Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

The Axiom of Choice and its implications in mathematics

Autor: Gina Garcia Tarrach

Director: Dr. Joan Bagaria

Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 29 de juny de 2017

Abstract

The Axiom of Choice is an axiom of set theory which states that, given a collection of non-empty sets, it is possible to choose an element out of each set of the collection. The implications of the acceptance of the Axiom are many, some of them essential to the development of contemporary mathematics. In this work, we give a basic presentation of the Axiom and its consequences: we study the Axiom of Choice as well as some of its equivalent forms such as the Well Ordering Theorem and Zorn's Lemma, some weaker choice principles, the implications of the Axiom in different fields of mathematics, some paradoxical results implied by it, and its role within the Zermelo-Fraenkel axiomatic theory.

Contents

Introduction

The Axiom of Choice states that, given a family X of non-empty sets, there exists a function f such that for all $A \in X$, $f(A) \in A$. In other words, it says that given a collection of non-empty sets, it is possible to choose an element out of each set in the collection. This may seem quite obvious, in fact, one can prove its truth when dealing with a finite collection of non-empty sets. However, if we want this principle to hold for all families of non-empty sets, and not just the finite ones, we have to postulate it as an axiom.

The role that the Axiom of Choice plays in contemporary mathematics is far from negligible: an important number of theorems and propositions that nowadays are regarded as essential for the development and study of different branches of mathematics depend on this axiom or are equivalent to it. It is not an overstatement to say that without the Axiom of Choice, contemporary mathematics would be very different as we know it today. However, the acceptance of the Axiom also leads to some counter-intuitive results. Although today almost all mathematicians accept the Axiom of Choice as a valid principle and use it whenever necessary to prove new results, the first explicit mention of the Axiom on Zermelo's proof of the Well Ordering Theorem was not free of controversy. The Axiom implies the existence of some mathematical objects that cannot be explicitly defined and this conflicts with a constructivist view of mathematics, which states that it is necessary to define or construct a mathematical object to prove it exists. Furthermore, accepting the Axiom leads to the Hausdorff Paradox or the Banach-Tarski Paradox which, although not being literally paradoxes (for they can be proved), may conflict with our intuition.

The aim of this work is to explore the Axiom of Choice and the implications and consequences of its acceptance in mathematics. In the introductory chapter we will give some important definitions and propositions regarding ordinals, cardinals and well-orders, for they will be relevant for some of the proofs given later on. In the next section, we will give a brief insight on the historical background of the Axiom: when did it first appear, what was the original controversy, how it was progressively accepted, etc. In chapter 2, we will give a formal definition of the Axiom and state some of its most important equivalent forms. In chapter 3, we will see some weaker choice principles and how they follow from the Axiom of Choice. In chapter 4, we will explore the consequences and implications of the Axiom of Choice in different branches of mathematics, namely set theory, algebra, analysis, topology, graph theory, and logic. In chapter 5, we will study those 'paradoxical' implications that follow from the acceptance of the Axiom. Finally, in chapter 6 we will see the relation of the Axiom with the other axioms of Zermelo-Fraenkel set theory, as well as a possible substitute for it, the Axiom of Determinacy.

Acknowledgments

I would like to thank Joan Bagaria, who has helped and oriented me during the process of understanding, reading, and writing about the Axiom of Choice and the universe that surrounds it.

I would also like to thank my brother, also a maths student, who has discussed with me many aspects of this work and who has 'most of the time' patiently listened to me and given me good advice.

0 Some preliminary notes on well-orders, ordinal and cardinal numbers

In this introductory chapter we will give some basic definitions and propositions regarding well-orders, ordinal numbers and cardinal numbers, that will be relevant for some of the proofs given during this work.

Definition 0.1. A *partial order* over a set non-empty A is a binary relation $\leq$ on A that satisfies:

(a) $\leq$ is reflexive, i.e., for all $a \in A$, $a \leq a$.

(b) $\leq$ is antisymmetric, i.e., for all $a, b \in A$, if $a \leq b$ and $b \leq a$ then $a = b$.

(c) $\leq$ is transitive, i.e., for all $a, b, c \in A$, if $a \leq b$ and $b \leq c$ then $a \leq c$.

The pair $(A, \leq)$ is called a *partially ordered set* or a *partial ordering.*

Definition 0.2. A *strict ordering* over a non-empty set A is a binary relation $<$ on A that satisfies:

(a) $<$ is asymmetric, i.e., for all $a, b \in A$, if $a < b$ then it is not the case that $b < a$. In other words, $a < b$ and $b < a$ cannot both be true.

(b) $<$ is transitive, i.e., for all $a, b, c \in A$, if $a \leq b$ and $b \leq c$ then $a \leq c$.

The following definitions make also sense for strict orderings.

Definition 0.3. Let $(A, \leq)$ be a partially ordered set. Two elements $a, b \in A$ are *comparable* in the ordening $\leq$ if either $a \leq b$ or $b \leq a$.

Definition 0.4. An ordering $\leq$ of A is called a *lineal ordering* or a *total ordering* if any two elements of A are comparable. The pair $(A, \leq)$ is then called a *linearly ordered set.*

Definition 0.5. Let $\leq$ be a partial ordering over A, and let $B \subseteq A$. An element $b \in B$ is the *least element* of B in the ordering $\leq$ if $b \leq x$ for every $x \in B$.

Definition 0.6. A set W is *well ordered* by the relation $\leq$ if

(a) $(W, \leq)$ is a linearly ordered set.

(b) Every nonempty subset of W has a least element.

Definition 0.7. A set T is *transitive* if every element of T is a subset of T.

Definition 0.8. A set α is an *ordinal number* if

(a) α is transitive.

(b) α is well-ordered by $\in_\alpha$ (the membership relation restricted to α).

An ordinal can be equivalently defined as a well-ordered set $(\alpha, <)$ such that $\beta = \{x \in \alpha \mid x < \beta\}$ for every $\beta \in \alpha$.

Theorem 0.1. *Every well-ordered set is isomorphic to a unique ordinal number.*
Proof. See [1], page 111.

Definition 0.9. If W is a well-ordered set, then the *order type* of W is the unique ordinal number isomorphic to W.

Lemma 0.1. *If α is an ordinal, then $\alpha \cup \{\alpha\}$ is an ordinal.*
Proof. If α is transitive and totally ordered by $\in$, so too is $\alpha \cup \{\alpha\}$. $\square$

Definition 0.10. An ordinal number α is called a *successor ordinal* if there exists an ordinal β such that $\alpha = \beta + 1 := \beta \cup \{\beta\}$. Otherwise, it is called a *limit ordinal.*

Definition 0.11. Two sets A and B are *equipotent* or have *the same cardinality* if there exists a bijection $f : A \to B$.

Definition 0.12. An ordinal number α is called an *initial ordinal* if it is not equipotent to any $\beta < \alpha$.

Theorem 0.2. *Each well-ordered set A is equipotent to a unique initial ordinal number.*
Proof. See [1], page 130

In Zermelo-Fraenkel set theory, the natural numbers are the finite ordinals, and $\mathbb{N}$ is identified with the first infinite ordinal, namely ω. The natural numbers and ω are initial ordinals. The next initial ordinals are ω_1, ω_2, ω_3, and so on.

Definition 0.13. If A is a well-ordered set, then the *cardinal number* of A, denoted $|A|$, is the unique initial ordinal equipotent to A.

In particular, $|A| = \omega$ for any infinite countable set A and $|A| = n$ for any finite set of n elements.

Theorem 0.3. *There are arbitrarily large initial ordinals.*
Proof. See [1], pages 130-131.

Definition 0.14. If α is an ordinal, an α-indexed sequence of the elements of a set A is a function from α to A. An α-indexed sequence is called *transfinite* if $\alpha \geq \omega$

Definition 0.15. Let ω_α be a transfinite sequence of infinite initial ordinal numbers where α ranges over all ordinal numbers. Infinite initial ordinals are, by definition, the cardinalities of infinite well-ordered sets, and are thus the infinite cardinal numbers. This cardinal numbers are called *alephs* and are defined as follows:

$$\aleph_0 = \omega$$
$$\aleph_\alpha = \omega_\alpha, \text{ for each } \alpha$$

1 Historical background

The first appearance of the Axiom of Choice, in a letter sent by Ernst Zermelo to David Hilbert in 1904, was surrounded by controversy. In it, Zermelo gave a proof of the Well-Ordering Theorem, which states that every set can be well-ordered, using what he first called the 'choice principle' (and would later be known as the Axiom of Choice). This result didn't go unnoticed amongst the mathematical community. The acceptance of this axiom led to significant consequences and even questioned the concept of existence in terms of mathematical objects. Most mathematicians of the time were not willing to accept what the axiom implied (even though some of them had implicitly used it in the past) and it wasn't until years later that it became widely accepted and recognised.

The question of whether any set can be well-ordered had been on the table for some years prior to Zermelo's publication. In 1882 Cantor introduced the notion of a well-ordering of a set and formulated the Well-Ordering Principle, which he believed was a *law of thought* that didn't require any proof. In order to 'label' the elements of a well-ordered set he brought in the notion of *ordinal numbers*, which extended the naturals in a simple way. Cantor then defined the second number-class as the set of all ordinals representing well-orderings of the natural numbers, which had to be of greater cardinality than the set of all natural numbers. The set of all ordinals representing well-orderings of the second number-class would then be the third number-class and so on. This process gives birth to an infinite sequence of cardinalities, the aleph-sequence: $\aleph_0$ (the set of all naturals), $\aleph_1$ (the second number-class), $\aleph_2$ (the third number-class), ...

In 1878, Cantor formulated the *Continuum Hypothesis*, which states that every infinite subset of $\mathbb{R}$ is either denumerable or has the power of the continuum. In other words, there is no set with cardinality greater than $\aleph_0$ and less than the continuum, that is, $2^{\aleph_0}$, the cardinality of the set of the real numbers. But how are the Continuum Hypothesis and the Well-Ordering Principle related? If every set can be well-ordered, then every set is bijectable with an ordinal and therefore must have an aleph-number representing its size. This means that any cardinality is represented in the aleph-sequence, all infinite cardinalities are comparable, and the Continuum Hypothesis could be rewritten in the form $2^{\aleph_0} = \aleph_1$.

Cantor, however, began to doubt that both the Continuum Hypothesis and the Well-Ordering Principle were true, since he tried unsuccessfully to prove them. Many mathematicians believed that $\mathbb{R}$ could not be well-ordered and therefore that the Well-Ordering Principle was false. Some years later, in 1900, Hilbert brought attention to the problem again, when he stated the Continuum Hypothesis as the first in the list of his famous 23 unsolved problems presented during the Second International Congress of Mathematicians. In 1904, the Hungarian mathematician Julius König claimed to have proven that the power of the continuum was not an aleph and hence that the set of the real numbers could not be well-ordered. Later that year, Hausdorff pointed out a mistake in Bernstein's lemma, which was used by König in his demonstration, invalidating, consequently, his proof. It was that same year that Zermelo published his proof of the Well-Ordering Principle (that became the Well-Ordering Theorem) using the Axiom of Choice as the basis for it.

Zermelo's publication was immediately controversial. Its consequences were not only mathematical, but also philosophical, for it postulated the existence of certain mathematical objects that were not explicitly defined (the axiom explicits no rule by which the choices are made), and therefore questioned the very notion of what a mathematical object is and what does it mean that it exists. The discussion about whether or not the Axiom of Choice and what it implied should be accepted arose a heated debate between constructivist mathematicians and non-constructivist ones. Borel, Baire and Lebesgue, French mathematicians who were on the constructivism side, opposed the axiom; Hilbert, Haussdorf, Hadamard and Keyser accepted the axiom and the proof; Hardy and Poincaré accepted the axiom but questioned the proof. However, none of Zermelo's critics were able to give a precise formulation of what it means to be definable. The axiom also questioned whether the use of infinitely many arbitrary choices was or wasn't a valid method in mathematics, although some of the axiom critics had unconsciously used this procedure in the past. To give just a few examples, Borel used it in his proof of Cantor's result that every infinite set has a denumerable subset, Baire used it to prove the Baire Category Theorem and W.H. Young, another future critic, used it implicitly to demonstrate Cantor's topological theorem that every family of disjoint open intervals on $\mathbb{R}$ is countable.

In 1908, Zermelo published a second proof of the Well-Ordering Theorem with the aim to reply to some of the criticism and clarify several aspects of the previous proof. In it he proposed to use the axiom only if there are no means to avoid it, and in such cases he recommended to state and study the proof's dependance on the axiom. Two weeks later, Zermelo published his famous axiomatic system, which included, of course, the Axiom of Choice. Zermelo's axiomatic system, along with Abraham Fraenkel's subsequent contributions, would later become a solid base for the future developments in set theory, although at first mathematicians were reluctant to accept it. Today, it is known by its abbreviation ZFC, or ZF if we take out the Axiom of Choice.

Regarding the Axiom of Choice's acceptance within the mathematical community, during the following years the abstract approach began to gain ground, mostly due to Hilbert's influence. Mathematicians began to increasingly recognise its implicit uses although some of them still raised objections to its use. During the first half of the century, Steinitz used it for his algebraic work and even made a plea for its acceptance. As modern algebra developed during the following years, the axiom was regarded as a crucial tool for the discipline. In 1918, the axiom found in Sierpiński a strong defender. He analysed the axiom and published a survey of its uses. Not only that, he also encouraged his students to do the same and continue with further study of the uses of the axiom and its equivalent forms. Over the next decades, mathematicians began to use it in topology, analysis, algebra, set theory and mathematical logic. In 1935, Zorn formulated Zorn's Lemma and claimed its equivalence with the Axiom of Choice. In that same decade, Gödel proved its consistency with ZF and, in 1963 Paul Cohen proved its independence by showing that it cannot be deduced from just ZF. Today, we can confirm that without the Axiom of Choice the nature of modern mathematics would be very different and some of the most important results of the last years would not have been possible.

2 The Axiom of Choice and its Equivalent Forms

2.1 The Axiom of Choice

There are many ways to formulate the Axiom of Choice. Accepting the axiom is the same as accepting any of its equivalent forms, some of which are well known and widely used in different areas of mathematics. In this chapter we will see one of the simplest formulations of the axiom, which is probably the most famous one and some of its equivalents, amongst which we find the Well Ordering Theorem and Zorn's Lemma.

In order to properly formulate the axiom in its simplest form, we need the following definition.

Definition 2.1.1. Let X be a collection of non-empty sets. A function f defined on X is called a *choice function* if $f(A) \in A$ for all $A \in X$.

Now, we may enunciate the axiom as follows:

Axiom of Choice (AC). *There exists a choice function for every collection of non-empty sets.*

What the axiom is essentially saying is that, given a collection of non-empty sets it is possible to choose an element out of each set in the collection. This might seem obvious when dealing with a finite collection of sets (in fact, the existence of a choice function for a finite family of sets can be proved without the need of the axiom, i.e., in ZF, as it shows the theorem below) but it is not so when it comes to infinite collections.

Theorem 2.1.1. *There exists a choice function for every finite collection of non-empty sets.*
Proof. See [1], page 139.

So the Axiom of Choice assures us that there exists a choice function for every collection of non-empty sets, a function that allows us to select an element of each set in the collection. However, it does not say anything about how the choice function is defined, that is, *how the choices are made*; it simply postulates its existence. That was the main reason why the introduction of the axiom was controversial, as we have seen in the chapter above. When one accepts the Axiom of Choice, one is in fact accepting that it is valid to use a hypothetical choice function, even in the cases when one is unable to give an explicit example of it.

2.2 The Well Ordering Theorem

The first explicit appearance of the axiom came together with the Well Ordering Theorem, which states the following:

Well Ordering Theorem (WO). *Every set can be well-ordered.*

In the proof of the equivalence between AC and WO we will use the concepts of transfinite recursion and the Hartogs Number of an ordinal.

Definition 2.2.1. For any set A, the *Hartogs number of* A or $h(A)$ is the least ordinal α such that there is no injection from α to A.

Lemma 2.1.1. *The Hartogs number of* A *exists for all* A.
Proof. See [1], pages 130-131.

Theorem 2.2.1. ***The Transfinite Recursion Theorem.*** *Given a class function*[1] $G : V \to V$ *where* V *is the class of all sets, there exists a unique transfinite sequence* $F : Ord \to V$, *where Ord is the class of all ordinals, such that for all ordinals* α, $F(\alpha) = G(F \upharpoonright \alpha)$, *where* $\upharpoonright$ *denotes the restriction of the domain of* F *to all ordinals* β *such that* $\beta < \alpha$.
Proof. See [1], pages 115-116.

[1] A class function is a rule, given by a logical formula, that assigns to each element in the lefthand class an element in the righthand class. It is not a function since its domain and codomain need not be sets, just classes.

Theorem 2.2.2. $AC \iff WO$

Proof. Let's see the left to right implication. Suppose AC holds and let A be a set. We want to well-order A, that is, construct a bijection G between A and some ordinal λ. We will proceed by transfinite recursion. Let a be a set such that $a \notin A$ and let f be the choice function on the set of all non-empty subsets of A. We now define the function G as follows:

$$\text{For all ordinals } \alpha, \qquad G(\alpha) = \begin{cases} f(A - \operatorname{ran}(G \restriction \alpha)) & \text{if } A - \operatorname{ran}(G \restriction \alpha) \neq \emptyset \\ a & \text{otherwise} \end{cases}$$

By transfinite recursion, G is defined for all ordinals α. G is the bijection we were looking for: it lists the elements of A one by one until A gets exhausted: at that point G has value a.

We observe that, for $\alpha < \beta$, if $G(\beta) \neq a$, then $G(\beta) \in A - \operatorname{ran}(G \restriction \beta)$ and $G(\alpha) \in \operatorname{ran}(G \restriction \beta)$ and, therefore, $G(\alpha) \neq G(\beta)$. Also, G gets exhausted at some point $\lambda < h(A)$, where $h(A)$ is the Hartogs number of A. If $G(\alpha) \neq a$ for all $\alpha \leq h(A)$, then G would be a one-to-one mapping of $h(A)$ into A, contradicting the definition of $h(A)$ as the least ordinal which cannot be mapped into A by a one-to-one function. So, let λ be the smallest ordinal number such that $\lambda \in \{\alpha : G(\alpha) = a\}$. G is a bijection between the set A and the ordinal λ and therefore, A is well-ordered and has order type λ.

Now let's see the left to right implication. Assuming WO, we want to see that any family of non-empty sets has a choice function. Let A be a family of non-empty sets and $<$ a well order in $\bigcup A$. We can define a choice function f in $\mathcal{P}(A)$ as follows:

$$f(B) = \begin{cases} \text{the least element of } B \text{ in the well order } < & \text{if } B \neq \emptyset \\ \emptyset & \text{if } B = \emptyset \end{cases}$$

□

As we have seen, the Well Ordering Theorem states that every set can be well-ordered, in particular, $\mathbb{R}$ can be well-ordered. Even if we cannot give an explicit well-order of the reals, by accepting the Axiom of Choice, we accept that there is, in fact, one.

2.3 Zorn's Lemma

Another important algebraic result equivalent to the Axiom of Choice is Zorn's Lemma.

Zorn's Lemma (ZL). *If $(A, <)$ is a partial ordering such that every chain of A has an upper bound, then A has at least a maximal element under $<$.*

Theorem 2.3.1. $AC \iff ZL$

Proof. Let's prove the left to right implication. Suppose AC holds and let $(A, \leq_A)$ be a partially ordered set that satisfies Zorns's Lemma's hypotheses. We will show that A has a maximal element. By Theorem 2.2.2., A is well-ordered, so there exists a bijection G between some ordinal λ and A. Then,

$$A = \{a_\gamma \mid \gamma < \lambda\}, \quad \text{where for each } \gamma < \lambda, \quad a_\gamma = G(\gamma).$$

We now define by recursion the function f: Ord $\to \lambda + 1$ where $f(0) = 0$ and for every ordinal $\alpha > 0$,

$$f(\alpha) = \begin{cases} \text{the least } \zeta \text{ such that } \gamma < \alpha \to a_{f(\gamma)} <_A a_\zeta & \text{if such a } \zeta \text{ exists} \\ \lambda & \text{otherwise} \end{cases}$$

Suppose there is no α such that $f(\alpha) = \lambda$. Then, $X = f[\text{Ord}]$ would be a well-defined subset of X. Since f is one-to-one, it has a well-defined inverse $f^{-1} : X \to \text{Ord}$, which is surjective. By the Axiom of Replacement[2], $f^{-1}[X]$ is a set, which leads to contradiction, because the set of all ordinals is not a set. Thus, we have that $f(\alpha) = \lambda$ at some point α. Then, let α be the least ordinal such that $f(\alpha) = \lambda$. If α were a limit ordinal, then the sequence $\langle a_{f(\gamma)} \mid \gamma < \alpha \rangle$ would be a chain in A with no upper bound, which contradicts the hypothesis of Zorn's Lemma. Therefore, α is a successor ordinal, that is $\alpha = \beta + 1$ for some β. Thus, $a_{f(\beta)}$ is a maximal element of A.

Now let's prove the other implication. Suppose ZL holds and let X be a collection of nonempty sets. We want to see that X has a choice function. Let F be the collection of all functions f such that dom $f \subseteq X$ and $f(A) \in A$ for all $A \in X$. The set F is ordered by

[2] See section 6.1

inclusion $\subseteq$ and if F_0 is a linearly ordered subset of $(F, \subseteq)$, then $f_0 = \bigcup F_0$ is a function such that $f_0 \in F$ and f_0 is an upper bound on F_0 in $(F, \subseteq)$. By Zorn's Lemma, we have that $(F, \subseteq)$ has a maximal element $\bar{f}$. We will see that dom $\bar{f} = X$, and consequently that $\bar{f}$ is the choice function for X. If dom $\bar{f} \neq X$, we could select a subset $A \in X-$ dom $\bar{f}$ and $a \in A$. In that case, $\bar{\bar{f}} = \bar{f} \cup \{(A, a)\} \in F$ and $\bar{f} \subset \bar{\bar{f}}$, contradicting the maximality of f. $\square$

2.4 Other equivalent forms

Some other equivalent forms of the axiom of choice are the following. In section 4 we will prove how the Axiom of Choice implies these results.

- **Tukey's Lemma.** Let X be a collection of nonempty sets. If X has finite character, then X has a maximal element with respect to inclusion $\subseteq$.
- Every infinite set A has the same cardinality as the cartesian product $A \times A$.
- An arbitrary cartesian product of nonempty sets is nonempty.
- Every non-empty set can be given a group structure.
- Every vector space has a basis.
- Every surjective function has a right inverse.
- **Krull's Theorem.** If A is a ring different from the trivial ring, then A has a maximal ideal.
- **Tychonoff's Theorem.** The product of compact topological spaces is compact.
- Every connected graph has a spanning tree.

It is important to point out that when we say that a certain proposition ϕ is equivalent to the axiom of choice what we are saying is that ZF[3] $+\phi$ proves AC and ZFC proves ϕ.

[3]The ZF axiomatic system will be presented in section 6.1

3 Weaker Forms of the Axiom of Choice

There are several statements or choice principles that are not equivalent to the Axiom of Choice, but are implied by it. These weaker forms of the axiom are of great importance: in many proofs the full strength of the Axiom of Choice is not needed, instead the use of one of its weaker forms is enough. Besides, some of this forms avoid some of the 'paradoxes' that the Axiom of Choice implies, for instance, when dealing with the Lebesgue measurability of all sets of real numbers[4]. As we have stated in the Historical Background section, Zermelo himself advised to avoid the use of the Axiom of Choice whenever possible. In this section some of the most important weaker forms of the AC are presented, as well as the proof that they follow from the axiom.

3.1 The Axiom of Dependent Choice

Axiom of Dependent Choice (DC). *For any nonempty set X and any entire*[5] *binary relation R on X, there exists a sequence (x_n) in X such that $x_n R x_{n+1}$ for each $n \in \mathbb{N}$.*

Theorem 3.1.1. $AC \Rightarrow DC$

Proof. Let X be a non empty set and let R be an entire binary relation on X. For each element $x \in X$ we define $R(x)$ as the range of x in R, that is, $R(x) = \{y \in X \mid xRy\}$. As R is an entire relation in X, for all $x \in X$ there exists some $y \in X$ such that xRy, so by assumption $R(x)$ is non-empty.

Now, we consider the indexed family of sets $\langle R(x) \rangle_{x \in X}$. By the Axiom of Choice, there exists a choice function f for this family of nonempty sets that satisfies $f(R(x)) \in R(x)$ for all $x \in X$. For simpler notation, we define now the function $g(x) = f(R(x))$, that in turn satisfies $g(x) \in R(x)$, that is, $xRg(x)$.

[4]See section 5.1

[5]An entire binary relation R on X is a relation R such that for all $a \in X$, there exists some $b \in X$ such that aRb.

We can now define the sequence $\langle x_n\rangle_{n\in\mathbb{N}} = \langle g^n(x)\rangle_{n\in\mathbb{N}}$, where g^n denotes the composition of g with itself n times. This sequence satisfies $x_n R x_{n+1}$ for all $n \in \mathbb{N}$, that is, what the Axiom of Dependent Choice states. $\square$

3.2 The Axiom of Countable Choice

Axiom of Countable Choice or Axiom of Denumerable Choice (CC). *If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a collection of non-empty sets, then there exists a sequence $\langle x_n\rangle_{n\in\mathbb{N}}$ such that $x_n \in X_n$ for all $n \in \mathbb{N}$. Equivalently, any countable collection of nonempty sets has a choice function.*

Theorem 3.2.1. *$DC \Rightarrow CC$*

Proof. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a collection of nonempty sets and $X = \bigsqcup\limits_{n\in\mathbb{N}} X_n = \bigcup\limits_{n\in\mathbb{N}} \{(x,n) \mid x \in X_n\}$ the disjoint union of the elements of the collection. We define the following relation R in X:

$$(x,m)\ R\ (y,n) \Leftrightarrow n = m+1$$

By definition, R is an entire relation and by the Axiom of Dependent Choice, there exists a sequence $\langle y_n\rangle_{n\in N}$ such that $y_n R y_{n+1}$ for all $n \in \mathbb{N}$. If $y_n = (a_n, N_n)$ for all $n \in \mathbb{N}$, by the definition of R, $N_{n+1} = N_n + 1$. By induction, we obtain that $N_n = n + N$ for a certain $N \in \mathbb{N}$ and thus $a_n \in X_{n+N}$.

We are now in a position to create a sequence $\langle x_n\rangle_{n\in\mathbb{N}}$ such that $x_n \in X_n$ for all $n \in \mathbb{N}$, as the Axiom of Countable Choice states. Since the cartesian product $X_0 \times X_1 \times ... \times X_{N-1}$ is non-empty[6], there exists a sequence $x_0, x_1, ..., x_{N-1}$ such that $x_n \in X_n$ for all $n < N$. For $n \geq N$, we define $x_n = a_{n-N} \in X_n$. $\square$

[6] See Theorem 4.1.5

3.3 The Boolean Prime Ideal Theorem

To introduce the Boolean Prime Ideal Theorem, another weaker form of the Axiom of Choice, we have to give first the following definitions:

Definition 3.3.1. A *partial ordering* $\leq$ *of a boolean algebra* B is defined by $a \leq b \iff a + b = b$.

Definition 3.3.2. An *ideal* I *of a Boolean algebra* B is a non-empty proper subset of B such that:

(a) If $a \in I$ and $b \leq a$, then $b \in I$.

(b) If $a, b \in I$, then $a + b \in I$.

Definition 3.3.3. An ideal I of a Boolean algebra B is a *prime ideal* if

(c) For each $a \in B$, either $a \in I$ or $-a \in I$.

Lemma 3.3.1. *In a Boolean algebra* B*, an ideal* I *is a prime ideal* $\iff$ I *is maximal.*
Proof. See [2], page 15.

Boolean Prime Ideal Theorem (BPI). *Every Boolean algebra has a prime ideal.*

Theorem 3.3.1. $AC \Rightarrow$ *BPI.*

Proof. Let B be a Boolean Algebra and let Σ be the set containing all the ideals of B, which is partially ordered by inclusion ($\subseteq$). Let C be a chain in Σ and we define $U := \bigcup_{S \in C} S$.

Let's see that U is an ideal: if $a \in U$ and $b \leq a$, then $a \in S$ for some $S \in C$. Since S is an ideal and, $b \in S \subseteq U$ we obtain that $b \in U$. If $a, b \in U$, there exist some $S_1, S_2 \in C$ such that $a \in S_1$ and $b \in S_2$. Since C is a chain, we have that either $S_1 \subseteq S_2$ or $S_2 \subseteq S_1$. Suppose the former: then $a, b \in S_2$ and, since S_2 is an ideal, $a + b \in S_2 \subseteq U$, which implies $a + b \in U$.

Therefore, $U \in \Sigma$ and for every $S \in C$, $S \subseteq U$. U is an upper-bound of C so, by Zorn's Lemma, Σ has a maximal element, M. In a Boolean Algebra, if an ideal is maximal then it is also a prime ideal, so M is a prime ideal of the Boolean Algebra B. $\square$

The Prime Ideal Theorem is equivalent to its stronger version: *In every Boolean algebra, every ideal can be extended to a prime ideal.* It has also another equivalent form, analogous to these two forms given but related to the notion of filters and ultrafilters, which we define below.

Definition 3.3.4. A *filter F of a Boolean algebra B* is a non-empty proper subset of B such that:

(a) If $a \in F$ and $b \geq a$, then $b \in F$.

(b) If $a, b \in I$, then $a \cdot b \in F$.

Definition 3.3.5. A filter F of a Boolean algebra B is an *ultrafilter* if

(c) For each $a \in B$, either $a \in F$ or $-a \in F$.

Given the duality between ideals and filters, we have also these two equivalent formulations of the Prime Ideal Theorem: *Every Boolean algebra has an ultrafilter* and *Every filter in a Boolean algebra can be extended to an ultrafilter.* It is also equivalent to the **Ultrafilter Theorem**, that applies to sets: *Every filter over a set X can be extended to an ultrafilter*[7]. The notions of filter and ultrafilter on a set are defined as follows:

Definition 3.3.6. Given a set X, a filter F on X is a non-empty subset of $\mathcal{P}(X)$ that satisfies:

(a) $\emptyset \notin F$.

(b) If $A, B \in F$, then $A \cap B \in F$. That is, F is closed under intersection.

(c) If $A \in F$ and $A \subseteq B \subseteq X$, then $B \in F$.

Definition 3.3.7. A filter F on a set X is an *ultrafilter* if

(d) For each $A \subseteq X$, either $A \in F$ or $X - A \in F$.

Note that a filter on a set has the finite intersection property.

[7]Proofs and clarifications on these statements can be found in [2], pages 14-16.

3.4 Other Weaker Forms

The Ordering Principle (OP). *For every set X there is a linear ordering for X.*

Theorem 3.4.1. *BPIT $\Rightarrow$ OP*

Proof. For this proof we will need to use the fact that the BPIT is equivalent to the Compactness Theorem for First Order Logic[8]. We will prove a stronger statement, namely that every partial ordering of a set X can be extended to a linear ordering of X.

Let $(X, \leq)$ be a partially ordered set and let $\mathcal{L}$ be a language that contains constants for all $x \in X$ and a binary predicate $\preceq$. Let Σ be the set containing the following sentences of $\mathcal{L}$:

$$x \preceq y \wedge y \preceq x \rightarrow x = y \text{ for every } x, y \in X$$
$$x \preceq y \wedge y \preceq z \rightarrow x \preceq z, \text{ for every } x, y, z \in X$$
$$x \preceq y \vee y \preceq x \text{ for every } x, y \in X$$
$$x \preceq y \text{ for every } x, y \in X \text{ such that } x \leq y.$$

Every finite subset of Σ has a model. This can be proved by induction on the cardinality of a finite subset of Σ. Suppose that all $A \subset \Sigma$ with $|A| = n - 1$ can be linearly ordered preserving the order $\leq$ of X. We will see that this also holds for $A \cup \{x\}$, where $x \in X$. For every element $a \in A$ such that $a \leq x$, we define $a \preceq x$ and $b \preceq x$ for all $b \in A$ such that $b \preceq a$. Otherwise, given $b \in A$, if there does not exist $a \in A$ such that $a \leq x$ and $b \preceq a$, we state $x \preceq b$. $A \cup \{x\}$ is also linearly ordered and preserves the order $\leq$ of X.

Since every finite subset of Σ has a model, then, the Compactness Theorem tells us that Σ has a model as well: it produces a linear ordering $\preceq$ of X, which extends $\leq$. Now, since every set has the trivial partial ordering ($x \leq y \iff x = y$), then it can be extended to a linear ordering, as we have just seen. Therefore, every set can be linearly-ordered. $\square$

[8] This equivalence is proved in [2]. In section 4.6 we will prove that the Axiom of Choice implies the Compactness Theorem.

Axiom of Choice for Finite Sets (ACF). *Every family of finite non-empty sets has a choice function.*

Theorem 3.4.2. *OP $\Rightarrow$ ACF*

Proof. Let X be a family of finite non-empty sets and let $U := \bigcup X$ be the generalised union of X. By OP, U has a linear ordering $\leq$. For each $A \in X$, A is a finite chain in U; therefore, each $A \in X$ has a minimal element. We can now define the function $f : X \to U$ such that $f(A) = min(A) \in A$ for all $A \in X$. f is a choice function for X. $\square$

Axiom of Choice for Finite Sets of n Elements ($\mathbf{C}_n$). *Every family of n-element non-empty sets has a choice function.*

Theorem 3.4.3. *ACF $\Rightarrow$ C_n.*

Proof. If every family of finite non-empty sets has a choice function, then in particular every family of n-element sets has a choice function. $\square$

Axiom of Choice for Well-Orderable Sets (ACWO). Every family of non-empty well-orderable sets has a choice function.

Theorem 3.4.4. AC $\Rightarrow$ ACWO

Proof. If every family of non-empty sets has a choice function (AC), then, in particular, every family of non-empty well-orderable sets has a choice function. $\square$

Theorem 3.4.5. ACWO $\Rightarrow$ ACF

Proof. Since ACWO holds and every finite set is well-orderable, then in particular, every family of finite non-empty sets has a choice function. $\square$

Having proved the theorems above we obtain the following implications, which may be shown to be irreversible (see, e.g., [2]).

$$\text{AC} \Rightarrow \text{DC} \Rightarrow \text{CC}$$
$$\text{AC} \Rightarrow \text{BPI} \Rightarrow \text{OP} \Rightarrow \text{ACF} \Rightarrow \text{C}_n$$
$$\text{AC} \Rightarrow \text{ACWO} \Rightarrow \text{ACF} \Rightarrow \text{C}_n$$

4 Consequences of the Axiom of Choice

As we have seen in the section above, the Axiom of Choice has many equivalent and weaker forms, some of which are essential and widely-used to prove statements in different fields of modern mathematics. Unquestionably, we can affirm that without the Axiom of Choice or some of its weaker forms, mathematics would have been very different as we know it today: more than a few results that nowadays are regarded as essential could not have been proved in a world without choice. In this section we will see some of the implications of the Axiom in different areas of mathematics and in which way it is involved, paying attention on which part of some well-known proofs does the axiom play its role.
The consequences of the Axiom of Choice that are also equivalent to it will be marked with an *.

4.1 The Axiom of Choice in Set Theory

Theorem 4.1.1. *$AC \Rightarrow$ The notion of cardinality defined in 0.13. can be extended for all infinite sets, i.e., for every infinite set A there exists a unique aleph $\aleph_\alpha$ such that A has cardinality $\aleph_\alpha$, that is, $|A| = \aleph_\alpha$.*

Proof. If the Axiom of Choice holds, A can be well-ordered. Therefore, it is equipotent to some infinite ordinal, and hence to a unique initial ordinal number $\omega_\alpha = \aleph_\alpha$. □

Corollary 4.1.1. *$AC \Rightarrow$ For any sets A and B either $|A| \leq |B|$ or $|B| \leq |A|$.*

Proof. By the AC, every set can be well-ordered, that is, is equipotent to some ordinal (hence, to some cardinal). Let α be the cardinality of A ($|A| = \alpha$) and β the cardinality of B ($|B| = \beta$). As α and β are ordinals, either $\alpha \leq \beta$ or $\beta \leq \alpha$. Therefore, $|A| \leq |B|$ or $|B| \leq |A|$. □

Without the Axiom of Choice we would not be able to prove that the ordering $|A| \leq |B|$ is a linear ordering, it would only be possible to prove that it is a partial ordering. However, thanks to the axiom we can guarantee that any two cardinalities are comparable. It also gives us a proper way to define the cardinality of a set, as the unique initial ordinal equipotent to it. The Axiom allows us to extend the notion of cardinality presented in 0.13.: since all sets can be well-ordered, the definition of the cardinality of a set can now be applied to all sets. Furthermore, the arithmetic of cardinal numbers is simplified with the help of the axiom; without it many formulas used to operate with infinite cardinals would be very hard to prove, or would just become false. One example of this simplification could be Theorem 4.1.6. Another consequence of the fact that each infinite cardinality is an aleph, is a reformulation of the Continuum Hypothesis:

Definition 4.1.1. The *Continuum Hypothesis (CH)* states that there exists no set X such that $\aleph_0 < |X| < 2^{\aleph_0}$.

Since $2^{\aleph_0}$ is the cardinality of $\mathbb{R}$, the Continuum Hypothesis asserts that every infinite subset of $X \subset \mathbb{R}$ must have either the cardinality of $\mathbb{R}$ or be countable.

Corollary 4.1.2 *$AC \Rightarrow$ The Continuum Hypothesis can be reformulated as* $2^{\aleph_0} = \aleph_1$.

Proof. Since CH states that there is no cardinality between $\aleph_0$ and $2^{\aleph_0}$ and, given AC, each infinite cardinality must be of the form $\aleph_\alpha$ for some ordinal α, we have $2^{\aleph_0} = \aleph_1$. □

Theorem 4.1.2. *$AC \Rightarrow$ Every infinite set has a countable subset.*

Proof. Let A be an infinite set. A can be well-ordered and therefore put in the form of a transfinite one-to-one sequence $\langle a_\alpha \mid \alpha < \lambda \rangle$, where the infinite ordinal λ is the order type of A. The range $X = \{a_\alpha \mid \alpha < \omega\}$ of the initial segment of this sequence is a countable subset of A. □

However, this proposition does not require the full strength of the Axiom of Choice to be proved, the weaker version CC (Axiom of Countable Choice) is sufficient. Let us see this.

Theorem 4.1.3. *$CC \Rightarrow$ Every infinite set has a countable subset.*

Proof. Let A be an infinite set. We consider all finite one-to-one sequences $A_k = \langle a_i \mid i < k \rangle$, where $a_i \in A$ for all i and $k < \omega$. By the Axiom of Countable Choice, we are able to pick one k-sequence (A_k) for every $k \in \mathbb{N}$ and construct the set $B = \{A_k : k < \omega\}$. B has a choice function f such that $f(A_k) \in A_k$ for all k. The set $X = \{f(A_k) \mid k < \omega\}$ is a countable subset of A. $\square$

From now on we will, whenever possible, use the weakest possible version of AC to prove the statements presented in this section. However, when the avoidance of the full Axiom derives in a significant complication of the proof, we will use AC and state that the proof can be done with a weaker version of it.

Theorem 4.1.4. *CC $\Rightarrow$ The union of a countable collection of countable sets is countable.*

Proof. Let A be a countable set whose elements are also countable and let $X = \bigcup A$. We will show that X is countable. A is countable, so there exists a one-to-one sequence $\langle X_n \mid n \in \mathbb{N}\rangle$ such that $A = \{X_n \mid n \in \mathbb{N}\}$. For each $n \in \mathbb{N}$, the set X_n is countable, so there exists a countable sequence whose range is X_n. By the Countable Axiom of Choice, it is possible to choose one such sequence for every n, that is, it is possible to pick one sequence $a_n = \langle x_n(k) \mid k \in \mathbb{N}\rangle$ out of all the sequences whose range is X_n. If we choose one a_n for every $n \in \mathbb{N}$, we can obtain a mapping f of $\mathbb{N} \times \mathbb{N}$ onto X by $f(n, k) = x_n(k)$. Since $\mathbb{N} \times \mathbb{N}$ is countable and X is its image under f, X is also countable. $\square$

Theorem 4.1.5.* *AC $\Rightarrow$ An arbitrary cartesian product of non-empty sets is non-empty.*

Proof. Let $X = \{X_i \mid i \in I\}$ a family of non-empty sets. By the Axiom of Choice, there exists a choice function f such that $f(X_i) \in X_i$. Then, the element $(f(X_i))_{i\in I}$ is an element of the cartesian product $\prod_{i\in I} X_i$. $\square$

Theorem 4.1.6.* *AC $\Rightarrow$ Every infinite set A has the same cardinality as the cartesian product $A \times A$. That is, $|A| = |A|^2$.*

Proof. By Theorem 4.1.1, and since A is an infinite set, we have that $|A| = \aleph_\alpha = \omega_\alpha$, where ω_α is an initial ordinal. We want to see that the order types of ω_α and $\omega_\alpha \times \omega_\alpha$ are the same. We define the following order $\prec$ of $\omega_\alpha \times \omega_\alpha$:

$$(\alpha_1,\alpha_2) \prec (\beta_1,\beta_2) \iff \begin{cases} max\{\alpha_1,\alpha_2\} < max\{\beta_1,\beta_2\}, \text{ or} \\ max\{\alpha_1,\alpha_2\} = max\{\beta_1,\beta_2\} \wedge \alpha_1 < \beta_1, \text{ or} \\ max\{\alpha_1,\alpha_2\} = max\{\beta_1,\beta_2\} \wedge \alpha_1 = \beta_1 \wedge \alpha_2 < \beta_2 \end{cases}$$

First, we show that $\prec$ defines a well order in $\omega_\alpha \times \omega_\alpha$:

(a) $\prec$ is asymmetric: It follows from the definition, if $(\alpha_1,\alpha_2) \prec (\beta_1,\beta_2)$, clearly the opposite does not hold.

(b) $\prec$ is transitive: Suppose $(\alpha_1,\alpha_2) \prec (\beta_1,\beta_2)$ and $(\beta_1,\beta_2) \prec (\gamma_1,\gamma_2)$ for some pair of ordinals. We have $max\{\alpha_1,\alpha_2\} \leq max\{\beta_1,\beta_2\} \leq max\{\gamma_1,\gamma_2\}$. If $max\{\alpha_1,\alpha_2\} < max\{\gamma_1,\gamma_2\}$, then clearly $(\alpha_1,\alpha_2) \prec (\gamma_1,\gamma_2)$. Suppose then, that $max\{\alpha_1,\alpha_2\} = max\{\beta_1,\beta_2\} = max\{\gamma_1,\gamma_2\}$. In this case we have $\alpha_1 \leq \beta_1 \leq \gamma_1$. If $\alpha_1 < \gamma_1$, then $(\alpha_1,\alpha_2) \prec (\gamma_1,\gamma_2)$. If $\alpha_1 = \beta_1 = \gamma_1$, in this case we have $\alpha_2 < \beta_2 < \gamma_2$. Since $\alpha_2 < \gamma_2$, $(\alpha_1,\alpha_2) \prec (\gamma_1,\gamma_2)$.

(c) Any two elements are comparable: For any two different pairs of ordinals $(\alpha_1,\alpha_2), (\beta_1,\beta_2)$ we want to see that either $(\alpha_1,\alpha_2) \prec (\beta_1,\beta_2)$ or $(\alpha_1,\alpha_2) \succ (\beta_1,\beta_2)$. Clearly, fone of the two conditions holds.

(d) There exists a least-element for every non-empty set X of pairs of ordinals. Let δ be the least-element of the set $\{max\{\alpha,\beta\} \mid (\alpha,\beta) \in X\}$ and let $Y = \{(\alpha,\beta) \in X \mid max\{\alpha,\beta\} = \delta\}$. Then, we define α_0 as the least-element of the set $\{\alpha \mid (\alpha,\beta) \in Y\}$ and the set $Z = \{(\alpha,\beta) \mid \alpha = \alpha_0\}$. If β_0 is the least-element of the set $\{\beta \mid (\alpha,\beta) \in Z\}$, then clearly and by definition, (α_0,β_0) is the least element of X.

Now, we will use transfinite induction[9] on α to show that $|\omega_\alpha \times \omega_\alpha| \leq |\omega_\alpha|$. Equivalently, $\aleph_\alpha \cdot \aleph_\alpha \leq \aleph_\alpha$. Since $\aleph_\alpha \leq \aleph_\alpha \cdot \aleph_\alpha$, by proving this we will show that $\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$.
The property holds for $\alpha = 0$, since for every set (n,m), the set of $\preceq$-predecessors of (m,n) is finite We have, then, $\aleph_0 \cdot \aleph_0 = \aleph_0$. Now consider $\alpha > 0$ and suppose that the property holds for every $\beta < \alpha$. If the order-type of the set $\omega_\alpha \times \omega_\alpha$ was greater than ω_α, then there would exist a pair (α_1,α_2) and a set $X = \{(\gamma_1,\gamma_2) \in \omega_\alpha \times \omega_\alpha \mid (\gamma_1,\gamma_2) \prec (\alpha_1,\alpha_2)\}$

[9]Transfinite induction is an extension of mathematical induction to well-ordered sets, for example ordinal or cardinal numbers. Suppose a property P defined for all ordinals α is true for all ordinals $\beta < \alpha$. If P holds for α, then it is true for all ordinals.

such that $|X| \geq \aleph_\alpha$. We will show that this does not happen, that is, that for any pair (α_1, α_2), the set X satisfies $|X| < \aleph_\alpha$.
Let $\beta = max\{\alpha_1, \alpha_2\} + 1$. We have that $\beta \in \omega_\alpha$ and $X \subseteq \beta \times \beta$, since $max\{\gamma_1, \gamma_2\} \leq max\{\alpha_1, \alpha_2\} < \beta$ and thus $\gamma_1, \gamma_2 \in \beta$. Now, consider $\lambda < \alpha$ such that $|\beta| \leq \aleph_\lambda$. By applying the inductive hypothesis we obtain that $|X| \leq |\beta \times \beta| = |\beta| \ \cdot |\beta| \leq \aleph_\lambda \cdot \aleph_\lambda \leq \aleph_\lambda < \aleph_\alpha$. □
By induction, this process can be generalised, and we can obtain that, for each $n \in \mathbb{N}$, $|A| = |A|^n$.

Definition 4.1.2. A family of non-empty sets X is of *finite character* if it satisfies:

(a) For every $A \in X$, every finite subset of A belongs to X

(b) If A is a set such that every finite subset of A belongs to X, then A belongs to X

Theorem 4.1.7.* *$AC \Rightarrow$ Tukey's Lemma: If X is a collection of non-empty sets of finite character, then X has a maximal element with respect to inclusion $\subseteq$.*

Proof. Let X be a collection of non-empty sets of finite character. X is partially ordered by inclusion $\subseteq$. Let C be a chain in X and we define $U := \bigcup_{A \in C} A$. Since for every finite subset $B \in U$, we have $B \in X$, it follows that $U \in X$. So, U is an upper bound of C. Applying Zorn's Lemma, we obtain that X has a maximal element. □

Definition 4.1.3. A set X is called *Dedekind infinite* if there exists some proper subset Y of X with $|X| = |Y|$. Otherwise, X is called *Dedekind finite.*

The usual definition of an infinite set (that is, a set with an infinite cardinality) is only equivalent to the definition of a Dedekind infinite set under the assumption of the Axiom of Countable Choice.

Theorem 4.1.8. *$CC \Rightarrow$ For every set X, X is Dedekind infinite iff X is infinite.*

This proposition was assumed by most mathematicians before the foundational crisis of mathematics, but it is in fact impossible to prove without the Axiom of Countable Choice. As a matter of fact, there exists a model of ZF where there exists an infinite set that is at the same time Dedekind-finite.

Proof. First, we will see that every Dedekind infinite set is infinite. For this implication we don't need CC. Let X be a finite set. So, there exists a bijection between X and some ordinal $n \in \mathbb{N}$, hence $|X| = n$. Let us see by induction on n that X is not Dedekind infinite. For $|X| = 0$, $X = \emptyset$, it has no proper subset. Suppose the property holds for n and let X be a set of cardinality $n+1$. Let a be an element in X and $\bar{X} = X \backslash \{a\}$. $\bar{X}$ has cardinality n and by the induction hypothesis, every proper subset has cardinality strictly less than n. Since every proper subset of X is either $\bar{X}$, of cardinality $n < n+1$, or of the form A, $A \cup \{a\}$, where A is a proper subset of $\bar{X}$ which has cardinality strictly less than n and $n+1$ respectively. In either case, we obtain that every proper subset B of X satisfies $|B| < n+1$.
Now, we will show that every infinite set is also Dedekind infinite. We define a function $g : \mathbb{N} \to \mathcal{P}(X)$ such that $g(n)$ is the set of finite subsets of X of size n. Since X is an infinite set, $g(n)$ is non-empty for each $n \in \mathbb{N}$. We have now the countable family of non-empty sets $\{g_n \mid n \in \mathbb{N}\}$ and, by the Axiom of Countable Choice, there exists a choice function f such that $f(g(n)) \in g(n)$. We may now define the set $F = \{f(g(n)) \mid n \in \mathbb{N}\}$, where for each n, $f(g(n))$ is a subset of X of size n and the set U as the union of members in F. U is a countable union of finite sets, therefore it is a countable set and a bijection $h : \mathbb{N} \to U$ can be defined. We may now define a bijection $B : X \to X\backslash\{h(0)\}$ as follows:

$$B(x) = \begin{cases} x & \text{if } x \notin U \\ h(n+1) & \text{if } x = h(n) \end{cases}$$

B is a bijection from X to a proper subset of X and, therefore, X is Dedekind infinite. □

4.2 The Axiom of Choice in Algebra

As we have seen in the chapter above, three equivalents of the Axiom of Choice are of huge importance for the development of modern algebra: Zorn's Lemma, Krull's Theorem and the fact that every vector space has a basis. We have seen that the axiom implies the first of these. Let us see now how the axiom implies the last two.

Theorem 4.2.1.* *$AC \Rightarrow$ Every vector space has a basis.*

Proof. Let V be a vector space and let L be the set containing all linearly independent subsets of V, which is partially ordered by inclusion ($\subseteq$). Now let C be a chain in L and we define $U := \bigcup_{S\in C} S$.

U is linearly independent: if $t_1, t_2, ..., t_n \in U$, then each t_i belongs to some $S_i \in C$ and, since C is a chain, there must be some $S_k \in C$ such that $t_1, t_2, ...t_n \in S_k$. As S_k is linearly independent there cannot be any non-trivial combination of $t_1, ..., t_n$ that equals zero. Therefore, $U \in L$ and U is an upper bound of C. If the Axiom of Choice holds, so does Zorn's Lemma and, hence, L has a maximal set, M. M is a linearly independent set of vectors of V, let us see that it is also a basis of the vector space.

Suppose M is not a basis of V, then there must exist some vector v which cannot be written as a linear combination of the elements in M. This implies that $M \cup \{v\}$ is a linearly independent set and hence $M \cup \{v\} \in L$. But $M \subsetneq M \cup \{v\}$, which contradicts the maximality of M in L. Therefore, M forms a basis for V. $\square$

Theorem 4.2.2.* *$AC \Rightarrow$ Krull's Theorem: If A is a ring different from the trivial ring, then A has a maximal ideal.*

Proof. Let A be a ring and let Σ be the set of all the proper ideals of A. Clearly, $\Sigma \neq \emptyset$, since $\{0\} \in \Sigma$ and Σ is partially ordered by $\subseteq$. Let C be a chain in Σ and we define $U := \bigcup_{B\in C} B$.

U is an ideal of A: if $a, b \in U$, then there are some ideals B_1, B_2 in C such that $a \in B_1 \in C$ and $b \in B_2 \in C$. Suppose $B_1 \subseteq B_2$, then $a, b \in B_2$, $a + b \in B_2 \subseteq U$ and, for every $\lambda \in A$, $\lambda a \in B_2 \subseteq U$. Furthermore, $U \neq \{1\}$, for if $1 \in U$, then $1 \in B_i$ for some B_i in C and this would result as $B_i = A$, which leads to contradiction, because B_i is a proper ideal of A. Therefore we have that $U \in \Sigma$ and for every $B \in C$, $B \subseteq U$. U is an upperbound of C, so by Zorn's Lemma, Σ has a maximal element. $\square$

Theorem 4.2.3.* *$AC \Rightarrow$ Every non-empty set can be given a group structure.*

Proof. Let X be a non-empty set. If X is finite, then it has a group structure as a cyclic group generated by any element $x \in X$. If $|X| = n$, then let f be a bijection between X

and $\mathbb{Z}/(n)$. We define the group operation in X as $x \star y = f^{-1}(f(x) + f(y))$. Suppose X is infinite and let F be the set whose elements are all the finite subsets of X. We define the following operation in F:

$$\text{For all } U, V \in F, \quad U \Delta V = (U - V) \cup (V - U)$$

(F, Δ) is a group, where $\emptyset$ is the identity for all $U \in F$, $U^{-1} = U$, since $U \Delta U = \emptyset$. Clearly, $U \Delta V \in F$ for each $U, V \in F$ and, given the properties of the union and difference in sets, one can verify that the associative products in the defined operation Δ holds. Thus, (F, Δ) is a group. Now, it remains to be seen that X is also a group. For that, we will see that assuming the Axiom of Choice, X can be put into bijection with F and become a group via the bijection.

We will show that $|X| = |F|$, and hence that such a bijection exists. For each $n \in \mathbb{N}$, let $F_n \subset F$ be the set of all subsets of F of cardinality n. F is then the disjoint union of the F_n. Since every subset with n elements is an element of the n-fold cartesian product X^n of X, then the number of subsets in F_n is at most $|X|^n$. We have $|F_n| \leq |X|^n$ and, by AC, we get $|X|^n = |X|$ for all n. To continue the proof we will first need the following Lemma,

Lemma 4.2.1. *AC $\Rightarrow$ For any set X, $|\bigcup X| \leq |X| \cdot$ sup $\{|Y| : Y \in X\}$*

Proof. See [3], page 49.

Now, taking into account the result of the previous lemma and the fact that $|F_n| \leq |X|$ for all n, we have $|F| = |\bigcup_n F_n| \leq |X| \cdot \aleph_0 = |X|$. But since F contains all singletons, we also have that $|X| \leq |F|$ and hence, $|F| = |X|$. There exists, then, a bijection $f : X \to F$. We can now define the group $(X, \star)$: for every $x, y \in X$, $x \star y = f^{-1}(f(x) \Delta f(y))$. $\square$

Definition 4.2.1. Let F be a group and let $X \subset F$. F is called a *free group generated by* X if for every function f from X to any group G, there exists a unique homomorphic extension of f, $\bar{f} : F \to G$. X is called a *set of free generators* of F.

Theorem 4.2.4. *AC $\Rightarrow$ Nielsen-Schreier Theorem. Every subgroup of a free group is free.*

Although different proofs of the Nielsen-Schreier Theorem are known, all of them depend on the Axiom of Choice. We will not present any proof here, for all of them are rather involved. For a proof of the theorem, see [4].

Definition 4.2.2. The *algebraic closure* of a field K is an extension $\bar{K}$ of K such that $\bar{K}$ is the smallest algebraically closed field that contains K.

Theorem 4.2.5. *$AC \Rightarrow$ For any field K, there exists a unique algebraic closure of K.*

Here we will use a proof that depends on Zorn's Lemma, but this theorem is in fact weaker than the Axiom of Choice. In fact, it can be proved with the Compactness Theorem (see Theorem 4.6.1.), which is a consequence of the Boolean Prime Ideal Theorem. We will only prove the existence of an algebraic closure, the proof of its uniqueness also requires the use of the Axiom of Choice, since it uses Zorn's Lemma.

Proof. To prove the existence, the following Lemma will be used:

Lemma. *4.2.2. Let E be an algebraically closed field and K a subfield of E. Then, $\{x \in E \mid x$ is algebraic over $K\}$ is an algebraic closure over K.*
Proof. See [5].

Now, applying the previous Lemma it will be sufficient to see that there exists an algebraically closed field E that contains K. Let F be the set of all polynomials in one variable with coefficients in K and of degree ≥ 1. Let's consider an indeterminate x_f for each polynomial $f \in F$ and let $K[x_f]$ be the polynomial ring in the indeterminates $\{x_f\}_{f \in F}$. Let I be the ideal of $K[x_f]$ generated by the polynomials $f(x_f) \in F[x_f]$.
I is a proper ideal: If it wasn't, there would exist a set of polynomials $g_1, ..., g_t \in K[x_f]$ such that $1 = g_1 f_1(x_{f_1}) + ... + g_t f_t(x_{f_t})$. Let L be a finite extension of K in which every f_i has a root α_i, for $1 \leq i \leq t$. If we evaluate the previous equation on $\alpha_1, ..., \alpha_t$, we end up with $1 = 0$. Hence, I is a proper ideal.
Now, by the Axiom of Choice there exists a maximal ideal M that contains I. We can now consider the field $K[x_f]/M = K_1$. The composition $K \hookrightarrow K[x_f] \to K[x_p]/M = K_1$ is injective and K identifies with a subfield of K_1. If $f \in K[x_f]$ has degree ≥ 1, then he class $\bar{x_f}$ in K_1 is a root of f.
Inductively, we obtain a succession of fields $K = K_0 \subseteq K_1 \subseteq ... \subseteq K_n \subseteq ...$, where K_n is obtained from K_{n-1} the same way K_1 is obtained from K. Let K' be the union of all the fields in the succession, $K = \bigcup K_i$. K' has a field structure (for every $a, b \in K$, there

exists some n such that $a, b \in K_n$). If $g \in K'[x]$, then $g \in K_n[x]$ for some n and hence has a root in $K_{n+1} \subseteq K'$. Since K' is algebraically closed and $K \subseteq K'$, we can conclude the that $\{x \in K' \mid x \text{ is algebraic over } K\}$ is an algebraic closure over K. □

Theorem 4.2.6. *$AC \Rightarrow$ The additive groups of $\mathbb{R}$ and $\mathbb{R}^2$ are isomorphic.*

Proof. Let B be a basis of $\mathbb{R}$ as a vector space over $\mathbb{Q}$. We have seen that the Axiom of Choice implies the existence of such basis in Theorem 4.2.1. Since $\mathbb{R}$ is uncountable, B is uncountable. Since B and $C = (B \times \{0\}) \cup (\{0\} \times B)$ have the same cardinality, there exists a bijection h between these sets. Furthermore, C is a basis of $\mathbb{R}^2$, again as a vector space over $\mathbb{Q}$. We define $f : \mathbb{R} \to \mathbb{R}^2$ as follows: if $x \in \mathbb{R}$, there exists a unique expression of x of the form $x = \sum\limits_{b \in B} a_b \cdot b$, where $a_b \in \mathbb{Q}$ for all $b \in B$ and $a_b \neq 0$ for a finitely many b. Now, $f(x) = \sum\limits_{b \in B} a_b \cdot h(b)$. f is an isomorphism between $\mathbb{R}$ and $\mathbb{R}^2$ as additive groups. □

4.3 The Axiom of Choice in Topology

Theorem 4.3.1. *$CC \Rightarrow$ The following are equivalent: for a subset X of a metric space M and every $x \in M$, x is in the closure of X if*

(a) *every neighborhoud of x intersects X.*

(b) $\lim\limits_{n \to \infty} x_n = x$ *for some sequence $(x_n)_n$ of points in X.*

Proof. Let us see (a) $\Rightarrow$ (b). Let x be a point in the closure of X. Let B_n be a ball of center x and radius $\frac{1}{n}$. By (a), $B_n \cap X \neq \emptyset$. By the Axiom of Countable Choice we are able to choose a point x_n for each B_n and we obtain that the sequence $(x_n)_n$ of points in X converges to x.

(b) $\Rightarrow$ (a) does not require the axiom. Let N be an open neighbourhood of x. There exists $\epsilon > 0$ such that the ball B_ϵ of center x and radius ϵ satisfies $B_\epsilon \subseteq N$. Since (b) holds, there exists n_ϵ such that $d(x, x_n) \leq \epsilon$ for all $n \geq n_\epsilon$, where $x_n \in X$ for all n. In fact, $x_{n_\epsilon} \in X$ and $x_{n_\epsilon} \in B_\epsilon \subseteq N$. Hence, $N \cap X \neq \emptyset$. □

Definition 4.3.1. A topological space X is *compact* if for every family $(U_i)_i$ of open sets such that $\bigcup\limits_i U_i = X$, there exists $i_1, ..., i_n$ such that $X = U_{i_1} \cup ... \cup U_{i_n}$.

Theorem 4.3.2.* *$AC \Rightarrow$ Tychonoff's Theorem: The product of compact topological spaces is compact.*

Proof. Let $\{K_i \mid i \in \mathcal{I}\}$ be a collection of compact spaces and let K be its product. Let $\mathcal{A}$ be a set of closed subsets of K having the finite intersection property (FIP), that is, that the intersection over any finite subcollection of $\mathcal{A}$ is not empty. We will show that $\bigcap \mathcal{A}$ is not empty (and hence, that K is compact).
Proceeding in the same way we have done in previous proofs, we can see that every chain in the set of all subsets (not necessarily closed) of K with the FIP has an upper bound and thus, by Zorn's Lemma, there exists a maximal set $\mathcal{B}$ of subsets of K with the FIP and such that $\mathcal{A} \subseteq \mathcal{B}$.
If $\pi_i(B)$ is the i-th canonical projection, we have that the $\pi_i(B)$, $B \in \mathcal{B}$, have the FIP and since the K_i are compact, we have that for each i there exists a b_i that belongs to the closure of $\pi_i(B)$ for all $B \in \mathcal{B}$. If we let $b = (b_i \mid i \in \mathcal{I})$, then it is sufficient to check that the neighbourhoods of b are contained in $\mathcal{B}$. This will imply that the neighbourhoods of b intersect all $B \in \mathcal{B}$ and thus that b is in the closure of B for all $B \in \mathcal{B}$ and hence in all $A \in \mathcal{A}$.
For each i, we pick a neighbourhood N_i of b_i, such that $N_i = K_i$ for almost all i. Let N be the cartesian product of all N_i, it is clear that N is a neighbourhood of b and it is enough to see that $N \in \mathcal{B}$. Even more, since N is the intersection of finitely many $\pi_i^{-1}(N_i)$, it suffices to see that $\pi_i^{-1}(N_i) \in \mathcal{B}$ for all i. But since b_i is in the closure of $\pi_i(B)$, we have that $N_i \cap \pi_i(B) \neq \emptyset$ for all $B \in \mathcal{B}$. Hence, $\pi_i^{-1}(N_i) \cap B \neq \emptyset$ for all $B \in \mathcal{B}$, and by the maximality of $\mathcal{B}$, we have $\pi_i^{-1}(N_i) \in \mathcal{B}$ □

4.4 The Axiom of Choice in Analysis

Theorem 4.4.1. *$CC \Rightarrow$ The following definitions are equivalent: A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point x if*

(a) $\forall \epsilon > 0$, $\exists \delta > 0$ *such that* $\forall y$, *if* $|x - y| < \delta$ *then* $|f(x) - f(y)| < \epsilon$.

(b) $\lim_{n\to\infty} x_n = x \Rightarrow \lim_{n\to\infty} f(x_n) = f(x)$.

Proof. The implication (a) $\Rightarrow$ (b) doesn't need the axiom: we pick $\epsilon > 0$ and find a $\delta > 0$ that satisfies (a). If we have a sequence $(x_n)_n$ that converges to x, then there exists n_0 such that $|x_n - x| < \delta$ for all $n \geq n_0$. Since (a) holds, then $|f(x_n) - f(x)| < \epsilon$ for all $n \geq n_0$, therefore, $(f(x_n))_n$ converges to $f(x)$.
Assume CC. Let's see (b) $\Rightarrow$ (a). If (a) is false, then there exists some $\epsilon > 0$ such that $\forall \delta > 0$, there exists some y such that $|x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$. Given that ϵ, for $\delta = \frac{1}{n}$, we can define the set $X_n = \{y : |x - y| \leq \delta\}$. By the CC we are able to choose $x_n \in X_n$ for each n, such that $|x - x_n| < \delta$ but $|f(x) - f(x_n)| \geq \epsilon$. Then, we got $\lim_{n\to\infty} x_n = x$ but $\lim_{n\to\infty} x_n = x$ is false. Contradiction, so (a) is true. $\square$

Theorem 4.4.2.* *AC $\Rightarrow$ every surjection has a right inverse.*

Proof. Let $f : X \to Y$ be a surjection, where X, Y are two sets. We want to see that there exists a function $g : Y \to X$ such that $f(g(y)) = y$ for all $y \in Y$. We consider the family of non-empty sets $\{f^{-1}(y)\}_{y \in Y}$, where f^{-1} denotes the preimage of y under f. By the Axiom of Choice, there exists a choice function $g : Y \to X$ such that for every $f^{-1}(y)$, $g(y) \in f^{-1}(y)$ for all $y \in Y$. Hence, $f(g(y)) = y$. $\square$

Definition 4.4.1. A function f defined on a vector space V over $\mathbb{R}$ is called a *linear functional* on V if $f(a \cdot u) + f(b \cdot v) = a \cdot f(u) + b \cdot f(v)$ for all $a, b \in \mathbb{R}$ and $u, v \in V$.

Definition 4.4.2. A function p defined on a vector space V is called a *sublinear functional* on V if $p(u+v) \leq p(u) + p(v)$ for all $u, v \in V$ and $p(a \cdot u) = a \cdot p(u)$ for all $a \geq 0$ and $u \in V$.

Theorem 4.4.3. *AC $\Rightarrow$ Hahn-Banach Theorem. Let p be a sublinear functional on a vector space V and let f_0 be a linear functional on a subspace V_0 of V such that $f_0(v) \leq p(v)$ for all $v \in V_0$. Then, there exists a linear functional f such that $f_0 \subseteq f$ and $f(v) \leq p(v)$ for all $v \in V$.*

The Hahn-Banach Theorem is an essential theorem in functional analysis. It is strictly weaker than the Axiom of Choice, in fact, it is implied by the Ultrafilter lemma, which as stated in 3.3, is an equivalent form of the Boolean Prime Ideal Theorem.

Proof. Let F be the set of all linear functionals g defined on some $W \subseteq V$ and that satisfy $f_0 \subseteq g$ and $g(w) \leq p(w)$ for all $w \in W$. We will show by Zorn's Lemma that $(F, \subseteq)$ has a maximal element f.

Let C be a chain in F and we define $\bar{g} := \bigcup_{g \in C} g$. Let's show that $\bar{g} \in F$. Clearly $\bar{g}$ is a function with values in $\mathbb{R}$ and $f_0 \subseteq \bar{g}$. Also, dom $\bar{g} = \bigcup_{g \in C}$ dom g, which is the union of $\subseteq$-chain of subspaces of V and hence, a subspace of V. Let's see now that $\bar{g}$ is a linear functional. Consider $a, b \in \mathbb{R}$ and $u, v \in$ dom $\bar{g}$. Then, there exist some $g_1, g_2 \in C$ such that $u \in$ dom $g_1 \in C$ and $v \in$ dom $g_2 \in C$. Since C is a chain, either $g_1 \subseteq g_2$ or $g_2 \subseteq g_1$. Consider the first case (the other one is analogous): then $u, v \in$ dom g_2 and $\bar{g}(a \cdot u + b \cdot v) = g_2(a \cdot u + b \cdot v) = a \cdot g_2(u) + b \cdot g_2(v) = a \cdot \bar{g}(u) + b \cdot \bar{g}(v)$, since g_2 is linear. Finally, $\bar{g}(u) = g(u) \leq p(u)$ for $u \in$ dom $\bar{g}$ and $g \in C$ such that $u \in$ dom g. $\bar{g}$ is an upper-bound of C, since for every $g \in C$, $g \subseteq \bar{g}$. Therefore, by Zorn's Lemma, F has a maximal element f.

Suppose now that dom $f \neq V$, that is, dom $f \subset V$ and let $u \in V-$ dom f and W be the subspace of V spanned by u and dom f. Every $w \in W$ can be uniquely written as $w = x + a \cdot u$, where $x \in$ dom f and $a \in \mathbb{R}$. Now, for every $c \in \mathbb{R}$, the function $f_c(x + a \cdot u) = f(x) + a \cdot c$ is a linear function in W (since f is linear) and satisfies $f \subset f_c$. If such function also satisfies $f_c(x + a \cdot u) \leq p(x + a \cdot u)$ for some $c \in \mathbb{R}$ and for all $x \in$ dom f and $a \in \mathbb{R}$, then it would contradict the maximality of f and by *reductio ad absurdum* we could conclude that, as a matter of fact, dom $f = V$ and f is the function that Hahn-Banach Theorem describes. To finish the proof, then, we need to find such c.

For all $a > 0$ and $x, y \in$ dom f, the c we are looking for has to satisfy $f(x) + a \cdot c \leq p(x + a \cdot u)$ and $f(y) - a \cdot c \leq p(y - a \cdot u)$ (if $a = 0$, then since $f \in F$, the property is immediately satisfied). This is equivalent to

$$f(y) - p(y - a \cdot u) \leq a \cdot c \leq p(x + a \cdot u) - f(x)$$

which in turn, is equivalent to

$$f(\frac{y}{a}) - p(\frac{y}{a} - u) \leq c \leq p(\frac{x}{a} + u) - f(\frac{x}{a})$$

But since $f \in F$, for all $v, t \in$ dom f, we have that $f(v) + f(t) = f(v + t) \leq p(u + v) \leq p(v - u) + p(t + u)$ and hence $f(v) - p(v - u) \leq p(t + u) - f(t)$. If we let $A =$ sup $\{f(v) - p(v - u) \mid v \in \text{dom } f\}$ and $B =$ inf $\{p(t + u) + f(t) \mid t \in \text{dom } f\}$, then clearly $A \leq B$. We can then choose $c \in \mathbb{R}$ such that $A \leq c \leq B$. $\square$

Theorem 4.4.4. *$AC \Rightarrow$ Every Hilbert space has an orthonormal basis.*

Proof. Let H be a Hilbert space and let O bethe set containing all orthonormal subsets of H, partially ordered by inclusion ($\subseteq$). Let C be a chain in O and we define $U := \bigcup\limits_{S \in C} S$. U is clearly an orthonormal set, since $||u|| = 1$ for all $u \in U$ and it is orthogonal (because all $S \in C$ are orthonormal). U is an upper bound of C and, by Zorn's Lemma O has a maximal element B. B is a maximal orthonormal set and this suffices to see that it is a basis for H. $\square$

4.5 The Axiom of Choice in Graph Theory

Definition 4.5.1. In Graph Theory, a *tree* is a connected graph without cycles. A *spanning tree* is a tree T of a graph G that contains all the vertices of G.

Theorem 4.5.1.* *$AC \Rightarrow$ Every connected graph has a spanning tree.*

Proof. Let $G = (V, E)$ be a connected graph. We choose $r \in V$ and for each $v \in V \backslash \{r\}$ we define $X(v) = \{v' \in V \mid v' \text{ is adjacent to } v \wedge d(v', r) < d(v, r)\}$, where d defines the distances between two vertices. Every $X(v)$ is non-empty, since for every v there is a path from r to v: $(r = v_0, ..., v_{n-1}, v_n = v)$ with minimum distance and hence $v_{n-1} \in X(v)$. By the Axiom of Choice, there exists a choice function f that allows us to pick a vertex for each set $X(v) : f(X(v)) = v' \in X(v)$. If we define X as the set of all edges $\{v, v'\}$, then $G^* = (V, X)$ is a spanning tree.
By induction on $d(v, r)$ from a vertex v to r, we will show that G^* is connected. If $d(v, r) = 1$, then $X(v) = \{r\}$, so $v' = r$. If $d(v, r) > 1$, suppose all $u \in V$ such that $d(u, r) < d(v, r)$ are connected to r. Since $v' \in X(v)$ satisfies $d(v', r) < d(v, r)$, v' is connected to r and since it is adjacent to v, v is also connected to r.
G^* has no cycles: Suppose $v_1 v_2 ... v_n v_1$ is a cycle C in G^*. Then $\{v_1, v_2\} \in C$, which means that $d(v_1, r) < d(v_2, r)$ or $d(v_2, r) < d(v_1, r)$. Suppose the former, this means that $X(v_2) = v_1$. Then, since $\{v_2, v_3\} \in C$, we have that $d(v_2, r) < d(v_3, r)$ and $X(v_3) = v_2$. If we iterate this process, we end up obtaining that $d(v_1, r) < d(v_2, r), ..., d(v_n, r) < d(v_1, r)$, which leads to contradiction. Since G^* is connected and has no cycles, we can conclude that it is a spanning tree of G. $\square$

4.6 The Axiom of Choice in Logic

Theorem 4.6.1. *$AC \Rightarrow$ The Compactness Theorem for First Order Logic. If every finite subset of Σ has a model, then Σ has a model. Equivalently, if every finite subset of Σ is satisfiable, then Σ is satisfiable.*

The Compactness Theorem is implied by the Boolean Prime Ideal Theorem (in fact, it is equivalent to it). A proof of this can be found in [2], pages 17-18. We will give here a simpler proof that depends on the Axiom of Choice. However, if we want to prove the Compactness Theorem for a countable language $\mathcal{L}$ where we have a recursive enumeration of the formulas, then neither the Axiom nor a weaker choice principle is needed. However, if we want to prove it (as we will do here) in a general way, for any language $\mathcal{L}$, we cannot do it without the Axiom or the Boolean Prime Ideal Theorem.

Proof. In order to prove the implication above we will use the Ultrafilter lemma, the equivalence of the BPI presented in section 3.3 which states that every filter can be extended to an ultrafilter and Łoś's theorem, that will be stated below. A proof of it can be found in [6].

Let Σ be a collection of sentences in a first order languace $\mathcal{L}$ such that every finite subset of Σ has a model. Let I be the collection of all finite subsets of Σ and, for every $i \in I$ let $\mathcal{M}_i$ be a model for the sentences in i (note that by choosing a model for every i we have already used the Axiom of Choice). Also, for every $i \in I$ we define $J_i = \{j \in I \mid i \subseteq j\}$ and let $F = \{\ J \subseteq I \mid\ J_i \subseteq J$ for some $i \in I\ \}$.

We will see that F is a filter. F is closed under finite intersection since the collection of all J_i is also closed under finite intersection (indeed: $J_{i_1} \cap J_{i_2} = J_{i_1 \cup i_2}$). Also, F is by definition closed under containment ($J \subseteq F$ and $J \subseteq K$, clearly $K \subseteq F$) and it does not contain the empty set since J_i is non-empty for all $i \in I$. So F is a filter and by the Ultrafilter Lemma, it can be extended to an ultrafilter U.

Now, $\prod_{i \in I}$ dom $\mathcal{M}_i$ is the cartesian product of the domains of the $\mathcal{M}_i$ and we can define the following equivalence relation on $\prod_{i \in I}$ dom $\mathcal{M}_i$:

$$f \sim_U g \Leftrightarrow \{i \in I \mid f(i) = g(i)\} \in U$$

We may now define the ultraproduct[10] $\mathcal{M} = \prod_{i \in I} \mathcal{M}_i / U$ as the quotient set with respect to $\sim_U$. Łoś's theorem states that for any formula σ, $\mathcal{M} \models \sigma \iff \{i \in I \mid \mathcal{M}_i \models \sigma\} \in U$. Now, in our case, for any $\sigma \in \Sigma$ we have that $\{\sigma\} \in I$ and $J_{\{\sigma\}} \subseteq \{i \in I \mid \mathcal{M}_i \models \sigma\}$. Since $J_{\{\sigma\}} \in F \subseteq U$, we have that $J_{\{\sigma\}} \in U$ and thus $\{i \in I \mid \mathcal{M}_i \models \sigma\} \in U$. By Łoś's theorem, $\mathcal{M} \models \sigma$. Since this is true for any $\sigma \in \Sigma$, we obtain $\mathcal{M} \models \Sigma$, so Σ has a model. □

Theorem 4.6.2. *AC ⇒ Gödel's Completeness Theorem for First Order Logic. Every consistent set of formulas is satisfiable.*

The Completeness Theorem for first order logic is equivalent to the Compactness Theorem for first order logic and both of them are equivalent to the Boolean Prime Ideal Theorem. It is enough for our purposes to prove that Compactness Implies Completeness. A proof of this can be found in [8].

Theorem 4.6.3.: Diaconescu's Theorem. *AC ⇒ Law of the Excluded Middle:* $p \vee \neg p$

This theorem states that the Axiom of Choice is sufficient to derive the Law of the Excluded Middle in constructive set theory. That is specially problematic in intuitionistic logic, which has a constructivist approach and does not accept the law of the excluded middle: one can only prove the truth of $p \vee \neg p$ for a specific p once either p or $\neg p$ has been proved.

Proof. Let $C = \{0, 1\}$ and let p be a proposition. Now, we define the two following sets:

$$A = \{x \in C \mid (x = 0) \vee p\}, \qquad B = \{x \in C \mid (x = 1) \vee p\}$$

Since $0 \in A$ and $1 \in B$ both sets are non-empty and $X = \{A, B\}$ is a collection of finite non-empty sets. By AC, X has a choice function $f : X \to C$ such that $f(A) \in A$ and $f(B) \in B$. So, by definition of A and B, we have $f(A) = 0 \vee p$ and $f(B) = 1 \vee p$. So, $(f(A) = 0 \wedge f(B) = 1) \vee p$. Since $p \to (A = B)$, we have $p \to (f(A) = f(B))$. So, if $f(A) = 0 \wedge f(B) = 1$, then $f(A) \neq f(B)$, hence $\neg p$. And if it is not the case that $f(A) = 0 \wedge f(B) = 1$, then p. □

[10] For more information on ultraproducts and a more detailed proof, see [7] and [6].

5 Paradoxical results implied by the Axiom of Choice

Besides implying the existence of mathematical objects that cannot be explicitly defined, such as a well order of the reals, the acceptance of the Axiom of Choice also leads to some paradoxical or counter-intuitive results, like the ones we will present in this section. For example, the Axiom of Choice implies that it is impossible to extend the Lebesgue measure to all subsets of $\mathbb{R}$ since with the help of the Axiom one can easilt construct non-measurable sets. These results are one of the main arguments against the Axiom of Choice. Let's see them in detail.

5.1 Existence of non Lebesgue mesurable sets in $\mathbb{R}$

Definition 5.1.1. Given an open interval $I = (a, b) \subset \mathbb{R}$, we define its length by $l(I) = b - a$. The *Lebesgue outer measure* of a set $A \subseteq \mathbb{R}$ is defined by

$$\lambda^*(A) = \inf \left\{ \sum_{k=1}^{+\infty} l(I_k) : I_k,\ k \in \mathbb{N},\ \text{is an open interval and } A \subseteq \bigcup_{k=1}^{\infty} I_k \right\}$$

Definition 5.1.2. The *Lebesgue σ-algebra* is the collection of all sets E that satisfy

$$\text{For all subsets } A \in \mathbb{R},\ \ \lambda^*(A) = \lambda^*(A \cap E) + \lambda^*(A \cap \mathbb{R} \backslash E)$$

Definition 5.1.3. Let E be a set contained in the Lebesgue $\sigma-$algebra. The *Lebesgue measure* of E, $\lambda(E)$ is given by $\lambda(E) = \lambda^*(E)$.

The Lebesgue measure satisfies the following properties:

(a) for all $a, b \in \mathbb{R}$ with $a \leq b$, $\lambda((a, b)) = b - a$

(b) $\lambda(\emptyset) = 0$ and $\lambda(\mathbb{R}) = +\infty$

(c) λ is countably additive, that is, if $\{E_i\}_{i\in\mathbb{N}}$ is a collection of pairwise disjoint sets in the Lebesgue $\sigma-$algebra, then $\lambda(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{+\infty} \lambda(E_i)$

(d) λ is translation invariant, that is, if $a \in \mathbb{R}$, $E \subseteq \mathbb{R}$ and we let $E + a = \{x + a \mid x \in E\}$, then $\lambda(E) = \lambda(E + a)$

At first sight, it may seem logical to want a function with these properties to be extended to all subsets of the real numbers, that is, to be defined over $\mathcal{P}(\mathbb{R})$. It is in fact an important problem in analysis to extend the notion of length of an interval to more complicated subsets of $\mathbb{R}$. However, the Axiom of Choice guarantees that such function does not exist by implying the existence of certain subsets of the reals that are non Lebesgue measurable. One example of such sets is the Vitali set.

Definition 5.1.4. The *Vitali set* is a subset $V \subset [0, 1]$ such that for all $x \in \mathbb{R}$ there exists a unique $v \in V$ such that $x - v$ is a rational number.

Theorem 5.1.1. *$AC \Rightarrow$ The Vitali set exists.*

Proof. We define the following relation on $[0, 1]$ by: $x \sim y$ if and only if $x - y$ is a rational number. By the Axiom of Choice, there exists a choice function that chooses an element out of each equivalent class. Let V be the set containing all these elements. Clearly $V \subset [0, 1]$ and for each $x \in \mathbb{R}$, there exists a unique $v \in V$ such that $x - v$ is a rational number. Therefore, V is a Vitali set. $\square$

Theorem 5.1.2. *The Vitali set V is not Lebesgue measurable.*

Proof. For each $q \in \mathbb{Q}$ we define $V_q = \{v + q \mid v \in V\}$. The V_q, $q \in \mathbb{Q}$, yield a partition of $\mathbb{R}$ into countably many disjoint sets: $\mathbb{R} = \bigcup_{q \in \mathbb{Q}} V_q$. Suppose now that V is Lebesgue measurable. If $\lambda(V) = 0$, then $\lambda(\mathbb{R}) = 0$, since λ is countably additive and translation invariant. Hence, $\lambda > 0$. However, this is not possible either, since in that case we would obtain

$$\lambda([0, 2]) \geq \lambda\left(\bigcup_{q \in [0,1] \cap \mathbb{Q}} V_q \right) = \sum_{q \in [0,1] \cap \mathbb{Q}} \lambda(V_q) = \infty$$

since for each $q \in [0, 1] \cap \mathbb{Q}$, $\lambda(V_q) = \lambda(V) > 0$. Having reached a contradiction, we must conclude that V is non Lebesgue measurable. $\square$

However, if instead of the Axiom of Choice we use the Axiom of Countable Choice, then, the existence of Vitali sets cannot be proved. Note that when we use AC to choose an element x out of each equivalent class of the relation $\sim$ we are, in fact, dealing with an uncountable number of classes. Indeed, each class is of the form $\{x + q_i \mid q_i \in \mathbb{Q}\}$ for some x. Clearly, each class has a countable number of elements, since $\mathbb{Q}$ is countable. Therefore, the total number of classes must be uncountable, since its union must be the set of all real numbers contained in $[0, 1]$, which is uncountable.

In fact, a stronger version of CC can also hold without implying the existence of Vitali sets. In 1970, Robert Solovay proved, assuming the consistency of an inaccessible cardinal[11], the existence of a model where all the axioms of ZF hold, the axiom of Dependent Choice (DC) holds and all sets of reals are Lebesgue measurable.

Another construction of a set that is not Lebesgue measurable, this time in $\mathbb{R}^2$ can be defined using the fact that the set of the reals is well-orderable. Indeed, Sierpiński proved that no well-ordering of a non-empty set of the reals is Lebesgue measurable. A proof of this can be found in [9].

5.2 The Hausdorff Paradox

Definition 5.2.1. Let $A, B \subset \mathbb{R}^3$. We say that A and B are *congruent* ($A \cong B$) if one can be obtained from the other by translation, rotation and/or reflection. Equivalently, $A \cong B$ if there exists an isometry $f : A \to B$.

Lemma 5.2.1. *$\cong$ is an equivalence relation.*

Proof. $\cong$ is obviously reflexive, it is symmetric since each isometry $f : A \to B$ has an inverse function $f^{-1} : B \to A$, and it is transitive since if we have $f : A \to B$ and $g : B \to C$, we have $f \circ g : A \to C$. □

Hausdorff Paradox. *The sphere $S^2 = \{x \in \mathbb{R}^3 \mid ||x|| = 1\}$ can be descomposed into four disjoint sets A, B, C, D such that A, B, C and $B \cup C$ are all congruent and D is countable.*

[11] An uncountable cardinal κ is called *inaccessible* if it is regular and $2^\lambda < \kappa$, for all cardinals $\lambda < \kappa$. If the Axiom of Choice holds, an infinite cardinal κ is *regular* if and only if it cannot be expressed as the cardinal sum of a set of cardinality less than κ.

The paradox shows that no finitely additive and isometry invariant measure on a sphere can be defined for all its subsets. If there existed a measure μ satisfying said conditions, then, since $B \cong C$, we would have $\mu(B) = \mu(C)$, and since $A \cong B \cup C$, it would also be true $\mu(A) = \mu(B \cup C) = \mu(B) + \mu(C) = 2\mu(B)$. But $B \cong A$ and, hence, $\mu(B) = \mu(A) = 2\mu(B)$. Therefore, such a measure cannot exist. The role of the Axiom of Choice in the proof of the paradox is crucial and again leads to an undesirable consequence: the impossibility to define such a measure on S^2.

Proof. In order to prove the 'paradox' we must first consider two axes of rotation a_ϕ and a_ψ going through the center of S^2 and the following rotations about them:

$$\phi = \begin{pmatrix} -cos\theta & 0 & sin\theta \\ 0 & -1 & 0 \\ sin\theta & 0 & cos\theta \end{pmatrix} \qquad \psi = \begin{pmatrix} \frac{-1}{2} & \frac{\sqrt{3}}{2} & 0 \\ \frac{-\sqrt{3}}{2} & \frac{-1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Note that ϕ is a rotation of π radians about a_ϕ and ψ is a rotation of $\frac{2\pi}{3}$ radians about a_ψ. θ is the angle between the axes.
Let G be the group obtained by the free product of the groups $\{1, \phi\}$ and $\{1, \psi, \psi^2\}$, that is, the group of all transformations obtained by successive applications of ϕ and ψ an arbitrary finite number of times, with the specification $\phi^2 = \psi^3 = 1$. Now, we want to choose θ in such a way that different elements of G represent different rotations generated by ϕ and ψ. In order to achieve that, we have to choose a θ that satisfies that no element of G different than 1 represents the identity rotation. Hausdorff proved in 1914 that when we pick θ such that cos θ is transcendental, then such condition is satisfied.
It suffices, then, to pick θ such that cos θ is transcendental. Under this condition, different elements of G represent different rotations. Indeed, suppose $\sigma_1 ... \sigma_n = \tau_1 ... \tau_m$, where σ_1 and τ_1 are of the form ϕ or $\psi^{\pm 1}$, properly alternated. Then, if $\sigma_n = \tau_m$ we can simplify them from the right until the term on the left differs (it is not possible to end up with $\sigma_1 ... \sigma_r = 1$, with $r < n$, since that would contradict Hausdorff's proof). If, on the contrary, $\sigma_n \neq \tau_m$, then we have that $\sigma_1 ... \sigma_n \tau_m^{-1} ... \tau_1^{-1} = 1$ and σ_n cannot be simplified with τ_m, so we end up with a contradiction to Hausdorff's result.
Since for every $n \in \mathbb{N}$, the set of rotations expressed by a combination of n rotations of the form ϕ, ψ or $\psi^2 = \psi^{-1}$ is finite and G is the countable union of these finite sets, G is a countable group.
Now let's consider a partition of G into three subgroups G_A, G_B and G_C constructed by

recursion on the lengths of the elements of G. We let $1 \in G_A$, $\phi, \psi \in G_B$ and $\psi^2 \in G_C$ and then, for any $\alpha \in G$,

$$\text{If } \alpha \text{ ends with } \psi^{\pm 1}, \quad \begin{cases} \alpha\phi \in G_B & \text{if } \alpha \in G_A \\ \alpha\phi \in G_A & \text{if } \alpha \in G_B \\ \alpha\phi \in G_A & \text{if } \alpha \in G_C \end{cases}$$

$$\text{If } \alpha \text{ ends with } \phi, \quad \begin{cases} \alpha\psi \in G_B \text{ and } \alpha\psi^{-1} \in G_C & \text{if } \alpha \in G_A \\ \alpha\psi \in G_C \text{ and } \alpha\psi^{-1} \in G_A & \text{if } \alpha \in G_B \\ \alpha\psi \in G_A \text{ and } \alpha\psi^{-1} \in G_B & \text{if } \alpha \in G_C \end{cases}$$

Lemma 5.2.2. *G_A, G_B and G_C satisfy*

$$G_A\phi = G_B \cup G_C, \quad G_A\psi = G_B, \quad G_A\psi^2 = G_C.$$

Proof. Let $\alpha \in G_A$. If α ends with $\psi^{\pm 1}$, then by definition $\alpha\phi \in G_B$. If α ends with ϕ, that is, $\alpha = \beta\phi$, then $\beta \notin G_A$ (in that case, $\beta\phi \in G_B$) and thus $\beta\phi\phi = \phi \in G_B \cup G_C$. We have $G_A\phi \subset G_B \cup G_C$.
If $\alpha \in G_B$ ends with $\psi^{\pm 1}$, then $\alpha\phi \in G_A$ by definition, and hence $\alpha\phi\phi = \alpha \in G_A\phi$. If α ends with ϕ, that is, $\alpha = \beta\phi$, then $\beta \notin G_B$ (in that case, $\beta\phi \in G_A$) and $\beta \notin G_C$ for the same reason. Consequently, $\beta \in G_A$ and $\alpha = \beta\phi \in G_A\phi$. So we get $G_B \subset G_A\phi$. An analogous argument can be applied to G_C and obtain $G_C \subset G_A\phi$. Hence, $G_B \cup G_C \subset G_A\phi$ and we obtain $G_B \cup G_C = G_A\phi$.
The proof in the other cases is analogous. $\square$

Now let's consider the sphere S^2. Each element of G different from the identity leaves unchanged at most two points of S. We call D the set of points that remain unchanged by at least one element of G. Since G is countable, D is also countable.
For all $\alpha \in G$, α is a rotation defined on $S^2 \backslash D$: $\alpha : S^2 \backslash D \to S^2 \backslash D$. Indeed, if $x \in S^2 \backslash D$, then $\alpha(x) \in S \backslash D$. If $\alpha(x) \notin S^2 \backslash D$, then there would exist some $\beta \in G$ such that $\beta(\alpha(x)) = \alpha(x)$ and hence $\alpha^{-1}(\beta(\alpha(x))) = x$, with $\alpha - 1\beta\alpha \in G$, contradicting the fact that $x \in S^2 \backslash D$.
Let's consider the following relation $\sim$ on $S \backslash D$: $x \sim y$ if and only if, there exists some $\alpha \in G$ such that $\alpha(x) = y$. Since G is a group, $\sim$ is an equivalence relation. Now, proceeding in a similar way as in the proof of the existence of Vitali sets, we use the Axiom

of Choice to pick an element out of every class of equivalence defined by the relationship $\sim$ and let M be the set containing all these elements. We now define the following sets:

$$A = \{\alpha(x) \mid x \in M \text{ and } \alpha \in G_A\}$$
$$B = \{\alpha(x) \mid x \in M \text{ and } \alpha \in G_B\}$$
$$C = \{\alpha(x) \mid x \in M \text{ and } \alpha \in G_C\}$$

With these sets defined, we can prove that $A \cup B \cup C = S^2 \backslash D$. Indeed, for all $y \in G$, there exists some $x \in M$ and some $\alpha \in G$ such that $y = \alpha(x)$. Hence, $y \in A, B$ or C, depending on whether $\alpha \in G_A, G_B$ or G_C. The three sets are disjoint: suppose there exists some $x \in S^2 \backslash D$ such that $x \in A$ and $x \in B$. Then, there would be some $y \in M, \alpha \in G_A$ and $z \in M, \beta \in G_B$ such that $\alpha(y) = \beta(z)$, and then we would have $y = \alpha^{-1}(\beta(z))$, which implies $y \sim z$. In that case, by the definition of M, we would have $y = z$, and thus $y = \alpha^{-1}(\beta(y))$, which means that $y \in D$ or that $\alpha^{-1}\beta = 1$, and thus $\alpha = \beta$. Both cases are impossible, since $y \in M$ and $\alpha \in G_A$ and $\beta \in G_B$, which are disjoint groups. Hence, $A \cup B \cup C \cup D = S^2$.

Lemma 5.2.3. *A, B and C satisfy:*

$$\phi[A] = B \cup C, \qquad \psi[A] = B, \qquad \psi^2[A] = C.$$

Proof. $x \in \phi(A) \iff$ there exists some $y \in A$ such that $x = \phi(y) \iff x = \phi(\alpha(z))$ for some $z \in M$ and $\alpha \in G_A \iff x = \beta(z)$ for some $\beta \in G_A\phi = G_B \cup G_C$ and some $z \in M \iff x \in B \cup C$.
$x \in \psi(A) \iff$ there exists some $y \in A$ such that $x = \psi(y) \iff x = \psi(\alpha(z))$ for some $z \in M$ and $\alpha \in G_A \iff x = \beta(z)$ for some $\beta \in G_A\psi = G_B$ and some $z \in M \iff x \in B$.
$x \in \psi^2(A) \iff$ there exists some $y \in A$ such that $x = \psi^2(y) \iff x = \psi^2(\alpha(z))$ for some $z \in M$ and $\alpha \in G_A \iff x = \beta(z)$ for some $\beta \in G_A\psi^2 = G_C$ and some $z \in M \iff x \in C$. $\square$

Therefore, $A \cong B \cup C$, $A \cong B$ and $A \cong C$. We have obtained a partition of the sphere $S^2 = A \cup B \cup C \cup D$ that satisfies the desired conditions. $\square$

Again, we see that the proof breaks down if instead of the full strength of AC we only use CC. The situation is analogous as in 5.1. The Axiom of Choice is used to choose an element out of each equivalence class. The equivalence classes are of the form $M_x = \{y \mid \alpha(x) = y$

for some $\alpha \in G\}$, where x is the representative of the class given by the choice function. But, since G is countable, the set of the elements in a certain equivalence class is also countable. However, $S^2 \backslash D$ is uncountable and, hence, there must be an uncountable number of equivalence classes. Therefore, CC is not enough to carry out the argument.

5.3 The Banach-Tarski Paradox

Definition 5.3.1. Two sets X, Y are *equidecomposable* ($X \approx Y$) is there is a finite decomposition of X into disjoint sets, $X = X_1 \cup ... \cup X_n$ and a finite decomposition of Y into the same number of disjoint sets $Y = Y_1 \cup ... \cup Y_n$ such that $X_i \cong Y_i$ for all $i = 1, ..., n$.

Lemma 5.3.1. *$\approx$ satisfies the following properties:*

(a) *$\approx$ is an equivalence relation*

(b) *If X is the disjoint union of X_1 and X_2, Y is the disjoint union of Y_1 and Y_2, $X_1 \approx Y_1$ and $X_2 \approx Y_2$, then $X \approx Y$.*

(c) *If $X \subseteq Y \subseteq Z$ and $Z \approx X$, then $Z \approx Y$.*

Proof. (a) By definition and since $\cong$ is an equivalence relation, it can clearly be seen that $\approx$ is reflexive, symmetric and transitive.
(b) If $X_1 = X_{11} \cup ... \cup X_{1n}$, $Y_1 = Y_{11} \cup ... \cup Y_{1n}$ where $X_{1i} \cong Y_{1i}$ for all $i = 1, ...n$ and $X_2 = X_{21} \cup ... \cup X_{2m}$, $Y_{21} \cup ... \cup Y_{2m}$ with $X_{2j} \cong Y_{2j}$ for all $j = 1, ..., m$ then, clearly the decompositions $X = X_{11} \cup ... \cup X_{1n} \cup X_{21} \cup ... \cup X_{2m}$ and $Y = Y_{11} \cup ... \cup Y_{1n} \cup Y_{21} \cup ... \cup Y_{2m}$ satisfy the requirements for equidecomposability.
(c) Let $X = X_1 \cup ... \cup X_n$ and $Z = Z_1 \cup ... \cup Z_n$ such that $Z_i \cong X_i$ for all $i = 1, ...n$. Let $f_i : Z_i \to X_i$ be an isometry for each $i = 1, ...n$ and $f : Z \to X$, where $f = \bigcup_i f_i$. Now we define the following succession by recursion:

$$Z_0 = Z, \quad Z_1 = f[Z_0] = f[Z] = X, \quad Z_2 = f[Z_1], \quad ...$$
$$Y_0 = Y, \quad Y_1 = f[Y_0], \quad Y_2 = f[Y_1], \quad ...$$

If we let $W = \bigcup_{i=0}^{\infty} Z_i - Y_i$, then $f[W]$ and $Z - W$ are disjoint sets, $W \approx f[W]$ and we get $Z = W \cup (Z - W)$ and $Y = f[W] \cup (Z - W)$. Therefore, applying (b), we get $Z \approx Y$. $\square$

The Banach-Tarski Paradox. *There exists a decomposition of the closed ball* $\mathbb{B} = \{x \in \mathbb{R}^3 \mid ||x|| \leq 1\}$ *into disjoint sets* X *and* Y *such that* $\mathbb{B} \approx X$ *and* $\mathbb{B} \approx Y$*.*

Proof. In order to prove the paradox, we will use the same sets A, B, C and D defined in Hausdorff's Paradox Proof, as well as the rotations ϕ and ψ. Now, we define A' as the set containing all the radii of $\mathbb{B}$ with one of the extremes in the center of $\mathbb{B}$ ($\{0\}$) and the other one in some point of A. Equivalently, A' is the set of all points of $\mathbb{B}\backslash\{0\}$ such that its projection onto the surface belongs to A. In an analogous way, we define B', C' and D' with the sets B, C and D from S^2 respectively. We have

$$\mathbb{B} = A' \cup B' \cup C' \cup D' \cup \{0\}$$
$$\phi[A'] = B' \cup C', \qquad \psi[A'] = B', \qquad \psi^2[A'] = C'$$

and clearly,

$$A' \approx B' \approx C' \approx B' \cup C'$$

Now we define the sets $X = A' \cup D' \cup \{0\}$ and $Y = \mathbb{B} - X = B' \cup C'$.
By lemma 5.3.1.(b), we have that $B' \cup C' \approx A' \cup B' \cup C'$, since $B' \approx A'$ and $C' \approx B \cup C'$. Hence, we also have $A' \approx A' \cup B' \cup C'$ and then,

$$X = A' \cup D \cup \{0\} \approx A' \cup B' \cup C' \cup D' \cup \{0\} = \mathbb{B}.$$

It remains to see that $Y \approx \mathbb{B}$. To do that, we must first prove the following Lemma,

Lemma 5.3.2. *There exists some rotation* α *such that* $\alpha[D'] \subset A' \cup B' \cup C'$*.*
Proof. It suffices to prove that there exists a rotation α such that $\alpha[D] \subset A \cup B \cup C$. We fix an axis a_α such that it does not intersect with points belonging to the countable set D. For every angle θ, let α_θ be the rotation of angle θ about the chosen axis. Then, if we enumerate the points in D, and let $D = \{x_n \mid n \in \mathbb{N}\}$, we can define $X_n = \{\theta \in [0, \pi] \mid \alpha_\theta(x_n) \in D\}$. Since the application $X_n \to D$, where $\theta \mapsto \alpha_\theta(x_n)$ is injective, X_n is countable, and so it is $\bigcup\limits_n X_n$. Therefore, it is enough to choose some $\theta \in [0, \pi]\backslash \bigcup n X_n$, which is an uncountable set, in order to obtain a rotation α_θ that satisfies the conditions required. $\square$

Since $C' \approx A' \cup B' \cup C'$, and using the fact that there exists some $\alpha[D'] \subset A' \cup B' \cup C'$, and thus, clearly $\alpha[D'] \approx D'$, we obtain that there exists a set $R \subset C$, such that $R' \approx D'$. Now, if we pick a point $p \in C'\backslash R'$, we obtain

$$B' \cup R' \cup \{p\} \approx A' \cup D' \cup \{0\} = X \approx \mathbb{B}.$$

And since

$$B' \cup R' \cup \{p\} \subseteq Y \subseteq \mathbb{B}$$

applying Lemma 5.3.1.(c), we end up with $Y \approx U$, as desired. $\square$

Obviously, in the proof of the Banach-Tarski paradox the Axiom of Choice is also essential and unavoidable, since it depends on the proof of Hausdorff paradox, which in turn, relies on the Axiom. As well as with Hausdorff paradox, the proof cannot be completed if instead of using the full strength we only use the Axiom of Countable Choice or even DC, for the same reasons. In the Solovay model mentioned above, there is also no paradoxical decomposition of the sphere, so we conclude that ZF + the Axiom of Dependent Choice does not imply the Banach-Tarski paradox.
The Banack-Tarski paradox can be generalised for $n > 3$, and so the impossibility of defining a finitely-additive and translation invariant measure also apply to all $\mathbb{R}^n$, $n \geq 3$.

6 The Axiom of Choice and the ZF axiomatic system

6.1 The ZF axiomatic system

As we have briefly seen in section 1, the publication of Zermelo's proof of the Well Ordering Theorem using the Axiom of Choice was followed by his publication in 1908 of a list of principles or axioms that attempted to axiomatize set theory and provide a foundation for mathematics. Zermelo's axioms tried to capture the intuitive notion of what it means to be a set but at the same time avoided the paradoxes that arose from Frege's naïve set theory.

The most famous of those paradoxes is *Russell's paradox.* One of Frege's basic axioms stated the existence of a set $X = \{Y \mid \phi(Y)\}$, where ϕ is any property definable in the theory. In other words, it asserted that for every ϕ, there exists a set whose elements are exactly the sets that satisfy ϕ. This leads to contradiction, for it postulates the existence of the set $A = \{X \mid X \notin X\}$, that is, the set of all sets that do not belong to themselves. If such set exists, then we end up with the contradiction $A \in A \iff A \notin A$. Russell himself, along with Alfred N. Whitehead, proposed in their *Principia Mathematica* another axiomatic system, apparently exempted of contradictions, but rather complicated. Zermelo's axioms were simpler and also avoided Frege's paradoxes, so it prevailed amongst set theorists.

However, the original list of axioms given by Zermelo had still some shortcommings: it lacked a proper notion of "definable property", and was unable to prove the existence of certain cardinals and sets that mathematicians took for granted. With the contributions made during the following years by Fraenkel, Skolem and von Neumann, the axiomatic system was improved and completed to avoid these shortcommings. The result of those additions to the original Zermelo's list of axioms is what we know today as the Zermelo-Fraenkel axiomatic system or **ZF**.

There are different equivalent formulations of the axioms in ZF, here we give one of them along with its formalisation in the language of first-order logic for sets: formal logic with quantifiers ($\exists, \forall$), equality ($=$) and the non-logical symbol $\in$ that represents the binary relation expressing membership relation.

1. **Axiom of Extensionality.** If two sets have the same elements, then they are equal.

$$\forall x \, \forall y \, (\forall z \, (z \in x \leftrightarrow z \in y) \rightarrow x = y)$$

2. **Axiom of Pairing.** For any sets x, y there exists a set $\{x, y\}$ containing exactly x and y.

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

3. **Axiom Schema of Separation.**[12] For any set x and any definable property ϕ, i.e., any formula of the first order language of set theory which may contain free variables other than z in this case, here exists a set whose elements are the elements of x that satisfy the property ϕ.

$$\forall x \, \exists y \, \forall z \, (z \in y \leftrightarrow z \in x \wedge \phi(z))$$

   Note that this is not 'an' axiom, but an axiom schema, for it postulates the existence of different axioms given by different properties.

4. **Axiom of Union.** For every set x, there exists a set $y = \bigcup x$ whose elements are the elements of the elements of x.

$$\forall x \, \exists y \, \forall z \, (z \in y \leftrightarrow \exists u \, (z \in u \wedge u \in x))$$

5. **Power Set Axiom.** For every set x, there exists a set $\mathcal{P}(x)$ whose elements are all and only the subsets of x.

$$\forall x \, \exists y \, \forall z \, (z \in y \leftrightarrow \forall u \, (u \in z \rightarrow u \in x))$$

6. **Axiom of Infinity.** There exists an infinite set i.

[12]The Axiom Schema of Separation is essential for avoiding Russell's paradox: it allows only the construction of subsets satisfying a given property, but it does not postulate the existence of the set $\{x \mid \phi(x)\}$, which led to paradoxes.

$$\exists i\ (\exists x\ (x \in i \wedge \forall y\ \neg(y \in x)) \wedge$$
$$\forall z\ (z \in i \to \exists u\ (u \in i \wedge \forall w\ (w \in u \leftrightarrow u \in z \vee u = z))))$$

Note that the existence of the empty set can be derived from the Axiom of Infinity, which postulates the existence of a certain set and the Axiom Schema of Separation: $\emptyset = \{x \in I \mid \neg(x = x)\}$. In some formulations of the axioms of ZF, the Axiom of the Empty Set, which postulates the existence of the empty set, is given as a basic axiom. The Axiom Schema of Separation may also be seen to follow from the Axiom of Replacement together with the remaining axioms.

7. **Axiom Schema of Replacement.** For every set X, if f is a class function restricted to X, then there exists a set $Y = f[X]$.
A class function is different from the usual notion of function in that it is defined on all sets and the set of all sets is not a set. To formalise it we consider a formula in the language of first order logic $\phi(x, y, u)$ such that if $\phi(x, y, u) = \phi(x, z, u)$, then $y = z$ (this represents that $y = z$ is the image of x).

$$\forall u\ (\forall x\ \forall y\ \forall z((\phi(x, y, u) \wedge \phi(x, z, u)) \to y = z) \to$$
$$\forall X\ \exists Y\ \forall y\ (y \in Y \leftrightarrow \exists x\ (x \in X \wedge \phi(x, y, u))))$$

8. **Axiom of Foundation.** Every non-empty set x has a $\in$-minimal element. Equivalently, for every non-empty set x there exists a $y \in x$ such that $x \cap y = \emptyset$.

$$\forall x\ (\exists u(u \in x) \to \exists y\ (y \in x \wedge \neg\exists z\ (z \in y \wedge z \in x)))$$

If we add to the given list of axioms the Axiom of Choice then we get the Zermelo-Fraenkel axiomatic system with Choice, or **ZFC**.

ZFC is today the standard axiomatic system used in set theory and it provides the standard foundation for mathematics: from those axioms it is possible to derive all the theorems of usual mathematics; mathematical objects may be regarded as sets and all theorems can be proved from the ZFC axioms using the logic rules.

6.2 Consistency and independence of the Axiom of Choice from ZF

Two questions arise when we consider the relationship of the Axiom of Choice with the ZF axioms. Since the acceptance of the Axiom of Choice leads to some problematic and counterintuitive consequences, it would be reasonable to ask if it is *consistent* with the axioms of ZF. That is, if accepting the axioms in ZF along with AC can lead to some contradiction. Another important question is if the Axiom of Choice is *independent* from ZF: can it be proved from ZF or, on the contrary, it is necessary to postulate it as a new axiom?

Regarding the first question, Gödel proved in 1938 that if ZF is consistent, then so is ZFC. He achieved that by constructing a certain model, (*the constructible universe*, which is represented by the letter L), and showing that, assuming ZF, the structure $(L, \in)$ is a model of both ZF and AC, proving therefore its relative consistency. However, due to Gödel's second incompleteness theorem, it is impossible to prove the consistency of ZF in ZF, and the same happens with ZFC (assuming they are consistent).

With regards to the second question, the independence of the Axiom of Choice from the other axioms in ZF was proved by Paul Cohen in 1963. Assuming that ZF is consistent, he used the forcing technique to construct a model of ZF in which there is a set of reals that cannot be well-ordered, and thus a model where both ZF and the negation of AC hold. This showed that, provided that ZF is consistent, ZF + ¬C (ZF with the negation of the Axiom of Choice) is also consistent. Thus, the results of Gödel and Cohen, taken together, prove the independence of AC from ZF, i.e., ZF, if consistent, can neither prove nor refute AC.

6.3 The Axiom of Choice and the Generalised Continuum Hypothesis

We already mentioned in 4.1. that the Axiom of Choice implies the reformulation of the Continuum Hypothesis into the form $2^{\aleph_0} = \aleph_1$. This formulation can be generalised, into what is known as the Generalised Continuum Hypothesis, or GCH.

Generalised Continuum Hypothesis. For every ordinal α, $\quad 2^{\aleph_\alpha} = \aleph_{\alpha+1}$.

Sierpńnski proved in 1945 that ZF together with GCH implies the Axiom of Choice. Thus, the Generalised Continuum Hypothesis and the Axiom of Choice are not independent of each other in ZF, there does not exist a model of ZF where the former holds and the latter fails. The Generalised Continuum Hypothesis is, therefore, a stronger version of the Axiom of Choice, relative to ZF.

GCH has been proved independent from ZFC, that is, within the ZFC system it is impossible to prove GCH nor its negation. If GCH is accepted as an axiom, then, since the Axiom of Choice is implied by it, it would no longer be a basic axiom, but a consequence of the other axioms.

6.4 The Axiom of Determinacy: a possible substitute for the Axiom of Choice?

Due to its problematic character, the Axiom of Choice has always had some detractors, the number of which has progressively decreased since its origins. However, there has always been a search for better alternatives, i.e., alternatives that maintain some of its most used and widely accepted implications but that at the same time escape from its counterintuitive consequences. One of those alternatives is the Axiom of Determinacy, introduced by Jan Mycielski and Hugo Steinhaus.

To state the Axiom of Determinacy we must first consider, for each set $A \subseteq \omega^\omega$ the following associated game G_A: two players I and II successively choose natural numbers

$$\begin{array}{lccccccc} I: & a_0 & & a_1 & & a_2 & & ... \\ II: & & b_0 & & b_1 & & b_2 & ... \end{array}$$

If the produced sequence $\langle a_0, b_0, a_1, b_1, ...\rangle$ belongs to A, then player I wins; otherwise player II wins the game.

Definition 6.4.1. A *strategy* σ for player I is a function defined on finite sequences of numbers with values in ω. A strategy for player II is defined analogously.

Definition 6.4.2. A strategy σ is a *winning strategy* for player I in G_A if, whenever he chooses $a_n = \sigma(b_0, b_1, ...b_{n-1})$, then I wins the game, independently of the numbers that player II chooses. A winning strategy for player II is defined analogously.

Definition 6.4.3. The game G_A is *determined* if either player I or II has a winning strategy.

We may now formulate the Axiom of Determinacy:

Axiom of Determinacy (AD). *For every $A \subseteq \omega^\omega$, the game G_A is determined.*

The acceptance of this axiom contradicts the Axiom of Choice, since if we accept AC we can prove that there exists an $A \subseteq \omega^\omega$ such that G_A is not determined.

Theorem 6.4.1. *AC $\Rightarrow$ There exists a set $A \subseteq \omega^\omega$ such that G_A is not determined.*

Proof. Since AC holds, there exists an ordinal number γ such that $2^{\aleph_0} = \aleph_\gamma$. That is, the cardinality of the set of real numbers is equal to a certain aleph. We now construct, by transfinite recursion, the sets X_α and Y_α both belonging to ω^ω , monotonically increasing (if $\beta < \alpha$, then $X_\beta \subseteq X_\alpha$ and $Y_\beta \subseteq Y_\alpha$), disjoint for each α and such that $|X_\alpha| \leq |\alpha|$ and $|Y_\alpha| \leq |\alpha|$.
We well-order the set of different strategies $\{\sigma_\alpha \mid \alpha < 2^{\aleph_0} = \aleph_\gamma\}$[13], to construct the sets X_α and Y_α. To construct such sets, we proceed as follows: if α is a limit ordinal, then $X_\alpha = \bigcup\limits_{\beta<\alpha} X_\beta$, $Y_\alpha = \bigcup\limits_{\beta<\alpha} Y_\beta$. And if $\alpha = \beta + 1$ we do the following: for all $b = (b_0, b_1, ...)$, consider $\sigma_{\beta,I}[b] = (a_0, b_0, a_1, b_1, ...)$ as the set of sequences played in a game where $a_0, a_1, ...$ are chosen according to the strategy $\sigma_{\beta,I}$. The set of all $\sigma_{\beta,I}[b]$ has cardinality $2^{\aleph_0}$, since that is the cardinality of all the possible sequences played by player II. Since $|X_\beta| \leq |\beta| < 2^{\aleph_0} = \aleph_\gamma$ we can choose $b \in \omega^\omega$ such that $\sigma_{\beta,I}[b] \notin X_\beta$. We pick the least b in the well-ordering of ω^ω such that it satisfies the said condition and let $Y_{\beta+1} = Y_\beta \cup \{\sigma_{\beta,I}[b]\}$. Analogously, we consider for all $a = (a_0, a_1, ...)$ the set $\sigma_{\beta,II}[a]$ as the set of all sequences played in a game where $b_0, b_1, ...$ are chosen according to $\sigma_{\beta,II}$. In that case, there exists

[13] The set of all functions $f : \mathbb{N} \to \mathbb{N}$ has cardinality $2^{\aleph_0}$.

some $a \in \omega^\omega$ such that $\sigma_{\omega,II}[a] \notin Y_{\beta+1}$. We pick the least a satisfying this condition and let $X_{\beta+1} = X_\beta \cup \{\sigma_{\beta,II}[a]\}$.
We now define $A = \bigcup\limits_{\alpha<\aleph_\gamma} X_\alpha$. By the construction of the X_α, A defines a game G_A that is not determined since the sequences generated by the strategies $\sigma_{\beta,II}$ of player II belong to A and the sequences generated by the strategies $\sigma_{\beta,I}$ are not. □

Therefore, if we want to incorporate AD to ZF we must renounce AC, since AD is inconsistent with ZFC. But what are the consequences of AD that make it such an interesting alternative? Firstly, the acceptance of AD does not led to undesirable consequences such as the Banach-Tarski paradox and it implies that every set of real numbers is Lebesgue measurable. Besides, it implies the Axiom of Countable Choice and the Axiom of Denumerable choice restricted to subsets of $\mathbb{R}$, so we are still able to maintain some weaker choice principles and those statements implied by them. Furthermore, AD implies the consistency of ZF, hence, by Gödel's second incompleteness theorem, the relative consistency of ZF+AD from ZF cannot be proved. It also implies a weaker version of the Continuum Hypothesis, namely that every uncountable set of reals has the same cardinality as $\mathbb{R}$. However, as it implies the negation of the Axiom of Choice, it also implies the negation of the Generalised Continuum Hypothesis.

AD is also intimately related to the study of large cardinals, such as the *measurable cardinals*. An uncountable cardinal κ is measurable if there exists a κ-additive[14], non-trivial and 0-1 valued measure on $\mathcal{P}(\kappa)$. Whereas the Axiom of Choice implies that a measurable cardinal must be very large and its existence cannot be proved in ZFC, the Axiom of Determinacy implies that $\aleph_1$ is in fact, measurable.

Nevertheless, if we renounced the Axiom of Choice in favour of the Axiom of Determinacy, we would also lose the AC equivalent forms, some of which, as we have previously stated, are of vital importance for contemporary mathematics: the fact that every vector space has a basis, the Well-Ordering Theorem, Zorn's Lemma, Krull's Theorem, etc. Although AD is an attractive substitute for AC, it wouldn't be able to compensate for the loss of these nowadays basic mathematical principles.

[14] A measure μ on κ is κ-additive if for any $\lambda < \kappa$ and any sequence A_α with $\alpha < \lambda$ of pairwise disjoint subsets of κ, $\mu(\bigcup\limits_\alpha A_\alpha) = \sum\limits_\alpha \mu(A_\alpha)$

7 Final Remarks

As we have seen along this work, the implications of the Axiom of Choice in mathematics are many. It does not only imply crucial and widely used propositions in different branches of mathematics, but it is also equivalent to some of them. Not all these statements, though, require the full strength of the Axiom to be proved; in some cases a weaker version is sufficient.

Within set theory, by implying that every set can be well-ordered, the Axiom provides a good way of defining the cardinality of an infinite set, as well as a simplification of cardinal arithmetic. It also implies that every infinite set has a countable subset or that the countable union of countable sets is countable, statements that mathematicians take for granted, but where the Axiom of Countable Choice plays an essential role. In Algebra, Zorn's Lemma, another important equivalent of the Axiom of Choice, is widely used to prove the existence of maximal elements, for example in Krull's Theorem, or to prove that every vector space has a basis, and every field an algebraic closure. In topology, Zorn's Lemma also plays a role in the proof of Tychonov's theorem, which states that the product of compact topological spaces is compact. The Boolean Prime Ideal Theorem is necessary to prove the Hahn-Banach Theorem, which plays an important role in functional analysis, and the Compactness and Completeness Theorems for first order logic (in their most general form). Within graph theory, the Axiom is equivalent to the fact that every graph has a spanning tree. These implications of the Axiom of Choice are of great importance for their respective fields and many other statements and further developments on those fields depend on them. That is why the implications of the Axiom of Choice in mathematics cannot be underestimated.

However, as we have also seen, the Axiom of Choice has also a problematic side. It implies the existence of mathematical objects that cannot or are not explicitly defined, such as a well-order of the reals. Also, by implying the existence of sets of the reals that are not Lebesgue measurable, the Hausdorff paradox, and the Banach-Tarski paradox, it conflicts with our intuition. One of its consequences is the impossibility of defining a measure which is countably additive and translation invariant measuring all subsets of $\mathbb{R}$, or $\mathbb{R}^n$. As we have also seen, these consequences can be avoided if we are willing to accept only a weaker version of the Axiom, namely the Axiom of Dependent Choice.

Although having such *non-desirable* consequences, as we have seen in section 5 the Axiom of Choice is relatively consistent with the other axioms of ZF, which means that if ZF is consistent, then so is ZFC. This some implications of AC may not be desirable, or they may conflict with our intuition, but they don't imply any contradiction (as long as the other axioms in ZF don't imply one either). That means that we can confidently accept the Axiom to establish a solid base for the foundation of mathematics. The Axiom of Determinacy presented in section 6.4. avoids such consequences, but as we have seen, it is incompatible with the Axiom of Choice. However, if we replaced the Axiom of Choice by the Axiom of Determinacy we would loose a great number of mathematical theorems that are essential in contemporary mathematics and for which the Axiom of Choice is essential.

References

[1] K. Hrbacek and T. Jech, *Introduction to Set Theory.* M. Dekker, 1999.

[2] T. Jech, *The Axiom of Choice.* North-Holland, 1973.

[3] T. Jech, *Set Theory.* Springer, 2003.

[4] P. Schupp and R. Lyndon, *Combinatorial Group Theory.* Springer, 1977.

[5] J. S. Milne, "Fields and galois theory (v4.53)," 2017. Available at www.jmilne.org/math/.

[6] H. Schoutens, *The Use of Ultraproducts in Commutative Algebra.* Springer, 1999.

[7] A. Paseau, "Proofs of the compactness theorem," *History and Philosophy of Logic*, 2009.

[8] J. Bagaria, "Models of set theory." `https://www.icrea.cat/en/Web/ScientificStaff/jbagaria/more#researcher-nav`.

[9] W. Rudin, *Real and Complex Analysis.* McGraw-Hill Education, 1966.

[10] K. Devlin, *The Joy of Sets.* Springer, 1997.

[11] G. Moore, *Zermelo's Axiom of Choice.* Springer, 1982.

[12] H. Rubin and J. Rubin, *Equivalents of the Axiom of Choice.* North-Holland, 1970.

[13] H. Rubin and J. Rubin, *Equivalents of the Axiom of Choice II.* North-Holland, 1985.

[14] J. Ferreirós, *Labyrinth of Thought.* Birkhäuser Basel, 2007.

[15] P. Howard and J. Rubin, *Consequences of the Axiom of Choice.* American Mathematical Society, 1991.

[16] M. Hallot, "Zermelo's axiomatization of set theory." `https://plato.stanford.edu/archives/win2016/entries/zermelo-set-theory/`.

[17] C. Ivorra, "La paradoja de banach-tarski." `http://www.uv.es/~ivorra/Libros/Banach_Tarski.pdf`.