

An Introduction to Axiomatic Set Theory

Joseph R. Mileti

February 6, 2007

1 Introduction

No one shall expel us from the paradise that Cantor has created. - David Hilbert

1.1 Why Set Theory?

Set theory originated in an attempt to understand and somehow classify “small” or “negligible” sets of real numbers. Cantor’s early explorations in the realm of the transfinite were motivated by a desire to understand the points of convergence of trigonometric series. The basic ideas quickly became a fundamental part of analysis.

Since then, set theory has become a way to unify mathematical practice and the way in which mathematicians deal with the infinite in all areas of mathematics. You’ve all seen the proof that the set of real numbers are uncountable, but what more can be said? Exactly how uncountable is the set of real numbers? Does this taming of the infinite give us any new tools to prove interesting mathematical theorems? Is there anything more that the set-theoretic perspective provides to the mathematical toolkit other than a crude notion of size and cute diagonal arguments?

We begin by listing a few basic questions from various areas of mathematics that can only be tackled with a well-defined theory of the infinite which set theory provides.

Algebra: A fundamental result in linear algebra is that every finitely generated vector space has a basis, and any two bases have the same size. We call the unique size of any basis of a vector space the dimension of that space. What can be said about vector spaces that aren’t finitely generated? Does every vector space have a basis? Is there a meaningful way to assign a “dimension” to every vector space in such a way that two vector spaces over the same field are isomorphic if and only if they have the same “dimension”? We need a well-defined and robust notion of infinite sets and infinite cardinality to deal with these questions.

Analysis: Lebesgue’s theory of measure and integration require an important distinction between countable and uncountable sets. Aside from this use, the study of the basic structure of the Borel sets or the projective sets (an extension of the Borel sets) require some sophisticated use of set theory, in a way that can be made precise.

Foundations: A remarkable side-effect of our undertaking to systematically formalize the infinite is that we can devise a formal axiomatic and finitistic system in which virtually of mathematical practice can be embedded in an extremely faithful manner. Whether this fact is interesting or useful depends on your philosophical stance about the nature of mathematics, but it does have an important consequence. It puts us in a position to prove that certain statements do not follow from the axioms (which have now been formally defined and are thus susceptible to a mathematical analysis), and hence can not be proven by the currently accepted axioms. For better or worse, this feature has become the hallmark of set theory. For example, we can ask questions like:

1. Do we really need the Axiom of Choice to produce a nonmeasurable set of real numbers?

2. Is there an uncountable set of real numbers which can not be in one-to-one correspondence with the set of all real numbers?

Aside from these ideas which are applicable to other areas of mathematics, set theory is a very active area of mathematics with its own rich and beautiful structure, and deserves study for this reason alone.

1.2 Motivating the Axioms

In every modern mathematical theory (say group theory, topology, the theory of Banach spaces), we start with a list of axioms, and derive results from these. In most of the fields that we axiomatize in this way, we have several models of the axioms in mind (many different groups, many different topological spaces, etc.), and we're using the axiomatization to prove abstract results which will be applicable to each of these models. In set theory, you may think that it is our goal to study one unique universe of sets, so our original motivation in writing down axioms is simply to state precisely what we are assuming in an area that can often be very counterintuitive. Since we will build our system in first-order logic, it turns out that there are many models of set theory as well (assuming that there is at least one...), and this is the basis for proving independence results, but this isn't our initial motivation. This section will be a little informal. We'll give the formal axioms (in a formal first-order language) and derive consequences starting in the next section.

Whether the axioms that we are writing down now are "obviously true", "correct", "justified", or even worthy of study are very interesting philosophical questions, but I will not spend much time on them here. Regardless of their epistemological status, they are now nearly universally accepted as the "right" axioms to use in the development of set theory. The objects of our theory are sets, and we have one binary relation \in which represents set membership. That is, we write $x \in y$ to mean that x is an element of y . We begin with an axiom which ensures that our theory is not vacuous.

Axiom of Existence: There exists a set.

We need to have an axiom which says how equality of sets is determined in terms of the membership relation. In mathematical practice using naive set theory, the most common way to show that two sets A and B are equal is to show that each is a subset of the other. We therefore define $A \subseteq B$ to mean that for all $x \in A$, we have $x \in B$, and we want to be able to conclude that $A = B$ from the facts that $A \subseteq B$ and $B \subseteq A$. That is, we want to think of a set as being completely determined by its members, thus linking $=$ and \in , but we need to codify this as an axiom.

Axiom of Extensionality: For any two sets A and B , if $A \subseteq B$ and $B \subseteq A$, then $A = B$.

The Axiom of Extensionality implicitly implies a few perhaps unexpected consequences about the nature of sets. First, if a is a set, then we should consider the two sets $\{a\}$ and $\{a, a\}$ (if we are allowed to assert their existence) to be equal because they have the same elements. Similarly, if a and b are sets, then we should consider $\{a, b\}$ and $\{b, a\}$ to be equal. Hence, whatever a set is, it should be inherently unordered and have no notion of multiplicity. Also, since the only objects we are considering are sets, we are ruling out the existence of "atoms" other than the empty set, i.e. objects a which are not the empty set but which have no elements.

We next need some rules about how we are allowed to build sets. The naive idea is that any property we write down determines a set. That is, for any property P of sets, we may form the set $\{x : P(x)\}$. For example, if you have a group G , you may form the center of G given by $Z(G) = \{x : x \in G \text{ and } xy = yx \text{ for all } y \in G\}$. Of course, this naive approach leads to the famous contradiction known as Russell's paradox. Let $P(x)$ be the property $x \notin x$, and let $z = \{x : P(x)\} = \{x : x \notin x\}$. We then have $z \in z$ if and only if $z \notin z$, a contradiction.

This gives our first indication that it may be in our best interest to tread carefully when giving rules about how to build sets. One now standard reaction to Russell’s Paradox and other similar paradoxes in naive set theory is that the set-theoretic universe is too “large” to encapsulate into one set. Thus, we shouldn’t allow ourselves the luxury of forming the set $\{x : P(x)\}$ because by doing so we may package too much into one set, and the set-theoretic universe is too “large” to make this permissible. In other words, we should only christen something as a set if it is not too “large”.

However, if we already have a set A and a property P , we should be allowed to form $\{x \in A : P(x)\}$ because A is a set (hence not too “large”), so we should be allowed to assert that the subcollection consisting of those sets x in A such that $P(x)$ holds is in fact a set. For example, if we have a group G (so G is already known to be a set), its center $Z(G)$ is a set because $Z(G) = \{x \in G : xy = yx \text{ for all } y \in G\}$. Therefore, we put forth the following axiom.

Axiom of Separation: For any set A and any property P of sets, we may form the set consisting of precisely those $x \in A$ such that $P(x)$, i.e. we may form the set $\{x \in A : P(x)\}$.

You may object to this axiom because of the vague notion of a “property” of sets, and that would certainly be a good point. We’ll make it precise when we give the formal first-order axioms in the next section. The Axiom of Separation allows us to form sets from describable subcollections of sets we already know exist, but we currently have no way to build larger sets from smaller ones. We now give axioms which allow us to build up sets in a permissible manner.

Our first axiom along these lines will allow us to conclude that for any two sets x and y , we may put them together into a set $\{x, y\}$. Since we already have the Axiom of Separation, we will state the axiom in the (apparently) weaker form that for any two sets x and y , there is a set with both x and y as elements.

Axiom of Pairing: For any two sets x and y , there is a set A such that $x \in A$ and $y \in A$.

We next want to have an axiom which allows us to take unions. However, in mathematics, we often want to take a union over a family of sets, possibly infinite. For example, we may have a set A_n for each natural number n , and then want to consider $\bigcup_{n \in \mathbb{N}} A_n$. By being clever, we can incorporate all of these ideas of taking unions into one axiom. The idea is the following. Suppose that we have two sets A and B , say $A = \{u, v, w\}$ and $B = \{x, z\}$. We want to be able to assert the existence of the union of A and B , which is $\{u, v, w, x, z\}$. First, by the Axiom of Pairing, we may form the set $\mathcal{F} = \{A, B\}$, which equals $\{\{u, v, w\}, \{x, z\}\}$. Now the union of A and B is the set of elements of elements of \mathcal{F} . In the above example, if we can form the set $\mathcal{F} = \{A_1, A_2, A_3, \dots\}$ (later axioms will justify this), then $\bigcup_{n \in \mathbb{N}} A_n$ is the set of elements of elements of \mathcal{F} . Again, in the presence of the Axiom of Separation, we state this axiom in the (apparently) weaker form that for any set \mathcal{F} , there is set containing all elements of elements of \mathcal{F} .

Axiom of Union: For any set \mathcal{F} , there is a set U such that for all sets x , if there exists $A \in \mathcal{F}$ with $x \in A$, then $x \in U$.

We next put forward two axioms which really allow the set-theoretic universe to expand. The first is the Power Set Axiom which tells us that if we have a set A , it is permissible to form the set consisting of all subsets of A .

Axiom of Power Set: For any set A , there is a set \mathcal{F} such that for all sets B , if $B \subseteq A$, then $B \in \mathcal{F}$.

Starting with the empty set \emptyset (which exists using the Axiom of Existence and the Axiom of Separation), we can build a very rich collection of finite sets using the above axioms. For example, we can form $\{\emptyset\}$ using the Axiom of Pairing. We can also form $\{\emptyset\}$ by applying the Axiom of Power Set to \emptyset . We can then go on

to form $\{\emptyset, \{\emptyset\}\}$ and many other finite sets. However, our axioms provide no means to build an infinite set.

Before getting to the Axiom of Infinity, we will lay some groundwork about ordinals. If set theory is going to serve as a basis for mathematics, we certainly need to be able to embed within it the natural numbers. It seems natural represent the number n as some set which we think of as having n elements. Which set should we choose? Let's start from the bottom-up. The natural choice to play the role of 0 is \emptyset because it is the only set without any elements. Now that we have 0, and we want 1 to be a set with one element, perhaps we should let 1 be the set $\{0\} = \{\emptyset\}$. Next, a canonical choice for a set with two elements is $\{0, 1\}$, so we let $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$. In general, if we have defined $0, 1, 2, \dots, n$, we can let $n + 1 = \{0, 1, \dots, n\}$. This way of defining the natural numbers has many advantages which we'll come to appreciate. For instance, we'll have $n < m$ if and only if $n \in m$, so we may use the membership relation to define the standard ordering of the natural numbers.

However, the \dots in the above definition of $n + 1$ may make you a little nervous. Fortunately, we can give another description of $n + 1$ which avoids this unpleasantness. If we've defined n , we let $n + 1 = n \cup \{n\}$, which we can justify the existence of using the Axiom of Pairing and the Axiom of Union. The elements of $n + 1$ will then be n , and the elements of n which should "inductively" be the natural numbers up to, but not including, n .

Using the above outline, we can use our axioms to justify the existence of any particular natural number n (or, more precisely, the set that we've chosen to represent our idea of the natural number n). However, we can't justify the existence of the set of natural numbers $\{0, 1, 2, 3, \dots\}$. To enable us to do this, we make the following definition. For any set x , let $S(x) = x \cup \{x\}$. We call $S(x)$ the *successor* of x . We want an axiom which says that there is a set containing $0 = \emptyset$ which is closed under successors.

Axiom of Infinity: There exists a set A such that $\emptyset \in A$ and for all x , if $x \in A$, then $S(x) \in A$.

With the Axiom of Infinity asserting existence, it's not too difficult to use the above axioms to show that there is a smallest (with respect to \subseteq) set A such that $\emptyset \in A$ and for all x , if $x \in A$, then $S(x) \in A$. Intuitively, this set is the collection of all natural numbers. Following standard set-theoretic practice, we denote this set by ω (this strange choice, as opposed to the typical \mathbb{N} , conforms with the standard practice of using lowercase greek letters to represent infinite ordinals).

With the set of natural numbers ω in hand, there's no reason to be timid and stop counting. We started with $0, 1, 2, \dots$, where each new number consisted of collecting the previous numbers into a set, and we've now collected all natural numbers into a set ω . Why not continue the counting process by considering $S(\omega) = \omega \cup \{\omega\} = \{0, 1, 2, \dots, \omega\}$? We call this set $\omega + 1$ for obvious reasons. This conceptual leap of counting into the so-called transfinite gives rise to the ordinals, the "numbers" which form the backbone of set theory.

Once we have $\omega + 1$, we can then form the set $\omega + 2 = S(\omega + 1) = \{0, 1, 2, \dots, \omega, \omega + 1\}$, and continue on to $\omega + 3, \omega + 4$, and so on. Why stop there? If we were able to collect all of the natural numbers into a set, what's preventing us from collecting these into the set $\{0, 1, 2, \dots, \omega, \omega + 1, \omega + 2, \dots\}$, and continuing? Well, our current axioms are preventing us, but we shouldn't let that stand in our way. If we can form ω , surely we should have an axiom allowing us to make this new collection a set. After all, if ω isn't too "large", this set shouldn't be too "large" either since it's just another sequence of ω many sets after ω .

The same difficulty arises when you want to take the union of an infinite family of sets. In fact, the previous problem is a special case of this one, but in this generality it may feel closer to home. Suppose we have sets A_0, A_1, A_2, \dots , that is, we have a set A_n for every $n \in \omega$. Of course, we should be able to justify making the union $\bigcup_{n \in \omega} A_n$ into a set. If we want to apply the Axiom of Union, we should first form the set $\mathcal{F} = \{A_0, A_1, A_2, \dots\}$ and apply the axiom to \mathcal{F} . However, in general, our current axioms don't justify forming this set despite its similarity to asserting the existence of ω .

To remedy these defects, we need a new axiom. In light of the above examples, we want to say something along the lines of "if we can index a family of sets with ω , then we can form this family into a set". Using this

principle, we should be able to form the set $\{\omega, \omega + 1, \omega + 2, \dots\}$ and hence $\{0, 1, 2, \dots, \omega, \omega + 1, \omega + 2, \dots\}$ is a set by the Axiom of Union. Similarly, in the second example, we should be able to form the set $\{A_0, A_1, A_2, \dots\}$. In terms of our restriction of not allowing sets to be too “large”, this seems justified because if we consider ω to not be too “large”, then any family of sets it indexes shouldn’t be too “large” either.

There is no reason to limit our focus to ω . If we have any set A , and we can index a family of sets using A , then we should be able to assert the existence of a set containing the elements of the family. We also want to make the notion of indexing more precise, and we will do it using the currently vague notion of a property of sets as used in the Axiom of Separation.

Axiom of Collection: Suppose that A is a set and $P(x, y)$ is a property of sets such that for every $x \in A$, there is a unique set y such that $P(x, y)$ holds. Then there is a set B such that for every $x \in A$, we have $y \in B$ for the unique y such that $P(x, y)$ holds.

Our next axiom is often viewed as the most controversial due to its nonconstructive nature and the sometimes counterintuitive results it allows us to prove. I will list it here as a fundamental axiom, but we will avoid using it in the basic development of set theory below until we get to a position to see its usefulness in mathematical practice.

The Axiom of Separation and the Axiom of Collection involved the somewhat vague notion of property, but whenever we think of a property (and the way we will make the notion of property precise using a formal language) we have a precise unambiguous definition which describes the property in mind. Our next axiom, the Axiom of Choice, asserts the existence of certain sets without the need for such a nice description. Intuitively, it says that if we have a set consisting only of nonempty sets, there is a function which picks an element out each of these nonempty sets without requiring that there be a “definable” description of such a function. We haven’t defined the notion of a function in set theory, and it takes a little work to do, so we will state the axiom in the following form: For every set \mathcal{F} of nonempty pairwise disjoint sets, there is a set C consisting of exactly one element from each element of \mathcal{F} . We think of C as a set which “chooses” an element from each of the elements of \mathcal{F} . Slightly more precisely, we state the axiom as follows.

Axiom of Choice: Suppose that \mathcal{F} is a set such every $A \in \mathcal{F}$ is nonempty, and for every $A, B \in \mathcal{F}$, if there exists a set x with $x \in A$ and $x \in B$, then $A = B$. There exists a set C such that for every $A \in \mathcal{F}$, there is a unique $x \in C$ with $x \in A$.

Our final axiom is in no way justified by mathematical practice because it never appears in arguments outside set theory. It is also somewhat unique among our axioms in that it asserts that certain types of sets do not exist. However, adopting it gives a much clearer picture of the set-theoretic universe and it will come to play an important role in the study of set theory itself. As with the Axiom of Choice, we will avoid using it in the basic development of set theory below until we are able to see its usefulness to us.

The goal is to eliminate sets which appear circular in terms of the membership relation. For example, we want to forbid sets x such that $x \in x$ (so there is no set x such that $x = \{x\}$). Similarly, we want to forbid the existence of sets x and y such that $x \in y$ and $y \in x$. In more general terms, we don’t want to have a set with an infinite descending chain each a member of the next, such as having sets x_n for each $n \in \omega$ such that $\dots \in x_2 \in x_1 \in x_0$. We codify this by saying every nonempty set A has an element which is minimal with respect to the membership relation.

Axiom of Foundation: If A is a nonempty set, then there exists $x \in A$ such that there is no set z with both $z \in A$ and $z \in x$.

1.3 Formal Axiomatic Set Theory

We now give the formal version of our axioms. We work in a first-order language \mathcal{L} with a single binary relation symbol \in . By working in this first-order language, we are able to make precise the vague notion of property discussed above by using first-order formulas instead. However, this comes at the cost of replacing the Axiom of Separation and the Axiom of Collection by infinitely many axioms (also called an axiom scheme) since we can't quantify over formulas within the theory itself. There are other more subtle consequences of formalizing the above intuitive axioms in first-order logic which we will discuss below.

Notice also that we allow parameters (denoted by \vec{p}) in the Axioms of Separation and Collection so that we will be able to derive statements which universally quantify over a parameter, such as "For all groups G , the set $Z(G) = \{x \in G : xy = yx \text{ for all } x \in G\}$ exists", rather than having to reprove that $Z(G)$ is a set for each group G that we know exists. Finally, notice how we can avoid using defined notions (like \emptyset , \subseteq , and $S(x)$ in the Axiom of Infinity) by expanding them out into our fixed language. For example, we replace $x \subseteq y$ by $\forall w(w \in x \rightarrow w \in y)$ and replace $\emptyset \in z$ by $\exists w(\forall y(y \notin w) \wedge w \in z)$ (we could also replace it by $\forall w(\forall y(y \notin w) \rightarrow w \in z)$).

In each of the following axioms, when we write a formula $\varphi(x_1, x_2, \dots, x_k)$, we implicitly mean that the x_i 's are distinct variables and that every free variable of φ is one of the x_i . We also use \vec{p} to denote a finite sequence of variables p_1, p_2, \dots, p_k . Notice that we don't need the Axiom of Existence because it is true in all \mathcal{L} -structures (recall that all \mathcal{L} -structures are nonempty).

Axiom of Extensionality:

$$\forall x \forall y (\forall w (w \in x \leftrightarrow w \in y) \rightarrow x = y)$$

Axiom (Scheme) of Separation: For each formula $\varphi(x, y, \vec{p})$ we have the axiom

$$\forall \vec{p} \forall y \exists z \forall x (x \in z \leftrightarrow x \in y \wedge \varphi(x, y, \vec{p}))$$

Axiom of Pairing:

$$\forall x \forall y \exists z (x \in z \wedge y \in z)$$

Axiom of Union:

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

Axiom of Power Set:

$$\forall x \exists z \forall y (\forall w (w \in y \rightarrow w \in x) \rightarrow y \in z)$$

Axiom of Infinity:

$$\exists z (\exists w (\forall y (y \notin w) \wedge w \in z) \wedge \forall x (x \in z \rightarrow \exists y (\forall w (w \in y \leftrightarrow (w \in x \vee w = x)) \wedge y \in z)))$$

Axiom (Scheme) of Collection: For each formula $\varphi(x, y, \vec{p})$ we have the axiom

$$\begin{aligned} \forall \vec{p} \forall w ((\forall x (x \in w \rightarrow \exists y \varphi(x, y, \vec{p})) \wedge \forall x (x \in w \rightarrow \forall u \forall v ((\varphi(x, u, \vec{p}) \wedge \varphi(x, v, \vec{p})) \rightarrow u = v))) \\ \rightarrow \exists z \forall x (x \in w \rightarrow \exists y (y \in z \wedge \varphi(x, y, \vec{p})))) \end{aligned}$$

Axiom of Choice:

$$\begin{aligned} \forall z ((\forall x (x \in z \rightarrow \exists w (w \in x)) \wedge \forall x \forall y ((x \in z \wedge y \in z \wedge \exists w (w \in x \wedge w \in y)) \rightarrow x = y)) \\ \rightarrow \exists c \forall x (x \in z \rightarrow (\exists w (w \in x \wedge w \in c) \wedge \forall u \forall v ((u \in x \wedge v \in x \wedge u \in c \wedge v \in c) \rightarrow u = v)))) \end{aligned}$$

Axiom of Foundation:

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

Let Ax_{ZFC} be the above set of sentences, and let $ZFC = Cn(Ax_{ZFC})$ (ZFC stands for Zermelo-Fraenkel set theory with Choice). Other presentations state the axioms of ZFC a little differently, but they all give the same theory. Some people refer to the Axiom of Separation as the Axiom of Comprehension, but Comprehension is sometimes also used to mean the contradictory statement (via Russell's Paradox) that we can always form the set $\{x : P(x)\}$, so I prefer to call it Separation. Also, some presentations refer to the Axiom of Collection as the Axiom of Replacement, but this name is more applicable to the statement that replaces the last \rightarrow in the statement of Collection with a \leftrightarrow , and this formulation implies the Axiom of Separation.

1.4 Working from the Axioms

We have set up ZFC as a first-order theory similar to the group axioms, ring axioms, or axioms of partial orderings. Since we have two notions of implication (semantic and syntactic), in order to show that $\sigma \in ZFC$, we can show that either $Ax_{ZFC} \models \sigma$ or $Ax_{ZFC} \vdash \sigma$. Given your experience with syntactic deductions, I'm guessing that you will jump on the first one.

When attempting to show that $Ax_{ZFC} \models \sigma$ we must take an arbitrary model of Ax_{ZFC} and show that it is a model of σ . Thus, we must be mindful of strange \mathcal{L} -structures and perhaps unexpected models. For example, let \mathcal{L} be the language of set theory (so we have one binary relation symbol \in) and let \mathfrak{N} be the \mathcal{L} -structure $(\mathbb{N}, <)$. Let's see which elements of Ax_{ZFC} hold in \mathfrak{N} .

- *Axiom of Extensionality:* In the structure \mathfrak{N} , this interprets as saying that whenever two elements of \mathbb{N} have the same elements of \mathbb{N} less than them, then they are equal. This holds in \mathfrak{N} .
- *Axiom (Scheme) of Separation:* This does not hold in \mathfrak{N} . Let $\varphi(x, y)$ be the formula $\exists w(w \in x)$. The corresponding instance of Separation is:

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

In the structure \mathfrak{N} , this interprets as saying that for all $n \in \mathbb{N}$, there is an $m \in \mathbb{N}$ such that for all $k \in \mathbb{N}$, we have $k < m$ if and only if $k < n$ and $k \neq 0$. This does not hold in \mathfrak{N} because if we consider $n = 2$, there is no $m \in \mathbb{N}$ such that $0 \not< m$ and yet $1 < m$.

- *Axiom of Pairing:* In the structure \mathfrak{N} , this interprets as saying that whenever $m, n \in \mathbb{N}$, there exists $k \in \mathbb{N}$ such that $m < k$ and $n < k$. This holds in \mathfrak{N} because given $m, n \in \mathbb{N}$, we may take $k = \max\{m, n\} + 1$.
- *Axiom of Union:* In the structure \mathfrak{N} , this interprets as saying that whenever $n \in \mathbb{N}$, there exists $\ell \in \mathbb{N}$ such that whenever $k \in \mathbb{N}$ has the property that there exists $m \in \mathbb{N}$ with $k < m$ and $m < n$, then $k < \ell$. This holds in \mathfrak{N} because given $n \in \mathbb{N}$, we may take $\ell = n$ since if $k < m$ and $m < n$, then $k < n$ by transitivity of $<$ in \mathbb{N} (in fact, we may take $\ell = n - 1$ if $n \neq 0$).
- *Axiom of Power Set:* In the structure \mathfrak{N} , this interprets as saying that whenever $n \in \mathbb{N}$, there exists $\ell \in \mathbb{N}$ such that whenever $m \in \mathbb{N}$ has the property that every $k < m$ also satisfies $k < n$, then $m < \ell$. This holds in \mathfrak{N} because given $n \in \mathbb{N}$, we may take $\ell = n + 1$ since if $m \in \mathbb{N}$ has the property that every $k < m$ also satisfies $k < n$, then $m \leq n$ and hence $m < n + 1$.
- *Axiom of Infinity:* In the structure \mathfrak{N} , this interprets as saying that there exists $n \in \mathbb{N}$ such that $0 < n$ and whenever $m < n$, we have $m + 1 < n$. This does not hold in \mathfrak{N} .
- *Axiom (Scheme) of Collection:* This holds in \mathfrak{N} , as we now check. Fix a formula $\varphi(x, y, \vec{p})$. Interpreting in \mathfrak{N} , we need to check that if we fix natural numbers \vec{q} and an $n \in \mathbb{N}$ such that for all $k < n$ there exists a unique $\ell \in \mathbb{N}$ such that $(\mathfrak{N}, k, \ell, \vec{q}) \models \varphi$, then there exists $m \in \mathbb{N}$ such that for all $k < n$ there

exists an $\ell < m$ such that $(\mathfrak{N}, k, \ell, \vec{q}) \models \varphi$. Let's then fix natural numbers \vec{q} and an $n \in \mathbb{N}$, and suppose that for all $k < n$ there exists a unique $\ell \in \mathbb{N}$ such that $(\mathfrak{N}, k, \ell, \vec{q}) \models \varphi$. For each $k < n$, let ℓ_k be the unique element of \mathbb{N} such that $(\mathfrak{N}, k, \ell_k, \vec{q}) \models \varphi$. Letting $m = \max\{\ell_k : k < n\} + 1$, we see that m suffices. Therefore, this holds in \mathfrak{N} .

- *Axiom of Choice*: In the structure \mathfrak{N} , this interprets as saying that whenever $n \in \mathbb{N}$ is such that
 - Every $m < n$ is nonzero.
 - For all $\ell, m < n$, there is no k with $k < \ell$ and $k < m$

then there exists $m \in \mathbb{N}$ such that for all $k < n$, there is exactly one $\ell \in \mathbb{N}$ with $\ell < m$ and $\ell < n$. Notice that the only $n \in \mathbb{N}$ satisfying the hypothesis (that is, the above two conditions) is $n = 0$. Now for $n = 0$, the condition is trivial because we may take $m = 0$ as there is no $k < 0$. Therefore, this holds in \mathfrak{N} .

- *Axiom of Foundation*: In the structure \mathfrak{N} , this interprets as saying that whenever $n \in \mathbb{N}$ has the property that there is some $m < n$, there there exists $m < n$ such that there is no k with $k < m$ and $k < n$. Notice that $n \in \mathbb{N}$ has the property that there is some $m < n$ if and only if $n \neq 0$. Thus, this holds in \mathfrak{N} because if $n \neq 0$, then we have that $0 < n$ and there is no $k < 0$ and $k < n$.

Is Ax_{ZFC} satisfiable? Can we somehow construct a model of Ax_{ZFC} ? These are interesting questions with subtle answers. For now, you'll have to live with a set of axioms with no obvious models.

Thus, when we develop set theory below, we will be arguing semantically via models. Rather than constantly saying “Fix a model \mathcal{M} of Ax_{ZFC} ” at the beginning of each proof, and proceeding by showing that $(\mathcal{M}, s) \models \varphi$ for various φ , we will keep the models in the background and assume that we are “living” inside one for each proof. When we are doing this, a “set” is simply an element of the universe M of our model \mathcal{M} , and given two “sets” a and b , we write $a \in b$ to mean that (a, b) is an element of $\in^{\mathcal{M}}$.

Also, although there is no hierarchy of sets in our axioms, we will often follow the practice of using lowercase letters a, b, c , etc. to represent sets that we like to think of as having no internal structure (such as numbers, elements of a group, points of a topological space), use capital letters A, B, C , etc. to represent sets whose elements we like to think of as having no internal structure, and use script letters \mathcal{A}, \mathcal{F} , etc. to represent sets of such sets.

1.5 ZFC as a Foundation for Mathematics

In the next 2 chapters we'll show how to develop mathematics quite faithfully within the framework of ZFC. This raises the possibility of using set theory as a foundation for mathematical practice. However, this seems circular because our development of logic presupposed normal mathematical practice and “naive” set theory (after all, we have the *set* of axioms of ZFC). It seems that logic depends on set theory and set theory depends on logic, so how have we gained anything from a foundational perspective?

It is indeed possible, at least in principle, to get out of this vicious circle and have a completely finitistic basis for mathematics. The escape is to buckle down and use syntactic arguments. Now there are infinitely many axioms of ZFC (because of the two axioms schemes), but instead of showing that $Ax_{ZFC} \vdash \tau$, we can instead show that $\Sigma \vdash \tau$ for a finite $\Sigma \subseteq Ax_{ZFC}$ (in which every line of the deduction has a finite collection of formulas on the left-hand side). In this way, it would be possible in principle to make every proof completely formal and finitistic where each line follows from previous lines by one of our proof rules. If we held ourselves to this style, then we could reduce mathematical practice to a game with finitely many symbols (if you insisted we could replace our infinite stock of variables Var with one variable symbol x and a new symbol $'$ and refer to x_3 as x''' , etc.) where each line could be mechanically checked according to our finitely many rules. Thus, it would even be possible to program a computer to check every proof.

In practice (for human beings at least), the idea of giving deductions for everything is outlandish. Leaving aside the fact that actually giving short deductions is often a painful endeavor in itself, it turns out that even the most basic statements of mathematics, when translated into ZFC, are many thousands of symbols long, and elementary mathematical proofs (such as say the Fundamental Theorem of Arithmetic) are many thousands of lines long. We'll discuss how to develop the real numbers below, but any actual formulas talking about real numbers would be ridiculously long and incomprehensible to the human reader. Due to these reasons, and since the prospect of giving syntactic deductions for everything gives me nightmares, I choose to argue everything semantically in the style of any other axiomatic subject in mathematics. It is an interesting and worthwhile exercise, however, to imagine how everything could be done syntactically.

2 Developing Basic Set Theory

2.1 First Steps

We first establish some basic set theoretic facts carefully from the axioms.

Definition 2.1. *If A and B are sets, we write $A \subseteq B$ to mean for all $c \in A$, we have $c \in B$.*

Although the symbol \subseteq is not part of our language, we will often use \subseteq in our formulas and arguments. This use is justified because it can always be transcribed into our language by replacing it with the corresponding formula as we did in the axioms.

Proposition 2.2. *There is a unique set with no elements.*

Proof. Fix a set b . By Separation applied to the formula $x \neq x$, there is a set c such that for all a , we have $a \in c$ if and only if $a \in b$ and $a \neq a$. For all a , we have $a = a$, hence $a \notin c$. Therefore, there is a set with no elements. If c_1 and c_2 are two sets with no elements, then by the Axiom of Extensionality, we may conclude that $c_1 = c_2$. \square

Definition 2.3. *We use \emptyset to denote the unique set with no elements.*

As above, we will often use \emptyset in our formulas and arguments despite the fact that there is no constant in our language representing it. Again, this use can always be eliminated by replacing it with a formula as we did in the axioms. We will continue to follow this practice without comment in the future when we introduce new definitions to stand for sets for which ZFC proves existence and uniqueness. In each case, be sure to understand how these definitions could be eliminated.

We now show how to turn the idea of Russell's Paradox into a proof that there is no universal set.

Proposition 2.4. *There is no set u such that $a \in u$ for every set a .*

Proof. Suppose that u is a set and $a \in u$ for every set a . By Separation applied to the formula $x \notin x$, there is a set c such that for all sets a , we have $a \in c$ if and only if $a \in u$ and $a \notin a$. Since $a \in u$ for every set a , we have $a \in c$ if and only if $a \notin a$ for every set a . Therefore, $c \in c$ if and only if $c \notin c$, a contradiction. \square

Proposition 2.5. *For all sets a and b , there is a unique set c such that, for all sets d , we have $d \in c$ if and only if either $d = a$ or $d = b$.*

Proof. Let a and b be sets. By Pairing, there is a set e such that $a \in e$ and $b \in e$. By Separation applied to the formula $x = a \vee x = b$ (notice that we are using parameters a and b in this use of Separation), there is a set c such that for all d , we have $d \in c$ if and only if both $d \in e$ and either $d = a$ or $d = b$. It follows that $a \in c$, $b \in c$, and for any $d \in c$, we have either $d = a$ or $d = b$. Uniqueness again follows from Extensionality. \square

Corollary 2.6. *For every set a , there is a unique set c such that, for all sets d , we have $d \in c$ if and only if $d = a$.*

Proof. Apply the previous proposition with $b = a$. □

Definition 2.7. Given two sets a and b , we use the notation $\{a, b\}$ to denote the unique set guaranteed to exist by the Proposition 2.5. Given a set a , we use the notation $\{a\}$ to denote the unique set guaranteed to exist by the Corollary 2.6.

Using the same style of argument, we can use Union and Separation to show that for every set \mathcal{F} , there is a unique set z consisting precisely of elements of elements of \mathcal{F} . The proof is an exercise.

Proposition 2.8. Let \mathcal{F} be a set. There is a unique set U such that for all a , we have $a \in U$ if and only if there exists $B \in \mathcal{F}$ with $a \in B$.

Definition 2.9. Let \mathcal{F} be a set. We use the notation $\bigcup \mathcal{F}$ to denote the unique set guaranteed to exist by the previous proposition. If A and B are sets, we use the notation $A \cup B$ to denote $\bigcup \{A, B\}$.

We now introduce some notation which conforms with the normal mathematical practice of writing sets.

Definition 2.10. Suppose that $\varphi(x, y, \vec{p})$ is a formula. Suppose that B and \vec{q} are sets. By Separation and Extensionality, there is a unique set C such that for all sets a , we have $a \in C$ if and only if $a \in B$ and $\varphi(a, B, \vec{q})$. We denote this unique set by $\{a \in B : \varphi(a, B, \vec{q})\}$.

With unions in hand, what about intersections? As in unions, the general case to consider is when we have a family of sets \mathcal{F} . We then want to collect those a such that $a \in B$ for all $B \in \mathcal{F}$ into a set. We do need to be a little careful however. What happens if $\mathcal{F} = \emptyset$? It seems that our definition would want to make the the intersection of the sets in \mathcal{F} consists of all sets, contrary to Proposition 2.4. However, this is the only case which gives difficulty because if $\mathcal{F} \neq \emptyset$, we can take the intersection to be a subset of one (any) of the elements of \mathcal{F} .

Proposition 2.11. Let \mathcal{F} be a set with $\mathcal{F} \neq \emptyset$. There is a unique set I such that for all a , we have $a \in I$ if and only if $a \in B$ for all $B \in \mathcal{F}$.

Proof. Since $\mathcal{F} \neq \emptyset$, we may fix $C \in \mathcal{F}$. Let $I = \{a \in C : \forall B \in \mathcal{F} (a \in B)\}$. For all a , we have $a \in I$ if and only if $a \in B$ for all $B \in \mathcal{F}$. Uniqueness again follows from Extensionality. □

Definition 2.12. Let \mathcal{F} be a set with $\mathcal{F} \neq \emptyset$. We use the notation $\bigcap \mathcal{F}$ to denote the unique set guaranteed to exist by the previous proposition. If A and B are sets, we use the notation $A \cap B$ to denote $\bigcap \{A, B\}$.

If A is a set, then we can not expect the complement of A to be a set because the union of such a purported set with A would be a set which has every set as an element, contrary to Proposition 2.4. However, if A and B are sets, and $A \subseteq B$, we can take the relative complement of A in B .

Proposition 2.13. Let A and B be sets with $A \subseteq B$. There is a unique set C such that for all a , we have $a \in C$ if and only if $a \in B$ and $a \notin A$.

Definition 2.14. Let A and B be sets with $A \subseteq B$. We use the notation $B \setminus A$ or $B - A$ to denote the unique set guaranteed to exist by the previous proposition.

2.2 Ordered Pairs and Cartesian Products

Since sets have no internal order to them, we need a way to represent ordered pairs. Fortunately (since it means we don't have to extend our notion of set), there is a hack which allows us to build sets which capture the notion of an ordered pair.

Definition 2.15. Given two sets a and b , we let $(a, b) = \{\{a\}, \{a, b\}\}$.

Proposition 2.16. *Let a, b, c, d be sets. If $(a, b) = (c, d)$, then $a = c$ and $b = d$.*

Proof. Suppose that a, b, c, d are sets and $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$. We first show that $a = c$. Since $\{c\} \in \{\{a\}, \{a, b\}\}$, either $\{c\} = \{a\}$ or $\{c\} = \{a, b\}$. In either case, we have $a \in \{c\}$, hence $a = c$. We now need only show that $b = d$. Suppose instead that $b \neq d$. Since $\{a, b\} \in \{\{c\}, \{c, d\}\}$, we have either $\{a, b\} = \{c\}$ or $\{a, b\} = \{c, d\}$. In either case, we conclude that $b = c$ (because either $b \in \{c\}$ or $b \in \{c, d\}$, and $b \neq d$). Similarly, since $\{c, d\} \in \{\{a\}, \{a, b\}\}$, we have either $\{c, d\} = \{a\}$ or $\{c, d\} = \{a, b\}$. In either case, we conclude that $d = a$. Therefore, using the fact that $a = c$, it follows that $b = d$. \square

We next turn to Cartesian products. Given two sets A and B , we would like to form the set $\{(a, b) : a \in A \text{ and } b \in B\}$. Justifying that we can collect these elements into a set takes a little work. The idea is as follows. For each fixed $a \in A$, we can assert the existence of $\{a\} \times B = \{(a, b) : b \in B\}$ using Collection (and Separation) because B is a set. Then using Collection (and Separation) again, we can assert the existence of $\{\{a\} \times B : a \in A\}$ since A is a set. The Cartesian product is then the union of this set. At later points, we will consider this argument sufficient, but we give a slightly more formal version here to really see how the axioms of Collection and Separation are applied and where the formulas come into play.

Proposition 2.17. *For any two sets A and B , there exists a unique set, denoted by $A \times B$, such that for all x , we have $x \in A \times B$ if and only if there exists $a \in A$ and $b \in B$ with $x = (a, b)$.*

Proof. Let $\varphi(\mathbf{b}, \mathbf{x}, \mathbf{a})$ be a formula expressing that “ $\mathbf{x} = (\mathbf{a}, \mathbf{b})$ ” (think about how to write this down). We have the statement

$$\forall \mathbf{a} \forall \mathbf{B} (\forall \mathbf{b} (\mathbf{b} \in \mathbf{B} \rightarrow \exists! \mathbf{x} \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a})))$$

where $\exists!$ is shorthand for “there is a unique”. Therefore, by Collection, we may conclude that

$$\forall \mathbf{a} \forall \mathbf{B} \exists \mathbf{C} \forall \mathbf{b} (\mathbf{b} \in \mathbf{B} \rightarrow \exists \mathbf{x} (\mathbf{x} \in \mathbf{C} \wedge \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a})))$$

Next using Separation and Extensionality, we have

$$\forall \mathbf{a} \forall \mathbf{B} \exists! \mathbf{C} \forall \mathbf{b} (\mathbf{b} \in \mathbf{B} \leftrightarrow \exists \mathbf{x} (\mathbf{x} \in \mathbf{C} \wedge \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a})))$$

From this it follows that

$$\forall \mathbf{A} \forall \mathbf{B} \forall \mathbf{a} (\mathbf{a} \in \mathbf{A} \rightarrow \exists! \mathbf{C} \forall \mathbf{b} (\mathbf{b} \in \mathbf{B} \leftrightarrow \exists \mathbf{x} (\mathbf{x} \in \mathbf{C} \wedge \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a}))))$$

Using Collection again, we may conclude that

$$\forall \mathbf{A} \forall \mathbf{B} \exists \mathcal{F} \forall \mathbf{a} (\mathbf{a} \in \mathbf{A} \rightarrow \exists \mathbf{C} (\mathbf{C} \in \mathcal{F} \wedge \forall \mathbf{b} (\mathbf{b} \in \mathbf{B} \leftrightarrow \exists \mathbf{x} (\mathbf{x} \in \mathbf{C} \wedge \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a}))))))$$

This implies

$$\forall \mathbf{A} \forall \mathbf{B} \exists \mathcal{F} \forall \mathbf{a} \forall \mathbf{b} ((\mathbf{a} \in \mathbf{A} \wedge \mathbf{b} \in \mathbf{B}) \rightarrow \exists \mathbf{C} (\mathbf{C} \in \mathcal{F} \wedge \exists \mathbf{x} (\mathbf{x} \in \mathbf{C} \wedge \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a}))))$$

Now let A and B be sets. From the last line above, we may conclude that there exists \mathcal{F} such that for all $a \in A$ and all $b \in B$, there exists $C \in \mathcal{F}$ with $(a, b) \in C$. Let $D = \bigcup \mathcal{F}$. Given any $a \in A$ and $b \in B$, we then have $(a, b) \in D$. Now applying Separation to the set D and the formula $\exists a \exists b (a \in A \wedge b \in B \wedge \varphi(\mathbf{b}, \mathbf{x}, \mathbf{a}))$, there is a set E such that for all x , we have $x \in E$ if and only if there exists $a \in A$ and $b \in B$ with $x = (a, b)$. As usual, Extensionality gives uniqueness. \square

2.3 Relations and Functions

Now that we have ordered pairs and Cartesian products, we can really make some progress.

Definition 2.18. *A relation is a set R such that every set $x \in R$ is an ordered pair. In other words, R is a relation if $\forall x (x \in R \rightarrow \exists a \exists b (x = (a, b)))$.*

Given a relation R , we want to define its domain to be the set of first elements of ordered pairs which are elements of R , and we want to define its range to be the set of second elements of ordered pairs which are elements of R . These are good descriptions which can easily (though not shortly) be turned into formulas, but we need to know that there is some set which contains all of these elements in order to apply Separation. Since the elements of an ordered pair $(a, b) = \{\{a\}, \{a, b\}\}$ are “two deep”, a good exercise is to convince yourself that $\bigcup\bigcup R$ will work. This justifies the following definitions.

Definition 2.19. *Let R be a relation*

1. $\text{dom}(R)$ is the set of a such that there exists b with $(a, b) \in R$.
2. $\text{ran}(R)$ is the set of b such that there exists a with $(a, b) \in R$.

Definition 2.20. *Let R be a relation. We write aRb if $(a, b) \in R$.*

Definition 2.21. *Let A be a set. We say that R is a relation on A if $\text{dom}(R) \subseteq A$ and $\text{ran}(R) \subseteq A$.*

We define functions in the obvious way.

Definition 2.22. *A function f is a relation which is such that for all $a \in \text{dom}(f)$, there exists a unique $b \in \text{ran}(f)$ such that $(a, b) \in f$.*

Definition 2.23. *Let f be a function. We write $f(a) = b$ if $(a, b) \in f$.*

Definition 2.24. *Let f be a function. f is injective (or an injection) if whenever $f(a_1) = b$ and $f(a_2) = b$, we have $a_1 = a_2$.*

Definition 2.25. *Let A and B be sets. We write $f: A \rightarrow B$ to mean that f is a function, $\text{dom}(f) = A$ and $\text{ran}(f) \subseteq B$.*

We are now in a position to define when a function f is surjective and bijective. Notice that surjectivity and bijectivity are not properties of a function itself because these notions depend on a set which you consider to contain $\text{ran}(f)$. Once we have a fixed such set in mind, however, we can make the definitions.

Definition 2.26. *Let A and B be sets, and let $f: A \rightarrow B$.*

1. f is surjective (or a surjection) if $\text{ran}(f) = B$.
2. f is bijective (or a bijection) if f is injective and surjective.

Definition 2.27. *Let A and B be sets.*

1. We write $A \preceq B$ to mean that there is an injection $f: A \rightarrow B$.
2. We write $A \approx B$ to mean that there is a bijection $f: A \rightarrow B$.

Proposition 2.28. *Let A , B , and C be sets.*

1. If $A \preceq B$ and $B \preceq C$, then $A \preceq C$.
2. $A \approx A$.
3. If $A \approx B$, then $B \approx A$.
4. If $A \approx B$ and $B \approx C$, then $A \approx C$.

2.4 Orderings

Definition 2.29. Let R be a relation on a set A .

1. R is reflexive on A if for all $a \in A$, we have aRa .
2. R is symmetric on A if for all $a, b \in A$, if aRb then bRa .
3. R is asymmetric on A if for all $a, b \in A$, if aRb then it is not the case that bRa .
4. R is antisymmetric on A if for all $a, b \in A$, if aRb and bRa , then $a = b$.
5. R is transitive on A if for all $a, b, c \in A$, if aRb and bRc , then aRc .
6. R is connected on A if for all $a, b \in A$, either aRb , $a = b$, or bRa .

Definition 2.30. Let R be a relation on a set A .

1. R is a partial ordering on A if R is transitive on A and asymmetric on A .
2. R is a linear ordering on A if R is a partial ordering on A and R is connected on A .
3. R is a well-ordering on A if R is a linear ordering on A and for every $X \subseteq A$ with $X \neq \emptyset$, there exists $m \in X$ such that for all $x \in X$, either $m = x$ or mRx .

2.5 The Natural Numbers and Induction

We specifically added the Axiom of Infinity with the hope that it captured the idea of the set of natural numbers. We now show how this axiom, in league with the others, allows us to embed the theory of the natural numbers into set theory. We start by defining the initial natural number and successors of sets.

Definition 2.31. $0 = \emptyset$

Definition 2.32. Given a set x , we let $S(x) = x \cup \{x\}$, and we call $S(x)$ the successor of x .

With 0 and the notion of successor, we can then go on to define $1 = S(0)$, $2 = S(1) = S(S(0))$, and continue in this way to define any particular natural number. However, we are seeking to form the set of all natural numbers.

Definition 2.33. A set I is inductive if $0 \in I$ and for all $x \in I$, we have $S(x) \in I$.

The Axiom of Infinity simply asserts the existence of some inductive set J . Intuitively, we have $0 \in J$, $S(0) \in J$, $S(S(0)) \in J$, and so on. However, J may very well contain more than just repeated applications of S to 0 . We now use the top-down approach to generation to define the natural numbers (the other two approaches will not work yet because their definitions rely on the natural numbers).

Proposition 2.34. There is a smallest inductive set. That is, there is an inductive set K such that $K \subseteq I$ for every inductive set I .

Proof. By the Axiom of Infinity, we may fix an inductive set J . Let $K = \{x \in J : x \in I \text{ for every inductive set } I\}$. Notice that $0 \in K$ because $0 \in I$ for every inductive set I (and so, in particular, $0 \in J$). Suppose that $x \in K$. If I is inductive, then $x \in I$, hence $S(x) \in I$. It follows that $S(x) \in I$ for every inductive set I (and so, in particular, $S(x) \in J$), hence $S(x) \in K$. Therefore, K is inductive. By definition of K , we have $K \subseteq I$ whenever I is inductive. \square

By Extensionality, there is a unique smallest inductive set, so this justifies the following definition.

Definition 2.35. We denote the unique smallest inductive set by ω .

We think that ω captures our intuitive idea of the set of natural numbers, and it is now our goal to show how to prove the basic statements about the natural numbers which are often accepted axiomatically. We first define a relation $<$ (along with a few related relation) on ω . Remember our intuitive idea is that \in captures the order relationship on the natural numbers.

Definition 2.36.

1. We define a relation $<$ on ω by setting $< = \{(n, m) \in \omega \times \omega : n \in m\}$.
2. We define a relation \leq on ω by setting $\leq = \{(n, m) \in \omega \times \omega : n < m \text{ or } n = m\}$.
3. We define a relation $>$ on ω by setting $> = \{(n, m) \in \omega \times \omega : m < n\}$.
4. We define a relation \geq on ω by setting $\geq = \{(n, m) \in \omega \times \omega : n > m \text{ or } n = m\}$.

Lemma 2.37. There is no $n \in \omega$ with $n < 0$.

Proof. Since $0 = \emptyset$, there is no set x such that $x \in 0$. Therefore, there is no $n \in \omega$ with $n < 0$. □

Lemma 2.38. Let $m, n \in \omega$. We have $m < S(n)$ if and only if $m \leq n$.

Proof. Let $m, n \in \omega$. We then have $S(n) \in \omega$ since ω is inductive, and

$$\begin{aligned}
 m < S(n) &\Leftrightarrow m \in S(n) \\
 &\Leftrightarrow m \in n \cup \{n\} \\
 &\Leftrightarrow \text{Either } m \in n \text{ or } m \in \{n\} \\
 &\Leftrightarrow \text{Either } m < n \text{ or } m = n \\
 &\Leftrightarrow m \leq n.
 \end{aligned}$$

This proves the lemma. □

Our primary objective is to show that $<$ is a well-ordering on ω . Due to the nature of the definition of ω , it seems that only way to prove nontrivial results about ω is “by induction”. We state the Step Induction Principle in two forms. The first is much cleaner and seemingly more powerful (because it immediately implies the second and we can quantify over sets but not over formulas), but the second is how one often thinks about induction is used in practice (using “properties” of natural numbers) and will be the only form that we can generalize to the collection of all ordinals.

Proposition 2.39 (Step Induction Principle on ω).

1. Suppose that X is a set, $0 \in X$, and for all $n \in \omega$, if $n \in X$ then $S(n) \in X$. We then have $\omega \subseteq X$.
2. For any formula $\varphi(n, \vec{p})$, we have the sentence

$$\forall \vec{p} ((\varphi(0, \vec{p}) \wedge (\forall n \in \omega)(\varphi(n, \vec{p}) \rightarrow \varphi(S(n), \vec{p}))) \rightarrow (\forall n \in \omega)\varphi(n, \vec{p}))$$

where $\varphi(S(n), \vec{p})$ is shorthand for the formula

$$\exists x (\forall y (y \in x \leftrightarrow (y \in n \vee y = n)) \wedge \varphi(x, \vec{p}))$$

Proof.

1. Let $Y = X \cap \omega$. Notice first that $0 \in Y$. Suppose now that $n \in Y = X \cap \omega$. We then have $n \in \omega$ and $n \in X$, so $S(n) \in \omega$ (because ω is inductive), and $S(n) \in X$ by assumption. Hence, $S(n) \in Y$. Therefore, Y is inductive, so we may conclude that $\omega \subseteq Y$. It follows that $\omega \subseteq X$.
2. Fix sets \vec{q} , and suppose $\varphi(0, \vec{q})$ and $(\forall n \in \omega)(\varphi(n, \vec{q}) \rightarrow \varphi(S(n), \vec{q}))$. Let $X = \{n \in \omega : \varphi(n, \vec{q})\}$. Notice that $0 \in X$ and for all $n \in \omega$, if $n \in X$ then $S(n) \in X$ by assumption. It follows from part 1 that $\omega \subseteq X$. Therefore, we have $(\forall n \in \omega)\varphi(n, \vec{q})$.

□

With the Step Induction Principle in hand, we can begin to prove the basic facts about the natural numbers. Our goal is to prove that $<$ is a well-ordering on ω , but it will take some time to get there. We first give a very simple inductive proof. For this proof only, we will give careful arguments using both versions of Step Induction to show how a usual induction proof can be formalized in either way.

Lemma 2.40. *For all $n \in \omega$, we have $0 \leq n$.*

Proof. The following two proofs correspond to the above two versions of the Induction Principle.

1. Let $X = \{n \in \omega : 0 \leq n\}$, and notice that $0 \in X$. Suppose now that $n \in X$. We then have $n \in \omega$ and $0 \leq n$, hence $0 < S(n)$ by Lemma 2.38, so $S(n) \in X$. Thus, by Step Induction, we have $\omega \subseteq X$. Therefore, for all $n \in \omega$, we have $0 \leq n$.
2. Let $\varphi(n)$ be the formula “ $0 \leq n$ ”. We clearly have $\varphi(0)$ because $0 = 0$. Suppose now that $n \in \omega$ and $\varphi(n)$. We then have $0 \leq n$, hence $0 < S(n)$ by Lemma 2.38. It follows that $\varphi(S(n))$. Therefore, by Step Induction, we have $0 \leq n$ for all $n \in \omega$.

□

We give a few more careful inductive proof using the second version of the Induction Principle to illustrate how parameters can be used. Afterwards, our later inductive proofs will be given in a more natural relaxed style.

Our relation $<$ is given by \in , but it is only defined on elements of ω . We thus need the following proposition which says that every element of a natural number is a natural number.

Proposition 2.41. *Suppose that $n \in \omega$ and $m \in n$. We then have $m \in \omega$.*

Proof. The proof is “by induction on n ”; that is, we hold m fixed by treating it as a parameter. Thus, fix $m \in \omega$. Let $X = \{n \in \omega : m \in n \rightarrow m \in \omega\}$. Notice that $0 \in X$ because $m \notin 0 = \emptyset$. Suppose now that $n \in X$. We show that $S(n) \in X$. Suppose that $m \in S(n) = n \cup \{n\}$. We then know that either $m \in n$, in which case $m \in \omega$ by induction (i.e. because $n \in X$), or $m = n$, in which case we clearly have $m \in \omega$. It follows that $S(n) \in X$. Therefore, by Step Induction, we may conclude that $X = \omega$. Since $m \in \omega$ was arbitrary, the result follows. □

Proposition 2.42. *$<$ is transitive on ω .*

Proof. We prove the result by induction on n . Fix $k \in \omega$. Let $X = \{n \in \omega : (k < m \wedge m < n) \rightarrow k < n\}$. We then have that $0 \in X$ vacuously because we do not have $m < 0$ by Lemma 2.37. Suppose now that $n \in X$. We show that $S(n) \in X$. Suppose that $k < m$ and $m < S(n)$ (if not, then $S(n) \in X$ vacuously). By Lemma 2.38, we have $m \leq n$, hence either $m < n$ or $m = n$. If $m < n$, then $k < n$ because $n \in X$. If $m = n$, then $k < n$ because $k < m$. Therefore, in either case, we have $k < n$, and hence $k < S(n)$ by Lemma 2.38. It follows that $S(n) \in X$. Thus, by Step Induction, we may conclude that $X = \omega$. Since $k, m \in \omega$ were arbitrary, the result follows. □

Lemma 2.43. *Let $m, n \in \omega$. We have $S(m) \leq n$ if and only if $m < n$.*

Proof. Suppose first that $m, n \in \omega$ and $S(m) \leq n$.

Case 1: Suppose that $S(m) = n$. We have $m < S(m)$ by Lemma 2.38, hence $m < n$.

Case 2: Suppose that $S(m) < n$. We have $m < S(m)$ by Lemma 2.38, hence $m < n$ by Proposition 2.42. Therefore, for all $n, m \in \omega$, if $S(m) \leq n$, then $m < n$.

We prove the converse statement that for all $m, n \in \omega$, if $m < n$, then $S(m) \leq n$ by induction on n . Fix $m \in \omega$. Let $X = \{n \in \omega : m < n \rightarrow S(m) \leq n\}$. We have $0 \in X$ vacuously because we do not have $m < 0$ by Lemma 2.37. Suppose now that $n \in X$. We show that $S(n) \in X$. Suppose that $m < S(n)$ (otherwise, $S(n) \in X$ vacuously). By Lemma 2.38, we have $m \leq n$.

Case 1: Suppose that $m = n$. We then have $S(m) = S(n)$, hence $S(n) \in X$.

Case 2: Suppose that $m < n$. Since $n \in X$, we have $S(m) \leq n$. By Lemma 2.38, we know that $n < S(n)$. If $S(m) = n$, this immediately gives $S(m) < S(n)$, while if $S(m) < n$, we may conclude that $S(m) < S(n)$ by Proposition 2.42. Hence, we have $S(n) \in X$.

Thus, by Step Induction, we may conclude that $X = \omega$. Since $m \in \omega$ was arbitrary, the result follows. \square

Lemma 2.44. *There is no $n \in \omega$ with $n < n$.*

Proof. This follows immediately from the Axiom of Foundation, but we prove it without that assumption. Let $X = \{n \in \omega : \neg(n < n)\}$. We have that $0 \in X$ by Lemma 2.37. Suppose that $n \in X$. We prove that $S(n) \in X$ by supposing that $S(n) < S(n)$ and deriving a contradiction. Suppose then that $S(n) < S(n)$. By Lemma 2.38, we have $S(n) \leq n$, hence either $S(n) = n$ or $S(n) < n$. Also by Lemma 2.38, we have $n < S(n)$. Therefore, if $S(n) = n$, then $n < n$, and if $S(n) < n$, then $n < n$ by Proposition 2.42 (since $n < S(n)$ and $S(n) < n$), a contradiction. It follows that $S(n) \in X$. Therefore, there is no $n \in \omega$ with $n < n$. \square

Proposition 2.45. *$<$ is asymmetric on ω .*

Proof. Suppose that $n, m \in \omega$, $n < m$, and $m < n$. By Proposition 2.42, it follows that $n < n$, contradicting Lemma 2.44. \square

Proposition 2.46. *$<$ is connected on ω .*

Proof. Fix $m \in \omega$. We prove that for all $n \in \omega$, either $m < n$, $m = n$, or $n < m$ by induction on n . Let $X = \{n \in \omega : (m < n) \vee (m = n) \vee (n < m)\}$. We have $0 \leq m$ by Lemma 2.40, hence either $m = 0$ or $0 < m$, and so $0 \in X$. Suppose then that $n \in X$, so that either $m < n$, $m = n$, or $n < m$.

Case 1: Suppose that $m < n$. Since $n < S(n)$ by Lemma 2.38, we have $m < S(n)$ by Proposition 2.42.

Case 2: Suppose that $m = n$. Since $n < S(n)$ by Lemma 2.38, it follows that $m < S(n)$.

Case 3: Suppose that $n < m$. We have $S(n) \leq m$ by Lemma 2.43. Hence, either $m = S(n)$ or $S(n) < m$.

Therefore, in all cases, either $m < S(n)$, $m = S(n)$, or $S(n) < m$, so $S(n) \in X$. The result follows by induction. \square

In order to finish off the proof that $<$ is a well-ordering on ω , we need a new version of induction. You may have heard it referred to as “Strong Induction”.

Proposition 2.47 (Induction Principle on ω).

1. Suppose that X is set and for all $n \in \omega$, if $m \in X$ for all $m < n$, then $n \in X$. We then have $\omega \subseteq X$.
2. For any formula $\varphi(n, \vec{p})$, we have the sentence

$$\forall \vec{p}((\forall n \in \omega)((\forall m < n)\varphi(m, \vec{p}) \rightarrow \varphi(n, \vec{p})) \rightarrow (\forall n \in \omega)\varphi(n, \vec{p}))$$

Proof.

1. Let $Y = \{n \in \omega : (\forall m < n)(m \in X)\}$. Notice that $Y \subseteq \omega$ and $0 \in Y$ because there is no $m \in \omega$ with $m < 0$ by Lemma 2.37. Suppose that $n \in Y$. We show that $S(n) \in Y$. Suppose that $m < S(n)$. By Lemma 2.38, we have $m \leq n$, hence either $m < n$ or $m = n$. If $m < n$, then $m \in X$ because $n \in Y$. For the case $m = n$, notice that $n \in X$ by assumption (because $m \in X$ for all $m < n$). Therefore, $S(n) \in Y$. By Step Induction, it follows that $\omega \subseteq Y$.

Now let $n \in X$. We have $n \in \omega$, hence $S(n) \in \omega$ because ω is inductive, so $S(n) \in Y$. Since $n < S(n)$ by Lemma 2.38, it follows that $n \in X$. Therefore, $\omega \subseteq X$.

2. This follows from part 1 using Separation. Fix sets \vec{q} , and suppose that

$$(\forall n \in \omega)((\forall m < n)\varphi(m, \vec{q}) \rightarrow \varphi(n, \vec{q}))$$

Let $X = \{n \in \omega : \varphi(n, \vec{q})\}$. Suppose that $n \in \omega$ and $m \in X$ for all $m < n$. We then have $(\forall m < n)\varphi(m, \vec{q})$, hence $\varphi(n, \vec{q})$ by assumption, so $n \in X$. It follows from part 1 that $\omega \subseteq X$. Therefore, we have $(\forall n \in \omega)\varphi(n, \vec{q})$. □

It is possible to give a proof of part 2 which makes use of part 2 of the Step Induction Principle, thus avoiding the detour through sets and using only formulas. This proof simply mimics how we obtained part 1 above, but uses formulas everywhere instead of working with sets. Although it is not nearly as clean, when we treat ordinals, there will times when we need to argue at the level of formulas.

Theorem 2.48. *$<$ is a well-ordering on ω*

Proof. By Proposition 2.42, Proposition 2.45, and Proposition 2.46, it follows that $<$ is a linear ordering on ω . Suppose then that $Z \subseteq \omega$ and there is no $n \in Z$ such that for all $m \in Z$, either $n = m$ or $n < m$. We show that $Z = \emptyset$. Notice that for every $n \in Z$, there exists $m \in Z$ with $m < n$ by Proposition 2.42.

Let $Y = \omega \setminus Z$. We show that $Y = \omega$ using the Induction Principle. Notice first that $0 \in Y$ because if $0 \in Z$, then there exists $m \in Z$ with $m < 0$ by the last sentence of the previous paragraph, contrary to Lemma 2.37. Suppose then that $n \in \omega$ is such that $m \in Y$, i.e. $m \notin Z$ for all $m < n$. If $n \notin Y$, we would then have that $n \in Z$, so by the last sentence of the previous paragraph, there exists $m \in Z$ with $m < n$, a contradiction. Therefore, $n \in Y$. Hence, by the Induction Principle, we have that $Y = \omega$ and so $Z = \emptyset$.

Therefore, if $Z \subseteq \omega$ and $Z \neq \emptyset$, there exists $n \in X$ such that for all $m \in Z$, either $n = m$ or $n < m$. It follows that $<$ is a well-ordering on ω . □

2.6 Sets and Classes

We know from Proposition 2.4 that there is no set u such that $a \in u$ for all sets a . Thus, our theory forbids us from placing every set into one universal set which we can then play with and manipulate. However, this formal impossibility within our theory does not prevent us from thinking about or referring to the “collection” of all sets or other “collections” which are too “large” to form into a set. After all, our universal quantifiers do indeed range over the “collection” of all sets. Also, if we are arguing semantically, then given a model \mathcal{M} of *ZFC*, we may “externally” work with the power set of M .

We want to be able to reason about such “collections” of sets in a natural manner within our theory without violating our theory. We will call such “collections” *classes* to distinguish them from sets. The idea is to recall that any first-order theory can say things about certain subsets of every model: the definable subsets. In our case, a formula $\varphi(x)$ is implicitly defining a certain collection of sets. Perhaps this collection is too “large” to put together into a set inside the model, but we may nevertheless use the formula in various ways within our theory. For example, for any formulas $\varphi(x)$ and $\psi(x)$, the sentence $\forall x(\varphi(x) \rightarrow \psi(x))$ says that every set which satisfies φ also satisfies ψ . If there exists sets C and D such that $\forall x(\varphi(x) \rightarrow x \in C)$ and $\forall x(\psi(x) \rightarrow x \in D)$, then we can use Separation to form the sets $A = \{x \in C : \varphi(x)\}$ and $B = \{x \in D : \psi(x)\}$,

in which case the sentence $\forall x(\varphi(x) \rightarrow \psi(x))$ simply asserts that $A \subseteq B$. However, even if we can't form these sets (intuitively because $\{x : \varphi(x)\}$ and $\{x : \psi(x)\}$ are too “large” to be sets), the sentence is expressing the same underlying idea. Allowing the possibility of parameters, this motivates the following “internal” definition.

Definition 2.49. A class \mathbf{C} is a formula $\varphi(x, \vec{p})$.

Our course, this isn't a very good way to think about classes. Externally, a class is simply a definable set (with the possibility of parameters). The idea is that once we fix sets \vec{q} to fill in for the position of the parameters, the formula describes the collection of those sets a such that $\varphi(a, \vec{q})$. The first class to consider is the class of all sets, which we denote by \mathbf{V} . Formally, we define \mathbf{V} to be the formula $x = x$, but we will content ourselves with defining classes in the following more informal “external” style.

Definition 2.50. \mathbf{V} is the class of all sets.

Here's a more interesting illustration of how classes can be used and why we want to consider them. Let \mathbf{C}_R be the class of all relations and let \mathbf{C}_F be the class of all functions. More formally, \mathbf{C}_R is the formula $\varphi_R(x)$ given by

$$\forall y(y \in x \rightarrow \exists a \exists b(y = (a, b)))$$

while \mathbf{C}_F is the formula $\varphi_F(x)$ given by

$$\forall y(y \in x \rightarrow \exists a \exists b(y = (a, b))) \wedge \forall a \forall b_1 \forall b_2(((a, b_1) \in x \wedge (a, b_2) \in x) \rightarrow b_1 = b_2)$$

With this shorthand in place, we can write things like $\mathbf{C}_F \subseteq \mathbf{C}_R$ to stand for the provable sentence $\forall x(\varphi_F(x) \rightarrow \varphi_R(x))$. Thus, by using the language of classes, we can express complicated formulas in a simplified, more suggestive, fashion. Of course, there's no real need to introduce classes because we could always just refer to the formulas, but it is psychologically easier to think of a class as some kind of ultra-set which our theory is able to handle, even if we are limited in what we can do with classes.

With the ability to refer to classes, why deal with sets at all? The answer is that classes are much less versatile than sets. For example, if \mathbf{C} and \mathbf{D} are classes, it makes no sense to write $\mathbf{C} \in \mathbf{D}$ because this doesn't correspond to a formula built from the implicit formulas giving \mathbf{C} and \mathbf{D} . This inability corresponds to the intuition that classes are too “large” to collect together into a set and then put into other collections. Hence, asking whether $\mathbf{V} \in \mathbf{V}$ is meaningless. Also, since classes are given by formulas, we are restricted to referring only to “definable” collections. Thus, there is no way to talk about or quantify over all “collections” of sets (something that is meaningless internally). However, there are many operation which do make sense on classes.

For instance, suppose that \mathbf{R} is a class of ordered pairs (with parameters \vec{p}). That is, \mathbf{R} is a formula $\varphi(x, \vec{p})$ such that the formula $\forall x(\varphi(x, \vec{p}) \rightarrow \exists a \exists b(x = (a, b)))$ is provable. We think of \mathbf{R} as a *class relation*. Using suggestive notation, we can then go on to define $\text{dom}(\mathbf{R})$ to be the class consisting of those sets a such that there exists a set b with $(a, b) \in \mathbf{R}$. To be precise, $\text{dom}(\mathbf{R})$ is the class which is the formula $\psi(a, \vec{p})$ given by $\exists x \exists b(x = (a, b) \wedge \varphi(x, \vec{p}))$. Thus, we can think of $\text{dom}(\cdot)$ as a operation on classes (given any formula $\varphi(x, \vec{p})$ which is a class relation, applying $\text{dom}(\cdot)$ results in the class given by the formula $\exists x \exists b(x = (a, b) \wedge \varphi(x, \vec{p}))$).

Similarly, we can talk about *class functions*. We can even use notation like $\mathbf{F}: \mathbf{V} \rightarrow \mathbf{V}$ to mean that \mathbf{F} is a class function with $\text{dom}(\mathbf{F}) = \mathbf{V}$. Again, each of these expressions could have been written out as formulas in our language, but the notation is so suggestive that it's clear how to do this without actually having to do it. An example of a general class function is $\mathbf{U}: \mathbf{V} \times \mathbf{V} \rightarrow \mathbf{V}$ given by $\mathbf{U}(a, b) = a \cup b$. Convince yourself how to write \mathbf{U} as a formula.

We can not quantify over classes within our theory in the same way that we can quantify over sets because there is no way to quantify over the formulas of set theory within set theory. However, we can, at the price of considering one “theorem” as infinitely many (one for each formula), make sense of a theorem which does universally quantify over classes. For example, consider the following.

Proposition 2.51. *Suppose that \mathbf{C} is a class, $0 \in \mathbf{C}$, and for all $n \in \omega$, if $n \in \mathbf{C}$ then $S(n) \in \mathbf{C}$. We then have $\omega \subseteq \mathbf{C}$.*

This proposition is what is obtained from the first version of Step Induction on ω by replacing the set X with the class \mathbf{C} . Although the set version can be written as one sentence which is provable in ZFC, this version can not because we can't quantify over classes in the theory. Unwrapping this proposition into formulas, it says that for every formula $\varphi(x, \vec{p})$, if we can prove $\varphi(0, \vec{p})$ and $(\forall n \in \omega)(\varphi(n, \vec{p}) \rightarrow \varphi(S(n), \vec{p}))$, then we can prove $(\forall n \in \omega)\varphi(n, \vec{p})$. That is, for each formula $\varphi(x, \vec{p})$, we can prove the sentence

$$\forall \vec{p}((\varphi(0, \vec{p}) \wedge (\forall n \in \omega)(\varphi(n, \vec{p}) \rightarrow \varphi(S(n), \vec{p}))) \rightarrow (\forall n \in \omega)\varphi(n, \vec{p}))$$

Thus, the class version is simply a neater way of writing the second version of Step Induction on ω which masks the fact that the quantification over classes requires us to write it as infinitely many different propositions (one for each formula $\varphi(x, \vec{p})$) in our theory.

Every set can be viewed as a class by making use of the class \mathbf{M} given by the formula $x \in p$. That is, once we fix a set p , the class $x \in p$ describes exactly the elements of p . For example, using \mathbf{M} in class version of Step Induction on ω , we see that the following sentence is provable:

$$\forall p((0 \in p \wedge (\forall n \in \omega)(n \in p \rightarrow S(n) \in p)) \rightarrow (\forall n \in \omega)(n \in p))$$

Notice that this is exactly the set version of Step Induction on ω .

On the other hand, not every class can be viewed as a set (look at \mathbf{V} , for example). Let \mathbf{C} be a class. We say that \mathbf{C} is a set if there exists a set A such that for all x , we have $x \in \mathbf{C}$ if and only if $x \in A$. At the level of formulas, this means that if \mathbf{C} is given by the formula $\varphi(x, \vec{p})$, then we can prove the formula $\exists A \forall x(\varphi(x, \vec{p}) \leftrightarrow x \in A)$. By Separation, this is equivalent to saying that there is a set B such that for all x , if $x \in \mathbf{C}$ then $x \in B$ (i.e. we can prove the formula $\exists B \forall x(\varphi(x, \vec{p}) \rightarrow x \in B)$). A class which is not set (that is, we can prove $\neg(\exists A \forall x(\varphi(x, \vec{p}) \leftrightarrow x \in A))$) is called a *proper class*. For example, \mathbf{V} is a proper class.

The following proposition will be helpful to us when we discuss transfinite constructions. Intuitively, it says that proper classes are too large to be embedded into any set.

Proposition 2.52. *Let \mathbf{C} be a proper class and let A be a set. There is no injective class function $\mathbf{F}: \mathbf{C} \rightarrow A$.*

Proof. Suppose that $\mathbf{F}: \mathbf{C} \rightarrow A$ is an injective class function. Let $B = \{a \in A : \exists c(c \in \mathbf{C} \wedge \mathbf{F}(c) = a)\}$ and notice that B is a set by Separation (recall that \mathbf{C} and \mathbf{F} are given by formulas). Since for each $b \in B$, there is a unique $c \in \mathbf{C}$ with $\mathbf{F}(c) = b$ (using the fact that \mathbf{F} is injective), we may use Collection and Separation to conclude that \mathbf{C} is a set, contradicting the fact that \mathbf{C} is a proper class. \square

We end this section by seeing how to simply restate the Axiom of Separation and the Axiom of Collection in the language of classes.

Axiom of Separation: Every subclass of a set is a set.

Axiom of Collection: If \mathbf{F} is a class function and A is a set, then there is a set containing the image of A under \mathbf{F} .

2.7 Finite Sets, Powers, and Products

2.7.1 Finite Sets

Definition 2.53. *Let A be a set. A is finite if there exists $n \in \omega$ such that $A \approx n$. If A is not finite, we say that A is infinite.*

Proposition 2.54. *Suppose that $n \in \omega$. Every injective $f: n \rightarrow n$ is bijective.*

Proof. The proof is by induction on $n \in \omega$. Suppose first that $n = 0$ and $f: 0 \rightarrow 0$ is injective. We then have $f = \emptyset$, so f is trivially bijective. Suppose now that the result holds for n so that every injective $f: n \rightarrow n$ is bijective. Suppose that $f: S(n) \rightarrow S(n)$. We then have $f(n) \leq n$, and we consider two cases.

Case 1: Suppose that $f(n) = n$. Since f is injective, we have $f(m) \neq n$ for every $m < n$, hence $f(m) < n$ for every $m < n$ (because $f(m) < S(n)$ for every $m < n$). It follows that $f \upharpoonright n: n \rightarrow n$. Notice that $f \upharpoonright n: n \rightarrow n$ is injective because f is injective, hence $f \upharpoonright n$ is bijective by induction. Therefore, $\text{ran}(f \upharpoonright n) = n$, and hence $\text{ran}(f) = S(n)$ (because $f(n) = n$). It follows that f is surjective, so f is bijective.

Case 2: Suppose that $f(n) < n$. We first claim that $n \in \text{ran}(f)$. Suppose instead that $n \notin \text{ran}(f)$. Notice that $f \upharpoonright n: n \rightarrow n$ is injective because f is injective, hence $f \upharpoonright n$ is bijective by induction. Therefore, $f(n) \in \text{ran}(f \upharpoonright n)$ (because $f(n) < n$), so there exists $\ell < n$ with $f(\ell) = f(n)$, contrary to the fact that f is injective. It follows that $n \in \text{ran}(f)$. Fix $k < n$ with $f(k) = n$. Define a function $g: n \rightarrow n$ by

$$g(m) = \begin{cases} f(m) & \text{if } m \neq k \\ f(n) & \text{if } m = k \end{cases}$$

Notice that if $m_1, m_2 < n$ with $m_1 \neq m_2$ and $m_1, m_2 \neq k$, then $g(m_1) \neq g(m_2)$ since $f(m_1) \neq f(m_2)$ (because f is injective). Also, if $m < n$ with $m \neq k$, then $g(m) \neq g(k)$ since $f(m) \neq f(n)$ (again because f is injective). It follows that $g: n \rightarrow n$ is injective, hence bijective by induction. From this we can conclude that $\text{ran}(f) = S(n)$ as follows. Notice that $f(n) \in \text{ran}(f)$ and $n \in \text{ran}(f)$ because $f(k) = n$. Suppose that $\ell < n$ with $\ell \neq f(n)$. Since $g: n \rightarrow n$ is bijective, there exists a unique $m < n$ with $g(m) = \ell$. Since $\ell \neq f(n)$, we have $m \neq k$, hence $f(m) = g(m) = \ell$, so $\ell \in \text{ran}(f)$. Therefore, $\text{ran}(f) = S(n)$, and hence f is bijective. \square

Corollary 2.55 (Pigeonhole Principle). *If $n, m \in \omega$ and $m > n$, then $m \not\leq n$.*

Proof. Suppose that $f: m \rightarrow n$ is injective. It then follows that $f \upharpoonright n: n \rightarrow n$ is injective, hence $f \upharpoonright n$ is bijective by Proposition 2.54. Therefore, since $f(n) \in n$, it follows that there exists $k < n$ with $f(k) = f(n)$, contradicting the fact that f is injective. Hence, $m \not\leq n$. \square

Corollary 2.56. *If $m, n \in \omega$ and $m \approx n$, then $m = n$.*

Proof. Suppose that $m \neq n$ so that either $m > n$ or $m < n$. If $m > n$, then $m \not\leq n$ by the Pigeonhole Principle, so $m \not\approx n$. If $m < n$, then $n \not\leq m$ by the Pigeonhole Principle, so $n \not\approx m$ and hence $m \not\approx n$. \square

Corollary 2.57. *If A is finite, there exists a unique $n \in \omega$ such that $A \approx n$.*

Definition 2.58. *If A is finite, the unique $n \in \omega$ such that $A \approx n$ is called the cardinality of A and is denoted by $|A|$.*

Proposition 2.59. *Let A be a nonempty set and let $n \in \omega$. The following are equivalent:*

1. $A \preceq n$.
2. There exists a surjection $g: n \rightarrow A$.
3. A is finite and $|A| \leq n$.

Proof. 1 implies 2: Suppose that $A \preceq n$ and fix an injection $f: A \rightarrow n$. Fix an element $b \in A$ (which exists since $A \neq \emptyset$). Define $g: n \rightarrow A$ by letting

$$g = \{(m, a) \in n \times A : f(a) = m\} \cup \{(m, a) \in n \times A : m \notin \text{ran}(f) \text{ and } a = b\}.$$

Notice that $g: n \rightarrow A$ and that g is a surjection.

2 implies 1: Suppose that $g: n \rightarrow A$ is a surjection. Define a set f by letting

$$f = \{(a, m) \in A \times n : g(m) = a \text{ and } g(k) \neq a \text{ for all } k < m\}$$

Using the fact $<$ well-orders ω and that g is a surjection, it follows that $f: A \rightarrow n$. Also, f is injective because g is a function.

1 implies 3: Suppose that $A \preceq n$. Let m be the least element of ω such that $A \preceq m$, and fix an injection $g: A \rightarrow m$. We claim that g is a bijection. Notice that $m \neq 0$ because A is nonempty, so we may fix $k \in \omega$ with $m = S(k)$. If g is not a bijection, we could construct an injective $h: A \rightarrow k$, a contradiction.

3 implies 1: Suppose that A is finite and $|A| \leq n$. Let $m = |A| \leq n$ and fix a bijection $f: A \rightarrow m$. We then have that $f: A \rightarrow n$ is an injection, so $A \preceq n$. \square

Corollary 2.60. *Suppose that $n \in \omega$. Every surjective $g: n \rightarrow n$ is bijective.*

Proof. Suppose that $g: n \rightarrow n$ is surjective. Define an injective $f: n \rightarrow n$ such that $f \circ g = id_n$ as above. We then have that f is bijective, hence g is bijective. \square

2.7.2 Finite Powers

It is possible to use ordered pairs to define ordered triples, ordered quadruples, and so on. For example, we could define the ordered triple (a, b, c) to be $((a, b), c)$. However, with the basic properties of ω in hand, we can give a much more elegant definition.

Proposition 2.61. *Let A be a set and let $n \in \omega$. There is a unique set, denoted by A^n , such that for all f , we have $f \in A^n$ if and only if $f: n \rightarrow A$.*

Proof. As usual, uniqueness follows from Extensionality, so we need only prove existence. The proof is by induction on n . Suppose that $n = 0$. Since for all f , we have $f: 0 \rightarrow A$ if and only if $f = \emptyset$, we may take $A^0 = \{\emptyset\}$. Suppose that the result holds for n , i.e. there exists a set A^n such that for all f , we have $f \in A^n$ if and only if $f: n \rightarrow A$.

Fix $a \in A$. Notice that for each $f \in A^n$, there is a unique function $f_a: S(n) \rightarrow A$ such that $f_a(m) = f(m)$ for all $m < n$ and $f_a(n) = a$ (let $f_a = f \cup \{(n, a)\}$ and use Lemma 2.38). Therefore, by Collection (since A^n is a set), Separation, and Extensionality, there is a unique set C_a such that for all g , we have $g \in C_a$ if and only if $g = f_a$ for some $f \in A^n$. Notice that for every $g: S(n) \rightarrow A$ with $g(n) = a$, there is an $f: n \rightarrow A$ such that $g = f_a$ (let $f = g \setminus \{(n, a)\}$). Therefore, for every g , we have $g \in C_a$ if and only if $g: S(n) \rightarrow A$ and $g(n) = a$.

By Collection (since A is a set), Separation, and Extensionality again, there is a set \mathcal{F} such that for all D , we have $D \in \mathcal{F}$ if and only if there exists $a \in A$ with $D = C_a$. Notice that for all g , we have $g \in \bigcup \mathcal{F}$ if and only if there exists $a \in A$ with $g \in C_a$. Let $A^{S(n)} = \bigcup \mathcal{F}$. For all g , we then have $g \in A^{S(n)}$ if and only if $g: S(n) \rightarrow A$. Therefore, by induction, for every $n \in \omega$, there is a set B such that for all f , we have $f \in B$ if and only if $f: n \rightarrow A$. \square

Proposition 2.62. *Let A be a set. There is a unique set, denoted by $A^{<\omega}$, such that for all f , we have $f \in A^{<\omega}$ if and only if $f \in A^n$ for some $n \in \omega$.*

Proof. By Collection (since ω is a set), Separation, and Extensionality, there is a unique set \mathcal{F} such that for all D , we have $D \in \mathcal{F}$ if and only if there exists $n \in \omega$ with $D = A^n$. Let $A^{<\omega} = \bigcup \mathcal{F}$. For every f , we then have $f \in A^{<\omega}$ if and only if $f \in A^n$ for some $n \in \omega$. \square

2.7.3 Finite Products

Suppose that f is a function with $\text{dom}(f) = n \in \omega$. We want to consider the Cartesian product of the sets indexed indexed by f .

$$\prod f = \{g \in \left(\bigcup \text{ran}(f)\right)^n : g(i) \in f(i) \text{ for all } i < n\}$$

2.8 Definitions by Recursion

Theorem 2.63 (Step Recursive Definitions on ω - Set Form). *Let A be a set, let $b \in A$, and let $g: \omega \times A \rightarrow A$. There exists a unique function $f: \omega \rightarrow A$ such that $f(0) = b$ and $f(S(n)) = g(n, f(n))$ for all $n \in \omega$.*

Proof. We first prove existence. Call a set $Z \subseteq \omega \times A$ *sufficient* if $(0, b) \in Z$ and for all $(n, a) \in Z$, we have $(S(n), g(n, a)) \in Z$. Notice that sufficient sets exist (since $\omega \times A$ is sufficient). Let

$$Y = \{(n, a) \in \omega \times A : (n, a) \in Z \text{ for every sufficient set } Z\}.$$

We first show that Y is sufficient. Notice that $(0, b) \in Y$ because $(0, b) \in Z$ for every sufficient set Z . Suppose now that $(n, a) \in Y$. For any sufficient set Z , we have $(n, a) \in Z$, hence $(S(n), g(n, a)) \in Z$. Therefore, $(S(n), g(n, a)) \in Z$ for every sufficient set Z , so $(S(n), g(n, a)) \in Y$. It follows that Y is sufficient.

We next show that for all $n \in \omega$, there exists a unique $a \in A$ such that $(n, a) \in Y$. Let

$$X = \{n \in \omega : \text{there exists a unique } a \in A \text{ such that } (n, a) \in Y\}.$$

Since Y is sufficient, we know that $(0, b) \in Y$. Suppose that $d \in A$ and $d \neq b$. Since the set $(\omega \times A) \setminus \{(0, d)\}$ is sufficient (because $S(n) \neq 0$ for all $n \in \omega$), it follows that $(0, d) \notin Y$. Therefore, there exists a unique $a \in A$ such that $(0, a) \in Y$ (namely, $a = b$), so $0 \in X$. Suppose now that $n \in X$, and let c be the unique element of A such that $(n, c) \in Y$. Since Y is sufficient, we have $(S(n), g(n, c)) \in Y$. Fix $d \in A$ with $d \neq g(n, c)$. We then have that $Y \setminus \{(S(n), d)\}$ is sufficient (otherwise, there exists $a \in A$ such that $(n, a) \in Y$ and $g(n, a) = d$, contrary to the fact that in this case we have $a = c$ by induction), so by definition of Y it follows that $Y \subseteq Y \setminus \{(S(n), d)\}$. Hence, $(S(n), d) \notin Y$. Therefore, there exists a unique $a \in A$ such that $(S(n), a) \in Y$ (namely, $a = g(n, c)$), so $S(n) \in X$. By induction, we conclude that $X = \omega$, so for all $n \in \omega$, there exists a unique $a \in A$ such that $(n, a) \in Y$.

Let $f = Y$ and notice that $f: \omega \rightarrow A$ from above. Since Y is sufficient, we have $(0, b) \in Y$, so $f(0) = b$. Let $n \in \omega$. Since $(n, f(n)) \in Y$ and Y is sufficient, it follows that $(S(n), g(n, f(n))) \in Y$, so $f(S(n)) = g(n, f(n))$.

We now prove uniqueness. Suppose that $f_1, f_2: \omega \rightarrow A$ are such that:

1. $f_1(0) = b$.
2. $f_2(0) = b$.
3. $f_1(S(n)) = g(n, f_1(n))$ for all $n \in \omega$.
4. $f_2(S(n)) = g(n, f_2(n))$ for all $n \in \omega$.

Let $X = \{n \in \omega : f_1(n) = f_2(n)\}$. Notice that $0 \in X$ because $f_1(0) = b = f_2(0)$. Suppose that $n \in X$ so that $f_1(n) = f_2(n)$. We then have

$$f_1(S(n)) = g(n, f_1(n)) = g(n, f_2(n)) = f_2(S(n))$$

hence $S(n) \in X$. It follows by induction that $X = \omega$, so $f_1(n) = f_2(n)$ for all $n \in \omega$. □

As an example of how to use this result (assuming we already know how to multiply - see below), consider how to define the factorial function. We want to justify the existence of a unique function $f: \omega \rightarrow \omega$ such that $f(0) = 1$ and $f(S(n)) = f(n) \cdot S(n)$ for all $n \in \omega$. We can make this work as follows. Let $A = \omega$, $b = 1$, and define $g: \omega \times \omega \rightarrow \omega$ by letting $g(a, n) = a \cdot S(n)$ (here we are thinking that the first argument of g will contain the ‘‘accumulated’’ value $f(n)$). The theorem now gives the existence and uniqueness of a function $f: \omega \rightarrow \omega$ such that $f(0) = 1$ and $f(S(n)) = f(n) \cdot S(n)$ for all $n \in \omega$.

However, this begs the question of how to define multiplication. Let’s start by thinking about how to define addition. The basic idea is to define it recursively. For any $m \in \omega$, we let $m + 0 = m$. If $m \in \omega$, and we know how to find $m + n$ for some fixed $n \in \omega$, then we should define $m + S(n) = S(m + n)$. It looks an

appeal to the above theorem is in order, but how do we treat the m that is fixed in the recursion? We need a slightly stronger version of the above theorem which allows a parameter to come along for the ride. The proof is basically the same so we just give a short sketch.

Theorem 2.64 (Step Recursive Definitions with Parameters on ω). *Let A and P be sets, let $h: P \rightarrow A$, and let $g: P \times \omega \times A \rightarrow A$. There exists a unique function $f: P \times \omega \rightarrow A$ such that $f(p, 0) = h(p)$ for all $p \in P$, and $f(p, S(n)) = g(p, n, f(p, n))$ for all $p \in P$ and all $n \in \omega$.*

Proof. One could reprove this from scratch following the above outline, but we give a simpler argument using Collection. For each $p \in P$, define $g_p: \omega \times A \rightarrow A$ by letting $g_p(n, a) = g(p, n, a)$ for all $(n, a) \in \omega \times A$. Using the above results without parameters, for each fixed $p \in P$, there exists a unique function $f_p: \omega \rightarrow A$ such that $f_p(0) = h(p)$ and $f_p(S(n)) = g_p(n, f_p(n))$ for all $n \in \omega$. By Collection and Separation, we may form the set $\{f_p : p \in P\}$. Let f be the union of this set. It is then straightforward to check that f is the unique function satisfying the necessary properties. \square

Definition 2.65. *Let $h: \omega \rightarrow \omega$ be defined by $h(m) = m$ and let $g: \omega \times \omega \times \omega \rightarrow \omega$ be defined by $g(m, n, a) = S(a)$. We denote the unique f from the previous theorem by $+$. Notice that $+: \omega \times \omega \rightarrow \omega$, that $m + 0 = m$ for all $m \in \omega$, and that $m + S(n) = S(m + n)$ for all $m, n \in \omega$.*

Now that we have the definition of $+$, we can prove all of the basic “axiomatic” facts about the natural numbers with $+$ by induction. Here’s a simple example.

Proposition 2.66. $0 + n = n$ for all $n \in \omega$.

Proof. The proof is by induction on n . For $n = 0$, simply notice that $0 + 0 = 0$. Suppose that $n \in \omega$ and $0 + n = n$. We then have $0 + S(n) = S(0 + n) = S(n)$. The result follows by induction. \square

A slightly more nontrivial example is a proof that $+$ is associative.

Proposition 2.67. *For all $k, m, n \in \omega$, we have $(k + m) + n = k + (m + n)$.*

Proof. We fix $k, m \in \omega$, and prove the result is by induction on n . Notice that $(k + m) + 0 = k + m = k + (m + 0)$. Suppose that we know the result for n , so that $(k + m) + n = k + (m + n)$. We then have

$$\begin{aligned} (k + m) + S(n) &= S((k + m) + n) \\ &= S(k + (m + n)) && \text{(by induction)} \\ &= k + S(m + n) \\ &= k + (m + S(n)) \end{aligned}$$

The result follows by induction. \square

Definition 2.68. *Let $h: \omega \rightarrow \omega$ be defined by $h(m) = 0$ and let $g: \omega \times \omega \times \omega \rightarrow \omega$ be defined by $g(m, a, n) = a + m$. We denote the unique f from the previous theorem by \cdot . Notice that $\cdot: \omega \times \omega \rightarrow \omega$, that $m \cdot 0 = 0$ for all $m \in \omega$, and that $m \cdot S(n) = m \cdot n + m$ for all $m, n \in \omega$.*

From now on, we will present our recursive definitions in the usual mathematical style. For example, we define iterates of a function as follows.

Definition 2.69. *Let B be a set, and let $h: B \rightarrow B$ be a function. We define, for each $n \in \omega$, a function h^n by letting $h^0 = id_B$ and letting $h^{S(n)} = h \circ h^n$ for all $n \in \omega$.*

For each fixed $h: B \rightarrow B$, this definition can be justified by appealing to the theorem with $A = B^B$, $b = id_B$, and $g: A \times \omega \rightarrow \omega$ given by $g(a, n) = h \circ a$. However, we will content ourselves with the above more informal style when the details are straightforward and uninteresting.

The above notions of recursive definitions can only handle types of recursion where the value of $f(S(n))$ depends just on the previous value $f(n)$ (and also n). Thus, it is unable to deal with recursive definitions such as that used in defining the Fibonacci sequence where the value of $f(n)$ depends on the two previous values of f whenever $n \geq 2$. We can justify these more general types of recursions by carrying along all previous values of f in the inductive construction. Thus, instead of having our iterating function $g: A \times \omega \rightarrow A$, where we think of the first argument of g as carrying the current value $f(n)$, we will have an iterating function $g: A^{<\omega} \rightarrow A$, where we think of the first argument of g as carrying the finite sequence consisting of all values $f(m)$ for $m < n$. Thus, given such a g , we are seeking the existence and uniqueness of a function $f: \omega \rightarrow A$ such that $f(n) = g(f \upharpoonright n)$ for all $n \in \omega$. Notice that in this framework, we no longer need to put forward a $b \in A$ as a starting place for f because we will have $f(0) = g(\emptyset)$. Also, we do not need to include a number argument in the domain of g because the current n in the iteration can be recovered as the domain of the single argument of g .

Theorem 2.70 (Recursive Definitions on ω). *Let A be a set and let $g: A^{<\omega} \rightarrow A$. There exists a unique function $f: \omega \rightarrow A$ such that $f(n) = g(f \upharpoonright n)$ for all $n \in \omega$.*

Proof. We first prove existence. Call a set $Z \subseteq \omega \times A$ *sufficient* if for all $n \in \omega$ and all $q \in A^n$ such that $(k, q(k)) \in Z$ for all $k < n$, we have $(n, g(q)) \in Z$. Notice that sufficient sets exist (since $\omega \times A$ is sufficient). Let

$$Y = \{(n, a) \in \omega \times A : (n, a) \in Z \text{ for every sufficient set } Z\}.$$

We first show that Y is sufficient. Suppose that $n \in \omega$, that $q \in A^n$, and that $(k, q(k)) \in Y$ for all $k < n$. For any sufficient set Z , we have $(k, q(k)) \in Z$ for all $k < n$, so $(n, g(q)) \in Z$. Therefore, $(n, g(q)) \in Z$ for every sufficient set Z , so $(n, g(q)) \in Y$. It follows that Y is sufficient.

We next show that for all $n \in \omega$, there exists a unique $a \in A$ such that $(n, a) \in Y$. Let

$$X = \{n \in \omega : \text{there exists a unique } a \in A \text{ such that } (n, a) \in Y\}.$$

Suppose that $n \in \omega$ is such that $k \in X$ for all $k < n$. Let $q = Y \cap (n \times A)$ and notice that $q \in A^n$. Since $(k, q(k)) \in Y$ for all $k < n$ and Y is sufficient, it follows that $(n, g(q)) \in Y$. Fix $b \in A$ with $b \neq g(q)$. We then have that $Y \setminus \{(n, b)\}$ is sufficient (otherwise, there exists $p \in A^n$ such that $(k, p(k)) \in Y$ for all $k < n$ and $g(p) = b$, but this implies that $p = q$ and hence $b = a$), so by definition of Y it follows that $Y \subseteq Y \setminus \{(n, b)\}$. Hence, $(n, b) \notin Y$. Therefore, there exists a unique $a \in A$ such that $(n, a) \in Y$, so $n \in X$. By induction, we conclude that $X = \omega$, so for all $n \in \omega$, there exists a unique $a \in A$ such that $(n, a) \in Y$.

Let $f = Y$ and notice that $f: \omega \rightarrow A$ from above. Suppose that $n \in \omega$. Let $q = Y \cap (n \times A)$ and notice that $q \in A^n$ and $q = f \upharpoonright n$. Since $(k, q(k)) \in Y$ for all $k < n$ and Y is sufficient, it follows that $(n, g(q)) \in Y$, so $f(n) = g(q) = g(f \upharpoonright n)$.

We now prove uniqueness. Suppose that $f_1, f_2: \omega \rightarrow A$ are such that:

1. $f_1(n) = g(f_1 \upharpoonright n)$ for all $n \in \omega$.
2. $f_2(n) = g(f_2 \upharpoonright n)$ for all $n \in \omega$.

Let $X = \{n \in \omega : f_1(n) = f_2(n)\}$. We prove by induction that $X = \omega$. Let $n \in \omega$ and suppose that $k \in X$ for all $k < n$. We then have that $f_1 \upharpoonright n = f_2 \upharpoonright n$, hence

$$f_1(n) = g(f_1 \upharpoonright n) = g(f_2 \upharpoonright n) = f_2(n)$$

hence $n \in X$. It follows by induction that $X = \omega$, so $f_1(n) = f_2(n)$ for all $n \in \omega$. □

As above, there is a similar version when we allow parameters. If $f: P \times \omega \rightarrow A$ and $p \in P$, we use the notation f_p to denote the function $f_p: \omega \rightarrow A$ given by $f_p(n) = f(p, n)$ for all $n \in \omega$.

Theorem 2.71 (Recursive Definitions with Parameters on ω). *Let A and P be sets and let $g: P \times A^{<\omega} \rightarrow A$. There exists a unique function $f: P \times \omega \rightarrow A$ such that $f(p, n) = g(p, f_p \upharpoonright n)$ for all $p \in P$ and $n \in \omega$.*

2.9 Infinite Sets, Powers, and Products

Theorem 2.72 (Cantor-Bernstein). *Let A and B be sets. If $A \preceq B$ and $B \preceq A$, then $A \approx B$.*

Proof. We may assume that A and B are disjoint (otherwise, we can work with $A \times \{0\}$ and $B \times \{1\}$, and transfer the result back to A and B). Fix injections $f: A \rightarrow B$ and $g: B \rightarrow A$. We say that an element $a \in A$ is *B-originating* if there exists $b_0 \in B$ and $n \in \omega$ such that $b_0 \notin \text{ran}(f)$ and $a = (g \circ f)^n(g(b_0))$. Similarly, we say that an element $b \in B$ is *B-originating* if there exists $b_0 \in B$ and $n \in \omega$ such that $b_0 \notin \text{ran}(f)$ and $b = (f \circ g)^n(b_0)$. Let

$$h = \{(a, b) \in A \times B : \text{Either } a \text{ is not } B\text{-originating and } f(a) = b \text{ or } a \text{ is } B\text{-originating and } g(b) = a\}$$

Notice that h is a function (because f is a function and g is injective), $\text{dom}(h) \subseteq A$, and $\text{ran}(h) \subseteq B$. We first show that $\text{dom}(h) = A$. Let $a \in A$. If a is not B -originating, then $(a, f(a)) \in h$, hence $a \in \text{dom}(h)$. Suppose that a is B -originating, and fix $b_0 \in B$ and $n \in \omega$ with $a = (g \circ f)^n(g(b_0))$. If $n = 0$, then $a = g(b_0)$, so $(a, b_0) \in h$ and hence $a \in \text{dom}(h)$. Suppose that $n \neq 0$ and fix $m \in \omega$ with $n = S(m)$. We then have $a = (g \circ f)^{S(m)}(g(b_0)) = (g \circ f)((g \circ f)^m(g(b_0))) = g(f((g \circ f)^m(g(b_0))))$. Therefore, $(a, f((g \circ f)^m(g(b_0)))) \in h$, and hence $a \in \text{dom}(h)$. It follows that $\text{dom}(h) = A$.

We now know that $h: A \rightarrow B$, and we need only show that h is a bijection. Let $a_1, a_2 \in A$ and suppose that $h(a_1) = h(a_2)$. We first show that either a_1 and a_2 are both B -originating or both a_1 and a_2 are both not B -originating. Without loss of generality, suppose that a_1 is B -originating and a_2 is not, so that $a_1 = g(h(a_1))$ and $h(a_2) = f(a_2)$. Since a_1 is B -originating, we may fix $b_0 \in B$ and $n \in \omega$ such that $b_0 \notin \text{ran}(f)$ and $a_1 = (g \circ f)^n(g(b_0))$. Notice that $(g \circ f)^n(g(b_0)) = a_1 = g(h(a_1)) = g(h(a_2)) = g(f(a_2)) = (g \circ f)(a_2)$. If $n = 0$, this implies that $g(b_0) = g(f(a_2))$, hence $f(a_2) = b_0$ (because g is injective), contrary to the fact that $b_0 \notin \text{ran}(f)$. Suppose that $n \neq 0$ and fix $m \in \omega$ with $S(m) = n$. We then have $(g \circ f)((g \circ f)^m(g(b_0))) = (g \circ f)^n(g(b_0)) = (g \circ f)(a_2)$, hence $(g \circ f)^m(g(b_0)) = a_2$ (because $g \circ f$ is injective), contrary to the fact that a_2 is not B -originating. Therefore, either a_1 and a_2 are both B -originating or both a_1 and a_2 are both not B -originating. If a_1 and a_2 are both not B -originating, this implies that $f(a_1) = f(a_2)$, hence $a_1 = a_2$ because f is injective. If a_1 and a_2 are both B -originating, we then have $a_1 = g(h(a_1)) = g(h(a_2)) = a_2$. It follows that h is injective.

We finally show that h is surjective. Fix $b \in B$. Suppose first that b is B -originating, and fix $b_0 \in B$ and $n \in \omega$ such that $b_0 \notin \text{ran}(f)$ and $b = (f \circ g)^n(b_0)$. We then have $g(b) = g((f \circ g)^n(b_0)) = (g \circ f)^n(g(b_0))$, hence $g(b) \in A$ is B -originating. It follows that $h(g(b)) = b$, so $b \in \text{ran}(h)$. Suppose now that b is not B -originating. We then must have $b \in \text{ran}(f)$, so we may fix $a \in A$ with $f(a) = b$. If a is B -originating, we may fix $b_0 \in B$ and $n \in \omega$ such that $b_0 \notin \text{ran}(f)$ and $a = (g \circ f)^n(g(b_0))$, and notice that $(f \circ g)^{S(n)}(b_0) = f((g \circ f)^n(g(b_0))) = f(a) = b$, contrary to the fact that b is not B -originating. Therefore, a is not B -originating, so $h(a) = f(a) = b$, and hence $b \in \text{ran}(h)$. It follows that h is surjective. \square

Definition 2.73. *Let A and B be sets. We write $A \prec B$ to mean that $A \preceq B$ and $A \not\approx B$.*

Theorem 2.74. *For any set A , we have $A \prec \mathcal{P}(A)$.*

Proof. First, define a function $f: A \rightarrow \mathcal{P}(A)$ by letting $f(a) = \{a\}$ for every $a \in A$. Notice that f is an injection, hence $A \preceq \mathcal{P}(A)$. We next show that $A \not\approx \mathcal{P}(A)$ by showing that there is no bijection $f: A \rightarrow \mathcal{P}(A)$. Suppose then that $f: A \rightarrow \mathcal{P}(A)$. Let $B = \{a \in A : a \notin f(a)\}$, and notice that $B \in \mathcal{P}(A)$. Suppose that $B \in \text{ran}(f)$, and fix $b \in A$ with $f(b) = B$. We then have $b \in f(b) \leftrightarrow b \in B \leftrightarrow b \notin f(b)$, a contradiction. It follows that $B \notin \text{ran}(f)$, hence f is not surjective. Therefore, $A \prec \mathcal{P}(A)$. \square

2.9.1 Countable Sets

Definition 2.75. *Let A be a set.*

1. A is countably infinite if $A \approx \omega$.
2. A is countable if A is either finite or countably infinite.
3. A is uncountable if A is not countable.

Proposition 2.76. *Let A be a set. The following are equivalent:*

1. A is countable.
2. $A \preceq \omega$.
3. There is a surjection $g: \omega \rightarrow A$.

Proof. □

2.9.2 General Powers

There is no reason to restrict to $n \in \omega$ in the above examples. In general, we want to define A^B to be the set of all functions from B to A . We can certainly make this definition, but it is the first instance where we really need to use Power Set.

Proposition 2.77. *Let A and B be sets. There is a unique set, denoted by A^B , such that for all f , we have $f \in A^B$ if and only if $f: B \rightarrow A$.*

Proof. Notice that if $f: B \rightarrow A$, then $f \subseteq B \times A$, hence $f \in \mathcal{P}(B \times A)$. Therefore, $A^B = \{f \in \mathcal{P}(B \times A) : f \text{ is a function, } \text{dom}(f) = B, \text{ and } \text{ran}(f) = A\}$. As usual, uniqueness follows from Extensionality. □

2.9.3 General Products

3 Doing Mathematics in Set Theory

3.1 The Basic Number Systems

3.2 Doing Logic In Set Theory

4 Well-Orderings, Ordinals, and Cardinals

4.1 Well-Orderings

The ability to do induction and make definitions by recursion on ω was essential to developing the basic properties of the natural numbers. With such success, we may wonder on which other kinds of structures we can do induction and recursion. Looking at the Step Induction Principle and Step Recursive Definitions on ω , it seems hard to generalize these ideas to anything more complicated than ω because by starting with zero and taking successors we can't get any further. However, the more general versions of induction and recursion which refer to the order on ω rather than just 0 and successors can be very fruitfully generalized to any well-ordering.

Proposition 4.1 (Induction on Well-Orderings). *Let $(W, <)$ be a well-ordering.*

1. Suppose that X is set and for all $z \in W$, if $y \in X$ for all $y < z$, then $z \in X$. We then have $W \subseteq X$.

2. For any formula $\varphi(z, \vec{p})$, we have the sentence

$$\forall \vec{p}((\forall z \in W)((\forall y < z)\varphi(y, \vec{p}) \rightarrow \varphi(z, \vec{p})) \rightarrow (\forall z \in W)\varphi(z, \vec{p}))$$

3. Suppose that \mathbf{C} is a class and for all $z \in W$, if $y \in \mathbf{C}$ for all $y < z$, then $z \in \mathbf{C}$. We then have $W \subseteq \mathbf{C}$.

Proof.

1. Suppose that $W \not\subseteq X$ so that $W \setminus X \neq \emptyset$. Since $(W, <)$ is a well-ordering, there exists $z \in W \setminus X$ such that for all $y \in W \setminus X$, either $z = y$ or $z < y$. Therefore, for all $y \in W$ with $y < z$, we have $y \in X$ (because $y \notin W \setminus X$). It follows from assumption that $z \in X$, contradicting the fact that $z \in W \setminus X$. Thus, it must be the case that $W \subseteq X$.

2. This follows from part 1 using Separation. Fix sets \vec{q} , and suppose that

$$(\forall z \in W)((\forall y < z)\varphi(y, \vec{q}) \rightarrow \varphi(z, \vec{q}))$$

Let $X = \{z \in W : \varphi(z, \vec{q})\}$. Suppose that $z \in W$ and $y \in X$ for all $y < z$. We then have $(\forall y < z)\varphi(y, \vec{q})$, hence $\varphi(z, \vec{q})$ by assumption, so $z \in X$. It follows from part 1 that $W \subseteq X$. Therefore, we have $(\forall z \in W)\varphi(z, \vec{q})$.

3. This is just a restatement of 2 using the language of classes. □

This is all well and good, but are there other interesting well-orderings other than ω (and every $n \in \omega$)? Well, any well-ordering has a smallest element. If there are any elements remaining, there must be a next smallest element. Again, if there are any elements remaining, there must be a next smallest element, and so on. Thus, any well-ordering begins with a piece that looks like ω .

However, we can build another “longer” well-ordering by taking ω , and adding a new element which is greater than every element of ω . This can be visualized by thinking of the set

$$A = \{1 - \frac{1}{n} \in \mathbb{R} : n \in \omega \setminus \{0\}\} \cup \{1\}.$$

It’s a simple exercise to check that A , ordered by inheritance from the usual order on \mathbb{R} , is a well-ordering. We can then add another new element which is greater than every element, and another and another and so on, to get a well-ordering that is a copy of ω with another copy of ω on top of the first. We can add a new element greater than all of these, and continue. These well-orderings “beyond” ω differ from ω (and all $n \in \omega$) in that they have points that are neither initial points nor immediate successors of other points.

Definition 4.2. Let $(W, <)$ be a well-ordering, and let $z \in W$.

1. If $z \leq y$ for all $y \in W$, we call z the *initial point* (such a z is easily seen to be unique).
2. If there exists $y \in W$ such that there is no $x \in W$ with $y < x < z$, we call z a *successor point*.
3. If z is neither an initial point nor a successor point, we call z a *limit point*.

A little thought will suggest that all well-orderings should be built up by starting at an initial point, taking successors (perhaps infinitely often), and then jumping to a limit point above everything previously. After all, if we already have an initial part that looks like ω , and we haven’t exhausted the well-ordering, then there must be a least element not accounted for, and this is the first limit point. If we still haven’t exhausted it, there is another least element, which is a successor, and perhaps another successor, and so on.

If this doesn't finish off the well-ordering, there is another least element not accounted for which will be the second limit point.

This idea makes it seem plausible that we can take any two well-orderings and compare them by running through this procedure until one of them runs out of elements. That is, if $(W_1, <_1)$ and $(W_2, <_2)$ are well-orderings, then either they are isomorphic, or one is isomorphic to an initial segment of the other. We now develop the tools to prove this result. We first show that we can make recursive definitions along well-orderings. The proof is basically the same as the proof of the Induction Principle on ω because the only important fact that allowed that argument to work was the property of the order $<$ on ω (not the fact that every element of ω was either an initial point or a successor point).

Definition 4.3. Let $(W, <)$ be a well-ordering, and let $z \in W$. We let $W(z) = \{y \in W : y < z\}$.

Definition 4.4. Let $(W, <)$ be a well-ordering. A set $I \subseteq W$ is called an initial segment of W if $I \neq W$ and whenever $x \in I$ and $y < x$, we have $y \in I$.

Proposition 4.5. Suppose that $(W, <)$ is a well-ordering and I is an initial segment of W . There exists $z \in W$ with $I = W(z)$.

Proof. Since I is an initial segment of W , we have $I \subseteq W$ and $I \neq W$. Hence, $W \setminus I \neq \emptyset$. Since $(W, <)$ is a well-ordering, there exists $z \in W \setminus I$ such that $z \leq y$ for all $y \in W \setminus I$. We claim that $I = W(z)$. If $y \in W(z)$, we then have $y \notin W \setminus I$ (because $y < z$), hence $y \in I$. Therefore, $W(z) \subseteq I$. Suppose that $y \in I$ and $y \notin W(z)$. We then have $y \geq z$, hence $z \in I$ because I is an initial segment, contradicting the fact that $z \in W \setminus I$. It follows that $I \subseteq W(z)$. Therefore, $I = W(z)$ by Extensionality. \square

Definition 4.6. Let $(W, <)$ be a well-ordering and let A be a set. We let

$$A^{<W} = \{f \in \mathcal{P}(W \times A) : f \text{ is a function and } f: W(z) \rightarrow A \text{ for some } z \in W\}$$

Theorem 4.7 (Recursive Definitions on Well-Orderings). Let $(W, <)$ be a well-ordering, let A be a set, and let $g: A^{<W} \rightarrow A$. There exists a unique function $f: W \rightarrow A$ such that $f(z) = g(f \upharpoonright W(z))$ for all $z \in W$.

Proof. We first prove existence. Call a set $Z \subseteq W \times A$ sufficient if for all $z \in W$ and all $q \in A^{W(z)}$ such that $(y, q(y)) \in Z$ for all $y < z$, we have $(z, g(q)) \in Z$. Notice that sufficient sets exist (since $W \times A$ is sufficient). Let

$$Y = \{(z, a) \in W \times A : (z, a) \in Z \text{ for every sufficient set } Z\}.$$

We first show that Y is sufficient. Suppose that $z \in W$, that $q \in A^{W(z)}$, and that $(y, q(y)) \in Y$ for all $y < z$. For any sufficient set Z , we have $(y, q(y)) \in Z$ for all $y < z$, so $(z, g(q)) \in Z$. Therefore, $(z, g(q)) \in Y$ for every sufficient set Z , so $(z, g(q)) \in Y$. It follows that Y is sufficient.

We next show that for all $z \in W$, there exists a unique $a \in A$ such that $(z, a) \in Y$. Let

$$X = \{z \in W : \text{there exists a unique } a \in A \text{ such that } (z, a) \in Y\}.$$

Suppose that $z \in W$ is such that $y \in X$ for all $y < z$. Let $q = Y \cap (W(z) \times A)$ and notice that $q \in A^{W(z)}$. Since $(y, q(y)) \in Y$ for all $y < z$ and Y is sufficient, it follows that $(z, g(q)) \in Y$. Fix $b \in A$ with $b \neq g(q)$. We then have that $Y \setminus \{(z, b)\}$ is sufficient (otherwise, there exists $p \in A^{W(z)}$ such that $(y, p(y)) \in Y$ for all $y < z$ and $g(p) = b$, but this implies that $p = q$ and hence $b = g(q)$), so by definition of Y it follows that $Y \subseteq Y \setminus \{(z, b)\}$. Hence, $(z, b) \notin Y$. Therefore, there exists a unique $a \in A$ such that $(z, a) \in Y$, so $z \in X$. By induction, we conclude that $X = W$, so for all $z \in W$, there exists a unique $a \in A$ such that $(z, a) \in Y$.

Let $f = Y$ and notice that $f: W \rightarrow A$ from above. Suppose that $z \in W$. Define $q \in A^{W(z)}$ by letting $q = Y \cap (W(z) \times A)$ and notice that $q = f \upharpoonright W(z)$. Since $(y, q(y)) \in Y$ for all $y < z$ and Y is sufficient, it follows that $(z, g(q)) \in Y$, so $f(z) = g(q) = g(f \upharpoonright W(z))$.

We now prove uniqueness. Suppose that $f_1, f_2: W \rightarrow A$ are such that:

1. $f_1(z) = g(f_1 \upharpoonright W(z))$ for all $z \in \omega$.
2. $f_2(z) = g(f_2 \upharpoonright W(z))$ for all $z \in \omega$.

Let $X = \{z \in W : f_1(z) = f_2(z)\}$. We prove by induction that $X = W$. Let $z \in W$ and suppose that $y \in X$ for all $y < z$. We then have that $f_1 \upharpoonright W(z) = f_2 \upharpoonright W(z)$, hence

$$f_1(z) = g(f_1 \upharpoonright W(z)) = g(f_2 \upharpoonright W(z)) = f_2(z)$$

hence $z \in X$. It follows by induction that $X = W$, so $f_1(z) = f_2(z)$ for all $z \in W$. □

Definition 4.8. Let $(W_1, <_1)$ and $(W_2, <_2)$ be well-orderings.

1. A function $f: W_1 \rightarrow W_2$ is order-preserving if whenever $x, y \in W_1$ and $x <_1 y$, we have $f(x) <_2 f(y)$.
2. A function $f: W_1 \rightarrow W_2$ is an isomorphism if it is bijective and order-preserving.
3. If W_1 and W_2 are isomorphic, we write $W_1 \cong W_2$.

Proposition 4.9. Suppose that $(W, <)$ is a well-ordering and $f: W \rightarrow W$ is order-preserving. We then have $f(z) \geq z$ for all $z \in W$.

Proof. We prove the result by induction on W . Suppose that $z \in W$ and $f(y) \geq y$ for all $y < z$. Suppose instead that $f(z) < z$, and let $x = f(z)$. Since f is order-preserving and $x < z$, it follows that $f(x) < f(z) = x$, contradicting the fact that $f(y) \geq y$ for all $y < z$. Therefore, $f(z) \geq z$. The result follows by induction. □

Corollary 4.10.

1. If $(W, <)$ is a well-ordering and $z \in W$, then $W \not\cong W(z)$.
2. If $(W, <)$ is a well-ordering, then its only automorphism is the identity.
3. If $(W_1, <_1)$ and $(W_2, <_2)$ are well-orderings, and $W_1 \cong W_2$, then the isomorphism from W_1 to W_2 is unique.

Proof.

1. Suppose that $W \cong W(z)$ for some $z \in W$ and let $f: W \rightarrow W(z)$ be a witnessing isomorphism. Then $f: W \rightarrow W$ is order-preserving and $f(z) < z$ (because $f(z) \in W(z)$), contrary to Proposition 4.9.
2. Suppose that $f: W \rightarrow W$ is an automorphism of W which is not the identity. By Proposition 4.9, we have $f(z) \geq z$ for all $z \in W$. Suppose that $z \in W$ and let $y = f(z)$. Since $f^{-1}: W \rightarrow W$ is also an automorphism of W , Proposition 4.9 implies that $f^{-1}(y) \geq y$, hence $z \geq f(z)$. Combining this with the above mentioned fact that $f(z) \geq z$, it follows that $z = f(z)$. Therefore, f is the identity.
3. Suppose that $f: W_1 \rightarrow W_2$ and $g: W_1 \rightarrow W_2$ are both isomorphisms. We then have that $g^{-1}: W_2 \rightarrow W_1$ is an isomorphism, hence $g^{-1} \circ f: W_1 \rightarrow W_1$ is an automorphism. Hence, by part b, we may conclude that $g^{-1} \circ f$ is the identity on W_1 . It follows that $f = g$.

□

Theorem 4.11. Let $(W_1, <_1)$ and $(W_2, <_2)$ be well-orderings. Exactly one of the following holds.

1. $W_1 \cong W_2$.
2. There exists $z \in W_2$ such that $W_1 \cong W_2(z)$.

3. There exists $z \in W_1$ such that $W_1(z) \cong W_2$.

In each of the above cases, the isomorphism and the z (if appropriate) are unique.

Proof. We first prove that one of the three options holds. Fix a set a such that $a \notin W_1 \cup W_2$ (such an a exists by Proposition 2.4). Our goal is to define a function $f: W_1 \rightarrow W_2 \cup \{a\}$ recursively. Define $g: (W_2 \cup \{a\})^{<W_1} \rightarrow W_2 \cup \{a\}$ as follows. Let $q \in (W_2 \cup \{a\})^{<W_1}$ and fix $z \in W_1$ such that $q: W_1(z) \rightarrow W_2 \cup \{a\}$. If $a \in \text{ran}(q)$ or $\text{ran}(q) = W_2$, let $g(q) = a$. Otherwise $\text{ran}(q)$ is a proper subset of W_2 , and we let $g(q)$ be the $<_2$ -least element of $W_2 \setminus \text{ran}(q)$. By Theorem 4.7, there is a unique $f: W_1 \rightarrow W_2 \cup \{a\}$ such that $f(z) = g(f \upharpoonright W_1(z))$ for all $z \in W_1$.

Suppose first that $a \notin \text{ran}(f)$ so that $f: W_1 \rightarrow W_2$. We begin by showing that $\text{ran}(f \upharpoonright W_1(z))$ is an initial segment of W_2 for all $z \in W_1$ by induction. Suppose that $z \in W_1$ and $\text{ran}(f \upharpoonright W_1(y))$ is an initial segment of W_2 for all $y < z$. If z is the initial point of W_1 , then $\text{ran}(f \upharpoonright W_1(z)) = \emptyset$ is certainly an initial segment of W_2 . Suppose that z is a successor point of W_1 , and let $y \in W_1$ be such that there is no $x \in W_1$ with $y < x < z$. By induction, we know that $\text{ran}(f \upharpoonright W_1(y))$ is an initial segment of W_2 . Since $f(y) = g(f \upharpoonright W_1(y))$ is the $<_2$ -least element of $W_2 \setminus (f \upharpoonright W_1(y))$, it follows that $\text{ran}(f \upharpoonright W_1(z)) = \text{ran}(f \upharpoonright W_1(y)) \cup \{f(y)\}$ is an initial segment of W_2 . Suppose finally that z is a limit point of W_1 . It then follows that $\text{ran}(f \upharpoonright W_1(z)) = \bigcup_{y < z} \text{ran}(f \upharpoonright W_1(y))$. Since every element of the union is an initial segment of W_2 , it follows that $\text{ran}(f \upharpoonright W_1(z))$ is an initial segment of W_2 (note that it can't equal W_2 because $f(z) \neq a$).

Therefore, $\text{ran}(f \upharpoonright W_1(z))$ is an initial segment of W_2 for all $z \in W_1$ by induction. It follows that for all $y, z \in W_1$ with $y < z$, we have $f(y) < f(z)$ (because $\text{ran}(f \upharpoonright W_1(z))$ is an initial segment of W_2 and $f(y) \in \text{ran}(f \upharpoonright W_1(z))$), so f is order-preserving. This implies that f is an injection, so if $\text{ran}(f) = W_2$, we have $W_1 \cong W_2$. Otherwise, $\text{ran}(f)$ is an initial segment of W_2 , so by Proposition 4.5 there is a $z \in W_2$ such that $W_1 \cong W_2(z)$.

Suppose now that $a \in \text{ran}(f)$. Let $z \in W_1$ be the $<_1$ -least element of W_1 such that $f(z) = a$. It then follows that $f \upharpoonright W_1(z): W_1(z) \rightarrow W_2$ is order-preserving by induction as above. Also, we must have $\text{ran}(f \upharpoonright W_1(z)) = W_2$ because $f(z) = a$. Therefore, $f \upharpoonright W_1(z): W_1(z) \rightarrow W_2$ is an isomorphism. This completes the proof that one of the above 3 cases must hold.

The uniqueness of the case, the isomorphism, and the z (if appropriate), all follow from Corollary 4.10 \square

With this result in hand, we now know that any well-ordering is uniquely determined by its “length”. The next goal is to find a nice system of representatives for the isomorphism classes of well-orderings. For that, we need to generalize the ideas that went into the construction of the natural numbers.

4.2 Ordinals

Our definition of the natural numbers had the advantage that the ordering was given by the membership relation \in . This feature allowed us to define successors easily and to think of a natural number n as the set of all natural numbers less than n . We now seek to continue this progression to measure well-orderings longer than ω . The idea is to define successors as in the case of the natural numbers, but now to take unions to achieve limit points.

The key property of ω (and each $n \in \omega$) that we want to use in our definition of ordinals is the fact that \in well-orders ω (and each $n \in \omega$). We need one more condition to ensure that there are no “holes” or “gaps” in the set. For example, \in well-orders the set $\{0, 2, 3, 5\}$, but we don't want to consider it as an ordinal because it skipped over 1 and 4. We therefore make the following definition.

Definition 4.12. A set z is transitive if whenever x and y are sets such that $x \in y$ and $y \in z$, we have $x \in z$.

Definition 4.13. Let z be a set. We define a relation \in_z on z by setting $\in_z = \{(x, y) \in z \times z : x \in y\}$.

Definition 4.14. An ordinal is a set α which is transitive and well-ordered by \in_α .

Our hard work developing the natural numbers gives us one interesting example of an ordinal.

Proposition 4.15. *ω is an ordinal.*

Proof. Proposition 2.41 says that ω is transitive, and Theorem 2.48 says that ω is well-ordered by $< = \in_\omega$. \square

Proposition 4.16. *If α is an ordinal and $\beta \in \alpha$, then β is an ordinal.*

Proof. We first show that β is transitive. Let x and y be sets with $x \in y$ and $y \in \beta$. Since $y \in \beta$, $\beta \in \alpha$, and α is transitive, it follows that $y \in \alpha$. Since $x \in y$ and $y \in \alpha$, it follows that $x \in \alpha$. Now since $x, y, \beta \in \alpha$, $x \in y$, $y \in \beta$, and \in_α is transitive on α , we may conclude that $x \in \beta$. Therefore, β is transitive.

Notice that $\beta \subseteq \alpha$ because $\beta \in \alpha$ and α is transitive. Therefore, \in_β is the restriction of \in_α to the subset $\beta \subseteq \alpha$. Since \in_α is a well-ordering on α , it follows that \in_β is a well-ordering on β . Hence, β is an ordinal. \square

Corollary 4.17. *Every $n \in \omega$ is an ordinal.*

Lemma 4.18. *If α is an ordinal, then $\alpha \notin \alpha$.*

Proof. Suppose that α is an ordinal and $\alpha \in \alpha$. Since $\alpha \in \alpha$, it follows that \in_α is not asymmetric on α , contradicting the fact that \in_α is a well-ordering on α . \square

Proposition 4.19. *If α is an ordinal, then $S(\alpha)$ is an ordinal.*

Proof. We first show that $S(\alpha)$ is transitive. Suppose that $x \in y \in S(\alpha)$. Since $y \in S(\alpha) = \alpha \cup \{\alpha\}$, either $y \in \alpha$ or $y = \alpha$. Suppose first that $y \in \alpha$. We then have $x \in y \in \alpha$, so $x \in \alpha$ because α is transitive. Hence, $x \in S(\alpha)$. Suppose now that $y = \alpha$. We then have $x \in \alpha$ because $x \in y$, so $x \in S(\alpha)$.

We next show that $\in_{S(\alpha)}$ is transitive on $S(\alpha)$. Let $x, y, z \in S(\alpha)$ with $x \in y \in z$. Since $z \in S(\alpha)$, either $z \in \alpha$ or $z = \alpha$. Suppose first that $z \in \alpha$. We then have $y \in \alpha$ (since $y \in z \in \alpha$ and α is transitive), and hence $x \in \alpha$ (since $x \in y \in \alpha$ and α is transitive). Thus, $x, y, z \in \alpha$, so we may conclude that $x \in z$ using the fact that \in_α is transitive on α . Suppose now that $z = \alpha$. We then have $x \in \alpha = z$ because $x \in y \in \alpha$ and α is transitive.

We next show that $\in_{S(\alpha)}$ is asymmetric on $S(\alpha)$. Let $x \in S(\alpha)$. If $x \in \alpha$, then $x \notin x$ because \in_α is asymmetric on α . If $x = \alpha$, then $x \notin x$ by Lemma 4.18.

We now show that $\in_{S(\alpha)}$ is connected on $S(\alpha)$. Let $x, y \in S(\alpha)$. If $x \in \alpha$ and $y \in \alpha$, then either $x \in y$, $x = y$, or $y \in x$ because \in_α is connected on α . If $x = \alpha$ and $y = \alpha$, we clearly have $x = y$. Otherwise, one of x, y equals α , and the other is an element of α , in which case we're done.

Finally, suppose that $X \subseteq S(\alpha)$ and $X \neq \emptyset$. If $X \cap \alpha = \emptyset$, then we must have $X = \{\alpha\}$, in which case X clearly has a $\in_{S(\alpha)}$ -least element. Suppose that $X \cap \alpha \neq \emptyset$. Since $X \cap \alpha \subseteq \alpha$ is nonempty and \in_α is a well-ordering on α , there exists a \in_α -least element β in $X \cap \alpha$. For any $\gamma \in X$, either $\gamma \in \alpha$ in which case we have either $\beta = \gamma$ or $\beta \in \gamma$ by choice of β , or $\gamma = \alpha$ in which case $\beta \in \gamma$ (because $\beta \in \alpha$). Therefore, X has a $\in_{S(\alpha)}$ -least element. \square

Proposition 4.20. *Suppose that α and β are ordinals. We then have $\alpha \subseteq \beta$ if and only if either $\alpha = \beta$ or $\alpha \in \beta$.*

Proof. (\Leftarrow) If $\alpha = \beta$, then clearly $\alpha \subseteq \beta$ and if $\alpha \in \beta$ we can use the fact that β is transitive to conclude that $\alpha \subseteq \beta$.

(\Rightarrow) Suppose that $\alpha \subseteq \beta$ and $\alpha \neq \beta$. Notice that $\beta \setminus \alpha$ is a nonempty subset of β , so there exists a \in_β -least element of $\beta \setminus \alpha$, call it z . We show that $\alpha = z$, hence $\alpha \in \beta$. We first show that $z \subseteq \alpha$. Let $x \in z$. Since $z \in \beta$ and β is transitive, we have $x \in \beta$. Since $x \in z$, we can not have $x \in \beta \setminus \alpha$ by choice of z , so $x \in \alpha$. Thus, $z \subseteq \alpha$. We next show that $\alpha \subseteq z$. Let $x \in \alpha$. Since $\alpha \subseteq \beta$, we have $x \in \beta$. Using the fact that $x, z \in \beta$ and \in_β is connected on β , we know that either $x \in z$, $x = z$, or $z \in x$. We can not have $x = z$ because $x \in \alpha$ and $z \in \beta \setminus \alpha$. Also, we can not have $z \in x$, because if $z \in x$ we can also conclude that $z \in \alpha$ (because $z \in x \in \alpha$ and α is transitive), contradicting the fact that $z \in \beta \setminus \alpha$. Thus, $\alpha \subseteq z$. It follows that $z = \alpha$ (by Extensionality), so $\alpha \in \beta$. \square

Proposition 4.21. *Suppose that α and β are ordinals. Exactly one of $\alpha \in \beta$, $\alpha = \beta$, or $\beta \in \alpha$ holds.*

Proof. We first show that at least one $\alpha \in \beta$, $\alpha = \beta$, $\beta \in \alpha$ holds. We first claim that $\alpha \cap \beta$ is an ordinal. If $x \in y \in \alpha \cap \beta$, then $x \in y \in \alpha$ and $x \in y \in \beta$, so $x \in \alpha$ and $x \in \beta$ (because α and β are transitive), and hence $x \in \alpha \cap \beta$. Thus, $\alpha \cap \beta$ is transitive. Notice that $\in_{\alpha \cap \beta}$ is the restriction of \in_α to the subset $\alpha \cap \beta \subseteq \alpha$. Since \in_α is a well-ordering on α , it follows that $\in_{\alpha \cap \beta}$ is a well-ordering on $\alpha \cap \beta$. Hence, $\alpha \cap \beta$ is an ordinal.

Now we have $\alpha \cap \beta \subseteq \alpha$ and $\alpha \cap \beta \subseteq \beta$. If $\alpha \cap \beta \neq \alpha$ and $\alpha \cap \beta \neq \beta$, then $\alpha \cap \beta \in \alpha$ and $\alpha \cap \beta \in \beta$ by Proposition 4.20, hence $\alpha \cap \beta \in \alpha \cap \beta$, contrary to Lemma 4.18. Therefore, either $\alpha \cap \beta = \alpha$ or $\alpha \cap \beta = \beta$. If $\alpha \cap \beta = \alpha$, we then have $\alpha \subseteq \beta$, hence either $\alpha = \beta$ or $\alpha \in \beta$ by Proposition 4.20. Similarly, if $\alpha \cap \beta = \beta$, we then have $\beta \subseteq \alpha$, hence either $\beta = \alpha$ or $\beta \in \alpha$ by Proposition 4.20. Thus, in any case, at least one $\alpha \in \beta$, $\alpha = \beta$, or $\beta \in \alpha$ holds.

We finish by showing that exactly one of $\alpha \in \beta$, $\alpha = \beta$, or $\beta \in \alpha$ holds. If $\alpha \in \beta$ and $\alpha = \beta$, then $\alpha \in \alpha$, contrary to Lemma 4.18. Similarly, if $\alpha = \beta$ and $\beta \in \alpha$, then $\beta \in \beta$, contrary to Lemma 4.18. Finally, if $\alpha \in \beta$ and $\beta \in \alpha$, then $\alpha \in \alpha$ (because α is transitive), contrary to Lemma 4.18. \square

Definition 4.22. *If α and β are ordinals, we write $\alpha < \beta$ to mean that $\alpha \in \beta$.*

Proposition 4.23. *Suppose that α and β are ordinals. If $\alpha \cong \beta$ as well-orderings, then $\alpha = \beta$.*

Proof. If $\alpha \neq \beta$, then either $\alpha < \beta$ or $\beta < \alpha$ by Proposition 4.21. Suppose without loss of generality that $\beta < \alpha$. We then have that the well-ordering β is an initial segments of the well-ordering α (in the notation for well-orderings, we have $\beta = \alpha(\beta)$), hence $\alpha \not\cong \beta$ by Corollary 4.10. \square

By the above results, it seems that we are in a position to say that $<$ is a linear ordering on the collection of all ordinals. However, there is a small problem here. We do not know that the class of all ordinals is a set. In fact, we will see below that the collection of all ordinals is a proper class.

Definition 4.24. **ORD** *is the class of all ordinals.*

We first establish that nonempty sets of ordinals have least elements.

Proposition 4.25. *If A is a nonempty subset of **ORD**, then A has a least element. Furthermore the least element is given by $\bigcap A$.*

Proof. Since $A \neq \emptyset$, we may fix an ordinal $\alpha \in A$. If $A \cap \alpha = \emptyset$, then for any $\beta \in A$, we can not have $\beta \in \alpha$, hence either $\alpha = \beta$ or $\alpha \in \beta$ by Proposition 4.21. Suppose that $A \cap \alpha \neq \emptyset$. Since $A \cap \alpha \subseteq \alpha$ is nonempty, it has an \in_α -least element, call it δ . Let $\beta \in A$ and notice that β is an ordinal. By Proposition 4.21, either $\beta \in \alpha$, $\beta = \alpha$, or $\alpha \in \beta$. If $\beta \in \alpha$, then $\beta \in A \cap \alpha$, so either $\delta = \beta$ or $\delta \in \beta$ by choice of δ . If $\beta = \alpha$, then $\delta \in \beta$ because $\delta \in \alpha$. If $\alpha \in \beta$, we then have $\delta \in \alpha \in \beta$, so $\delta \in \beta$ because β is transitive. It follows that δ is the least element of A .

Therefore, we know that A has a least element, call it δ . Since $\delta \in A$, we certainly have $\bigcap A \subseteq \delta$. For all $\alpha \in A$, we then have either $\delta = \alpha$ or $\delta \in \alpha$, hence $\delta \subseteq \alpha$ by Proposition 4.20. Therefore, $\delta \subseteq \bigcap A$. It follows that $\delta = \bigcap A$. \square

Proposition 4.26. *If A is a subset of **ORD**, then $\bigcup A$ is an ordinal. Furthermore, we have $\bigcup A = \sup A$, i.e. $\alpha \leq \bigcup A$ for all $\alpha \in A$ and $\bigcup A \leq \beta$ whenever β is an ordinal with $\beta \geq \alpha$ for all $\alpha \in A$.*

Proof. We first show that $\bigcup A$ is transitive. Suppose that $x \in y \in \bigcup A$. Since $y \in \bigcup A$, there exists $\alpha \in A$, necessarily an ordinal, such that $y \in \alpha \in A$. Since α is transitive and $x \in y \in \alpha$, we can conclude that $x \in \alpha$. It follows that $x \in \bigcup A$. Hence, $\bigcup A$ is transitive.

We next show that $\in_{\bigcup A}$ is transitive on $\bigcup A$. Let $x, y, z \in \bigcup A$ with $x \in y \in z$. Since $z \in \bigcup A$, there exists $\alpha \in A$, necessarily an ordinal, such that $z \in \alpha \in A$. Since $z \in \alpha$ and α is an ordinal, we may use Proposition 4.16 to conclude that z is an ordinal. Thus, z is transitive, so we may use the fact that $x \in y \in z$ to conclude that $x \in z$.

We next show that $\in_{\cup A}$ is asymmetric on $\cup A$. Let $x \in \cup A$ and fix $\alpha \in A$, necessarily an ordinal, such that $x \in \alpha \in A$. Using Proposition 4.16 again, it follows that x is an ordinal, hence $x \notin x$ by Lemma 4.18.

We now show that $\in_{\cup A}$ is connected on $\cup A$. Let $x, y \in \cup A$. Fix $\alpha, \beta \in A$, necessarily ordinals, such that $x \in \alpha \in A$ and $y \in \beta \in A$. Again, using Proposition 4.16, we may conclude that x and y are ordinals, hence either $x \in y$, $x = y$, or $y \in x$ by Proposition 4.21.

Finally, suppose that $X \subseteq \cup A$ and $X \neq \emptyset$. Notice that for any $y \in X$, there exists $\alpha \in A$, necessarily an ordinal, such that $y \in \alpha \in A$, and hence y is an ordinal by Proposition 4.21. Therefore, X is a nonempty subset of **ORD**, so by Proposition 4.25 we may conclude that X has a least element (with respect to $\in_{\cup A}$).

We now show that $\cup A = \sup A$. Suppose that $\alpha \in A$. For any $\beta \in \alpha$, we have $\beta \in \alpha \in A$, hence $\beta \in \cup A$. It follows that $\alpha \subseteq \cup A$, hence $\alpha \leq \cup A$ by Proposition 4.20. Thus, $\cup A$ is an upper bound for A . Suppose that γ is an upper bound for A , i.e. γ is an ordinal and $\alpha \leq \gamma$ for all $\alpha \in A$. For any $\beta \in \cup A$, we may fix $\alpha \in A$ such that $\beta \in \alpha$ and notice that $\beta \in \alpha \subseteq \gamma$, so $\beta \in \gamma$. It follows that $\cup A \subseteq \gamma$, hence $\cup A \leq \gamma$ by Proposition 4.20. Therefore, $\cup A = \sup A$. \square

Proposition 4.27. ***ORD** is a proper class.*

Proof. Suppose that **ORD** is a set, so that there is a set O such that α is an ordinal if and only $\alpha \in O$. In this case, O is a transitive set (by Proposition 4.16) which is well-ordered by \in_O (transitivity follows from the fact that ordinals are transitive sets, asymmetry follows from Lemma 4.18, connectedness follows from Proposition 4.21, and the fact that every nonempty subset has a least element is given by Proposition 4.25). Therefore, O is an ordinal and so it follows that $O \in O$, contrary to Lemma 4.18. Hence, **ORD** is not a set. \square

Since **ORD** is a proper class, there are subclasses of **ORD** which are not subsets of **ORD**. We therefore extend Proposition 4.25 to the case of nonempty subclasses of **ORD**. The idea is that if we fix an $\alpha \in \mathbf{C}$, then $\alpha \cap \mathbf{C}$ becomes a set of ordinals, so we can apply the above result.

Proposition 4.28. *If \mathbf{C} is a nonempty subclass of **ORD**, then \mathbf{C} has a least element.*

Proof. Since $\mathbf{C} \neq \emptyset$, we may fix an ordinal $\alpha \in \mathbf{C}$. If $\mathbf{C} \cap \alpha = \emptyset$, then for any $\beta \in \mathbf{C}$, we can not have $\beta \in \alpha$, hence either $\alpha = \beta$ or $\alpha \in \beta$ by Proposition 4.21. Suppose that $\mathbf{C} \cap \alpha \neq \emptyset$. In this case, $\mathbf{C} \cap \alpha$ is a nonempty set of ordinals by Separation, hence $\mathbf{C} \cap \alpha$ has a least element δ by Proposition 4.25. It now follows easily that δ is the least element of \mathbf{C} . \square

Proposition 4.29 (Induction on **ORD**). *Suppose that $\mathbf{C} \subseteq \mathbf{ORD}$ and that for all ordinals α , if $\beta \in \mathbf{C}$ for all $\beta < \alpha$, then $\alpha \in \mathbf{C}$. We then have $\mathbf{C} = \mathbf{ORD}$.*

Proof. Suppose that $\mathbf{C} \subsetneq \mathbf{ORD}$. Let $\mathbf{B} = \mathbf{ORD} \setminus \mathbf{C}$ and notice that \mathbf{B} is a nonempty class of ordinals. By Proposition 4.28, it follows that \mathbf{B} has a least element, call it α . For all $\beta < \alpha$, we then have $\beta \notin \mathbf{B}$, hence $\beta \in \mathbf{C}$. By assumption, this implies that $\alpha \in \mathbf{C}$, a contradiction. It follows that $\mathbf{C} = \mathbf{ORD}$. \square

This gives a way to do “strong induction” on the ordinals, but there is a slightly more basic version. We can’t get around looking at many previous values at limit ordinals, but we can by with just looking at the previous ordinal in the case of successors.

Proposition 4.30 (Step/Limit Induction on **ORD**). *Suppose that $\mathbf{C} \subseteq \mathbf{ORD}$ and that*

1. $0 \in \mathbf{C}$.
2. Whenever $\alpha \in \mathbf{C}$, we have $S(\alpha) \in \mathbf{C}$.
3. Whenever α is a limit ordinal and $\beta \in \mathbf{C}$ for all $\beta < \alpha$, we have $\alpha \in \mathbf{C}$.

We then have $\mathbf{C} = \mathbf{ORD}$.

Proof. Suppose that $\mathbf{C} \subsetneq \mathbf{ORD}$. Let $\mathbf{B} = \mathbf{ORD} \setminus \mathbf{C}$ and notice that \mathbf{B} is a nonempty class of ordinals. By Proposition 4.28, it follows that \mathbf{B} has a least element, call it α . We can't have $\alpha = 0$ because $0 \in \mathbf{C}$. Also, it is not possible that α is a successor, say $\alpha = S(\beta)$, because if so, then $\beta \notin \mathbf{B}$ (because $\beta < \alpha$), so $\beta \in \mathbf{C}$, hence $\alpha = S(\beta) \in \mathbf{C}$. Finally, suppose that α is a limit. Then for all $\beta < \alpha$, we have $\beta \notin \mathbf{B}$, hence $\beta \in \mathbf{C}$. By assumption, this implies that $\alpha \in \mathbf{C}$, a contradiction. It follows that $\mathbf{C} = \mathbf{ORD}$. \square

Theorem 4.31 (Recursive Definitions on \mathbf{ORD}). *Let $\mathbf{G}: \mathbf{V} \rightarrow \mathbf{V}$ be a class function. There exists a unique class function $\mathbf{F}: \mathbf{ORD} \rightarrow \mathbf{V}$ such that $\mathbf{F}(\alpha) = \mathbf{G}(\mathbf{F} \upharpoonright \alpha)$ for all $\alpha \in \mathbf{ORD}$.*

Theorem 4.32 (Recursive Definitions with Parameters on \mathbf{ORD}). *Let \mathbf{P} be a class and let $\mathbf{G}: \mathbf{P} \times \mathbf{V} \rightarrow \mathbf{V}$ be a class function. There exists a unique class function $\mathbf{F}: \mathbf{P} \times \mathbf{ORD} \rightarrow \mathbf{V}$ such that $\mathbf{F}(p, \alpha) = \mathbf{G}(\mathbf{F}_p \upharpoonright \alpha)$ for all $p \in \mathbf{P}$ and all $\alpha \in \mathbf{ORD}$.*

Theorem 4.33. *Let $(W, <)$ be a well-ordering. There exists a unique ordinal α such that $W \cong \alpha$.*

Proof. Fix a set a such that $a \notin W$ (such an a exists by Proposition 2.4). We define a class function $\mathbf{F}: \mathbf{ORD} \rightarrow W \cup \{a\}$ recursively as follows. If $a \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$ or $\text{ran}(\mathbf{F} \upharpoonright \alpha) = W$, let $\mathbf{F}(\alpha) = a$. Otherwise, $\text{ran}(\mathbf{F} \upharpoonright \alpha) \subsetneq W$, and we let $\mathbf{F}(\alpha)$ be the least element of $W \setminus \text{ran}(\mathbf{F} \upharpoonright \alpha)$.

Since \mathbf{ORD} is a proper class, it follows from Proposition 2.52 that \mathbf{F} is not injective. From this it follows that $a \in \text{ran}(\mathbf{F})$ (otherwise, a simple inductive proof gives that \mathbf{F} would have to be injective). Let α be the least ordinal such that $\mathbf{F}(\alpha) = a$. Now it is straightforward to prove (along the lines of the proof of Theorem 4.11) that $\mathbf{F} \upharpoonright \alpha: \alpha \rightarrow W$ is an isomorphism.

Uniqueness follows from Proposition 4.23 \square

Definition 4.34. *Let $(W, <)$ be a well-ordering. The unique ordinal α such that $W \cong \alpha$ is called the order-type of $(W, <)$.*

4.3 Arithmetic on Ordinals

Definition 4.35. *We define ordinal addition (that is a class function $+: \mathbf{ORD} \times \mathbf{ORD} \rightarrow \mathbf{ORD}$) recursively as follows.*

1. $\alpha + 0 = \alpha$.
2. $\alpha + S(\beta) = S(\alpha + \beta)$.
3. $\alpha + \beta = \bigcup \{\alpha + \gamma : \gamma < \beta\}$ if β is a limit ordinal.

Similarly, we define ordinal multiplication recursively as follows.

1. $\alpha \cdot 0 = 0$.
2. $\alpha \cdot S(\beta) = \alpha \cdot \beta + \alpha$.
3. $\alpha \cdot \beta = \bigcup \{\alpha \cdot \gamma : \gamma < \beta\}$ if β is a limit ordinal.

Finally, we define ordinal exponentiation recursively as follows.

1. $\alpha^0 = 1$.
2. $\alpha^{S(\beta)} = \alpha^\beta \cdot \alpha$.
3. $\alpha^\beta = \bigcup \{\alpha^\gamma : \gamma < \beta\}$ if β is a limit ordinal.

Proposition 4.36. *Let α , β , and γ be ordinals. If $\beta \leq \gamma$, then $\alpha + \beta \leq \alpha + \gamma$.*

Proof. Fix ordinals α and β . We prove by induction on γ that if $\beta \leq \gamma$, then $\alpha + \beta \leq \alpha + \gamma$. If $\gamma = \beta$, this is trivial. Suppose that $\beta \leq \gamma$ and we know the result for γ . We then have

$$\begin{aligned}\alpha + \beta &\leq \alpha + \gamma \\ &< S(\alpha + \gamma) \\ &= \alpha + S(\gamma)\end{aligned}$$

Suppose now that $\gamma > \beta$ is a limit ordinal. We then have

$$\begin{aligned}\alpha + \beta &\leq \bigcup \{\alpha + \delta : \delta < \gamma\} && \text{(since } \beta < \gamma\text{)} \\ &= \alpha + \gamma\end{aligned}$$

□

Proposition 4.37. *Let α , β , and γ be ordinals. We have $\beta < \gamma$ if and only if $\alpha + \beta < \alpha + \gamma$.*

Proof. Notice first that

$$\alpha + \beta < S(\alpha + \beta) = \alpha + S(\beta)$$

Now for any ordinal $\gamma > \beta$, we have $S(\beta) \leq \gamma$, hence

$$\alpha + \beta < \alpha + S(\beta) \leq \alpha + \gamma$$

□

Proposition 4.38. *Let α and β be ordinals. If β is a limit ordinal, then $\alpha + \beta$ is a limit ordinal.*

Proof. Since β is a limit ordinal, we have

$$\alpha + \beta = \bigcup \{\alpha + \gamma : \gamma < \beta\}$$

Suppose now that $\delta < \alpha + \beta$, and fix an ordinal $\gamma < \beta$ such that $\delta < \alpha + \gamma$. We then have that $S(\gamma) < \beta$ because β is a limit ordinal, hence

$$\begin{aligned}S(\delta) &< S(\alpha + \gamma) \\ &= \alpha + S(\gamma) \\ &\leq \alpha + \beta\end{aligned}$$

It follows that $\alpha + \beta$ is a limit ordinal.

□

Proposition 4.39. *Let α , β , and γ be ordinals. We have $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.*

Proof. Fix ordinals α and β . We prove that $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ for all ordinals γ by induction. Suppose first that $\gamma = 0$. We then have

$$\begin{aligned}(\alpha + \beta) + 0 &= \alpha + \beta \\ &= \alpha + (\beta + 0)\end{aligned}$$

Suppose now that γ is an ordinal and we know that $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$. We then have

$$\begin{aligned}(\alpha + \beta) + S(\gamma) &= S((\alpha + \beta) + \gamma) \\ &= S(\alpha + (\beta + \gamma)) && \text{(by induction)} \\ &= \alpha + S(\beta + \gamma) \\ &= \alpha + (\beta + S(\gamma))\end{aligned}$$

Suppose now that γ is a limit ordinal and we know that $(\alpha + \beta) + \delta = \alpha + (\beta + \delta)$ for all $\delta < \gamma$. We then have

$$\begin{aligned} (\alpha + \beta) + \gamma &= \bigcup\{(\alpha + \beta) + \delta : \delta < \gamma\} \\ &= \bigcup\{\alpha + (\beta + \delta) : \delta < \gamma\} \\ &= \bigcup\{\alpha + \varepsilon : \varepsilon < \beta + \gamma\} \\ &= \alpha + (\beta + \gamma) \end{aligned}$$

where the last line follows because $\beta + \gamma$ is a limit ordinal. \square

4.4 Cardinals

Definition 4.40. A cardinal is an ordinal α such that $\alpha \not\approx \beta$ for any $\beta < \alpha$.

Proposition 4.41. Every $n \in \omega$ is a cardinal, and ω is a cardinal.

Proposition 4.42. Every infinite cardinal is a limit ordinal.

Proposition 4.43. Let A be a set. There is an ordinal α such that $\alpha \not\leq A$.

Proof. Let $\mathcal{F} = \{(B, R) \in \mathcal{P}(A) \times \mathcal{P}(A \times A) : R \text{ is a well-ordering on } B\}$. By Collection and Separation, $A = \{\text{order-type}(B, R) : (B, R) \in \mathcal{F}\}$ is a set of ordinals. Let α be an ordinal such that $\alpha > \bigcup A$ (such an α exists because **ORD** is a proper class). Notice that $\alpha \not\leq A$ because if $f: \alpha \rightarrow A$ were an injection, we could let $B = \text{ran}(f)$ and let R be the well-ordering on B obtained by transferring the ordering of α . We would then have $\alpha \in A$ since $(B, R) \in \mathcal{F}$ and (B, R) has order-type α , a contradiction. It follows that $\alpha \not\leq A$. \square

Definition 4.44. Let A be a set. The least ordinal α such that $\alpha \not\leq A$ is called the Hartogs number of A , and is denoted by $H(A)$.

Proposition 4.45. $H(A)$ is a cardinal for every set A .

Proof. Let A be a set and let $\alpha = H(A)$. Suppose that $\beta < \alpha$ and $\alpha \approx \beta$. Let $f: \alpha \rightarrow \beta$ be a bijection. Since $\beta < \alpha = H(A)$, there exists an injection $g: \beta \rightarrow A$. We then have that $g \circ f: \alpha \rightarrow A$ is an injection, contrary to the fact that $\alpha \not\leq A$. It follows that $\alpha \not\approx \beta$ for any $\beta < \alpha$, so $H(A) = \alpha$ is a cardinal. \square

Definition 4.46. If κ is a cardinal, we let $\kappa^+ = H(\kappa)$.

Definition 4.47. We define \aleph_α for $\alpha \in \mathbf{ORD}$ by

1. $\aleph_0 = \omega$.
2. $\aleph_{\alpha+1} = \aleph_\alpha^+$.
3. $\aleph_\alpha = \bigcup\{\aleph_\beta : \beta < \alpha\}$ if α is a limit ordinal.

The following proposition can be proved by a straightforward induction.

Proposition 4.48. Let α and β be ordinals.

1. $\alpha \leq \aleph_\alpha$.
2. If $\alpha < \beta$, then $\aleph_\alpha < \aleph_\beta$.

Proposition 4.49. Let κ be an ordinal. κ is an infinite cardinal if and only if there exists $\alpha \in \mathbf{ORD}$ with $\kappa = \aleph_\alpha$.

Proof. We first prove that \aleph_α is an infinite cardinal for all $\alpha \in \mathbf{ORD}$ by induction. Notice that $\aleph_0 = \omega$ is a cardinal by Proposition 4.41. Also, if \aleph_α is a cardinal, then $\aleph_{\alpha+1} = \aleph_\alpha^+ = H(\aleph_\alpha)$ is a cardinal by Proposition 4.45. Suppose then that α is a limit ordinal and that \aleph_β is a cardinal for all $\beta < \alpha$. Notice that \aleph_α is an ordinal by Proposition 4.26. Suppose that $\gamma < \aleph_\alpha$. Since $\gamma < \aleph_\alpha = \bigcup\{\aleph_\beta : \beta < \alpha\}$, there exists $\beta < \alpha$ such that $\gamma < \aleph_\beta$. Notice that $\beta + 1 < \alpha$ since $\beta < \alpha$ and α is a limit ordinal. Since $\aleph_{\beta+1} \not\leq \aleph_\beta$, it follows that $\aleph_{\beta+1} \not\leq \gamma$, so $\aleph_\alpha \not\leq \gamma$. Therefore $\aleph_\alpha \not\approx \gamma$ for any $\gamma < \aleph_\alpha$, hence \aleph_α is a cardinal.

Suppose now that κ is an infinite cardinal. By Proposition 4.48, we have $\kappa \leq \aleph_\kappa$. If $\kappa = \aleph_\kappa$, we are done. Suppose then that $\kappa < \aleph_\kappa$ let α be the least ordinal such that $\kappa < \aleph_\alpha$. Notice that $\alpha \neq 0$ because κ is infinite and α can not be a limit ordinal (otherwise, $\kappa < \aleph_\beta$ for some $\beta < \alpha$). Thus, there exists β such that $\alpha = S(\beta)$. By choice of α , we have $\aleph_\beta \leq \kappa$. If $\aleph_\beta < \kappa$, then $\aleph_\beta < \kappa < \aleph_{S(\beta)} = H(\aleph_\beta)$, contradicting the definition of $H(\aleph_\beta)$. It follows that $\kappa = \aleph_\beta$. \square

Proposition 4.50. *Let A be a set. There exists an ordinal α such that $A \approx \alpha$ if and only if A can be well-ordered.*

Proof. Suppose first that there exists an ordinal α such that $A \approx \alpha$. We use a bijection between A and α to transfer the ordering on the ordinals to an ordering on A . Let $f: A \rightarrow \alpha$ be a bijection. Define a relation $<$ on A by letting $a < b$ if and only if $f(a) < f(b)$. It is then straightforward to check that $(A, <)$ is a well-ordering (using the fact that (α, \in_α) is a well-ordering).

For the converse direction, suppose that A can be well-ordered. Fix a relation $<$ on A so that $(A, <)$ is a well-ordering. By Theorem 4.33, there is an ordinal α such that $A \cong \alpha$. In particular, we have $A \approx \alpha$. \square

Of course, this leaves open the question of which sets can be well-ordered. Below, we will use the Axiom of Choice to show that every set can be well-ordered.

Definition 4.51. *Let A be a set which can be well-ordered. We define $|A|$ to be the least ordinal α such that $A \approx \alpha$.*

Lemma 4.52. *If A can be well-ordered, then $|A|$ is a cardinal.*

4.5 Addition and Multiplication Of Cardinals

Lemma 4.53. *Let A_1, A_2, B_1, B_2 be sets with $A_1 \approx A_2$ and $B_1 \approx B_2$.*

1. $(A_1 \times \{0\}) \cup (B_1 \times \{1\}) \approx (A_2 \times \{0\}) \cup (B_2 \times \{1\})$.
2. $A_1 \times B_1 \approx A_2 \times B_2$.

This lemma gives us a reasonable way to define the sum and product of two cardinals. Let κ and λ be cardinals. Notice that $(\kappa \times \{0\}) \cup (\lambda \times \{1\})$ and $\kappa \times \lambda$ can be well-ordered. This allows us to make the following definition.

Definition 4.54. *Let κ and λ be cardinals.*

1. $\kappa + \lambda = |(\kappa \times \{0\}) \cup (\lambda \times \{1\})|$.
2. $\kappa \cdot \lambda = |\kappa \times \lambda|$.

Proposition 4.55. *Let κ and λ be cardinals.*

1. $\kappa + \lambda = \lambda + \kappa$.
2. $\kappa \cdot \lambda = \lambda \cdot \kappa$.

Definition 4.56. *We define an ordering $<$ on $\mathbf{ORD} \times \mathbf{ORD}$ as follows. Let $\alpha_1, \beta_1, \alpha_2, \beta_2$ be ordinals. We set $(\alpha_1, \beta_1) < (\alpha_2, \beta_2)$ if one of the following holds.*

1. $\max\{\alpha_1, \beta_1\} < \max\{\alpha_2, \beta_2\}$.
2. $\max\{\alpha_1, \beta_1\} = \max\{\alpha_2, \beta_2\}$ and $\alpha_1 < \alpha_2$.
3. $\max\{\alpha_1, \beta_1\} = \max\{\alpha_2, \beta_2\}$, $\alpha_1 = \alpha_2$, and $\beta_1 < \beta_2$.

Lemma 4.57. $<$ is a well-ordering on $\mathbf{ORD} \times \mathbf{ORD}$.

Proof. Transitivity, asymmetry, and connectedness are easily shown by appealing to the transitivity, asymmetry, and connectedness of the ordering on \mathbf{ORD} . Let \mathbf{C} be a nonempty subclass of $\mathbf{ORD} \times \mathbf{ORD}$. Notice that $\mathbf{D} = \{\max\{\alpha, \beta\} : (\alpha, \beta) \in \mathbf{C}\}$ is a nonempty subclass of \mathbf{ORD} , hence has a least element δ by Proposition 4.28. Now let $A = \{\alpha \in \delta : (\alpha, \delta) \in \mathbf{C}\}$.

Suppose first that $A \neq \emptyset$, and let α_0 be the least element of A (which exists by Proposition 4.25). Let $(\alpha, \beta) \in \mathbf{C}$. Notice that if $\max\{\alpha, \beta\} > \delta$, we then have $(\alpha_0, \delta) < (\alpha, \beta)$. Suppose then that $\max\{\alpha, \beta\} = \delta$. If $\alpha = \delta$, we then have $(\alpha_0, \delta) < (\alpha, \beta)$ because $\alpha_0 < \delta$. If $\alpha \neq \delta$ and $\beta = \delta$, we then have $\alpha_0 \leq \alpha$ by choice of α_0 , hence $(\alpha_0, \delta) \leq (\alpha, \beta)$.

Suppose now that $A = \emptyset$. Let $B = \{\beta \in S(\delta) : (\delta, \beta) \in \mathbf{C}\}$ and notice that $B \neq \emptyset$. Let β_0 be the least element of B (which exists by Proposition 4.25). Let $(\alpha, \beta) \in \mathbf{C}$. Notice that if $\max\{\alpha, \beta\} > \delta$, we then have $(\delta, \beta_0) < (\alpha, \beta)$. Suppose then that $\max\{\alpha, \beta\} = \delta$. Notice that we must have $\alpha = \delta$ because $A = \emptyset$. It follows that $\beta_0 \leq \beta$ by choice of β_0 , hence $(\delta, \beta_0) \leq (\alpha, \beta)$. \square

Theorem 4.58. For all $\alpha \in \mathbf{ORD}$, we have $\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$.

Proof. The proof is by induction on $\alpha \in \mathbf{ORD}$. Suppose α is an ordinal and that $\aleph_\beta \cdot \aleph_\beta = \aleph_\beta$ for all $\beta < \alpha$. Notice that if we restrict the $<$ relation on $\mathbf{ORD} \times \mathbf{ORD}$ to $\aleph_\alpha \times \aleph_\alpha$, we still get a well-ordering. Given $(\gamma, \delta) \in \aleph_\alpha \times \aleph_\alpha$, we let

$$P_{\gamma, \delta} = \{(\theta_1, \theta_2) \in \aleph_\alpha \times \aleph_\alpha : (\theta_1, \theta_2) < (\gamma, \delta)\}.$$

Let $(\gamma, \delta) \in \aleph_\alpha \times \aleph_\alpha$. Let $\varepsilon = \max\{\gamma, \delta\} + 1$. Since $\gamma, \delta < \aleph_\alpha$, and \aleph_α is an infinite cardinal by Proposition 4.49, it follows that $\varepsilon < \aleph_\alpha$ and hence $|\varepsilon| < \aleph_\alpha$. Fix $\beta < \alpha$ such that $|\varepsilon| = \aleph_\beta$. We then have $P_{\gamma, \delta} \subseteq \varepsilon \times \varepsilon \approx \aleph_\beta \times \aleph_\beta \approx \aleph_\beta$, by induction. Therefore, $|P_{\gamma, \delta}| < \aleph_\alpha$ for every $(\gamma, \delta) \in \aleph_\alpha \times \aleph_\alpha$.

Since $\aleph_\alpha \times \aleph_\alpha$ is well-ordered by $<$, it follows from Theorem 4.33 that $\aleph_\alpha \times \aleph_\alpha \cong \theta$ for some ordinal θ . Let $f: \aleph_\alpha \times \aleph_\alpha \rightarrow \theta$ be a witnessing isomorphism. Then f is injective, so we must have $\aleph_\alpha \preceq \theta$, and hence $\aleph_\alpha \leq \theta$. Suppose that $\aleph_\alpha < \theta$. Since f is an isomorphism, there exists $(\gamma, \delta) \in \aleph_\alpha \times \aleph_\alpha$ such that $f((\gamma, \delta)) = \aleph_\alpha$. We then have $|P_{\gamma, \delta}| = \aleph_\alpha$, a contradiction. It follows that $\theta = \aleph_\alpha$, so f witnesses that $\aleph_\alpha \times \aleph_\alpha \approx \aleph_\alpha$. Hence $\aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha$. \square

Corollary 4.59. *Suppose that κ and λ are cardinals, $1 \leq \kappa \leq \lambda$, and $\lambda \geq \aleph_0$. We then have*

1. $\kappa + \lambda = \lambda = \lambda + \kappa$.
2. $\kappa \cdot \lambda = \lambda = \lambda \cdot \kappa$.

Proof. Fix α such that $\lambda = \aleph_\alpha$ by Proposition 4.49. Notice that

$$\kappa \cdot \lambda \leq \lambda \cdot \lambda = \aleph_\alpha \cdot \aleph_\alpha = \aleph_\alpha = \lambda.$$

Since we clearly have $\lambda \leq \kappa \cdot \lambda$, it follows that $\kappa \cdot \lambda = \lambda$. Also, notice that

$$\kappa + \lambda \leq \lambda + \lambda = 2 \cdot \lambda = \lambda.$$

Since we clearly have $\lambda \leq \kappa + \lambda$, it follows that $\kappa + \lambda = \lambda$. □

5 The Axiom Of Choice

5.1 Use of the Axiom of Choice in Mathematics

Definition 5.1. *Let \mathcal{F} be a family of nonempty sets. A choice function on \mathcal{F} is a function $h: \mathcal{F} \rightarrow \bigcup \mathcal{F}$ such that $h(A) \in A$ for all $A \in \mathcal{F}$.*

Proposition 5.2. *The following are equivalent (over ZF).*

1. *The Axiom of Choice: If \mathcal{F} is a family of nonempty pairwise disjoint sets, then there is a set C such that there is a unique element of $C \cap A$ for every $A \in \mathcal{F}$.*
2. *Every family \mathcal{F} of nonempty sets has a choice function.*
3. *Every family \mathcal{F} of nonempty pairwise disjoint sets has a choice function.*

Proof. 1 implies 2: Let \mathcal{F} be a family of nonempty sets. Let $\mathcal{G} = \{\{A\} \times A : A \in \mathcal{F}\}$, and notice that \mathcal{G} is a set by Collection and Separation. Furthermore, \mathcal{G} is a family of nonempty pairwise disjoint sets. By 1, there is a set C such that there is unique element of $C \cap B$ for every $B \in \mathcal{G}$. By Separation, we may assume that $C \subseteq \bigcup \mathcal{G}$. Letting $h = C$, it now follows that $h: \mathcal{F} \rightarrow \bigcup \mathcal{F}$ and $h(A) \in A$ for every $A \in \mathcal{F}$. Therefore, \mathcal{F} has a choice function.

2 implies 3: Trivial.

3 implies 1: Let \mathcal{F} be a family of nonempty pairwise disjoint sets. By 3, there is choice function h for \mathcal{F} . Let $C = \text{ran}(h)$ and notice that there is a unique element of $C \cap A$ for every $A \in \mathcal{F}$ (because the sets in \mathcal{F} are pairwise disjoint). □

Here are some examples where the Axiom of Choice is implicitly used in mathamtics.

Proposition 5.3. *If $f: A \rightarrow B$ is a surjection, there exists an injection $g: B \rightarrow A$ such that $f \circ g = id_B$.*

Proof. The idea of constructing such a g is to let $g(b)$ be an arbitrary $a \in A$ such that $f(a) = b$. When you think about it, there doesn't seem to be a way to define g without making all of these arbitrary choices.

Define a function $H: B \rightarrow \mathcal{P}(A)$ by letting $H(b) = \{a \in A : f(a) = b\}$. Notice that $H(b) \neq \emptyset$ for every $b \in B$ because f is surjective. Let $h: \mathcal{P}(A) \setminus \{\emptyset\} \rightarrow A$ be a choice function, so $h(D) \in D$ for every $D \in \mathcal{P}(A) \setminus \{\emptyset\}$. Set $g = h \circ H$ and notice that $g: B \rightarrow A$. We first show that $(f \circ g)(b) = b$ for every $b \in B$. Let $b \in B$. Since $h(H(b)) \in H(b)$, it follows that $f(h(H(b))) = b$, hence $(f \circ g)(b) = f(g(b)) = f(h(H(b))) = b$. Therefore, $f \circ g$ is the identity function on B . We finally show that g is injective. Let $b_1, b_2 \in B$ with $g(b_1) = g(b_2)$. We then have $b_1 = (f \circ g)(b_1) = f(g(b_1)) = f(g(b_2)) = (f \circ g)(b_2) = b_2$. □

Proposition 5.4. *If $f: \mathbb{R} \rightarrow \mathbb{R}$ and $y \in \mathbb{R}$, then f is continuous at y if and only if for every sequence $\{x_n\}_{n \in \omega}$ with $\lim_{n \rightarrow \infty} x_n = y$, we have $\lim_{n \rightarrow \infty} f(x_n) = f(y)$.*

Proof. The left-to-right direction is unproblematic. For the right-to-left direction, the argument is as follows. Suppose that f is not continuous at y , and fix $\varepsilon > 0$ such that there is no $\delta > 0$ such that whenever $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$. We define a sequence as follows. Given $n \in \omega$, let x_n be an arbitrary real number with $|x_n - y| < \frac{1}{n}$ such that $|f(x_n) - f(y)| \geq \varepsilon$. Again, we're making infinitely many arbitrary choices in the construction.

Suppose that f is not continuous at y , and fix $\varepsilon > 0$ such that there is no $\delta > 0$ such that whenever $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$. Define a function $H: \mathbb{R}^+ \rightarrow \mathcal{P}(\mathbb{R})$ by letting $H(\delta) = \{x \in \mathbb{R} : |x - y| < \delta \text{ and } |f(x) - f(y)| \geq \varepsilon\}$. Notice that $H(\delta) \neq \emptyset$ for every $\delta \in \mathbb{R}^+$ by assumption. Let $h: \mathcal{P}(\mathbb{R}) \setminus \{\emptyset\} \rightarrow \mathbb{R}$ be a choice function. For each $n \in \omega$, let $x_n = h(H(\frac{1}{n}))$. One then easily checks that $\lim_{n \rightarrow \infty} x_n = y$ but it's not the case that $\lim_{n \rightarrow \infty} f(x_n) = f(y)$. \square

Another example is the proof is the countable union of countable sets is countable. Let $\{A_n\}_{n \in \omega}$ be countable sets. The first step is to fix injections $f_n: A_n \rightarrow \omega$ for each $n \in \omega$ and then build an injection $f: \bigcup_{n \in \omega} A_n \rightarrow \omega$ from these. However, we are again making infinitely many arbitrary choices when we fix the injections. We'll prove a generalization of this fact using the Axiom of Choice below.

Example. Let $\mathcal{F} = \mathcal{P}(\omega) \setminus \{0\}$. Notice that $\bigcup \mathcal{F} = \omega$. We can prove the existence of a choice function for \mathcal{F} without the Axiom of Choice as follows. Define $g: \mathcal{F} \rightarrow \omega$ by letting $g(A)$ be the $<$ -least element of A for every $A \in \mathcal{P}(\omega) \setminus \{0\}$. More formally, we define $g = \{(A, a) \in \mathcal{F} \times \omega : a \in A \text{ and } a \leq b \text{ for all } b \in A\}$ and prove that g is a choice function on \mathcal{F} . \square

Proposition 5.5. *Without the Axiom of Choice, one can prove that if \mathcal{F} is a family of nonempty sets and \mathcal{F} is finite, then \mathcal{F} has a choice function.*

5.2 Equivalents of the Axiom of Choice

Theorem 5.6 (Zermelo). *The following are equivalent.*

1. *The Axiom of Choice.*
2. *Every set can be well-ordered.*

Proof. 2 implies 1: We show that every family of nonempty sets has a choice function. Let \mathcal{F} be a family of nonempty sets. By 2, we can fix a well-ordering $<$ of $\bigcup \mathcal{F}$. Define $g: \mathcal{F} \rightarrow \bigcup \mathcal{F}$ by letting $g(A)$ be the $<$ -least element of A . Notice that g is a choice function on \mathcal{F} .

1 implies 2: Let A be a set. It suffices to show that there is an ordinal α such that $\alpha \approx A$. Our goal is to define a class function $\mathbf{F}: \mathbf{ORD} \rightarrow A$ recursively. First, let $g: \mathcal{P}(A) \setminus \{\emptyset\} \rightarrow A$ be a choice function. Fix $x \notin A$. We now define \mathbf{F} as follows. If $x \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$ or $\text{ran}(\mathbf{F} \upharpoonright \alpha) = A$, let $\mathbf{F}(\alpha) = x$. Otherwise, $\text{ran}(\mathbf{F} \upharpoonright \alpha) \subsetneq A$, and we let $\mathbf{F}(\alpha) = g(A \setminus \text{ran}(\mathbf{F} \upharpoonright \alpha))$. Since A is a set and \mathbf{ORD} is a proper class, we know that \mathbf{F} is not injective. It follows that we must have $x \in \text{ran}(\mathbf{F})$ (otherwise, a simple induction shows that \mathbf{F} is injective). Let α be the least ordinal such that $\mathbf{F}(\alpha) = x$. A straightforward induction now shows that $\mathbf{F} \upharpoonright \alpha: \alpha \rightarrow A$ is injective, and we notice that it is surjective because $\mathbf{F}(\alpha) = x$. It follows that $A \approx \alpha$. \square

Definition 5.7. *Zorn's Lemma is the statement that if $(P, <)$ is nonempty partially ordered set with the property that each chain in P has an upper bound in P , then P has a maximal element.*

Theorem 5.8. *The following are equivalent.*

1. *The Axiom of Choice.*

2. Zorn's Lemma.

Proof. 1 implies 2: Let $(P, <)$ be nonempty partially ordered set with the property that each chain in P has an upper bound in P . Let $g: \mathcal{P}(P) \setminus \{\emptyset\} \rightarrow P$ be a choice function. Fix $x \notin P$. We define a class function $\mathbf{F}: \mathbf{ORD} \rightarrow P$ recursively as follows. If $x \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$, let $\mathbf{F}(\alpha) = x$. Also, if $\text{ran}(\mathbf{F} \upharpoonright \alpha) \subseteq A$ and there is no $q \in P$ such that $q > p$ for every $p \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$, let $\mathbf{F}(\alpha) = x$. Otherwise, $\text{ran}(\mathbf{F} \upharpoonright \alpha) \subseteq A$ and $\{q \in P : q > p \text{ for every } p \in \text{ran}(\mathbf{F} \upharpoonright \alpha)\} \neq \emptyset$, and we let $\mathbf{F}(\alpha) = g(\{q \in P : q > p \text{ for every } p \in \text{ran}(\mathbf{F} \upharpoonright \alpha)\})$. We know that \mathbf{F} can not be injective, so as above we must have $x \in \text{ran}(\mathbf{F})$. Fix the least ordinal α such that $\mathbf{F}(\alpha) = x$. A straightforward induction shows that $\text{ran}(\mathbf{F} \upharpoonright \alpha)$ is injective and that $\text{ran}(\mathbf{F} \upharpoonright \alpha)$ is a chain in P .

Notice that $\alpha \neq 0$ because $P \neq \emptyset$. Suppose that α is a limit ordinal. Since $\text{ran}(\mathbf{F} \upharpoonright \alpha)$ is a chain in P , we know by assumption that there exists $q \in P$ with $q \geq p$ for all $p \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$. Notice that we can not have $q = \mathbf{F}(\beta)$ for any $\beta < \alpha$ because we would then have $\beta + 1 < \alpha$ (because α is a limit ordinal) and $q < \mathbf{F}(\beta + 1)$ by definition of \mathbf{F} , contrary to the fact that $q \geq p$ for all $p \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$. It follows that $q > p$ for all $p \in \text{ran}(\mathbf{F} \upharpoonright \alpha)$, hence $\mathbf{F}(\alpha) \neq x$, a contradiction. It follows that α is a successor ordinal, say $\alpha = S(\beta)$. Since $\mathbf{F}(\beta) \neq x$ and $\mathbf{F}(S(\beta)) = x$, it follows that $\mathbf{F}(\beta)$ is a maximal element of P .

2 implies 1: Let \mathcal{F} be a family of nonempty sets. We use Zorn's Lemma to show that \mathcal{F} has a choice function. Let $P = \{q : q \text{ is a function, } \text{dom}(q) \subseteq \mathcal{F}, \text{ and } q(A) \in A \text{ for every } A \in \text{dom}(q)\}$. Given $p, q \in P$, we let $p < q$ if and only if $p \subsetneq q$. It is easy to check that $(P, <)$ is a partial ordering. Notice that $P \neq \emptyset$ because $\emptyset \in P$. Also, if H is a chain in P , then $\bigcup H \in P$, and $p \leq \bigcup H$ for all $p \in H$. It follows that every chain in P has an upper bound in P . By Zorn's Lemma, P has a maximal element which we call g . We need only show that $\text{dom}(g) = \mathcal{F}$. Suppose instead that $\text{dom}(g) \subsetneq \mathcal{F}$, and fix $A \in \mathcal{F} \setminus \text{dom}(g)$. Fix $a \in A$. We then have $g \cup \{(A, a)\} \in P$ and $g < g \cup \{(A, a)\}$, a contradiction. It follows that $\text{dom}(g) = \mathcal{F}$, so g is a choice function on \mathcal{F} . \square

5.3 The Axiom of Choice and Cardinal Arithmetic

Once we adopt the Axiom of Choice, it follows that every set can be well-ordered. Therefore, $|A|$ is defined for every set A .

Proposition 5.9. *Let A and B be sets.*

1. $A \preceq B$ if and only if $|A| \leq |B|$.
2. $A \approx B$ if and only if $|A| = |B|$.

Proof.

1. Suppose first that $|A| \leq |B|$. Let $\kappa = |A|$ and let $\lambda = |B|$, and fix bijections $f: A \rightarrow \kappa$ and $g: \lambda \rightarrow B$. Since $\kappa \leq \lambda$, we have $\kappa \subseteq \lambda$ and so we may consider $g \circ f: A \rightarrow B$. One easily checks that this is an injective function.
Suppose now that $A \preceq B$, and fix an injection $h: A \rightarrow B$. Let $\kappa = |A|$ and let $\lambda = |B|$, and fix bijections $f: \kappa \rightarrow A$ and $g: B \rightarrow \lambda$. We then have that $g \circ h \circ f: \kappa \rightarrow \lambda$ is an injection, hence $\kappa \leq \lambda$.
2. Suppose first that $A \approx B$. We then have that $|A| \leq |B|$ and $|B| \leq |A|$ by part 1, hence $|A| = |B|$.
3. Suppose now that $|A| = |B|$. By part 1, we then have that $A \preceq B$ and $B \preceq A$, hence $A \approx B$ by the Cantor-Bernstein Theorem.

\square

Proposition 5.10. $|A \times A| = |A|$ for every infinite set A .

Proof. Since A is infinite, there exists α such that $|A| = \aleph_\alpha$. We then have $A \times A \approx \aleph_\alpha \times \aleph_\alpha \approx \aleph_\alpha$, hence $|A \times A| \leq \aleph_\alpha$. We clearly have $\aleph_\alpha \leq |A \times A|$, hence $|A \times A| = \aleph_\alpha = |A|$. \square

Proposition 5.11. *Let \mathcal{F} be a family of sets. Suppose that $|\mathcal{F}| \leq \kappa$ and that $|A| \leq \lambda$ for every $A \in \mathcal{F}$. We then have $|\bigcup \mathcal{F}| \leq \kappa \cdot \lambda$.*

Proof. Let $\mu = |\mathcal{F}|$ (notice that $\mu \leq \kappa$), and fix a bijection $f: \mu \rightarrow \mathcal{F}$. Also, for each $A \in \mathcal{F}$, fix an injection $g_A: A \rightarrow \lambda$ (using the Axiom of Choice). We define an injection $h: \bigcup \mathcal{F} \rightarrow \kappa \times \lambda$ as follows. Given $b \in \bigcup \mathcal{F}$, let α be the least ordinal such that $b \in f(\alpha)$, and set $h(b) = (\alpha, g_{f(\alpha)}(b))$. Suppose that $b_1, b_2 \in \bigcup \mathcal{F}$ and $h(b_1) = h(b_2)$. Let α_1 be the least ordinal such that $b_1 \in f(\alpha_1)$ and let α_2 be the least ordinal such that $b_2 \in f(\alpha_2)$. Since $h(b_1) = h(b_2)$, it follows that $\alpha_1 = \alpha_2$, and we call their common value α . Therefore, using the fact that $h(b_1) = h(b_2)$ again, we conclude that $g_{f(\alpha)}(b_1) = g_{f(\alpha)}(b_2)$. Since $g_{f(\alpha)}$ is an injection, it follows that $b_1 = b_2$. Hence, $h: \mathcal{F} \rightarrow \kappa \times \lambda$ is an injection, so we may conclude that $|\mathcal{F}| \leq \kappa \cdot \lambda$. \square

Proposition 5.12. $|A^{<\omega}| = |A|$ for every infinite set A .

Proof. Using Proposition 5.10 and induction (on ω), it follows that $|A^n| = |A|$ for every $n \in \omega$ with $n \geq 1$. Since $A^{<\omega} = \bigcup \{A^n : n \in \omega\}$, we may use Proposition 5.11 to conclude that $|A^{<\omega}| \leq \aleph_0 \cdot |A| = |A|$. We clearly have $|A| \leq |A^{<\omega}|$, hence $|A^{<\omega}| = |A|$. \square

Definition 5.13. *Let A and B be sets. We let A^B be the set of all functions from B to A .*

Proposition 5.14. *Let A_1, A_2, B_1, B_2 be sets with $A_1 \approx A_2$ and $B_1 \approx B_2$. We then have $A_1^{B_1} \approx A_2^{B_2}$.*

Now that we've adopted the Axiom of Choice, we know that A^B can be well-ordered for any sets A and B , so it makes sense to talk about $|A^B|$. This gives us a way to define cardinal exponentiation.

Definition 5.15. *Let κ and λ be cardinals. We use κ^λ to also denote the cardinality of the set κ^λ . (So, we're using the same notation κ^λ to denote both the set of functions from λ to κ and also its cardinality).*

Proposition 5.16. *Let κ, λ , and μ be cardinals.*

1. $\kappa^{\lambda+\mu} = \kappa^\lambda \cdot \kappa^\mu$.
2. $\kappa^{\lambda \cdot \mu} = (\kappa^\lambda)^\mu$.
3. $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$.

Proof. Fix sets A, B, C such that $|A| = \kappa, |B| = \lambda$, and $|C| = \mu$ (we could use κ, λ , and μ , but it's easier to distinguish sets from cardinals).

1. It suffices to find a bijection $F: A^{B \times \{0\} \cup C \times \{1\}} \rightarrow A^B \times A^C$. We define F as follows. Given $f: B \times \{0\} \cup C \times \{1\} \rightarrow A$, let $F(f) = (g, h)$ where $g: B \rightarrow A$ is given by $g(b) = f((b, 0))$ and $h: C \rightarrow A$ is given by $h(c) = f((c, 1))$.
2. It suffices to find a bijection $F: (A^B)^C \rightarrow A^{B \times C}$. We define F as follows. Given $f: C \rightarrow A^B$, let $F(f): B \times C \rightarrow A$ be the function defined by $F(f)((b, c)) = f(c)(b)$ for all $b \in B$ and $c \in C$.
3. It suffices to find a bijection $F: A^C \times B^C \rightarrow (A \times B)^C$. We define F as follows. Given $g: C \rightarrow A$ and $h: C \rightarrow B$, let $F((g, h)): C \rightarrow A \times B$ be the function defined by $F((g, h))(c) = (g(c), h(c))$ for all $c \in C$. \square

Proposition 5.17. $2^\kappa = |\mathcal{P}(\kappa)|$ for all cardinals κ .

Proof. Fix a cardinal κ . We define a function $F: 2^\kappa \rightarrow \mathcal{P}(\kappa)$ as follows. Given $f: \kappa \rightarrow 2$, let $F(f) = \{\alpha \in \kappa : f(\alpha) = 1\}$. We then have that F is a bijection, hence $2^\kappa = |\mathcal{P}(\kappa)|$. \square

Corollary 5.18. $\kappa < 2^\kappa$ for all cardinals κ .

Proof. We know that $\kappa \prec \mathcal{P}(\kappa)$ from above. □

Proposition 5.19. If $2 \leq \lambda \leq \kappa$, then $\lambda^\kappa = 2^\kappa$

Proof. We have

$$2^\kappa \leq \lambda^\kappa \leq \kappa^\kappa \leq (2^\kappa)^\kappa = 2^{\kappa \cdot \kappa} = 2^\kappa$$

□

6 Set-theoretic Methods in Analysis and Model Theory

6.1 Subsets of \mathbb{R}

6.1.1 The Reals

Proposition 6.1. $|\mathbb{R}| = 2^{\aleph_0}$.

Proof. The function $f: \mathbb{R} \rightarrow \mathcal{P}(\mathbb{Q})$ given by $f(x) = \{q \in \mathbb{Q} : q < x\}$ is injective, so

$$|\mathbb{R}| \leq |\mathcal{P}(\mathbb{Q})| = 2^{|\mathbb{Q}|} = 2^{\aleph_0}$$

The function $f: 2^{\aleph_0} \rightarrow \mathbb{R}$ given by

$$f(q) = \sum_{n=0}^{\infty} \frac{q(n)}{10^n}$$

is injective, so $2^{\aleph_0} \leq |\mathbb{R}|$. □

Proposition 6.2. If $a, b \in \mathbb{R}$ and $a < b$, then $|(a, b)| = 2^{\aleph_0}$.

Proof. The above injection shows that $|(0, 1)| = 2^{\aleph_0}$ and for all $a, b \in \mathbb{R}$ with $a < b$, we have $(0, 1) \approx (a, b)$. □

Proposition 6.3. If O is a nonempty open subset of \mathbb{R} , then $|O| = 2^{\aleph_0}$.

Proof. Every nonempty open subset of \mathbb{R} contains an open interval. □

6.1.2 Perfect Sets

Definition 6.4. Let $P \subseteq \mathbb{R}$. We say that P is perfect if it is closed and has no isolated points.

Example. $[a, b]$ is perfect for all $a, b \in \mathbb{R}$ with $a < b$. □

Proposition 6.5. The Cantor Set C defined by

$$C = \left\{ \sum_{n=1}^{\infty} \frac{q(n-1)}{3^n} : q \in \{0, 2\}^{\omega} \right\}$$

is perfect.

Proof. Consult your favorite analysis book. □

Proposition 6.6. If $P \subseteq \mathbb{R}$ is perfect and $a, b \in \mathbb{R}$ with $a < b$ and $a, b \notin P$, then $P \cap [a, b]$ is perfect.

Proof. Since both P and $[a, b]$ are closed, it follows that $P \cap [a, b]$ is closed. Fix $x \in P \cap [a, b]$, and notice that $x > a$ and $x < b$ since $a, b \notin P$. Let $\varepsilon > 0$. Since P is perfect, we know that x is not isolated in P , so there exists $y \in P$ such that $0 < |x - y| < \min\{\varepsilon, x - a, b - x\}$. We then have that $0 < |x - y| < \varepsilon$ and also that $y \in [a, b]$ (by choice of ε). Therefore, x is not isolated in $P \cap [a, b]$. It follows that $P \cap [a, b]$ is perfect. \square

Proposition 6.7. *If $P \subseteq \mathbb{R}$ is a nonempty perfect set and $\varepsilon > 0$, then there exists nonempty perfect sets $P_1, P_2 \subseteq \mathbb{R}$ such that*

1. $P_1 \cap P_2 = \emptyset$.
2. $P_1 \cup P_2 \subseteq P$.
3. $\text{diam}(P_1), \text{diam}(P_2) < \varepsilon$.

Proof. Let $P \subseteq \mathbb{R}$ be a nonempty perfect set and let $\varepsilon > 0$. Since P is nonempty, we may fix $x \in P$.

Case 1: There exists $\delta > 0$ such that $[x - \delta, x + \delta] \subseteq P$. We may assume (by making δ smaller if necessary) that $\delta < \varepsilon$. In this case, let $P_1 = [x - \delta, x - \frac{\delta}{2}]$ and let $P_2 = [x + \frac{\delta}{2}, x + \delta]$.

Case 2: Otherwise, for every $\delta > 0$, there exists infinitely many $y \in [x - \delta, x + \delta] \setminus P$. Thus, there exists points $a, b, c, d \in [x - \frac{\varepsilon}{4}, x + \frac{\varepsilon}{4}] \setminus P$ such that $a < b < c < d$. In this case, let $P_1 = P \cap [a, b]$ and let $P_2 = P \cap [c, d]$. \square

Proposition 6.8. *If $P \subseteq \mathbb{R}$ is a nonempty perfect set, then $|P| = 2^{\aleph_0}$.*

Proof. Since $P \subseteq \mathbb{R}$, we know that $|P| \leq 2^{\aleph_0}$. By the previous Proposition, there exists a nonempty perfect set $Q \subseteq P$ such that $\text{diam}(Q) < 1$. We can now use the previous Proposition to recursively define a function $f: 2^{<\omega} \rightarrow \mathcal{P}(P)$ such that:

1. $f(\lambda) = Q$.
2. $f(\sigma)$ is a nonempty perfect set for all $\sigma \in 2^{<\omega}$.
3. $\text{diam}(f(\sigma)) < \frac{1}{2^{|\sigma|}}$ for all $\sigma \in 2^{<\omega}$.
4. $f(\sigma * 0) \cup f(\sigma * 1) \subseteq f(\sigma)$ for all $\sigma \in 2^{<\omega}$.
5. $f(\sigma * 0) \cap f(\sigma * 1) = \emptyset$ for all $\sigma \in 2^{<\omega}$.

Now define $g: 2^{\aleph_0} \rightarrow P$ by letting $g(q)$ be the unique element of $\bigcap_{n \in \omega} f(q \upharpoonright n)$ for all $q \in 2^{\aleph_0}$ (notice that such an element must exist because the intersection is of a nested sequence of compact sets, and that the element is unique because the diameters go to 0). Finally, notice that g is injective by virtue of property 5 of the function f . \square

6.1.3 Closed Sets

Definition 6.9. *Suppose that $C \subseteq \mathbb{R}$ is a closed set. We define C' to be the set $C - \{x \in \mathbb{R} : x \text{ is an isolated point of } C\}$. We call C' the Cantor-Bendixson derivative of C .*

Notice that a closed set C is perfect if and only if $C = C'$.

Proposition 6.10. *If $C \subseteq \mathbb{R}$ is a closed set, then $C' \subseteq C$ is also closed.*

Proof. Recall that a set is closed if and only if its complement is open. We show that $\overline{C'}$ is open. Fix $x \in \overline{C'}$. If $x \notin C$, then since C is closed, we may fix $\delta > 0$ such that $(x - \delta, x + \delta) \subseteq \overline{C} \subseteq \overline{C'}$. Suppose then that $x \in C$. Since $x \notin C'$, we know that x is an isolated point of C . Fix $\delta > 0$ such that $C \cap (x - \delta, x + \delta) = \{x\}$. We then have that $(x - \delta, x + \delta) \subseteq \overline{C'}$. Therefore, $\overline{C'}$ is open, hence C' is closed. \square

Proposition 6.11. *If $C \subseteq \mathbb{R}$ is a closed set, then $C \setminus C' = \{x \in \mathbb{R} : x \text{ is an isolated point of } C\}$ is countable.*

Proof. Define a function $f: C \setminus C' \rightarrow \mathbb{Q} \times \mathbb{Q}$ by letting $f(x) = (q, r)$ where (q, r) is least (under some fixed well-ordering of $\mathbb{Q} \times \mathbb{Q}$) such that $C \cap (q, r) = \{x\}$. We then have that f is injective, hence $C \setminus C'$ is countable because $\mathbb{Q} \times \mathbb{Q}$ is countable. \square

Definition 6.12. *Let $C \subseteq \mathbb{R}$ be a closed set. We define a sequence $C^{(\alpha)}$ for $\alpha < \omega_1$ recursively as follows.*

1. $C^{(0)} = C$.
2. $C^{(\alpha+1)} = (C^{(\alpha)})'$.
3. $C^{(\alpha)} = \bigcap \{C^{(\beta)} : \beta < \alpha\}$ if α is a limit.

Notice that each $C^{(\alpha)}$ is closed and that $C^{(\beta)} \subseteq C^{(\alpha)}$ whenever $\alpha < \beta < \omega_1$ be a trivial induction.

Proposition 6.13. *Let $C \subseteq \mathbb{R}$ be a closed set. There exists an $\alpha < \omega_1$ such that $C^{(\alpha+1)} = C^{(\alpha)}$.*

Proof. Suppose that $C^{(\alpha+1)} \neq C^{(\alpha)}$ for all $\alpha < \omega_1$. Define a function $f: \omega_1 \rightarrow \mathbb{Q} \times \mathbb{Q}$ by letting $f(\alpha) = (q, r)$ where (q, r) is least (under some fixed well-ordering of $\mathbb{Q} \times \mathbb{Q}$) such that there is a unique element of $C^{(\alpha)} \cap (q, r)$. We then have that f is injective, contrary to the fact that $|\mathbb{Q} \times \mathbb{Q}| = \aleph_0$. \square

Theorem 6.14. *Let $C \subseteq \mathbb{R}$ be a closed set. There exists a perfect set $P \subseteq \mathbb{R}$ and a countable $A \subseteq \mathbb{R}$ such that $C = A \cup P$ and $A \cap P = \emptyset$.*

Proof. Let $\alpha < \omega_1$ be least such that $C^{(\alpha+1)} = C^{(\alpha)}$. Let $P = C^{(\alpha)}$ and let $A = \bigcup_{\beta < \alpha} (C^{(\beta)} \setminus C^{(\beta+1)})$. Notice that $C = A \cup P$ and $A \cap P = \emptyset$, that P is perfect because $P = P'$, and that A is countable because it is the countable union of countable sets. \square

Corollary 6.15. *If $C \subseteq \mathbb{R}$ is an uncountable closed set, then $|C| = 2^{\aleph_0}$.*

Proof. Let $C \subseteq \mathbb{R}$ be an uncountable closed set. We have $|C| \leq 2^{\aleph_0}$ because $C \subseteq \mathbb{R}$. Let P be perfect and A countable such that $C = A \cup P$ and $A \cap P = \emptyset$. Since C is uncountable, we have $P \neq \emptyset$, hence $|P| = 2^{\aleph_0}$, and so $|C| \geq 2^{\aleph_0}$. \square

6.1.4 Borel Sets

Definition 6.16. *Let \mathcal{O} be the set of open subsets of \mathbb{R} . We define the set \mathcal{B} of Borel sets to be the smallest subset of $\mathcal{P}(\mathbb{R})$ such that*

1. $\mathcal{O} \subseteq \mathcal{B}$.
2. If $A \in \mathcal{B}$, then $\mathbb{R} \setminus A \in \mathcal{B}$.
3. If $A_n \in \mathcal{B}$ for all $n \in \omega$, then $\bigcup_{n \in \omega} A_n \in \mathcal{B}$.

Definition 6.17. *We define a sequence $(\Sigma_\alpha, \Pi_\alpha)$ for $\alpha < \omega_1 \setminus \{0\}$ recursively as follows.*

1. $\Sigma_1 = \mathcal{O}$ and $\Pi_1 = \{\mathbb{R} \setminus A : A \in \mathcal{O}\}$.
2. $\Sigma_{\alpha+1} = \{\bigcup_{n \in \omega} A_n : \text{each } A_n \in \Pi_\alpha\}$ and $\Pi_{\alpha+1} = \{\mathbb{R} \setminus A : A \in \Sigma_{\alpha+1}\}$.
3. $\Sigma_\alpha = \{\bigcup_{n \in \omega} A_n : \text{each } A_n \in \Pi_\beta \text{ for some } \beta < \alpha\}$ and $\Pi_\alpha = \{\mathbb{R} \setminus A : A \in \Sigma_\alpha\}$ if α is a limit.

Proposition 6.18. $\mathcal{B} = \bigcup_{\alpha < \omega_1} \Sigma_\alpha$.

Corollary 6.19. $|\mathcal{B}| = 2^{\aleph_0}$.

Corollary 6.20. $\mathcal{B} \neq \mathcal{P}(\mathbb{R})$.

Proof. We have $|\mathcal{P}(\mathbb{R})| = 2^{|\mathbb{R}|} = 2^{2^{\aleph_0}} > 2^{\aleph_0} = |\mathcal{B}|$. \square

6.1.5 Measurable Sets

Definition 6.21. Let \mathcal{M} be the set of all (Lebesgue) measurable subsets of \mathbb{R} .

Proposition 6.22. $|\mathcal{M}| = 2^{2^{\aleph_0}}$.

Proof. The Cantor set \mathcal{C} is measurable with measure 0, and $|\mathcal{C}| = 2^{\aleph_0}$. Since every subset of a set of measure 0 is measurable (with measure 0), it follows that $\mathcal{P}(\mathcal{C}) \subseteq \mathcal{M}$. Therefore, $|\mathcal{M}| \geq 2^{|\mathcal{C}|} = 2^{2^{\aleph_0}}$. \square

Corollary 6.23. There is a measurable set which is not Borel.

6.2 The Size of Models

6.2.1 Controlling the Size of Models

Proposition 6.24. Let \mathcal{L} be a language and suppose that $\Gamma \subseteq \text{Form}_{\mathcal{L}}$ is satisfiable. There exists a model (\mathcal{M}, s) of Γ such that $|\mathcal{M}| \leq |\mathcal{L}| + \aleph_0$.

Proof. We already proved this last quarter when \mathcal{L} was countable (and particular when \mathcal{L} was finite). Suppose that \mathcal{L} is infinite and let $\kappa = |\mathcal{L}|$.

Recall the proof of the Completeness Theorem. Notice that if \mathcal{L} is consistent, then \mathcal{L}' formed in the first step of adding witnesses satisfies $|\mathcal{L}'| = \kappa$ because $|\text{Form}_{\mathcal{L}} \times \text{Var}| = \kappa^{<\omega} \cdot \aleph_0 = \kappa$. Thus, each \mathcal{L}_n achieved by iteratively adding witnesses satisfies $|\mathcal{L}_n| = \kappa$, so the final $\mathcal{L}' = \bigcup_{n \in \omega} \mathcal{L}_n$ satisfies $|\mathcal{L}'| = \kappa$. It follows that $|\text{Term}_{\mathcal{L}'}| = \kappa$, and since the \mathcal{L}' -structure \mathcal{M}' we constructed in the proof of the Completeness Theorem is formed by taking the quotient from an equivalence relation on the countable $\text{Term}_{\mathcal{L}'}$, we can conclude that $|\mathcal{M}'| \leq \kappa$. Therefore, the \mathcal{L} -structure $\mathcal{M}' \upharpoonright \mathcal{L}$ from the proof of the Completeness Theorem has cardinality at most κ . \square

Theorem 6.25 (Lowenheim-Skolem Theorem). Let \mathcal{L} be a language and suppose that $\Gamma \subseteq \text{Form}_{\mathcal{L}}$ has an infinite model. Let $\kappa \geq |\mathcal{L}| + \aleph_0$. There exists a model (\mathcal{M}, s) of Γ such that $|\mathcal{M}| = \kappa$.

Proof. Suppose that $\kappa \geq |\mathcal{L}|$. Let \mathcal{L}' be \mathcal{L} together with new constant symbols c_α for all $\alpha < \kappa$. Notice that $|\mathcal{L}'| = |\mathcal{L}| + \kappa = \kappa$. Let

$$\Gamma' = \Gamma \cup \{c_\alpha \neq c_\beta : \alpha, \beta < \kappa \text{ and } \alpha \neq \beta\}$$

Notice that every finite subset of Γ' has a model by using an infinite model of Γ and interpreting the constants which appear in the finite subset as distinct elements. Therefore, by Compactness, we know that Γ' is a model. By Proposition 6.24, there exists a model (\mathcal{M}', s) of Γ' such that $|\mathcal{M}'| \leq |\mathcal{L}'| + \aleph_0 = \kappa$. Notice that we must also have $|\mathcal{M}'| \geq \kappa$, hence $|\mathcal{M}'| = \kappa$. Letting \mathcal{M} be the restriction of the structure \mathcal{M}' to the language \mathcal{L} , we see that (\mathcal{M}, s) is a model of Γ and that $|\mathcal{M}| = \kappa$. \square

6.2.2 Counting Models

Definition 6.26. Given a theory T in a language \mathcal{L} and a cardinal κ , let $I(T, \kappa)$ be the number of models of T of cardinality κ up to isomorphism.

Proposition 6.27. Let T be a theory in a language \mathcal{L} with $|\mathcal{L}| = \lambda$. For any infinite cardinal κ , we have $I(T, \kappa) \leq 2^{\kappa \cdot \lambda}$. In particular, if $\kappa \geq \lambda$ is infinite, then $I(T, \kappa) \leq 2^\kappa$.

Proof. Let κ be an infinite cardinal. We have

$$\begin{aligned}
I(T, \kappa) &\leq \kappa^{|\mathcal{L}|} \cdot |\mathcal{P}(\kappa^{<\omega})|^{|\mathcal{R}|} \cdot |\mathcal{P}(\kappa^{<\omega})|^{|\mathcal{F}|} \\
&\leq \kappa^{|\mathcal{L}|} \cdot |\mathcal{P}(\kappa)|^{|\mathcal{R}|} \cdot |\mathcal{P}(\kappa)|^{|\mathcal{F}|} \\
&\leq \kappa^\lambda \cdot (2^\kappa)^\lambda \cdot (2^\kappa)^\lambda \\
&\leq (2^\kappa)^\lambda \cdot (2^\kappa)^\lambda \cdot (2^\kappa)^\lambda \\
&= 2^{\kappa \cdot \lambda}
\end{aligned}$$

□

Proposition 6.28. *If T is the theory of groups, then $I(T, \aleph_0) = 2^{\aleph_0}$.*

Proof. Let P be the set of primes. Notice that the set of finite subsets of P is countable, so the set of infinite subsets of P has cardinality 2^{\aleph_0} . For each infinite $A \in \mathcal{P}(P)$, let

$$G_A = \bigoplus_{p \in A} \mathbb{Z}/p\mathbb{Z}$$

Notice that if $A, B \in \mathcal{P}(P)$ are infinite, then $G_A \not\cong G_B$. □

Proposition 6.29. *Let T be the theory of vector spaces over \mathbb{Q} . We have $I(T, \aleph_0) = \aleph_0$ and $I(T, \kappa) = 1$ for all $\kappa \geq \aleph_1$.*

Proof. Notice first that if V is a vector space over \mathbb{Q} and $\dim_{\mathbb{Q}}(V) = n \in \omega$, then

$$|V| = |\mathbb{Q}^n| = \aleph_0$$

Now if V is a vector space over \mathbb{Q} and $\dim_{\mathbb{Q}}(V) = \kappa \geq \aleph_0$, then since every element of V is a finite sum of scalar multiples of elements of a basis, it follows that

$$|V| \leq |(\mathbb{Q} \times \kappa)^{<\omega}| = |(\aleph_0 \cdot \kappa)^{<\omega}| = |\kappa^{<\omega}| = \kappa.$$

and we clearly have $|V| \geq \kappa$, so $|V| = \kappa$.

Since two vector spaces over \mathbb{Q} are isomorphic if and only if they have the same dimension, it follows that $I(T, \aleph_0) = \aleph_0$ (corresponding to dimensions in $\omega \cup \{\aleph_0\}$) and $I(T, \kappa) = 1$ for all $\kappa \geq \aleph_1$ (corresponding to dimension κ). □

Proposition 6.30. *For any p , we have $I(ACF_p, \aleph_0) = \aleph_0$ and $I(ACF_p, \kappa) = 1$ for all $\kappa \geq \aleph_1$.*

Proof. Notice that if F is an algebraically closed field and $\text{tr.deg.}(F) = \kappa$, then $|F| = \kappa$. □

Definition 6.31. *Let T be a theory and let κ be a cardinal. We say that T is κ -categorical if $I(T, \kappa) = 1$.*

Proposition 6.32 (Los-Vaught Test). *Suppose that T is a theory such that all models of T are infinite. If there exists $\kappa \geq |\mathcal{L}| + \aleph_0$ such that T is κ -categorical, then T is complete.*

Proof. Let T be a theory such that all models of T are infinite. Suppose that T is not complete and fix $\sigma \in \text{Sent}_{\mathcal{L}}$ such that $\sigma \notin T$ and $\neg\sigma \notin T$. We then have that $T \cup \{\sigma\}$ and $T \cup \{\neg\sigma\}$ are both satisfiable with infinite models (because all models of T are infinite), so by the Lowenheim-Skolem Theorem we may fix a model \mathcal{M}_1 of $T \cup \{\sigma\}$ and a model \mathcal{M}_2 of $T \cup \{\neg\sigma\}$ such that $|\mathcal{M}_1| = \kappa = |\mathcal{M}_2|$. We then have that \mathcal{M}_1 and \mathcal{M}_2 are models of T which are not isomorphic, hence $I(T, \kappa) \geq 2$. □

Corollary 6.33. *DLO and each ACF_p are complete.*

Theorem 6.34 (Morley's Theorem). *Let \mathcal{L} be a countable language and let T be a theory. If T is κ -categorical for some $\kappa \geq \aleph_1$, then T is κ -categorical for all $\kappa \geq \aleph_1$.*

6.3 Ultraproducts and Compactness

Let \mathcal{L} be a language. Let I be a set, and suppose that for each $i \in I$ we have an \mathcal{L} -structure \mathcal{M}_i . For initial clarity, think of $I = \omega$, so that we have \mathcal{L} -structures $\mathcal{M}_0, \mathcal{M}_1, \mathcal{M}_2, \dots$. We want a way to put together all of the \mathcal{M}_i which somehow “blends” the properties of the \mathcal{M}_i together into one structure. An initial thought is to form a product of the structures \mathcal{M}_i with underlying set $\prod_{i \in I} M_i$. That is, M consists of all functions $g: I \rightarrow \bigcup_{i \in I} M_i$ such that $g(i) \in M_i$ for all $i \in I$. Interpreting the constants and functions would then be straightforward. For example, suppose that $\mathcal{L} = \{e, f\}$ where e is a constant symbol and f is a binary relation symbol. Suppose that $I = \omega$ and each \mathcal{M}_i is a group. Elements of M would then be sequences $\langle a_i \rangle_{i \in \omega}$, we would interpret e as the sequence of each identity in each group, and we would interpret f as the componentwise group operation (i.e. $f^{\mathcal{M}}(\langle a_i \rangle_{i \in \omega}, \langle b_i \rangle_{i \in \omega}) = \langle f^{\mathcal{M}_i}(a_i, b_i) \rangle_{i \in \omega}$). In general, we would let $c^{\mathcal{M}}$ be the function $i \mapsto c^{\mathcal{M}_i}$ for each constant symbol c , and given $f \in \mathcal{F}_k$ we would let $f^{\mathcal{M}_i}(g_1, g_2, \dots, g_k)$ be the function $i \mapsto f^{\mathcal{M}_i}(g_1(i), g_2(i), \dots, g_k(i))$.

This certainly works, but it doesn’t really “blend” the properties of the structures together particularly well. For example, if each \mathcal{M}_i is a group and all but one is abelian, the product is still nonabelian. Also, if we have relation symbols, it’s not clear what the “right” way to determine how to interpret the relation on \mathcal{M} . For example, if $\mathcal{L} = \{R\}$ where R is a binary relation symbol and $I = \omega$, do we say that the pair $(\langle a_i \rangle_{i \in \omega}, \langle b_i \rangle_{i \in \omega})$ is an element of $R^{\mathcal{M}}$ if *some* $(a_i, b_i) \in R^{\mathcal{M}_i}$, or if *all* $(a_i, b_i) \in R^{\mathcal{M}_i}$, or something else? Which is the “right” definition? In other words, if each \mathcal{M}_i is a graph, do we put an edge between the sequences if some edge exists between the components, or if every pair has an edge?

We thus want a more “democratic” approach of forming \mathcal{M} which also gives a way to nicely interpret the relation symbols. If I were finite, perhaps we could do a majority rules (if *most* of the pairs were in the relation), but what if I is infinite?

6.3.1 Ultrafilters

Definition 6.35. Let X be a set. A filter on X is a set $\mathcal{F} \subseteq \mathcal{P}(X)$ such that

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

Example. Let X be a nonempty set, and let $x \in X$. The set

$$\mathcal{F} = \{A \in \mathcal{P}(X) : x \in A\}$$

is a filter on X . Such a filter is called a *principal* filter on X generated by x . □

Proposition 6.36. Let X be an infinite set. The set

$$\mathcal{F} = \{A \in \mathcal{P}(X) : A \text{ is cofinite}\}$$

is a filter on X .

Proposition 6.37. Let X be a set and let \mathcal{F} be a filter on X . For every finite $\mathcal{T} \subseteq \mathcal{F}$, we have $\bigcap \mathcal{T} \neq \emptyset$.

Proof. By induction on $|\mathcal{T}|$. □

Definition 6.38. Let X be a set and suppose that $\mathcal{S} \subseteq \mathcal{P}(X)$. We say that \mathcal{S} has the finite intersection property if $\bigcap \mathcal{T} \neq \emptyset$ for all finite $\mathcal{T} \subseteq \mathcal{S}$.

Proposition 6.39. Let X be a set and suppose that $\mathcal{S} \subseteq \mathcal{P}(X)$. The following are equivalent

1. \mathcal{S} has the finite intersection property.
2. There exists a filter \mathcal{F} on X such that $\mathcal{S} \subseteq \mathcal{F}$.

Proof. 1 implies 2: Let

$$\mathcal{F} = \{A \in \mathcal{P}(X) : \bigcap \mathcal{T} \subseteq A \text{ for some finite } \mathcal{T} \subseteq \mathcal{S}\}$$

We claim that \mathcal{F} is a filter on X . Notice that we clearly have $X \in \mathcal{F}$, and that $\emptyset \notin \mathcal{F}$ because \mathcal{S} has the finite intersection property. Now if $A \in \mathcal{F}$, say $\bigcap \mathcal{T} \subseteq A$ where $\mathcal{T} \subseteq \mathcal{S}$ is finite, and $A \subseteq B \subseteq X$, then $\bigcap \mathcal{T} \subseteq B$, so $B \in \mathcal{F}$. Finally, suppose that $A, B \in \mathcal{F}$, and fix finite $\mathcal{T}_1, \mathcal{T}_2 \subseteq \mathcal{S}$ such that $\bigcap \mathcal{T}_1 \subseteq A$ and $\bigcap \mathcal{T}_2 \subseteq B$. We then have that $\bigcap(\mathcal{T}_1 \cup \mathcal{T}_2) \subseteq A \cap B$, hence $A \cap B \in \mathcal{F}$.

2 implies 1: Fix a filter \mathcal{F} on X with $\mathcal{S} \subseteq \mathcal{F}$. Let \mathcal{T} be a finite subset of \mathcal{S} . We then have that \mathcal{T} is a finite subset of \mathcal{F} , hence $\bigcap \mathcal{T} \in \mathcal{F}$ because \mathcal{F} is a filter. Since $\emptyset \notin \mathcal{F}$, it follows that $\bigcap \mathcal{T} \neq \emptyset$. \square

Definition 6.40. Let X be a set. An ultrafilter on X is filter \mathcal{U} on X such that for all $A \subseteq X$, either $A \in \mathcal{U}$ or $X \setminus A \in \mathcal{U}$.

Example. Every principal filter is an ultrafilter. \square

Proposition 6.41. Let \mathcal{F} be a filter on X . \mathcal{F} is an ultrafilter on X if and only if \mathcal{F} is a maximal filter on X (i.e. there is no filter \mathcal{G} on X with $\mathcal{F} \subsetneq \mathcal{G}$).

Proof. Suppose that \mathcal{F} is not a maximal filter on X . Fix a filter \mathcal{G} on X such that $\mathcal{F} \subsetneq \mathcal{G}$. Fix $A \in \mathcal{G} \setminus \mathcal{F}$. Notice that $X \setminus A \notin \mathcal{F}$ because otherwise we would have $X \setminus A \in \mathcal{G}$ and hence $\emptyset = A \cap (X \setminus A) \in \mathcal{G}$, a contradiction. Therefore, $A \notin \mathcal{F}$ and $X \setminus A \notin \mathcal{F}$, so \mathcal{F} is not an ultrafilter on X .

Conversely, suppose that \mathcal{F} is not an ultrafilter on X . Fix $A \in \mathcal{P}(X)$ such that $A \notin \mathcal{F}$ and $X \setminus A \notin \mathcal{F}$. We claim that $\mathcal{F} \cup \{A\}$ has the finite intersection property. Fix a filter \mathcal{G} on X such that $\mathcal{F} \cup \{A\} \subseteq \mathcal{G}$. Since $\mathcal{F} \subsetneq \mathcal{G}$, it follows that \mathcal{F} is not a maximal filter on X . \square

Proposition 6.42. Let \mathcal{F} be a filter on X . There exists an ultrafilter \mathcal{U} on X such that $\mathcal{F} \subseteq \mathcal{U}$.

Proof. Zorn's Lemma. \square

Corollary 6.43. Let X be an infinite set. There exists a nonprincipal ultrafilter on X .

Proof. Let \mathcal{F} be the filter on X consisting of all cofinite subsets of X . Fix an ultrafilter \mathcal{U} on X such that $\mathcal{F} \subseteq \mathcal{U}$. For all $x \in X$, we have $X \setminus \{x\} \in \mathcal{F} \subseteq \mathcal{U}$, hence $\{x\} \notin \mathcal{U}$. \square

6.3.2 Ultraproducts

Ultrafilters (or even just filters) solve our democratic blending problem for relation symbols beautifully. Suppose that $\mathcal{L} = \{\mathbf{R}\}$ where \mathbf{R} is a binary relation symbol and $I = \omega$. Suppose also that \mathcal{U} is an ultrafilter on ω . Given elements $\langle a_i \rangle_{i \in \omega}$ and $\langle b_i \rangle_{i \in \omega}$ of M , we could then say that the pair $(\langle a_i \rangle_{i \in \omega}, \langle b_i \rangle_{i \in \omega})$ is an element of \mathbf{R}^M if the set of indices $i \in I$ such that $(a_i, b_i) \in \mathbf{R}^{M_i}$ is “large”, i.e. if $\{i \in I : (a_i, b_i) \in \mathbf{R}^{M_i}\} \in \mathcal{U}$. Of course, our notion of “large” depends on the ultrafilter, but that flexibility is the beauty of the construction!

However, we have yet to solve the dictatorial problem of function symbols (such as the product of groups in which each is abelian save one ending up nonabelian regardless of what we consider “large”). Wonderfully, and perhaps surprisingly, the ultrafilter can be used in another way to save the day. For concreteness, consider the situation where $\mathcal{L} = \{\mathbf{e}, \mathbf{f}\}$ where \mathbf{e} is a constant symbol and \mathbf{f} is a binary relation symbol, $I = \omega$, and each \mathcal{M}_i is a group. The idea is to flat out ignore variations on “small” sets by considering two sequences $\langle a_i \rangle_{i \in \omega}$ and $\langle b_i \rangle_{i \in \omega}$ to be the same if the set of indices in which they agree is “large”, i.e. if $\{i \in I : a_i = b_i\} \in \mathcal{U}$. In other words, we should define an equivalence relation \sim in this way and take a quotient! This is completely analagous to considering two function $f, g: \mathbb{R} \rightarrow \mathbb{R}$ to be the same if the set $\{x \in \mathbb{R} : f(x) \neq g(x)\}$ has measure 0. What does this solve? Suppose that \mathcal{M}_0 was our rogue nonabelian group, and each \mathcal{M}_i for

$i \neq 0$ was an abelian group. Suppose also that $\omega \setminus \{0\} \in \mathcal{U}$ (i.e. our ultrafilter is not the principal ultrafilter generated by $\{0\}$, and thus we are considering $\{0\}$ to be a “small” set). Given a sequence $\langle a_i \rangle_{i \in \omega}$, let $[\langle a_i \rangle_{i \in \omega}]$ be the equivalence class of $\langle a_i \rangle_{i \in \omega}$ under the relation. Assuming that everything is well-defined (see below), we then have that $\langle f^{\mathcal{M}_i}(a_i, b_i) \rangle_{i \in \omega} \sim \langle f^{\mathcal{M}_i}(b_i, a_i) \rangle_{i \in \omega}$ and so

$$\begin{aligned} f^{\mathcal{M}}([\langle a_i \rangle_{i \in \omega}], [\langle b_i \rangle_{i \in \omega}]) &= [\langle f^{\mathcal{M}_i}(a_i, b_i) \rangle_{i \in \omega}] \\ &= [\langle f^{\mathcal{M}_i}(b_i, a_i) \rangle_{i \in \omega}] \\ &= f^{\mathcal{M}}([\langle b_i \rangle_{i \in \omega}], [\langle a_i \rangle_{i \in \omega}]) \end{aligned}$$

and so we have saved abelianess by ignoring problems on “small” sets!

To summarize before launching into details, here’s the constuction. Start with a language \mathcal{L} , a set I , and \mathcal{L} -structures \mathcal{M}_i for each $i \in I$. Form the product $\prod_{i \in I} \mathcal{M}_i$, but take a quotient by considering two elements of this product to be equivalent if the set of indices on which they agree is “large”. Elements of our structure are now equivalence classes, so we need to worry about things being well-defined, but the fundamental idea is to interpret constant symbols and functions componentwise, and interpret relation symbols by saying that that an k -tuple is in the interpretation of some $R \in \mathcal{R}_k$ if the set of indices on which the corresponding k -tuple is in $R^{\mathcal{M}_i}$ is “large”. Amazingly, this process behave absolutely beautifully with regards to first-order logic. For example, if we denote this “blended” structure by \mathcal{M} , we will prove below that for any $\sigma \in \text{Sent}_{\mathcal{L}}$ we have

$$\mathcal{M} \models \sigma \text{ if and only if } \{i \in I : \mathcal{M}_i \models \sigma\} \in \mathcal{U}$$

That is, an arbitrary sentence σ is true in the “blended” stucture if and only if the set of indices $i \in I$ in which σ is true in \mathcal{M}_i is “large”!

Onward to the details. The notation is painful and easy to get lost in, but keep the fundamental ideas in mind and revert to thinking of $I = \omega$ whenever the situation looks hopelessly complicated. First we have the proposition saying that the \sim defined in this way is an equivalence relation and that our definitions are well-defined.

Proposition 6.44. *Let I be a set, and suppose that for each $i \in I$ we have an \mathcal{L} -structure \mathcal{M}_i . Let \mathcal{U} be an ultrafilter on I . Define a relation \sim on $\prod_{i \in I} \mathcal{M}_i$ by saying that $g \sim h$ if $\{i \in I : g(i) = h(i)\} \in \mathcal{U}$.*

1. \sim is an equivalence relation on $\prod_{i \in I} \mathcal{M}_i$.
2. Suppose that $g_1, g_2, \dots, g_k, h_1, h_2, \dots, h_k \in \prod_{i \in I} \mathcal{M}_i$ are such that $g_j \sim h_j$ for all j .
 - (a) $\{i \in I : (g_1(i), g_2(i), \dots, g_k(i)) = (h_1(i), h_2(i), \dots, h_k(i))\} \in \mathcal{U}$.
 - (b) For each $R \in \mathcal{R}_k$, the following are equivalent
 - $\{i \in I : (g_1(i), g_2(i), \dots, g_k(i)) \in R^{\mathcal{M}_i}\} \in \mathcal{U}$.
 - $\{i \in I : (h_1(i), h_2(i), \dots, h_k(i)) \in R^{\mathcal{M}_i}\} \in \mathcal{U}$.
 - (c) For each $f \in \mathcal{F}_k$, we have $\{i \in I : f^{\mathcal{M}_i}(g_1(i), g_2(i), \dots, g_k(i)) = f^{\mathcal{M}_i}(h_1(i), h_2(i), \dots, h_k(i))\} \in \mathcal{U}$.

Proof. □

Definition 6.45. *Let I be a set, and suppose that for each $i \in I$ we have an \mathcal{L} -structure \mathcal{M}_i . Let \mathcal{U} be an ultrafilter on I . We define an \mathcal{L} -structure $\mathcal{M} = \prod_{i \in I} \mathcal{M}_i / \mathcal{U}$ as follows. Define the relation \sim on $\prod_{i \in I} \mathcal{M}_i$ as above, and we let the universe of \mathcal{M} be the corresponding quotient. We interpret the symbols of \mathcal{L} as follows.*

1. For each $c \in \mathcal{C}$, let $c^{\mathcal{M}} = [i \mapsto c^{\mathcal{M}_i}]$.
2. For each $R \in \mathcal{R}_k$, let $R^{\mathcal{M}} = \{([g_1], [g_2], \dots, [g_k]) \in M^k : \{i \in I : (g_1(i), g_2(i), \dots, g_k(i)) \in R^{\mathcal{M}_i}\} \in \mathcal{U}\}$.

3. For each $f \in \mathcal{F}_k$, let $f^{\mathcal{M}}([g_1], [g_2], \dots, [g_k]) = [i \mapsto f^{\mathcal{M}_i}(g_1(i), g_2(i), \dots, g_k(i))]$.

We call \mathcal{M} the ultraproduct of the \mathcal{M}_i over the ultrafilter \mathcal{U} .

Definition 6.46. In the above situation, given variable assignments $s_i: \text{Var} \rightarrow M_i$ for each $i \in I$, we let $\langle s_i \rangle_{i \in I}$ denote the variable assignment $\text{Var} \rightarrow M$ given by $\langle s_i \rangle_{i \in I}(\mathbf{x}) = [i \mapsto s_i(\mathbf{x})]$.

Lemma 6.47. Let \mathcal{L} be a language, let I be a set, and let \mathcal{U} be an ultrafilter on I . Suppose that for each $i \in I$, we have an \mathcal{L} -structure \mathcal{M}_i , and let $\mathcal{M} = \prod_{i \in I} \mathcal{M}_i / \mathcal{U}$. For all $t \in \text{Term}_{\mathcal{L}}$ and all $s_i: \text{Var} \rightarrow M_i$, we have

$$\overline{\langle s_i \rangle_{i \in I}}(t) = [i \mapsto \overline{s_i}(t)]$$

In other words, for all $t(x_1, x_2, \dots, x_k) \in \text{Term}_{\mathcal{L}}$ and all $g_1, g_2, \dots, g_k \in \prod_{i \in I} M_i$, we have

$$t^{\mathcal{M}}([g_1], [g_2], \dots, [g_k]) = [i \mapsto t^{\mathcal{M}_i}(g_1(i), g_2(i), \dots, g_k(i))]$$

Proof. Suppose that $\mathbf{c} \in \mathcal{C}$. Let $s_i: \text{Var} \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} \overline{\langle s_i \rangle_{i \in I}}(\mathbf{c}) &= \mathbf{c}^{\mathcal{M}} \\ &= [i \mapsto \mathbf{c}^{\mathcal{M}_i}] \\ &= [i \mapsto \overline{s_i}(\mathbf{c})] \end{aligned}$$

Suppose that $\mathbf{x} \in \text{Var}$. Let $s_i: \text{Var} \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} \overline{\langle s_i \rangle_{i \in I}}(\mathbf{x}) &= \langle s_i \rangle_{i \in I}(\mathbf{x}) \\ &= [i \mapsto s_i(\mathbf{x})] \\ &= [i \mapsto \overline{s_i}(\mathbf{x})] \end{aligned}$$

Suppose that $f \in \mathcal{F}_k$ and $t_1, t_2, \dots, t_k \in \text{Term}_{\mathcal{L}}$ are such that the result holds for the t_i . Let $s_i: \text{Var} \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} \overline{\langle s_i \rangle_{i \in I}}(ft_1 t_2 \cdots t_k) &= f^{\mathcal{M}}(\overline{\langle s_i \rangle_{i \in I}}(t_1), \overline{\langle s_i \rangle_{i \in I}}(t_2), \dots, \overline{\langle s_i \rangle_{i \in I}}(t_k)) \\ &= f^{\mathcal{M}}([i \mapsto \overline{s_i}(t_1)], [i \mapsto \overline{s_i}(t_2)], \dots, [i \mapsto \overline{s_i}(t_k)]) \\ &= [i \mapsto f^{\mathcal{M}_i}(\overline{s_i}(t_1), \overline{s_i}(t_2), \dots, \overline{s_i}(t_k))] \\ &= [i \mapsto \overline{s_i}(ft_1 t_2 \cdots t_k)] \end{aligned}$$

□

Theorem 6.48. Let \mathcal{L} be a language, let I be a set, and let \mathcal{U} be an ultrafilter on I . Suppose that for each $i \in I$, we have an \mathcal{L} -structure \mathcal{M}_i , and let $\mathcal{M} = \prod_{i \in I} \mathcal{M}_i / \mathcal{U}$. For all $\varphi \in \text{Form}_{\mathcal{L}}$ and all $s_i: \text{Var} \rightarrow M_i$, we have

$$(\mathcal{M}, \langle s_i \rangle_{i \in I}) \models \varphi \text{ if and only if } \{i \in I : (\mathcal{M}_i, s_i) \models \varphi\} \in \mathcal{U}$$

In other words, for all $\varphi(x_1, x_2, \dots, x_k) \in \text{Form}_{\mathcal{L}}$ and all $g_1, g_2, \dots, g_k \in \prod_{i \in I} M_i$, we have

$$(\mathcal{M}, [g_1], [g_2], \dots, [g_k]) \models \varphi \text{ if and only if } \{i \in I : (\mathcal{M}_i, g_1(i), g_2(i), \dots, g_k(i)) \models \varphi\} \in \mathcal{U}$$

In particular, for any $\sigma \in \text{Sent}_{\mathcal{L}}$, we have

$$\mathcal{M} \models \sigma \text{ if and only if } \{i \in I : \mathcal{M}_i \models \sigma\} \in \mathcal{U}$$

Proof. The proof is by induction.

Suppose that $t_1, t_2 \in Term_{\mathcal{L}}$. Let $s_i: Var \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models t_1 t_2 &\Leftrightarrow \overline{\langle s_i \rangle_{i \in I}}(t_1) = \overline{\langle s_i \rangle_{i \in I}}(t_2) \\ &\Leftrightarrow [i \mapsto \overline{s_i}(t_1)] = [i \mapsto \overline{s_i}(t_2)] \\ &\Leftrightarrow \{i \in I : \overline{s_i}(t_1) = \overline{s_i}(t_2)\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models t_1 t_2\} \in \mathcal{U} \end{aligned}$$

Suppose that $R \in \mathcal{R}_k$ and $t_1, t_2, \dots, t_k \in Term_{\mathcal{L}}$. Let $s_i: Var \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models Rt_1 t_2 \cdots t_k &\Leftrightarrow (\overline{\langle s_i \rangle_{i \in I}}(t_1), \overline{\langle s_i \rangle_{i \in I}}(t_2), \dots, \overline{\langle s_i \rangle_{i \in I}}(t_k)) \in R^{\mathcal{M}} \\ &\Leftrightarrow ([i \mapsto \overline{s_i}(t_1)], [i \mapsto \overline{s_i}(t_2)], \dots, [i \mapsto \overline{s_i}(t_k)]) \in R^{\mathcal{M}} \\ &\Leftrightarrow \{i \in I : (\overline{s_i}(t_1), \overline{s_i}(t_2), \dots, \overline{s_i}(t_k)) \in R^{\mathcal{M}_i}\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models Rt_1 t_2 \cdots t_k\} \in \mathcal{U} \end{aligned}$$

Suppose that the result holds for φ and ψ . Let $s_i: Var \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models \varphi \wedge \psi &\Leftrightarrow (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models \varphi \text{ and } (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models \psi \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models \varphi\} \in \mathcal{U} \text{ and } \{i \in I : (\mathcal{M}_i, s_i) \models \psi\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models \varphi \text{ and } (\mathcal{M}_i, s_i) \models \psi\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models \varphi \wedge \psi\} \in \mathcal{U} \end{aligned}$$

Suppose that the result holds for φ . Let $s_i: Var \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models \neg\varphi &\Leftrightarrow (\mathcal{M}, \langle s_i \rangle_{i \in I}) \not\models \varphi \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models \varphi\} \notin \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \not\models \varphi\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models \neg\varphi\} \in \mathcal{U} \end{aligned}$$

Suppose that the result holds for φ . Let $s_i: Var \rightarrow M_i$ be variable assignments. We then have

$$\begin{aligned} (\mathcal{M}, \langle s_i \rangle_{i \in I}) \models \exists y \varphi &\Leftrightarrow \text{There exists } a \in M \text{ such that } (\mathcal{M}, \langle s_i \rangle_{i \in I} [y \Rightarrow a]) \models \varphi \\ &\Leftrightarrow \text{There exists } g \in \prod_{i \in I} M_i \text{ such that } (\mathcal{M}, \langle s_i \rangle_{i \in I} [y \Rightarrow [g]]) \models \varphi \\ &\Leftrightarrow \text{There exists } g \in \prod_{i \in I} M_i \text{ such that } (\mathcal{M}, \langle s_i [y \Rightarrow g(i)] \rangle_{i \in I}) \models \varphi \\ &\Leftrightarrow \text{There exists } g \in \prod_{i \in I} M_i \text{ such that } \{i \in I : (\mathcal{M}_i, s_i [y \Rightarrow g(i)]) \models \varphi\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : \text{There exists } a \in M_i \text{ such that } (\mathcal{M}_i, s_i [y \Rightarrow a]) \models \varphi\} \in \mathcal{U} \\ &\Leftrightarrow \{i \in I : (\mathcal{M}_i, s_i) \models \exists y \varphi\} \in \mathcal{U} \end{aligned}$$

□

Theorem 6.49. *If every finite subset of Σ has a model, then Σ has a model.*

Proof. Let I be the set of all finite subsets of Σ . For each $\Psi \in I$, fix a model \mathcal{M}_{Ψ} of Ψ . For each $\sigma \in \Sigma$, let $A_{\sigma} = \{\Psi \in I : \sigma \in \Psi\}$. Let $\mathcal{S} = \{A_{\sigma} : \sigma \in \Sigma\} \subseteq \mathcal{P}(I)$ and notice that \mathcal{S} has the finite intersection property because

$$\{\sigma_1, \sigma_2, \dots, \sigma_n\} \in A_{\sigma_1} \cap A_{\sigma_2} \cap \cdots \cap A_{\sigma_n}$$

Fix an ultrafilter \mathcal{U} on I such that $\mathcal{S} \subseteq \mathcal{U}$ and let $\mathcal{M} = \prod_{\Psi \in I} \mathcal{M}_\Psi / \mathcal{U}$. For any $\sigma \in \Sigma$, we then have that $A_\sigma \subseteq \{\Psi \in I : \mathcal{M}_\Psi \models \sigma\}$, hence $\{\Psi \in I : \mathcal{M}_\Psi \models \sigma\} \in \mathcal{U}$, and so $\mathcal{M} \models \sigma$. Therefore, \mathcal{M} is a model of Σ . \square