give you all a test. Each of you gets a hat to wear, which is either black or white. You can see everyone in front of you, including the colors of their hats; you can't see your own hat, nor can you see anyone behind you. Starting at the back of the line, the host will ask each person to guess whether their own hat is black or white. You'll be able to hear the guesses, and whether they're right or wrong. Before the game starts, you all get a few minutes to talk and plan out your strategy. What should you do to get as many correct guesses as possible?

The Solution:

Let's call the person at the back of the line Xi, they don't have any information so we cannot guarantee if they'll get it right but the other 8 can get it right with certainty.

Xi counts up the number of white hats, w in front of them, if it is even they say "black"; if it is odd they say "white."

Now the second person (say Gee) knows if Xi saw an even or odd number of white hats. But Gee can count up all the black hats they see. If Xi saw an even number of hats, but Gee sees an odd number, that means Gee must be wearing the remaining black hat. This process continues down the line.[17]

Okay so back to our question; we have an infinite number of people and no prior information, so this method would definitely not work. But we can guarantee that infinitely many people get the right answer—better yet, only finitely many people will get it wrong. In our endless infinite line, there will be a last wrong person; all the endless people in front of them will guess right.

This can be done using the μ-stratergy.

6.1.2 The μ-stratergy

[HT08]

Let A be the set of people wearing the hat, S be the set of all possible numbers sequences on said hats (in our case $\mathbb{R}^{\mathbb{N}}$), and $\prec$ be a binary, transitive relation on A.

Definition 6.1. X denotes the set of all $A \to S$. We will call the elements of this set *scenarios*. For each $a \in A$ we define the equivalence relation as $\sim$ on X by $f \sim g$ iff f and g agree on $\prec a = \{s \in A \mid s \prec a\}$, i.e. the set of predecessors of a. $[f]_a$ denotes the equivalence class of f under $\sim$.

We would fix an element $\alpha \in X$ which we consider to be the true scenario.

[16] You can watch a video about the same here

[17] Not related to the topics we are discussing, but parity bits work on the same logic.

For agents s and a, $s \prec a$ means that a can see the number on agent s's hat, and $f \sim g$ means that scenarios f and g are indistinguishable to agent a, such that f and g assign the same number to the hats that a can see. Given a scenario f for the numbers on the hats, $[f]_a$ is the set of scenarios consistent with what a can see.

A strategy for guessing is as follows:

Definition 6.2. Let $\mathbb{O} = \{[\alpha]_a \mid \alpha \in X \, and \, a \in A\}$. The strategy is the function $g : \{\mathbb{O} \to X \mid \forall g([\alpha])_a \in [\alpha]_a\}$.

A well-ordering $\preceq$ on X (established using AC) such that, $\mu : \mathbb{O} \to X$ and $\mu([f]_a)$ is the $\preceq$ least element of $[f]_a$.[18]

We also abbreviate $\mu([\alpha]_a)$ to $\langle\alpha\rangle_a$.

An equivalence relation for this problem would look like $\mathbb{N} \sim \mathbb{R}^{\mathbb{N}}$ such that they're different in only finitely many places.

We'll say that two scenarios are equivalent if they are different in only finitely many places. For example $(1, 2, 3, 4, 5, 6, \ldots)$ and $(17, 2000\pi, -345e, 4, 5, 6, \ldots)$.

Let W_0 be the people who guess the number on their hat incorrectly.

$$W_0 = \{a \in A \mid \langle\alpha\rangle_a(a) \neq \alpha(a)\}.$$

Corollary 6.3. *If $(A, \prec)$ is a strict linear order with no infinite increasing chains, then W_0 is finite.*

Proof. An infinite linear order must have an infinite increasing chain or an infinite decreasing chain; W_0 can have neither.

Now each person (a) can look in front of them discern which equivalent sequence matches the the numbers on the hats in front of them ($[f]_a$) and tell the number for their position from the representative sequence that they had initially chosen.

They might not be right; we have no way to know until the host tells us. But since we're using the same representative sequence, and the sequence is only different from the "true" sequence finitely many times, an infinite number of us will get the answer correctly. And only a finite number will fail (by 6.3). Therefore, a co-finite set of $\mathbb{N}$ guesses correctly.

In our endless infinite line, there will be a *last* wrong person because they cannot see any of the hats and hence cannot discern the sequence.

[18] This strategy may be viewed as a formalization of Occam's Razor if we interpret $f \prec g$ as meaning that f is "simpler" than g in some sense.

6.2 The Banach-Tarski Paradox

This is a paradox that mathematicians have used to tease physicists for ages; true in theory and counter-intuitive in practice. The Banach-Tarski paradox essentially argues (and very well proves) that one can conjure two completely identical objects—impeccable counterfeits—out of one.

It is not really a paradox, in the sense it is not self-contradictory; rather, it is a highly unintuitive theorem. Simply, it states that one can take a 3-dimensional ball, cut it into a finite number of disjoint sets consisting of infinitesimally small pieces (points), perform simple rotations on said sets, and then put them back together to yield two copies of the original ball. By extension it can be shown that one can start with any bounded set with nonempty interior and reassemble it into any other such set of any volume, so that one could, in principle, begin with a pea and end up with a ball as large as the Sun.

Here is an example of a similar paradox in $\mathbb{R}^2$ (which exists without AC).

6.2.1 The Circle Trick

Let C be the unit circle in $\mathbb{R}^2$, r be the segment $(0, 1)$ along the x-axis. (r is merely a radius of C), ρ be a counterclockwise rotation by $\frac{1}{5} rads$ arond the origin, so $p(r)$ is another radius of C, and since 2π is irrational we can safely saw that $p^n(r)$ will never coincide with r again.

Definition 6.4. So we define set $\mathcal{D}$ as

$$\mathcal{D} = \bigcup_{n=0}^{\infty} = \rho^n(r).$$

Remark 6.5. It is worth noting that the union of sets here is disjoint.

The set $C \bigcup \mathcal{D}$ looks like a wheel with an infinite number of spokes on it.

If we now rotate $\mathcal{D}$ clockwise by $\frac{1}{5} rads$ then we get a similar wheel. And because $\rho^n(r) \neq r$ for any n,

$$\rho^{n-1}(r) \neq \rho^{-}1(r)$$

for any n. I.E. the $\rho^{-}1(r)$ doesn't intersect the set $C \bigcup \mathcal{D}$ at all. We have apparently, cut $C \bigcup \mathcal{D}$ into two pieces, moved one of them, and ended up with the original set plus a line ($\rho^{-}1(r)$),

$$C \bigcup \rho^{-}1(\mathcal{D}) = C \bigcap \mathcal{D} \bigcap \rho^{-}1(r).$$

Before we delve into the nitty-gritties of the theorem we should contrast the clean theory of polygonal dissections with dissection into arbitrary sets, and understand some necessary group theory.

6.2.2 Polygonal Dissections

[Kas07]

We have dealt with dissection and reassembly of figures in classical geometry. It is the basis of several important theorems, including the Pythagorean Theorem.[19]

Definition 6.6. We say that two polygons are congruent by dissection if they can be dissected (ignoring boundaries) into the same finite number of polygonal pieces, which can be put into correspondence such that each pair of corresponding pieces are congruent (that is, some isometry of the plane sends one to the other).

Theorem 6.7. *Any two polygons with the same area are congruent by dissection.*

Proof. Congruence by relation is an equivalence relation.

- Reflexivity: A polygon A can be dissected and rearranged to form itself.
- Symmetry: If polygon A can be dissected and rearranged into B, then polygon B can also be dissected and rearranged into A.

[19] Uses the fact that every polygon is congruent by dissection to a square.

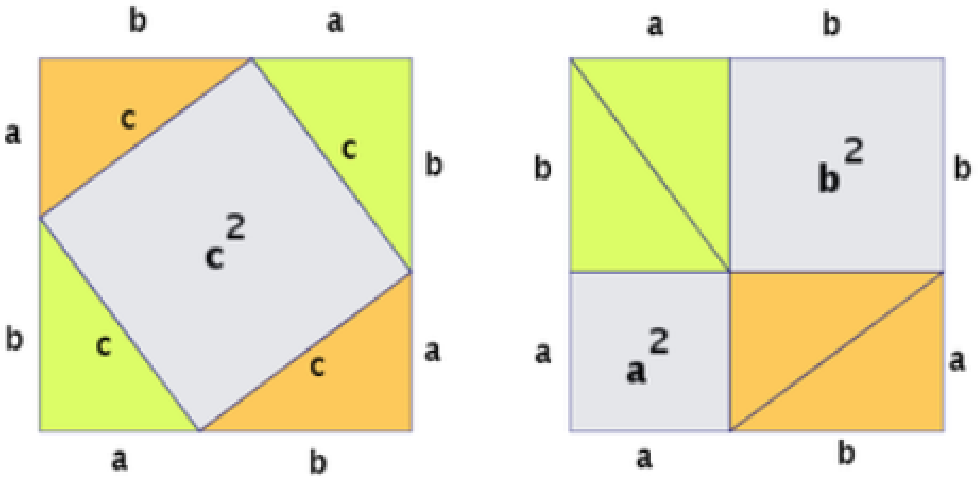


Figure 4. Polygonal Dissection in the Pythagorean Theorem.

- Transitivity: If polygon A can be dissected and rearranged into B, and can also be dissected and rearranged into a different polygon, C, then by making both sets of cuts we end up with pieces that can be arranged to form either B or C.

6.2.3 Equidecomposability

Although it may feel somewhat weird and unstatisfying that boundaries are ignored in the discussion above, considering line segments have zero area it makes sense that they do not matter (here).

However, it has been shown that it is possible to perform these dissections, while taking boundaries into consideration. A few extra, non-polygonal pieces take care of the unwanted segments. An analogue of this idea is called equidecomposition, which can also be defined on arbitrary subsets of $\mathbb{R}^{\mathbb{N}}$.

Definition 6.8. Two subsets $A, B \subseteq \mathbb{R}^{\mathbb{N}}$ are called equidecomposable if they can be partitioned into the same finite number of pieces, which can be matched such that each pair of corresponding pieces are congruent.

6.2.4 Some Group Theory

We generalize the notion of equidecomposability to action of a group G on a set X. (We say G acts on X if each element $g \in G$ corresponds to an action $X \to X$, denoted by $x \mapsto gx$, such that they respect the group laws: $(g_1 g_2)x = g_1(g_2 x)$ and $ex = x$.

Definition 6.9. If G acts on X, we say that $A, B \subseteq X$ are G-equidecomposable if they can be partitioned into the same finite number of pieces, which can be matched such that each pair of corresponding pieces A_i, B_i are related by the action of some $g_i \in G \mid B_i = g_i A_i = \{g_i a | a \in A_i\}$.

An abstract version of the paradox would require us to prove some $A \subset X$ is equidecompsable with $B = g_1 A \cup g_2 A$, where $g_1 A$ and $g_2 A$ are disjoint copies of A.

For instance, we can let $X = \mathbb{Z}$, G equal the group of all bijections $\mathbb{Z} \to \mathbb{Z}$, $A = E \cup O = \{0, 2, 4, \dots\} \cup \{1, 3, 5, \dots\}$, and $B = \mathbb{N} \cup \mathbb{Z}^{+}$. And bijections $E \to \mathbb{N}$ defined by $e \mapsto \frac{e}{2}$, $O \to \mathbb{Z}^{-}$ defined by $o \mapsto -\frac{o+1}{2}$.

Hence, the natural numbers can be put into correspondence with the integers, which to be honest, is not particularly interesting or uncanny. To make it more interesting, let's consider free-groups.

Remark 6.10. If Y is a group, and $S \subset Y$, then a word in S is defined as any expression of the form $s_1^{\varepsilon_1} s_2^{\varepsilon_2} \ldots s_n^{\varepsilon_n}$, where s_n are elements of S, each ε_i is ± 1, and n is the length of the word.[20]

A free-group F_S over a given set S consists of all the words that can be built from members of S[21] (aka generators). The number of generators is the *rank* of the free-group.

Definition 6.11. The free group on k generators $\alpha_1, \ldots, \alpha_k$ is the group of finite words on the symbols $\alpha_1, \ldots, \alpha_k \alpha_1^{-}1, \ldots, \alpha_k^{-}1$ under composition, modulo the equivalences $\alpha_i \alpha_i^{-}1 =, \alpha_i^{-}1 \alpha_i = \epsilon$ (the empty word).

Theorem 6.12. *Consider the group $G = F \times \{1, r\}(r^2 = 1)$ acting on itself, where F is the free group on two generators. Then $F \times \{1\}$ is G-equidecomposable with all of G.*

Less formally, one copy of F is equidecomposable with two copies of F. The most straightforward approach uses the following generalization of the Banach-Schröder-Bernstein Theorem from set theory.

Theorem 6.13 (Banach-Schröder-Bernstein Theorem). *Suppose group G acts on X. If each of $A, B \subseteq X$ is G-equidecomposable with a subset of the other, then A is G-equidecomposable with B.*

Proof. Consider the two given G-equidecompositions as maps $\alpha : A \to \alpha A \subseteq B$ and $\beta : B \to \beta B \subseteq A$. Let $A_1 = \bigcup_{i=0}^{\infty} (\beta\alpha)^i (A \; \beta B)$; since $\beta\alpha A \subseteq \beta B \subseteq A$, we have by induction that $(\beta\alpha)^i A \subseteq A$, so $A_1 \subseteq A$.

Let $A_2 = A \setminus A_1$, $B_1 = \alpha A_1 \subseteq B$, $B_2 = B \backslash B_1$. Observe that $A_1 = (A \setminus \beta B) \cup \beta\alpha A_1$, so $A_2 = \beta B \cap (A \setminus \beta\alpha A_1) = \beta B \setminus \beta\alpha A_1 = \beta B \setminus \beta B_1 = \beta B_2$. We can therefore construct a G-equidecomposition of A and B by cutting A_1 from A_2, then performing α on A_1 and $\beta^{-}1$ on A_2.

Proof of theorem. 6.12

Write F as the free group on α, τ . The identity map is a G-equidecomposition from $F \times 1$ to a subset of G. To construct a G-equidecomposition in the other direction, define $W_(\rho)$ for any symbol ρ to be the set of words (in shortest form) that begin with ρ, and write $G = G_1 \cup G_2 \cup G_3 \cup G_4$, where

$$\begin{array}{ll} G_1 = W(\alpha) \times \{1\} & F_1 = (\epsilon, 1)G_1 = W(\alpha) \times \{1\} \\ G_2 = (F \backslash W(\alpha)) \times \{1\} & F_2 = (\alpha^{-}1, 1)G_2 = W(\alpha^{-}1) \times \{1\} \\ G_3 = W(\tau) \times \{r\} & F_1 = (\epsilon, r)G_3 = W(\tau) \times \{1\} \\ G_4 = (F \backslash W(\tau)) \times \{r\} & F_4 = (\tau^{-}1, 1)G_4 = W(\tau^{-}1) \times \{1\}. \end{array}$$

Then F_1, F_2, F_3, F_4 are disjoint subsets of $F \times 1$. By the Banach-Schröder-Bernstein Theorem, $F \times 1$ is G-equidecomposable with G.

6.2.5 Now to $\mathbb{R}^3$ and Constructing the Paradox

Interestingly the free group with two generators can be realized as a group of rotations of our sphere.

Theorem 6.14. *There exist rotations ϕ, ρ, of the unit sphere in $\mathbb{R}^3$ that are independent; i.e., no nontrivial word on $\phi, \rho, \phi^{-}1, \rho^{-}1$ corresponds to the identity rotation. Consequently, these rotations generate the free group F on two generators.*

Proof. There are several ways to find such rotations. In fact, it is not hard to prove topologically that almost every pair of rotations will do. One constructive solution is to use rotations of angle $\arccos \frac{1}{3}$ about perpendicular axes:

[20] By convention, the identity (unique) element can be represented by the empty word, which is the unique word of length zero.

[21] trivial variations such as $st = suu^{-}t$ for $s, u, t \in S$ are disregarded since they are the same word

$$\phi = \begin{bmatrix} \frac{1}{3} & -\frac{2\sqrt{2}}{3} & 0 \\ \frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \rho = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{2\sqrt{2}}{3} \\ 0 & \frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix}$$

It can now be shown that any word w on $\phi, \rho, \phi^-1, \rho^-1$ sends $(1, 0, 0)$ to a vector of the form $(a, b\sqrt{2}, c)/3^k$, where k is the length of the word w and b is not divisible by 3. In particular, $(1, 0, 0)$ is never sent to itself by a nontrivial word, so the rotations are independent.

We have to show that our ball B^3 is equidecomposable with two unit balls under the isometry of $\mathbb{R}^3$. We well start by doing this for a spherical shell S^2 and then we can apply the same method to all the shells of the sphere, where radii r is $0 < r < 1$.

In this manner we can transform all points except the center.

Lemma 6.15. *B^3 is equidecomposable with $B^3 \setminus \{\beta\}$ where β is the center of the ball.*

Proof. Consider a circle that passes through β and is contained within B^3. Let ρ be a $1rad$ rotation of the circle, then the points $\beta, \rho\beta, \rho^2\beta, \rho^3\beta, \ldots$ are distinct and we get the same set except β upon applying ρ on them.

This yields an analogous to the 2-piece decomposition of the B^3 with $B^3\backslash 0$: one piece being $\{\rho\beta, \rho^2\beta, \rho^3\beta, \ldots\}$,[22] and the being the rest of the ball to itself.

Transformation of such a spherical shell involves several (countably many) troublesome points. We need to remove them.

Lemma 6.16. *Let D be the set of points on the sphere S^2 that are fixed by some non-trivial element of F. Then S^2 is equidecomposable with $S^2 \setminus D$.*

Proof. We can generalise the concept in the previous lemma.

D is countable, because F is a countable set of words, and each nontrivial rotation of F fixes only the two points of S^2 on the rotation's axis. Therefore, there is an axis through the center that does not pass through any point of D. Furthermore, there exists a rotation ρ about this axis such that $D, \rho D, \rho^2 D, \rho^3 D, \ldots$ are disjoint, because there are at most countably many rotations ρ about this axis that satisfy $\rho^i d = d'$ for some $i \in \mathbb{Z}^+, d, d' \in\in D$. If we apply this rotation to just the set $D \cup \rho D \cup \rho^2 D \cup \rho^3 D \cup \ldots$, the result is that D disappears. So as in the last proof, we get a two-piece equidecomposition of S^2 with $S^2 \setminus D$.

Now to duplicate $S^2 \setminus D$.

Lemma 6.17 (***AC***)**.** *$S^2 \setminus D$ can be partitioned into two sets, both being equidecomposable with $S^2 \setminus D$.*

Proof.

Remark 6.18. Distinct rotations $f_1, f_2 \in f$ send any point $x \in S^2 \setminus D$ to different images, because if $f_1 x = f_2 x$, then $(f_1 f_2^{-1})x = x$, which contradicts $x \notin D$.

The F-orbit of a point $x \in S^2 \setminus D$ is the set$Fx = \{fx \mid f \in F\}$ of all such images of x. These F-orbits partition $S^2 \setminus D$, because they are equivalence classes under the equivalence relation$x \sim y \Leftrightarrow y \in Fx$.

By the Axiom of Choice, there is a set M consisting of exactly one member from each F -orbit. Then every point $x \in S^2 \setminus D$ can be written uniquely as $x = fm$ for some $f \in F, m \in M$. So the sets $fM(f \in F)$ also partition $S^2 \setminus D$.

By 6.12, F is equidecomposable with two copies of itself, i.e.it can be partitioned into two subsets F_1, F_2 that are each F -equidecomposable with F.

[22] from $\{\beta, \rho\beta, \rho^2\beta, \rho^3\beta, \ldots\}$ under ρ

We should look at the actions that these equidecompositions perform on the points f, and by logic, apply equivalent actions to the sets fM .

Let $\phi_i : F_i \to F$ be F -equidecompositions ($i = 1, 2$), which we think of as maps $f \mapsto \phi_i f$.

The sets F_1M, F_2M partition $FM = S^2 \setminus D$, and we claim that there are equidecompositions $\nu_i : F_iM \to FM = S^2 \setminus D$ given by $f \mapsto (\phi_i f)m$.

To see that ν_i is actually an equidecomposition, let $A_{ik} \subseteq F_i$ be a piece that goes to $B_{ik} \subseteq F$ under the F-equidecomposition ϕ_i, by the action of $\phi_{ik} \in F$. Then there is a corresponding piece of ν_i : the action of ϕ_{ik} sends the piece $A_{ik}M$ to $B_{ik}M$

Finally, we have the Banach-Tarski Paradox.

Theorem 6.19 (Banach-Tarski Paradox)**.** *(AC). The unit ball B^3 is equidecomposable with two copies of itself.*

Proof. Upon combining lemmas 6.16, 6.17, and 6.16 again, we realize that S^2 is equidecomposable with two copies of itself. Scaling this construction about the center allows us to do this for the spherical shells of radius r, for any $0 < r < 1$, with the same rotations. If we do this for all shells simultaneously, we get an equidecomposition of $B^3 \backslash \{\beta\}$ with two copies of itself.

Finally, using 6.15 before and after this, we see that B^3 is equidecomposable with two copies of itself.

Corollary 6.20 (Banach-Tarski Paradox, Strong Form)**.** *(AC). If $A, B \subset \mathbb{R}^3$ are any bounded sets with nonempty interior, then A and B are equidecomposable.*

Proof. Using the Banach-Schröder-Bernstein Theorem, it is sufficient to prove that a subset of A is equidecomposable with B, and vice versa.

Let L be an open ball in A. By 6.19 and transfinite induction, L is equidecomposable with any finite number of copies of L. Since B is bounded, we can position finitely many copies of L such that they cover B. Then B is the disjoint union of subsets of these copies of$L$, hence is equidecomposable with the pre-image in A of these subsets under the copying procedure.

At first what appeared to be a direct contradiction to both our intuition and measure theory, we now realise is definitely not the latter. Because the pieces are infinitesimally small; *they have degenerate volume.* Or, to be more precise they are not measurable. Or, to be even more precise they are not Lebesgue measurable.[23] In fact, any function that satisfies the properties of Lebesgue measure cannot be defined on these pieces.

[23] The concept of Lebesgue measures is explained in the appendix, section 10.1.

It gets more eerie when we translate it to probability theory.

Let A be one of the infinitesimally small, non-measurable pieces of B^3. If we pick a random point in B^3, clearly either it is in A or it is not. Intuitively, we should be able to assign a probability to the event of the point being in A, and this probability should be invariant under rotations of the ball. However, if we tried to add up the probabilities assigned to all the pieces, before and after the rotations, we would get a contradiction. We are forced to admit the existence of an event with undefined probability.

But well, at least now you are a part of the cool group of mathematicians who understand this joke :)

6.3 De Bruijn–Erdős Theorem

We started the paper with this theorem, and now as we near its end, it's time to prove it. But before that, some definitions:

Definition 6.21. Given a graph G, a sub-graph H of G consists of some of the vertices of G and some of the edges of G that connect those vertices.

Definition 6.22. A proper coloring is an assignment of colors to the vertices of a graph so that no two adjacent vertices have the same color.

Definition 6.23. The chromatic number of a graph is the minimum number of colors in a proper coloring of that graph

The De Bruijn–Erdős theorem is a beautiful, compactness theorem to calculate what the least chromatic number of an infinite graph can be.

Theorem 6.24. *Suppose $G = (V, E)$ is a graph, d is a natural number, and, for every finite subgraph H of G, $\chi(H) \leq d$. Then $\chi(G) \leq d$.*

Proof. Let X be a collection of all finite subsets of V.

$$X = \{W \subseteq V \mid W \text{ is finite}\},$$

where $(\forall W \in X)(X_W = \{U \in X \mid W \subseteq U\}$.

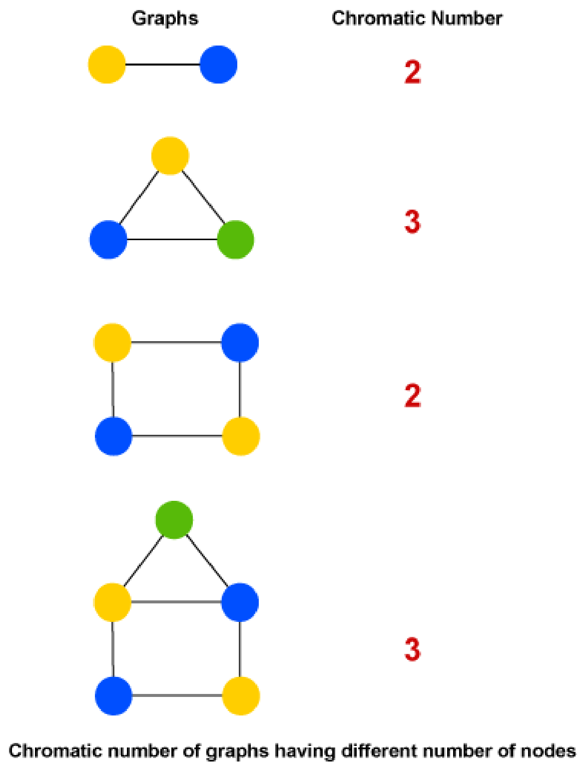


Figure 5. Some graphs with different chromatic numbers.

Let $\mathcal{F}$ be a filter on X for some $W \in X, X_W \subseteq Y$. Now, applying AC we can find an ultrafilter $\mathcal{G}$ on X such that $\mathcal{F} \subseteq \mathcal{G}$. For every $W \in X$, we can find a chromatic coloring $f_W : W \to \{0, 1, \ldots, d-1\}$.

Note: $f_W : W \to d, (\forall v, w \in W)$*, if* $\{v, w\} \in E$*, then* $f_W(v) \neq f_W(w)$*.*

For $w \in V, i < d$, let

$$X_{w,i} = \{W \in X \mid w \in W \text{ and } f_W(w) = i\}.$$

$\mathcal{G}$ as an ultrafilter ensures that $i_w < d$ is unique such that $X_{w,i_w} \in \mathcal{G}$.

Now to define a function $f : V \to d$ by letting $f(w) = i_w forallw \in V$.

Claim. *f is a chromatic coloring. $(\forall \{u, v\} \in E)$, we have $f(u) \neq f(v)$.*

Proof. Fix $\{u, v\} \in E$. $\mathcal{G}$ being an ultrafilter, we can find a set W in $X_{u,i_u} \cap X_{v,i_v}$.

Now we must have $u, v \in W, f_W(u) = i_u$, and $f_W(v) = i_v$. Since f_W is a chromatic coloring and $\{u, v\} \in E$, it follows that $i_u \neq i_v$ and hence $f(u) \neq f(v)$.

[LH17]

This proves both the claim and the theorem.

Is it possible to replace the finite number d in the statement of the De Bruijn-Erdős theorem by an infinite cardinal? Recent work by Chris Lambie-Hanson says, the answer to that question is "a resounding no."

However, the existence of large cardinals does yield infinitary versions of the De Bruijn-Erdős theorem, such as the following:

Theorem 6.25. *Suppose κ is a strongly compact cardinal, G is a graph, $\mu < \kappa$, and, for all subgraphs H of G of size less than κ, we have $\chi(H) \leq \mu$. Then $\chi(G) \leq \mu$.*

6.4 Spanning Trees

Another interesting and somewhat visually-appealing result of AC is the fact that every connected graph has a spanning tree.

Definition 6.26. A connected graph is graph that is connected in the sense of a topological space, i.e., there is a path from any point to any other point in the graph. Hence, null graph and singleton graph are connected graphs.

A graph that is not connected is said to be disconnected.

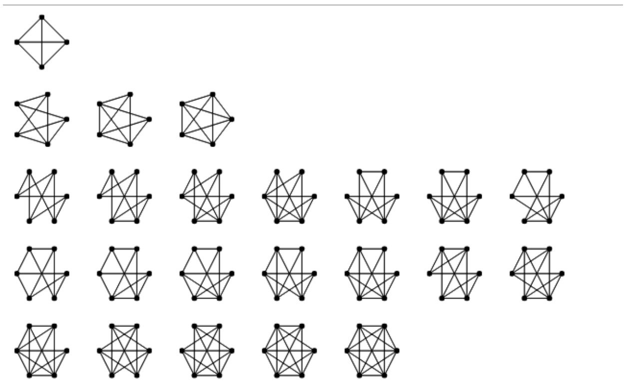


Figure 6. A few examples of connected graphs.

Definition 6.27. A spanning tree is a subset of Graph G, which has all the vertices covered with minimum possible number of edges.

Hence, a spanning tree does not have cycles and it cannot be disconnected.

For connected graphs, spanning trees can be defined either as the minimal set of edges that connect all vertices or as the maximal set of edges that contains no cycle.

Back to the claim that, *every connected graph has a spanning tree.* This can be practically constructed like so, let G be a connected graph,

1. if G has zero edges, it is already a tree.
2. if G has $1 \leq m$ edges
   (a) if G contains a cycle, remove one edge of this cycle, the resulting graph would still be connected.
   (b) otherwise, G is a tree, hence it is it's own spanning tree.

7 Looking at the World without Choice

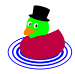
The negation of the Axiom of Choice has several weird, mind-boggling, and somewhat barmy consequences. The Zermelo-Fraenkel Set Theory plus the Negation of AC is denoted by $ZF\neg C$.

The following models of $ZF\neg C$ are also models of ZF, so for each of them there exists a model of ZF where that statement is true.[24]

1. The Cartesian product of two non-empty sets is empty. Let X be a collection of non-empty sets, and P be a set in X. Intuitively, we know that

   $$(\forall X, \varnothing \notin X \Rightarrow \exists f : X \to \bigcup X \; \forall P \in X(f(P)) \in P)].$$

   But if we consider the negation of AC, there exists no choice function such as f. Because of which a Cartesian product cannot be constructed, and is empty. $\forall p \in P$.

   Proof. $(p \to q \Leftrightarrow \neg[p \wedge (\neg q)]$

2. In all models of $ZF\neg C$ there is a vector space with no basis.

   Andreas Blass in his prove that every vector space has a basis used the Axiom of Multiple Choice(AMC)[25], an equivalent statement to AC and defines a vector space using this a family of non-empty sets and by the existence of a basis he constructs a function F. But if we assume the negation of AC, then AMC fails in this model. Therefore there is a family of sets each containing at least two elements, but there is no function, as required. Using this family we can construct the same vector space, but now we can prove that it has no basis.

3. In some model, there is a vector space with two bases of different Cardinalities.

4. In some model, the real numbers are a countable union of countable sets.[26]

5. There exists a model in which every set in $\mathbb{R}^{\mathbb{N}}$ is measurable. Hence, it would be possible to exclude the Banach-Tarski Paradox in that model.[27]

6. In all models of $ZF\neg C$, the generalized continuum hypothesis does not hold.

   The generalized continuum hypothesis (GCH) states that if an infinite set's Cardinality lies between that of an infinite set A and that of the power set $P(A)$, then it has the same Cardinality as either A or $P(A)$. I.e. for any infinite cardinal α there is no cardinal β such that $\alpha < \beta < 2^{\alpha}$.

   GCH is also independent of ZFC, but Sierpiński proved that ZF + GCH implies the AC. Hence, there are no models of ZF in which GCH holds and AC fails.

[24] This paper would not be discussing all of the results in detail.

[25] The axiom of multiple choice asserts that given a family A of non-empty sets, such that every $B \in A$ has at least two elements, then there is a function F such that $F(B)$ is a non-empty, finite, proper subset of B for all $B \in A$.

[26] This does not imply that the real numbers are countable: To show that a countable union of countable sets is itself countable requires the Axiom of countable choice.

[27] This is also possible whilst assuming the Axiom of dependent choice, which is weaker than AC but sufficient to develop most of real analysis.

8 Moving On

The Axiom of Choice is independent of the other ZF axioms, it has weird consequences, and it's negation has absurd results. I wrote a 30-page paper discussing it, and all we can take away that it is decidedly non-constructive in nature. Then why do we care so much about it?

For pretty much the same reason we care about infinity, or complex numbers, or even mathematics. Okay the last part is pretty controversial. But in essence all of these are made up, but they are also tremendously useful.

In Jay Daigle's words, *"Just Relax, if we're trying to model the world, any infinite set we have to deal with will be a limit of finite sets. And any infinite family of infinite sets will be a limit of finite families of finite sets. And we know we have choice for finite sets of finite sets. So we can always get choice for these specific infinite sets, if we really need it—-just by taking the limit of the elements we chose from our finite families.*

What the Axiom of Choice says is: don't worry about it. You don't have to explain how your family of sets came from a finite family. You don't have to explain how you're choosing elements. We'll just assume you can make it work somehow.

That's what axioms are for. They tell us what we want to just assume we can do, without really explaining how. Our axioms are a list of things we don't want to have to think about. And in practice, we don't have to think about whether we can make choices. Any time it really matters, we can."

The Axiom of Choice makes life easier, and more interesting. It's consequences are beautiful and IT MAKES SENSE; that's why so many mathematicians accept it. But if it seems like hogwash to you, that is okay too.

9 Bibliography

References

[Dai21] Jay Daigle. What is the axiom of choice?, 2021.

[HT08] Christopher S Hardin and Alan D Taylor. A peculiar connection between the axiom of choice and predicting the future. *The American Mathematical Monthly*, 115(2):91–96, 2008.

[Kas07] Anders Kaseorg. The banach-tarski paradox. 2007.

[LH17] Chris Lambie-Hanson. 3 thoughts on "ultrafilters viii: Chromatic compactness", 2017.

[Sim63] George F Simmons. *Introduction to topology and modern analysis*, volume 44. Tokyo, 1963.

[Tör15] Robin Törnkvist. Tychonoff's theorem and its equivalence with the axiom of choice, 2015.

10 Appendix

10.1 Lebesgue Measures

Definition 10.1. For any subset $A \subseteq \mathbb{R}^2$, define $\lambda(A)$ to be the infimum of all $k \in [0, \infty]$ such that A can covered by a countable set of hypercubes with total measure k. Then we say that $B \subseteq \mathbb{R}^2$ is Lebesgue measurable if

$$\lambda(A) = \lambda(B \cap A) + \lambda(B \setminus A)$$

for all $A \subseteq \mathbb{R}^2$; then its Lebesgue measure is defined to be $\lambda(A)$.

All polygons, circles, a wide variety of other figures, and that line segments have measure 0 in the plane are Lebesgue measurable.

However,the Banach-Tarski Paradox is a striking demonstration of this shortcoming of Lebesgue measure, or any measure that purports to satisfy the properties above.

10.2 The Generalized Continuum Hypothesis and Gödel's Constructible Universe

The continuum hypothesis talks about the Cardinality of infinite sets. The continuum hypothesis was advanced by Georg Cantor in 1878, and establishing its truth or falsehood is the first of Hilbert's 23 problems presented in 1900 CH states: *There is no set whose Cardinality is strictly between that of the integers and the real numbers.*

In ZFC, this is equivalent to the following equation in aleph numbers: $2^{\aleph_0} = \aleph_1$.

The Generalised Continuum hypothesis on the other hand, states that f, an infinite set's Cardinality lies between that of an infinite set S and that of the power set $P(S)$ of S, then it has the same Cardinality as either S or $P(S)$.

Gödel's Constructible Universe is a something Gödel conceptulized to create a model of the ZF set theory in which the continuum hypothesis holds. Here is a resource that succinctly explains the topic.

This is an interesting, awesome development in the same field. Definitely an article worth checking out.