University of Warwick

MA262 Second Year Essay

Introduction to ZFC

Zermelo-Fraenkel Set Theory with Choice

Sean Tan

Contents

Introduction

Anyone who has done any form of rigorous mathematics understands the vital importance of the notion of a set, and the pivotal role they play in almost any construction one can think of. For example, functions are mappings between sets; groups and rings are sets with certain structure embedded into them; and the roots of real analysis lie in the intricacies of the real line $\mathbb{R}$. Even the colloquial numbers one learns since young can be rigorously constructed using sets, as we will briefly show in this paper. This motivates us to ensure a rigorous, contradiction-free set theory for mathematicians to use in their various works.

Naive Set Theory is an informal, but very intuitive and natural way of defining sets which uses the *Axiom of Unrestricted Comprehension*. Unrestricted comprehension states that for any property P, there exists a set X whose elements x are exactly those that satisfy P (we may say that $P(x)$ holds in this case). This seems like a fairly reasonable assumption, the intuition being that we should be able to group things certain objects together that are similar in some sense. In 1891, Georg Cantor, who is widely considered the founder of set theory, published a paper "Über eine elementare Frage der Mannigfaltigkeitslehre" [MV92, p.75-78] which contains what is widely known now as Cantor's Theorem.

Theorem 0.0.1 (Cantor's Theorem)**.** *Let X be a set and let $\mathcal{P}(X)$ denote the set of all subsets (power set) of X. Let $f : X \to \mathcal{P}(X)$ be a map. Then f is not a surjection.* ◊

Proof. Suppose f is a surjection. Let

$$A = \Big\{ x \in X \;\Big|\; x \notin f(x) \Big\}.$$

Since A is a subset of X, it follows that $A \in \mathcal{P}(X)$, so by surjectivity of f, there exists an $a \in X$ such that $f(a) = A$. However, this implies that

$$a \in A \iff a \in f(A) \iff a \notin A$$

which is a contradiction, so f cannot be a surjection. □

This may seem like a triviality, since the power set must obviously have more elements than the original set, but Cantor's Theorem doesn't just apply to finite sets – **it applies to infinite sets too**. Observe that if X is infinite, then $\mathcal{P}(X)$ is an even bigger infinite set. Repeatedly taking power sets this way allows us to construct an infinite hierarchy of infinities, each strictly larger than its predecessor in size. This invoked the study of infinite cardinal numbers, denoted $\aleph_0, \aleph_1, \aleph_2, \ldots$ to rank the sizes of different infinities. Although the study of this is very interesting, we will not be discussing this. What we will focus on is the catastrophe that arose from Cantor's Theorem. After staring at the proof long enough, one might realise that we can tweak the argument slightly and use Unrestricted Comprehension to define a set X such that

$$X = \Big\{ x \;\Big|\; x \notin x \Big\}$$

which implies that $X \in X \Leftrightarrow X \notin X$, a contradiction to Unrestricted Comprehension. This is widely known as Russell's Paradox,[1] which threatened the consistency of set theory and by extension, all of mathematics. This brings us to the main topic of this paper; to introduce the *Zermelo-Fraenkel Axioms (abbrevaited as ZF) with Choice (this extends the abbreviation to ZFC)* which first gives a rigorous framework of set theory, then proceeds to define every object in mathematics as some set.

[1] Although Bertrand Russell received credit for publishing this in 1901, both Zermelo and Cantor had already realised this contradiction a few years back.

1 Primitive Set Theory

We want to forget everything we know about maths and assume absolutely nothing[2] here, so a natural place to start is to define what a set truly is. Many textbooks define a set to be a *collection* of objects, possibly needing to obey certain rules, but clearly the word *collection* has not yet been defined, leading us back to the use of informal intuitive language, which we are trying our best to avoid. Unfortunately, there is no formal, abstract definition of what a set is, because we care more about how to interact with sets, rather than how to label them. If anything, sets are simply just any mathematical object constructed in ZF! This might be unsatisfying, but the reader should note that sets aren't the only victim of this. Vectors, for example, take many different forms in maths, physics and computer science, but fundamentally there is no 'proper' definition of what a vector is. Vectors are simply just elements of a vector space, which we already know, is not even necessarily the conventional vectors one might think of.

Another pill that might be even harder to swallow is the fact that we can't really define the membership property (denoted $\in$) as well, since it is technically a relation, which we can't define without basic set theory. It might be best to see the statements "$x \in y$", "$x \notin y$" not just as "x is or is not an element of y" but more abstractly as a *bridge* that connects first order logic to actual mathematics. We need all statements to be strictly either true or false without any other third possibility, and the membership property helps bring this to life. We shall now confidently assert the existence of the idea of sets (not the existence of sets themselves!) without a proper definition and assume now that for all sets x, y, we must strictly have that either $x \in y$ or $x \notin y$.

Definition 1.0.1. We define *equality* of x and y, denoted $x = y$ to be such that

$$x = y \implies \forall z \left[(z \in x \Leftrightarrow z \in y) \wedge (x \in z \Leftrightarrow y \in z) \right]$$

where $\wedge$ denotes the "AND" Logical (Boolean) Operator. We say that x is *equal* to y. $\lozenge$

Note that this is merely the definition according to first order logic with equality, and that we have only shown what equality implies, and not *how* to achieve equality. This will be dealt with shortly after.

Definition 1.0.2. Suppose that X, Y are sets.

- We say that X is a *subset* of Y if every element of X belongs to Y. We denote this as $X \subseteq Y$.
- We say that X is a *strict subset* of Y if X is a subset not equal to Y. We denote this as $X \subset Y$.

$\lozenge$

We have talked a lot about sets so far, but have not yet guaranteed the existence of even one. This leads us to the introduction of our very first axiom.

Axiom of Existence: There exists a set with no elements.

$$\exists x : \forall y \Big[y \notin x \Big]$$

This asserts the existence of a set, or possibly even more than one set. We want to confirm that the latter is not the case, in which we will need another axiom.

[2] If we are to be really pedantic we are still assuming first order logic with equality. Those who would like to look at this first can consult Steven G. Krantz' *Handbook of Logic and Proof Techniques for Computer Science* [Kra02]

Axiom of Extensionality: Two sets are equal if they are subsets of each other.

$$\forall x \forall y \, : \, \forall z \Big[z \in x \Leftrightarrow z \in y \Big] \implies x = y$$

Many might wonder why we don't make this the definition of equality instead. Note that without Definition 1.0.1, there would be no implication to the statement $x = y$. In other words, we would be able to reach "$x = y$" but would not be able to do anything further with it.

Proposition 1.0.3. *There exists only one set with no elements.* ◊

Proof. [HJ99, p.8] Suppose that x and y are both sets with no elements. For any z we see that the statement "$z \in x$" is obviously false by definition[3], and the same goes for y, so the statement

$$z \in x \iff z \in y$$

is true for all z because "F $\Leftrightarrow$ F" is a tautology[4], so $x = y$ by the Axiom of Extensionality. □

We have thus, shown that the set containing no elements is unique, motivating us to bestow upon it a name for distinction. We shall call this set the *empty set* and denote it as $\varnothing$. We now proceed to amend the Axiom of Unrestricted Comprehension so that we can construct sets without needing to worry about any contradictions like Russell's Paradox.

Axiom Schema of Restricted Comprehension:

For any set x, there exists a subset y of x such that ϕ holds for all elements in y.

$$\forall x \forall w_1, w_2, \ldots, w_n \exists y \forall z \, : \, z \in y \Leftrightarrow \Big[(z \in x) \wedge \phi \left(z, w_1, w_2, \ldots, w_n, x \right) \Big]$$

We say that this is an *axiom schema* and not just an ordinary axiom, since there are an infinite number of choices for ϕ, meaning that we have just asserted an infinite number of axioms. It should be clear that Russell's Paradox cannot be as easily invoked now, although we cannot rigorously prove that this contradiction can never occur in ZF yet; this will be done soon.

Proposition 1.0.4. *Let X be a set, ϕ a property and let Y be the set constructed through restricted comprehension using X and ϕ. Then Y is unique.* ◊

Proof. Let Y' be another set constructed through restricted comprehension the same way. We see that for any x

$$x \in Y \iff x \in X \wedge \phi(x) \iff x \in Y'$$

so by extensionality, $Y = Y'$. □

Any set constructed through restricted comprehension is thus unique, motivating us to denote the set as

$$\Big\{ x \in X \;\Big|\; \phi(x) \Big\}.$$

Definition 1.0.5. Let X, Y be sets. We define the *set-theoretic difference* of X and Y to be the set

$$X \setminus Y = \Big\{ x \in X \;\Big|\; x \notin Y \Big\}.$$

We may also denote this as $X - Y$. Its uniqueness follows from above. ◊

[3] Check the logical statement of the Axiom of Existence!

[4] That is, it is a true statement.

So far we've only been able to construct 'smaller' sets by taking subsets of already existing sets, and since there is no set of all sets, we will naturally need to find a way to *build* 'bigger' sets too. After all, the only set we know that truly exists so far is the empty set, and using restricted comprehension on that isn't very interesting! We now introduce new axioms that will allow us to build new sets from 'smaller' sets. Recall that $\vee$ denotes the "OR" Logical Operator.

Axiom of Pairing: For any two sets x and y, there exists a set that exactly contains x and y.

$$\forall x \forall y \exists z \forall w \ : \ w \in z \Leftrightarrow \Big[w = x \vee w = y \Big]$$

Axiom of Union: For any set x, there is a set y such that if $w \in x$, then all of the elements of w are in the set y.

$$\forall x \exists y \forall z \ : \ z \in y \Leftrightarrow \exists w \Big[w \in x \wedge z \in w \Big]$$

Using similiar arguments above, we can easily show that the sets constructed from these are unique as well, so for any set A we denote the *union* of A (the set constructed by the Axiom of Union) as $\bigcup A$, and the set constructed from Axiom of Pairing will be denoted, as expected, by $\{A, B\}$. We can now use restricted comprehension to define the *intersection* as

$$\bigcap A = \Big\{ x \in \bigcup A \ \Big| \ x \in a \text{ for every } a \in A \Big\}.$$

(*definition from [AM]*)

Definition 1.0.6. Let X, Y be sets. We define the *union* $X \cup Y$ and the *intersection* $X \cap Y$ of X and Y as

$$X \cup Y = \bigcup \Big\{ X, Y \Big\} \qquad \text{and} \qquad X \cap Y = \bigcap \Big\{ X, Y \Big\}$$

respectively. We also say that X and Y are *disjoint* if $X \cap Y = \varnothing$. $\Diamond$

Note that the set $\{X, Y\}$ indeed exists by the Axiom of Pairing, and the union/intersection follows from the Axiom of Union. We now move on to the introduction of the last two axioms of this section.

Axiom of Power Set: The power set of any set exists.

$$\forall x \exists y \forall z \ : \ z \in y \Leftrightarrow z \subseteq x$$

Axiom of Foundation: Every non-empty set x contains a member y such that x and y are disjoint.

$$\forall x \ : \ x \neq \varnothing \implies \exists y \Big[y \in x \wedge x \cap y = \varnothing \Big]$$

Although the power set axiom seems to be a fairly reasonable assertion, one might question why the Axiom of Foundation is necessary. There are in fact many important implications, but we shall only discuss the most elementary one.

Proposition 1.0.7. *There is no set x such that $x \in x$.* $\Diamond$

Proof. Let x be any set. Then the set $\{x\}$ exists by the Axiom of Pairing (by taking both sets to be the same x), and since $\{x\}$ is non-empty, the Axiom of Foundation tells us that there must be some member in $\{x\}$ that is disjoint from it. Clearly the only member in $\{x\}$ is x, so we deduce that x must be the member such that $x \cap \{x\} = \varnothing$. If x is such that $x \in x$, then $x \in x \cap \{x\}$, implying that $x \cap \{x\}$ is non-empty which is a contradiction, so we must have that $x \notin x$. $\square$

It thus follows that $\{x \in A \mid x \notin x\} = A$ for any set A and that $A \notin A$, getting rid of Russell's Paradox once and for all. We postpone the introduction of the remaining axioms for further sections where we can introduce them with proper context.

2 The Natural Numbers

2.1 Constructing the Naturals

We now explore a direct application of the axioms as we construct the most primitive and fundamental set of numbers, the *Natural Numbers*. How do we define them? How will we guarantee their existence? We can't just assume that they just exist without any justification, as we would then be implicitly adding more axioms to our construction. We will instead stick to our set-theoretic roots, and define numbers to be sets so that we can use the set operations we have constructed so far to manipulate them. The first candidate we will need to define is 0. The most natural candidate is obviously $\varnothing$, so we let $0 = \varnothing$. We can then continue on by defining *inductively*:

$$1 = \{0\} = \{\varnothing\}$$

$$2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\}$$

$$3 = \{0, 1, 2\} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$$

$$\vdots$$

so $n = \{0, 1, 2, \ldots, n-1\}$, where $n-1$ is the number before n.[5] Observe the power of the ZF Axioms in action. The existence of $\varnothing$ is axiomatic, $\{\varnothing\}$ exists by the Power Set Axiom (or Pairing) and we can construct $\{\varnothing, \{\varnothing\}\}$ by Axiom of Pairing again. In this way we can systematically construct an infinite number of sets that have some sort of ascending[6] pattern, and denote them with numbers $0, 1, 2, \ldots$ and so on, so numbers are just 'labels' for sets in our construction. We now want to give a proper definition to what the natural numbers are. Can we define a natural number to be of the form $n = \{0, 1, 2, \ldots, n-1\}$? No, as Hrbacek and Jech point out in *Introduction to Set Theory* [HJ99, p.40], we cannot define a natural number to be a set of all the smaller natural numbers, because such a "definition" would involve the very concept being defined. We shall instead define them in this manner:

$$1 = 0 \cup \{0\} = \varnothing \cup \{\varnothing\} = \{\varnothing\}$$

$$2 = 1 \cup \{1\} = \{\varnothing\} \cup \{\{\varnothing\}\} = \{\varnothing, \{\varnothing\}\}$$

$$3 = 2 \cup \{2\} = \{\varnothing, \{\varnothing\}\} \cup \{\{\varnothing, \{\varnothing\}\}\} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$$

$$\vdots$$

Defining them this way doesn't change anything and we will still have that $n = \{0, 1, 2, \ldots, n-1\}$, although we won't be able to prove this yet as we need the *principle of induction*, which will be proved shortly. Nevertheless, we save the proof of them being equivalent definitions for the addendum (see Proposition 4.1.1). This method of constructing the natural numbers is commonly known as the *von Neumann Construction* (see [Gol96]).

Definition 2.1.1. We define the *successor* of a "number" n to be the set

$$n + 1 = n \cup \{n\}.$$ $\Diamond$

Again, $n + 1$ serves as a simple notational convention to drive intuition.

[5] Note that we have not yet actually defined what subtraction is, and $n-1$ simply serves as a notational convention in this case.

[6] Whatever that means!

Definition 2.1.2. A set I is said to be *inductive* if

- $0 \in I$,
- $n \in I \implies n+1 \in I$. ◊

We now have the terminology we need to properly define the natural numbers.

Definition 2.1.3. We define the *natural numbers*, denoted $\mathbb{N}$ to be the set such that

$$\mathbb{N} = \bigcap \Big\{ I \subseteq S \;\Big|\; I \text{ is inductive} \Big\} = \bigcap_{I \text{ is inductive}} I$$

where S is an already existing inductive set. ◊

Having $I \subseteq S$ is perfectly justified here as we have that $I \in \mathcal{P}(S)$ in this case, so restricted comprehension is obeyed. However, we have run into a problem here. How do we know that an inductive set exists? If it doesn't, we are in big trouble, as it then follows that $\mathbb{N} = \varnothing$. As of right now, we can construct any finite subset $\{0, 1, \cdots, n\}$ of natural numbers as large as we like, but it seems that the structure of $\mathbb{N}$ (according to how we're trying to define it at least) possesses something special that these finite subsets don't — the notion of an infinity. This leads us to the introduction of another axiom that brings infinity to life in the world of mathematics.

Axiom of Infinity: An inductive set exists.

$$\exists x \;:\; \varnothing \in x \;\wedge\; \forall y \Big[y \in x \Rightarrow y + 1 \in x \Big]$$

Again quoting Karel Hrbacek and Thomas Jech's *Introduction to Set Theory* [HJ99, p.41] "Infinite sets are basic tools of modern mathematics and the essence of set theory. No contradiction resulting from their use has ever been discovered in spite of the enormous body of research founded on them. Therefore, we treat the Axiom of Infinity on par with our other axioms." It follows now, from the existence of an inductive set, that $\mathbb{N} = \{0, 1, 2, \cdots\}$ and goes on *forever*. We now state a well-known corollary of the Axiom of Infinity.

Corollary 2.1.4 (Principle of (Weak) Mathematical Induction)**.** *If $P(n)$ is a property such that*

- *$P(0)$ holds,*
- *for all $n \in \mathbb{N}$, $P(n) \implies P(n+1)$,*

then $P(n)$ holds true for all $n \in \mathbb{N}$. ◊

Proof. Let

$$N = \Big\{ n \in \mathbb{N} \;\Big|\; P(n) \text{ holds} \Big\}.$$

Clearly, $0 \in N$, and if $n \in N$, then $n+1 \in N$ by the initial hypothesis, so N is inductive. Therefore, $\mathbb{N} \subseteq N$ since $\mathbb{N}$ is the intersection of all inductive sets, and since $N \subseteq \mathbb{N}$ by definition, we deduce that $N = \mathbb{N}$, so $P(n)$ holds for all n. □

With the ZF Axioms that we have introduced so far, not only have we completely resolved Russell's Paradox, but we have also rigorously proven the existence and validity of all common set-theoretic operations, the natural numbers, and even induction. There are many more objects we can easily construct now, such as ordered tuples (x, y), inequalities, functions, equivalence relations, partial/strict orders, and much more. For the sake of length, we will not include this (rather exciting) construction in the main body of the paper, and save this for the addendum. We now assume knowledge of MA1 and MA2 modules from this point onwards, as it will be necessary when bringing up the next axiom.

3 Axiom of Choice

3.1 Motivation

We now explore what is known as arguably the most controversial (and thus optional) axiom in ZFC Set Theory. We will denote a collection of sets[7] as $\{X_i\}_{i\in I}$ (where I is some index set that could be uncountably infinite), and may even abbreviate this simply as $\{X_i\}$.

Definition 3.1.1. Let $\{X_i\}_{i\in I}$ be a collection of non-empty sets. The function $f : \{X_i\} \to \bigcup\{X_i\}$ is said to be a *choice function* if $f(X_i) \in X_i$ for all $i \in I$. $\Diamond$

A choice function takes an element out of each X_i and defines the chosen element to be $f(X_i)$. In simple terms, given a collection of sets, we simply choose one element out of each set, and call this a choice function. Now we ask the reader: **Does a choice function always exist?** If there are only finitely many X_i, we can use the fact that if X_i is non-empty, it must contain some element in it. We can then simply denote this element as x_i, for example, and repeat for every X_i, inductively constructing a choice function f where $f(X_i) = x_i$ (the full proof of this will be in the addendum). Consider now the case where I is countably infinite, or even uncountably infinite. If we try to do the same thing here, the process will continue indefinitely (as there are infinitely many X_i) so we will never be able to inductively complete the choice function.[8] So how do we prove that a choice function always exists in the infinite case? Before we answer this, we first look at some examples.

Example 3.1.2. Consider the collection of all non-empty subsets of $\mathbb{N}$, namely, $\mathcal{P}(\mathbb{N}) \setminus \{\varnothing\}$. Then we can define $f : \mathcal{P}(\mathbb{N}) \setminus \{\varnothing\} \to \mathbb{N}$ to be such that

$$f(A) = \min(A) \quad \forall A \subseteq \mathbb{N},\ A \neq \varnothing$$

where $\min(A)$ denotes the least element of A, so f is a choice function as $f(A) \in A$. $\Diamond$

The *Well-Ordering Principle*[9] tells us that every non-empty subset of $\mathbb{N}$ must have a least element, so we can just choose this minimum for each subset and construct a choice function this way. This shows that it is possible to find choice functions for infinite families of sets, if we can systematically choose an element out of each one. What about a choice function for $\mathcal{P}(\mathbb{R}) \setminus \{\varnothing\}$? We can't take the smallest element of every set now as the open interval $(0, 1)$ for example, has no least element. What about taking the median of the set? Unfortunately, the subset $[0, 1) \cup (1, 2]$ does not have a median. This suggests that no matter what "algorithm" we propose, there will always be a way of constructing a counter example subset of $\mathbb{R}$ in which the algorithm fails, i.e. there is no possible way to systematically choose a *singular* element from every non-empty subset of $\mathbb{R}$. Does this imply that a choice function need not exist when there are infinitely many sets? This is where the Axiom of Choice comes into play.

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

$$\forall x \ :\ \varnothing \notin x \implies \exists f : x \to \bigcup x \Big[\forall y (y \in x \implies f(y) \in y)\Big]$$

AC asserts the idea that we can always arbitrarily choose elements out of each set without needing to know what the elements are. In 1963, Paul Cohen showed that AC is logically independent of ZF [Coh64], meaning that it is impossible to prove or disprove AC using ZF. We will now discuss the most important consequences of AC, and demonstrate its amazing, yet terrifying power.

[7] A *collection* or a *family* of sets simply means a set containing sets. Apparently mathematicians don't really like the phrase "set of sets".

[8] Induction only works for finite cases, not infinite.

[9] A proof of this is at Theorem 3.3.3.

3.2 Zorn's Lemma

Zorn's Lemma will be the first implication of AC that we will discuss; it is an important theorem which says something about how sets that are bounded above (in some sense) must behave. Recall that a set X is *partially ordered* by a relation $\preceq$ if it is *reflexive, antisymmetric* and *transitive*.

Definition 3.2.1. Let X be partially ordered by $\preceq$ and let $a \in X$. We say that a is

- *maximum* (or a *greatest element*) if for all $x \in X$, $x \preceq a$,
- *maximal* if for any $x \in X$ we have $x \succeq a$, then $x = a$. ◊

One can view a maximal as a weaker notion of a maximum. The key difference that separates them is that a maximal does not necessarily need to be comparable[10] to *every element* in X. It just needs to be a maximum with respect to everything it is comparable to.

Definition 3.2.2. Let X be partially ordered by $\preceq$ and let $A \subseteq X$.

- We say that A is a *chain* of[11] X if it is a totally-ordered (that is, every element in A is comparable to one another).
- We say that $u \in X$ is an *upper bound* of A if $a \preceq u$ for all $a \in A$. ◊

Note in particular that the upper bound of A need not be in A.

Example 3.2.3. Consider the set $X = \{1, 2, 3, a, b, c\}$. We partially order X by $\preceq$ such that

$$1 \preceq 2 \preceq 3 \quad \text{and} \quad a \preceq b \preceq c$$

and $1, 2, 3$ are incomparable with a, b, c. Note that 3 is not a maximum, as we need $x \preceq 3$ for all $x \in X$, but it is not true that $a \preceq 3$ as they are incomparable. However, 3 is in fact, a maximal, as it is indeed the case that if x is such that $x \succeq 3$ (which implies that x must be comparable with 3), then x must be equal to 3. The same applies for c. Examples of chains of X are $\{a, b\}$ and $\{1, 3\}$, but $\{b, 3\}$ is not a chain. The subset $\{a, b\}$ has upper bound c, even though it is not in the subset. ◊

We are now equipped with the terminology to state Zorn's Lemma.

Theorem 3.2.4 (Zorn's Lemma)**.** *Let X be non-empty and partially ordered. If every chain of X has an upper bound, then X must have a maximal element.* ◊

If X is finite, then we can employ a simple induction argument to find a maximal[12]. If X is totally-ordered, then X is a chain in its own right, so by the hypothesis, it must have an upper bound in X, and this must clearly be a maximum (maximal) element! The usefulness of Zorn's Lemma appears when X is both infinite, and not totally-ordered. We will present the proof of Zorn's Lemma from Paul Halmos' *Naive Set Theory* [Hal60, p.64], which is quite long (3 pages), but the overall gameplan is simple: we use the Axiom of Choice to pick out elements from X and construct a *maximal chain* of X. That is, a chain of X that can't have more elements added to it, without breaking the fact that it is a chain. This maximal chain must then contain a maximal element in X.

[10] Elements x and y are *comparable* (with respect to $\prec$) if either $x \prec y$, $x \succ y$, or $x = y$ must hold.

[11] Most books use the phrase 'chain in X', which I personally find slightly misleading as it could imply that the chain is a member of the set, and not a subset of it, so we will stick with the phrase "chain of X".

[12] This is analagous to the fact that choice functions for finite families of sets always exist.

In order to find a maximal chain of X, we start by defining a new set $\mathcal{X}$ to be

$$\mathcal{X} = \Big\{ A \subseteq X \;\Big|\; A \text{ is a chain of } X \Big\}$$

and let $\mathcal{X}$ be partially ordered by set inclusion. Note that a maximal of $\mathcal{X}$ is a maximal chain of X described above. We now need to show that finding a maximal in $\mathcal{X}$ naturally realises a maximal in X. We will in fact, show that $\mathcal{X}$ satisfies all the hypotheses of Zorn's Lemma.

Lemma 3.2.5. *Let X and $\mathcal{X}$ be defined as above. Then*

(a) $\mathcal{X}$ is non-empty,

(b) if $\mathcal{C}$ is a chain of $\mathcal{X}$, then $\bigcup \mathcal{C}$ is an upper bound of $\mathcal{C}$ (with respect to inclusion), and

(c) if $A \in \mathcal{X}$ is maximal, then X has a maximal element. ◊

Proof. (a) The empty set $\varnothing$ being a chain of $\mathcal{X}$ is a vacuous truth, so $\varnothing \in \mathcal{X}$, and $\mathcal{X}$ is non-empty.

(b) We just need to show that $\bigcup \mathcal{C}$ is in $\mathcal{X}$, since we already know that $C \subseteq \bigcup \mathcal{C}$ for all $C \in \mathcal{C}$ by definition. It suffices to show that $\bigcup \mathcal{C}$ is a chain of X. If $x, y \in \bigcup \mathcal{C}$, then there must exist $C_x, C_y \in \mathcal{C}$ that contain x and y respectively. Since $\mathcal{C}$ is a chain of $\mathcal{X}$, we must have either $C_x \subseteq C_y$ or $C_y \subseteq C_x$. WLOG, we can let $C_x \subseteq C_y$, so both $x, y \in C_y$ and since C_y is a chain of X, we conclude that x, y must be comparable so $\bigcup \mathcal{C}$ is a chain of X.

(c) If A is a maximal element of $\mathcal{X}$, then it is a chain of X, so under the hypothesis of Zorn's Lemma, it must have an upper bound in X. We denote this as $u \in X$, and we will show that u is maximal in X. Suppose there was some $z \in X$ such that $z \succeq u$. For all $a \in A$, we have that

$$a \preceq u \text{ and } u \preceq z \implies a \preceq z$$

so z is also an upper bound of A, and since it is comparable to every element in A, we conclude that $A \cup \{z\}$ is a chain. However, $A \subseteq A \cup \{z\}$ and A is maximal in $\mathcal{X}$, so $A = A \cup \{z\}$ and thus $z \in A$. Since $z \preceq u$ as u is an upper bound of A, we conclude by antisymmetry, that $z = u$.

□

It thus suffices to consider the structure of $\mathcal{X}$, as it analagous to structure of X, meaning that we just need to find a maximal in $\mathcal{X}$ to prove Zorn's Lemma. Let $f : \mathcal{P}(X) \setminus \{\varnothing\} \to X$ be a choice function[13]. For every $A \in \mathcal{X}$, define

$$\hat{A} = \{z \in X \setminus A \mid A \cup \{z\} \in \mathcal{X}\}.$$

That is, the set $\hat{A}$ describes all non-trivial[14] members of X such that when added to A, the resulting set is still a chain of X. Now, define $g : \mathcal{X} \to \mathcal{X}$ to be such that

$$g(A) = \begin{cases} A \cup \{f(\hat{A})\} & \text{if } \hat{A} \neq \varnothing, \\ A & \text{if } \hat{A} = \varnothing. \end{cases}$$

This function g takes a chain of X, and adds a new arbitrary element to it such that the output is still a chain. It thus suffices to find $A \in \mathcal{X}$ such that $g(A) = A$, as it then follows that A is a maximal of X. We start off by introducing some new terminology.

[13] Here is where we invoke the Axiom of Choice.

[14] Members that are not in A.

Definition 3.2.6. We say that $\tau \subseteq \mathcal{X}$ is a *tower* if it satisfies the following:

- $\varnothing \in \tau$,
- if $A \in \tau$, $g(A) \in \tau$, and
- if $\mathcal{C}$ is a chain of τ, then $\bigcup \mathcal{C} \in \tau$. ◊

By definition, $\mathcal{X}$ is a tower, so towers indeed exist. So we can let

$$\tau_0 = \bigcap_{\tau \text{ is a tower}} \tau \ .$$

Note that τ_0 is also a tower and is in fact the smallest one. This might be a bit much to absorb in at first glance, but it will in fact be the case that the only tower we actually need to consider is τ_0. For the sake of clarity, let us try to construct this smallest tower. Firstly, we know that $\varnothing \in \tau_0$, so $g(\varnothing) \in \tau_0$. Recalling our definition of $\hat{A}$ from above, we may then note that

$$\hat{\varnothing} = \{z \in X \setminus \varnothing \mid \varnothing \cup \{z\} \in \mathcal{X}\} = \{z \in X \mid \{z\} \in \mathcal{X}\} = X$$

so $g(\varnothing) = \varnothing \cup \{f(\hat{\varnothing} \setminus \varnothing)\} = \{f(X)\}$. We must now arbitrarily take an element from X to be in the tower τ_0. Recall now, Example 3.2.3 where $X = \{1, 2, 3, a, b, c\}$. If we take $f(X) = 1$ for example, then $g(\varnothing) = \{1\}$, and from repeated iterations of g, we could end up with something like

$$g(\{1\}) = \{1, 3\} \implies g(\{1, 3\}) = \{1, 2, 3\} \implies g(\{1, 2, 3\}) = \{1, 2, 3\}$$

so $\tau_0 = \{\varnothing, \{1\}, \{1, 3\}, \{1, 2, 3\}\}$. Observe now that τ_0 is in fact, a maximal chain of $\mathcal{X}$ (with respect to inclusion), and that the union of this set, $\bigcup \tau_0 = \{1, 2, 3\}$ is maximal in $\mathcal{X}$, and therefore a maximal chain of X. This example hopefully convinces the reader that τ_0 always contains a maximal element of $\mathcal{X}$, and the union of τ_0 gives us a maximal element of $\mathcal{X}$, which is what we want. In summary:

$$\text{maximal chain of } \mathcal{X} \implies \text{maximal element in } \mathcal{X} = \text{ maximal chain of } X \implies \text{maximal element in } X$$

which implies Zorn's Lemma. Since X is finite in this case, we can inductively construct τ_0. However, if X is infinite, then the job is not so easy, as a maximal chain of X could be infinitely large, meaning that we can't inductively build it. This is where the Axiom of Choice comes into play. We have set up our choice function in a way such that we can keep adding elements from $\hat{A}$ to A by applying g, letting us somehow "induct to infinity". This is why AC is so powerful. It essentially allows us bypass any finite restrictions that induction has. We now must confirm that τ_0 is in indeed a chain of $\mathcal{X}$, and to do this, we first need to show that if $C \in \tau_0$ is comparable (that is, comparable with every element in τ_0), then $g(C) \in \tau_0$ is also comparable (as it should!). We require some basic lemmas first.

Lemma 3.2.7. *Let $C \in \tau_0$ be comparable. If $A \in \tau_0$ is such that $A \subset C$, then $g(A) \subseteq C$.* ◊

Proof. Suppose, for a contradiction, that $A \in \tau_0$ is such that $A \subset C$, but $g(A) \not\subseteq C$. Note that $g(A) \in \tau_0$ by definition of a tower, so comparability of C in τ_0 tells us that we must have $C \subset g(A)$. However, it then follows that

$$A \subset C \subset g(A)$$

which implies that $g(A)$ has at least 2 more elements than A, contradicting how g is defined. Therefore, no such A can exist.

□

Lemma 3.2.8. *Fix a comparable element $C \in \tau_0$ and define*

$$\mathcal{U} = \{A \in \tau_0 \mid A \subseteq C \text{ or } g(C) \subseteq A\}.$$

Then $\mathcal{U} = \tau_0$. ◊

Proof. Since $\mathcal{U} \subseteq \tau_0$ by definition, it suffices to show that $\tau_0 \subseteq \mathcal{U}$. Recalling that τ_0 is the intersection of all towers, it suffices to show that $\mathcal{U}$ is a tower. Note that $\varnothing \subseteq C$, so $\varnothing \in \mathcal{U}$.

Choose any $A \in \mathcal{U}$. If $g(C) \subseteq A$, then $g(C) \subseteq g(A)$ so $g(A) \in \mathcal{U}$. If $A \subseteq C$, we have 2 cases. If $A \subset C$ then $g(A) \subseteq C$ by the previous lemma, and if $A = C$ then $g(A) = g(C)$, so either way $g(A) \in \mathcal{U}$.

Suppose $\mathcal{C}$ is a chain of $\mathcal{U}$. We consider 2 cases: If every $A \in \mathcal{C}$ is a subset of C, then $\bigcup \mathcal{C} \subseteq C$ so $\bigcup \mathcal{C} \in \mathcal{U}$. Suppose now that there exists some $A \in \mathcal{C}$ that is not a subset of C. Note that $A \in \mathcal{C} \subseteq \mathcal{U} \subseteq \tau_0$, so comparability of C in τ_0 tells us that

$$C \subset A \implies g(C) \subseteq A \implies g(C) \subseteq \bigcup \mathcal{C}$$

so $\bigcup \mathcal{C} \in \mathcal{U}$, and it follows that $\mathcal{U}$ is a tower. □

Corollary 3.2.9. *If $C \in \tau_0$ is comparable then $g(C)$ is comparable.* ◊

Proof. Let $A \in \tau_0$, so either $A \subseteq C$ or $g(C) \subseteq A$ by definition of $\mathcal{U}$. In the latter case $g(C)$ is comparable so we are done. If $A \subseteq C$, then $A \subseteq g(C)$. □

We can now show that every element in τ_0 is comparable, as it should.

Lemma 3.2.10. *As expected, τ_0 is indeed a chain of $\mathcal{X}$.* ◊

Proof. Let

$$\mathcal{V} = \{A \in \tau_0 \mid A \text{ is comparable in } \mathcal{X}\}.$$

It suffices to show that $V = \tau_0$. In other words, that $\mathcal{V}$ is a tower. The empty set $\varnothing$ is obviously comparable with every set (since $\varnothing \subseteq A$ for any A), so $\varnothing \in \mathcal{V}$. Suppose $A \in \mathcal{V}$, then $g(A) \in \mathcal{V}$ by Corollary 3.2.9. If $\mathcal{C}$ is a chain of $\mathcal{V}$, we need to show that $\bigcup \mathcal{C}$ is comparable with every $A \in \mathcal{V}$. We consider 2 cases. If every $C \in \mathcal{C}$ is a subset of A then $\bigcup \mathcal{C} \subseteq A$. If there is a $C \in \mathcal{C}$ that is not a subset of A, then by comparability of A we deduce that $A \subset C \Rightarrow A \subset \bigcup \mathcal{C}$. So $\mathcal{V}$ is a tower. □

Now that we have shown that τ_0 is a chain of $\mathcal{X}$, we can easily show that $\bigcup \tau_0$ is maximal in $\mathcal{X}$.

Proof of Zorn's Lemma. Let $A = \bigcup \tau_0$ be an upper bound of τ_0 (with respect to inclusion). Note that τ_0 is a chain of τ_0 (itself), since τ_0 is a chain of $\mathcal{X}$, so by definition of a tower, its union is also in τ_0, so $A \in \tau_0$. Therefore $g(A) \in \tau_0$, and since A is an upper bound of τ_0, we have that $g(A) \subseteq A$. By definition of g, $A \subseteq g(A)$, so we finally deduce that $g(A) = A$, thus proving Zorn's Lemma. □

After such a long proof, one must certainly be craving for some examples. We now present the most famous application of Zorn's Lemma: Let V be a vector space. Does V have a basis? The obvious answer seems to be yes, and if V is finite-dimensional, we have learned in linear algebra that there are many ways to show the existence of a basis, using *sifting*[15] or some inductive argument.

[15] Go through the list of vectors and remove the vectors linearly dependent of the previous ones, inductively constructing a linearly independent set.

Suppose now that V is *infinite-dimensional*. That is, V is only spanned by an infinite number of elements. Must a basis of V necessarily exist? Let us look at some examples first.

1. Let $V = \mathbb{R}[x]$ be the vector space of real polynomials over the field $\mathbb{R}$, so elements in V are of the form $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$, where every $a_i \in \mathbb{R}$. It should be clear that a basis for V is $\{1, x, x^2, \cdots\}$, so this shows that bases *can* exist for infinite-dimensional vector spaces.

2. Suppose now that $V = \mathbb{R}$ is the vector space of real numbers over the field of rationals $\mathbb{Q}$. We can list many linearly independent sets, such as $\{\sqrt{2}, \sqrt{3}, \sqrt{5}, \ldots\}$ or $\{\pi, \pi^2, \pi^3, \ldots\}$, but clearly none of these span $\mathbb{R}$. The set of all irrationals (and 1), namely the set $(\mathbb{R} \setminus \mathbb{Q}) \cup \{1\}$ indeed spans $\mathbb{R}$, but is not a linearly independent set, as both π and $\pi + 1$ are in it, for example. One should hopefully see that there is no easy way to find a basis for this vector space, possible hinting at the possibility that a basis might not even exist.

3. Consider now, the vector space V of functions $f : \mathbb{R} \to \mathbb{R}$ over the field $\mathbb{R}$. There are the standard well-known functions (polynomials, exponentials, trigonometric), crazy piecewise functions, and even infinite series of functions! This set is so overwhelmingly large, one must be tempted to say that a basis certainly cannot exist for such as vector space. Does this mean that bases *need not exist* for infinite-dimensional vector spaces? Zorn's Lemma (and in turn, the Axiom of Choice) can be used to show something about vector spaces that is truly shocking.

Theorem 3.2.11. *Every vector space has a basis.* ◊

Proof. [Joh09] Let V be a vector space. Define $\mathcal{L}$ to be a collection of sets such that

$$\mathcal{L} = \Big\{ \{\mathbf{v}_i\} \subseteq V \;\Big|\; \{\mathbf{v}_i\} \text{ is a linearly independent set} \Big\}$$

and partially ordered $\mathcal{L}$ by inclusion. We will use Zorn's Lemma to show that $\mathcal{L}$ has a maximal element, that is, a maximal linearly independent set of vectors, which we will show to be in fact, a basis of V. Note that $\varnothing \in \mathcal{L}$ by definition[16], so $\mathcal{L}$ is non-empty.

Let $\mathcal{C}$ be a chain of $\mathcal{L}$, we will show that $\bigcup \mathcal{C}$ is an upper bound of $\mathcal{C}$. It suffices to show that $\bigcup \mathcal{C} \in \mathcal{L}$, in other words, that $\bigcup \mathcal{C}$ is a linearly independent set. Suppose not, then there exists a finite subset of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\} \subseteq \bigcup \mathcal{C}$ that have some non-trivial linear combination of the zero vector. Note that for every $\mathbf{v}_i$, there must exist some $S_i \in \mathcal{C}$ such that $\mathbf{v}_i \in \mathcal{S}_i$, and since $\mathcal{C}$ is totally ordered, it can then be shown[17] that there must exist some S_n that contains all the $\mathbf{v}_i$. This implies in particular, that S_n is not a linearly independent set, contradicting the structure of $\mathcal{L}$. We thus conclude that $\bigcup \mathcal{C} \in \mathcal{L}$, meaning that all the conditions of Zorn's Lemma are satisfied, so $\mathcal{L}$ must have a maximal linearly independent subset of V, which we shall denote as $\mathcal{B}$.

Suppose $\mathcal{B}$ does not span V. Then, there exists some $\mathbf{v} \in V$ that is not a linear combination of the vectors in $\mathcal{B}$, meaning that the set $\mathcal{B} \cup \{\mathbf{v}\}$ is a linearly independent set and thus a member of $\mathcal{L}$, which contradicts the maximality of $\mathcal{B}$, so it is indeed a basis. □

No matter how crazy and unfathomable your vector space is, the Axiom of Choice allows us to navigate through an uncountably infinite family of linearly independent sets, and assert that a basis must **always** exist.

[16] The statement "All vectors in $\varnothing$ are linearly independent" is a vacuous truth.

[17] look back at the proof of Lemma 3.2.5, and note that we can extend the argument used there to inductively find this S_n, since there is only a finite number of v_i.

3.3 Well-Ordering Theorem

Although Zorn's Lemma is highly respected as a powerful tool in mathematics, there is another famous implication of AC (which we will show through Zorn's Lemma) that has been regarded as controversial. Recall that X has a least element a if for all $x \in X$, $a \preceq x$.

Definition 3.3.1. Let X be partially ordered. We say that X is *well-ordered* if every non-empty subset of X has a least element. ◊

This is in fact stronger than being totally-ordered, as we will now show.

Proposition 3.3.2. *Every well-ordered set is totally-ordered.* ◊

Proof. Let X be well-ordered (by $\preceq$) and pick any $x, y \in X$. Then $\{x, y\}$ is a subset of X, so it must have a least element. If x is the least element, then $x \preceq y$. Similiar if y is the least element, so x and y are comparable, and X is totally-ordered. □

We now use *Strong Induction* (see Proposition 4.1.2) to prove a well-known property of $\mathbb{N}$.

Theorem 3.3.3 (Well-Ordering Principle)**.** $\mathbb{N}$ *is well-ordered with respect to* $\leq$. ◊

Proof. Suppose A is a subset of $\mathbb{N}$ that isn't well-ordered. It suffices to show that $A = \varnothing$. If $0 \in A$, then A is well-ordered as for all $n \in A \subseteq \mathbb{N}$, we have that $n \geq 0$, so we must have that $0 \notin A$. Suppose now that natural numbers $\{0, 1, \ldots, n-1\}$ are all not in A for some fixed $n \in \mathbb{N}$. If $n \in A$, then n is a least element of A, since all naturals smaller than n are not in A, so we deduce that $n \notin A$. It follows from strong induction that $n \notin A$ for all $n \in \mathbb{N}$. In other words, $A = \varnothing$. □

Note that this also shows us that $\mathbb{N}$ is totally-ordered. With AC we can in fact show that **any set can be well-ordered**. This is called the Well-Ordering Theorem and should not be confused with the Well-Ordering Principle[18]. We shall prove this using the construction from Paul Halmos' *Naive Set Theory* [Hal60, p.67-69].

Definition 3.3.4. Let X be partially ordered and let $a \in X$. We say that A is an *initial segment* of X if there exists an $a \in X$ such that

$$A = \Big\{ x \in X \;\Big|\; x \preceq a \Big\}.$$

If X and A are well-ordered (by the same ordering), then we say that X is a *continuation* of A. ◊

Continuation is a partial ordering stronger than set inclusion, in the sense that every set must have a least element. Note that a family of sets is a chain with respect to continuation if every set is a continuation of another.

Lemma 3.3.5. *Let $\mathcal{C}$ be a chain with respect to continuation. Then $\bigcup \mathcal{C}$ has a unique well-ordering such that it is a continuation of every element in $\mathcal{C}$.* ◊

Proof. Choose any $x, y \in \bigcup \mathcal{C}$. Then, there must exist $C_x, C_y \in \mathcal{C}$ that contain x and y respectively, and since one must be a continuation of the other, we can conclude that one of C_x or C_y contains both x and y. We now copy the well-ordering of this set as our ordering for $\bigcup \mathcal{C}$. It should be clear that this is a partial order, so we now just need to show that it is a well-ordering in $\bigcup \mathcal{C}$.

[18] Although many still miscall the Well-Ordering Theorem as the Well-Ordering Principle, which should be noted to the reader.

Suppose $A \subseteq \bigcup \mathcal{C}$ is non-empty. The members of A can only be from sets in $\mathcal{C}$, so there exists some $B \in \mathcal{C}$ such that $A \cap B \neq \varnothing$. Note that $A \cap B \subseteq B$ and B is well-ordered so $A \cap B$ has a least element, which we shall denote as u. We want to show that u is a least element of A. Suppose not, then there exists some $x \in A$ such that $x \prec u$ (A is totally-ordered so the least element must be comparable with u). Note that $x \notin B$ since if x was in B, then $x \in A \cap B$, contradicting the minimality of u. Let $C \in \mathcal{C}$ be such that $x \in C$. Then $C \not\subseteq B$, and since $\mathcal{C}$ is totally-ordered with respect to continuation, we must have that $B \subseteq C$, but this implies that $x \in B$, which is a contradiction, so x cannot exist and it follows that $\bigcup \mathcal{C}$ is well-ordered.

We now just need to show that the well-ordering is unique. Observe that if we took any other well-ordering that is not the ordering in the sets of $\mathcal{C}$, then $\bigcup \mathcal{C}$ fails to be a continuation of all the members in $\mathcal{C}$, so we must choose this specific ordering. □

Those who thoroughly understand the proofs of Lemma 3.2.5 and Theorem 3.2.11 should see where we are going with this. Everytime we wish to invoke Zorn's Lemma, we want to show that every chain $\mathcal{C}$ is bounded above. We almost always do this by considering its union $\bigcup \mathcal{C}$. Once we have found a maximal, we use some contradiction argument on it's maximality to deduce the properties that we are looking for.

Theorem 3.3.6 (Well-Ordering Theorem)**.** *Every non-empty set can be well-ordered.* ◊

Proof. Fix a non-empty set X and let

$$\mathcal{W} = \left\{ A \subseteq X \;\middle|\; A \text{ can be well-ordered} \right\}$$

which we partially order by continuation. For any chain $\mathcal{C}$ of $\mathcal{W}$, we note that $\bigcup \mathcal{C}$ can be uniquely well-ordered such that it is a continuation of $\mathcal{C}$ by Lemma 3.3.5, so it must be in $\mathcal{W}$ and also an upper bound of $\mathcal{C}$ (with respect to continuation). Every chain is thus bounded above by its union so by Zorn's Lemma, a maximal must exist which we shall denote as M.

We wish to show now that $M = X$, so $X \in \mathcal{W}$ and thus can be well-ordered. Suppose not, then noting that $M \subseteq X$, the set $X \setminus M$ must be non-empty, so we let $x \in X \setminus M$. Now we let $M' = M \cup \{x\}$ and well order M' the same way as M, adding on an extra condition that $y \preceq x$ for all $y \in M$. Clearly M' is now a continuation of M, which contradicts the maximality of M. □

Note that the Well-Ordering Theorem doesn't state that every set has a least element, but that every set **can be ordered to have one**. For example, the set of integers $\mathbb{Z}$ doesn't have a least element under the conventional inequality, but we can define a new partial order $\preceq$ such that

$$0 \preceq -1 \preceq 1 \preceq -2 \preceq 2 \preceq -3 \preceq 3 \preceq \cdots$$

and this is clearly a well-ordering of $\mathbb{Z}$. The rationals can also be well-ordered similiarly:

$$0 \preceq -\frac{1}{1} \preceq \frac{1}{1} \preceq -\frac{1}{2} \preceq \frac{1}{2} \preceq -\frac{1}{3} \preceq -\frac{2}{3} \preceq \frac{1}{3} \preceq \cdots .$$

If X is countably infinite, then we can simply take your favourite bijection from $\mathbb{N}$ to X, and the well-orderingness of $\mathbb{N}$ naturally induces a well-ordering of X. What if X is uncountable, like $\mathbb{R}$ for example? Sure we could let 0 be the least element, but how do we proceed from there? We would need to effectively *list* out all the real numbers somehow, which doesn't seem to be a very possible task! The Well-Ordering Theorem asserts that a well-ordering of $\mathbb{R}$ must exist, and that well-orderings for sets much larger than $\mathbb{R}$ must exist as well. The idea of this should be just as disturbing as the notion of every vector space having a basis!

3.4 Conclusion

We have now shown that AC implies Zorn's Lemma and that Zorn's Lemma implies the Well-Ordering Theorem. Does this mean that Zorn's Lemma and the Well-Ordering Theorem are weaker statements than AC? We now present the most important result of this entire section, and arguably the most important result regarding the topic of AC.

Corollary 3.4.1. *The following are equivalent:*

(a) Axiom of Choice,

(b) Zorn's Lemma,

(c) Well-Ordering Theorem. ◊

Proof. It suffices to show that $(c) \Rightarrow (a)$.

Let $\{X_i\}$ be a family of non-empty sets. If every X_i can be well-ordered, then every X_i must have a least element. We use the same argument in Example 3.1.2 and let $f(X_i) = \min(X_i)$. □

In conclusion, these 3 powerful statements cannot exist without one another. Furthermore, it was shown by Andreas Blass in 1984 [Bla84] that every vector space having a basis implies the Axiom of Choice, hence making it a 4th equivalent to AC.

We would now like to take a moment and ask the reader again the same question we asked at the beginning of Section 3. If we have a family of non-empty sets, **does a always a choice function always exist**? As AC is logically independent of ZF, we might never be able to answer this question. Although we have not found any contradictions from the use of AC, some disturbing results have emerged, such as the Well-Ordering Theorem. On the other hand, if we decide to reject AC we are forfeiting many powerful tools, like the existence of a basis. The Axiom of Choice appears in many different areas, such as *Group Theory, Topology, Functional Analysis, Measure Theory, Graph Theory, etc*, and in fact has many more equivalents in the context of those areas! AC is also really important in *Category Theory*, a topic widely regarded as the *mathematics of mathematics*. Keen readers who are interested in this may refer to Horst Herrlich's *Axiom of Choice* [Her06]. It talks about all the disasters without choice and with choice,[19] and will convince you that both AC and its negation must be true. For any readers who still feel at unease, it should be noted that ZFC is widely regarded as the standard form of Set Theory today, and that many mathematicians use The Axiom of Choice without reservation.[20] On the topic of AC, Jerry Bona once jokingly said, quoted in Steven G. Krantz' *Handbook of Logic and Proof Techniques for Computer Science* [Kra02, p.121-126]:

> "The Axiom of Choice is obviously true, the Well–Ordering Principle is obviously false; and who can tell about Zorn's Lemma?" J.L. Bona

[19] Although it should be noted that the book contains far lesser disasters with choice than without. Make of that what you will.

[20] Some might not even realise they are invoking AC when they do.

4 Addendum

4.1 Debts to pay off

Proposition 4.1.1. *We have that for all $n \in \mathbb{N}$,*

$$n \cup \{n\} = \{0, 1, 2, \ldots, n\}.$$

$\Diamond$

Proof. We first note that we are not allowed to use Lemma 4.2.7 as this would lead to circular reasoning. We proceed with induction on n. Clearly $0 \cup \{0\} = \varnothing \cup \{0\} = \{0\}$ so true for $n = 0$. If $n \cup \{n\} = \{0, 1, 2, \ldots, n\}$, then we have that

$$\begin{aligned} k \in n+1 \cup \{n+1\} &\iff k \in n+1 \text{ or } k \in \{n+1\} \\ &\iff k \in n \cup \{n\} \text{ or } k = n+1 \\ &\iff k \in \{0, 1, 2, \ldots, n\} \text{ or } k = n+1 \\ &\iff k \in \{0, 1, 2, \ldots, n+1\} \end{aligned}$$

so $n+1 \cup \{n+1\} = \{0, 1, 2, \ldots, n+1\}$ by the Axiom of Extensionality. □

Proposition 4.1.2 (Strong Induction)**.** *If $P(n)$ is a property such that*

- *$P(0)$ holds,*
- *for some fixed $n \in \mathbb{N}$, $P(k)$ holds true for all $k < n \implies P(n)$ holds,*

then $P(n)$ holds for all $n \in \mathbb{N}$. $\Diamond$

Proof. [HJ99, p.44] Let $Q(n)$ be the property "$P(k)$ holds for all $k < n$". $Q(0)$ is vacuously true. If $Q(n)$ held true, then $P(k)$ holds true for all $k < n$ and thus $P(n)$ would be true as well by the initial hypothesis. We thus see that $P(k)$ holds for all $k \leq n \Rightarrow k < n+1$ so $Q(k+1)$ holds and thus by induction, $Q(n)$ holds for all $n \in \mathbb{N}$.

For any $n \in \mathbb{N}$, $Q(n+1)$: "$P(k)$ holds for all $k < n+1$" is true so $P(n)$ holds. □

4.2 Orderings and Relations

In this section we give a brief introduction to the rigorous construction of ordered tuples. We then show how we can use this to define relations, functions, and construct the conventional inequality on $\mathbb{N}$ as an example. The Axiom of Extensionality tells us that $\{x, y\} = \{y, x\}$, so $\{x, y\}$ is not a good candidate for the ordered pair. We want to construct the ordered pair (x, y) such that $(x, y) \neq (y, x)$, and in 1921 Kuratowski proposed the following definition that is now well-accepted.

Definition 4.2.1. We define the *ordered pair* (x, y) to be

$$(x, y) = \Big\{\{x\}, \{x, y\}\Big\}.$$

$\Diamond$

We can then define the *ordered triple* $(a, b, c) = \big((a, b), c\big)$, and extend to define any finite *n-tuple*. We now prove that this definition possesses its (expected) ordered property.

Proposition 4.2.2. *$(x, y) = (x', y')$ if and only if $x = x'$ and $y = y'$.* $\Diamond$

Proof. If $x = x'$ and $y = y'$ then obviously $(x, y) = (x', y')$.

Suppose now that $\{\{x\}, \{x, y\}\} = \{\{x'\}, \{x', y'\}\}$. If $\{x\} = \{x'\}$ then we trivially have that $x = x'$ and consequently $y = y'$. If $\{x\} = \{x', y'\}$ then we must have that $x = x' = y'$ so $\{\{x'\}, \{x', y'\}\} = \{\{x\}, \{x, x\}\} = \{\{x\}\}$ and this is equal to $\{\{x\}, \{x, y\}\}$. We deduce that $\{x\} = \{x, y\}$ so $x = y$ and thus $x = x' = y = y'$. $\square$

Definition 4.2.3. Suppose that X, Y are sets. We define the *cartesian product* of X and Y to be such that

$$X \times Y = \Big\{(x, y) \in \mathcal{P}(\mathcal{P}(X \cup Y)) \;\Big|\; x \in X \wedge y \in Y\Big\}.$$

$\Diamond$

Similiar to the n-tuple, we can define the cartesian product of three sets to be $A \times B \times C = (A \times B) \times C$ and extend this to the cartesian product of any finite number of sets. The sudden appearance of $\mathcal{P}(\mathcal{P}(X \cup Y))$ might be disturbing, but a good sanity check is to note that

$$x \in X, y \in Y \implies x, y \in X \cup Y \implies \{x\}, \{x, y\} \in \mathcal{P}(X \cup Y) \implies \Big\{\{x\}, \{x, y\}\Big\} \in \mathcal{P}(\mathcal{P}(X \cup Y)).$$

Now that we have introduced order into the world of mathematics, we can define relations, an extremely fundamental and important concept that appears essentially everywhere.

Definition 4.2.4. Suppose that X, Y are sets. A set R is a *relation* from X to Y if R is a subset of $X \times Y$. We write xRy to denote $(x, y) \in R$. If $X = Y$, then we say that R is a relation on X. $\Diamond$

With this, we can now give a proper definition of what a function is.

Definition 4.2.5. Let X, Y be sets. We say that $f : X \to Y$ is a function if f is a relation from X to Y such that for every $x \in X$ there exists a unique $y \in Y$ such that xfy, and in this case we write that $y = f(x)$. $\Diamond$

A good example of a relation is the natural inequality $<$. We will prove that $<$ is a *strict order* (asymmetric and transitive) and that $\leq$ is a partial order (reflexive, antisymmetric and transitive).

Definition 4.2.6. We define the *strict inequality*, denoted $<$ to be a relation on $\mathbb{N}$ such that

$$m < n \iff m \in n.$$

We define the *non-strict inequality*, denoted $\leq$ to be a relation on $\mathbb{N}$ such that

$$m \leq n \iff m < n \text{ or } m = n.$$

The relations $>$ and $\geq$ can be defined similiarly. ◊

Lemma 4.2.7. *Suppose that $n \in \mathbb{N}$. Then,*

(a) $n \geq 0$,

(b) for any $k \in \mathbb{N}$, $n < k + 1$ if and only if $n \leq k$. ◊

Proof. (a) Let $P(n)$ be the statement "$n \geq 0$". We proceed with induction on n. $P(0)$ is obviously true since $0 = 0$. Suppose now $P(n)$: "$n \geq 0$" is true, if $n > 0$ then $0 \in n$ and thus $0 \in n \cup \{n\} = n + 1$ so $n + 1 > 0$. If $n = 0$ then $0 \in \{n\} \subseteq n \cup \{n\} = n + 1$ by the same argument, so $n + 1 > 0$ and thus $P(n + 1)$ holds in either case.

(b) For any $k \in \mathbb{N}$, we have that

$$n < k + 1 \iff n \in k + 1 = \{k\} \cup k \iff n \in \{k\} \text{ or } n \in k \iff n = k \text{ or } n < k.$$

□

Theorem 4.2.8. *$<$ is a strict order on $\mathbb{N}$ and $\leq$ is a partial order on $\mathbb{N}$.* ◊

Proof. We want to show that $<$ is a strict order, so will first prove that $<$ is transitive by induction. Let $P(n)$ be the property "for all $k, m \in \mathbb{N}$, $k < m$ and $m < n \implies k < n$". $P(0)$ is vacuously true since there is no $m \in \mathbb{N}$ such that $m < 0$. Suppose now that $P(n)$ holds and that $k < m$ and $m < n + 1$. Then, $m \leq n$ by Lemma 4.2.7 and thus by $P(n)$ we have that $(k < m) \wedge (m < n) \implies k < n$. This again implies that $k < n + 1$ by Lemma 4.2.7, so $P(n + 1)$ holds, and thus $<$ is transitive.

We will now show that $<$ is asymmetric. Suppose not, then there exists $n, k \in \mathbb{N}$ such that $n < k$ and $k < n$. By transitivity we deduce that $n < n$, but this is a contradiction to Proposition 1.0.7 as there is no n that satisfies $n \in n$, so $<$ is a strict order.

$=$ is an equivalence relation, so $\leq$ is obviously reflexive and transitive. Suppose now that $n \leq m$ and $n \geq m$. If $n \neq m$, then we have that $n < m$ and $n > m$ which is a contradiction to the asymmetry of $<$, so we must have that $n = m$ and thus $\leq$ is asymmetric and is hence a partial order.

□

Using relations, we can in fact use the equivalence relation $(a, b)R(c, d) \Leftrightarrow a + d = b + c$ on $\mathbb{N}^2$ (notational convention for $\mathbb{N} \times \mathbb{N}$) to define the set of integers $\mathbb{Z}$ (as done in *MA132 Foundations*) and further use the equivalence relation $(a, b)R(c, d) \Leftrightarrow ad = bc$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ to construct the set of rationals $\mathbb{Q}$. We can then use the *Least Upper Bound Axiom* for example to 'complete the gaps' and construct the real numbers $\mathbb{R}$. We can also define basic arithmetic (addition, multiplication,

exponentiation) as functions and prove their expected properties (e.g. associativity, commutativity, distributivity, etc). The orderings $<$ and $\leq$ extend naturally from $\mathbb{N}$ to $\mathbb{R}$ as well. We will not be going through the construction of this and instead cite Van Wyk's *'An Axiomatic Construction of the Real Number System'* [VW21] that does exactly this, and uses rational Cauchy Sequences to construct the reals.

4.3 The Axiom Schema of Replacement

There is one last axiom that we did not mention in this paper. It is the Axiom Schema of Replacement.

Axiom Schema of Replacement:

The image of any set under a definable mapping ϕ is also a set.

$$\forall A \forall w_1, \ldots, w_n \ : \ \forall x \in A \ \exists! y \ \phi(x, y, w_1, \ldots, w_n, A) \implies \exists B \Big[\forall y \in B \ \exists x \in A \ \phi(x, y, w_1, \ldots, w_n, A) \Big]$$

In simple terms, if you have a function $\phi(x)$ and some specified domain, then the image of ϕ exists. Although this is not used much most in ordinary mathematics, an important implication of this axiom is the existence of limit ordinals, denoted ω.

The natural numbers are $0, 1, 2, 3, 4, \ldots$, and every natural number (which is a set) has the property that all smaller naturals are members of it. We now try and attempt to count beyond the natural numbers to infinity. We let $\omega = \mathbb{N}$, to be an infinite number larger than any other natural number ($n < \omega$ as $n \in \omega$), and continue by defining its successor $\omega + 1 = \omega \cup \{\omega\}$ and so on. We call these limit ordinals and we generalize them alongside the natural numbers more simply as ordinal numbers. In order to justify the existence of $2\omega = \omega + \omega$, we need the Axiom Schema of Replacement to construct the set $\{\omega + n \mid n \in \mathbb{N}\}$, and proceed to take its union.

It turns out that ordinal numbers are extremely useful in many areas of set theory. Ordinal numbers, for example can be used to greatly shorten the proof of AC implies Zorn's Lemma, in which we cite Ken Brown's adaptation of it [Bro10].

References

[AM] (https://math.stackexchange.com/users/742/arturo-magidin) Arturo Magidin. How to prove from zfc that the intersection of any non-empty set exists? Mathematics Stack Exchange. https://math.stackexchange.com/q/3884139(version:2020-10-28).

[Bla84] Andreas Blass. Existence of bases implies the axiom of choice. *Contemporary mathematics*, 31:31–33, 1984. https://dept.math.lsa.umich.edu/~ablass/bases-AC.pdf.

[Bro10] Ken Brown. Mathematics 6310, zorn's lemma. 2010. https://pi.math.cornell.edu/~kbrown/6310/zorn.pdf.

[Coh64] Paul J Cohen. *The independence of the axiom of choice*. Stanford University, 1964. https://cir.nii.ac.jp/crid/1130000795022127104.

[Gol96] DC Goldrei. *Classic Set Theory: For Guided Independent Study*. Routledge, 1996.

[Hal60] Paul Richard Halmos. *Naive set theory*. van Nostrand, 1960. https://link.springer.com/book/10.1007/978-1-4757-1645-0.

[Her06] Horst Herrlich. *Axiom of choice*, volume 1876. Springer, 2006. https://link.springer.com/content/pdf/10.1007/11601562.pdf.

[HJ99] Karel Hrbacek and Thomas Jech. *Introduction to set theory, revised and expanded*. Crc Press, 3rd edition, 1999. https://doi.org/10.1201/9781315274096.

[Joh09] Johnson, C. Mathematics Prelims - Every Vector Space Has a Basis, 2009. https://mathprelims.wordpress.com/2009/06/10/every-vector-space-has-a-basis/.

[Kra02] Steven G Krantz. *Handbook of logic and proof techniques for computer science*. Springer Science & Business Media, 2002. https://link.springer.com/book/10.1007/978-1-4612-0115-1.

[MV92] Deutsche Mathematiker-Vereinigung. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, volume 1. Springer, 1892. https://books.google.com/books?hl=en&lr=&id=RC8oSE9M7iEC&oi=fnd&pg=PA1&dq=Jahresbericht+der+Deutschen+Mathematiker-Vereinigung&ots=-H6IyNT4EJ&sig=h2KBcTnvI9AzatFLmtQTQd7g2Uw.

[VW21] Leonard Van Wyk. *An Axiomatic Construction of the Real Number System*. JMU Libraries, 2021. https://commons.lib.jmu.edu/letfspubs/203/?utm_source=commons.lib.jmu.edu%2Fletfspubs%2F203&utm_medium=PDF&utm_campaign=PDFCoverPages.