Zermelo-Frankel Set Theory and Well Orderings

Menaka Lashitha Bandara

16 May 2006

Abstract

In 1883, Georg Cantor proposed that it was a valid law of thought that every set can be well ordered. This *Well Ordering Principle* remained at the heart of Cantor's cardinal numbers, which he had constructed to investigate infinite sets. However, this *Well Ordering Principle* transcended itself into the *Well Ordering Problem* when within a decade, Cantor himself was hunting for a proof. In 1908, Ernst Zermelo produced a solution. Here, Zermelo formulated and made use of a new and extremely powerful mathematical tool - *The Axiom of Choice*.

In this essay, we explore the *Zermelo-Frankel Axioms of Set Theory*, prove *Zorn's Lemma* and the *Well Ordering Theorem*, and consider some of the consequences of *The Axiom of Choice*.

Introduction

A *Well Ordered* set is a set in which every subset attains a minimum. The most common well ordered set is the set of natural numbers. Well ordering is a desirable property because such sets somewhat resemble the natural numbers [3, p66]. There is hope, then, that we can work with well ordered sets the way we work with natural numbers. There are many wonderful consequences, but here are two: extend counting beyond the natural numbers and extend the process of mathematical induction.

Well orderings are not, as Cantor discovered, a trivial "law of thought." The *Well Ordering Theorem*: *Every set can be Well Ordered* is a bold and daring claim. It is a deep result in set theory. This is the reason and the motivation for the first part of this essay - to understand set theory - in particular axiomatic set theory.

"Beyond classical analysis[1], there is an infinity of different mathematics," writes Jean Diedonné [6, p4]. Mathematics is set theory itself[2]. The first step towards understanding this statement is to consider: what do we mean by mathematics? We have an intuitive understanding of numbers, arithmetic, counting, and so on. However, the idea of set theory is to construct a theory in which these observations are a consequence, and in which all other wonderful entities can be constructed rigorously. Rephrasing this, we can say that the motivation of set theory was to create mathematics from something fundamental and unquestionable.

The first of these theories was naïve set theory, a theory in which sets were considered to be the intuitive objects. The need for the axiomatising of set theory began at the turn of the 20th Century, when philosophers and mathematicians began to find contradictions in naïve set theory. To illustrate the point, consider this problem suggested by Bertrand Russell in 1903. Russell's paradox [2, p9] runs as follows: if $\mathcal{U}$ is a set that contains all sets (the universal set), and $\mathcal{M} = \{A \in \mathcal{U} : A \notin A\}$, then is $\mathcal{M} \in \mathcal{M}$? Well, suppose that $\mathcal{M} \in \mathcal{M}$. Then we have $\mathcal{M} \notin \mathcal{M}$. If $\mathcal{M} \notin \mathcal{M}$, then $\mathcal{M} \in \mathcal{M}$. This problem highlights an inconsistency - the existence of a universal set was taken for granted in naïve set theory [3, p7]. This problem is grave: Mathematics can be only as accurate at the theory from which it is constructed. The set theory which we explore is the Zermelo-Frankel set theory - the theory on which classical analysis is based - and in fact, the accepted theory for defining mathematics.

Ideally, the axioms themselves should be "unobjectionable truths." By the very nature of axioms, however, there is always debate. There has, in fact, been one axiom of set theory which has been subjected to severe criticism

[1] Based on Zermelo-Frankel-Choice axioms.

[2] Where this is seemingly false, the theory can be easily extended to consistently deal with "classes". See [2, p9].

and debate. This has been the *Axiom of Choice*. In his day, David Hilbert writes that this axiom is the "most attacked up to the present in mathematical literature" [6, p1]. This axiom can be used to non-constructively prove the existence of some fantastic objects - some of which are seeming paradoxes. We will consider the choice axiom to some depth, and in fact, prove that it is equivalent to the *Well Ordering Theorem*.

1 The Language of Set Theory

Preceding a discussion of the non-logical Zermelo-Frankel Axioms, we consider the logic of our language.

The language of set theory consists of the following:

1. *Variables*: usually denoted by $x, X, \eta, \mathcal{M}$ and so on. In general, we shall used capitalised letters for denoting sets, and capitalised script characters for denoting sets of sets.
2. *The predicate* $\in$ *(of belonging)*: We say x belongs to X by writing $x \in X$.
3. *The predicate* $=$ *(equality)*: If x equals y, we write $x = y$.
4. *Predicate logic*: negation ($\neg$, not), conjunction ($\wedge$, and), disjunction ($\vee$, or), implies ($\implies$, implies), equivalence ($\iff$, iff).
5. *Logical quantifiers*: Universal quantifier ($\forall$, for all) and Existential quantifier ($\exists$, there exists).
6. *Scope symbols*: $()$, $[]$, where the scope is limited to the brackets.

A *sentence* for our purposes is characterised by the three following rules:

1. Let x, y be variables, then $x \in y$ is a sentence and we write $S(x, y) =$ "$x \in y$".
2. If S, T are sentences, then $\neg S$, $S \vee T$, $S \wedge T$, $S \implies Y$, $S \iff Y$, $S = Y$ are also sentences.
3. Let $S(a)$ be a sentence in a. Then $\forall a[S(a)]$ and $\exists a[S(a)]$ are also sentences.

It is worthwhile noting that we have not been rigorous and formal in constructing this underlying logic. For instance, we have been naïve in using the relation $=$, in defining our sentence $S(x, y)$. It is possible to be axiomatic and formal, yet this is extremely tedious and almost inappropriate for our purposes. Furthermore, we shall refrain from the practice of checking that our expressions are in fact sentences, since in most cases, our sentences are simple.

Sometimes, we will also be slack in our use of the universal quantifier. Rather than writing $\forall x \in A$, we shall simply write $x \in A$. To avoid ambiguity, we shall always be explicit with the existential quantifier.

We shall call an arbitrary "collection" of "things" by the term *collection*. For those collections which are sets, we shall explicitly call them *sets*. In general, a collection is not a set, but a set is a collection. However, to avoid any confusion, we shall never refer to a set as a collection.

A formal, axiomatic and a more complete treatment can be found in [2, §2,3,4].

2 Zermelo-Frankel Axioms and Mathematics

Prior to the Zermelo-Frankel axioms, Russell and Whitehead attempted to axiomatise set theory and resolve paradoxes by introducing a "doctrine" of types [9, p523]. It seems, however, that this theory obscured the simplicity one seeks from axioms. The Zermelo-Frankel system delivered a more obvious and tangible set of axioms.

In this section, we introduce the Zermelo-Frankel axioms of set theory. Simultaneously, we attempt to construct some of the mathematical machinery necessary to discuss the axioms in a more mathematical, rather than a purely logical setting. This also demonstrates how some mathematics can indeed be built from set theory. We achieve this by phrasing and discussing the axioms using the machinery that we develop, taking care to avoid any logical circularity.

2.1 Axiom of Extension

We begin our discussion of set theory by asking a rather philosophical question: what does it mean for two sets to be equal? Now, this may seem a little trivial, and we have an intuitive answer. However, intuition is insufficient in when we seek rigour, and so we make explicit what we feel.

Axiom 1 (The Axiom of Extension) *Let A, B be sets. Then $A = B$ if and only if they contain the same elements.*

We highlight an immediate consequence of this definition.

Theorem 2.1

1. $\{a, a\} = \{a\}$
2. $\{a, b\} = \{b, a\}$

Proof For the first result, note that $x \in \{a, a\} \iff x \in \{a\}$. The second result follows by the same argument. □

This short but insightful result highlights that repeated elements are ignored in sets, and that sets are unordered. These are crucial properties of sets, and they are worthy of attention. Also, Extension (Axiom 1) is used to guarantee the uniqueness of sets. Since we have mentioned this fact here, we shall not be explicit in all but the proofs where it is considered necessary.

We now define the notion of a subset:

Definition 2.2 (Subset, Proper Subset) *Let A and B be sets. If $x \in A \implies x \in B$, then we say that A is a subset of B and write $A \subseteq B$ or $B \supseteq A$. If $A \subseteq B$ but $A \neq B$, then we write $A \subsetneq B$ or $B \supsetneq A$, and we say that A is a proper subset of B.*

With the use of this notation, we can give another formulation of Extension (Axiom 1). We can say that sets $A = B$ iff $A \subseteq B$ and $B \subseteq A$.

We also note the following useful result about the subset relation.

Theorem 2.3 (Transitivity of Subsets) *If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.*

Proof Since $A \subseteq B$, we know that $x \in A \implies x \in B$. And since $B \subseteq C$, $x \in B \implies x \in C$. But then trivially, $x \in A \implies x \in C$. □

2.2 Axiom of Replacement

We now consider the schema for replacement. The power and importance of this axiom will be highlighted as we begin to delve deeper into set theory, and more specifically extending counting to an infinite setting. The motivation behind this axiom is to characterise when we can create a new set by substituting values of the old set.

Axiom 2 (The Axiom of Replacement) *Let $S(a,b)$ be a sentence, and A a set. If for all $a \in A$, there exists a unique b such that $S(a,b)$ is true, then there exists a set B such that $b \in B \iff$ for some $a \in A$, $S(a,b)$ is true.*

This axiom might seem somewhat cryptic. But it tells us this: If we have some sentence $S(a,b)$, and for each a we there is exactly one b for $S(a,b)$ true, then the collection $X = \{b : S(a,b)\}$ is a set.

We now state a theorem which will be of frequent use:

Theorem 2.4 (Theorem of Specification) *Let $S(a)$ be a sentence on a and let A be a set. Then $B = \{a \in A : S(a)\}$ is a set.*

Proof We consider the sentence $S'(a,b) =$ "$a \in A \land S(a) \land a = b$".

Now, by Replacement (Axiom 2), we are guaranteed a set C such that $b \in B \iff S(a,b)$. But we have $a = b$, which implies that $a \in B \iff a \in A \land S(a)$. By Extension (Axiom 1), this is exactly the set B. □

Initially, this theorem was introduced as an axiom - the Axiom of Specification (also called Separation). When Replacement became necessary to obtain deeper results, Specification was made redundant.

We illustrate how this theorem can be used to solve the problem which served as a primary motivation for the axiomatising of set theory: Russell's Paradox.

Theorem 2.5 (Nonexistence of the Universal Set) *There exists no Universal set.*

Proof Suppose that there exists a universal set. Let $\mathcal{U}$ be that set. We note that the sentence $S(a) =$ "$a \notin a$" is a valid sentence. Now by Theorem 2.4, $A = \{a \in \mathcal{U} : a \notin a\}$, is a set. Now note that $x \in A \iff x \in \mathcal{U} \land x \notin x$.

Now if $A \in A$, then $A \in \mathcal{U}$ and $A \notin A$. If $A \notin A$, then $A \in \mathcal{U}$ (by hypothesis) and $A \in A$ by construction of A. Either way, if we assume $A \in \mathcal{U}$ we get a contradiction. The only way out of the contradiction is to assume that $A \notin \mathcal{U}$. But then $\mathcal{U}$ does not contain all sets. □

The proof shows that $\mathcal{U}$ is not a set, and consequently A is not guaranteed to be a set [2, p10]. It may be called a class, but discussion of classes are beyond the scope of this essay.

While this may seem a contradiction to the logic of our proof, it is not. Here is the reason: we've assumed that the collection which we call the universal set is in fact a set, and from this it follows that we can legitimately apply Specification (Theorem 2.4) to obtain the object in question as a set. The fact that it is not a set only gives the proof greater validity - if we investigated the theory of classes, we could use this fact as the contradiction in our proof.

Another point worth noting is that generally, when we cannot construct sets just by stating a property. For instance $\{x : x$ is an elephant$\}$ is not a set.

In general, when we use Specification, given a sentence $P(x)$, we will either write $\{x \in A : P(x)\}$ or if we write $\{x : P(x)\}$, then we assume that $P(x)$ specifies that x belongs to some set. Note, this is not a contradiction to our formulation of Replacement (Axiom 2). In this axiom, we still need an initial set to "replace."

2.3 Axiom of Empty Set

So far, we have discussed what we mean by set equality, and to some extent, how to create a new set out of a pre-existing one. But for all we know, we may have been working in a vacuum. This next axiom guarantees that there is at least one set. Later, we will see that this axiom (along with the other axioms of set theory) are in fact sufficient to create mathematics.

Axiom 3 (The Empty Axiom) *There exists a set, denoted $\varnothing = \{\}$, that contains no elements.*

As always, the uniqueness of this set is guaranteed by Extension (Axiom 1).

In order to construct the empty set, we could have equally well have assumed the existence of some set[3, p8]. Then by the use of Specification (Theorem 2.4), we could use the sentence: $S(a) =$ "$a \in S \wedge a \neq a$", which would result in a set containing no elements - or $\varnothing$. The preference of approach is a question of aesthetics. The approach taken feels more minimal.

There is one more point to emphasise. The empty set ties in beautifully with logic. This is the notion of vacuous truth. Informally, it means that any universally quantified implication about the empty set is always true. Without proof, we use the following identity from logic: $(p \implies q) \equiv \neg p \vee q$.

Theorem 2.6 (The principle of vacuous truth) *Let $P(x)$ be some sentence. Then $\forall x \, [x \in \varnothing \implies P(x)]$ is always true.*

Proof Using the identity which we discussed in the previous paragraph, note the following:

$$\forall x \, [x \in \varnothing \implies P(x)] \equiv \forall x \, [\neg \, (x \in \varnothing) \vee P(x)] \equiv \forall x \, [x \notin \varnothing \vee P(x)]$$

But it is always true that $x \notin \varnothing$ by Empty Axiom (Axiom 3). □

What follows from this principle the following fact:

Theorem 2.7 *Let A be a set. Then $\varnothing \subseteq A$.*

Proof Given a set A, consider sentence $S(a) =$ "$a \in A$". Then, we consider $\forall x \in \varnothing \implies S(x)$. This is vacuously true (Theorem 2.6). The result follows by Definition 2.2. □

2.4 Axiom of Pairs

Suppose we have two sets A, and B. The question remains: are these sets elements of some set? The theory we have built is to weak to resolve this question. This motivates the following axiom.

Axiom 4 (The Axiom of Pairing) *For any two sets A, and B, there exists a set X that contains A, and B.*

By the use of Specification (Theorem 2.4), we can find a set $Y \subseteq X$ such that $Y = \{A, B\}$. Consider setting $S(a) =$ "$a = A \vee a = B$".

We give names to two types of sets that are special:

Definition 2.8 (Unordered Pair) *Let $X = \{a, b\}$. Then X is an unordered pair.*

Definition 2.9 (Singleton) *The set $A = \{a\}$ is called a singleton.*

2.5 Axiom of Unions

It follows naturally that given an arbitrary set of sets $\mathcal{C}$, we want a superset in which every set in $\mathcal{C}$ is a subset.

Axiom 5 (The Axiom of Unions) *Let $\mathcal{C}$ be a set of sets. The there exists a set C such that for all $A \in \mathcal{C} \iff A \subseteq C$.*

Alternatively, we could relax our "if and only if" condition, and use Specification (Theorem 2.4) to create the set guaranteed by Unions (Axiom 5).

Now we define the notation for unions.

Definition 2.10 (Union) *Let $\mathcal{C}$ be a set of sets. Let C be the set guaranteed by Unions (Axiom 5). Then we say that C is the union of the sets of $\mathcal{C}$ and write:*

$$C = \bigcup \mathcal{C} = \bigcup_{A \in \mathcal{C}} A$$

We state two trivial facts about unions:

Theorem 2.11

1. $\bigcup \varnothing = \bigcup_{A \in \varnothing} A = \varnothing$
2. $\bigcup \{A\} = \bigcup_{X \in \{A\}} X = A$

Proof Trivial. The first result is a vacuous argument, and the second by a simple application of Extension (Axiom 1). □

Suppose we have a pair of sets. Then the union of these sets can be written in a particular way. Note that the definition does not highlight an exception, but rather, it emphasises our definition in this special case.

Definition 2.12 (Union of a Pair) *Let $\mathcal{C} = \{A, B\}$. Then we write:*

$$\bigcup \mathcal{C} = A \cup B = \{x : x \in A \vee x \in B\}$$

We present some basic and trivial facts about unions of pairs. We omit the proofs since they are straightforward.

Theorem 2.13 (Properties of Union of Pairs) *Let A, B, C be sets. Then,*

1. $A \cup \varnothing = A$
2. $A \cup B = B \cup A$
3. $A \cup (B \cup C) = (A \cup B) \cup C$
4. $A \cup A = A$
5. $A \subseteq B \iff A \cup B = B$

2.6 Intersections

Given a set of sets, we want to create a collection of elements of the sets such that the collection is a set. This is the motivation for creating the intersections of sets. The results obtained so far are in fact strong enough to assert the collection in the following definition is a set.

Definition 2.14 (Intersection) *Let $\varnothing \neq \mathcal{C}$ be a set of sets. The we define the intersection of the sets in $\mathcal{C}$ by:*

$$\bigcap \mathcal{C} = \bigcap_{A \in \mathcal{C}} A = \{x \in X : \forall X \in \mathcal{C}\}$$

The sharp mathematician would immediately ask: why do we seek $\mathcal{C} \neq \varnothing$? Well, suppose that $\mathcal{C} = \varnothing$. Then, consider the intersection:

$$\begin{aligned} C &= \bigcap \mathcal{C} \\ &= \bigcap \varnothing \\ &= \{x \in X : \forall X \in \varnothing\} \end{aligned}$$

Now consider the statement $a \notin C$, which implies that $x \notin X$ and $X \in \varnothing$. Vacuously (Theorem 2.6) it is true that that every $x \in C$. But that would mean that C is the universal collection that is a set, and it would result in a contradiction in our theory.

This does not cause a huge problem. We just need to be careful that the set of sets we have is not empty when we take intersections. Where we do not make this explicit, it is always assumed that such a set is nonempty.

As with intersections we note that pairs of sets give rise to the following definition.

Definition 2.15 (Intersection of a Pair) *Let $\mathcal{C} = \{A, B\}$. Then we define the intersection by:*

$$\bigcap \mathcal{C} = A \cap B = \{x : x \in A \land x \in B\}$$

We state some rudimentary properties of intersections. The proofs of the following statements are a trivial application of Extension (Axiom 1).

Theorem 2.16 (Properties of the Intersection of a Pair)

1. $A \cap \varnothing = \varnothing$
2. $A \cap B = B \cap A$
3. $A \cap (B \cap C) = (A \cap B) \cap C$
4. $A \subseteq B \iff A \cap B = A$

And further, we note that intersections and unions are distributive.

Theorem 2.17 (Distributivity of Intersections over Unions)

1. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
2. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

2.7 Complementation

Here, we define the notion of what it means to "subtract" a set of things from a set. So, when we speak of complement, we are always talking about relative complement.

Definition 2.18 (Complement) *Let A and B be sets. Then we define the complement of B by A:*

$$A \setminus B = \{x \in A : x \notin B\}$$

Sometimes, this is also referred to as set difference. Again, we present some important properties of complement. These are straightforward results and thus the proofs are omitted.

Theorem 2.19 (Properties of Complement)

1. $x \notin B \iff x \notin A \setminus B$ *whenever* $A \subseteq B$
2. $A \setminus (A \setminus B) = A \cap B$
3. $A \setminus \varnothing = A$
4. $A \setminus A = \varnothing$
5. $A \cap (B \setminus A) = \varnothing$
6. $A \cup (B \setminus A) = A \cup B$
7. $A \subseteq B \subsetneq E \iff E \setminus B \subseteq E \setminus A$ *whenever* $A, B \subsetneq E$

The following theorem is also straightforward. However, the first part of it is used frequently, so we present a proof. The second part is a tedious, but trivial application of Extension (Axiom 1).

Theorem 2.20

1. $A \subseteq B \iff A \setminus B = \varnothing$
2. $A \cap (B \setminus C) = (A \cap B) \setminus (A \cap C)$

Proof

1. Let $A \subseteq B$. Then $A \setminus B = \{x \in A : x \notin B\}$. But $A \subseteq B$ and from the contrapositive of the definition, $x \notin B \implies x \notin A$. It follows that $A \setminus B = \varnothing$.

To prove the converse, suppose that $A \setminus B = \varnothing$. Trivially, whenever $x \notin B \implies x \notin A$, and the contrapositive yields $x \in A \implies x \in B$. So, $A \subseteq B$.

□

2.8 Axiom of Powers

It is a consequence of Specification (Theorem 2.4) that a subset of a set is itself a set. However, is the collection of all subsets of a set itself a set? We would certainly hope so - but the axioms we have discussed so far are not powerful enough provide an answer.

Axiom 6 (The Power Axiom) *For each set X the collection that contains all and only the subsets of X denoted $\wp(X)$ is a set.*

Alternatively, we could have claimed the existence of some set that contains all subsets of the set, but this definition is minimal and the uniqueness of $\wp(X)$ is guaranteed by Extension (Axiom 1). We call $\wp(X)$ the power set of X.

We present some fundamental results about the power set.

Theorem 2.21 (Properties of the Power Set) *Let A, B be sets. Then:*

1. $\varnothing \in \wp(A)$
2. $\bigcap \wp(A) = \varnothing$
3. $A \subseteq B \implies \wp(A) \subseteq \wp(B)$
4. $\wp(A) \cap \wp(A) = \wp(A \cap B)$
5. $\wp(A) \cup \wp(B) \subseteq \wp(A \cup B)$

Proof

1. By Theorem 2.7, $\varnothing \subseteq A$. It follows then that $\varnothing \in \wp(A)$.

2. Since $\varnothing \in \wp(A)$, by Theorem 2.16 the result follows.

3. Let $X \in \wp(A)$. Then $X \subseteq A \implies X \subseteq B$. Trivially, it follows that $X \in \wp(B)$.

4. Note: $X \in \wp(A) \cap \wp(B) \iff X \subseteq A \wedge X \subseteq B \iff X \in \wp(A \cap B)$.

5. Let $X \in \wp(A) \cup \wp(B) \implies X \in \wp(A) \vee X \in \wp(B) \implies X \subseteq A \cup B \implies X \in \wp(A \cup B)$.

□

The following result relates complements to unions and intersections.

Theorem 2.22 (Generalised De Morgan's Laws) *Let X be a set and let $\varnothing \neq \mathcal{C} \subseteq \wp(X)$. Then:*

1. $X \setminus \left(\bigcup_{A\in\mathcal{C}} A\right) = \bigcap_{A\in\mathcal{C}} (X \setminus A)$
2. $X \setminus \left(\bigcap_{A\in\mathcal{C}} A\right) = \bigcup_{A\in\mathcal{C}} (X \setminus A)$

Proof

1. Let $x \in X \setminus \left(\bigcup_{A\in\mathcal{C}} A\right)$. Then $x \in X$ and $x \notin \bigcup_{A\in\mathcal{C}A}$. So, for all $A \in \mathcal{C}$, $x \notin A$. By Theorem 2.19 (1), we have $x \in X \setminus A$. It follows then that $x \in \bigcap_{A\in\mathcal{C}} (X \setminus A)$.

Conversely, let $x \in \bigcap_{A\in\mathcal{C}} (X \setminus A)$. Then for all $A \in \mathcal{C}$, $x \in X \setminus A$. Again, we invoke Theorem 2.19 (1), and it follows that $x \notin \bigcup_{A\in\mathcal{C}} A$ which implies $x \in X \setminus \bigcup_{A\in\mathcal{C}} A$.

2. Similarly, let $x \in X \setminus \left(\bigcap_{A \in \mathcal{C}} A\right)$. So, $x \in X$ and $x \notin \bigcap_{A \in \mathcal{C}} A$, for all $A \in \mathcal{C}$. It follows then $x \in X \setminus A$ for all $A \in \mathcal{C}$ which implies $x \in \bigcap_{A \in \mathcal{C}} X \setminus A$.

Conversely, let $x \in \bigcap_{A \in \mathcal{C}} X \setminus A$. Then, $x \in X \setminus A$ for all $A \in \mathcal{C}$, which implies that $x \notin A$ for all $A \in \mathcal{C}$. It follows then that $x \notin \bigcup_{A \in \mathcal{C}} A$ which implies $x \in X \setminus \left(\bigcup_{A \in \mathcal{C}} A\right)$.

□

A natural question to ask is why this proof is presented here rather than in the previous section. This question highlights an important point. In order for De-Morgan's Laws to work, we need at least a "virtual" universe. The complementation here is in some sense "absolute," since all operations are restricted to some larger set. It is customary in mathematics to take a set "large enough" in areas where a universal set may seem necessary.

Another point: we begin with a set X, and we consider collections of subsets of X. However, without the Power Axiom (Axiom 6), we cannot deduce that such collections are indeed sets. Indeed, if they are not sets, then our definitions of union, intersection, and compliment only hold in the vacuum. Axiom 6 has been aptly named - it is indeed powerful.

3 Ordered Pairs

We have previously discussed that sets are unordered. Theorem 2.1 is a proof of this fact. However, in order to construct relations, functions, and all the rest, we need a notion of order.

The construction here may seem somewhat artificial. However, it's not a high price to pay when considering the only other alternative - creating a redundant axiom.

With the aid of the Power Axiom (Axiom 6), we know that given a pair $\{a, b\}$, the collection $\{\{a\}, \{a, b\}\}$ is indeed a set. This motivates the following definition.

Definition 3.1 (Ordered Pair) *Let $\{a, b\}$ be a set. Then we define the ordered pair:*

$$(a, b) = \{\{a\}, \{a, b\}\}$$

The following result tells us that such orderings are unique. A proof can be found in [3, p24].

Theorem 3.2 (Uniqueness of Ordered Pairs) *If $(a, b) = (x, y)$, then $a = x$ and $b = y$.*

Now we assert the existence of a set that contains exactly all of the ordered pairs (a, b) of sets $a \in A$ and $b \in B$.

Theorem 3.3 (Cartesian Product Theorem) *Let A, B be sets. Then there exists a set denoted $A \times B$ such that $(a, b) \in A \times B \iff a \in A \wedge b \in B$.*

Proof Fix $a \in A$ and $b \in B$. The, note that:

$$\begin{aligned}
\{a\} \subseteq A \wedge \{b\} \subseteq B &\implies \{a\}, \{a, b\} \subseteq A \cup B \\
&\iff \{a\}, \{a, b\} \in \wp(A \cup B) \\
&\implies \{\{a\}, \{a, b\}\} \subseteq \wp(A \cup B) \\
&\iff \{\{a\}, \{a, b\}\} \in \wp(\wp(A \cup B))
\end{aligned}$$

We define $A \times B \subseteq \wp(\wp(A \cup B))$ by using Specification (Theorem 2.4):

$$A \times B = \{(a, b) \in \wp(\wp(A \cup B)) : a \in A \wedge b \in B\}$$

With uniqueness of this set is guaranteed by Extension (Axiom 1). □

The next result tells us that for any set of ordered pairs, we can always find a Cartesian product that contains them. A proof sketch can be found in [3, p24].

Theorem 3.4 *Let R be a set of ordered pairs. Then there exists an $A \times B$ such that $R \subseteq A \times B$.*

We now list some properties of ordered pairs. Again, the proofs are straightforward, but tedious, and hence they are omitted.

Theorem 3.5 *Let A, B, X, Y be sets. Then:*

1. $(A \cup B) \times X = (A \times X) \cup (B \times X)$
2. $(A \cap B) \times (X \cap Y) = (A \times X) \cap (B \times Y)$
3. $(A \setminus B) \times X = (A \times X) \setminus (B \times X)$
4. $A \times B = \varnothing \iff A = \varnothing \vee B = \varnothing$
5. $A \times B \subseteq X \times Y \iff A \subseteq X \wedge B \subseteq Y$ *provided* $A \times B \neq \varnothing$

It is rather important to emphasise that the preceding theorem does not imply that $(A \cup B) \times (X \cup Y) = (A \times X) \cup (B \times Y)$. This statement is in fact false.

3.1 Relations and Functions

Having a concept of an ordered pair, we can now move on to describe relations.

Definition 3.6 (Relation) *Let X be a set, and let $R \subseteq \wp(\wp(X))$. Then R is a relation on X and we write xRy (x stands in relation to y in R) if $(x, y) \in R$. We define the domain and range of the relation R by:*

$$\operatorname{dom}(R) = \{x \in X : \exists y \in X, (x, y) \in R\}$$
$$\operatorname{ran}(R) = \{y \in X : \exists x \in X, (x, y) \in R\}$$

There is a special type of relation which will be of importance to us at a later stage. This is the the notion of an equivalence relation.

Definition 3.7 (Equivalence Relation) *Let R be a relation on X. Then if:*

1. $\forall x \in X, (x, x) \in R$ *(Reflexive)*
2. $\forall x, y \in X, (x, y) \in R \implies (y, x) \in R$ *(Symmetric)*
3. $\forall x, y, z \in X, (x, y), (y, z) \in R \implies (x, z) \in R$ *(Transitive)*

Then R is an equivalence relation. For an equivalence relation, we write $x \pmod R = [x] = \{y : xRy\}$, the equivalence class of x.

Definition 3.8 (Partition of a Set) *Let X be a set, and let $\mathcal{P} \subseteq \wp(X)$. Then if:*

1. $\bigcup \mathcal{P} = X$

2. $\forall X, Y \in \mathcal{P} \implies X \cap Y = \varnothing$

Then $\mathcal{P}$ is partition of of X.

This following theorem draws the connection between partitions and equivalence relations.

Theorem 3.9 *Let X be a set. Then:*

1. *If R is an equivalence relation on X, then $S \pmod{R}$ is a partition of X.*
2. *If $\mathcal{P}$ is a partition of X, then there exists a relation R which induces $\mathcal{P}$.*

Proof

1. Since R is an equivalence relation, $(x, x) \in R$, for all $x \in X$. Then, it follows that $\bigcup_{x \in X} x = X$.

Now take $[x], [y] \in X \pmod{R}$. Suppose $z \in [x]$ and $z \in [y]$. Then $(x, z) \in R$ and $(z, y) \in R$ which implies $(x, y) \in R$. It follows that $[x] = [y]$.

2. Let $\mathcal{P}$ be a partition of X. We define a relation R:

$$R = \{(x, y), (y, x) \in X \times X : x, y \in P_\alpha \in \mathcal{P}\}$$

Trivially, R is symmetric and reflexive. We show that transitivity holds. Suppose $(x, y), (y, z) \in R$. This holds if and only if $x, y \in P_\alpha$ and $y, z \in P_\beta$, for some $P_\alpha, P_\beta \in \mathcal{P}$. Since $\mathcal{P}$ is a partition, we have that $\alpha = \beta$, and it follows that $x, z \in P_\alpha \iff (x, z) \in R$.

□

Now we define a function in the language of set theory. Intuitively, we would like a function to "map" one value to another. We do not, however, want a function to map one value to possibly two values. The following definition makes this informal discussion rigorous.

Definition 3.10 (Function) *Let X, Y be sets, and let f be a relation in $X \times Y$. Then if $\operatorname{dom}(f) = X$ and if $(x, y) \wedge (x, z) \in f \implies y = z$, then we say that f is a function $f : X \to Y$ and write $f(x) = y$ when $(x, y) \in f$.*

Firstly, we mention how to create a new function from another.

Definition 3.11 (Restriction, Extension) *Let $f : X \to Y$. Then given $A \subseteq X$, $f|_A : A \to Y$ is called the restriction of f to A, and f is called an extension of $f|_A$.*

Now we answer a fundamental question about collections of functions. Given sets X and Y, is collection of all functions from X to Y itself a set?

Theorem 3.12 (Set of functions) *Let X, Y be sets. Then the collection of all functions from X to Y denoted X^Y is a set.*

Proof We note that given any function f, $f \subseteq X \times Y$, since every function is a relation. Thus, $f \in \wp(X \times Y)$. Then $X^Y \subseteq \wp(X \times Y)$, and can be created explicitly by using Specification (Theorem 2.4). □

With functions, sometimes we call the range of the function its image, and write $\operatorname{ran}(f) = \operatorname{im}(f)$. Furthermore, we will be sloppy with notation and denote $\operatorname{im}(f) = f(X)$, where $X = \operatorname{dom}(f)$.

Definition 3.13 (Inverse Image) *Let $f : X \to Y$. Then we define the inverse image of f at y by:*

$$f^{-1}(y) = \{x \in X : f(x) = y\}$$

It turns out that there are two types of functions that are especially important.

Definition 3.14 (Injective) *Let $f : X \to Y$. If $f(x) = f(y) \implies x = y$, then we say that f is an injection or injective.*

Definition 3.15 (Surjective) *Let $f : X \to Y$. If for all $y \in Y$, there exists an $x \in X$ such that $f(x) = y$, then we say that f is a surjection or surjective.*

These definitions motivates the following important theorem about functions.

Theorem 3.16 (Existence of Unique Inverse) *Suppose f is injective and surjective. Then, there exists a unique f^{-1} inverse such that $f^{-1}(f(x)) = x$, for all $x \in X$.*

Definition 3.17 (Bijection) *If $f : X \to Y$ is both injective and surjective, then it is called a bijection or bijective.*

The proof of this theorem is elementary and readily accessible. The fact that we have a conflict with our notation for the unique inverse and the inverse image may seem somewhat sloppy. However, this is not quite an accident. Note that when f is bijective, each inverse image $f^{-1}(y) = \{x : f(x) = y\} = \{x\}$, and justifies our use of this notation.

There is one more important concept about functions. This is the notion of composing two functions together:

Definition 3.18 (Composition) *Let $f : X \to Y$ and $g : Y \to Z$ be functions. Then consider the function $h : X \to Z$ defined by:*

$$h(x) = (g \circ f)(x) = g(f(x))$$

Then h is the composition of g and f.

We state two useful results. The proofs are trivial and are omitted.

Theorem 3.19 *Let $f : X \to Y$, $g : Y \to Z$, $h : Z \to W$. Then:*

1. $(h \circ g) \circ f = h \circ (g \circ f)$
2. $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$

Theorem 3.20 *If $f : X \to Y$ and $g : Y \to Z$ then:*

1. *If f, g injective then $g \circ f$ injective*
2. *If f, g surjective then $g \circ f$ surjective*

3.2 Families of Sets

Functions are important fundamental objects in all areas of mathematics. The discussion to follow will highlight their importance.

Definition 3.21 (Family) *Let Φ, X be sets. Suppose $x : \Phi \to X$. Then we write $x(\alpha) = x_\alpha$, and we call x_α a family in X, and Φ the index set.*

Naturally, we can replace X with a power set, and we can begin to consider families of subsets. In fact, any set of sets itself is a family - the index set is itself. For this reason, we shall often talk about families rather than sets of sets.

In general, we will be loose with the notation, and omit supplying an index set. However, it is important to emphasise that there is always an implied index set.

Definition 3.22 (Intersection, Union of Families) *Let $\mathcal{C} = \{X_\alpha\}$, $\alpha \in \Phi$, a family of sets. Then,*

1. *$\bigcup \mathcal{C} = \bigcup_{\alpha\in\Phi} X_\alpha$ is the Union of the family.*
2. *$\bigcap \mathcal{C} = \bigcap_{\alpha\in\Phi} X_\alpha$ is the Intersection of the family ($\mathcal{C} \neq \varnothing$).*

We have seen in Theorem 3.3 that we can make sense of products when we have pairs. It turns out that families is the exact way in which arbitrary products can be characterised. This definition will gain more merit when we consider the Choice Axiom.

Definition 3.23 (Product of Family) *Let $\mathcal{C} = \{A_\alpha\}$ be a family of sets, with index Φ. Then define product:*

$$\prod \mathcal{C} = \prod_{\alpha\in\Phi} A_\alpha = \left\{ \{x_\alpha\} \in \Phi^{\bigcup \mathcal{C}} : x_\alpha \in A_\alpha \right\}$$

Now, is $\prod \mathcal{C}$ indeed a set? We reason as follows: By Theorem 3.12, we know that the collection of all functions from Φ to $\bigcup \mathcal{C}$ is indeed a set. Then, by invoking Specification (Theorem 2.4), we can assert that $\prod \mathcal{C}$ is indeed a set.

3.3 Axiom of Infinity

Let us now turn to the problem of constructing the natural numbers. We only really know of the existence of one set which the Empty Axiom (Axiom 3) guarantees. Somehow, we should construct numbers from it. The following definition may shed some light on a possible approach.

Definition 3.24 (Successor) *Let A be a set. The we define the successor of A by:*

$$A^+ = A \cup \{A\}$$

Note that, such a construction is a valid set, since $\{A\} \in \wp(A)$.

Definition 3.25 (Natural Numbers)

$$\begin{aligned}
&0 = \varnothing \\
&1 = 0^+ = \{0\} = \{\varnothing\} \\
&2 = 1^+ = \{0, 1\} = \{\varnothing, \{\varnothing\}\} \\
&3 = 2^+ = \{0, 1, 2\} = \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\} \\
&\vdots
\end{aligned}$$

It may seem awkward that every number contains all the numbers before it. But bear with us for a moment. This very fact shall become of great importance when we consider Ordinal numbers.

A more profound question is this: is the collection of all natural numbers a set? Our theory is too weak to provide an answer. So, we axiomatise.

Axiom 7 (Axiom of Infinity) *There exists a set that contains* 0 *and all its successors.*

What we really want, however, is the existence of a set that contains only the natural numbers. The following result tells us that such a set does indeed exist.

Theorem 3.26 (Existence of minimal successor set) *There exists a set* ω *that contains only* 0 *and all its successors.*

Proof Let A_∞ be the set guaranteed by Infinity (Axiom 7). Then, let:

$$\mathcal{C} = \{A \in \wp(A_\infty) : A \text{ contains 0 and all its successors}\}$$

We have $\mathcal{C} \neq \varnothing$, since $A_\infty \in \mathcal{C}$. Then, we let $\omega = \bigcap \mathcal{C}$.

Note, then that $0 \in \omega$, since $0 \in A, \forall A \in \mathcal{C}$. Also, let $n \in \omega$. Then $n \in A, \forall A \in \mathcal{C}$ which implies $n^+ \in A, \forall A \in \mathcal{C}$. But then $n^+ \in \omega$. The minimality of ω is a simple proof by contradiction. □

From this point onwards, we shall denote this minimal set ω.

Theorem 3.27 (The Principle of Mathematical Induction) *If* $S \subseteq \omega$ *and* $0 \in S$ *and* $n^+ \in S$ *whenever* $n \in S$*, then* $S = \omega$*.*

Proof Suppose $S \neq \omega$. But then, $S \subsetneq \omega$, and this contradicts the minimality of ω guaranteed by Theorem 3.26. □

The result above gives more merit to axiomatic set theory. The Principle of Mathematical Induction is indeed a very powerful technique of proof. Historically, it has been just what it is named - a principle. However, it is reassuring to know that in the light of set theory, it in fact becomes a consequence.

4 Axiom of Regularity

We present the following axiom for completeness of our discussion of the Zermelo-Frankel system. This axiom will not be quoted or used elsewhere in the essay. It is, however, required for deeper results in the theory. We state it and examine two critical consequences.

Axiom 8 (Axiom of Regularity) *For any set* $X \neq \varnothing$*, there exists a* $Y \in X$ *such that* $X \cap Y = \varnothing$*.*

This axiom is also sometimes called the Foundation Axiom. We illustrate two profound consequences.

Theorem 4.1 *Let* $X \neq \varnothing$ *be a set. Then* $X \notin X$*.*

Proof Suppose that X is a set and $X \in X$. Consider the set $\{X\}$. This set contradicts Regularity, since there exist no $Y \in \{X\}$ such that $X \cap Y = \varnothing$. □

Theorem 4.2 *For any family of sets $\{X_n : n \in \omega\}$, it is never true that $X_{n^+} \in X_n$ for all $n \in \omega$.*

Proof Suppose there does exist such a family $\mathfrak{F}$. Trivially, there exist no $X \in \mathfrak{F}$ such that $X \cap \mathfrak{F} = \varnothing$, which again contradicts Regularity. □

These results resolve two important questions: can a set be a member of itself, and can the $\in$ predicate give rise to an infinite chain. Although results are a consequence of this axiom, they were in fact the very motivation for Regularity. A more complete discussion can be found in [8, p56], [2, p19], and [6, p153].

5 The Peano Postulates and Arithmetic

We consider the Peano theory of arithmetic defined by the postulates which Peano presented in 1889 [10, p83]. Initially, they were considered to be axioms of their own right. However, in the presence of set theory, they are consequences! This is most desirable and enlightening. Our model is strong enough to imply the statements we consider the "very fountainhead of all mathematical knowledge" [3, p47].

5.1 The Peano Postulates

Before we state and prove the Peano axioms, we shall introduce two auxiliary results necessary to prove the postulates.

Lemma 5.1 *No natural number is a subset of any of its elements.*

Proof Let $S = \{n \in \omega : \forall z \in n, n \not\subseteq z\}$.

Then $0 \in S$ by Theorem 2.6. By construction, we have $n \subseteq n \implies n \notin n$. This implies that $n^+ \notin n$. Let $n^+ \subseteq x \implies n \subseteq x$. Then we have $x \notin n$. Also, $n^+ \not\subseteq n$ and $n^+ \not\subseteq z$, for all $z \in n$ (since $z \in n$). But this implies $n^+ \not\subseteq z$ for all $z \in n$ and $n^+ \not\subseteq n$. It follows that $n^+ \in S$.

By Induction (Theorem 3.27), $S = \omega$ and the proof is complete. □

Lemma 5.2 *Every element of a natural number is a subset of that number.*

Proof Let $S = \{n \in \omega : x \subseteq n, \forall x \in n\}$.

Then $0 \in S$ by Theorem 2.6.

Let $n \in S$. Then for all $x \in n$, $x \subseteq n$. Consider $n^+ = n \cup \{n\}$. Now $x \in n \implies x \in n^+ \implies x \subseteq n^+$. Also, $n \in n^+$, and $\{n\} \subseteq n^+$. But this exhausts all the elements of n^+, and it follows that $n^+ \in S$. By Induction (Theorem 3.27) we have that $S = \omega$. □

We now state the postulates:

Theorem 5.3 (The Peano Postulates)

1. $0 \in \omega$
2. $n \in \omega \implies n^+ \in \omega$

3. *(Principle of Mathematical Induction): If $S \subseteq \omega$ and $0 \in S$ and $n^+ \in \omega$ whenever $n \in \omega$ then $S = \omega$*

4. $n^+ \neq 0, \forall n \in \omega$.

5. *If $n, m \in \omega$ and if $n^+ = m^+$, then $n = m$.*

Proof

1,2,3. Trivially, the results follow from Theorem 3.26 and Theorem 3.27.

4. By definition, $n^+ = n \cup \{n\}$. Since $n \in n^+$, it is impossible that $n^+ = \varnothing$.

5. We note that since $n^+ = m^+$ it follows that $m, n \in n^+$ and $m, n \in m^+$. So we have either $n \in m$ or $n = m$ or $m \in n$. If $n \neq m$, then $n \in m$ and $m \in n$, which implies $n \in n$ and Lemma 5.2, $n \subseteq n$. But this contradicts Lemma 5.1.

□

We have seen that Induction (Theorem 3.27) is indeed a very powerful tool in proving statements about the natural numbers. But this process can be used to also define functions. In particular we shall define addition, multiplication, and exponentiation using the following result.

With this motivation, we prove the following remarkable result.

Theorem 5.4 (The Recursion Theorem) *Let X be a set, and let $f : X \to X$. Fix $a \in X$. Then there exists a function $u : \omega \to X$ such that:*

$$\begin{cases} u(0) = a \\ u(n^+) = f(u(n)) \end{cases}$$

Proof Let:

$$\mathcal{C} = \left\{ A \subseteq \omega \times X : 0 \in A \wedge (n, x) \in A \implies \left(n^+, f(x)\right) \in A \right\}$$

Now $\mathcal{C} \neq \varnothing$, since $\omega \times X \in \mathcal{C}$. We define $u = \bigcap \mathcal{C}$. We prove that u is indeed a function.

Define:

$$S = \{n \in \omega : (n, x), (n, y) \in u \implies x = y\}$$

Firstly, $0 \in S$. We prove by contradiction. Suppose $0 \notin S$. Then there are $a \neq b \in X$ such that $(0, a), (0, b) \in u$. But certainly, $u \setminus \{(0, b)\} \in \mathcal{C}$ is guaranteed since $n^+ \neq 0$. But this contradicts the minimality of u.

Now, let $n \in S$. So $(n, x) \in u$. Suppose $n^+ \notin S$. Then, there exist some $(n^+, y) \in u$ with $f(x) \neq y$. We consider $u \backslash n^+, y$. Since $n^+ \neq 0$, $(0, a) \in u \backslash n^+, y$. And given $(m, z) \in u \backslash (n^+, y) \implies (m^+, f(z)) \in u \backslash (n^+, y)$, since $y \neq f(x)$. But again, this contradicts the minimality of u.

By Induction (Theorem 3.27), $S = \omega$. □

5.2 Arithmetic

We are now ready to define sums, products, and exponents. The existence of the recursive functions in the following definitions are guaranteed by Theorem 5.4.

Definition 5.5 (Sum) *Let $f : \omega \to \omega$ be defined by $f(n) = n^+$. Then for any $m \in \omega$, we find $u_m : \omega \to \omega$ such that $u_m(0) = m$ and $u_m(n^+) = f(u_m(n))$. We write this as $m + n$.*

Definition 5.6 (Product) *Let $f_m : \omega \to \omega$ be defined by $f_m(n) = m + n$. We find $p_m(0) = 0$ and $p_m(n^+) = f_m(p_m(n))$, for all $m, n \in \omega$. We denote this as $mn = m \cdot n$*

Definition 5.7 (Exponent) *Let $f_m : \omega \to \omega$ be defined by $f_m(n) = n \cdot m$, for all $m, n \in \omega$. Then, we find $e_m(0) = 1$ and $e_m(n^+) = f_m(e_m(n))$. We write this as m^n.*

It is straightforward, yet extremely tedious and beyond the scope of this essay to illustrate that these definitions actually do give us the familiar laws and properties of sums, products and exponents.

We simply recite the following theorems which highlight some of the important aspects of arithmetic. Their proofs are almost always an application of Induction (Theorem 3.27).

Definition 5.8 (Associative, Commutative) *We say that a function $* : A \times A \to A$ is associative if given for any $x, y, z \in A$, $(x * y) * z = x * (y * z)$. We say it is commutative if $x * y = y * x$, for all $x, y \in A$*

Theorem 5.9 (Associativity, Commutativity) *Sums, products, and exponents of natural numbers are associative and commutative.*

5.3 Comparability of the Natural Numbers

The motivation behind the following discussion is to define a notion of one number being greater or lesser than the other. We know from the use of natural numbers that any two are indeed lesser or greater than the other. We show that this is, in fact, a consequence of our construction.

Definition 5.10 (Comparability of Numbers) *We say that two numbers $m, n \in \omega$ are comparable if $m \in n$ or $n \in m$ or $n = m$.*

Theorem 5.11 (Comparability Theorem for Numbers) *Any two $n, m \in \omega$ are comparable.*

Proof Define:

$$\begin{aligned} S_n &= \{m \in \omega : n \text{ comparable to } m\} \\ S &= \{n \in \omega : S_n = \omega\} \end{aligned}$$

We show $S_0 \in \omega$. Take $0 \in \omega$. Then for any $0 \neq m \in \omega$, $0 \in m$. Otherwise $0 = m$. So $0 \in S_0$. Now suppose $k \in S_0$. So either $0 = k$ or $0 \in k$. Either way, $0 \in k^+$ and $k^+ \in S_0$. By Theorem 3.27, the result follows.

Now, suppose for some $0 \neq n \in \omega$, we have $S_n = \omega$. Now, $n \in n^+$, so we have $0 \in S_{n^+}$. Now take any $k \in S_{n^+}$. So, either $k \in n^+$ or $k = n^+$ or $n^+ \in k$. If $n^+ \in k$, then $n^+ \in k^+$. Now if $k \in n^+$, either $k \in n$ or $k \in \{n\}$. If $k \in \{n\}$, then $k = n$ and $k^+ = n^+$. Otherwise, $k \in n$ and we have that since $k \in S_n = \omega$ implies $k^+ \in S_n$. If, $k^+ = n^+$, there is nothing to do. Otherwise, we have $k^+ \neq n$. Then if $k^+ \in n \implies k^+ \in n^+$.

But since both $k^+, n \in S_n$, we must have $n \in k^+$. We cannot have $n^+ = k^+$ (which would imply $n = k$ by Lemma 5.1 and contradict our assumption), so we must have that $n^+ \in k^+$. But we have shown that n^+ and k^+ are comparable whenever k is. By Induction (Theorem 3.27) $S_{n^+} = \omega$.

We have also shown that $0 \in S$ and whenever $n \in S$, $n^+ \in S$. By Induction (Theorem 3.27) again, and we have that $S = \omega$. □

From this result, we have the following spectacular consequence.

Corollary 5.12 *If $m, n \in \omega$, then only one of the following hold:*

1. $m \in n$
2. $m = n$
3. $n \in m$

Proof Fix $m, n \in \omega$. By Theorem 5.11, we have that they are comparable. Now, suppose that $m \in n$ and $m = n$. But then $m \in m$ which is a contradiction to Lemma 5.1. A similar contradiction occurs when if we suppose $n \in m$ and $m = n$. Suppose that $m \in n$ and $n \in m$. But by Lemma 5.2, $m \subseteq n$. But this implies $n \in n$ which again contradicts Lemma 5.1. □

These asserts provide a sound motivation for attempting to define ordering of natural numbers. We want that for any two natural numbers m, n, only one of the following are true: $m = n$, $m < n$ or $m > n$. This looks very similar to Corollary 5.12, and highlights the importance of this result. We characterise this informal discussion into a rigorous definition.

Definition 5.13 (Lesser, Greater for Natural Numbers) *Let $m, n \in \omega$. If $m \in n$, we say m is less than n or n greater than m and write $m < n$ or $n > m$ respectively. If $m < n$ or $m = n$, we write $m \leq n$.*

5.4 Infinity

We have seen an axiom called "infinity," and we have an intuitive understanding of saying ad infinitum. This, however, lacks the mathematical rigour we seek. First, we characterise a way to "compare" two arbitrary sets.

Definition 5.14 (Set Equivalence) *We say that sets A and B are equivalent if there exists a bijection $f : A \to B$. We write $A \sim B$.*

Theorem 5.15 *Set Equivalence is an equivalence relation.*

Proof Let A, B, C be sets. Trivially, $A \sim A$, just take function $f : A \to A$ defined by $f(x) = x$. Again, trivially, if $A \sim B$ then $B \sim A$, by Theorem 3.16.

Now, suppose $A \sim B$ and $B \sim C$. Then, let $f : A \to B$ and $g : B \to C$ be the promised bijections. Trivially, $g \circ f : A \to C$ is a bijection by Theorem 3.20. □

The following results highlight some important facts. Their proofs can be found in [3, §13]

Theorem 5.16

1. *Every proper subset of a natural number is equivalent to a smaller natural number.*
2. *For $n \in \omega$, there exist no $k \subsetneq n$ with $n \sim k$.*

This notion of set equivalence can be used to define infinity.

Definition 5.17 (Infinity) *A set S is finite if it is equivalent to some natural number. Otherwise it is called infinite.*

We embrace our first infinite set:

Theorem 5.18 *The set ω is infinite.*

Proof Let $S = \{n \in \omega : \omega \not\sim n\}$. Trivially, $0 \not\sim \omega$.

Now suppose $n \in S$, and suppose $n^+ \sim \omega$. So there exists a bijection $f : n^+ \to \omega$. Now consider the restriction $f|_n : n \to \omega$. So $\operatorname{im}(f|_n) = \omega \setminus x$, where $f(n) = x$. Now consider the function $s : \omega \to \omega$ defined by $s(y) = y, \forall 0 \neq y \neq x$. And $s(0) = x$ and $s(x) = 0$. Certainly, this is a bijection, since we are simply swapping 0 and x. Now, the function $g : \omega \to \omega \setminus \{0\}$ defined by $g(y) = y^+$ is trivially a bijection. But then, $s \circ g : \omega \to \omega \setminus x$ is a bijection. Now, by Theorem 5.15, equivalence is transitive and it follows $\omega \sim n$, a contradiction. So, $n^+ \in S$. Our result follows from Induction (Theorem 3.27). □

The following results are important, but their proofs are straightforward and are omitted.

Theorem 5.19

1. *A finite set is equivalent to a unique natural number.*
2. *Every subset of a natural number is finite.*

These results motivate a way we can "count" the number of elements in a set. The following definition makes such a characterisation clear.

Definition 5.20 (Number of Elements) *Let F be a finite set. Then let $n \in \omega$ be the unique natural number such that $F \sim n$. Then we say that F has n elements and write $|F| = n$.*

The following result confirms what we expect.

Theorem 5.21 *If $E \subseteq F$, then $|E| \leq |F|$.*

6 The Axiom of Choice

6.1 The Axiom

We present the celebrated and widely debated Axiom of Choice.

Axiom 9 (The Axiom of Choice) *Let $\varnothing \neq \mathcal{C} = \{A_\alpha\}$ where $\alpha \in \Phi$ be a family of nonempty sets. Then the product*

$$\prod \mathcal{C} = \prod_{\alpha \in \Phi} A_\alpha$$

is nonempty.

We present the following result to assert fact that this axiom is a consequence when we are dealing with finite families.

Theorem 6.1 (Finite Axiom of Choice) *Let $\varnothing \neq \mathcal{F} = \{F_i\}$ $i \in n^+ \in \omega$ be a finite family of nonempty sets. Then $\prod_{i \in n} F_i \neq \varnothing$.*

Proof Let $S = \left\{n \in \omega : \prod_{i \in n^+} X_i \neq 0, \forall X_i \neq 0\right\}$.

Now, suppose $n = 0$. Then $i = 0$ and then $\mathcal{F} = X_0$. Trivially $\prod \mathcal{F} \neq \varnothing$ since $X_0 \neq \varnothing$ by hypothesis.

Suppose $n \in S$. So, given $\{X_i\}$ nonempty family for $i \in (n^+)^+$, we know that $\prod_{i \in n^+} X_i \neq \varnothing$. It follows that:

$$\prod_{i \in (n^+)^+} X_i = \prod_{i \in n^+} X_i \times X_{n^+}$$

which must be nonempty since $X_{n^+} \neq \varnothing$ and by invoking Theorem 3.3(4).

The expected result follows from Induction (Theorem 3.27). □

Our formulation of Axiom 9 is not, in fact, the original way in which this axiom was stated [10, p186]. In our notation, this can be thought of as guaranteeing the existence of a choice function [5, p9]. This motivates the following results.

Theorem 6.2 (Existence of a Choice Function) *For every set $\mathcal{C} \neq \varnothing$ of nonempty subsets of a set X there exists a function $f : \mathcal{C} \to X$ such that $f(A) \in A$.*

Proof By Axiom 9, $\prod_{X \in \mathcal{C}} X \neq \varnothing$. So, $\{x_X\}_{X \in \mathcal{C}} \in \prod \mathcal{C}$. But by Definition 3.23, $x : \mathcal{C} \to X$, such that $x(X) \in X$. □

Corollary 6.3 (Equivalence of Axiom of Choice to Choice Function) *Theorem 6.2 $\iff$ Axiom 9*

Proof We need to only show ($\implies$). Assume Theorem 6.2 is true. Let $f : \mathcal{C} \to X$ be the promised function. Since $f(A) \in A$ by hypothesis, we put $f_A = f(A)$, and by Definition 3.23 $\{f_A\}_{A \in \mathcal{C}} \in \prod \mathcal{C}$. But this implies that $\prod \mathcal{C}$ is nonempty. □

The following two important results are a direct application of the Choice axiom (Axiom 9) and the Recursion Theorem (Theorem 5.4). A proof sketch can be found in [3, p61].

Theorem 6.4 *Every infinite set has a subset equivalent to ω.*

The following consequence was used by Dedekind as the very definition of infinity [3, p61].

Corollary 6.5 *A set is infinite if and only if it is equivalent to a proper subset of itself.*

6.2 The Debate

We shall discuss the motivations for this axiom informally. The motivation comes from the fact that we want to be able to make choices - i.e., allow us to "choose" elements from a set. As we have seen in Theorem 6.1, the finite case is an almost trivial consequence. It is really in the infinite case that this process of choosing becomes non-trivial, because the question becomes philosophical: what does it mean to choose an infinite number of elements?

Without raising too many eyebrows, we attempt to provide an answer. We look beyond the problem of choosing and concentrate instead on the Choice Axiom (Axiom 9). Our result in Corollary 6.3 tells us that this is equivalent to the problem of making arbitrary choices, so it is valid that we shall instead concentrate on this more "tangible" formulation. We have seen by Theorem 6.1 that it is indeed true that the finite product of a nonempty family of nonempty sets is nonempty. Given that our current model is not strong enough to satisfactorily answer in the general case, and on the assumption that a generalised axiom (Axiom 9) does not lead to logical contradictions, why should we not accept it? If we know something is true in the finite case, why should we dismiss a generalisation to the infinite without good reason?

Underlying this reasoning was the assumption that the axiom does not give rise to contradictions. Earlier in this century, when the debate was at its peak, the problem of consistency was the greatest philosophical objection. In 1938 Kurt Gödel obtained results with regards to models of set theory and showed that Zermelo-Frankel Axioms were consistent[3] with The Choice Axiom [6, p3]. This dispelled any fears of contradictions. It was also shown that The Choice Axiom was independent of the Zermelo-Frankel system. A surprising consequence was that the negation of the axiom is also consistent with the Zermelo-Frankel system!

It is interesting to note that this axiom had been widely used, but implicitly, preceding its formal characterisation by Zermelo [6, p8]. In fact, up until the 1960's, instances of the implicit use of the axiom were discovered. A more thorough survey can be found in [6].

Invoking the Choice Axiom seems unavoidable whenever we enumerate a set. Consider proving that the union of a countable family of sets is countable. There is a sense of enumeration at some level in such a proof. We need to order the elements, and immediately, we make choice. In fact, a more restricted version of this axiom can then be employed. This axiom known as the Countable Axiom of Choice restricts the family to a countable set of sets. In light of our preceding discussion, there are models of set theory in which the negation of the axiom is assumed, and the real numbers - an uncountable set - is then the union of a countable set of sets [6, p9]. This is an example of the fact that the Choice Axiom is indeed independent.

The Tychonoff Product Theorem [7, §37] is an example of a result that uses the full strength of Axiom 9. This theorem guarantees that the product of a family of compact sets is also compact. Such a result is deep and profound, and can be used to prove many extraordinary existential results in analysis. Another is the Vitali Covering Theorem [1, p142], an extremely important result in Measure Theory, which guarantees the existence of a countable set of closed balls that "fills" a measurable set and misses only a set of points of zero measure [6, p246]. This theorem is generally proved by invoking *Zorn's Lemma*, which as we shall see later, happens to be an equivalence of the Axiom of Choice. Out of all such fantastic results, probably one of the most startling is known as the Banach-Tarski Paradox [11, p27]. Informally, this paradox tells us that we can cut a sphere into finite parts, reassemble them into two spheres with the sum of the volumes being unequal to the volume of the original!

All these results have something in common. Their existence is proven in a non-constructive way. It is the constructivists who are not persuaded by Gödel's 1938 result in embracing Choice. Their criticism is that all existence proofs should explicitly construct the object in question. Yet, such a demand would simply reduce mathematics to a collection of algorithms.

Gödel's result eased the anxieties of many philosophers and mathematicians of the potential danger of contradictions in assuming the Choice Axiom. It is indeed a sigh of relief that the axiom is now widely accepted and used by a large section of the mathematical community.

[3] Assuming that Zermelo-Frankel Axioms themselves are consistent. This cannot, in fact, be proved.

7 Zorn's Lemma

In this section, we shall prove an extremely useful result that is equivalent to the Choice Axiom. It is a maximal principle published by Max Zorn in 1935, where such a principle was used for the first time in Algebra [5, p35]. Time and time again, this principle has become wonderfully useful. It seems that it is often easier to phrase many existential problems in the language of Zorn.

The language of Zorn is somewhat more complex than the ideas that we have so far discussed. We introduce the machinery of *order* to understand and prove this remarkable result.

7.1 Order

We begin with the most general notion of order.

Definition 7.1 (Partial Order) *Let $(X, \leq)$ denote a set X with relation $\leq$. Suppose that:*

1. *$x \leq x$, for all $x \in X$ (Reflexive)*
2. *If $x \leq y$ and $y \leq x$ then $x = y$ (Anti-Symmetric)*
3. *If $x \leq y$ and $y \leq z$, then $x \leq z$ (Transitive)*

The we say that $(X, \leq)$ or simply X is a partially ordered set.

Definition 7.2 (Chain/Total Order) *Let $(X, \leq)$ be a partially ordered set. If for all $x, y \in X$, either $x \leq y$ or $y \leq x$, then $(X, \leq)$ is a chain or total order.*

We have used the notation symbol $\leq$ explicitly. We are familiar with the use of less and greater with orderings and we shall generalise this language to arbitrary orderings. The connection between $\leq$ and $<$ should also be obvious. In fact, we could have equally formulated our definitions in terms of $<$ rather than $\leq$.

Definition 7.3 ((Weak) Initial Segment) *Let $(X, \leq)$ be a partial order. Then, let $S_a = \{x \in X : x < a\}$, for some $a \in X$. Then S_a is the initial segment/section of X by a. Similarly $S'_a = \{x \in X : x \leq a\}$ is called weak.*

Sometimes, a segment is also called a section.

Definition 7.4 (Minimal, Maximal) *Let $(X, \leq)$ be a partial order. Suppose there exists $l, m \in X$ such that there exist no $x \in X$ with $x > l$ or $m < x$. Then l is a minimal element of X and m is a maximal element in X.*

Definition 7.5 (Minimum, Maximum) *Let $(X, \leq)$ be a chain. If there exist $l, m \in X$ such that for all $x \in X$, either $l \leq x$ or $m \geq x$, then l is called a minimum, and m the maximum.*

In these two definitions, it is important to emphasise that there can be ordered sets that can have a maximum or maximal without a minimum or a minimal element. And conversely.

We make a connection now between these two types of elements.

Theorem 7.6 (Minimality in Chains) *Let $(X, \leq)$ be a chain. Then:*

1. *l is minimal if and only if l is the minimum*
2. *m is maximal if and only if m is the maximum*

Proof Suppose that l is minimal. Then for all $x \in X$, we have $x \nless a$. But since we have a total order, this implies $x \geq a$.

Now for the converse, assume that l is minimum. But this implies that for all $x \in X$, $x \geq l$, which implies that there exist no $x \in X$ such that $x < l$.

The proof is trivially similar for the maximal case. □

Definition 7.7 (Upper/Lower Bound) *Let $(X, \leq)$ be a partially ordered set, and let $A \subseteq X$. Then, if for all $a \in A$, there exist a $u \in X$ such that $a \leq u$, we say that u is an upper bound for A and that A is bounded above. Similarly, if for all $a \in A$, there exists a l such that $l \leq a$, then we say that l is a lower bound and that A is bounded below.*

A quick remark: a set may be bounded above, below, neither or both.

Definition 7.8 (Infimum, Supremum) *Let $(X, \leq)$ a partially ordered set, and let $E \subseteq X$. Then let U be the set of upper bounds on E and L be the set of lower bounds on E. If U attains a minimum, the we say $\sup(E) = \min(U)$ the supremum of E. If L attains a maximum, then we say $\inf(E) = \max(L)$ the infimum of E.*

The following theorem, although trivial in proof, is an important result.

Theorem 7.9 *Let $(X, \leq)$ be a partially ordered set, and $E \subseteq X$. Then:*

1. *If E attains a minimum l, then $\inf(E) = l$.*
2. *If E attains a maximum m, then $\sup(E) = m$.*

And finally, we introduce the notion of comparability for arbitrary sets.

7.2 The Lemma

The proof is non-trivial. The proof here is an adaptation of the one presented in [3, §16] In order to make the proof more tangible, we shall present a collection of lemmas and definitions. This may seem dry and abstract, but the other approach - of present the proof in whole - is overwhelmingly confusing.

Lemma 7.10 *Let $(X, \leq)$ be a partially ordered set. Define $s : X \to \wp(\wp(X))$ by $s(x) = \{z \in X : z \leq x\}$. Let $(\mathcal{S} = \operatorname{im}(s), \subseteq)$ ordered by set inclusion. Then $x \in X$ is maximal if and only if $s(x)$ is maximal in $\mathcal{S}$.*

Proof Suppose that $m \in X$ is maximal. Then, there is no $x \in X$ such that $x > m$. Suppose that $s(m)$ was not maximal in $\mathcal{S}$. Suppose $s(m) \subsetneq s(x)$. But $s(m)$ is a weak initial segment, which implies $m \in s(m)$ and we conclude that $m < x$, contradicting the maximality of m.

Now for the converse, assume that $s(m)$ is maximal. That is, there is no $s(x)$ such that $s(m) \subsetneq s(x)$. If m was not maximal in X, we find $x > m$, and trivially $s(m) \subsetneq s(x)$ contradicting maximality of $s(m)$. □

In the following definition, we have arbitrarily chosen its name for convenience.

Definition 7.11 (Fairy Set) *Let $(X, \leq)$ be a partially ordered set, and define $\mathcal{X} = \{\mathcal{C} \in \wp(X) : \mathcal{C}$ is a chain $\}$, with $(\mathcal{X}, \subseteq)$ ordered by set inclusion. Then $(\mathcal{X}, \subseteq)$ is the Fairy Set of X.*

Lemma 7.12 *Let $(X, \leq)$ be a partially ordered set. Let $(\mathcal{X}, \subseteq)$ be the Fairy of X. Then if $\mathcal{M} \subseteq \mathcal{X}$ is a chain in $\mathcal{X}$, then $\bigcup \mathcal{M}$ is a chain in X.*

Proof Let $x, y \in \bigcup \mathcal{M}$. Then, there are elements $\mathcal{C}_x, \mathcal{C}_y \in \mathcal{M}$ such that $x \in \mathcal{C}_x$ and $y \in \mathcal{C}_y$. But by hypothesis, we've ordered $\mathcal{X}$ by set inclusion and $\mathcal{M}$ is a chain. So, either $\mathcal{C}_x \subseteq \mathcal{C}_y$ or $\mathcal{C}_y \subseteq \mathcal{C}_x$. Without loss of generality, assume the latter. Then $x, y \in \mathcal{C}_x$. But again, by hypothesis, $\mathcal{C}_x$ is a chain in X. It follows then that either $x \leq y$ or $y \leq x$. □

We introduce some new terminology.

Definition 7.13 (Adjunction Set) *Let $(X, \leq)$ be a partially ordered set, and $(\mathcal{X}, \subseteq)$ the Fairy of X. Then for any $A \in \wp(X)$, the set $\hat{A} = \{x \in X : A \cup \{x\} \in \mathcal{X}\}$ is called the adjunction of A.*

Definition 7.14 (Adjunction Function) *Let $(X, \leq)$ be a partially ordered set and $(\mathcal{X}, \subseteq)$ the Fairy of X. We denote the adjunction of $A \in \wp(X)$ by $\hat{A}$. Let $c : \wp(X) \setminus \varnothing \to X$ be a choice function. Define $\phi : \mathcal{X} \to \mathcal{X}$ by*

$$\phi(A) = \begin{cases} A \cup \left\{ c(\hat{A} \setminus A) \right\} & , \hat{A} \setminus A \neq \varnothing \\ A & , \hat{A} \setminus A = \varnothing \end{cases}$$

Lemma 7.15 *Let $(X, \leq)$ be a partially ordered set with Fairy $(\mathcal{X}, \subseteq)$. Let $\phi : \mathcal{X} \to \mathcal{X}$ be the Adjunction function. Then $\phi(A) = A \iff A$ is maximal in $\mathcal{X}$.*

Proof Suppose that $\phi(A) = A$. Then, $\hat{A} \setminus A = \varnothing$. That is, $x \in \hat{A} \iff x \in A$. Then we can find no $x \in X \setminus A$ such that $A \cup \{x\} \in \mathcal{X}$. But that is exactly saying that for no $x \in X \setminus A$ is $A \cup \{x\}$ a chain. It follows that A is maximal in $\mathcal{X}$.

Now assume that A is maximal in $\mathcal{X}$. So, there exist no $A' \supsetneq A$ with A' a chain in X. That is, there exist no $x \in X \setminus A$ such that $A \cup \{x\} \in \mathcal{X}$. It follows then that if $x \in \hat{A} \implies x \in A$. By construction of the Adjunct, trivially $A \subseteq \hat{A}$. It follows that $A = \hat{A}$ and $\hat{A} \setminus A = \varnothing$. So, $\phi(A) = A$. □

Definition 7.16 (Tower) *Let $(X, \leq)$ be a partially ordered set, and let $(\mathcal{X}, \subseteq)$ be the Fairy of X. Let $\phi : \mathcal{X} \to \mathcal{X}$ be the adjunction set function. Then if there exists a $\mathcal{T} \in \mathcal{X}$ satisfying:*

1. *$\varnothing \in \mathcal{T}$*
2. *$A \in \mathcal{T} \implies \phi(A) \in \mathcal{T}$*
3. *$\mathcal{C} \in \mathcal{T}$ a chain in $X \implies \bigcup \mathcal{C} \in \mathcal{T}$*

Then $\mathcal{T}$ is a tower in $\mathcal{X}$.

Definition 7.17 (Comparability) *If $(\mathcal{X}, \subseteq)$ is a partial ordering, then $C \in \mathcal{X}$ is called comparable if for all $A \in \mathcal{X}$, either $C \subseteq A$ or $A \subseteq C$.*

Lemma 7.18 *Let $(X, \leq)$ be a partially ordered set, with Fairy $(\mathcal{X}, \subseteq)$. Let $\mathcal{M} = \{\mathcal{T} \subseteq \mathcal{X} : \mathcal{T}$ is a tower $\}$. Then $\varnothing \neq \mathcal{T}_0 = \bigcap \mathcal{C}$ is a chain in $\mathcal{X}$.*

Proof Observe that $\mathcal{X}$ is a tower itself, so $\mathcal{M} \neq \varnothing$. Then, note that $\varnothing \in \mathcal{T}$ for any tower $\mathcal{T}$. It follows then that $\mathcal{T}_0 \neq \varnothing$.

Let $A \in \mathcal{T}_0$. Let $\phi : \mathcal{X} \to \mathcal{X}$ be the Adjunction function. Fix $C \in \mathcal{T}_0$, a comparable set. We note that comparable sets exist in $\mathcal{T}_0$ The set $\varnothing \in \mathcal{T}_0$ is an example.

Now, we prove: if $A \in \mathcal{T}_0$ and $A \subsetneq C$, then $\phi(A) \subseteq C$. We know that C is comparable. So either $\phi(A) \subseteq C$ or $C \subsetneq \phi(A)$. If the latter is true, then $\phi(A) \setminus A$ will differ by more than one element which is impossible by the definition of the Adjunction function.

Consider the set $\mathcal{U} = \{A \in \mathcal{T}_0 : A \subseteq C \vee \phi(C) \subseteq A\}$. It is obvious by construction that $\mathcal{U}$ is indeed a chain. We claim that $\mathcal{U}$ is a tower.

Now trivially, $\varnothing \in \mathcal{T}_0$. Let $A \in \mathcal{T}_0$. Consider $\phi(A)$. We split into three cases. Case 1: $A \subsetneq C$. But our previous result gives us that $\phi(A) \subseteq C \implies \phi(A) \in \mathcal{U}$. Case 2: $A = C$. Then, $\phi(A) = \phi(C) \implies \phi(C) \subseteq \phi(A)$, and it follows that $\phi(A) \in \mathcal{U}$. Case 3: $\phi(C) \subseteq A$. Then since $A \subseteq \phi(A) \implies \phi(A) \in \mathcal{U}$.

Now, let $\mathcal{C} \in \mathcal{U}$ be a chain. So, for all $A \in \mathcal{C}$, either $A \subseteq C$ or $\phi(C) \subseteq A$. If $\phi(C) \in \mathcal{C}$ then we have $\phi(C) \in \bigcup \mathcal{C}$. If $A \subseteq C \implies A \subseteq \phi(C)$. Otherwise, consider when all $A \in \mathcal{C}$ implies $A \subseteq C$. Then, $\bigcup \mathcal{C} \subseteq C$. In either case, we have $\bigcup \mathcal{C} \in \mathcal{U}$.

By construction $\mathcal{U} \subseteq \mathcal{T}_0$. But we must conclude $\mathcal{U} = \mathcal{T}_0$ to avoid contradicting the minimality of $\mathcal{T}_0$. It follows that $\mathcal{T}_0$ is indeed a chain in $\mathcal{X}$. □

We present the celebrated *Zorn's Lemma*.

Theorem 7.19 (Zorn's Lemma) *If $(X, \leq)$ is a partially ordered set in which every chain is bounded above, then there exists a maximal element in X.*

Proof Let $s : X \to \wp(\wp(X))$ be defined as the weak initial segment by x. By Lemma 7.10, $m \in X$ maximal if and only if $s(m)$ maximal in $(\mathrm{im}(s), \subseteq)$.

Let $(\mathcal{X}, \subseteq)$ denote the Fairy of X. Note that for any $\mathcal{C} \in \mathcal{X}$, there exists an $x \in X$ such that $\mathcal{C} \subseteq s(x)$. But trivially, $s(m)$ maximal in $(\mathrm{im}(s), \subseteq)$ if and only if $s(m)$ maximal in $(\mathcal{X}, \subseteq)$.

In this construction, we can forget X altogether, and consider the problem of finding a maximal element in $(\mathcal{X}, \subseteq)$. By Lemma 7.12, for a chain $\mathcal{M} \subseteq \mathcal{X}$, $\bigcup \mathcal{M}$ is an upper bound for $\mathcal{M}$.

Let $\phi : \mathcal{X} \to \mathcal{X}$ denote the Adjunction function, and let $\mathcal{T}_0$ be the promised chain in $\mathcal{X}$ by Lemma 7.18. Trivially, $\mathcal{T}_0$ is also a tower, so for any $A, C \in \mathcal{T}_0$ $A \subseteq C \implies A \subseteq \phi(C)$ or $\phi(C) \subseteq A$.

Now consider the set $\mathcal{U} = \bigcup \mathcal{T}_0$. By definition of ϕ, $\phi(\mathcal{U}) \supseteq \mathcal{U}$. But for all $\mathcal{A} \in \mathcal{T}_0$, $\mathcal{A} \subseteq \mathcal{U}$. So, either $\mathcal{A} \subsetneq \mathcal{U} \implies \phi(\mathcal{U}) \subseteq \mathcal{U}$ (sub-result within Lemma 7.18) or $\mathcal{A} = \mathcal{U}$. In either case, $\phi(\mathcal{U}) = \mathcal{U}$. By Lemma 7.15, $\mathcal{U}$ is maximal. □

8 The Well Ordering Theorem

In this section, we prove the Well Ordering Theorem by applying Zorn's Lemma (Theorem 7.19). Before we begin to consider technicalities, we note that Zorn's Lemma was formulated much after Zermelo's proof of the Well Ordering Theorem. The original proof directly involves the Choice Axiom. The original proof can be found in [10, p183].

Preceding the proof, we shall present some definitions and theorems regarding well orderings.

Definition 8.1 (Well Ordered Set) *Let $(W, \leq)$ be a partially ordered set. Then if every subset of W attains a minimum, we say that W is well ordered.*

In the light of this definition we highlight a simple, yet insightful result.

Theorem 8.2 *Every well ordered set is a chain.*

Proof Let W be a well ordered set. Then for any $x, y \in W$, we the set $\{x, y\} \subseteq W$. But then either x or y is the minimum. That is exactly, $x \leq y$ or $y \leq x$. □

Definition 8.3 (Continuation) *Let $(A, \leq)$ and $(B, \leq)$ be well ordered sets, such that $A \subseteq B$. Then if A is an initial segment of B, we say that B is a continuation of A and write $A \Subset B$.*

We present some important auxiliary results regarding continuations in order to prove the Well Ordering Theorem.

Lemma 8.4 *Let $(\mathcal{C}, \Subset)$ be a chain of well ordered sets ordered by continuation. Then each $A \in \mathcal{C}$ has the same minimum.*

Proof Assume the converse is true. Let $B \in \mathcal{C}$ and let m_B be minimum of B. Let $A \in \mathcal{C}$ with $B \Subset A$ with a minimum $m_A \neq m_B$. Such a set must exist, otherwise the result is trivial. By continuation hypothesis, there exists an $b \in A$ such that $B = s(b)$.

Now, either $m_B < m_A$ or $m_B > m_A$ by Theorem 8.2. Note that $m_B < m_A \implies B \not\subseteq A$. Otherwise $m_B > m_A \implies s(B) \neq B$. In either case, we have a contradiction, and the result follows. □

Lemma 8.5 *Let $(\mathcal{C}, \Subset)$ a chain of well ordered sets ordered by continuation. Then $\bigcup \mathcal{C}$ is an upper bound for $\mathcal{C}$.*

Proof For every $A \in \mathcal{C}$, we construct $A' = \{(x, y) : x \leq y \in A\}$ (we have this automatically by definition, but we do this for emphasis). Now, let $\mathcal{C}'$ be the collection of all A'. By Lemma 8.4, let m be the minimum of each $A \in \mathcal{C}$.

Now, consider $\bigcup \mathcal{C}'$. Then $(m, y) \in \bigcup \mathcal{C}'$, for all $y \in \bigcup \mathcal{C}$. If $(m', m) \in \bigcup \mathcal{C}'$ with $m \neq m'$, then there exists some $B \in \mathcal{C}$ with minimum $m' \in B$ with $m' < m$. But this is impossible by Lemma 8.4.

Then, given any $A \in \mathcal{C}$, either $A = \bigcup \mathcal{C}$ or there exists a $B \in \mathcal{C}$ such that $A \Subset B$. In the former case, we're done - nothing to prove. In the latter case, there must exist a $b \in B \in \bigcup \mathcal{C}$ such that $s(b) = A \iff s'(b) = A' = \{(x, y) : x < y < b\} \in \mathcal{C}'$. So, $\bigcup \mathcal{C}'$ gives the required well ordering for $\bigcup \mathcal{C}$, and for all $A \in \mathcal{C}$ such that $A \neq \bigcup \mathcal{C}$, $A \Subset \bigcup \mathcal{C}$. □

Now we present the celebrated result. It has been adapted from [3, p69].

Theorem 8.6 (The Well Ordering Theorem) *Every set can be Well Ordered.*

Proof Let X be a set, and let $\mathcal{W}$ be a collection of well ordered subsets of X. Now, $\mathcal{W} \neq 0$ since $\varnothing \in \mathcal{W}$ or if $X \neq \varnothing$, for any $x \in X$, $\{x\} \in \mathcal{W}$. We order $\mathcal{W}$ by continuation.

Now, for any chain $\mathcal{C} \subseteq \mathcal{W}$, $\bigcup \mathcal{C}$ is an upper bound for $\mathcal{C}$. We apply Zorn's Lemma (Theorem 7.19), and find a maximal well ordered set $W \subseteq X$.

Our claim is that $W = X$. Suppose this is not true. Then, $X \setminus W \neq \varnothing$. Take any $z \in X \setminus W$, and construct $W' = W \cup \{z\}$. We order W' by $x \leq y$ if $x \leq y \in W$. For all $x \in W$, we set $x < z$. Then W' is a well ordered set, $W \Subset W'$ and $W' \in \mathcal{W}$. But this contradicts the maximality of W. □

We shall emphasise that this theorem does not guarantee anything about an order structure which may pre-exist in the set. In fact, the well ordering may be quite different to the usual ordering of a set. For instance, consider the real numbers $\mathbb{R}$. Surely, $\mathbb{R}$ under its usual ordering is not a well ordering. The well ordering theorem simply states that we can find an ordering for $\mathbb{R}$ such that it is a well ordering.

Theorem 8.7 *Well Ordering Theorem (Theorem 8.6)* $\implies$ *Choice Axiom (Axiom 9)*

Proof We show that Theorem 8.6 implies Theorem 6.2. Then, by Corollary 6.3, we obtain the desired result.

Let $\mathcal{C}$ be a family of nonempty sets indexed by Φ. We well order each $X_\Phi \in \mathcal{C}$. The, we define the function $c : \mathcal{C} \to \bigcup \mathcal{C}$:

$$c(X_\alpha) = \min(X_\alpha)$$

Now, c is defined since we well order each X_α. Trivially, for every $\alpha \in \Phi$, $c(X_\alpha) \in X_\alpha$. This is the desired choice function. □

Now we come to a hallmark result, which illustrates how the process of making arbitrary choice are linked to maximality and order.

Corollary 8.8 (Equivalence of Choice, Zorn, and Well Ordering) *The following are equivalent:*

1. *The Axiom of Choice (Axiom 9)*
2. *Zorn's Lemma (Theorem 7.19)*
3. *Well Ordering Theorem (Theorem 8.6)*

Proof Trivially, Theorem 7.19 shows that the Choice Axiom implies Zorn's Lemma. Then Theorem 8.6 shows that Zorn's Lemma implies Well Ordering Theorem. Then Theorem 8.7 shows that the Well Ordering Theorem implies the Choice Axiom. □

9 The Transfinite

One of the most fantastic aspects of well ordered sets is the fact that they resemble the structure of the natural numbers. Although this resemblance is not exact, it is sufficient to generalise some of the most wonderful facts about the natural numbers. The most natural starting place is to extend the process of mathematical induction to well ordered sets - which by the Well Ordering Theorem - is any set.

We shall also briefly look at extending the process of counting using well orderings. Disclaimer: The detailed examination of such a theory is beyond the scope of this essay. Our discussion will be informal, and rarely shall we provide proof.

9.1 Transfinite Induction and Recursion

We state and prove the *Principle of Transfinite Induction*.

Theorem 9.1 (Principle of Transfinite Induction) *Let $(X, \leq)$ be a well ordered set, and let $S \subseteq X$. Then if for all $x \in X$, $s(x) \subseteq X \implies x \in S$, then $S = X$.*

Proof Suppose $S \neq X$. So, $X \setminus S \neq \varnothing$. Then take the $m = \min(X \setminus S)$. We have that $s(m) \subseteq S \implies m \in S$. But by construction $m \in X \setminus S$, which is a contradiction. □

There are two points to emphasise here. Firstly, there is no starting element. But given $(X, \leq)$ a well ordered set with a minimum m and a set S as in the hypothesis of Theorem 9.1. Then $s(m) = \varnothing$ and $\varnothing \subseteq S$ from Theorem 2.7, which implies that $m \in S$ by hypothesis. In other words, our formulation gives us the initial element in S for free. Secondly, rather than jumping from a predecessor to a successor, we jump from a set of predecessors to successors. This highlights one of the differences of arbitrary well ordered sets and the natural numbers. Every natural number has a predecessor, whereas elements of well ordered sets need not necessarily attain a strict predecessor.

The following result tells us that the *Principle of Transfinite Induction* is indeed a generalisation of Induction. That is, they are both equivalent on ω.

Theorem 9.2 *Transfinite Induction on $\omega \iff$ (Finite) Induction*

Proof Let $S \subset \omega$ such that the Transfinite Hypothesis holds. Then, $s(0) = \varnothing \implies 0 \in S$. Also, let $s(x^+) \subseteq S \implies x^+ \in S$. But $s(x^+) \subseteq S \iff x \in s(x^+)$. So, $x \in S \implies x^+ \in S$ and we have $S = \omega$.

Now we prove the converse. Suppose the Induction hypothesis holds. Fix $y \in S \implies y^+ \in S$. Consider the set $s(y^+)$. Trivially, by the Induction hypothesis, $0 \in s(y^+)$ and in general we have for all $x \leq y \implies x^+ \in S$. So, $s(y^+) \subseteq S \iff y \in S \implies y^+ \in S$. Since Induction holds, $S = \omega$ and we're done. □

As we were able to define functions using the process of Induction, we show a similar theorem in the light of Transfinite Induction. First, we introduce some language to express our theorem.

Definition 9.3 (Sequence of type a) *Let $(W, \leq)$ be a well ordered set, and X an arbitrary set. Then for any a, a function $f_a : s(a) \to X$ is called a sequence of type a.*

Definition 9.4 (Sequence function of type W in X) *Let $(W, \leq)$ be a well ordered set and X an obituary set. Let $\mathfrak{U} = \{f_a : s(a) \to X : \forall a \in W\}$. If there exists a $f : \mathfrak{U} \to X$, then f is called a sequence of type W in X.*

These definitions may seem rather obscure. We state them for rigour. Informally speaking, sequence functions tell us how to extend a sequence [3, p70].

We now state the Transfinite Recursion Theorem. We only prove uniqueness. The existence proof is a long and tedious construction of the function as a set of ordered pairs. The proof is explored in [3, p71].

Theorem 9.5 (Transfinite Recursion Theorem) *Let $(W, \leq)$ be a well ordered set, X some arbitrary set, and f a sequence of type W in X. Then there exists a unique $u : W \to X$ such that $u(a) = f(u|_{s(a)})$.*

Proof We show that u is unique, so assume u exists. Let $S = \big\{a \in W : u|_{s(a)} \text{ is unique } \big\}$. We note that $S \neq \varnothing$, since $m \in S$ for $m = \min(W)$. Trivially, whenever $s(a) \subseteq S$, $u|_{s(a)}$ unique, and it follows that $a \in S$. By Transfinite Induction (Theorem 9.1), $S = W$, and we conclude $u : W \to X$ is unique. □

9.2 Ordinals

We have mentioned before that we would in fact like to extend the process of counting beyond the natural numbers. But what exactly does this mean? Cantor's work revealed that we can in a sense "measure" sizes of infinity. That is to say, we can generalise what it means for two sets to contain the same "number" of elements to infinite sets. The motivation in this section is to investigate how to give meaning to this notion of "number."

Definition 9.6 (Similarity) *Let $(X, \leq_X)$ and $(Y, \leq_Y)$ be two partially ordered sets. Then if there exists a bijection $f : X \to Y$ such that $x \leq_X y \iff f(x) \leq_Y f(y)$, then we say that X is similar to Y, and write $X \simeq Y$. We call f a similarity.*

Although we've stated this definition in general for partially ordered sets, its importance becomes apparent when we consider well ordered sets. We state without proof, the following startling result. The proof can be found in [3, p73].

Theorem 9.7 (Comparability Theorem for Well Ordered Sets) *Let $(X, \leq_X)$ and $(Y, \leq_Y)$ be well ordered sets. Then only one of the following are true:*

1. *$X \simeq Y$*
2. *$X \simeq s(a)$ for some $a \in Y$*
3. *$Y \simeq s(b)$ for some $b \in X$.*

In the light of our present language, we can observe that the crucial factor in characterising natural numbers was that every $n \in \omega$ contains all its predecessors. This can be used to motivate the definition of an ordinal number.

Definition 9.8 (Ordinal Number) *Suppose α is a well ordered set such that for every element $\xi \in \alpha$, $s(\xi) = \xi$. Then α is an ordinal number.*

Trivially, we can see that all natural numbers are ordinal numbers. But consider the set ω itself. This is certainly a well ordered set, and given any $n \in \omega$, $s(n) = n$. So, ω itself is an ordinal number. We characterise this in the following definition.

Definition 9.9 (Finite/Infinite Ordinal Numbers) *If α is an ordinal number and $\alpha \in \omega$, then we say that α is finite. Otherwise, α is infinite.*

How do we know that there are in fact ordinal numbers other than ω? The question has an easy answer. We can consider taking $\omega^+ = \omega \cup \{\omega\}$. We can order ω^+ by keeping the usual order when $m, n \in \omega$ and for all $n \in \omega$, we write $n < \omega$. Certainly, this is a well ordering and ω^+ is an ordinal. We shall write $\omega + n$ for the number constructed by taking n successors.

Now for another question: We required the Infinity Axiom (Axiom 7) to guarantee that the set of all natural numbers are in fact a set. Can we be guaranteed that all successors of ω is indeed a set? The question has an easy answer, and for the first time, we use the full strength of the Replacement Axiom (Axiom 2). Consider the sentence $S(n, p) =$ "$n \in \omega \wedge p = \omega + n$". Then, the axiom guarantees us that the collection $\omega' = \{p : p = \omega + n\}$ is indeed a set. Then again, we can start with $\omega' \cup \omega$, and take successors, and continue this process ad infinitum. The following result should become more accessible in the light of this discussion.

Theorem 9.10 (Burali-Forti Paradox) *There exists no set that contains all ordinals numbers.*

We know that every natural number is comparable, and this motivated our definition of order. It is not unusual to attempt to do the same with ordinals - for we know that they are structurally somewhat similar to their natural counterparts. We quote the following result to motivate our definition. The proof is application of Theorem 9.7.

Theorem 9.11 (Properties of Ordinal Numbers) *Let α, β be ordinal numbers. Then, either $\alpha \in \beta$, $\alpha = \beta$ or $\beta \in \alpha$.*

This motivates our definition:

Definition 9.12 (Ordinal Order) *Let α, β be ordinal numbers. We say that the ordinal $\alpha < \beta$ if $\alpha \in \beta$.*

By The Comparability Theorem for Ordinals (Theorem 9.7), we know that we can associate with every well ordered set a ordinal number. Further, we can show that this is unique. However, we interrupt our development of ordinals as a consequence of the following problem: If two sets are equivalent, can we associate the same ordinal number to them? It is shocking to know that we cannot! Consider the set ω^+. Clearly, the function $f : \omega \to \omega^+$ defined by $f(0) = \omega, f(n^+) = n$ is a bijection. But by construction, we know that $\omega \not\simeq \omega^+$, which implies that they do not have the same ordinality.

In fact, this should not come as a surprise. The Comparability Theorem talks in the language of similarities, not equivalences. The problem is that similarities concern themselves too much with the underlying order structure, when, for our purposes, it really should be ignored.

9.3 Comparability and Countability

In order to associate two sets that are equivalent the same number, we need to somehow look at building our theory from equivalences. We begin with a new definition.

Definition 9.13 (Set Dominance) *Let X and Y be sets. We say that Y dominates X if there exists an injection $f : X \to Y$. We write $X \precsim Y$. If $X \not\sim Y$ we say that Y strictly dominates X and write $X \prec Y$.*

We have the following two useful results:

Theorem 9.14 *Set Dominance is an order relation.*

Theorem 9.15 (Comparability Theorem for Sets) *For any two sets X,Y, either $X \precsim Y$, or $X = Y$ or $Y \precsim X$.*

Now we shall classify the "smallest" of infinities.

Definition 9.16 (Countable/Uncountable) *Let X be a set. If $X \precsim \omega$, then X is countable. If $X \sim \omega$, we say it is countably infinite. Otherwise we say X is uncountable.*

We know that there exist sets that are countable. Surely ω itself is such a set. But for a non-trivial example, consider the set $E = \{2n \in \omega : n \in \omega\}$. Certainly $f(n) = 2n$ between E and ω. But the question remains: are there sets which are uncountable? This famous theorem provides an answer.

Theorem 9.17 (Cantor Theorem) *Let X be a set. Then $X \prec \wp(X)$.*

Proof Suppose there exists a bijection $f : X \to \wp(X)$. Then consider the set $S = \{x \in X : x \notin f(x)\}$. Now, trivially, $S \in \wp(X)$. So, there exist some $s \in X$ such that $f(s) = S$. Now, if $s \in S$, then $s \notin f(s) = S$. And if $s \notin S$, then $s \in f(s) = S$. Either way, we get a contradiction and the result follows. □

The importance of this result should not require any further validation. But we shall soon see how the theory of cardinal numbers certainly celebrates it.

9.4 Cardinals

Our discussion of ordinal numbers was not a wasteful effort to demonstrate that we can construct weird numbers that measure infinities. They are convenient because the essence which we seek - the notion of equivalence - is hidden somewhere in the wide ocean of ordinals. The whole ocean, it seems, is too numerous. So we consider the better behaved ones which suit our purposes.

Theorem 9.11 tells us that ordinals are always comparable. It follows from the fact: that every element of an ordinal is itself an initial segment, that any set of ordinals is well ordered. This is crucial to make the following definition valid. Also note that the collection we form is indeed a set - the collection of all well orderings of a set X is a subset of $\wp(\wp(X \times X))$ and it follows that it is indeed a set. With each such well ordering, we can associate a unique ordinal number, and by using the strength of Replacement (Axiom 2), we are guaranteed that collection we are about to describe is indeed a set.

Definition 9.18 (Cardinal Number) *Let X be a set, and let $\mathcal{S} = \{\alpha : \alpha \sim X\}$. Then we write* $\mathrm{card}(X) = \min(\mathcal{S})$*, the cardinality or cardinal number of X.*

We have seen and encountered one infinite cardinal number already. This is ω, and was the motivation behind our definition of Countability. In the presence of the language we've developed, we can say that a set X is countable infinite if $\mathrm{card}(X) = \omega$. It also happens that ω is the smallest infinite cardinal, since by the very characterisation of infinite, if $n < \omega$, then n is finite.

How can we give meaning to addition and multiplication of cardinals? The answer is elegant. We can define the sum of two cardinal numbers by considering the union of two disjoint sets of the corresponding cardinality. Then the cardinal sum is the cardinality of the resulting set. Then products follow from by considering Cartesian products. A deeper discussion of cardinal numbers can be found in [4, p28].

Cardinal numbers are either finite or infinite. We have briefly and informally discussed that ω is the first infinite cardinal number. Now, $\mathrm{card}(\wp(\omega)) \neq \omega$ by Cantor's Theorem (Theorem 9.17). We call the first infinite cardinal number $\aleph_0 = \omega$. Then we list the subsequent cardinals $\aleph_1, \aleph_2, \ldots$ - the infinity of infinities.

The assertion $\aleph_1 = \mathrm{card}(\wp(\aleph_0))$ is known as the Continuum Hypothesis and was formulated by Cantor [2, p2] who searched unsuccessfully for a proof. The Generalised Continuum Hypothesis states that $\aleph_{n^+} = \mathrm{card}(\wp(\aleph_n))$.

Along with the independence of the Axiom of Choice, Kurt Gödel also proved that the Generalised Continuum Hypothesis was consistent with the Axioms of Set Theory. In 1963, Paul Cohen established the independence of the Continuum Hypothesis [6, p3]. Conclusion: no proof exists.

These theories lie at the heart of modern mathematics. They have been the work of many brilliant and wonderful minds. The consequences of these theories are still not complete, and they may never will be. Mathematics does indeed seem infinite. We bring our discussion to an end by quoting David Hilbert - "No one shall expel us from the paradise that Cantor has created for us."

Acknowledgements

Firstly, I would like to thank John Stillwell (Dept. Maths, Monash University) for his patience, encouragement and assistance. I am compelled to believe that his knowledge is uncountably infinite! I should also thank Lloyd Humberstone (Dept. Philosophy, Monash University) for his wisdom in logic and assisting me in understanding vacuous truth. Also, Tom Hall (Dept. Maths, Monash University) deserves thanks, for his encouragement and enthusiasm regarding this topic and Maria Athanassenas (Dept. Maths, Monash University) for her encouragement and simply putting up with me as an annoying student! Lastly, I'd like to thank my parents for their *Food for Maths* program - the "relief drops" curbed the tax on my time from having to cook.

References

[1] Herbert Federer. *Geometric Measure Theory*. Springer-Verlag Berlin, 1996.

[2] W.M. Zaring G. Takeuti. *Introduction to Axiomatic Set Theory*. Springer-Verlag New York Inc., 1971.

[3] Paul R. Halmos. *Naive Set Theory*. Springer-Verlag New York Inc., 1974.

[4] Felix Hausdorff. *Set Theory (Mengenlehre)*. Chelsea Publishing Company, 1957.

[5] Jean E. Rubin Herman Rubin. *Equivalents of the Axiom of Choice, 2*. Elsevier Science Publishers, 1985.

[6] Gregory H. Moore. *Zermelo's Axiom of Choice*. Springer-Verlag New York Inc., 1982.

[7] James R. Munkres. *Topology, 2nd Edition*. Pearson Education, 1996.

[8] Judith Roitman. *Introduction to Modern Set Theory*. John Wiley and Sons, Inc., 1990.

[9] Bertrand Russell. *The Principles of Mathematics*. Bradford and Dickens, 1956.

[10] Jean van Heijenoort. *From Frege to Godel*. Oxford University Press, 1967.

[11] Stan Wagon. *The Banach-Tarski Paradox*. Cambridge University Press, 1985.