Notes for MATH 582

David J. Fernández-Bretón

(based on the handwritten notes from January-April 2016,
typed in January-April 2017,
polished in January-April 2018)

Table of Contents

Chapter 1

Basic set theory

Set theory is sometimes considered to be a branch of Mathematical Logic. In particular, it might be relevant not just for mathematics, but for computer science and philosophy as well. It is often said that set theory was born "on that December of 1873 in which Cantor proved that there are uncountably many reals". It is certainly true that Cantor must have been one of the first few people which considered abstract collections of mathematical objects at the level of abstraction that we now do in set theory. Over the course of the years, set theory has blossomed and developed to the point where it now constitutes a whole branch of mathematics on its own. This course is an introduction to this exciting area of mathematics.

This course has two main objectives. The first is to convince students that every mathematical object (including: functions, (real and complex) numbers, points in n-dimensional space, polynomials, and virtually everything that you can think of) can be viewed as a set of some sort (some people like to say that "everything *is* a set", although I consider it to be more accurate if we say that everything *can be implemented* within set theory). Thus we will spend the first half of this course introducing the commonly accepted axioms of set theory, and explaining how, starting from these axioms, it is possible to implement most everyday mathematical objects by defining appropriate sets that behave, when interpreted in the appropriate way, as we would intuitively expect these objects to behave. The culmination of this will be the construction of the Real Line $\mathbb{R}$. The second objective of this course is to introduce the student to a few basic topics of set theory proper, most notably Ordinal Numbers (and Transfinite Induction and Recursion), Infinite Cardinal Arithmetic (with some level of sophistication, beyond the elementary and simple-minded distinction between countable and uncountable) and some equivalences and consequences of the Axiom of Choice (including applications to other areas of Mathematics). These topics will roughly constitute the second half of the course. Recently, I have started referring to the topics from the first half as the "foundational" part of set theory, whereas the topics from the second half constitute a fragment of the "mathematical" part of set theory.

1.1 What is a Set?

Informally, we can think of a set as some collection of objects, called its *elements*, where the word "collection" is intended in a very abstract way. A set is usually written either explicitly as the list of its elements (e.g. the set $\{6, 28\}$), or as a description of what its elements look like (e.g. the set $\{n|n \text{ is a perfect number and } n \leq 100\}$). The special symbol $\in$ is introduced as a binary relation, stating that the object to the left of the symbol is an element of the set to the right of it: for example, $6 \in \{6, 28\}$.

Formally speaking, though, it is not necessary to explicitly write down any definition of what a set is: a set would just be any object belonging to set theory, and all that is therefore relevant is that these objects behave like the axioms of set theory say that they do. In other words, the axioms themselves play, to a certain extent, the rôle of definitions (in the same sense that, for example, the group axioms constitute the very definition of what a group is, rather than being some "obvious" statements about some predetermined mathematical object). So the really important part, to begin with our study of sets, is to explicitly state what are the axioms of set theory.

1.2 Axiomatizing set theory

There are a number of possible axiomatizations of set theory that have been proposed over the course of the years. The main ones among these are:

- The axioms of Zermelo-Fraenkel, along with the Axiom of Choice (abbreviated ZFC),
- the axioms of von Neumann-Bernays-Gödel (abbreviated NBG),

- the axioms of Morse-Kelley (abbreviated MK),
- Quine's New Foundations (abbreviated NF),
- the modification of NF that allows objects known as *urelements* (abbreviated NFU).

Mostly for historical reasons, the axiom system that is currently the most widespread and commonly accepted is ZFC, and so these are the axioms that we will use in this course, although I will briefly mention the other axiom systems, and how they differ from ZFC, in appropriate moments throughout the course.

So in theory, we should now proceed to state what the ZFC axioms are. However, the list of ZFC axioms consists of seven axioms and two axiom schemas (resulting in infinitely many axioms overall), which might look quite arbitrary at first sight. So in order to motivate these axioms, we will first spend some time working in what I like to call "Cantor's set theory", see what are its advantages and also its disadvantages, and what problems arise from this way of doing set theory. After this, the ZFC axioms should (for the most part) look quite natural.

What is Cantor's set theory, then? Cantor used to do set theory based on only two very basic principles. The first one is the following:

Axiom 1 (Principle of Extensionality)**.** *A set is determined by its elements.*

This means that, if A and B are two sets that have the exact same elements, then they are actually the same set, symbolised $A = B$. This principle is central to the concept of a set, as it explicitly expresses that the identity of a set is given only by the objects that are elements of this set, irrespective of any other property (such as the order in which its elements are written, or thought of; or the number of times that we write such elements). For example, it is based on this principle that we can conclude that the following four sets are actually the same:

$$\{6, 28\} = \{28, 6\} = \{6, 28, 6\} = \{n \big| n \text{ is a perfect number and } n \leq 100\}.$$

The second basic principle in Cantor's set theory is the following:

Axiom 2 (Principle of Set Creation)**.** *Given any property, there exists a set whose elements are exactly those objects that satisfy the given property.*

Thus, if we denote a property of some object x by $P(x)$, then the Principle of Set Creation ensures the existence of a set A such that $x \in A$ if and only if $P(x)$. As we said before, this set is denoted by $\{x \big| P(x)\}$, and the Principle of Extensionality ensures that it is unique. For example, if the property $P(x)$ is "x is a prime number and $x \leq 30$", then the Principle of Set Creation ensures the existence of the set

$$\{x \big| P(x)\} = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29\}.$$

1.3 The Language of set theory

Now, there are two problems with Cantor's set theory, which we will mention in reverse (according to when they were historically pointed out) order. So for us, the first problem (although historically it was the second) is that the notion of "property" has not been properly defined, and it turns out that this notion can bring serious trouble, if used without care. If we just allow any English expression, then we might end up considering properties like the one below:

Paradox 1. *Consider the property $P(x)$ given by*

x is the least natural number that cannot be described with less than twenty words.

This property is contradictory.

Proof. Since the English language has only finitely many words[1], there are only finitely many sequences of less than twenty words (there is a very big number of them, but that number is finite), of which some will be actual descriptions of natural numbers (others will be meaningful, but not actual descriptions of a specific natural number; and most of them will be ~~garbage~~ meaningless). Hence there are only finitely many natural numbers that can be described with less than twenty words, but there are infinitely many natural numbers overall and so there must exist at least one natural number that cannot be described in less than twenty words. Thus we should be able to take the least such number, but clearly "the least natural number that cannot be described with less than twenty words" constitutes a description of this number, that employs less than twenty words, and this is a contradiction (this is known as "Berry's paradox"). □

[1] I have had at least one person complain about this assumption, because new words get added to the language all the time, and since there's no limit in the number of letters that words can have, we could think that the English language actually has (at least potentially) infinitely many words. If this is something that concerns you, you can substitute the above sentence by the sentence "x is the least natural number that cannot be described with less than thirty words that appeared in print before the end of the year 2016", or else by the sentence "x is the least natural number that can be described in less than 100 ASCII characters" (the latter will work even if we allow for things like LaTeX code for mathematical expressions within our descriptions).

It looks like this contradiction arises from the fact that English can reference many things, including pieces of the language itself[2]. The most elegant way out of this dilemma, which was proposed by Skolem (in 1923), is to consider a language where there is a clear distinction between the objects that the language talks about, and the language itself; a language so restricted that one cannot speak about the language itself by using it. The language that we will use for this purpose is known as *the Language of Set Theory* (which is a variety of what is known in Logic as a *first-order language*), which will be abbreviated as $\mathscr{L}_{\mathrm{ST}}$. It is given as follows:

Definition 1.

- The **alphabet** of $\mathscr{L}_{\mathrm{ST}}$ is the collection of the following symbols:
  - *Logical connectives:* $\wedge, \vee, \neg, \Rightarrow, \iff$;
  - *quantifiers:* $\forall, \exists$;
  - *variables:* $a, b, c, \dots, A, B, C, \dots, \alpha, \beta, \gamma, \dots$;
  - *parentheses:* $), (,], [$;
  - *specifically set-theoretic symbols:* $=$ and $\in$.
- A **formula** of $\mathscr{L}_{\mathrm{ST}}$, or $\mathscr{L}_{\mathrm{ST}}$**-formula** (which eventually we will just call **formula**, for short) is a finite string (or sequence) of symbols of the alphabet of $\mathscr{L}_{\mathrm{ST}}$ that makes sense (in that we can make sense of the question of whether it is true or false).

Clearly the definition of formula above is not an entirely formal one. It is possible to formally define, by induction of the length of a string of symbols, specifically which such strings constitute formulas and which ones do not (those who took Math 481 know this too well!); but for this course, the intuitive notion of "string of symbols that makes sense" will do: for example,

$$(\forall x)(\exists y)(x = y \iff x \in y)$$

makes sense, and hence is a formula of $\mathscr{L}_{\mathrm{ST}}$; whereas

$$x(\forall y)))\neg(x = \wedge)(((
$$

does not make sense, and hence it is not a formula of $\mathscr{L}_{\mathrm{ST}}$.

Intuitively, the semantic idea is that our universe of discourse contains only sets. Variables stand for sets and sets only, and quantifiers range over sets only (that is, $(\forall x)$ means "for every **set** x" and $(\exists x)$ means "there exists a **set** x such that"), and so it is not possible for an $\mathscr{L}_{\mathrm{ST}}$-formula to make reference to another formula, or to any object extraneous to set theory, but rather only to sets. On the flip side, every object that we talk about when doing set theory will be a set, and this includes also the objects that happen to be elements of other sets. So every set that we consider in this course will have the property that all of its elements are also sets (whose elements, in turn, are also sets, and so on; intuitively we say that we will only work with *hereditary sets*).

In spite of its simplicity, the Language of set theory will turn out to be quite powerful, as everything that we want to say about sets (and hence, because of what we said at the beginning, about essentially every mathematical object) is something that we will be able to say by means of an $\mathscr{L}_{\mathrm{ST}}$-formula. These formulas can get quite lengthy very fast, so we will frequently introduce new symbols or English expressions as abbreviations of longer $\mathscr{L}_{\mathrm{ST}}$-formulas (and we typically call these introductions of abbreviations *definitions*). We can think of these abbreviations (or definitions) as macros. For example,

Definition 2.

1. any expression of the form $A \neq B$ will be thought of as an abbreviation of the $\mathscr{L}_{\mathrm{ST}}$-formula $\neg(A = B)$,
2. expressions of the form $x \notin X$ will be thought of as abbreviations of $\neg(x \in X)$,
3. the expression $A \subseteq B$ is an abbreviation of $(\forall x)(x \in A \Rightarrow x \in B)$,
4. the expression $A \nsubseteq B$ is an abbreviation of $\neg(A \subseteq B)$.

Thus we can now write $\mathscr{L}_{\mathrm{ST}}$-formulas using not only the symbols that properly belong to the alphabet of $\mathscr{L}_{\mathrm{ST}}$, but also the symbols $\neq, \notin, \subseteq, \nsubseteq$. The idea is that every time we encounter one of these symbols, we know how to replace it by another (typically much longer) expression so that in the end the formula ends up being an "actual" $\mathscr{L}_{\mathrm{ST}}$-formula (that is, one that only uses symbols that belong to the alphabet of $\mathscr{L}_{\mathrm{ST}}$). Most of the time, we do not care to actually do

[2] For further examples of this phenomenon, consider the sentence "This sentence is false", or the pair of sentences "The following sentence is true. The preceding sentence is false". Another cute example would be the meme where Pinocchio says "my nose will grow now".

the replacement (or "macro expansion"); rather, we only care that in theory, we would be able to do such replacement if we wanted to. We will keep defining new abbreviations, and abbreviations over abbreviations (note, for example, that if we expand the expression $A \nsubseteq B$ we obtain another expression which still needs to be expanded once more before it can actually be a legit $\mathscr{L}_{\mathrm{ST}}$-formula), throughout the course. This way, we will very quickly wind up being able to state our theorems in plain English, the way we usually do in other courses, although this will not mean that we are back to the situation where every English expression (even vague or contradictory ones) is acceptable. Rather, we will be using a carefully curated version of English, one in which we will always know, at least in theory, that everything we say could be expanded to eventually be an actual $\mathscr{L}_{\mathrm{ST}}$-formula (although we certainly would not want to actually see, let alone parse, such a formula, as it is bound to be unbearably long). We can think of this "careful" version of "mathematical English" as a "user-friendly interface" for doing set theory, as opposed to the actual (fully expanded) $\mathscr{L}_{\mathrm{ST}}$-formulas, which are more like "machine language".

1.4 Cantor's set theory

Hence, we will vow from now on that we will stick to $\mathscr{L}_{\mathrm{ST}}$-formulas in order to express facts about sets. In particular, we will state again the two principles of Cantor's set theory as $\mathscr{L}_{\mathrm{ST}}$-formulas. The first principle is easy:

Axiom 1 (Extensionality)**.**

$$(\forall A, B)[(\forall x)(x \in A \iff x \in B) \Rightarrow A = B]$$

Given the meaning of the symbol $\subseteq$, it is easy to see that another equivalent way of stating the Axiom of Extensionality above is by means of the following $\mathscr{L}_{\mathrm{ST}}$-formula:

$$(\forall A, B)[(A \subseteq B) \wedge (B \subseteq A) \Rightarrow A = B],$$

which explicitly justifies the most common method of proving that two sets are equal. Also notice that the last implication in the above formula can very easily be replaced by a biconditional, as the implication $A = B \Rightarrow (\forall x)(x \in A \iff x \in B)$ is logically valid (this much should be intuitively very clear, and those who took Math 481 should be able to actually provide a formal proof of it!).

This is probably a good moment to make a parenthesis and emphasize the difference between $\in$ and $\subseteq$ (which should really be obvious, since $\in$ is a primitive symbol, whereas $A \subseteq B$ is an abbreviation of $(\forall x)(x \in A \Rightarrow x \in B)$, but I know that students sometimes get confused at first), as well as the difference that brackets make. For example, $\varnothing \neq \{\varnothing\}$, since the left-hand side contains no elements, whereas the right-hand side contains one element (namely the set $\varnothing$). For the exact same reason, $\varnothing \neq \{\{\varnothing\}\}$ (the left-hand side contains no elements, the right-hand side contains one, namely $\{\varnothing\}$. Now, it is also the case that $\{\varnothing\} \neq \{\{\varnothing\}\}$: both sets under comparison contain exactly one element, but the unique element of the left-hand side is $\varnothing$, and the unique element of the right-hand side is $\{\varnothing\}$, and we already saw that $\varnothing \neq \{\varnothing\}$. Now, it should be clear that $\varnothing \notin \varnothing$ (in fact, $X \notin \varnothing$ for every X), whereas $\varnothing \subseteq \varnothing$ (because the implication $(\forall x)(x \in \varnothing \Rightarrow x \in \varnothing)$ is vacuously true), in fact, $\varnothing \subseteq X$ for every X. As another example, $\{\varnothing\} \in \{\{\varnothing\}\}$ but $\{\varnothing\} \nsubseteq \{\{\varnothing\}\}$ (why? can anyone tell me?). Note also that, for every set X, it is the case that $X \subseteq X$ (whereas it would be really weird if we had $X \in X$, although we still don't have the tools to rule out this possibility!!!).

In order to state Cantor's second principle, we still need to explain one more thing about $\mathscr{L}_{\mathrm{ST}}$, namely what it means for a variable to occur free in a formula.

Definition 3. If φ is an $\mathscr{L}_{\mathrm{ST}}$-formula, and x is a variable, we say that x **occurs free** in φ if x is not affected by any quantifier in φ. When we want to emphasize that x occurs free in φ, we will typically write $\varphi(x)$.

(Note that this definition is not a definition within $\mathscr{L}_{\mathrm{ST}}$ (in the sense that it is not a specification for a new abbreviation), but rather a definition *about* $\mathscr{L}_{\mathrm{ST}}$. As such, it belongs to the *metatheory* that we use in order to do set theory, rather than to set theory proper.)

Once again, the above definition is not a completely formal one, although those who took Math 481 should know that it is possible to make it fully formal. For this course, however, the intuitive notion will be enough. For example, the variable x is not free in the formula

$$(\forall x)(x \in A),$$

because it is affected (*bound*) by the $\forall$, whereas the same variable x occurs free in the formula

$$(\exists A)(x \in A)$$

because there is no quantifier affecting x. Intuitively speaking, a variable x occurs free in the formula φ if $\varphi(x)$ looks like the template of a statement, which can be sometimes true and sometimes false depending of what object we "plug in" instead of x. On the other hand, when a variable occurs, but not free, in φ, then it should look more like a "dummy variable", in the sense that consistently replacing all occurences of that variable (that are bound by the corresponding quantifier, including the occurence right next to the quantifier) with any other variable should not change the meaning of the formula.

Axiom 2 (Unrestricted Comprehension)**.** *Let $\varphi(x)$ be an $\mathscr{L}_{\mathrm{ST}}$-formula with the free variable x. Then*

$$(\exists A)(\forall x)(x \in A \iff \varphi(x))$$

Note that, strictly speaking, the axiom above is not really an axiom, but an *axiom schema*, that is, a formula for building axioms (one axiom per each $\mathscr{L}_{\mathrm{ST}}$-formula with one free variable, $\varphi(x)$), and so the whole of Cantor's set theory consists now of infinitely many axioms. This is fine, however, as it is possible to algorithmically detect, given an $\mathscr{L}_{\mathrm{ST}}$-formula ψ, whether it constitutes an instance of this Axiom Schema or not (one only needs to chech whether ψ is of the form $(\exists A)(\forall x)(x \in A \iff \varphi(x))$, for some $\mathscr{L}_{\mathrm{ST}}$-formula φ with one free variable). In technical terms, we say that the set of axioms of Cantor's set theory is *recursive*[3]. Recursiveness is something that we demand of any axiom system, for historical reasons, since Hilbert's strong advocacy of the axiomatic method relied on the possibility of mechanically checking whether a proof is correct or not, which as a particular case implies that it should possible to mechanically check whether a formula invoked within a proof is an axiom or not.

As before, given an $\mathscr{L}_{\mathrm{ST}}$-formula $\varphi(x)$, we denote the set whose existence is ensured by the corresponding instance of this axiom scheme by $\{x \big| \varphi(x)\}$, which is unique by the Axiom of Extensionality. In practice, then, in order to prove that a given set exists, all we need to do is to explicitly write such set in the form $\{x \big| \varphi(x)\}$, carefully enough that $\varphi(x)$ constitutes a legitimate $\mathscr{L}_{\mathrm{ST}}$-formula. For example, we can prove the following theorem:

Theorem 4. *In Cantor's set theory, one can prove the following:*

1. *There exists a unique set containing all sets as elements. We define V to stand for such set,*

2. *there exists a unique set with no elements. We define $\varnothing$ to stand for such set,*

3. *there exists a unique set whose only element is the set $\varnothing$ defined above. We define $\{\varnothing\}$ to denote such set,*

4. *if we are given finitely many sets $x_1, \dots, x_n$, then there exists a unique set whose only elements are $x_1, \dots, x_n$. We define $\{x_1, \dots, x_n\}$ to denote such set;*

5. *there exists a unique set whose only element is the set $\{\varnothing\}$ (whose only element, in turn, is $\varnothing$). We define $\{\{\varnothing\}\}$ to denote this set,*

6. *if we are given two sets A and B, then there exists a unique set whose elements are exactly those sets that belong to either A or B. We define $A \cup B$ (the union of A and B) to denote such set,*

7. *if we are given two sets A and B, then there exists a unique set whose elements are exactly those sets that belong to both A and B. We define $A \cap B$ (the intersection of A and B) to denote such set,*

8. *if we are given two sets A and B, then there exists a unique set whose elements are exactly those sets that belong to A, but not to B. We define $A \setminus B$ (the set difference of A minus B) to denote such set.*

9. *given a set A, there exists a unique set whose elements are exactly those sets that are subsets of A. We define $\mathfrak{P}(A)$ (the powerset of A) to denote such set.*

10. *given a set A, there exists a unique set whose elements are exactly those sets that do not belong to A ("the complement of A").*

Proof. In all cases, uniqueness follows from the Axiom of Extensionality. Now for existence, just apply the corresponding instances of the Axiom Scheme of Comprehension:

1. $V = \{x \big| x = x\}$,

2. $\varnothing = \{x \big| x \neq x\}$,

3. $\{\varnothing\} = \{x \big| x = \varnothing\}$,

4. $\{x_1, \dots, x_n\} = \{x \big| x = x_1 \vee \dots \vee x = x_n\}$,

5. this is a particular case of item 4 above, with $n = 1$ and $x_1 = \{\varnothing\}$,

[3] A set is *recursive* if there exists an algorithm that, upon being given any input, halts in finite time and correctly answers whether or not the given input belongs to the set.

6. $A \cup B = \{x \big| x \in A \vee x \in B\}$,

7. $A \cap B = \{x \big| x \in A \wedge x \in B\}$,

8. $A \setminus B = \{x \big| x \in A \wedge x \notin B\}$,

9. $\mathfrak{P}(A) = \{X \big| X \subseteq A\}$.

10. by part 8, the set $V \setminus A$ exists.

(recall that every time we "define" a symbol to denote a given set, we are setting up a new macro that can be expanded to an $\mathscr{L}_{\mathrm{ST}}$-formula: if we "define" A as $A = \{x \big| \varphi(x)\}$, it means that the expression $y = A$ is an abbreviation of $(\forall x)(x \in y \iff \varphi(x))$, and the expression $x \in A$ is an abbreviation of $\varphi(x)$. For example, the full expansion of item 3 above would be given by $\{\varnothing\} = \{y \big| (\forall x)(x \in y \iff x \neq x)\}$. In item 4 we are assuming that we have already defined appropriate macros for $x_1, \dots, x_n$; in items 6–8 our assumption is that we have already defined macros for A and B, and in item 9 we should already have available a macro for A.) □

All in all, Cantor's set theory looks great: it allows us prove the existence of sets that we intuitively feel should exist, and so far it looks like it succeeds at allowing us to do set theory in the way that we are used to (when we are at the pre-axiomatic stage, i.e. in the style of the"Joy of Sets" pamphlet). In fact, it could be argued that a very basic desideratum of any axiomatization of set theory is that it should allow us (at the very least) to prove all ten items in Theorem 4. It turns out, however, that there is a sense in which this is too much to ask. In other words, Cantor's set theory is "too good to be true", because it is inconsistent. There is a sense in which all the axiomatizations of set theory that have been proposed, ever since Cantor, have been an attempt to keep as much as possible from Theorem 4, while at the same time remaining consistent and preserving as much as possible of the naïve intuition that arises from Cantor's view of set theory. Each of these axiomatizations has had to forgo either one or two of those ten items, or some significant part of the naïve intuition.

Paradox 2 (Russell's Paradox)**.** *Consider the set*

$$W = \{x \big| x \notin x\}.$$

This leads to a contradiction.

Proof. In Cantor's set theory, an instance of the Axiom Schema of Comprehension ensures that W exists. But then, it should be possible to decide, for any set x, whether it is the case that $x \in W$ or not. In particular, taking $x = W$, we can see that, by definition, $W \in W$ must mean that $W \notin W$, and $W \notin W$ means that $\neg(W \notin W)$, i.e. $W \in W$. So in either case we obtain a contradiction. □

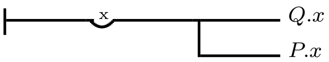
This paradox (the second problem with Cantor's set theory, although historically it was the first) was pointed out by Russell (who apparently discovered it in 1901) to Frege in 1902, who was using a slightly different axiom system for his set theory (this is an interesting story, because Frege actually interrupted the printing of his second book (Volume II of the *Begriffsschrift*) because this paradox had been found and so most of the stuff in that book was no longer valid[4]). Cantor, on the other hand, seems to have been aware of this paradox early on, but this didn't bother him that much: his take on this seems to have been simply that every instance of the Comprehension Schema that leads to contradiction can be removed from the axioms. The reason why we now don't consider that position reasonable is, of course, that by doing that we would obtain a non-recursive (and possibly even not-well-defined!) collection of axioms.

1.5 Patching up Cantor's set theory

Thus we would like to arrive at an axiom system that allows us to prove most of the things that we were able to prove in Cantor's set theory, but without the paradoxes, especially Rusell's Paradox[5]. One way to do this, the one that historically ended up being the most accepted (up to today) is by means of the ZFC axioms, which we will proceed to explain now. (At this point, the students should have the sheet containing the Axioms of set theory, which in this document can be found as Appendix A.)

Let us have a quick look at the axioms, one by one. The first three (two?) are

[4] Also, Frege's notation for logical formulas was quite unusual, completely unlike the one that we're used to. For example, to say that $(\forall x)(P(x) \Rightarrow Q(x))$ Frege would write



To type such symbols, one needs the LaTeX package `begriff`.

[5] When referring to the way that the principle of Comprehension is restricted to avoid these paradoxes, Barr and Wells (in *Category Theory for Computing Science*, available online at `http://www.math.mcgill.ca/triples/Barr-Wells-ctcs.pdf`, 1998) state that "[t]his prophylaxis guarantees safe sets". If the reader did not burst up with laughter while reading those words, then reading them again, out loud, is encouraged.

Axiom of Existence: $(\exists x)(x = x)$; informally, *there exists a set.* This axiom is here just for completeness, to make sure that there is at least one object to talk about (in logic, it is typically accepted as a logical axiom, so in theory we shouldn't even need to mention this axiom, but we do so just to emphasize that our theory is nonempty).

Axiom of Extensionality: $(\forall x)(\forall y)((\forall z)(z \in x \iff z \in y) \Rightarrow x = y)$; informally, *a set is determined by its elements.* This one is exactly the same as in Cantor's set theory, and it works in the exact same way.

Axiom Schema of Comprehension: For every $\mathscr{L}_{\mathrm{ST}}$-formula φ with one free variable, $(\forall x)(\exists y)(\forall z)(z \in y \iff (z \in x \wedge \varphi(z)))$; informally, *for every set x, the set $y = \{z \in x \big| \varphi(z)\}$ exists.* This is what we do in order to avoid Russell's paradox. First of all, notice that this is a schema that generates infinitely many axioms. It looks a lot like the "unrestricted comprehension" of Cantor's set theory, except that it no longer allows us to "create sets out of nothing": given a property φ, one can no longer just ensure the existence of the set $\{x \big| \varphi(x)\}$, but rather only of the sets of the form $\{x \in A \big| \varphi(x)\}$, where A is some set whose existence has been proved previously. Thus this axiom only allows us to create subsets of sets that are already given (for this reason, it is sometimes called the *Separation Axiom*, or the *Specification Axiom*, or the *Subset Axiom*, since instead of creating sets out of a void, it only "separates" subsets of already given sets).

We are now in a position of proving our first few theorems in ZFC. Our objective is to prove as many items from Theorem 4 as possible. Compare the proofs of the following two theorems with the proofs of the analogous results in Cantor's set theory.

Theorem 5.

1. *There exists a unique set with no elements (which we will denote from now on with the letter $\varnothing$),*

2. *given sets a, b, there exists a unique set whose elements are exactly those sets that belong to both a and b (which we will denote from now on with the symbol $a \cap b$),*

3. *given sets a, b, there exists a unique set whose elements are exactly those sets that belong to a but not to b (which we will denote from now on with the symbol $a \setminus b$).*

Proof. (Uniqueness, as always, directly follows from extensionality.)

1. By the axiom of existence, there exists a set x. Now an instance of the axiom schema of comprehension gives us the existence of the set
$$\{y \in x \big| y \neq y\},$$
which clearly has no elements.

2. We are assuming that a and b are given, thus an instance of the Axiom Schema of Comprehension ensures that $a \cap b = \{x \in a \big| x \in b\}$ exists.

3. Once again, appealing to the Axiom Schema of Comprehension gives us the set $a \setminus b = \{x \in a \big| x \notin b\}$.

□

Note how, in all instances of the previous proof, it is no longer sufficient to just invoke a property, but we also need some set from which we can possibly extract the subset of elements satisfying the property. Interestingly, in ZFC the argument for Russell's paradox becomes a proof by contradiction that there is no "universal set" (so unlike Cantor's set theory, ZFC disproves the existence of V).

Theorem 6. *There is no set containing all sets as elements.*

Proof. Suppose by contradiction that V is a set such that $(\forall x)(x \in V)$. Then an instance of the Comprehension Axiom would imply that the set
$$W = \{x \in V \big| x \notin x\}$$
exists, but (just as in the proof of Russell's paradox) since $W \in V$, it is meaningful to ask whether $W \in W$, and analysing the definition of W we get that $W \in W$ iff $W \notin W$, a contradiction. □

This proof looks strange at first, so perhaps we should try to analyse it a bit further. Certainly $x \notin x$ is an $\mathscr{L}_{\mathrm{ST}}$-formula with one free variable, so for any given set A, the corresponding instance of Comprehension gives us the set $W = \{x \in A \big| x \notin x\}$. Russell's paradox arises from the question of whether $W \in W$, but in this case, since W is constructed as a subset of A, the question only arises in case $W \in A$. But then W should have been available to us at the moment where we constructed A, which seems unlikely (especially since we are trying to define W after already being in possession of A).

The argument of the previous paragraph is not a formal argument, rather it is something that relies on a certain intuition that students might or might not have developed at this stage. Another, more formal, way of thinking about this is as follows: given A, Comprehension provides us with the set $W = \{x \in A | x \notin x\}$. So in order to belong to W, a set x must satisfy *both* statements $x \in A$ and $x \notin x$. When we ponder the question of whether $W \in W$, we have to consider two cases:

- If $W \in W$, then we must have that $W \in A$ and $W \notin W$, which contradicts our assumption.
- On the other hand, if $W \notin W$, it is because it is not the case that $W \in A \wedge W \notin W$. So we must have either $W \notin A$, or $W \in W$. Since the latter option contradicts our current assumption, the conclusion must be simply that $W \notin A$, and that's it... the contradiction is nowhere to be found.

In particular, given any set A we just proved the existence of this other set, W, which (among other things) has the property that $W \notin A$. So we have just proved that for every set, there exists some other set that is not an element of the first set; in other words, we have proved that no set can possibly contain all sets.

So all of this looks like a rather sketchy move, but it seems that it does accomplish the objective of avoiding Russell's paradox.

Two questions might naturally arise now, namely:

- Could there be some other way to make Russell's paradox work in ZFC?
- Is it possible that, even if Russell's paradox is avoided by the ZFC axioms, some other contradiction arise?

The answer to these two questions is, unfortunately, "we do not know"; in fact, arguably the answer is actually "we cannot know". By Gödel's Second Incompleteness Theorem, we cannot prove in ZFC that ZFC has no contradictions (unless ZFC actually has contradictions, since in this case we can prove everything from those axioms!). All we can be certain of, as of right now, is that there has a significant amount of people (some of which have been extremely smart) that have worked in the axiom system ZFC, for close to a whole century now, and to this day, no contradiction has been found. This situation might seem unsatisfactory at first sight, but it is in fact no different than what we experience in our everyday life: we will never be completely, 100% sure that the world exists (for all we know, everything could be a dream; or to put it differently, just remember Descartes's evil genius[6]), yet we do not let that uncertainty paralyze us, and somehow we just carry on with our lives. For a slightly less extreme example, I can never be completely certain that the engineer or the architect in charge of designing this building (or the construction workers!) did the job correctly (or, for that matter, that the laws of physics in which the design is based will keep holding in the next moment), yet I still confidently go about my daily business without (too much) fear that the building will suddenly collapse over my head[7]. Thus we take the pragmatic position that our uncertainty about whether or not ZFC is consistent will not stop us from doing math.

There is actually an interesting story related to this: in 2011 Edward Nelson claimed to have found a contradiction (in Peano Arithmetic, which would imply also a contradiction in ZFC), but it took about two or three days for Terence Tao to find an essential gap in Nelson's purported proof of an inconsistency, so eventually Nelson withdrew his claim. The most exciting result in 2011 in foundations of mathematics was having no result![8]

So now we have been able to prove a few things using the first couple of ZFC axioms, and it looks like we have (barely) managed to avoid Russell's paradox. But we cannot prove the existence of sets like $\{\varnothing\}$ with these axioms only, since these two axioms only allow us to extract subsets of given sets, and this is too restrictive. Similarly, we cannot prove the existence of $a \cup b$, for this would require that the elements that are either in a or in b all belong simultaneously to some given set. And something entirely similar happens with powersets. This is the reason that we have axioms 3, 4, and 5:

Axiom of Pairing: $(\forall x)(\forall y)(\exists z)(\forall w)(w \in z \iff (w = x \vee w = y))$; informally, *for every two sets x and y, the set $z = \{x, y\}$ exists.* There is nothing requiring a and b to be distinct, so in particular if we are given a set a, then this axiom allows us to ensure the existence of the singleton $\{a\} = \{a, a\}$.

Axiom of Union: $(\forall x)(\exists y)(\forall z)(z \in y \iff (\exists w)(w \in x \wedge z \in w))$, informally: *for every set x, the set $y = \bigcup\limits_{w \in x} w$ exists* (recall that all elements of x must be themselves sets, so each set is at the same time "a family of sets", and this axiom ensures the existence of the union of such family). The union of all elements of x is sometimes also denoted by

[6] Or, for a more contemporary example, remember the movie *The Matrix*, which talks about a variety of matrix that you won't encounter in your typical Linear Algebra class!!!

[7] Those of you sitting in the front rows of this classroom might never have complete certainty that your classmate sitting behind will not stab you in the back, yet that does not prevent you from focusing on listening to me and taking notes (or the sweeter alternative, falling asleep) without much fear for your lives.

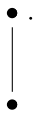
[8] I should probably give credit to Peter Krautzberger (who was also a postdoc at UofM a few years ago) for that last sentence. He said as much, with essentially the same words, in a blog post which is available online at `https://www.peterkrautzberger.org/0113/`.

$\bigcup x$; intuitively, taking such a union amounts to removing the "2nd level brackets" if we were to write x explicitly: for example,

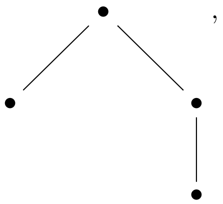
$$\bigcup\{\{a,b,c\},\{b,d\}\{e\}\} = \{a,b,c,b,d,e\} = \{a,b,c,d,e\}.$$

A useful way of thinking of sets within set theory is as trees. Given a set, draw a "top" node representing it, and let this node have a child for each element of the set. These elements are themselves sets, so their corresponding nodes should have one child node per element, and so on. Repeating this process for as long as necessary, we obtain the tree representation of the corresponding set. For example, the tree representation of $\varnothing$ would be

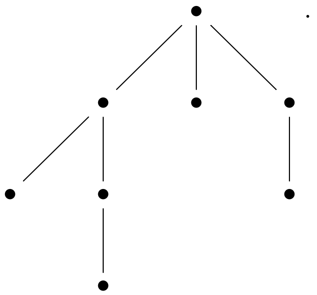
whereas the tree representation of $\{\varnothing\}$ would be



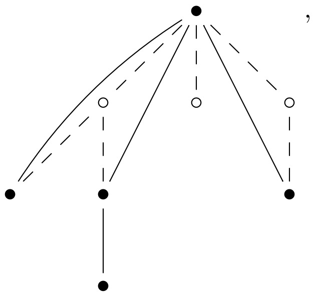
A more complicated set, such as $\{\varnothing,\{\varnothing\}\}$ would be represented by the following tree,



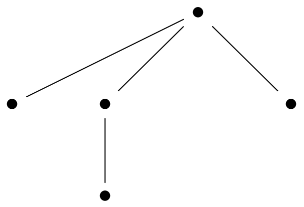
whereas, as a final example, the tree representation of the set $\{\{\varnothing,\{\varnothing\}\},\varnothing,\{\varnothing\}\}$ would be



Taking the union of a set amounts to removing the "second level nodes", and connecting the root directly to the "third level nodes" on the tree diagram of our set. For example, computing the union of the set in our last example should look as follows:



which simplifies to



,

the diagram of the set $\{\varnothing, \{\varnothing\}, \varnothing\} = \{\varnothing, \{\varnothing\}\}$.

Axiom of Powerset: $(\forall x)(\exists y)(\forall z)(z \in y \iff z \subseteq x)$; informally, *for every set x, the powerset $y = \mathfrak{P}(x) = \{z \big| z \subseteq x\}$ exists.* I believe this one is pretty much self-explanatory.

The following theorem takes care of the remaining items in Theorem 4 that we haven't been able to prove (or refute) so far (note the order in which we list our statements, which is influenced by the order in which we must prove them):

Theorem 7.

1. *there exists a unique set whose only element is $\varnothing$ (denoted $\{\varnothing\}$),*
2. *there exists a unique set whose only element is the set $\{\varnothing\}$ (denoted $\{\{\varnothing\}\}$),*
3. *given two sets a, b, there exists a unique set whose elements are exactly those sets that belong to either a or b (denoted $a \cup b$),*
4. *given finitely many sets $x_1, \ldots, x_n$, there exists a unique set whose only elements are $x_1, \ldots, x_n$ (denoted $\{x_1, \ldots, x_n\}$),*
5. *given a set a, there exists a set whose elements are exactly those sets that are subsets of a (denoted $\mathfrak{P}(a)$),*
6. *for every set a, it is not the case that there exists a set whose elements are all sets that do not belong to a (that is, in* ZFC *complements don't exist).*

Proof. (By now you should know what to do for uniqueness.)

1. Since we've already proved that $\varnothing$ exists, Pairing ensures the existence of $\{\varnothing\}$.
2. By 1 above, $\{\varnothing\}$ exists, so Pairing ensures the existence of $\{\{\varnothing\}\}$.
3. Given a and b, pairing ensures that the set $\{a, b\}$ exists. Thus by Union, the set $\bigcup\{a, b\}$ exists. This set consists of those x that belong to some element of $\{a, b\}$, that is, this set is exactly $a \cup b$!
4. This one needs to be done by induction[9]: if $n = 1$, the existence of $\{x_1\}$ follows directly from Pairing. Suppose that we have the result for n sets, and now we are given $n + 1$ sets $x_1, \ldots, x_n, x_{n+1}$. By induction hypothesis, we can prove the existence of $\{x_1, \ldots, x_n\}$; by Pairing, we can prove the existence of $\{x_{n+1}\}$. Thus by 3 above, we obtain that the set
$$\{x_1, \ldots, x_n\} \cup \{x_{n+1}\} = \{x_1, \ldots, x_n, x_{n+1}\}$$
exists, and we are done.
5. This one is precisely the statement of the Powerset Axiom.
6. Note that this is another example of a statement where ZFC proves the opposite of what Cantor's set theory does. Let a be a set, and suppose by contradiction that there exists a set a' satisfying $(\forall x)(x \in a' \iff x \notin a)$. By item 3 above, the set $a \cup a'$ exists, and consists of all x such that $x \in a$ or $x \in a'$. But $x \in a' \iff x \notin a$, and since for every x it is the case that either $x \in a$ or $x \notin a$, we conclude that $a \cup a'$ is a set containing all sets as elements. But the existence of such a set has been disproved in Theorem 6 above.

□

We have used the Axiom of Union to prove the existence of the union of two sets, but the axiom is of course more powerful and ensures the existence of the union of an arbitrary family of sets. We don't need any axiom for ensuring the existence of the intersection of an arbitrary nonempty family, since this one follows from Comprehension (note that we have been, and will continue to be, always very careful with our uses of Comprehension; by only invoking subsets of sets that we have obtained previously).

[9] There is a subtlety here, because we still have not justified the principle of induction within ZFC. Rather, this is an induction that gets done in the metatheory, and this theorem is really a metatheorem, stating that for every n, it is possible to prove the statement $(\forall x_1, \ldots, x_n)(\exists y)(\forall z)(z \in y \iff (z = x_1 \vee \cdots \vee z = x_n))$ in ZFC.

Proposition 8. *If x is nonempty, then there is a unique set whose elements are precisely those sets that belong to all elements of x, and we will denote such set by $\bigcap x$ or, if we want to be more explicit, by $\bigcap_{y\in x} y$ (and if x is empty, we declare that $\bigcap\varnothing=\varnothing$ by convention, which, although counterintuitive, is still less worse than having $\bigcap\varnothing=V$, since V does not exist in* ZFC*).*

Proof. Whether or not x is nonempty, we know that $\bigcup x$ exists by Union. Now, use an instance of Comprehension to ensure that

$$\bigcap x=\left\{y\in\bigcup x\middle|(\forall z\in x)(y\in z)\right\}$$

exists (notice that, in case $x=\varnothing$, then $\bigcup x=\varnothing$, which means that the above equation uniformly defines $\bigcap x$ according to our convention, regardless of whether or not x is nonempty). □

It is now time to make a remark about the Axiom of Powerset. Suppose that x is a finite set, say $x=\{x_1,\dots,x_n\}$. Then, we can explicitly write out every possible subset of x by using item 4 in Theorem 7. And once we have all of the 2^n possible subsets, we can certainly collect them together in a set, again by item 4 in Theorem 7. For example, if we are given the set $\{a,b,c\}$, we can individually prove the existence of each of the sets $\varnothing,\{a\},\{b\},\{c\},\{a,b\},\{b,c\},\{a,c\},\{a,b,c\}$ (by Theorem 7 part 4), and once we have these sets, once again Theorem 7 part 4 guarantees the existence of the set $\{\varnothing,\{a\},\{b\},\{c\},\{a,b\},\{b,c\},\{a,c\},\{a,b,c\}\}=\mathfrak{P}(x)$. So we should be aware that the Powerset Axiom is not necessary if we want to only work with finite sets; the reason that we need this axiom is specifically because we want to consider infinite sets as well.

In a similar vein as in the previous paragraph, the Axiom of Pairing (by mirroring the reasoning that we carried out in the first two items of Theorem 7) allows us to ensure the existence of the set $\{\{\{\varnothing\}\}\}$, and then subsequently of $\{\{\{\{\varnothing\}\}\}\}$, and so on *ad infinitum*; note that all of these sets are pairwise distinct (this is a problem in the current assignment!). Thus the pairing axiom allows us to ensure not only that our universe of discourse is nonempty, but in fact infinite. Hence there are infinitely many sets. However, this still does not ensure that there is any set containing infinitely many elements. Although we can certainly think of infinite collections of sets (just like we can think about the collection of all sets), that doesn't mean that there should be a set *within* the theory that exactly corresponds to this collection (just like there is no set within the theory that exactly corresponds to the collection of all sets). This is the reason that we need to introduce the following axiom.

Axiom of Infinity: $(\exists x)(\varnothing\in x\land(\forall y)(y\in x\Rightarrow y\cup\{y\}\in x))$; informally, *there exists an infinite set.* The idea here is the following: let x be the set whose existence is asserted by this axiom (by the way, notice that, in particular, this axiom implies the Axiom of Existence, and thus we might as well completely delete Existence from our list of axioms). Then $y_0=\varnothing\in x$. But we must also have $y_1=\varnothing\cup\{\varnothing\}\in x$, and then, $y_2=y_1\cup\{y_1\}\in x$, and so on. It is possible to prove that the y_n are pairwise distinct (this will also be a part of your assignment), and thus the set x whose existence is asserted must contain infinitely many elements.

In the preceding paragraph, the issue that the y_n are pairwise disjoint arose. In general, how would we go about proving that $x\neq x\cup\{x\}$ for an arbitrary set x? By extensionality, and since we have that $x\subseteq x\cup\{x\}$, we would need to find a y such that $y\in x\cup\{x\}$ but $y\notin x$, and the natural (and only possible) candidate for this would be to take $y=x$. It turns out, however, that proving $x\notin x$ is harder than it would seem at first sight –in fact, so hard that it is impossible, unless we use the following axiom.

Axiom of Foundation: $(\forall x)((\exists y)(y\in x)\Rightarrow(\exists y)(y\in x\land\neg(\exists z)(z\in x\land z\in y)))$; informally, *for every nonempty set x there exists a $y\in x$ such that y and x are disjoint.* This is actually a very deep axiom, although almost never explicitly used by anyone who is not a set theorist, and we will discuss its full implications towards the end of the course (when we deal with well-founded induction). For now, however, we can prove a couple facts, that are not entirely disgusting (maybe we could even say that they are nice), and that follow from this axiom.

Proposition 9.

1. *There is no set x with the property that $x\in x$,*

2. *there are no two sets a,b with the property that $a\in b$ and $b\in a$,*

3. *there are no sets $a_1,\dots,a_n$ satisfying $a_{i+1}\in a_i$ for all i, and $a_1\in a_n$.*

Proof. 1. Suppose that x is such that $x\in x$. By pairing, $\{x\}$ exists, and it is clearly nonempty. Thus by Foundation, there is a $y\in\{x\}$ such that $\{x\}\cap y=\varnothing$, but the only possibility for y is $y=x$, and since $x\in x$, we have that $x\in\{x\}\cap x=\varnothing$, a contradiction.

2. Suppose that a, b are such that $a \in b$ and $b \in a$. By pairing, the set $\{a, b\}$ exists, and it is clearly nonempty, thus by Foundation there is a $y \in \{a, b\}$ such that $y \cap \{a, b\} = \varnothing$. But there are only two cases for y, namely either $y = a$ or $y = b$; in the first case, we have $b \in a \cap \{a, b\}$, in the second, $a \in b \cap \{a, b\}$. In both cases we obtain that $y \cap \{a, b\} \neq \varnothing$, a contradiction.

3. This one, you guys will prove in the next assignment!

□

Proposition 9 tells us (among other things) that, in ZFC, every set x satisfies $x \notin x$. This yields a new perspective on how to avoid Russell's Paradox using the ZFC axioms: since every x satisfies $x \notin x$, it follows that, for every A, the set $\{x \in A \big| x \notin x\} = A$, and no contradiction arises. This also gives us a different proof that there is no universal set in ZFC: if there was such set V, containing all sets as elements, then in particular we would have $V \in V$, which directly contradicts Proposition 9.

The last two axioms are much more subtle, and will only be discussed later on in the course (one of them occupies a section, and the other a full chapter, in these notes), but we state them here for completeness –so that we now know the full "official" list of axioms, and there are no further surprises later.

Axiom Schema of Replacement: For every formula φ of the Language of set theory with two free variables, $(\forall z)((\forall x)(x \in z \Rightarrow (\exists! y)\varphi(x, y)) \Rightarrow (\exists w)(\forall x)(x \in z \Rightarrow (\exists y)(y \in w \wedge \varphi(x, y))))$; informally, *for every set x, if what φ describes behaves like a function with domain x, then the range of that function (the set of images of elements of x under the function) exists.*

Axiom of Choice: $(\forall x)((\varnothing \notin x \wedge (\forall y)(\forall z)((y \in x \wedge z \in x \wedge y \neq z) \Rightarrow y \cap z = \varnothing)) \Rightarrow (\exists w)(\forall y)(y \in x \Rightarrow (\exists! z)(z \in y \cap w)))$; informally, *for every set x whose elements are pairwise disjoint and nonempty, there exists a set w that contains exactly one element from each $y \in x$.*

The Axiom of Choice was somewhat controversial for some time, to the extent that nowadays, there is a whole area of research within set theory devoted to the analysis of what propositions and theorems require the use of this axiom in order to be proved. Replacement is almost never outside of set theory (although in set theory it is crucial for many arguments, including the existence of most of the so-called *ordinal numbers*). There is only one result that I am aware of, outside of set theory, where the Axiom of Replacement is used in an essential way (in the sense that we know that it is not possible to prove the statement without using the axiom), and that is Donald Martin's result that every Borel game in a Polish space is determined (this is a complicated proof by induction on the Borel complexity of the payoff set of the corresponding game, where the base case –that open games are determined– is what is known as the *Gale-Stewart theorem.* An amenable exposition of this proof can be found in a series of blog posts by Tim Gowers, the first of which is available at `https://gowers.wordpress.com/2013/08/23/determinacy-of-borel-games-i/`).

It is worth mentioning that the theory ZFC is not finitely axiomatizable, that is, there is no finite collection of $\mathscr{L}_{\mathrm{ST}}$-formulas allowing us to prove the exact same collection of theorems that ZFC does (those of you who took Math 481 should realize that this is equivalent to the statement that the class of all models of ZFC is not elementary; if you ever take Math 682, you will see that the proof of this fact follows easily from the Reflection Theorem). A brief historical account of how the axiom system of ZFC came to be would have to start with Zermelo's proof of the well-ordering principle (stating that every set can be well-ordered). Zermelo devoted a portion of his paper to explicitly state each and every assumption that is used for his proof; this collection of assumptions is essentially the ZC^- axioms (that is, all of the axioms in our current list, except for Foundation and Replacement). However, Comprehension was still formulated with the vague words "for every property"; the usage of first-order logic to state Comprehension as an $\mathscr{L}_{\mathrm{ST}}$-formula arose later, by a suggestion of Skolem's. Fraenkel suggested the addition of the Axiom of Replacement (hence the "F" in ZFC stands for this axiom), and von Neumann introduced the Axiom of Foundation (and provided a proof of its relative consistency with the remaining axioms, by using a technique which is clearly reminiscent of Gödel's later work with the so-called Constructible Universe). There have been other attempts at different axiom systems that avoid Russell's Paradox, most notably Russell's theory of types (which predates Zermelo's paper) and Quine's New Foundations (which dates from much later, well after the ZFC axioms were already widely used), which we proceed to explain in the next section.

1.6 Other ways of dodging Russell's Paradox

Let us now say a few words about Russell's theory of types and Quine's New Foundations. In Russell's theory of types, one takes the notion of natural number as a starting point (intuitively speaking, although I believe that formally this can be avoided), and then associates to every element of the theory a natural number (its "type"), which is supposed to denote the "level" of the corresponding object within the hierarchy of sets. So objects of type 0 are to be thought of as elements only, not sets, since they do not have any elements. Objects of type 1 are sets that can only have objects of type 0 as elements, and in general the objects of type $n + 1$ are only allowed to have objects of type n as elements.

Note, for example, that the axiom of extensionality is dropped here, or at the very least, it should be replaced with a variant of it stating that, *within each positive type*, sets are determined by their elements. Thus, this theory has, e.g., many different empty sets (one on each type); on the other hand, there is no such thing as an Axiom of Extensionality for objects of type 0, since these are considered to be pairwise different even though they have the same elements –namely no elements. This theory never really gained much traction among mathematicians (although it became very famous among philosophers), mostly because it is very cumbersome to have to keep track of the types, and because there is this nebulous feeling, when utilizing this theory, that we are gratuitiously multiplying entities more than would normaly be desirable (for example, like we just mentioned, there are multiple different empty sets; for another example, there would also be many different representatives of the number "2"[10], one for each type $n \geq 2$. Everyone who likes to use Ockham's razor as a methodological principle is much more likely to feel at ease with ZFC (with its extreme simplicity of assuming only one kind of entity, namely sets) than with Russell's theory of types (in which essentially every object that one can care to define will have multiple counterparts, one on each different type).

Another attempt at dodging Russell's Paradox, partly inspired in Russell's Theory of Types, but without the complications associated to having multiple different types, is Quine's axiom system known as the New Foundations (usually denoted NF). Basically, NF consists of one axiom (the Axiom of Extensionality), together one axiom scheme (Comprehension), just like Cantor's set theory. The key difference here is that not all formulas of the form "$\{x \big| \varphi(x)\}$ exists" are axioms, but only those for which $\varphi(x)$ satisfies a very specific condition, known as *stratification.* An $\mathscr{L}_{\mathrm{ST}}$-formula φ is said to be **stratified** if there is a function f mapping each of the variables occuring in φ to a natural number, in such a way that the following two things happen for every pair of variables x, y:

- Whenever the subformula $x = y$ occurs in φ, then we must have that $f(x) = f(y)$; and
- whenever the subformula $x \in y$ occurs in φ, then we must have that $f(y) = f(x) + 1$.

Intuitively, we can think of a stratified formula as one that would "make sense" even if we were working in Russell's theory of types, that is, where we could make sense of the idea that each variable x belongs to a given type (the assignment of types is coded by the function f), and where we demand that variables x can only belong to other variables y with $f(y) = f(x) + 1$, that is, the "type" of y has to be exactly one more than the "type" of x. So the axiom scheme of Comprehension in NF establishes that the formula "$\{x \big| \varphi(x)\}$ exists" is an axiom if and only if φ is stratified. In this axiom system, we can prove all of the items in Theorem 4, including the existence of a universal set V, and there is no way of invoking Comprehension to obtain Russell's Paradox, because the formula $x \notin x$ is not stratified. This theory has many features that seem quite counterintuitive at first sight (for example, not every subset of a set that can be described by an $\mathscr{L}_{\mathrm{ST}}$-formula is guaranteed to exist, for the formula might not be stratified), of which the most outrageous one is that NF implies the negation of the Axiom of Choice, which perhaps explains why mathematicians are, in general, not too keen on this axiom system. It is possible to modify the axiom system NF to allow for objects known as "urelements" (objects that are not sets, but only elements of sets) to coexist together with sets. The resulting axiom system is denoted by NFU (New Foundations with Urelements), and this system is slightly less counterintuitive in many respects, even though it still has some idiosyncratic features (for example, there is the notion of a "Cantorian set", which is a set X that happens to be in bijection with the set $\{\{x\} \big| x \in X\}$, and non-Cantorian sets might exist; in contrast, in ZFC it is trivially easy to prove that every set satisfies the definition of a Cantorian set). On the other hand, this axiom system seems to have a relatively low consistency strength (meaning that it is not "as powerful" as ZFC, in the sense that it does not allow us to prove many of the more difficult and substantial mathematical results that can be proved in ZFC). The system NFU was proved (by Jensen in 1969) to be consistent, provided ZFC is; the question of whether we could say the same about NF remained open for a really long time. Recently (since 2010) Randall Holmes has announced a solution (in the affirmative) to this question.

1.7 The last few words about the axiomatic method

Let us say a few final words about axiomatics. It is important to remember that there are two meanings of the word "axiom". The first meaning, the original one, is that an axiom should be some intuitively obvious truth (traditionally, this is the sense in which the word is used when one mentions Euclid's axioms for geometry, or the axioms for the Real Line). The second meaning is that of an axiom as a "definitional axiom": when using the word with this meaning, an axiom is just an assumption that one takes as a starting point, in order to deductively proceed to draw consequences of these assumptions (it is in this sense that the "group axioms", for example, are axioms: these are typically introduced to students before they even have any intuition about what a group is, and their rôle is not that of an intuitively obvious fact, but rather the axioms themselves constitute the very definition of a group, an object which is fundamentally unknown before the axioms are stated). A gradual shift from the first meaning to the second one is a dialectical move that happens

[10] Where the object of type $n + 1$ that represents the number "2" in this theory would be defined be the set of all sets of type n that have exactly two elements.

fairly frequently: for example, in Linear Algebra one typically starts with the axioms of a vector space, which tend to reflect more or less obvious truths about elements of $\mathbb{R}^n$ (so at the beginning these axioms are justified by taking the word "axiom" with the first meaning); but once the list of axioms is laid out, one proceeds to draw consequences of those axioms purely formally, even if some of these consequences clash with our previous intuitions, in which case we revise those intuitions rather than discard the formal consequences of the axioms (so in the end, we wind up using the word "axiom" with its second meaning). It can be argued that a similar shift takes place also when doing geometry, and when studying the real line $\mathbb{R}$ axiomatically. In fact, I would argue that even group theory is the outcome of such a shift (since it seems that, historically speaking, the axioms for a group were originally motivated by the fact that many people had already been working with several examples of groups –groups of permutations, of matrices, etc.–, so that in the beginning the concept of "group" really arose from an attempt to unify several objects for which people had already developed some intuition), even though that is not the typical experience of a generic undergraduate student nowadays.

What happens in set theory is entirely analogous: at first we started with some intuitions about "abstract collections of objects", and attempted to encapsulate these intuitions within a bunch of axioms. As we saw throughout the previous pages, the way in which this intuition guides us through the formulation of axioms has to be amended and corrected every now and then, because otherwise some contradictions might arise, and at the same time we need to explicitly keep adding new axioms guaranteeing the existence of certain sets that we intuitively feel "should exist", but whose existence we can no longer prove after the corresponding amendments. By means of this process, after performing the necessary corrections to our first clumsy attempts at appropriately choosing some axioms, we settle on a specific collection of them –the ZFC axioms. But once we have agreed that this will be the "official" list of axioms, we start thinking of the concept of set (and more specifically, of the meaning of the symbol $\in$) as if it was an undefined notion, defined exclusively in terms of the axioms that need to be satisfied, and we just formally derive consequences of our axioms. Along the way we develop new intuitions about what can be proved from the axioms, and these intuitions wind up replacing whatever original ("naïve", if you will) intuitions we formerly had regarding sets.

Chapter 2

Relations, functions, and objects of that ilk

Parts of this topic would normally not qualify at first sight as set-theoretic. However, we normally include them in this course for three main reasons. The first is that this material will be needed in order to carry out, later on, the construction of the real line $\mathbb{R}$. Secondly, recall that we are also in the business of illustrating how to implement all of Mathematics in ZFC. We will, among other things, show explicitly how to implement our first few mathematical objects (relations and functions) within ZFC. Lastly, multiple other courses (such as Modern Algebra) extensively use some particular cases of some of the concepts introduced here (for example, equivalence relations); it is useful to present a unified way of thinking about all of these isolated particular examples.

We make a point of noting that, throughout this chapter, we will only use the first five axioms (Extensionality, Comprehension, Pairing, Union, and Powerset)[1].

2.1 Ordered pairs, relations and functions

So far, we have seen how to define a few different simple sets, always taking the empty set $\varnothing$ as a starting point. We will now start implementing other kinds of objects within set theory.

To begin with, suppose that we have two objects (i.e. two sets) x and y. The Axiom of Pairing guarantees the existence of the *unordered pair* $\{x, y\}$, which is an object that, in a sense, carries all the information that both x and y carry, but has the disadvantage that it does not allow us to distinguish x from y (since, by the Axiom of Extensionality, it is the case that $\{x, y\} = \{y, x\}$). We would like to define (implement) an object that, in addition to carrying all the information about x and y, allows us to distinguish between both objects. If we succeed in finding such an object, it will only be natural to denote this object by the symbol (x, y), and to call it an *ordered pair*.

In other words, given x and y we are seeking for the definition of an object, whose existence can be proved in ZFC and which we will denote with (x, y), such that $(x, y) = (z, w)$ if and only if $x = z$ and $y = w$ for all x, y, z, w. We emphasize that there are multiple ways of defining such an object (some of them depend on the axiom system at hand, for example the usual definition of an ordered pair in NF does not make sense in ZFC), and our specific choice is not really that relevant, as long as we choose a definition that does its job properly. In this course, we will officially adopt Kuratowski's definition of ordered pair, which for better or for worse is the one most widely used nowadays.

Definition 10. Given two sets x, y, we define the **ordered pair** as follows:

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

We will say that a set a is an ordered pair if there exist x, y such that $a = (x, y)$.

Recall that we vowed to only use the language $\mathscr{L}_{\mathrm{ST}}$ when speaking within our theory. This means that, if we want to be justified in writing the words "a is an ordered pair", we need to show that there is an $\mathscr{L}_{\mathrm{ST}}$-formula $\varphi(a)$ which holds true exactly when a is an ordered pair. So we proceed to show such a formula explicitly. Start by noting that the $\mathscr{L}_{\mathrm{ST}}$-formula

$$(\exists x)(x \in a \wedge (\forall y)(y \in a \Rightarrow y = x))$$

is true if and only if a is a singleton; hence we are now justified in writing the words "a is a singleton" when doing set

[1] Plus, of course, the Axiom of Existence.

theory; those words are just an abbreviation of the above formula[2]. Similarly, the formula

$$(\exists x)(\exists y)(x \in a \wedge y \in a \wedge (\forall z)(z \in a \Rightarrow (z = x \vee z = y)))$$

holds true if and only if a has at most two elements, hence from now on we can say that "a has at most two elements", and take this to be an abbreviation of that formula. Piecing all of this together, we can think of the words "a is an ordered pair" as if they were the abbreviation of the following $\mathscr{L}_{\mathrm{ST}}$-formula (technically, of the following thing which can be turned into an $\mathscr{L}_{\mathrm{ST}}$-formula if we appropriately expand all of the relevant macros, or abbreviations):

$$(a \text{ has at most two elements}) \wedge \left(\bigcap a \text{ is a singleton}\right) \wedge \left(\bigcap a \in a\right) \wedge (\forall x \in a)(x \text{ has at most two elements})$$

Therefore, talking about whether an object is an ordered pair or not is something that can be done in the language $\mathscr{L}_{\mathrm{ST}}$. Let us now proceed to prove that our definition of an ordered pair is adequate, that is, that an ordered pair really captures its two entries along with the information about the order in which these entries occur.

Theorem 11. *Suppose that a is an ordered pair. Then the x, y such that $a = (x, y)$ are uniquely determined. In particular, for all x, y, z, w, we have that $(x, y) = (z, w)$ if and only if $x = z$ and $y = w$.*

Proof. Note that $\bigcup(x, y) = \{x, y\}$ and $\bigcap(x, y) = \{x\}$. Thus if a is the ordered pair (x, y) then we can recover x by means of the formula

$$x = \bigcup \left(\bigcap a\right),$$

(that is, if a is an ordered pair then $\bigcap a$ is a singleton, whose unique element is x); similarly, noting that $\bigcup a$ is a singleton if and only if $x = y$, we see that we can recover y by means of the formula

$$y = \begin{cases} x = \bigcup(\bigcap a) & \text{if } \bigcup a \text{ is a singleton,} \\ \bigcup(\bigcup a \setminus \bigcap a) & \text{otherwise.} \end{cases}$$

□

The above proof contains a definition by cases, so we need to justify that definitions by cases can actually be properly made within the language $\mathscr{L}_{\mathrm{ST}}$. In fact, as long as there are only finitely many cases, this will be possible. For if we have $\mathscr{L}_{\mathrm{ST}}$-formulas $\varphi_1(x), \dots, \varphi_n(x)$ that are mutually exclusive for every x (that is, such that for every choice of x and indices i, j, the formulas $\varphi_i(x)$ and $\varphi_j(x)$ cannot be both simultaneously true) and objects (whose definition might depend on x) $y_1, \dots, y_n$, then we can define the object

$$z = \begin{cases} y_1 \text{ if } \varphi_1(x), \\ y_2 \text{ if } \varphi_2(x) \\ \vdots \\ y_n \text{ if } \varphi_n(x), \end{cases}$$

which will depend on x, by means of the formula

$$(\varphi_1(x) \wedge z = y_1) \vee (\varphi_2(x) \wedge z = y_2) \vee \cdots \vee (\varphi_n(x) \wedge z = y_n).$$

We introduce a special notation, which we might use occasionally, for denoting the first and second entries of an ordered pair. We will use the symbols π_1 and π_2 (first and second projection) for this. That is, if a is an ordered pair and $a = (x, y)$, then we define $\pi_1(a) = x$ and $\pi_2(a) = y$.

Now that we have a working definition of ordered pair that runs smoothly, we can proceed to prove, in ZFC, that cartesian products of two sets exist.

Proposition 12. *Given two sets A and B, there exists a unique set whose elements are exactly those ordered pairs (a, b) such that $a \in A$ and $b \in B$; we denote such set by $A \times B$.*

Proof. Notice that, for $a \in A$ and $b \in B$, we have that $(a, b) = \{\{a\}, \{a, b\}\} \subseteq \mathfrak{P}(A \cup B)$. Hence, an instance of Comprehension ensures the existence of

$$\{x \in \mathfrak{P}(\mathfrak{P}(A \cup B)) \big| (\exists a \in A)(\exists b \in B)(x = (a, b))\} = A \times B.^3$$

□

[2] There is a really nice trick that we can do with singletons in ZFC. Let x be a singleton, which means that there exists a unique y such that $x = \{y\}$. An easy calculation shows that $\bigcup x = y$. That is, every time we have a singleton, if we want to talk about the unique element of that singleton, it suffices to take the union (guaranteed to exist by the Axiom of Union) of that singleton. In other words, the symbol $\bigcup$ acts as an "operator" that "extracts" the unique element of every singleton. This allows us to use the words "let y be the unique element of the singleton x" with the peace of mind that those words can actually be expressed by means of an $\mathscr{L}_{\mathrm{ST}}$-formula.

[3] Equivalently, we could have written $A \times B = \{x \in \mathfrak{P}(\mathfrak{P}(A \cup B)) \big| (x \text{ is an ordered pair}) \wedge \pi_1(x) \in A \wedge \pi_2(x) \in B)\}$.

Now that we have cartesian products of two sets at our convenience, we can try to generalize to larger cartesian products. Given three sets A, B, C, we define $A \times B \times C = (A \times B) \times C$, and in general, we recursively define

$$A_1 \times \cdots \times A_n \times A_{n+1} = (A_1 \times \cdots \times A_n) \times A_{n+1}$$

This definition of the n-fold cartesian product defines implicitly, at the same time, what n-tuples will be. An n-tuple has to be an element of the corresponding n-fold cartesian product. Hence, an ordered triple (a, b, c) is by definition equal to the ordered pair $((a, b), c)$; and inductively we can see that our definitions yield

$$(a_1, \dots, a_n, a_{n+1}) = ((a_1, \dots, a_n), a_{n+1}).$$

Definition 13. Given sets $A_1, \dots, A_n$, we define an n**-ary** relation between these sets to be just an arbitrary subset of $A_1 \times \cdots \times A_n$.

A particular case of the above definition, which will be used extensively in this course, is when $n = 2$. In this case, we have two sets A, B, and we are looking at **binary relations**. So a binary relation between A and B (or just "on A" if $A = B$) is just a subset of $A \times B$. Whenever $R \subseteq A \times B$ is a binary relation, we will typically write $a \ R \ b$ instead of $(a, b) \in R$, to emphasize that we are thinking of R not just as a set, but as our implementation of a binary relation within set theory.

Example 14.

1. Suppose that we know who $\mathbb{Z}$ is[4], and that we allow ourselves to include the divisibility symbol $\mid$ in our formulas. Then the following relation R on $\mathbb{Z}$,
$$R = \{(n, m) \in \mathbb{Z} \times \mathbb{Z} \big| n \mid m\}$$
is just the relation "n is a divisor of m". On the other hand, the following relation S on $\mathbb{Z}$,
$$S = \{(n, m) \in \mathbb{Z} \times \mathbb{Z} \big| m \mid n\}$$
corresponds to the relation "n is a multiple of m".

2. Assuming that we know who $\mathbb{R}$ is, and that we allow ourselves to include the symbol $<$ in our formulas, the relation "is less than" on $\mathbb{R}$ corresponds to the set
$$\{(x, y) \in \mathbb{R} \times \mathbb{R} \big| x < y\}.$$

3. Given any two sets A and B, we always have the *empty relation* $\varnothing \subseteq A \times B$.

Definition 15. If $R \subseteq A \times B$ is a binary relation between A and B, we define the following.

1. The **domain** of R is the set
$$\mathrm{dom}(R) = \{a \in A \big| (\exists b \in B)(a \ R \ b)\}.$$

2. The **range** of R is the set
$$\mathrm{ran}(R) = \{b \in B \big| (\exists a \in A)(a \ R \ b)\}.$$

3. If $X \subseteq A$, the **image** of X under R is the set
$$R[X] = \{b \in B \big| (\exists a \in X)(a \ R \ b)\}.$$

4. If $Y \subseteq B$, the **preimage** of Y under R is the set
$$R^{-1}[Y] = \{a \in A \big| (\exists b \in Y)(a \ R \ b)\}.$$

5. If $X \subseteq A$, the **restriction** of R to X is the set
$$R \restriction X = \{(a, b) \in R \big| a \in X\}.$$

[4] We still have not defined many sets directly from the axioms, so if we were to stick to only these, we would be short on examples. Hence, when listing examples of recently defined concepts, we might resort to our old intuitions about everyday mathematical objects, with the caveat that these are only mentioned for the sake of having some examples available, but with the warning that we still have not implemented these sets within set theory. Later on, we will learn how to implement all of these sets within set theory, and at that moment these examples will take on a completely new meaning, for they will transition from being informal accounts to being completely formal objects whose existence is provable in ZFC.

6. If $S \subseteq B \times C$ is another binary relation, this time between B and C, the **composition** of R and S is the set
$$S \circ R = \{(a,c) \in A \times C \big| (\exists b \in B)(a \ R \ b \wedge b \ S \ c)\}.$$

7. The **inverse** of R is the set
$$R^{-1} = \{x \in \operatorname{ran}(R) \times \operatorname{dom}(R) \big| (\exists a \in A)(\exists b \in B)(x = (b,a) \wedge a \ R \ b)\}.$$
(Notice that we wrote every set in a way that makes it evident how to prove its existence via an instance of Comprehension.)

We now proceed to remark a few easy, but interesting, facts. First of all, let us note that the mention to the sets A and B in the sentence "R is a relation between A and B" is superfluous. This is because any set R consisting of ordered pairs satisfies $R \subseteq X \times X$, where $X = \bigcup(\bigcup R)$. Thus for any set R that consists of ordered pairs, there is an X such that R is a binary relation on X. Given this, we are from now on justified in uttering the words "R is a relation" to mean "every element of R is an ordered pair". In particular, this allows us to define all seven items above for any set R consisting of ordered pairs (in particular, every R all of whose elements are ordered pairs can be thought of as a relation between $\operatorname{dom}(R)$ and $\operatorname{ran}(R)$).

Another fact to notice is that, given any two relations R and S, it always makes sense to consider the set $S \circ R$, simply by considering R to be a relation between $\operatorname{dom}(R)$ and $\operatorname{ran}(R) \cup \operatorname{dom}(S)$, and S to be a relation between $\operatorname{ran}(R) \cup \operatorname{dom}(S)$ and $\operatorname{ran}(S)$. Note that $S \circ R = \varnothing$ (the empty relation) whenever $\operatorname{ran}(R) \cap \operatorname{dom}(S) = \varnothing$.

We now mention a short list of useful trivialities, given a relation R:

- $\operatorname{ran}(R) = R[\operatorname{dom}(R)]$,
- $R[X] = \operatorname{ran}(R \upharpoonright X)$,
- $\operatorname{dom}(R^{-1}) = \operatorname{ran}(R)$,
- $\operatorname{ran}(R^{-1}) = \operatorname{dom}(R)$,
- $(S \circ R)^{-1} = R^{-1} \circ S^{-1}$.
- The symbol $R^{-1}[Y]$ might *a priory* denote two different sets (depending on whether one is thinking of item 4 or item 7 of Definition 15); however these two coincide.

There is a plethora of properties that relations might have, and several combinations of these will be of crucial importance in this course. The following definition lists some of these properties.[5]

Definition 16. Let R be a binary relation.

1. We say that R is **symmetric** if $(\forall a)(\forall b)(a \ R \ b \Rightarrow b \ R \ a)$ (equivalently, $R^{-1} \subseteq R$).
2. We say that R is **asymmetric** if $(\forall a)(\forall b)(a \ R \ b \Rightarrow \neg(b \ R \ a))$.
3. We say that R is **antisymmetric** if $(\forall a)(\forall b)((a \ R \ b \wedge b \ R \ a) \Rightarrow a = b)$.
4. Given a set A, we say that R is **reflexive over** A if $(\forall a \in A)(a \ R \ a)$.
5. We say that R is **irreflexive** if $(\forall a)(\neg(a \ R \ a))$.
6. We say that R is **transitive** if $(\forall a)(\forall b)(\forall c)((a \ R \ b \wedge b \ R \ c) \Rightarrow a \ R \ c)$ (equivalently, $R \circ R \subseteq R$).
7. Given a set A, we say that R **satisfies trychotomy over** A if $(\forall a, b \in A)(a \ Rb \vee b \ R \ a \vee a = b)$.

It is now time to look at some particularly well-behaved relations. The following definition constitutes the implementation of our first highly nontrivial object within set theory.

Definition 17. A binary relation f is said to be a **function** if $(\forall x \in \operatorname{dom}(f))(\exists! y \in \operatorname{ran}(f))((x,y) \in f)$.[6]

[5] In practice, I have never listed these conditions at this moment during lecture. It seems pointless to state a long list of definitions, none of which will be used immediately. Doing this would entail that, by the time we finally use these definitions, they would have been largely forgotten and it would be necessary to state them again.

[6] The quantifier $\exists!$ (read "there exists a unique") is just one more of our abbreviations. The string of symbols $(\exists! x)(\varphi(x))$ will be thought of as an abbreviation of the $\mathscr{L}_{\text{ST}}$-formula $(\exists x)(\varphi(x) \wedge (\forall y)(\varphi(y) \Rightarrow x = y))$.

Notice that by definition, if f is a function and $x \in \mathrm{dom}(f)$, then the set $\{y \in \mathrm{ran}(f) \big| (x, y) \in f\}$ is a singleton. Hence we can introduce the new notation (i.e. abbreviation of an $\mathscr{L}_{\mathrm{ST}}$-formula)

$$f(x) = \bigcup \{y \in \mathrm{ran}(f) \big| (x, y) \in f\}.$$

The definitions of $\mathrm{dom}(f), \mathrm{ran}(f), f[X], f^{-1}[Y], f \upharpoonright X$ are not different for functions than they are for binary relations; after all, a function is just a particular case of a binary relation. However, the definitions of some of these sets can be notationally streamlined when the relation under consideration is a function, as we can simply write $\mathrm{ran}(f) = \{f(x) \big| x \in \mathrm{dom}(f)\}$, $f[X] = \{f(x) \big| x \in X\}$, $f^{-1}[Y] = \{x \in \mathrm{dom}(f) \big| f(x) \in Y\}$, and $f \upharpoonright X = \{(x, f(x)) \big| x \in X\}$. We should probably say a few words regarding our "square-bracket" notation for set images and set pre-images, that is, for the difference between $f(x)$ and $f[X]$. In most other areas of mathematics, it is common to use round brackets (as in "$f(X)$") to denote the object that in these notes we are denoting by $f[X]$. We need to be careful when doing set theory, because everything is a set, and in particular any element X in the domain of the function f is itself a set, and as such it might very well happen to be a subset of $\mathrm{dom}(f)$, in which case there is a clear difference between $f(X)$ (the unique y such that $(X, y) \in f$) and $f[X]$ (the collection $\{f(x) \big| x \in X\}$). If we were using round parentheses for both of those objects, the way most non-set-theorists do, we would be faced with some unbearable ~~lightness of being~~ ambiguity, as the following example illustrates.

Example 18. Suppose that a and b are two arbitrary nonempty sets, with $a \notin b$, and consider the following function:

$$f = \{(\varnothing, a), (\{\varnothing\}, b)\}.$$

Then $\{\varnothing\} \subseteq \mathrm{dom}(f)$, but also $\{\varnothing\} \in \mathrm{dom}(f)$, and moreover

$$f(\{\varnothing\}) = b \neq \{a\} = f[\{\varnothing\}].$$

Having clarified that, we now proceed to introduce a piece of notation that is fairly commonly used in mathematics, and which from now on will officially be part of our $\mathscr{L}_{\mathrm{ST}}$ abbreviations.

Definition 19. Let A, B, f be sets. We declare that the sequence of symbols $f : A \longrightarrow B$ will abbreviate the following $\mathscr{L}_{\mathrm{ST}}$-formula:

$$(f \text{ is a function}) \wedge (\mathrm{dom}(f) = A) \wedge (\mathrm{ran}(f) \subseteq B).$$

Notice that, with this definition, if $f : A \longrightarrow B$ and $B \subseteq C$ then it is also the case that $f : A \longrightarrow C$. The whole notion of "codomain" is, generally speaking, quite foggy in set theory, since our definitions do not require us to specify a concrete codomain when talking about a function. As a result of this, we may accept any superset of the range of a function as a possible codomain for that function, and therefore we need to be very careful when stating the definition of surjectivity. For this concept to make sense, we will need to always specify explicitly the codomain of f in any sentence stating whether or not f is onto. The notion of injectivity, on the other hand, does not present any of these difficulties, and can be stated without having to specify any extra information about the function f.

Definition 20. Let f be a function.

1. We say that f is **injective** or **one-to-one** if $(\forall y \in \mathrm{dom}(f))(\exists! x)(y = f(x))$, or equivalently, $(\forall x, z \in \mathrm{dom}(f))(f(x) = f(z) \Rightarrow x = z)$.
2. Given a set Y satisfying $\mathrm{ran}(f) \subseteq Y$, we say that f is (surjective) **onto** Y if $Y = \mathrm{ran}(f)$, or equivalently, if $(\forall y \in Y)(\exists x \in \mathrm{dom}(f))(y = f(x))$. Sometimes we shall write "$f : \mathrm{dom}(f) \longrightarrow Y$ is onto" rather than "f is a function onto Y". This, as well as any other reasonable alternative, can be used, as long as one is careful to explicitly mention the set Y with respect to which the function f is onto.

Example 21. Assuming that we will eventually be able to implement the whole machinery of Calculus I within set theory (which we will definitely be able to do later in this course), we will then have the following examples:

1. $\sin : \mathbb{R} \longrightarrow \mathbb{R}$,
2. $\sin : \mathbb{R} \longrightarrow \mathbb{R}$ is not surjective, or equivalently,
3. $\sin$ is not onto $\mathbb{R}$,
4. $\sin : \mathbb{R} \longrightarrow [-1, 1]$,
5. $\sin : \mathbb{R} \longrightarrow [-1, 1]$ is surjective, or equivalently,

6. sin is onto $[-1,1]$,
7. sin is not injective,
8. $\sin\upharpoonright\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ is injective.

Just as it was the case for domain, range, images, pre-images and restrictions, our definition of the composition of two functions $g\circ f$ will be the same that we use for binary relations. We proceed right away to prove that this definition works in this context as intended.

Proposition 22. *If f and g are functions, then so is $g\circ f$.*

Proof. Certainly $g\circ f$ is a binary relation. Now let $x\in\mathrm{dom}(g\circ f)$, and suppose that there are z,z' such that $(x,z),(x,z')\in g\circ f$. This means that there exist y,y' such that $(x,y),(x,y')\in f$ and $(y,z),(y',z')\in g$. Since f is a function, we must have $y=y'$ and therefore we have $(y,z),(y,z')\in g$. Now since g is a function, we must have $z=z'$, and we are done. □

Given two functions f,g, Proposition 22 ensures that $g\circ f$ is a function. Thus, given an $x\in\mathrm{dom}(g\circ f)$, there is a unique z such that $(x,z)\in g\circ f$, and this in turn implies that there is a (unique) y such that $(x,y)\in f$ and $(y,z)\in g$. We can therefore write $y=f(x)$ and $z=g(y)=g(f(x))$. In other words, we have just recovered the formula $(g\circ f)(x)=g(f(x))$ that we have all known since we were little children.

It is necessary to alert the reader that dealing with compositions of functions might require some caution. Given any two functions f and g, one is typically (when working in any mathematics that is not set theory) urged to ensure that $\mathrm{ran}(f)\subseteq\mathrm{dom}(g)$ before mentioning the composition function $f\circ g$. In set theory, however, when one uses the definitions that we have provided in these notes, it turns out that the composition $f\circ g$ is always defined (and it will be a function by Proposition 22), regardless of any relation or lack thereof between $\mathrm{ran}(f)$ and $\mathrm{dom}(g)$. Notice, however, that it might sometimes be the case that $\mathrm{dom}(f\circ g)\subsetneq\mathrm{dom}(f)$. In fact, if we use consistently the definitions from these notes, we can see that $\mathrm{dom}(f\circ g)=\{x\in\mathrm{dom}(f)\big|f(x)\in\mathrm{dom}(g)\}=\mathrm{dom}(f)\cap f^{-1}[\mathrm{dom}(g)]$; in particular, if $\mathrm{ran}(f)\cap\mathrm{dom}(g)=\varnothing$ then $g\circ f=\varnothing$ (the empty function).

There is also no need to introduce a special definition for the inverse function f^{-1}, since we will, once again, continue to use the one that we have for arbitrary relations. However, if we are consistent with these definitions, we will see that f^{-1} always exists and is a relation. We urge the reader to check (by using the facts that $\mathrm{dom}(f^{-1})=\mathrm{ran}(f)$ and $\mathrm{ran}(f^{-1})=\mathrm{dom}(f)$) that the relation f^{-1} will be a function precisely when f is injective.

Definition 23. Given any set A we define $\mathrm{id}_A=\{(a,a)\big|a\in A\}$ (this relation is easily seen to be a function).

Note that, if f is injective, then $f\circ f^{-1}=\mathrm{id}_{\mathrm{ran}(f)}$ and $f^{-1}\circ f=\mathrm{id}_{\mathrm{dom}(f)}$. The following proposition gives us a generalization of the previous statement for functions for which we are considering an arbitrary codomain.

Proposition 24. *Let $f:A\longrightarrow B$. Then f is injective if and only if there exists $g:B\longrightarrow A$ satisfying $g\circ f=\mathrm{id}_A$.*

Proof. If there is a g satisfying the required condition, and $x,z\in\mathrm{dom}(f)$ are such that $f(x)=f(z)$, then we must have that $x=\mathrm{id}_A(x)=(g\circ f)(x)=g(f(x))=g(f(z))=(g\circ f)(z)=\mathrm{id}_A(z)=z$, which shows that f is injective.

Conversely, we know that if f is injective then $f^{-1}:\mathrm{ran}(f)\longrightarrow\mathrm{dom}(f)$. In order to obtain a function whose domain is all of B, rather than just $\mathrm{ran}(f)$, we pick an arbitrary (but fixed) $a\in A$ and we define

$$g=f^{-1}\cup\{(b,a)\in B\times\{a\}\big|b\in B\setminus\mathrm{ran}(f)\}.$$

It is straightforward to check that g is as required[7]. □

The attempt at stating an analogous (or *dual*) result for surjective functions will lead, in due time, to careful considerations regarding the Axiom of Choice.

Now we know hot to implement (or model, or mimic) functions of one variable, with domain A, as certain specific subsets of $A\times B$ for some B. In order to do the same with functions of several variables, all we need to do is to consider functions whose domain consists of ordered tuples. For example, the sum operation $+$ over the real numbers $\mathbb{R}$, which is a function of two real variables, can be thought of as a function $+:\mathbb{R}\times\mathbb{R}\longrightarrow\mathbb{R}$ (rather than taking "two real numbers" as input, we think of this as a function that takes ordered pairs of real numbers, and outputs another real number).

Another object that we might want to model is an "indexed family" $\{a_i\big|i\in I\}$. Our intuitive (pre-set-theoretic) idea of an indexed family is that of an object which does not satisfy extensionality. Rather, two indexed families $\{a_i\big|i\in I\}$ and $\{b_i\big|i\in I\}$ are considered to be equal if and only if $I=J$ and $(\forall i\in I)(a_i=b_i)$, in other words, both repetitions and

[7] Note that all we did is just to define the function g by $g(y)=\begin{cases}f^{-1}(y); & \text{if } y\in\mathrm{ran}(f)\\ a; & \text{otherwise.}\end{cases}$

the order of enumeration matter when it comes to indexed families. Ultimately, the information that uniquely specifies an indexed family is the precise assignment of an object (i.e. a set) a_i to each $i \in I$, so it makes sense to implement indexed families as functions. In other words, the object that we will call "the indexed family" $\{a_i \big| i \in I\}$ will be the function f with domain I such that $(\forall i \in I)(f(i) = a_i)$. This function f will be called the **indexing** of the family (and formally, within our theory, this indexing function will *be* the family). Thus any function f with domain I can be denoted (and thought of) as the indexed family $\{f(i) \big| i \in I\}$, although typically we will use symbols with subscripts, such as a_i, to denote the element $f(i)$. The result of "forgetting the indexing" of such a family (to only keep the elements that are enumerated in the indexing, while forgetting about repetitions and order in which those elements are) is just the set $\mathrm{ran}(f)$. Given an indexed family $\{a_i \big| i \in I\}$, with indexing function f, we can write down symbols such as $\bigcup_{i\in I} a_i$ to denote $\bigcup \mathrm{ran}(f)$, or $\bigcap_{i\in I} a_i$ to denote $\bigcap \mathrm{ran}(f)$.

Given two sets A, B, we will frequently need to consider the set of all functions $f : A \longrightarrow B$, which is usually denoted by either the symbol B^A, or by the symbol ${}^A B$. That is,

$$B^A = \{f \in \mathfrak{P}(A \times B) \big| f : A \longrightarrow B\},$$

which exists by Comprehension.

Example 25.

1. $\varnothing^A = \varnothing$ if $A \neq \varnothing$,
2. $\varnothing^\varnothing = \{\varnothing\}$,
3. $A^\varnothing = \{\varnothing\}$,
4. $\{0, 1\}^{\mathbb{N}} =$ collection of all infinite sequences of bits.

We finalize this section by providing the reader with a supply of fun facts that will have to be proven in the next homework assignment. Assuming that f is a function, it is the case that:

- $f^{-1}[\bigcup \mathscr{Y}] = \bigcup\{f^{-1}[Y] \big| Y \in \mathscr{Y}\}$,
- $f^{-1}[\bigcap \mathscr{Y}] = \bigcap\{f^{-1}[Y] \big| Y \in \mathscr{Y}\}$,
- $f^{-1}[\mathrm{ran}(f) \setminus Y] = \mathrm{dom}(f) \setminus f^{-1}[Y]$,
- $f[\bigcup \mathscr{X}] = \bigcup\{f[X] \big| X \in \mathscr{X}\}$,
- $f[\bigcap \mathscr{X}] \subseteq \bigcap\{f[X] \big| X \in \mathscr{X}\}$ (but the reverse inclusion is, in general, false),
- $f[\mathrm{dom}(f) \setminus X] = \mathrm{ran}(f) \setminus f[X]$.

2.2 Equivalence relations

We now proceed to discuss a specific kind of relations that will be of crucial importance in what follows, namely equivalence relations. The intuition behind the idea of an equivalence relation is that this will be a relation that behaves like some sort of "generalized equality". More concretely, our definition of an equivalence relation will simply be that of a relation satisfying the properties that are typically ascribed to the relation of equality. Recall that, for a relation R, we will write $a\ R\ b$ instead of $(a, b) \in R$.

Definition 26. Let A be a set. A relation $E \subseteq A \times A$ is said to be an **equivalence relation on** A if

- E is reflexive on A (recall that this means $(\forall a \in A)(a\ E\ a)$, or equivalently, $\mathrm{id}_A \subseteq E$),
- E is symmetric (recall that this means $(\forall a, b)(a\ E\ b \Rightarrow b\ E\ a)$, or equivalently, $E^{-1} \subseteq E)$[8],
- E is transitive (recall that this means that $(\forall a, b, c)((a\ E\ b \wedge b\ E\ c) \Rightarrow a\ E\ c)$, or equivalently, $E \circ E \subseteq E$).

Example 27. Assuming that we have successfully implemented objects such as $\mathbb{Z}$ and $\mathbb{R}$, we have the following examples.

1. Given an $n \in \mathbb{Z}$, the relation of congruence modulo n is given by

$$\equiv_{\mathrm{mod} n} = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \big| n \mid b - a\}$$

and it is an equivalence relation on $\mathbb{Z}$.

[8] Note that this implies, in fact, that $E = E^{-1}$. For if $(a, b) \in E$, that is, $a\ E\ b$, by reflexivity $b\ E\ a$ which means that $a\ E^{-1}\ b$, i.e. $(a, b) \in E^{-1}$. Thus $E \subseteq E^{-1}$, and we are done.

2. Given an arbitrary group G and a subgroup $H \subseteq G$, we can consider the relation of congruence modulo H (i.e. "belonging to the same coset"), which is given by

$$\equiv_H = \{(g,h) \in G \times G \big| h^{-1}g \in H\},$$

and is an equivalence relation on G.

3. If we let $\mathcal{T}$ be the set of all triangles in $\mathbb{R}^2$, the relation $\cong$ described by "being congruent to" (in other words, $\cong = \{(T_1,T_2) \in \mathcal{T} \times \mathcal{T} \big| T_1 \text{ is congruent to } T_2\}$) is an equivalence relation on $\mathcal{T}$.

4. Using the same set $\mathcal{T}$ of all triangles in $\mathbb{R}^2$ as above, the relation $\sim$ that corresponds to "being similar to" (that is, $\sim = \{(T_1,T_2)\mathcal{T} \times \mathcal{T} \big| T_1 \text{ is similar to } T_2\}$) is also an equivalence relation on $\mathcal{T}$.

We will next see that, in a sense, we do not need to require reflexivity for a relation to be an equivalence relation (as long as we are careful about *on which set* this will be an equivalence relation).

Theorem 28. *Let R be a symmetric and transitive relation. Then R is an equivalence relation on* $\operatorname{dom}(R)$.

Proof. We only need to prove that R is reflexive in $\operatorname{dom}(R)$, so let $a \in \operatorname{dom}(R)$. Then by definition, there exists a $b \in \operatorname{ran}(R)$ such that $a\ R\ b$. The assumption that R is reflexive yields $b\ R\ a$ as well, and then the last two relations imply that $a\ R\ a$ by transitivity. □

The following definition captures the general idea behind the concept of a quotient structure (such as a quotient group, quotient ring, quotient space, etc., that you might remember from your algebra or topology classes).

Definition 29. Let A be a set, and let E be an equivalence relation on A.

1. Given $a \in A$, we define the **equivalence class of a modulo E** to be the set of elements that are E-related (E-equivalent) to a:
$$[a]_E = \{b \in A \big| a\ E\ b\}.$$

2. We define the **quotient set of A modulo E** to be the set of all equivalence classes modulo E:
$$A/E = \{[a]_E \big| a \in A\} = \{x \in \mathfrak{P}(A) \big| (\exists a \in A)(x = [a]_E)\}$$
(the only purpose of the last equality is to show that the existence of this set follows from the Comprehension Axiom Scheme).

3. Lastly, we define the **canonical projection**, denoted by π_E, to be the function mapping each $a \in A$ to its equivalence class $[a]_E$ (writing out this set as $\pi_E = \{(a,y) \in A \times (A/E) \big| y = [a]_E\}$ ensures its existence, via the Comprehension Axiom scheme).

We now proceed to prove that equivalence classes have a very special property that makes them particularly well-behaved.

Proposition 30. *Let A be a set and let E be an equivalence relation on A. Then, any two E-equivalence classes are either equal or disjoint, that is, $(\forall a,b \in A)([a]_E = [b]_E \vee [a]_E \cap [b]_E = \varnothing)$.*

Proof. Let $a,b \in A$, and suppose that $[a]_E \cap [b]_E \neq \varnothing$. This means that there exists some $c \in [a]_E \cap [b]_E$, that is, $a\ E\ c$ and $b\ E\ c$. Taking an arbitrary element $d \in [a]_E$, we have that $a\ E\ d$, so by symmetry we also have $d\ E\ a$, which by transitivity (since $a\ E\ c$) implies that $d\ E\ c$; using symmetry again we get $c\ E\ d$, which by transitivity (since $b\ E\ c$) implies that $b\ E\ d$, that is, $d \in [b]_E$. We have thus shown that $[a]_E \subseteq [b]_E$; since this was shown for any two a,b such that $[a]_E \cap [b]_E \neq \varnothing$, it must also be true for b,a which means that the same argument establishes at once that $[b]_E \subseteq [a]_E$ as well. Thus $[a]_E = [b]_E$, and we are done. □

This means that, if E is an equivalence relation on the set A, then every element of A belongs to one and only one equivalence class. As we will see in a moment, this will provide us with another equivalent characterization of equivalence relations. We first state the following definition.

Definition 31. Given a set X, a **partition** of X is a family $\mathcal{P} \subseteq \mathfrak{P}(X)$ such that

1. $(\forall Y \in \mathcal{P})(Y \neq \varnothing)$,
2. $\bigcup \mathcal{P} = X$,
3. $(\forall Y,Z \in \mathcal{P})(Y \neq Z \Rightarrow Y \cap Z = \varnothing)$.

Thus, a partition is a collection of (nonempty) subsets of X such that every element of X belongs to one and only one element of the partition; in particular, note that the quotient set of a set A modulo an equivalence relation E constitutes a partition of A. Before we proceed to establish our characterization of equivalence relations, we mention the following example as the motivation for yet another characterization that we will be interested in.

Example 32. Let $f : A \longrightarrow B$. It is not hard to check that the relation E_f defined by

$$E_f = \{(a, b) \in A \times A \big| f(a) = f(b)\}$$

is an equivalence relation.

We are now in conditions to state our main theorem regarding equivalence relations, which consists in establishing that the concept of an equivalence relation is essentially equivalent on the one hand to the concept of a partition, and on the other hand to the concept of utilizing a function as we did in Example 32.

Theorem 33. *Let A be a set, and let $E \subseteq A \times A$ be a relation. The following are equivalent:*

1. *E is an equivalence relation,*
2. *there is a partition $\mathcal{P}$ of A such that $(\forall a, b \in A)(a \ E \ b \iff (\exists X \in \mathcal{P})(a, b \in X))$,*
3. *there exists a function f with $\mathrm{dom}(f) = A$ such that $(\forall a, b \in A)(a \ E \ b \iff f(a) = f(b))$.*

Proof.

1⇒2 If E is an equivalence relation, it follows from Proposition 30 that A/E is a partition of A satisfying the required condition.

2⇒3 If $\mathcal{P}$ is a partition of A, then the requirement that $\bigcup \mathcal{P} = A$ implies that every $a \in A$ belongs to some $X \in \mathcal{P}$, while the requirement that $(\forall X, Y \in \mathcal{P})(X \neq Y \Rightarrow X \cap Y = \varnothing)$ implies that such X is unique. Thus we can define $f : A \longrightarrow \mathcal{P}$ by $f(a) = \bigcup \{X \in \mathcal{P} \big| a \in X\}$ (that is, f maps each $a \in A$ to the unique $X \in \mathcal{P}$ to which a belongs). Given this f, we have that, for all $a, b \in A$, $a \ E \ b$ if and only if $(\exists X \in \mathcal{P})(a, b \in X)$ (by assumption), which happens if and only if $f(a) = f(b)$ (by definition of f).

3⇒1 This is just Example 32.

□

Regarding the equivalence between statements 1 and 2 in Theorem 33, we should note that there are very explicit descriptions of how to obtain a partition out of an equivalence relation and vice versa, in such a way that these "conversions" are inverses of each other. More concretely, given an equivalence relation E on the set A, the relevant partition $\mathcal{P}_E$ is just the quotient set A/E. Conversely, if $\mathcal{P}$ is a partition of A, the relevant equivalence relation $E_{\mathcal{P}} \subseteq A \times A$ is the one given by $a \ E \ b$ iff a and b belong to the same element of the partition. Note that $E_{\mathcal{P}_E} = E$ and $\mathcal{P}_{E_{\mathcal{P}}} = \mathcal{P}$.

We now finalize this section by stating a couple of results that provide information about defining functions whose domain and/or range is a quotient set[9].

Theorem 34. *Suppose that E is an equivalence relation on the set A and F is an equivalence relation on the set B. Let $f : A \longrightarrow B$. Then the following two conditions are equivalent:*

1. *$(\forall a, a' \in A)(a \ E \ a' \Rightarrow f(a) \ F \ f(a'))$,*
2. *there exists a function $\hat{f} : A/E \longrightarrow B/F$ such that the following diagram commutes:*

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\pi_E"] & B \arrow[d, "\pi_F"] \\
A/E \arrow[r, "\hat{f}"] & B/F
\end{tikzcd}$$

Proof. (2) ⇒ (1) Suppose that $\hat{f} : A/E \longrightarrow B/F$ is such that $\pi_F \circ f = \hat{f} \circ \pi_E$, and let $a, b \in A$ be such that $a \ E \ b$. Then $[a]_E = [b]_E$, and therefore $[f(a)]_F = \pi_F(f(a)) = (\pi_F \circ f)(a) = (\hat{f} \circ \pi_E)(a) = \hat{f}(\pi_E(a)) = \hat{f}([a]_E) = \hat{f}([b]_E) = \hat{f}(\pi_E(b)) = (\hat{f} \circ \pi_E)(b) = (\pi_F \circ f)(b) = \pi_F(f(b)) = [f(b)]_F$, hence $f(a)$ and $f(b)$ both lie in the same F-equivalence class, which means that $f(a) \ F \ f(b)$.

[9] These results are here for completeness, and to satisfy the curiosity of whomever finds herself reading these notes. But I have never proved them, or even stated them, in class, due to some very obvious time constraints.

(1) $\Rightarrow$ (2) Suppose that $(\forall a, b \in A)(a \; E \; b \Rightarrow f(a) = f(b))$. Let

$$\hat{f} = \{(x, y) \in (A/E) \times (B/F) \big| (\exists a \in A)(x = [a]_E \wedge y = [f(a)]_F\}.$$

Clearly $\hat{f}$ is a binary relation between A/E and B/F, with domain A/E. Now, to check that it is a function, we just need to assume that $(x, y), (x, y') \in \hat{f}$. This means that, for some $a, a' \in A$, we have $x = [a]_E$, $x = [a']_E$, $y = [f(a)]_F$, and $y' = [f(a')]_F$. However, the fact that $[a]_E = x = [a']_E$ means that $a \; E \; a'$, which by assumption implies that $f(a) \; F \; f(a')$. This means that $y = [f(a)]_F = [f(a')]_F = y'$, thus $\hat{f}$ satisfies the definition of a function.

Now, given an arbitrary $a \in A$, then by definition of $\hat{f}$ we have that $([a]_E, [f(a)]_F) \in \hat{f}$, which implies that $\hat{f}([a]_E) = [f(a)]_F$. Hence we have that $(\pi_F \circ f)(a) = \pi_F(f(a)) = [f(a)]_F = \hat{f}([a]_E) = \hat{f}(\pi_E(a)) = (\hat{f} \circ \pi_E)(a)$. This finishes the proof that $\pi_F \circ f = \hat{f} \circ \pi_E$, and we are done.

□

The following result is an immediate consequence of Theorem 35, in the particular case where the equivalence relation F on B is just plain equality, that is, when $F = \mathrm{id}_B = \{(b, b) \big| b \in B\}$ (modulo composing the obtained map $\hat{f}$ with the bijection $b \longmapsto [b]_F = \{b\}$ between B and B/F).

Theorem 35. *Suppose that $g : A \longrightarrow B$, and let E be an equivalence relation on A. Then the following two conditions are equivalent:*

1. $(\forall a, b \in A)(a \; E \; b \Rightarrow f(a) = f(b))$,
2. *there exists a function $\tilde{g} : A/E \longrightarrow B$ such that the following diagram commutes:*

$$\begin{tikzcd}
A \arrow[r, "g"] \arrow[d, "\pi_E"] & B \\
A/E \arrow[ur, "\tilde{g}"'] &
\end{tikzcd}$$

2.3 Partial orders

We now start to carefully analyze a notion which generalizes, in a very abstract setting, the concept of an order.

Definition 36. Let X be a set, and $R \subseteq X \times X$.

1. R is said to be a **partial order** if it is reflexive, antisymmetric and transitive.
2. R is said to be a **strict partial order** if it is irreflexive and transitive.

To remind the reader of the definitions that have not been used so far, we recall that a relation R is antisymmetric if $(\forall x, y)((x \; R \; y \wedge y \; R \; x) \Rightarrow x = y)$, equivalently, $R \cap R^{-1} \subseteq \mathrm{id}_{\mathrm{dom}(R)}$; and a relation R is irreflexive if $(\forall x)\neg(x \; R \; x)$, equivalently, $R \cap \mathrm{id}_{\mathrm{dom}(R)} = \varnothing$.

As a clear example of the contrast between these two definitions, we provide the following example.

Example 37. Assume that we already know what the set $\mathbb{R}$ consists of, as well as its usual order. Then,

1. $\{(x, y) \in \mathbb{R} \times \mathbb{R} \big| x \leq y\}$ is an example of a partial order,
2. $\{(x, y) \in \mathbb{R} \times \mathbb{R} \big| x < y\}$ is an example of a strict partial order.

A tedious but straightforward fact (which you will be asked to prove in the next assignment) is that the notions of a partial order and strict partial order convey essentially the same idea (which is not the same as saying that they are mathematically equivalent). Concretely, if R is a partial order, then the relation $\overline{R}$ given by $(x \; \overline{R} \; y) \iff ((x \; R \; y) \wedge (x \neq y))$ is a strict partial order. Conversely, given a strict partial order S, we can define the relation $\hat{S}$ given by $(x \; \hat{S} \; y) \iff ((x \; S \; y) \vee (x = y))$, which turns out to be a partial order. Moreover, $\hat{\overline{R}} = R$ and $\overline{\hat{S}} = S$. Hence these two concepts (partial order and strict partial order), though formally not the same, are equivalent; or rather, we can think of them as two different implementations of the same idea.

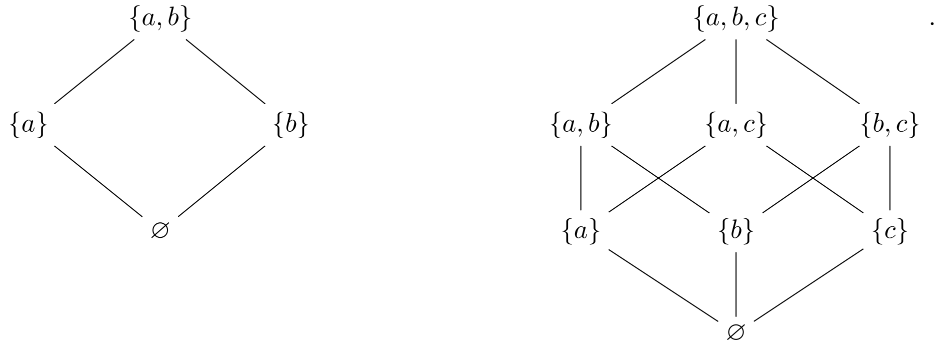
For the general theory, we will only use partial orders. When applying this theory in other contexts, sometimes it will be more useful to think about strict partial orders, in which case we will feel free to utilize the facts that we have proved about partial orders, appropriately adapted to strict partial orders. Partial orders will usually be denoted by symbols such as $\leq, \leqq, \preceq, \lesssim, \lessapprox, \precsim, \precapprox, \trianglelefteq$; when we want to refer to the corresponding strict partial orders, we will use the symbols $<, \lneqq, \prec, \lnsim, \lnapprox, \precnsim, \precnapprox, \triangleleft$, respectively.

Let us have a look at some more examples:

Example 38. As usual, we will assume that we know how to implement several everyday mathematical objects within set theory, for the sake of having more examples available.

1. $\{(n,m) \in \mathbb{N} \times \mathbb{N} \big| n \mid m\}$ is a partial order on $\mathbb{N}$.
2. Given a set X, $\{(Y,Z) \in \mathfrak{P}(X) \times \mathfrak{P}(X) \big| Y \subseteq Z\}$ is a partial order on $\mathfrak{P}(X)$.
3. If G is a group, then $\{(H,K) \in \mathfrak{P}(G) \times \mathfrak{P}(G) \big| (H, K \leq G) \wedge (H \subseteq K)\}$ is a partial order on the collection of all subgroups of G.
4. $\{(a,b) \in \mathbb{Z} \times \mathbb{Z} \big| (0 \leq a \leq b) \vee (b \leq a < 0) \vee (a < 0 \leq b)\}$ is a partial order[10] on $\mathbb{Z}$.
5. If $(X, \lesssim)$ and $(Y, \precsim)$ are two partial orders, we define the partial order $\trianglelefteq = \lesssim \times_{\text{lex}} \precsim$ by $((x,y) \trianglelefteq (x',y')) \iff ((x \lnsim x') \vee ((x = x') \wedge (y \precsim y'))$; this is known as the **lexicographic partial order** on $X \times Y$.

A graphical representation of the second example above, for the particular cases where $X = \{a,b\}$ is a set with two elements, and $X = \{a,b,c\}$ is a set with three elements, would be as follows:



Remark 39.

- Note that, if $\leq$ is a partial order, then it is not possible to have "cycles", that is, distinct elements $x_1, \ldots, x_n$ such that $x_1 \leq x_2 \leq \cdots \leq x_n \leq x_1$. If we had any such cycle, transitivity together with antisymmetry of $\leq$ would imply that $x_1 = \cdots = x_n$.
- Notice also that, given any partial order $\leq$, the relation $\leq^{-1}$ is also a partial order. The latter will typically be denoted with the symbol $\geq$.

Definition 40. A **partially ordered set** is an ordered pair $(X, \lesssim)$ such that $X \neq \varnothing$ and $\lesssim \subseteq X \times X$ is a partial order.

Definition 41. Let $(X, \leq)$ be a partially ordered set.

1. The element $x \in X$ is **minimal** if $(\forall y \in X)(y \leq x \Rightarrow y = x)$.
2. The element $x \in X$ is **maximal** if it is minimal with respect to $\geq$.
3. Given $Y \subseteq X$, the element $x \in Y$ is a **minimum** for Y if $(\forall y \in Y)(x \leq y)$.
4. Given $Y \subseteq X$, the element $x \in Y$ is a **maximum** for Y if it is a minimum for Y with respect to $\geq$.

Remark 42. Notice that, if $Y \subseteq X$ has a minimum, then this minimum is unique. For if $x, y \in Y$ are both minimums, then by definition we have $x \leq y$ and $y \leq x$, thus $x = y$. A completely analogous phenomenon happens with maximums[11]. Thus we introduce the notation

$$\min_{\leq}(Y) = \bigcup \{y \in Y \big| y \text{ is a minimum for } Y\},$$

if the set whose union we are taking in the right-hand side is nonempty, and with the exact same caveat we define

$$\max_{\leq}(Y) = \bigcup \{y \in Y \big| y \text{ is a maximum for } Y\}.$$

If the relation $\leq$ in question is clear from the context, we might omit its mention as a subindex and simply write $\min(Y)$ and $\max(Y)$.

[10] Pictorially, this partial order corresponds to arranging the elements of $\mathbb{Z}$ in the order $-1, -2, -3, \ldots, 0, 1, 2, 3, \ldots$.

[11] Some people like to say *minima* and *maxima* to refer to the plural forms of the words "minimum" and "maximum", like it would normally be done in latin. I have to admit that I do not know enough about the English language to properly assess whether carbon-copying grammar structures from another language is a legitimate thing to do, but I will point out that this practice looks highly suspicious to me.

Unlike minimums and maximums, minimal and maximal elements need not be unique. Notice also that the maximum of X must be a maximal element, and similarly the minimum is also a minimal element. For an example of a partially ordered set without minimal or maximal elements, consider the partially ordered set $(\mathbb{Z}, \leq)$ (the set of integers equipped with the usual ordering); for an example of a partially ordered set with minimal elements but without a minimum, consider the partially ordered set $(\mathbb{N} \setminus \{1\}, |)$ (the set of natural numbers greater than or equal to 2, equipped with the divisibility relation).

Definition 43. Let $(X, \leq)$ be a partially ordered set, and let $Y \subseteq X$.

1. The element $x \in X$ is a **lower bound** for Y if $(\forall y \in Y)(x \leq y)$,
2. The element $x \in X$ is an **upper bound** if it is a lower bound for Y with respect to $\geq$.
3. The element $x \in X$ is an **infimum** for Y if it is the greatest lower bound for Y, that is, if is equal to the element $\max\{y \in Y \big| y \text{ is a lower bound for } Y\}$. Hence if Y has an infimum then such infimum is unique, and so we denote that unique element by $\inf(Y)$.
4. The element $x \in X$ is a **supremum** for Y, denoted by $\sup(X)$ (if it exists, since in this case it is unique) if it is an infimum for Y with respect to $\geq$.

Example 44.

1. In the partially ordered set $(\mathfrak{P}(X), \subseteq)$, we have that, for any two $A, B \subseteq X$, $\sup\{A, B\} = A \cup B$ and $\inf\{A, B\} = A \cap B$. More generally, if $\mathcal{A} \subseteq \mathfrak{P}(X)$ and $\mathcal{A} \neq \varnothing$, then $\sup(\mathcal{A}) = \bigcup \mathcal{A}$ and $\inf(\mathcal{A}) = \bigcap \mathcal{A}$.
2. In the partially ordered set $(\mathbb{Z}, |)$, we have that for any two $n, m \in \mathbb{Z}$, $\sup\{n, m\} = \operatorname{lcm}(n, m)$ and $\inf\{n, m\} = \gcd(n, m)$.
3. In the partially ordered set $(\mathbb{Q}, \leq)$, consider the set $\{x \in \mathbb{Q} \big| x^2 < 2\}$. This set has upper bounds but no supremum. The same set, viewed as a subset of the partially ordered set $(\mathbb{R}, \leq)$, has a supremum, namely $\sqrt{2}$.

Definition 45.

1. In a partially ordered set $(X, \leq)$, two elements $x, y \in X$ are said to be **comparable** if $(x \leq y) \vee (y \leq x)$.
2. The partial order $\leq \subseteq X \times X$ is said to be a **total order** or a **linear order** if every two elements of X are comparable. In this case, the partially ordered set $(X, \leq)$ is said to be a **linearly ordered set**.

Note that, in the case of a strict partial order $<$ on the set X, the appropriate adaptation of the concept of comparability corresponds to the property that $(\forall x, y \in X)((x < y) \vee (y < x) \vee (x = y))$, and so a strict linear order is defined to be a strict partial order $< \subseteq X \times X$ that satisfies this property, known as *trychotomy*.

For examples of linearly order sets, we can consider $(\mathbb{R}, \leq)$ or $(\mathbb{N}, \leq)$; on the other hand, for examples of partially ordered sets that are not linearly ordered we can consider $(\mathfrak{P}(X), \subseteq)$ (as long as X has at least two elements) or $(\mathbb{N}, |)$.

2.4 Well-orders

The following will be one of the most important concepts in this course. Its usage throughout all of mathematics is extremely ubiquitous.

Definition 46. A partial order relation $\leq \subseteq X \times X$ will be said to be a **well-order** if it satisfies the property that $(\forall Y \subseteq X)(Y \neq \varnothing \Rightarrow (\exists y \in Y)(y = \min(Y)))$. The ordered pair $(X, \leq)$, where $\leq \subseteq X \times X$ is a well-order, will be called a **well-ordered set**.

Notice that every well-order is also a linear order, since for every two $x, y \in X$, the existence of $\min\{x, y\}$ means that either $x \leq y$ or $y \leq x$.

Example 47.

1. $(\mathbb{N}, \leq)$ is a well-ordered set.
2. $(\mathbb{Z}, \leq)$ is not a well-ordered set.

We need to be careful not to confuse the existence of minimums with the existence of infimums. For example, $[0,1] \subseteq \mathbb{R}$ is not well-ordered because the set $(0,1)$ does not have a minimum (even though it clearly has an infimum, namely 0).

Informally, we can think of well-ordered sets as sets that can be "counted", in the sense that we can go through its elements "one by one": if $(X, \leq)$ is a well-ordered set, then we can let $x_0 = \min(X)$, then $x_1 = \min(X \setminus \{x_0\})$, and so on, up until the point where we have all x_n, and after that we take $x_\omega = \min(X \setminus \{x_0, \ldots, x_n, \ldots\})$, and just generally keep going, until the process halts (we will see later that the fact that the process eventually halts is a consequence of the Axiom of Replacement). We emphasize that the procedure described in this paragraph is not something that we can formally carry out just yet, rather it is at this point just an intuitive idea that helps get a feeling for what well-ordered sets might look like.

Example 48. Consider (once again, assuming that we have successfully implemented the set $\mathbb{R}$ and its main features within set theory) the linearly ordered set $(\mathbb{R}, \leq)$.

1. The set $\{1 - \frac{1}{2^n} \big| n \in \mathbb{N}\}$ is a well-ordered set.
2. The set $\{m - \frac{1}{2^n} \big| m, n \in \mathbb{N}\}$ is also a well-ordered set.
3. The set $\{m - \frac{1}{2^n} - \frac{1}{2^k} \big| (k, m, n \in \mathbb{N}) \wedge (n < k)\}$ is also a well-ordered set.

Remark 49. The property of being a well-ordered set is *hereditary.* That is, if $(X, \leq)$ is a well-ordered set and $Y \subseteq X$, then $(Y, \leq\restriction Y)$ will also be a well-ordered set (normally, the symbol $\leq\restriction Y$ would denote the set $\{(x,y) \in \leq \big| x \in Y\}$, but in this context we take it to mean $\leq \cap (Y \times Y)$, which is a partial order relation on Y).

Definition 50. Given a well-ordered set $(X, \leq)$ and $x \in X$, we define the **initial segment** of x in X as follows:

$$\operatorname{seg}(x) = \{y \in X \big| y < x\}$$

(note that the inequality here is strict!).

One of the main features of well-ordered sets is the fact that it is possible to prove statements using induction on them. In this sense, they can be thought of as a generalization of the set $(\mathbb{N}, \leq)$.

Theorem 51 (Principle of Transfinite Induction)**.** *Let $(X, \leq)$ be a well-ordered set, and let $Y \subseteq X$. If Y satisfies the property that $(\forall x \in X)(\operatorname{seg}(x) \subseteq Y \Rightarrow x \in Y)$, then $X = Y$.*

Before proceeding to prove Theorem 51, let us analyze what this principle establishes when applied to the particular case of the well-ordered set $(\mathbb{N}, \leq)$. The hypothesis of the theorem for the subset $Y \subseteq \mathbb{N}$, that $(\forall n \in \mathbb{N})(\operatorname{seg}(n) \subseteq Y \Rightarrow n \in Y)$, can be split in to cases, namely $n = 1$ and $n > 1$. When $n = 1$, since $\operatorname{seg}(n) = \varnothing \subseteq Y$, the hypothesis just states that $1 \in Y$. On the other hand, for $n > 1$, this hypothesis can be rewritten as $(\forall k < n)(k \in Y) \Rightarrow (n \in Y)$, which is none other than the usual inductive step in a proof by "strong" induction. The conclusion is just that $Y = \mathbb{N}$, and so Theorem 51 subsumes the usual principle of induction in $\mathbb{N}$.

Proof of Theorem 51. Suppose not, that is, suppose that Y satisfies the property that $(\forall x \in X)(\operatorname{seg}(x) \subseteq Y \Rightarrow x \in Y)$[12], yet $Y \neq X$. This means that $X \setminus Y \neq \varnothing$, so since $\leq$ is a well-order, there must exist a minimum for that set, say $y = \min(X \setminus Y)$. Then we have that $y \notin Y$. However we also have that $\operatorname{seg}(y) \subseteq Y$: if $x \in \operatorname{seg}(y)$, it means that $x < y$ which in particular means that $\neg(y \leq x)$, so $x \notin (X \setminus Y)$ (since y is the minimum of the latter) and this must mean that $x \in Y$. Thus $\operatorname{seg}(y) \subseteq Y$, and so by our assumption on Y we conclude that $y \in Y$, a contradiction. □

We next state a result which is, in a sense, the converse of Theorem 51 for linear orders. This result can be interpreted as saying that the only linear orders in which you can perform proofs by induction are precisely the well-orders.

Theorem 52. *Let $\leq$ be a linear order on X whose only inductive set is X itself (that is, the only $Y \subseteq X$ with the property that $(\forall x \in X)(\operatorname{seg}(x) \subseteq Y \Rightarrow x \in Y)$ is X). Then $\leq$ is a well-order.*

Proof. Let $Y \subseteq X$ be a nonempty subset, and let $Z = \{x \in X \big| (\forall y \in Y)(x < y)\}$ be the set of all *strict* lower bounds for Y (notice the strict inequality!). In particular, $Z \cap Y = \varnothing$. If Z is inductive, this means $Z = X$, which can only be the case if $Y = \varnothing$, a contradiction. Hence $Z \neq X$ and so Z is not inductive, therefore there is a $z \in X$ such that $\operatorname{seg}(z) \subseteq Z$ but $z \notin Z$. We claim that z is a minimum for Y. To see this, we start by arguing that z is a lower bound for Y. For any $y \in Y$, since $\leq$ is linear we must have $y \leq z$ or $z \leq y$. If $y < z$ then $y \in \operatorname{seg}(z) \subseteq Z$, meaning that y is a strict lower bound for Y, contradicting that $y \in Y$. Hence we must have $z \leq y$, and since $y \in Y$ was arbitrary, we conclude that z is a lower bound for Y. If this lower bound was a strict lower bound, we would be able to conclude that $z \in Z$, contradicting the choice of z. Therefore it must be the case that this lower bound is not strict, in other words, $z \in Y$, and so $z = \min(Y)$. □

[12] From now on, we will call this property "Y is inductive on X", for short.

Now that we have seen that it is possible to prove statements by induction over well-ordered sets, we would like to also be able to define functions by recursion. In order to be able to state the appropriate theorem without making a notational mess, we introduce the following notation: whenever $(X, \leq)$ is a well-ordered set, and Y is any set, we let $Y^{<X} = \{f \subseteq X \times Y \big| (\exists x \in X)(f : \text{seg}(x) \longrightarrow Y)\}$.

Theorem 53 (Principle of Transfinite Recursion)**.** [13] *Let $(X, \leq)$ be a well-ordered set, let Y an arbitrary set, and suppose that $G : Y^{<X} \longrightarrow Y$. Then there exists a unique $F : X \longrightarrow Y$ such that $(\forall x \in X)(F(x) = G(F \restriction \text{seg}(x)))$.*

In order to be able to comprehend how this generalizes the usual principle of building functions by recursion on $\mathbb{N}$, we will explicitly explain the contents of Theorem 53 in the particular case when our well-ordered set is $\mathbb{N}$. Suppose, for example, that $G : \mathbb{N}^{<\mathbb{N}} \longrightarrow \mathbb{N}$ is the function given by $G(\{(1, a_1), \ldots, (n, a_n)\}) = (n + 1)a_n$, and $G(\varnothing) = 1$. Looking at the unique F whose existence is guaranteed by Theorem 53, we see that $F(1) = G(F \restriction \varnothing) = G(\varnothing) = 1$, and $F(n + 1) = G(F \restriction \{1, \ldots, n\}) = G(\{(1, F(1)), \ldots, (n, F(n))\}) = (n + 1)F(n)$. Those who have seen the usual recursive definition of the factorial function will immediately recognize that $F(n) = n!$ for each $n \in \mathbb{N}$.
For another example, let $G : \mathbb{N}^{<\mathbb{N}} \longrightarrow \mathbb{N}$ be the function given by $G(\{(1, a_1), \ldots, (n, a_n)\}) = a_n + a_{n-1}$ if $n \geq 2$, $G(\varnothing) = 1$, and $G(\{(1, a)\}) = 1$. Then the corresponding unique F whose existence is guaranteed by Theorem 53 would be given by $F(1) = G(F \restriction \varnothing) = G(\varnothing) = 1$, $F(2) = G(F \restriction \{1\}) = G(\{(1, F(1))\}) = 1$, and $F(n + 1) = G(F \restriction \{1, \ldots, n\}) = F(\{(1, F(1)), \ldots, (n - 1, F(n - 1)), (n, F(n))\}) = F(n) + F(n - 1)$ for $n \geq 1$. The reader should be able to recognize that $F(n)$ is thus the n-th Fibonacci number for all $n \in \mathbb{N}$.

Proof of Theorem 53. We will start by proving uniqueness under the assumption of existence, both because it is easier, and because uniqueness (of parts of the final function) will be used when constructing the actual function in order to prove existence. So suppose F, F' both satisfy the required property. We will now show that the set $Y = \{x \in X \big| F(x) = F'(x)\}$ is inductive, which will imply that $Y = X$ and so $F = F'$. To see this, assume that $x \in X$ is such that $\text{seg}(x) \subseteq Y$. Then for each $y \in \text{seg}(x)$, we have $F(y) = F'(y)$, and so we can conclude that $F \restriction \text{seg}(x) = F' \restriction \text{seg}(x)$. But then $F(x) = G(F \restriction \text{seg}(x)) = G(F' \restriction \text{seg}(x)) = F'(x)$ and so $x \in Y$, and we are done.

We now proceed to prove existence. The main difficulty is that defining F by means of the instance of the Comprehension Axiom that corresponds to the formula $F(x) = G(F \restriction \text{seg}(t))$ is not possible, since the symbol F itself occurs in that formula and so the definition would be circular. In other words, before F has been defined, we do not have any well-defined "macro" to replace the symbol F with a more extended $\mathscr{L}_{\text{ST}}$-formula, and so the symbols $F(x) = G(F \restriction \text{seg}(t))$ do not constitute yet a legit string of symbols for using in our theory. The way out of this dilemma is the same one that was first proposed by Dedekind, when he was justifying the analog of this theorem on the well-ordered set $\mathbb{N}$. Namely, replace the references to F with a reference to some other object that is only identified by the fact that it satisfies the property that F is should in the end satisfy. In other words, we define

$$Y = \{x \in X \big| (\exists F_x : \text{seg}(x) \longrightarrow Y)(\forall y \in \text{seg}(x))(F_x(y) = G(F_x \restriction \text{seg}(y)))\}.$$

Note that, for each $x \in Y$, the corresponding F_x witnessing that $x \in Y$ must be unique. This follows from the uniqueness argument in the first paragraph of this proof, along with the fact that $\text{seg}(x)$ is a well-ordered set (since the property of being well-ordered is hereditary). Now let us show that the set Y is inductive. Suppose that x is such that $\text{seg}(x) \subseteq Y$. There are three cases:

1. The first case is when $\text{seg}(x) = \varnothing$. Then the condition $\text{seg}(x) \subseteq Y$ is vacuous, and we just need to show the existence of F_x. For this, taking $F_x = \varnothing$ suffices.

2. Another possibility is that $\text{seg}(x)$ is nonempty and has a maximum element, which we will denote by y. Then $\text{seg}(x) = \text{seg}(y) \cup \{y\}$ and since we already have an appropriately behaved function $F_y : \text{seg}(y) \longrightarrow Y$, all we need to do is just extend this function to an F_x having y in its domain and satisfying that $F_x(y) = G(F_x \restriction \text{seg}(y)) = G(F_y)$. In other words, it suffices to let $F_x = F_y \cup \{(y, G(F_y))\}$.

3. Finally, we need to consider the case when $\text{seg}(x)$ is nonempty and has no maximum element. Then for each $y \in \text{seg}(x)$, there is a $z \in \text{seg}(x)$ such that $y \in \text{seg}(z)$ (if $y < x$, since $\text{seg}(x)$ has no maximum, there must be some $z \in \text{seg}(x)$ with $y < z$). This means that, letting

$$F_x = \left(\bigcup_{y \in \text{seg}(x)} F_y \right),$$

[13] During my first two iterations of teaching Math 582, I noticed that the proof of this theorem was particularly painful: it is much more complicated than both the immediately preceding and the immediately succeeding theorems, and it is not used for anything until after we have stated and proved an even more general version of this theorem, one whose proof is essentially the same anyways. Consequently, the proof of this result was omitted in the third iteration of this course, although the result itself was mentioned.

we will have that $F_x : \mathrm{seg}(x) \longrightarrow Y$ (the fact that F_x is a function is something that follows from the uniqueness argument in the first paragraph of this proof), and furthermore it is straightforward to verify that F_x satisfies the property required to make it into a witness that $x \in Y$. There is only one little detail left to check, namely that we need to be able to prove that this F_x exists. This will follow from the Axiom of Union as soon as we manage to prove that the set $\{F_y \big| y \in \mathrm{seg}(x)\}$ exists. But such an existential statement follows from the following instance of the Axiom of Comprehension:

$$\{F \in \mathfrak{P}(X \times Y) \big| (\exists y \in \mathrm{seg}(x))(F = F_y)\}$$

(where "$F = F_y$" is really a shorthand for the $\mathscr{L}_{\mathrm{ST}}$-formula $(\forall z \in \mathrm{seg}(y))(F(z) = G(F \restriction \mathrm{seg}(z))))$, as soon as we notice that this set contains exactly all F_y for $y \in \mathrm{seg}(x)$.

Having considered all three cases, we have finished showing that Y is an inductive subset of X, and hence $Y = X$. We now just need to construct F, which is something that can be done in exactly the same way that we built each of the F_x (namely, we need to consider three cases, according to whether $X = \varnothing$, $X \neq \varnothing$ has a maximum, or $X \neq \varnothing$ does not have a maximum, and in each case we essentially do the same that we just did). □

Chapter 3

The natural numbers, the integers, the rationals, the reals... and more

We have successfully implemented some mathematical objects of a general nature, such as functions and relations. It is now time to show how to implement some more concrete mathematical objects within set theory.

3.1 Peano Systems

We state the axioms for Peano Systems (not to be confused with the axioms of Peano Arithmetic[1]). Intuitively, a Peano System is a structure that behaves in exactly the way that we would expect the set of natural numbers to behave. We will see that our definition is "categorical", in the sense that there will be exactly one Peano System up to isomorphism. Hence, as soon as we manage to prove that there exists at least one Peano system (which will follow from the Axiom of Infinity), the possibility of having available an object that we can identify with the set of natural numbers will finally materialize.

Definition 54. A **Peano system** is a triple (N, s, i) such that $i \in N$ and $s : N \longrightarrow N$, satisfying:

1. $i \notin \operatorname{ran}(s)$,
2. s is injective,
3. whenever $X \subseteq N$ satisfies the following
   (a) $i \in X$,
   (b) $(\forall x \in X)(s(x) \in X)$,

   then in must be the case that $X = N$ (this is essentially second order induction).

Example 55. We will show a couple of non-examples, and one example. In both cases, we will think of a diagram representing the Peano system (N, s, i). Our diagram has one vertex per each element of N, and contains, at the base of each element $n \in N$, an arrow that points toward $s(n)$.

1. Triples (N, s, i) whose diagram is cyclic, or eventually cyclic, are non-examples of Peano systems.

$$\begin{tikzcd}
 & a \arrow[r, mapsto] & b \arrow[dr, mapsto] & \\
f \arrow[ur, mapsto] & & & c \arrow[dl, mapsto] \\
 & e \arrow[ul, mapsto] & d \arrow[l, mapsto] &
\end{tikzcd}$$

or

$$\begin{tikzcd}
a \arrow[r, mapsto] & b \arrow[r, mapsto] & c \arrow[dr, mapsto] & \\
 & h \arrow[ur, mapsto] & & d \arrow[d, mapsto] \\
 & g \arrow[u, mapsto] & f \arrow[l, mapsto] & e \arrow[l, mapsto]
\end{tikzcd} .$$

[1] The Axioms of Peano Arithmetic are first-order, and in particular the Axiom Scheme of Induction only applies to definable subsets. In contrast, the axioms that we will state here are, in a sense, second-order axioms, and in particular, our principle of Induction will apply to all sets.

2. If (N, s, i) is a Peano system, then we know that in the corresponding diagram, no arrow will be pointing towards i (because $i \notin \operatorname{ran}(s)$, and that no vertex has two incoming arrows (because s is injective). So essentially, there is a unique example of a (diagram of a) Peano system which looks as follows:

$$\begin{tikzcd} i \arrow[r, mapsto] & s(i) \arrow[r, mapsto] & s(s(i)) \arrow[r, mapsto] & s(s(s(i))) \arrow[r, mapsto] & {} & \cdots & {} \arrow[r, mapsto] & s^n(i) \arrow[r, mapsto] & {} & \cdots & . \end{tikzcd}$$

Proposition 56. *Let (N, s, i) be a Peano system, and let $x \in N$. Then either $x = i$ or $(\exists y \in N)(x = s(y))$ (that is, $N = \{i\} \cup \operatorname{ran}(s)$).*

Proof. It is easy to check that the set $X = \{x \in N \big| x = i \vee x \in \operatorname{ran}(s)\}$ satisfies that $i \in X$, and $(\forall x \in X)(s(x) \in X)$, so by property 3 of the definition of a Peano system (which we will from now on refer to as "the property of induction"), we conclude that $X = N$. □

We will now show how the property of induction will enable us to define functions by recursion, whenever the domain of the desired function is a Peano system[2].

Theorem 57 (Recursion Theorem for Peano systems)**.** *Let (N, s, i) be a Peano system, and let Z be any set. For any choice of $z \in Z$ and any $g : Z \longrightarrow Z$, there exists a unique $f : N \longrightarrow Z$ satisfying that $f(i) = z$ and $(\forall x \in N)(f(s(x)) = g(f(x)))$. This statement can be summarized by means of the following commutative diagram.*

$$\begin{tikzcd} & N \arrow[rr, "s"] \arrow[dd, "f"] & & N \arrow[dd, "\exists! f"] \\ \{*\} \arrow[ur, "i"] \arrow[dr, "\forall z"] & & & \\ & Z \arrow[rr, "\forall g"] & & Z \end{tikzcd}$$

Proof. As with the recursion theorem for well-orders, we find a difficulty in that we cannot use the property that $f(s(x)) = g(f(x))$ to define f using Comprehension, since this would require us to mention f inside its very definition. Hence we will need to do a few somersaults and cartwheels in order to prove this theorem; to this effect, we introduce a definition. A function $h \subseteq N \times Z$ will be called **acceptable** if $\operatorname{dom}(h) \subseteq N$, $\operatorname{ran}(h) \subseteq Z$, $(i, z) \in h$, and $(\forall x \in N)(s(x) \in \operatorname{dom}(h) \Rightarrow (x \in \operatorname{dom}(h) \wedge h(s(x)) = g(h(x))))$.

Now, an instance of the Axiom of Comprehension, plus one usage of the Axiom of Union, ensure the existence of

$$f = \bigcup \{h \subseteq N \times Z \big| h \text{ is an acceptable function}\}.$$

Clearly, $f \subseteq N \times Z$ is a binary relation, consisting of those ordered pairs (x, y) such that $y = h(x)$ for some acceptable function h satisfying $x \in \operatorname{dom}(h)$. We will use the principle of induction for Peano systems to show that $\operatorname{dom}(f) = N$. For this, first note that $i \in \operatorname{dom}(f)$ because $\{(i, z)\}$ is an acceptable function. Now for the inductive step, assume that $x \in \operatorname{dom}(f)$, which means that $x \in \operatorname{dom}(h)$ for some acceptable function h. If $s(x) \in \operatorname{dom}(h)$ we are done, otherwise let $h' = h \cup \{(s(x), g(h(x)))\}$ and notice that h' is also an acceptable function that satisfies $s(x) \in \operatorname{dom}(h')$. Hence in either case we get that $s(x) \in \operatorname{dom}(f)$, thus by the principle of induction we obtain $\operatorname{dom}(f) = N$.

Now, to show that f is actually a function, we define $X = \{x \in N \big| (\exists! y \in Z)((x, y) \in f)\}$, and we will show (using the principle of induction on a Peano system) that $X = N$. For the base case, just note that every acceptable function h must satisfy, by definition, that $i \in \operatorname{dom}(h)$ and $h(i) = z$, hence $(i, y) \in f$ implies $y = z$ and so $i \in X$. Now for the inductive step, suppose that $x \in X$, and let y be the unique element such that $(x, y) \in f$. This means that every acceptable function h with $x \in \operatorname{dom}(h)$ must satisfy $h(x) = y$. Now, whenever $(s(x), y') \in f$, it must be because for some acceptable function h with $s(x) \in \operatorname{dom}(h)$, it is the case that $y' = h(s(x))$. By the definition of acceptability, we must have that $x \in \operatorname{dom}(h)$ and $y' = h(s(x)) = g(h(x)) = g(y)$. Hence $g(y)$ is the unique element such that $(s(x), g(y)) \in f$, and this shows that $s(x) \in X$. Thus by the principle of induction, we can conclude that $X = N$, and so $f : N \longrightarrow Z$. To see that f satisfies the property required by the statement of the theorem, note that, for every acceptable function h, we must have that $f(i) = h(i) = z$; and for every $x \in N$, taking any acceptable function h with $s(x) \in \operatorname{dom}(h)$ we have that $x \in \operatorname{dom}(h)$ and $f(s(x)) = h(s(x)) = g(h(x)) = g(f(x))$.

Now to prove uniqueness of f, suppose that there is another f' satisfying the same conditions, and let $Y = \{x \in N \big| f(x) = f'(x)\}$. Clearly $i \in Y$, since $f(i) = z = f'(i)$. Now if $x \in Y$, that is, $f(x) = f'(x)$, then we will have that $f(s(x)) = g(f(x)) = g(f'(x)) = f'(s(x))$ and consequently $s(x) \in Y$. Thus by the principle of induction for Peano systems, we have that $Y = N$, meaning that $f = f'$. □

[2] Note that, since our Peano systems do not come equipped with any notion of partial order, it does not make sense to utilize any of the theorems that we have about well-ordered sets. Note also that the principle of induction for well-orders corresponds roughly to what is commonly known as the "principle of strong induction"; whereas the property of induction for Peano systems roughly corresponds to what is commonly known as the "principle of weak induction".

As a particular case of Theorem 57, suppose that (N, s, i) and (M, t, j) are two Peano systems. Then we have guaranteed the existence of functions $f : N \longrightarrow M$ and $g : M \longrightarrow N$ making the following couple of diagrams commutative:

$$\begin{tikzcd}
& N \arrow[r, "s"] \arrow[dd, "f"] & N \arrow[dd, "f"] \\
\{*\} \arrow[ur, "i"] \arrow[dr, "j"'] & & \\
& M \arrow[r, "t"] & M
\end{tikzcd}$$

and

$$\begin{tikzcd}
& M \arrow[r, "t"] \arrow[dd, "g"] & M \arrow[dd, "g"] . \\
\{*\} \arrow[ur, "j"] \arrow[dr, "i"'] & & \\
& N \arrow[r, "s"] & N
\end{tikzcd}$$

On the other hand, for every Peano system (N, s, i), we have that id_N is the unique function making the diagram

$$\begin{tikzcd}
& N \arrow[r, "s"] \arrow[dd, "\mathrm{id}_N"] & N \arrow[dd, "\mathrm{id}_N"] \\
\{*\} \arrow[ur, "i"] \arrow[dr, "i"'] & & \\
& N \arrow[r, "s"] & N
\end{tikzcd}$$

commutative. However, $g \circ f$ is another function making the previous diagram commutative, so by uniqueness we must have $g \circ f = \mathrm{id}_N$. On the other hand, replacing N with M and $g \circ f$ with $f \circ g$, we conclude also that $f \circ g = \mathrm{id}_M$. This shows that (N, s, i) is isomorphic to (M, t, j), since f and g are both morphisms, and inverses of each other. Therefore every two Peano systems must be isomorphic, in other words, there exists at most one Peano system, unique up to isomorphism.

3.1.1 Arithmetic in a Peano system

Definition 58. Let (N, s, i) be a Peano system. Given a fixed $x \in N$, we use the recursion theorem, Theorem 57, to define a function A_x (denoted this way because the function will "add to x") by specifying that $A_x(i) = s(x)$ and $A_x(s(y)) = s(A_x(y))$. For each $x \in N$, this gives us a unique function $A_x : N \longrightarrow N$; we can now define a binary operation $+ : N \times N \longrightarrow N$ given by $+(x, y) = A_x(y)$[3].

We will usually write $x + y$ instead of $+(x, y)$; this binary operation will be called **addition**. A more intuitive way of writing down this definition (which is the way in which it is normally written) would be as follows:

- $x + i = s(x)$,
- $x + s(y) = s(x + y)$.

As an example of a property of addition that we can prove using the principle of induction for Peano systems is associativity. That is, we will prove, by induction on z, that $(\forall x, y, z \in N)((x + y) + z = x + (y + z))$.

Base case If $z = i$, then $(x + y) + i = s(x + y) = x + s(y) = x + (y + i)$.

Inductive step Suppose that $(x+y)+z = x+(y+z)$. Now $(x+y)+s(z) = s((x+y)+z) = s(x+(y+z)) = x+s(y+z) = x + (y + s(z))$, and we are done.

[3] Strictly speaking, we would need to mention the use of an instance of the Axiom of Comprehension, in order to justify the existence of the set

$$+ = \{((x, y), z) \in (N \times N) \times N \big| (\exists A \in N^N)(A(i) = s(x) \wedge (\forall w \in N)(A(s(w)) = s(A(w))) \wedge A(y) = z)\}.$$

Having the associativity property under our belt, we can now prove that $+$ is commutative, that is, we will prove by induction on y that $(\forall x, y \in N)(x + y = y + x)$.

Base case We need to show that $x + i = i + x$ for all $x \in N$. This will, in and of itself, be a proof by induction on x.

Base case If $x = i$, we have $i + i = i + i$ and we're done.

Inductive step Suppose that $x + i = i + x$, now notice that $s(x) + i = s(s(x)) = s(x + i) = s(i + x) = i + s(x)$, and we are done.

The above induction completes the base case of our main induction.

Recursive step Suppose we have now that $x + y = y + x$, and let us look at $x + s(y)$. We have that $x + s(y) = s(x + y) = s(y + x) = y + s(x) = y + (x + i) = y + (i + x) = (y + i) + x = s(y) + x$.

Definition 59. We will now proceed to define the binary operation of **product** $\cdot : N \times N \longrightarrow N$ in a Peano system (N, s, i). We do this using the recursion theorem, Theorem 57, by[4]:

- $x \cdot i = x$,
- $x \cdot s(y) = x \cdot y + x$.

We now provide a list of properties of addition and multiplication that can be proved, using the recursive definitions and what we have proved before, by means of the principle of induction on a Peano system. You will now spend the rest of today's class proving these.

1. $(\forall x, y, z \in N)(x + z = y + z \Rightarrow x = y)$, (*hint*: induction on z, this one doesn't really involve $\cdot$)
2. $(\forall x, y, z \in N)(x \cdot (y + z) = x \cdot y + x \cdot z)$, (*hint*: induction on z)
3. $(\forall x, y, z \in N)((x \cdot y) \cdot z = x \cdot (y \cdot z))$, (*hint*: induction on z)
4. $(\forall x, y, z \in N)((x + y) \cdot z = x \cdot z + y \cdot z)$, (*hint*: induction on z)
5. $(\forall x, y, z \in N)(x \cdot y = y \cdot x)$, (*hint*: induction on y, base step requires induction on x)

Extra Exercise: can you define exponentiation using the same ideas?

3.1.2 The order relation in a Peano system

Having defined the usual arithmetic operations, we now proceed to define the ordering on a Peano system.

Definition 60. Let (N, s, i) be a Peano system. We define the binary relation $<$ by

$$<= \{(x, y) \in N \times N \big| (\exists z \in N)(y = x + z)\}.$$

Proposition 61. *The binary relation $<$ defined above is a (strict) linear order on N.*

Proof. We need to prove three things about $<$, namely, that it is irreflexive, that it is transitive, and that any two (distinct) elements are comparable.

To see that $<$ is irreflexive, let us proceed by proving, by induction on $x \in N$, that $\neg(x < x)$. The base case is easy, if $x = i$ then having $i < i$ would mean that, for some $x \in N$, $i = i + x = x + i = s(x)$, which is impossible. Now suppose that $\neg(x < x)$. To show that $\neg(s(x) < s(x))$, suppose otherwise. Then we have that $s(x) < s(x)$, which means that, for some z, it is the case that $s(x) = s(x) + z = z + s(x) = s(z + x) = s(x + z)$. Since s is injective, we obtain that $x = x + z$, meaning that $x < x$, contradicting our inductive hypothesis.

Now to show that $<$ is transitive, suppose that $x < y$ and $y < z$. Then there are $v, w \in X$ such that $y = x + v$ and $z = y + w = (v + x) + w = x + (v + w)$, which readily means that $x < z$.

Let us now proceed to show, by induction on $x \in N$, that $(\forall y \in N)(x < y \lor x = y \lor y < x)$.

Base case If $x = i$, any $y \in N$ satisfies that either $y = i$, or $y = s(z)$ for some z, by Proposition 56. In the first case we have $i = x = y = i$; in the second case we have that $y = s(z) = z + i = i + z$, which means by definition that $x = i < y$.

[4] Formally, we would need to define first a function that multiplies times x, for each fixed x, and then somehow glue all of these functions together, along the lines of what we did in Definition 58. We will, however, write our recursive definition for this binary operation in the most economical way, knowing that this could be fully formalized, if we really wanted to, in the same way that is done for addition.

Recursive step Suppose that x is such that $(\forall y \in N)(x < y \vee x = y \vee y < x)$. Letting $y \in N$ be arbitrary, we now want to show that $s(x)$ and y are comparable. By induction hypothesis, one of $x < y$, $x = y$ or $y < x$ holds; each of these gives rise to a different case to be considered.

- If $x = y$, then $s(x) = s(y) = y + i$, which means that $y < s(x)$.
- If $y < x$, then this means that $x = y + z$ for some $z \in N$. Therefore, $s(x) = s(y + z) = y + s(z)$, which implies that $y < s(x)$.
- $x < y$, which means that $y = x + z$ for some z. By Proposition 56, either $z = i$, or $z = s(w)$ for some $w \in N$. In the first case we have that $y = x + i = s(x)$, whereas in the second case we obtain $y = x + s(w) = s(x + w) = s(w + x) = w + s(x) = s(x) + w$, which implies that $s(x) < y$. In either case we get that $s(x) = y$ or $s(x) < y$, and we are done.

□

Hence $<$ is a linear order. This will allow us to tackle the following exercise, and we will proceed afterwards to show that $<$ is actually a (strict) well-ordering.

Exercise: $(\forall x, y, z \in N)(x \cdot z = y \cdot z \Rightarrow x = y)$.

Remark 62. Suppose that we have a Peano system (N, s, i), and consider its corresponding order $<$. Let us analyze what initial segments (that is, sets of the form $\{y \in N \big| y < x\}$ for some $x \in N$) look like. Let us first look at $\mathrm{seg}(i)$. In order for $x \in N$ to satisfy $x < i$, one must have $i = x + z$ for some $z \in N$. By Proposition 56, there are two cases, either $z = i$ or $z = s(w)$ for some $w \in N$. If $z = i$, then $i = x + z = x + i = s(x)$, which is impossible; otherwise, if $z = s(w)$, then $i = x + s(w) = s(x + w)$, which is also impossible. Therefore $\mathrm{seg}(i) = \varnothing$.

Now, every element of N that does not equal i is of the form $s(x)$. Note that, if $y < s(x)$, then $s(x) = y + z$ for some $z \in N$. Once again we have two cases, according to whether $z = i$ or $z = s(w)$ for some $w \in N$. In the second case we have that $s(x) = y + z = y + s(w) = s(y + w)$, since s is injective, this imples that $x = y + w$, i.e. $y < x$. In the first case, if $z = i$ then $s(x) = y + z = y + i = s(y)$, which by injectivity of s implies that $x = y$. Hence $\mathrm{seg}(s(x)) = \{y \in N \big| y < x \vee y = x\} = \mathrm{seg}(x) \cup \{x\}$.

The previous remark will be instrumental for the following proof that $<$ is a well-order.

Theorem 63. *Let (N, s, i) be a Peano system, and let $<$ be its corresponding linear order defined as above. Then $<$ is a (strict) well-order.*

Proof. Our strategy will consist in proving that the only inductive subset of N (that is, the only $X \subseteq N$ with the property that $(\forall x \in N)(\mathrm{seg}(x) \subseteq X \Rightarrow x \in X)$) is N itself. Since $(N, <)$ is a (strictly) linearly ordered set, by Theorem 52 we will conclude that $<$ is a (strict) well-ordering.

Following our plan, let us assume that $X \subseteq N$ is an inductive set, that is, a set with the property that $(\forall x \in N)(\mathrm{seg}(x) \subseteq X \Rightarrow x \in X)$. Let $Y = \{x \in N \big| \mathrm{seg}(x) \subseteq X\}$. We will show, using the principle of induction for Peano systems, that $Y = N$. This will conclude our proof, since $Y = N$ implies that, for each $x \in N$, $s(x) \in Y$, meaning that $x \in \mathrm{seg}(s(x)) \subseteq X$ and so $X = N$. In order to carry out our proof that $Y = N$, we need to show that $i \in Y$ and that $(\forall x \in N)(x \in Y \Rightarrow s(x) \in Y)$.

Base case It is vacuously the case that $\mathrm{seg}(i) = \varnothing \subseteq X$, so $i \in Y$.

Inductive step Suppose that $x \in Y$, that is, $\mathrm{seg}(x) \subseteq X$. Our assumption about X then implies that $x \in X$. Since $\mathrm{seg}(s(x)) = \mathrm{seg}(x) \cup \{x\} \subseteq X$, we can conclude that $s(x) \in Y$, and we are done with our induction.

□

To conclude this subsection, we will remark that the order $<$ is compatible with the sum. This means that, for each $x, y, z \in N$, $x < y \iff z + x = z + y$. To see this, note that if $x < y$, meaning that $y = x + w$ for some $w \in N$, then $y + z = x + w + z = (x + z) + w$, and therefore $x + z < y + z$. Conversely, if $x + z < y + z$, meaning that $y + z = x + z + w$ for some $w \in N$, then by the cancellation rule proved earlier we will have that $y = x + w$, and thus $x < y$. The order $<$ is also compatible with the product, but we will not prove this here. Rather, this will occur as an exercise in the next problem set.

3.1.3 There exists one Peano system

So far, we have proved many properties of Peano systems. Amongst these, we have uniqueness of Peano systems up to isomorphism, but we still have not said a single word about the existence of a Peano system. If we succeed in proving that there exists a Peano system, we would have implemented, within set theory, an object that behaves the way we would normally expect the set of natural numbers to behave. Before we do that, however, we will first introduce an implementation of each individual natural number within our theory. To this effect, keep in mind that the following "definition" is really a statement that we utter from the viewpoint of the metatheory, in which we relate each of the natural numbers from our metatheory (that is, each of the "actual" natural numbers, the ones that live in the Platonic world of ideas) to some specific element of our theory ZFC. To emphasize this, we will (for a very short time) use Quine's quotation marks to differenciate between the "actual" natural number n, and the set $\ulcorner n \urcorner$ that will represent (or "implement") the number n within set theory.

Definition 64. Recursively define $\ulcorner n \urcorner$, for a non-negative integer n, as follows:

- $\ulcorner 0 \urcorner = \varnothing$,
- $\ulcorner n+1 \urcorner = \ulcorner n \urcorner \cup \{\ulcorner n \urcorner\}$.

An induction in the metatheory allows us to see that $\ulcorner n \urcorner$ is the set whose elements are precisely the sets of the form $\ulcorner k \urcorner$, for k a predecessor of n. This is clearly true if $n = 0 = \varnothing$; now if $\ulcorner n \urcorner = \{\ulcorner 0 \urcorner, \dots, \ulcorner n-1 \urcorner\}$, then the set $\ulcorner n+1 \urcorner = \ulcorner n \urcorner \cup \{\ulcorner n \urcorner\} = \{\ulcorner 0 \urcorner, \dots, \ulcorner n-1 \urcorner, \ulcorner n \urcorner\}$ contains precisely those elements $\ulcorner k \urcorner$ for those k that precede $n+1$.

We now state the Axiom of Infinity, which is essentially the assumption that there is a set containing all of the $\ulcorner n \urcorner$ at once.

Axiom of Infinity There exists a set X satisfying: $\varnothing \in X$ and $(\forall x \in X)(x \cup \{x\} \in X)$.

Any set satisfying the conditions ensured by the axiom above will be called an **inductive set**. Thus the Axiom of Infinity ensures the existence of an inductive set. A straightforward induction shows that, if X is an inductive set, then $\ulcorner n \urcorner \in X$ for all natural numbers n. The definition of an inductive set does not preclude, however, the possibility that said set contains other extraneous (spurious?) elements that do not have the form $\ulcorner n \urcorner$. If we want a set that (intuitively) contains exclusively the $\ulcorner n \urcorner$ for n an natural number, it looks as though we need to define the **smallest inductive set**.

Definition 65. Take an inductive set X by the axiom of infinity. Define

$$\omega = \{x \in X \big| (\forall Y)(Y \text{ is inductive} \Rightarrow x \in Y)\},$$

and define $\mathbb{N} = \omega \setminus \{0\}$.

Note that it is easy to prove that ω itself must be an inductive set, and that it must be a subset of every inductive set. Intuitively, the idea is that ω contains all of the $\ulcorner n \urcorner$ (because every inductive set must contain those), and nothing more. From now on we will drop the Quinean quotes $\ulcorner 0 \urcorner, \ulcorner 1 \urcorner, \dots$ and write simply $0, 1, \dots$, hoping that context will make it clear whether we are referring to the numbers in the metatheory, or the corresponding sets that we have defined within our theory. We proceed to exhibit a Peano system.

Definition 66. Define the **successor function** $S : \omega \longrightarrow \omega$ to be the one given by $S(n) = n \cup \{n\}$. Abusing notation, we will also denote by S the restriction $S \restriction \mathbb{N}$ of the function S just defined to the set $\mathbb{N}$.

Note that the fact that ω is an inductive set is what implies that $S(n) \in \omega$ whenever $n \in \omega$. The following lemma shows that this function S will have a left inverse, and in particular, it will be injective.

Lemma 67. *Given any $n \in \omega$, it is the case that $\bigcup S(n) = n$. In other words, the function $\bigcup : \omega \longrightarrow \omega$ is a left inverse for the function S, in the sense that $\bigcup \circ S = \mathrm{id}_\omega$.*

Proof. This will be done, essentially, by induction. Formally, we will show that the set

$$X = \{n \in \omega \big| \bigcup S(n) = n\},$$

is inductive, from which it will follow that $\omega \subseteq X$, which implies (since we are operating under the assumption that $X \subseteq \omega$) that $\omega = X$. In particular, we will have $(\forall n \in \omega)(\bigcup S(n) = n)$.

To see that X is an inductive set, note first that

$$\bigcup S(\varnothing) = \bigcup(\varnothing \cup \{\varnothing\}) = \bigcup\{\varnothing\} = \varnothing,$$

thus $\varnothing \in X$. Now suppose that $\bigcup S(x) = x$, and observe that

$$\bigcup S(S(x)) = \bigcup(S(x) \cup \{S(x)\} = \left(\bigcup S(x)\right) \cup S(x) = x \cup S(x) = x \cup (x \cup \{x\}) = (x \cup x) \cup \{x\} = x \cup \{x\} = S(x),$$

thus $S(x) \in X$ and we are done. □

Corollary 68. *The function S is injective.*

Proof. If $n, m \in \omega$ are such that $S(n) = S(m)$, then $n = \bigcup S(n) = \bigcup S(m) = m$ by Lemma 67. □

Theorem 69. $(\mathbb{N}, S, 1)$ *is a Peano system.*

Proof. We need to check the three conditions in the definition of a Peano system. The previous corollary shows that S is injective, which is one of the conditions. Now if $\{0\} = 1 = S(n)$ for some n, then it must be the case that $n = \bigcup S(n) = \bigcup 1 = \bigcup\{0\} = 0 \notin \mathbb{N}$, thus $1 \notin S[\mathbb{N}]$ (equivalently $1 \notin \operatorname{ran}(S \restriction \mathbb{N})$). This takes care of another condition.

Finally, we need to check that $(\mathbb{N}, S, 1)$ satisfies the principle of induction for Peano systems, so suppose that $A \subseteq \mathbb{N}$ is such that $1 \in A$ and $(\forall n \in \mathbb{N})(n \in A \Rightarrow S(n) \in A)$. Essentially, A is an inductive set except for the fact that it does not contain $\varnothing$. So we adjoin this element, and we proceed to show that $A \cup \{\varnothing\}$ is an inductive set. Clearly $\varnothing \in A \cup \{\varnothing\}$. Now, for every $n \in A \cup \{\varnothing\}$, it is the case (if $n \in A$, by our assumption on A, and if $n = 0$, by our assumption that $S(0) = 1 \in A$) that $S(n) \in A \subseteq A \cup \{\varnothing\}$. Hence $\omega \subseteq A \cup \{\varnothing\}$, and therefore $\omega = A \cup \{\varnothing\}$, which means that $A = (A \cup \{\varnothing\}) \setminus \{\varnothing\} = \omega \setminus \{\varnothing\} = \mathbb{N}$. This finishes the proof. □

We have seen that the Axiom of Infinity implies the existence of a Peano system. In a sense, the Axiom of Infinity is in fact equivalent to the existence of a Peano system (that is, if we were to replace the Axiom of Infinity with the statement "there exists a Peano system", then we would still be able to prove the exact same theorems). In other words, if we drop the Axiom of Infinity from our list of axioms, and assume the existence of a Peano system instead, then it is possible to prove the statement of the Axiom of Infinity as a theorem –but we will not prove this here and now, since the proof employs nontrivially the Axiom of Replacement–.

3.2 The integers

Our objective now is to implement the set of integers within set theory. In other words, we would like to define a structure (a set equipped with addition and multiplication operations, along with an appropriate linear order relation), containing an isomorphic copy of the natural numbers, where every element has an additive inverse. We would like to define this structure so that it follows the idea of taking the natural numbers and then just "adjoining their negatives".

There are many options for how to do this. A very intuitive one would be to define the set

$$\mathbb{Z} = (2 \times \mathbb{N}) \cup \{0\} = (\{0\} \times \mathbb{N}) \cup (\{1\} \times \mathbb{N}) \cup \{0\},$$

so that elements of the form $(0, n)$ are identified with n, and elements of the form $(1, n)$ are identified with $-n$. We would now need to define addition between two elements of $\mathbb{Z}$, and this would need to be done by cases, as in

$$\begin{cases} x + 0 = 0 + x = x \text{ for every } x \in \mathbb{Z}; \\ (i, n) + (j, m) = \begin{cases} (i, n + m) \text{ if } i = j; \\ 0 \text{ if } i \neq j \text{ and } n = m; \\ (i, n - m) \text{ if } i \neq j \text{ and } m < n; \\ (j, m - n) \text{ if } i \neq j \text{ and } n < m \end{cases} \end{cases}$$

The reader should observe that defining addition in this implementation of $\mathbb{Z}$ is really cumbersome, due to the need to break things up into so many cases. Proving that this addition satisfies the relevant properties (commutativity, associativity, etc.) would multiply the number of cases to consider, and defining multiplication would raise the level of cumbersomeness by a few orders of magnitude. In other words, this particular implementation, although technically doable, would be extremely painful. Thus we will take an alternative path, which the reader (student?) should work out by herself through the exercises that conform Appendix B.

3.3 The rational numbers

Just like we did for the integers, we now attempt to implement the rational numbers within set theory. In the case of $\mathbb{Z}$, we defined an equivalence relation that intuitively reflected the idea of taking differences of natural numbers, and that gave us a system with natural numbers plus their negatives. Now, for $\mathbb{Q}$, we will do the analogous thing –defining an equivalence relation reflecting the idea that we are taking quotients of integers, and this will give us a system with the integers plus their multiplicative inverses–. The student (reader?) should proceed to work out the details by herself, using the problem set which appears in this document as Appendix C.

3.4 The real numbers

It is time now to implement, within set theory, a structure that behaves the way the set of real numbers (with its addition, multiplication, and order) should. There are (at least) two different ways of doing this; choosing either one of them does not make an important difference, since the corresponding structures that arise from each of these constructions are isomorphic. We will describe the first one somewhat briefly, and then we will look at the second one with much more detail.

3.4.1 Dedekind cuts

The first construction consists in taking the so-called *Dedeking completion* of $\mathbb{Q}$. The main idea is that $\mathbb{Q}$ is somewhat "incomplete" because it has nonempty bounded subsets without a supremum (for example, the set $\{q \in \mathbb{Q} \big| q < 0 \vee q^2 < 2\}$, whose supremum should intuitively be $\sqrt{2}$, but $\mathbb{Q}$ has no such element). Thus we aim to embed $\mathbb{Q}$ within a larger structure that has the property that every nonempty bounded subset has a supremum. This is done by means of Dedekind cuts. A Dedekind cut is just an initial segment of $\mathbb{Q}$ that is closed downward, in other words:

Definition 70.

1. A set $x \subseteq \mathbb{Q}$ will be called a **Dedekind cut** if
   (a) $\varnothing \neq x \neq \mathbb{Q}$,
   (b) $(\forall q \in x)(\forall r \in \mathbb{Q})(r < q \Rightarrow r \in x)$, and
   (c) x has no maximum.
2. We define the set $\mathbb{R} = \{x \subseteq \mathbb{Q} \big| x \text{ is a Dedekind cut}\}$.
3. The partial order relation $\leq \subseteq \mathbb{R} \times \mathbb{R}$ is defined by $x \leq y \iff x \subseteq y$. In other words, $\leq = \subseteq \cap (\mathbb{R} \times \mathbb{R})$.

The embedding of $\mathbb{Q}$ into $\mathbb{R}$ is given by $q \longmapsto \{r \in \mathbb{Q} \big| r < q\}$. Moreover, it is easy to check that, if $\varnothing \neq A \subseteq \mathbb{R}$ is bounded, then $\bigcup A$ is a Dedekind cut, and $\bigcup A = \sup(A)$. Thus, from the perspective of "completing" $\mathbb{Q}$ as a linearly ordered set, the Dedekind cuts construction works quite straightforwardly. On the other hand, from the perspective of defining appropriate arithmetic operations on $\mathbb{R}$, things are a bit more complicated.

Definition 71. We define $+ : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ by $x + y = \{q \in \mathbb{Q} \big| (\exists r \in x)(\exists s \in y)(q < r + s)\}$,

It needs to be shown that $x + y$ is well-defined, in the sense that the set $x + y$ as defined above is a Dedekind cut whenever x and y are. At this point we will start omitting most details, but ultimately it turns out that $(\mathbb{R}, +)$ is an abelian group (with neutral element $0_{\mathbb{R}} = \{q \in \mathbb{Q} \big| q < 0\}$, and where the cut x has additive inverse given by $-x = \{q \in \mathbb{Q} \big| (\exists r > q)(-r \notin x)\}$).

The definition of multiplication is significantly more cumbersome: given $x, y \in \mathbb{R}$, we define

$$xy = 0_{\mathbb{R}} \cup \{qr \big| (0 \leq q \in x) \wedge (0 \leq s \in y)\}$$

if $x, y \geq 0_{\mathbb{R}}$, with other different definitions for the remaining cases that arise, depending on how x and y compare to $0_{\mathbb{R}}$. Quite clearly, proving with full details that this operation has the necessary properties (commutativity, associativity, distributivity over $+$, etc.) is bound to provoke severe discomfort upon the innocent and unsuspecting detail-checker. It is partly for this reason that we do not develop the full details of this particular way of constructing $\mathbb{R}$ here. But the interested reader can consult Enderton's textbook for a more detailed exposition.

3.4.2 Cauchy sequences

In order to talk about completing the field of rational numbers by means of Cauchy sequences, we first talk about **norms**.

Definition 72. If F is a field, a **norm** of F is a function $N : F \longrightarrow \mathbb{Q}$ satisfying the following properties:

1. $(\forall x \in F)(N(x) \geq 0 \wedge (N(x) = 0 \iff x = 0))$,
2. $(\forall x, y \in F)(N(xy) = N(x)N(y))$,
3. $(\forall x, y \in F)(N(x + y) \leq N(x) + N(y))$.

This is a very abstract definition, which consequently has its great generality as an advantage. The first observation is that Definition 72 above is a provisional definition, which will work for now. Once we have successfully constructed the real line $\mathbb{R}$, then we will have access to the definition of norm that is regularly used throughout mathematics –namely, the exact same as in Definition 72, but replacing each occurence of $\mathbb{Q}$ with $\mathbb{R}$–. The second observation is that there are many norms that can be defined on $\mathbb{Q}$, of which we will consider one to be especially important, since it is the one that will allow us to construct the real line $\mathbb{R}$.

Definition 73. We define a norm in the field $\mathbb{Q}$ that will be called the **absolute value**, denoted by N_0, and defined as follows:

$$N_0(q) = \begin{cases} q \text{ if } 0 \leq q, \\ -q \text{ if } q < 0. \end{cases}$$

It is routine to check that the function N_0 satisfies the conditions stipulated in Definition 72.

From now on, we will proceed with a general description of how to build a "completion" of a field F using an arbitrary norm function N. This has the advantage of providing us with multiple constructions at once, one for each possible field F and norm in F. For example, if we were to take the p-adic norm N_p on the field $\mathbb{Q}$[5], then the construction that we will describe gives rise to the field $\mathbb{Q}_p$ of p-adic numbers. On the other hand, our implementation of the real line $\mathbb{R}$ will be none other than the completion of $\mathbb{Q}$ with respect to the absolute value norm N_0 from Definition 73. So from now on, we fix an arbitrary field F and an arbitrary norm function N on F, and proceed to describe the completion construction in general terms.

Remark 74. If F is a field and $N : F \longrightarrow \mathbb{Q}$ is a norm function, then the following properties hold:

1. $N(1) = 1$: This is because $N(1) = N(1 \cdot 1) = N(1)N(1)$, and since $N(1) \neq 0$ we conclude that $N(1) = 1$.
2. $N(-1) = 1$: This is because $1 = N(1) = N((-1) \cdot (-1)) = N(-1)N(-1) = N(-1)^2$; now since $N(-1) \geq 0$, we can conclude that $N(-1) = 1$.
3. For every $x \in F$, we have $N(-x) = N((-1) \cdot x) = N(-1)N(x) = N(x)$.

To proceed with our construction, we need to develop a theory very similar to the theory of convergence that you guys probably learned in your analysis (calculus?) course. The first difficulty would be to somehow get rid of the quantifier "for all positive real numbers", since we do not know yet what real numbers are. However, this difficulty is not hard to overcome, as requiring that something can be made "$< \varepsilon$ for all $\varepsilon > 0$" is equivalent to requiring that it can be made $< \frac{1}{n}$ for all possible $n \in \mathbb{N}$.

Definition 75.

- An element $\vec{x} = \langle x_n \big| n \in \mathbb{N} \rangle \in F^{\mathbb{N}}$ will be said to be a **Cauchy sequence** if

$$(\forall n \in \mathbb{N})(\exists M \in \mathbb{N})(\forall m, k \in \mathbb{N}) \left((m \geq M \wedge k \geq M) \Rightarrow N(x_m - x_k) < \frac{1}{n} \right)$$

- Let $\mathscr{C} = \{\vec{x} \in F^{\mathbb{N}} \big| \vec{x} \text{ is a Cauchy sequence}\}$. Define the relation $\sim \subseteq \mathscr{C} \times \mathscr{C}$ by $\vec{x} \sim \vec{y}$ if and only if

$$(\forall n \in \mathbb{N})(\exists M \in \mathbb{N})(\forall m \in \mathbb{N}) \left(m \geq M \Rightarrow N(x_m - y_m) < \frac{1}{n} \right).$$

Proposition 76. *The relation $\sim$ is an equivalence relation on $\mathscr{C}$.*

Proof. Reflexivity and symmetry are quite trivial. To see transitivity, suppose that $\vec{x} \sim \vec{y} \sim \vec{z}$, and take an arbitrary $n \in \mathbb{N}$. then there are M_1, M_2 such that if $m \geq M_1$ then $N(x_m - y_m) < \frac{1}{2n}$, and if $m \geq M_2$ then $N(y_m - z_m) < \frac{1}{2n}$. Thus if we let $M = \max\{M_1, M_2\}$, then whenever $m \geq M$ we will have that

$$N(x_m - z_m) = N((x_m - y_m) + (y_m - z_m)) \leq N(x_m - y_m) + N(y_m - z_m) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n},$$

thus $\vec{x} \sim \vec{z}$ and we are done. □

We are now ready to define the completion of the field F with respect to the norm N.

[5] Recall that, if p is a prime number, then the p-adic norm N_p is defined as follows: Given $q = \frac{a}{b} \in \mathbb{Q}$, with $a, b \in \mathbb{Z}$ and $(a, b) = 1$, let n (respectively m) be the exponent of the prime number p in the decomposition of a (respectively b) in prime numbers (note, in particular, that only one of n, m can be nonzero, since we are assuming that a and b are coprime). Then the p-adic norm of q is defined to be $p^{m-n} = \frac{1}{p^{n-m}}$.

Definition 77.

- We define the **completion** of the field F with respect to the norm N by $F_N = \mathscr{C}/\sim$.
- We define the binary operation $+ : F_N \times F_N \longrightarrow F_N$ by

$$[\vec{x}]_\sim + [\vec{y}]_\sim = [\langle x_n + y_n \big| n \in \mathbb{N}\rangle]_\sim.$$

Proposition 78. *The binary operation $+$ thus defined on F_N is well-defined.*

Proof. We first proceed to prove that the sequence $\langle x_n + y_n \big| n \in \mathbb{N}\rangle$ is indeed a Cauchy sequence, so let $n \in \mathbb{N}$. Pick M_1, M_2 such that $m, k \geq M_1 \Rightarrow N(x_m - x_k) < \frac{1}{2n}$ and $m, k \geq M_2 \Rightarrow N(y_m - y_k) < \frac{1}{2n}$. Then if $m, k \geq \max\{M_1, M_2\}$, we have that

$$N((x_m + y_m) - (x_k + y_k)) = N((x_m - x_k) + (y_m - y_k)) \leq N(x_m - x_k) + N(y_m - y_k) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$

Now, to prove that $+$ is well-defined, suppose that $\vec{x}, \vec{y}, \vec{z}, \vec{w} \in \mathscr{C}$ are such that $\vec{x} \sim \vec{z}$ and $\vec{y} \sim \vec{w}$. To show that $\vec{x} + \vec{y} \sim \vec{z} + \vec{w}$, take an arbitrary $n \in \mathbb{N}$. Pick M_1, M_2 such that $m \geq M_1 \Rightarrow N(x_m - z_m) < \frac{1}{2n}$ and $m \geq M_2 \Rightarrow N(y_m - w_m) < \frac{1}{2n}$. Then if $m \geq \max\{M_1, M_2\}$, we have that

$$N((x_m + y_m) - (z_m + w_m)) = N((x_m - z_m) + (y_m - w_m)) \leq N(x_m - z_m) + N(y_m - w_m) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$

□

It is not extremely complicated to prove that the addition operation, thus defined on the set F_N, behaves the way we expect it would.

Proposition 79. $(F_N, +)$ *is an abelian group.*

Proof. Straighforward: commutativity and associativity follow easily from the commutativity and associativity of addition in the field F. The additive identity is just the Cauchy sequence with constant term 0, and given $[\vec{x}]_\sim$, its additive inverse is just $[\langle -x_n \big| n < \omega\rangle]_\sim$. □

Before we work out the definition of product in F_N, and start proving its properties, we will need a lemma.

Lemma 80. *Every Cauchy sequence is bounded.*

Proof. Let $\vec{x}$ be a Cauchy sequence, and grab $M \in \mathbb{N}$ such that $m \geq M \Rightarrow N(x_m) - N(x_M) \leq N(x_m - x_M) < 1$. Then for every $m \in \mathbb{N}$, we have that

$$N(x_m) < \max\{1 + N(x_M), N(x_{M-1}), \dots, N(x_1)\}.$$

□

Definition 81. We define the binary operation $\cdot : F_N \times F_N \longrightarrow F_n$ by

$$[\vec{x}]_\sim \cdot [\vec{y}]_\sim = [\langle x_n y_n \big| n \in \mathbb{N}\rangle]_\sim$$

Proposition 82. *The operation $\cdot : F_N \times F_N \longrightarrow F_N$ is well-defined.*

Proof. We begin by proving that, if $\vec{x}$ and $\vec{y}$ are Cauchy sequences, then the sequence $\langle x_n y_n \big| n \in \mathbb{N}\rangle$ is also a Cauchy sequence. By Lemma 80, there are bounds K_1 for $\vec{x}$ and K_2 for $\vec{y}$, so let $K = \max\{K_1, K_2\}$. The usual reasoning allows us to obtain, given an $n \in \mathbb{N}$, some M such that whenever $m, k \geq M$, $N(x_m - x_k) < \frac{1}{2Kn}$ and $N(y_m - y_k) < \frac{1}{2Kn}$. Thus we obtain that $N(x_m - x_k)N(y_m) < \frac{1}{2n}$ and $N(y_m - y_k)N(x_k) < \frac{1}{2n}$. Therefore

$$N(x_m y_m - x_k y_k) \leq N(x_m y_m - x_k y_m) + N(y_m x_k - y_k x_k) = N(x_m - x_k)N(y_m) + N(y_m - y_k)N(x_k) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{n}.$$

Now, to prove that $\cdot$ is well-defined, suppose that $\vec{x} \sim \vec{z}$ and $\vec{y} \sim \vec{w}$. Let $n \in \mathbb{N}$ be arbitrary. Pick K to be a bound for both $\vec{y}$ and $\vec{z}$, and grab an M such that $m \geq M$ implies $N(x_m - z_m) < \frac{1}{2Kn}$ and $N(y_m - w_m) < \frac{1}{2Kn}$. Then whenever $m \geq M$, we will have

$$N(x_m y_m - z_m w_m) \leq N(x_m y_m - z_m y_m) + N(z_m y_m - z_m w_m) = N(x_m - z_m)N(y_m) + N(y_m - w_m)N(z_m) < \frac{1}{2n} + \frac{1}{2n} = \frac{1}{2}.$$

□

Proposition 83. $(F_N, +, \cdot)$ *is a commutative ring with identity.*

Proof. Straightforward: commutativity and associativity of $\cdot$, as well as distributivity of $\cdot$ over $+$, follow from the analogous properties of the analogous operations in F. The multiplicative identity is just the Cauchy sequence with constant term equal to 1. □

In fact, much more than that is true.

Theorem 84. $(F_N, +, \cdot)$ *is a field.*

Proof. All we need to do is to prove the existence of multiplicative inverses, so let $[\vec{x}]_\sim \in F_N$ be distinct from $[\langle 0 \big| n < \omega \rangle]_\sim$. Then there exists an n such that for infinitely many m, we have $N(x_m) \geq \frac{1}{n}$. If we take M such that $m, k \geq M \Rightarrow N(x_m - x_k) < \frac{1}{2n}$, then picking any $m \geq M$ such that $N(x_m) \geq \frac{1}{2n}$ we have that, for arbitrary $k \geq M$, it must be the case that $N(x_k) > 0$, in fact, $N(x_k) > \frac{1}{2n}$. Thus we can define the sequence

$$[\vec{x}]_\sim^{-1} = \langle \underbrace{1, \dots, 1}_{m \text{ times}}, \frac{1}{x_{m+1}}, \frac{1}{x_{m+2}}, \dots \rangle].$$

Then clearly $[\vec{x}]_\sim \cdot [\vec{x}]_\sim^{-1} = 1$ (the obvious representative sequence is eventually equal to 1), provided we can show that the above sequence is in fact a Cauchy sequence. To show this, let $n' \in \mathbb{N}$ be arbitrary. Begin by picking n', and remember that $\frac{1}{2n}$ is a lower bound for the terms $N(x_k)$, $k \geq M$, which implies that $\frac{1}{N(x_k)} \leq 2n$ for such k. Grab an M' such that $k, k' \geq M' \Rightarrow N(x_k - x_{k'}) < \frac{1}{4n^2 n'}$. Then whenever $k, k' \geq \max\{m, M'\}$ we will have

$$N\left(\frac{1}{x_k} - \frac{1}{x_{k'}}\right) = N\left(\frac{x_{k'} - x_k}{x_k x_{k'}}\right) = \frac{N(x_{k'} - x_k)}{N(x_k)N(x_{k'})} < \frac{1}{4n^2 n'}(2n)^2 = \frac{1}{n'}.$$

□

A more straightforward way of proving that F_N is a field, provided that one knows some more theory (concretely, the theory of rings and quotients), is to take the set $\mathscr{C}$ of Cauchy sequences on F and define ring operations there just by taking the ring operations on F coordinatewise. Then, one can define a **null sequence** in $\mathscr{C}$ to be a sequence $\langle x_n \big| n < \omega \rangle$ such that $(\forall n \in \mathbb{N})(\exists M \in \mathbb{N})(\forall m \geq M)(N(x_n) < \frac{1}{n})$ (in other words, null sequences are precisely those that are related to the sequence which is constantly zero, according to our equivalence relation $\sim$). Denoting the set of null sequences by $\mathscr{N}$, it turns out that $\mathscr{N}$ is actually an ideal of $\mathscr{C}$, and a maximal ideal at that (to prove maximality, one essentially follows the proof of Theorem 84). Hence the quotient ring $\mathscr{C}/\mathscr{N}$ is a field, which we define to be F_N; in fact, our equivalence relation $\sim$ is exactly the same as being congruent modulo the ideal $\mathscr{N}$.

Be that as it may, the field F embeds into the field F_N by means of the mapping that sends each $x \in F$ to the constant Cauchy sequence $\langle x \big| n < \omega \rangle$. It is pretty obvious that these field operations are properly preserved by such mapping, which additionally is clearly injective.

The above finishes the construction of the completion of a field F with respect to the norm $N : F \longrightarrow \mathbb{Q}$. We will now assume that the field F is ordered, that is, that it comes equipped with a partial order relation $\leq$ which is compatible with the field operations (in the sense that $x \leq y$ iff $x + z \leq y + z$ for all $z \in F$, and $x \leq y$ iff $xz \leq yz$ whenever $z \geq 0$). In this case, we will also be able to turn the completion F_N of F into an ordered field.

Definition 85. Define $<\subseteq F_N \times F_N$ by $[\vec{x}]_\sim \leq [\vec{y}]_\sim$ if and only if $(\vec{x} \nsim \vec{y}) \wedge (\exists M \in \mathbb{N})(\forall m \in \mathbb{N})(m \geq M \Rightarrow x_m < y_m)$.

Proposition 86. *The relation $\leq$ is well-defined and trychotomous.*

Proof. For well-definedness, suppose that $\vec{x} \sim \vec{z}$ and $\vec{y} \sim \vec{w}$, and that $[\vec{x}]_\sim < [\vec{y}]_\sim$. Since $\vec{x} \nsim \vec{y}$, there exists an n such that for infinitely m, we have $|x_m - y_m| \geq \frac{1}{n}$. Pick an M sufficiently large so that, whenever $m, k \geq M$, we have

- $N(x_m - x_k) < \frac{1}{4n}$,
- $N(y_m - y_k) < \frac{1}{4n}$,
- $N(x_k - z_k) < \frac{1}{4n}$,
- $N(y_k - w_k) < \frac{1}{4n}$, and
- $x_m < y_m$.

Then, letting $m \geq M$ be such that $N(x_m - y_m) \geq \frac{1}{n}$, and $k \geq M$ arbitrary, we have that

$$N(x_m - z_k) \leq N(x_m - x_k) + N(x_k - z_k) < 2\frac{1}{8n} = \frac{1}{2n},$$

and in an entirely analogous way we can compute that $N(y_m - w_k) < \frac{1}{2n}$. These two inequalities, together with $N(x_m - y_m) \geq \frac{1}{n}$, are easily seen to imply that $z_k < w_k$, which takes care of well-definedness.

Now for trychotomy, if we suppose that $\vec{x} \nsim \vec{y}$, then for certain $n \in \mathbb{N}$ it is the case that $N(x_m - y_m) \geq \frac{1}{n}$ for infinitely many m. Choose M such that $m, k \geq M \Rightarrow N(x_m - x_k) < \frac{1}{2n} \wedge N(y_m - y_k) < \frac{1}{2n}$, so that if such an m satisfies $N(x_m - y_m) \geq \frac{1}{n}$, and $k \geq M$ is arbitrary, then we will have $x_m < y_m \iff x_k < y_k$. □

From now on, although the constructions that we will develop work well for any ordered field F, we run into some technical difficulties when the codomain of our norms is just $\mathbb{Q}$, rather than $\mathbb{R}$. For this reason, we will conclude our general treatment of field completions, and specialize now to the case where we complete the ordered field $\mathbb{Q}$ with respect to the absolute value norm N_0.

Definition 87. We define $\mathbb{R}$ to be the completion $\mathbb{Q}_{N_0}$ of $\mathbb{Q}$ with respect to the absolute value norm N_0. The set $\mathbb{R}$ will be equipped with the addition and multiplication operations, as well as with the ordering relation, that have been described above.

The following is a property somewhat specific to the absolute value norm N_0. It no longer translates automatically to the more general setting.

Proposition 88. *$\mathbb{R}$ has the **Archimedean property**, that is, whenever $[\vec{x}]_\sim, [\vec{y}]_\sim \in \mathbb{R}$, we have $[\vec{x}]_\sim > 0 \Rightarrow (\exists n \in \mathbb{N})(n[\vec{x}]_\sim > [\vec{y}]_\sim)$.*

Proof. Since $\vec{x} > \langle 0 \big| n < \omega\rangle$, then there is an n such that for infinitely many m, $x_m > \frac{1}{n}$. Thus if K is an upper bound for $\vec{y}$, we will have that, for infinitely many m, $(2Kn)x_m > 2K$, which for sufficiently large k will imply $(2Kn)x_k > y_k$ since for such k we will have $N(x_m - x_k) < K$, meaning that $x_k > K$. □

We now define the absolute value function $|\cdot| : \mathbb{R} \longrightarrow \mathbb{R}$ as usual (i.e. $|x|$ equals x if $x \geq 0$, and $-x$ otherwise). In your assignment for this week, you will be asked to prove that, once one identifies $\mathbb{Q}$ with its image under the embedding $: \mathbb{Q} \longrightarrow \mathbb{R}$, it is the case that $\mathbb{Q}$ is dense in $\mathbb{R}$, in the sense that for every $\varepsilon > 0$ and every $x \in \mathbb{R}$, there exists a $q \in \mathbb{Q}$ such that $|x - q| < \varepsilon$.

Now, to conclude the construction of $\mathbb{R}$ and the proofs of its main properties, we set out to outline the proof that every Cauchy sequence in $\mathbb{R}$ converges to a limit in $\mathbb{R}$. First notice that the definition of a Cauchy sequence has to be restated for sequences that are elements of $\mathbb{R}^{\mathbb{N}}$ rather than $\mathbb{Q}^{\mathbb{N}}$ (we do this verbatim, except we use the absolute value function instead of a norm). Also remember that, if $\vec{x} \in \mathbb{R}^{\mathbb{N}}$ is a sequence, then $\vec{x}$ said to converge to $l \in \mathbb{R}$ if and only if $(\forall n \in \mathbb{N})(\exists M \in \mathbb{N})(\forall m \geq M)(N(x_m - l) < \frac{1}{n})$; and $\vec{x}$ is said to converge if there exists some $l \in \mathbb{R}$ such that $\vec{x}$ converges to l.

Theorem 89. *$\mathbb{R}$ is complete. That is, every Cauchy sequence in $\mathbb{R}^{\mathbb{N}}$ converges.*

Proof. We begin by pointing out a sequence of key points (perhaps we should call each of these a "lemma") about Cauchy sequences. Each of these, except for the last one, apply equally well to sequences of rational numbers and to sequences of real numbers (and the proofs are identical in both cases). The last key point, however, only applies to Cauchy sequences of rational numbers.

1. If $\vec{x} = \langle x_n \big| n < \omega\rangle$ is a Cauchy sequence, and $\langle x_{n_k} \big| k < \omega\rangle$ is a subsequence[6] of $\vec{x}$ that converges to l, then $\vec{x}$ itself converges to l as well. For, given $n \in \mathbb{N}$, one chooses an M such that on the one hand, $k \geq M \Rightarrow N(x_{n_m} - l) < \frac{1}{2n}$, and on the other hand, $k, m \geq M \Rightarrow N(x_m - x_{n_m}) < \frac{1}{2n}$. These two inequalities, when put together, and after a routine application of the triangle inequality, yield that $N(x_m - l) < \frac{1}{n}$, for all $m \geq M$.

2. If $\vec{x}$ is a sequence such that $(\forall n \in \mathbb{N})(N(x_n - x_{n+1}) < \frac{1}{2^n})$, then $\vec{x}$ is a Cauchy sequence: for if $n \in \mathbb{N}$ is given and M is such that $\frac{1}{2^{M-1}} < \frac{1}{n}$, then whenever $m, k \geq M$ we have that

$$\begin{aligned} N(x_m - x_k) &\leq N(x_m - x_{m+1}) + N(x_{m+1} - x_{m+2}) + \cdots + N(x_{k-1} - x_k) \\ &< \frac{1}{2^m} + \frac{1}{2^{m+1}} + \cdots + \frac{1}{2^k} \\ &\leq \sum_{i=m}^{\infty} \frac{1}{2^i} = \frac{1}{2^{m-1}} \leq \frac{1}{2^{M-1}} < \frac{1}{n}. \end{aligned}$$

[6] Formally speaking, we a subsequence of $\vec{x}$ is given by considering $\vec{x} \circ \vec{n}$, where $\vec{n} = \langle n_k \big| k \in \mathbb{N}\rangle \in \mathbb{N}^{\mathbb{N}}$ is a strictly increasing sequence. Suggestively, this new sequence is denoted by $\langle x_{n_k} \big| n < \omega\rangle$.

3. If $\vec{x}$ is a Cauchy sequence, then there is a subsequence $\langle x_{n_k} \big| k \in \mathbb{N}\rangle$ satisfying $(\forall k \in \mathbb{N})(N(x_{n_k} - x_{n_{k+1}}) < \frac{1}{2^k})$: to see this, we just need to recursively define

$$n_1 = \min \left\{ M \in \mathbb{N} \middle| (\forall m, k \geq M) \left(N(x_m - x_k) < \frac{1}{2} \right) \right\},$$

and

$$n_{k+1} = \min \left\{ M \in \mathbb{N} \middle| (M \geq n_k) \wedge (\forall m, m' \geq M) \left(N(x_m - x_{m'}) < \frac{1}{2^{k+1}} \right) \right\}.$$

4. A Cauchy sequence of rational numbers $\vec{x} = \mathbb{Q}^{\mathbb{N}}$ is equivalent to (that is, it determines the same real number as) all of its subsequences. This one follows directly from writing down what it means for $\vec{x}$ to be a Cauchy sequence, and from reading what it means for $\vec{x}$ to be equivalent to a given subsequence.

We now proceed to simply outline the end of the proof, leaving the gory details to the reader. Start by taking a Cauchy sequence $\langle x^n \big| n < \omega\rangle \in \mathbb{R}^{\mathbb{N}}$. Using the facts above, assume without loss of generality that for every $n < \omega$ we have $|x^n - x^{n+1}| < 2^n$. Now suppose that the enumeration $\langle q_n \big| n < \omega\rangle$ constitutes a bijection between the natural numbers and the rational numbers (we will justify in Chapter 5 that such a bijection exists), and using density of the rationals, for each $n < \omega$ let k_n be the least integer such that $|q_{k_n} - x_n| < 2^{n+1}$, now define the sequence of rational numbers $\vec{y} = \langle y_n \big| n < \omega\rangle$ to be given by $y_n = q_{k_n}$. Playing with the inequalities, conclude that $(\forall n < \omega)(|y_n - y_{n+1}| < 2^n)$; by the list of facts at the beginning of this proof, this will imply that the sequence $\vec{y}$ is a Cauchy sequence. Furthermore, some more straightforward play with the inequalities will make it apparent that the sequence $\langle x^n \big| n < \omega\rangle$ of real numbers converges to the real number whose equivalence class is the Cauchy sequence $\vec{y}$. □

3.5 Other mathematical objects

We will now discuss, very briefly, how to carry out the implementation of other mathematical objects within set theory. We saw how to recursively define n-fold cartesian products, however, a more elegant choice, now that we have defined the natural numbers, is to implement the cartesian product of an indexed family $\langle A_i \big| i \in n\rangle$ to be given by the set $\{f : n \longrightarrow \bigcup_{i\in n} A_i \big| (\forall i \in n)(f(i) \in A_i)\}$, where the intention is that the function f with $\mathrm{dom}(f) = n$ represents the n-tuple $\langle f(0), f(1), \dots, f(n-1)\rangle$. In particular, the n-dimensional cartesian space is implemented as $\mathbb{R}^n$, the set of all functions from $n = \{0, \dots, n-1\}$ to $\mathbb{R}$.

Similarly, suppose we are given a ring R. Then the polynomial ring $R[X]$ is implemented in a very natural way as the set $\{f : \omega \longrightarrow R \big| (\exists n \in \omega)(\forall m > n)(f(m) = 0)\}$. Here the idea is that the function $f : \omega \longrightarrow R$, that satisfies $f(m) = 0$ for $m \geq n$, represents the polynomial given by $f(0) + f(1)X + \cdots + f(n)X^n$. In the same spirit, the ring $R[[X]]$ of power series with coefficients in R is implemented by the set R^ω, whereas the ring of Laurent series $R((X))$ is naturally implemented as $\{f : \mathbb{Z} \longrightarrow R \big| (\exists n \in \mathbb{Z})(\forall m < n)(f(n) = 0)\}$, with the obvious interpretations.

Most other mathematical objects tend to be defined very explicitly as sets, possibly modulo taking for granted some other elementary objects such as $\mathbb{R}$. Thus, presumably by now the reader should know how to embed all of the mathematics that she currently knows within set theory. And, if she so wishes, it should be possible to keep track of how all the mathematics that she subsequently learns can similarly be implemented within set theory. So from now on, the statement "everything is a set" should make perfect sense. This finishes the exposition of the "foundational" part of set theory.

Chapter 4

Ordinal numbers

It is now time to discuss ordinal numbers, which constitute the first of Cantor's attempts at systematically comparing different kinds of infinity. Our approach is to define ordinal numbers in the style of von Neumann. Along the way, we will need to also discuss classes, and to provide a detailed account of the Axiom of Replacement.

4.1 Definition and basic facts

The idea behind ordinal numbers is that we would like to consider the collection of equivalence classes of well-ordered sets, under the equivalence relation of isomorphism. Unfortunately, such equivalence classes do not exist within ZFC, and so the collection of all of these proper classes is not something that we can meaningfully consider in our set theory. Fortunately, von Neumann designed a very elegant way of obtaining a "canonical" representative for each of these equivalence classes. The first step is defining transitive sets.

Definition 90. A set x will be said to be **transitive** if $(\forall y)(y \in x \Rightarrow y \subseteq x)$.

Proposition 91. *The following are equivalent for any set x,*

- *x is transitive,*
- *$(\forall z, y)(z \in y \in x \Rightarrow z \in x)$ (thus we can think of transitive sets as "points of transitivity of the relation $\in$"),*
- *$\bigcup x \subseteq x$,*
- *$x \subseteq \mathfrak{P}(x)$.*

Proof. Obvious. □

Example 92. $\varnothing$, each $n \in \mathbb{N}$, and ω are examples of transitive sets.

Lemma 93. *Suppose that X is a set such that every $x \in X$ is transitive. Then $\bigcup X$ and $\bigcap X$ are also transitive sets.*

Proof. If $z \in y \in \bigcup X$ (respectively $z \in y \in \bigcap X$) then for some (respectively for all) $x \in X$, $z \in y \in x$; since by assumption each $x \in X$ is transitive, we conclude that $z \in x \in X$ and thus $z \in \bigcup X$ (respectively $z \in \bigcap X$). Hence $\bigcup X$ (respectively $\bigcap X$) is transitive. □

Definition 94. An **ordinal number** is a transitive set which is strictly well-ordered by the relation $\in$ (that is, α is an ordinal number if and only if α is transitive and the relation $\{(\beta, \gamma) \in \alpha \times \alpha \big| \beta \in \gamma\}$ is a well-order on α).

Example 95. $\varnothing$, each $n \in \mathbb{N}$, and ω are all examples of ordinal numbers.

Definition 96. If α, β are two ordinal numbers, we will say that $\alpha < \beta$ if and only if $\alpha \in \beta$ (that is, between ordinal numbers we will stipulate the symbol $<$ to be synonymous with the symbol $\in$). We also define the expression $\alpha \leq \beta$ to have the obvious meaning (this obvious meaning, in case it is not so obvious, is just $\alpha < \beta \vee \alpha = \beta$).

As the next theorem shows, the relation $<$ (respectively $\leq$) thus defined among ordinals behaves in the way that we would expect a strict partial order (respectively partial order) to behave.

Theorem 97. *Let α, β, γ be ordinals.*

1. *$\alpha \nless \alpha$ (because if $\alpha < \alpha$ then we would have $\alpha \in \alpha \in \alpha$, contradicting that $\in$ is irreflexive in α),*

2. *$\alpha < \beta < \gamma \Rightarrow \alpha < \gamma$ (this is just the fact that γ is a transitive set).*

□

We will now prove a few important properties of ordinal numbers. The first one tells us that the collection of all ordinal numbers satisfies the definition of a transitive set, if only this collection was a set (which, as we will see later, it is not, but it will still be a "transitive collection"); the second one provides a nice characterization of the non-strict well-order relation $\leq$, and the third one will imply that the collection of ordinals satisfies the definition of a well-ordered set, if only it was a set (it will only be a "well-ordered collection"). The fourth one is a just nice extra[1].

Theorem 98. *Let α, β be ordinals, and let $X \neq \varnothing$ be a set all of whose elements are ordinals.*

1. *$(\forall z \in \alpha)(z$ is an ordinal$)$,*
2. *$\alpha \subseteq \beta \iff \alpha \leq \beta$.*
3. *$\bigcap X$ is an ordinal, moreover $\min(X) = \bigcap X$,*
4. *$\bigcup X$ is an ordinal, moreover $\sup(X) = \bigcup X$.*

Proof.

1. Let $z \in \alpha$. Since $z \subseteq \alpha$, α is well-ordered by $\in$, and the property of being well-ordered is hereditary, we conclude that z is well-ordered by $\in$ as well. Now to prove that z is transitive, let $x \in y \in z$. Since $\in$ is transitive for elements of α (as α is well-ordered, in particular partially ordered, by $\in$), we have that $x \in z$ and thus z is a transitive set.

2. $\Leftarrow$ Easy,

   $\Rightarrow$ Suppose that $\alpha \subseteq \beta$ and $\alpha \neq \beta$. Then $\beta \setminus \alpha \neq \varnothing$ so we can let $\xi = \min(\beta \setminus \alpha)$. We claim that $\xi = \alpha$ (and once the claim is established, of course α will be an ordinal).

   To see that $\xi \subseteq \alpha$: if $\delta \in \xi$ then $\delta \notin \beta \setminus \alpha$ (since ξ is the minimum of $\beta \setminus \alpha$) which implies that $\delta \in \alpha$. Now, to see that $\alpha \subseteq \xi$, let $\mu \in \alpha$ be arbitrary. Since $\in$ linearly orders β, we must have that either $\mu \in \xi$ or $\xi \in \mu$ or $\xi = \mu$. Notice that $\xi \in \mu$ implies $\xi \in \alpha$(since $\mu \in \alpha$ and α is transitive), contradicting that $\xi \notin \alpha$. Similarly $\xi = \mu \in \alpha$ carries the same contradiction. Therefore it must be the case that $\mu \in \xi$, and we are done.

3. By Lemma 93, $\bigcap X$ is transitive. To show that it is well-ordered by $\in$, take any $\alpha \in X$ and note that $\bigcap X \subseteq \alpha$, thereby ensuring that $\bigcap X$ is well-ordered by $\in$, since the property of being well-ordered is hereditary. By part 2 above, since among ordinals $\leq$ is the same as $\subseteq$ then it is clear that $\bigcap X = \inf(X)$. So now we just need to show that the ordinal $\bigcap X$ belongs to X, which will ensure that it is the minimum of the set X (since we already know that it is its infimum). So suppose otherwise, then $\bigcap X < x$ for all $x \in X$ (since $\bigcap X \leq x$ and it cannot equal x since it does not belong to X). Remembering the meaning of the symbol $<$, we see that $(\forall \alpha \in X)(\bigcap X \in \alpha)$, which implies that $\bigcap X \in \bigcap X$. This contradicts the fact that $\in$ is an irreflexive relation (i.e. Theorem 97 part 1) among ordinals.

4. By Lemma 93, we know that $\bigcup X$ is transitive. In order to show that it is well-ordered by $\in$, just remember that $\bigcup X$ is a set consisting of ordinals (by point 1 above), and start by using Theorem 97 to infer that $\in$ partially orders this (in fact, every) set of ordinals. Now, to show that this partial order is a well-order, let $\varnothing \neq Y \subseteq \bigcup X$. Then Y is a set of ordinals and so it has a minimum by point 3 above. This shows that $\bigcup X$ is well-ordered by $\in$ and hence it is an ordinal. By point 2 above, we already know that among ordinals $\leq$ is the same as $\subseteq$, which clearly implies that $\bigcup X$ is the smallest upper bound for X.

□

So the collection of all ordinals is, in fact, a "well-ordered (and hence linearly ordered, in particular) collection". This fact, in and of itself, will lead to the conclusion that this collection cannot actually be a set. This, which in the olden days was considered to be a paradox, was discovered by Cantor even before he found out about Russell's paradox. It was Russell, however, the one who made a big fuss about it (just like he did with Russell's paradox), and somehow the paradox ended up being named after Cesare Burali-Forti, due to the fortuitious fact that one of Burali-Forti's theorems contradicted another result of Cantor, which is what lead Russell to think about this paradox.

Corollary 99 (Burali-Forti paradox)**.** *There is no set containing all ordinals.*

[1] As we would say in Spanish, "esa nomás es de pilón".

Proof. If there was a set X such that $(\forall\alpha)(\alpha \text{ is an ordinal} \Rightarrow \alpha \in X)$, then by Comprehension we would have that the set

$$O = \{\alpha \in X \big| \alpha \text{ is an ordinal}\}$$

exists and contains precisely all ordinal numbers. Note that, under these assumptions, Theorem 98 part 1 can be interpreted as saying that O is a transitive set. On the other hand, Theorem 97 along with part 3 of Theorem 98 yields that O is well-ordered by $\in$. Hence O itself is an ordinal number, which implies that $O \in O$, but this is impossible because it contradicts irreflexivity of $\in$ among ordinals (i.e. Theorem 97 part 1). □

We will see now that every ordinal has an immediate successor. The rule for generating the successor of an ordinal is exactly the same that we have already used to generate the successor of a natural number.

Definition 100. If α is an ordinal, then we define its **successor**, denoted by $S(\alpha)$, to be $\alpha \cup \{\alpha\}$.

Remark 101. It is easy to see that $S(\alpha)$ is also an ordinal. To see that $S(\alpha)$ is transitive, notice that $\xi \in S(\alpha) = \alpha \cup \{\alpha\}$ means that either $\xi \in \alpha$, in which case $\xi \subseteq \alpha \subseteq \alpha \cup \{\alpha\}$, or $\xi \in \{\alpha\}$, in which case $\xi = \alpha \subseteq \alpha \cup \{\alpha\}$. Now, since all elements of $S(\alpha)$ are ordinals, and sets of ordinals have minimums, it immediately follows that $S(\alpha)$ is well-ordered by $\in$.

Note also that calling the ordinal $S(\alpha)$ the "successor" of α is a very much appropriate choice: for $\beta < S(\alpha)$ if and only if either $\beta \in \alpha$ (i.e. $\beta < \alpha$) or $\beta \in \{\alpha\}$ (i.e. $\beta = \alpha$). So $S(\alpha)$ is an ordinal strictly larger than α, with nothing else in between α and $S(\alpha)$. This analysis yields yet another proof of the Burali–Forti paradox.

Second proof of the Burali–Forti paradox. Using comprehension like the previous proof, the existence of a set containing all ordinals would imply the existence of the set O whose elements are precisely all ordinals (and nothing else). Let $\xi = \sup(O)$, and consider the ordinal $S(\xi)$. On the one hand, $(\forall\alpha \in O)(\alpha \le \xi < S(\xi))$. However, $S(\xi) \in \mathbf{Ord}$, thus $S(\xi) < S(\xi)$, a contradiction. □

Definition 102. Let α be an ordinal.

1. α is said to be a **successor ordinal** if $(\exists\beta)(\alpha = S(\beta))$,
2. α is said to be a **limit ordinal** if it is not a successor ordinal.

Definition 102 provides us with a conceptual way of understanding ordinal numbers as divided into two fundamentally different classes. The main difference when working with ordinal numbers (as opposed to just working with natural numbers) is the presence of limit ordinals. Note that ω is the first (the smallest) limit ordinal. This motivates the following definition.

Definition 103. We will say that an ordinal α is a **finite ordinal** if $\alpha < \omega$.

Remark 104. There are several equivalent definitions for a finite ordinal. One possibility is to define a finite ordinal to be an ordinal on which the relation $\in^{-1}$ is a well-order (or in other words, not only do nonempty subsets have a minimum, but also a maximum). You guys will prove the equivalence of these two definitions in your next assignment. Another equivalent definition is that a finite ordinal is an ordinal α satisfying $(\forall\beta \le \alpha)(\beta = 0 \vee \beta \text{ is a successor})$. Let us proceed to prove that this last definition is, in fact, equivalent to Definition 103.

First note that, if α is an ordinal satisfying $(\forall\beta \le \alpha)(\beta = 0 \vee \beta \text{ is a successor})$, then we must have $\alpha < \omega$ (since if $\alpha \ge \omega$ then ω, a limit ordinal, witnesses the failure of the assumed statement). So now we just need to prove that every $n < \omega$ satisfies $(\forall k \le n)(k = 0 \vee k \text{ is a successor})$. To see this, define

$$X = \{n \in \omega \big| (\forall k \le n)(k = 0 \vee k \text{ is a successor})\}.$$

Clearly $0 \in X$; and it is also straightforward to verify that if $n \in X$ then $S(n) \in X$ as well (since every $k \le S(n)$ satisfies that either $k = S(n)$ (so k is a successor) or $k \le n$ (so k is either 0 or a successor, by inductive hypothesis)). Thus X is an inductive set, which shows that $X = \omega$.

Now that we have discussed finite ordinals, as well as successors, we will proceed to say a word about limit ordinals.

Remark 105. If α is a limit ordinal, then for every $\beta < \alpha$ we have that $S(\beta) < \alpha$ as well. In other words, the well-ordered set α does not have a maximum. This implies that $\alpha = \sup\{\xi \big| \xi < \alpha\} = \bigcup \alpha$. So we now know that an ordinal α is limit if and only if $\alpha = \bigcup \alpha$.

The structure of the first few ordinals looks as follows:

$$0, 1, 2, 3, \dots, n, \dots, \omega, S(\omega), S(S(\omega)), \dots, \underbrace{S(\cdots(S}_{n \text{ times}}(\omega)\cdots), \dots$$

Suggestively, we denote $S(\omega)$ by $\omega+1$, $S(S(\omega))$ by $\omega+2$, and so on. What comes after all of the $\omega+n$ are over? That would be the ordinal number that, set-theoretically, equals $\{0, 1, \dots, \omega, \omega+1, \omega+2, \dots\}$, and we would probably like to suggestively denote this ordinal with the symbol $\omega+\omega$. However, in order to be able to prove the existence of such a set, we will need to use the Axiom of Replacement. This axiom will also be required to prove that every well-ordered set is isomorphic to some ordinal number. In order to properly understand the Axiom of Replacement, one needs a thorough understanding of classes, task that we will proceed to undertake in the next section.

4.2 Classes

Some of you might have heard that there are objects called "classes", in addition to sets. At the same time, you might have noticed that it is sometimes useful to refer to certain collections of sets (such as the collection of all sets, or the collection of all ordinals) that are not sets themselves. These kinds of collections will be what we call classes. Although there is no set in ZFC that corresponds to a collection like this, we find that we can still talk, in the metatheory, about the collection, because we have written down a definition of what it means to belong to the collection. Therefore, intuitively speaking, a class will be just a "well-defined collection of sets", whether or not such collection has a counterpart within the theory ZFC.

More formally, let us move over to the metatheory so that we can talk about (rather than *in*) the language $\mathscr{L}_{\text{ST}}$. Every $\mathscr{L}_{\text{ST}}$-formula with one parameter, $\varphi(x)$, determines the collection $\mathbf{C}_\varphi$ of sets that satisfy the formula $\varphi(x)$ (we can think of each such formula as providing us with a new definition, and we can consider the collection of objects satisfying the definition). We can write something suggestive, such as

$$\mathbf{C}_\varphi = \{x \big| \varphi(x)\},$$

to declare a symbol for the class $\mathbf{C}_\varphi$ of all sets satisfying $\varphi(x)$. The fact that we write this does not by any means imply that we are claiming the existence of $\mathbf{C}_\varphi$ "as a set", but rather it should be understood that this is just an indication that a new abbreviation has been introduced (you can think about this as "introducing a new macro"). Implicitly, every time we write such a declaration, we are also introducing a number of related abbreviations, such as the following:

1. The expression $x \in \mathbf{C}_\varphi$ abbreviates $\varphi(x)$,
2. the expression $\mathbf{C}_\varphi = \mathbf{C}_\psi$ abbreviates $(\forall x)(\varphi(x) \iff \psi(x))$,
3. the expression $\mathbf{C}_\varphi \subseteq \mathbf{C}_\psi$ abbreviates $(\forall x)(\varphi(x) \Rightarrow \psi(x))$.

On the other hand, expressions such as $\mathbf{C}_\varphi \in X$ (or any expression where $\mathbf{C}_\varphi$ occurs immediately to the left of the symbol $\in$) will not always be accepted. Sometimes it is possible to prove that "the class $\mathbf{C}_\varphi$ is a set"; in other words, sometimes we can prove the statement $(\exists y)(\forall x)(x \in y \iff \varphi(x))$. In these cases, it is considered valid to use the symbol $\mathbf{C}_\varphi$ the same way that we use any other symbol introduced to denote sets whose existence has been proved from the axioms. On the other hand, there are cases where it is not possible to prove that the class $\mathbf{C}_\varphi$ is a set, and sometimes it is in fact possible to prove that this class is not a set (those two possibilities are not the same, as a careful re-reading of them reveals). In other words, it is sometimes the case that the formula $(\exists x)(\forall y)(y \in x \iff \varphi(x))$ is either not provable, or downright refutable, from the ZFC axioms. In the latter case, we will say that $\mathbf{C}_\varphi$ is a *proper class* (so, a proper class is a class that is not a set), and under these circumstances the symbol $\mathbf{C}_\varphi$ should never occur immediately to the left of a membership symbol $\in$. Any such occurence will be completely forbidden, and treated as nonsense with the highest severity.

Thus, as long as we are careful before writing the symbol $\mathbf{C}_\varphi$ to the left of the symbol $\in$, we are always able to write a number of statements "about the class $\mathbf{C}_\varphi$", which intuitively make sense if we think of such class as a collection. This will be fine as long as we always remember that these statements are not, properly speaking, formulas in $\mathscr{L}_{\text{ST}}$, but rather they are abbreviations of actual $\mathscr{L}_{\text{ST}}$-formulas. As an example of this, we explain below how to interpret operations such as the union or intersection of two classes:

1. $\mathbf{C}_\varphi \cap \mathbf{C}_\psi = \mathbf{C}_{\varphi \wedge \psi}$,
2. $\mathbf{C}_\varphi \cup \mathbf{C}_\psi = \mathbf{C}_{\varphi \vee \psi}$,
3. $\mathbf{C}_\varphi \setminus \mathbf{C}_\psi = \mathbf{C}_{\varphi \wedge \neg\psi}$,

4. $\mathbf{C}_\varphi \times \mathbf{C}_\psi = \{x \big| (\exists a)(\exists b)(x = (a,b) \wedge \varphi(a) \wedge \varphi(b))\}$,

5. $\mathbf{C}_\varphi \restriction \mathbf{C}_\psi = \mathbf{C}_\varphi \cap (\mathbf{C}_\psi \times \mathbf{V})$,

6. $\bigcup \mathbf{C}_\varphi = \{x \big| (\exists y)(x \in y \wedge \varphi(y))\}$,

7. $\bigcap \mathbf{C}_\varphi = \{x \big| (\forall y)(\varphi(y) \Rightarrow x \in y)\}$ (which is actually a set if $\mathbf{C}_\varphi \neq \varnothing$, as you guys will prove in your assignment, because of the Axiom of Comprehension[2]),

8. $\mathfrak{P}(\mathbf{C}_\varphi) = \{x \big| (\forall y)(y \in x \Rightarrow \varphi(y))\}$.

Basically, any other expression that we can make sense of will be fine. Note, however, that we should not write things like, for example, $\{\mathbf{C}_\varphi, \mathbf{C}_\psi\}$, if we know that either of $\mathbf{C}_\varphi$ or $\mathbf{C}_\psi$ might be proper classes.

We introduce the following standard notations for certain specific classes:

- $\mathbf{V} = \{x \big| x = x\}$,
- $\in = \{(x,y) \big| x \in y\}$,
- $\mathbf{Ord} = \{x \big| (\forall z)(\forall y)(z \in y \in x \Rightarrow z \in x) \wedge \in\restriction x \text{ is a well-order}\}$.

In order to show how the introduction of all the terminology regarding classes can provide us with a useful conceptual tool, we will state (and prove) a couple of principles that allow us to say that the class of all ordinals is well-ordered in a strong sense. That is, not only will every nonempty subset of ordinals have a minimum, but in fact every nonempty class will. Note that this is not really a theorem that can be stated in $\mathscr{L}_{\text{ST}}$, but rather a metatheorem: for each $\mathscr{L}_{\text{ST}}$-formula (i.e. for each class) there is a corresponding theorem, whose proof just needs to follow the proof scheme below.

Theorem 106 (Transfinite Induction Theorem Scheme, Version 1)**.** *Let $\mathbf{C}$ be a class [i.e. let $\varphi(x)$ be an $\mathscr{L}_{\text{ST}}$-formula such that $(\forall x)(\varphi(x) \Rightarrow x \in \mathbf{Ord})$]. If $\mathbf{C} \neq \varnothing$ [i.e. if $(\exists \alpha \in \mathbf{Ord})(\varphi(\alpha))$], then $\mathbf{C}$ has a minimum [i.e. $(\exists \alpha \in \mathbf{Ord})(\varphi(\alpha) \wedge (\forall \beta < \alpha)(\neg\varphi(\beta)))$].*

Proof. Since $\mathbf{C} \neq \varnothing$, we can pick $\alpha \in \mathbf{C}$ [i.e. pick α such that $\varphi(\alpha)$]. If no $\beta < \alpha$ is such that $\varphi(\beta)$, then we're done, otherwise just pick $\min\{\beta < \alpha \big| \varphi(\beta)\}$ and we are done. □

We now turn this into the possibility of performing proofs by induction over the ordinal numbers. Once again, the statement below is not a theorem, but a metatheorem or theorem scheme.

Theorem 107 (Transfinite Induction Theorem Scheme, Version 2)**.** *Let $\mathbf{C}$ be a class of ordinals [i.e. let $\varphi(x)$ be an $\mathscr{L}_{\text{ST}}$-formula such that $(\forall x)(\varphi(x) \Rightarrow x \in \mathbf{Ord})$] such that $(\forall \alpha \in \mathbf{Ord})(\alpha \subseteq \mathbf{C} \Rightarrow \alpha \in \mathbf{C})$ [i.e. suppose that $(\forall \alpha \in \mathbf{Ord})((\forall \beta < \alpha)(\varphi(\beta)) \Rightarrow \varphi(\alpha))$] holds. Then $\mathbf{C} = \mathbf{Ord}$ [i.e. $(\forall \alpha)(\alpha \in \mathbf{Ord} \Rightarrow \varphi(\alpha))$].*

Proof. Suppose that $\mathbf{C} \neq \mathbf{Ord}$, then we can let $\beta = \min(\mathbf{Ord} \setminus \mathbf{C})$ [i.e. let β be the least ordinal such that $\neg\varphi(\beta)$]. Then $\beta \subseteq \mathbf{C}$ but $\beta \notin \mathbf{C}$, a contradiction. □

As per the previous theorem, in order to prove that some property φ holds of all ordinals, we should formally prove, for each $\alpha \in \mathbf{Ord}$, the statement $(\forall \beta < \alpha)(\varphi(\beta)) \Rightarrow \varphi(\alpha)$. In practice, most of the time this is broken up into three cases:

1. Prove that $\varphi(0)$ holds,

2. prove that $(\forall \alpha)(\varphi(\alpha) \Rightarrow \varphi(S(\alpha)))$,

3. prove that, for each limit ordinal α, $(\forall \beta < \alpha)(\varphi(\beta)) \Rightarrow \varphi(\alpha)$.

The fact that we can do induction over the ordinal numbers suggests that we should also be able to construct functions by recursion over them. There is a sense in which we will in fact be able to do this. In order to understand in which sense, and be able to prove the corresponding result, we will need to wait until after the Axiom of Replacement has been explained.

[2] Or, for an even better definition, $\bigcap \mathbf{C}_\varphi = \{x \in \bigcup \mathbf{C}_\varphi \big| (\forall y)(\varphi(y) \Rightarrow x \in y)\}$, which is *always* a set.

4.3 A few words about NBG

We will now briefly explain a couple of axiomatic theories, alternative to ZFC, where classes actually exist as objects in the theory, and not just as circumlocutions in the metatheory.

The first one is the axiom system of von Neumann, Bernays and Gödel, abbreviated NBG. In this system, "class" is the basic undefined notion (i.e. we think of classes as being the objects of our theory). Formally we have a first-order language with one binary relation $\in$ and a unary predicate Set, which is intended to denote which objects, among all classes, are sets. The axioms are the same as in ZFC, except that appropriate modifications are made to reflect the fact that those axioms speak only about sets: each instance of $(\forall x)(\varphi)$ gets replaced by $(\forall x)(\text{Set}(x) \Rightarrow \varphi)$, and each instance of $(\exists x)(\varphi)$ gets replaced by $(\exists x)(\text{Set}(x) \wedge \varphi(x))$ throughout the Axioms of Pairing, Union, Powerset, Infinity, Foundation, and Choice. Extensionality still applies to classes, i.e. the formula $(\forall X)(\forall Y)(X = Y \iff (\forall z)(z \in X \iff z \in Y))$ is an axiom. The Axiom of Comprehension no longer needs to be an infinite schema, for it can now be concisely stated by means of the formula:

$$(\forall Y)(\forall x)(\text{Set}(x) \Rightarrow \text{Set}(x \cap Y)).$$

A similar modification allows us to state the Axiom of Replacement in one single formula, which we will not exhibit here, since we still have not explained what this axiom says.

Additionally, we have the following two axioms (or rather, one axiom plus one axiom scheme) which go beyond ZFC in order to impose some restrictions on how classes behave:

- $(\forall x)(\text{Set}(x) \iff (\exists y)(x \in y))$,
- for each formula φ whose quantifiers range only over sets,

  $$(\exists X)(\forall y)(y \in X \iff \varphi(y)),$$

  schema known as *class comprehension*, which reflects the fact that, for each formula, we expect the class of sets satisfying the formula to exist.

With these axioms, the system NBG is a conservative extension of ZFC, which means that for every formula φ containing only set variables (i.e. every variable occuring in φ corresponds to a set, as appropriately expressed by using the unary predicate Set), we have that $\mathsf{NBG} \vdash \varphi$ if and only if $\mathsf{ZFC} \vdash \varphi$. In spite of the axiom schema, NBG is known to be finitely axiomatizable.

Another axiom system that has been considered, although by now it is essentially not used by anyone (we only mention it here for the historical reason that it exists), is the system of Morse and Kelley, abbreviated MK. The axioms of MK are the same as the axioms of NBG, except that when it comes to Class Comprehension, we drop from the schema the restriction on φ having only set variables. This makes MK into a more powerful axiom system, where much more statements can be proved (compared to either NBG or ZFC). This system is not finitely axiomatizable.

4.4 The Axiom of Replacement

Let us recall the statement of the Axiom Schema of Replacement, as it appears on our axiom list: for each formula φ with (at least) two free variables x, y, the following $\mathscr{L}_{\text{ST}}$-formula is an axiom:

$$(\forall z)((\forall x)(x \in z \Rightarrow (\exists! y)(\varphi(x,y))) \Rightarrow (\exists w)(\forall x)(x \in z \Rightarrow (\exists y)(y \in w \wedge \varphi(x,y)))).$$

Let us have a look at the antecedent of the conditional statement: $(\forall x)(x \in z \Rightarrow (\exists! y)(\varphi(x,y))$. This means that the class[3]

$$\mathbf{C}_\varphi = \{(x,y) \big| \varphi(x,y)\}$$

behaves like a function, at least when restricted to z. From now on we will talk about *class functions* (remembering that this is an utterance that only makes sense if and when we stand in the metatheory). In this case, we can denote such class by a symbol such as $\mathbf{F}_\varphi$, an write the symbols $\mathbf{F}_\varphi : z \longrightarrow \mathbf{V}$ to conveniently abbreviate the fact that this class is a class function. Then, what the Axiom of Replacement does is to guarantee the existence of the set $w = \mathbf{F}_\varphi[z] = \{\mathbf{F}(x) \big| x \in z\}$.

Thus, we can more concisely express the Axiom Schema of Replacement, in the metatheory (this will not be a single $\mathscr{L}_{\text{ST}}$-formula, but a scheme for infinitely many formulas, one for each class function), by means of the following statement:

For every set z and every class function $\mathbf{F} : z \longrightarrow \mathbf{V}$, the set $\mathbf{F}[z]$ exists.

[3] In general, if an $\mathscr{L}_{\text{ST}}$-formula has two free variables, we think of it as defining a class of ordered pairs rather than just a class of sets. For example, $\mathbf{C}_{x \in y} = \{(x,y) \big| x \in y\}$, $\mathbf{C}_{x \subseteq y} = \{(x,y) \big| x \subseteq y\}$, and $\mathbf{C}_{y = \mathfrak{P}(x)} = \{(x,y) \big| y = \mathfrak{P}(x)\}$. As another example, we use id to denote the class $\mathbf{C}_{x=y}$ containing precisely all ordered pairs of the form (x,x). Hence from now on, it is possible to use the symbol $\text{id} \upharpoonright X$ for the object that we formerly used to call id_X, for every set X.

Intuitively, given any class function $\mathbf{F} : z \longrightarrow \mathbf{V}$, we can “replace”, element by element, each member of the set $z = \{x \big| x \in z\}$ by its $\mathbf{F}$-images, to obtain the new set $\mathbf{F}[z] = \{\mathbf{F}(x) \big| x \in z\}$. This follows the intuition that sets are “relatively small”, whereas proper classes are collections that are “too large”. So this axiom states that, if we know that z is small enough to be a set, then so is $\mathbf{F}[z] = \{\mathbf{F}(x) \big| x \in z\}$, which has at most as many elements as z does (but possibly less, if $\mathbf{F}$ is not injective).

For example, the $\mathscr{L}_{\mathrm{ST}}$-formula $\varphi(x, y)$ given by $(\exists a)(\exists b)(x = (a, b) \wedge y = a)$ represents the class function whose domain is the class of all ordered pairs, and which maps each ordered pair to its first coordinate. So if R is a relation, the Axiom of Replacement ensures the existence of $\mathrm{dom}(R) = \mathbf{F}_\varphi[R]$. Similarly, the Axiom of Replacement ensures the existence of $\{\mathfrak{P}(x) \big| x \in X\}$ for each set X. These uses of Replacement might not look very impressive, since the sets whose existence we just ensured can also be derived from the Axiom of Comprehension, but we are slowly getting close to the point where new, powerful consequences of the Axiom of Replacement will be derived.

As a first observation, we notice that having the axiom of Replacement makes a couple of the other axioms redundant. In what follows, we use the letters ZF_0 to abbreviate the Axioms of Extensionality, Union, Powerset, and Infinity, along with the stronger version of the Axiom of Existence which ensures the existence of $\varnothing$ (as opposed to just the existence of some set).

Theorem 108. ZF_0 + *Replacement* $\vdash$ *Pairing.*

Proof. In ZF_0, $2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\} = \mathfrak{P}(\mathfrak{P}(\varnothing))$ exists by the Axiom of Powerset. Now, given any two sets a, b, consider the $\mathscr{L}_{\mathrm{ST}}$-formula $\varphi(x, y)$ given by $(x = 0 \wedge y = a) \vee (x = 1 \wedge y = b)$. This formula defines the function $\mathbf{F}_\varphi : 2 \longrightarrow \mathbf{V}$ given by $\mathbf{F}_\varphi(0) = a$ and $\mathbf{F}_\varphi(1) = b$. By replacement, $\mathbf{F}_\varphi[2] = \{a, b\}$ exists. □

Theorem 109. ZF_0 + *Replacement* $\vdash$ *Comprehension.*

Proof. Given an $\mathscr{L}_{\mathrm{ST}}$-formula φ, consider the $\mathscr{L}_{\mathrm{ST}}$-formula $\psi(x, y)$ given by

$$(\varphi(x) \wedge y = \{x\}) \vee (\neg\varphi(x) \wedge y = \varnothing).$$

That is, this formula describes a class function $\mathbf{F}_\psi : \mathbf{V} \longrightarrow \mathbf{V}$ given by

$$\mathbf{F}_\psi(x) = \begin{cases} \{x\} \text{ if } \varphi(x); \\ \varnothing \text{ if } \neg\varphi(x). \end{cases}$$

Then given any set A, replacement ensures the existence of $\mathbf{F}_\psi[A] = \{\{x\} \big| x \in A \wedge \varphi(x)\} \cup \{\varnothing \big| x \in A \wedge \neg\varphi(x)\}$. Hence by union, the set

$$\bigcup \mathbf{F}_\psi[A] = \left(\bigcup_{\substack{x \in A \\ \varphi(x)}} \{x\} \right) \cup \left(\bigcup_{\substack{x \in A \\ \neg\varphi(x)}} \varnothing \right) = \{x \big| x \in A \wedge \varphi(x)\}$$

exists[4]. □

We have mentioned before that, since **Ord** is a well-ordered class, it is to be expected that we can also perform recursive constructions along this class. This result, which we state below, is not really a theorem (in the sense that it is not an $\mathscr{L}_{\mathrm{ST}}$-formula provable from the ZFC axioms), but rather a metatheorem, or theorem scheme; and this will be our first important application of the Axiom of Replacement.

Theorem 110 (Transfinite Recursion Theorem Scheme)**.** *Suppose that $\varphi(x, y)$ is an $\mathscr{L}_{\mathrm{ST}}$-formula such that $(\forall x)(\exists! y)(\varphi(x, y))$ [intuitively, we have a class function $\mathbf{G} : \mathbf{V} \longrightarrow \mathbf{V}$]. Then we can write down an $\mathscr{L}_{\mathrm{ST}}$-formula ψ such that:*

1. *$(\forall x)(\exists! y)(\psi(x, y))$ [that is, ψ defines a class function, which we will denote by $\mathbf{F}$], and*

2. *for all ordinals α, $\psi(\alpha, y)$ if and only if $\varphi(\{(\xi, z) \big| \xi < \alpha \wedge \psi(\xi, z)\}, y)$ [that is, $\mathbf{F}(\alpha) = \mathbf{G}(\mathbf{F} \restriction \alpha)$, for all ordinal numbers α].*

[4] There is an alternative proof for this that does not use the Axiom of Union, at the cost of considering two different cases. Given an $\mathscr{L}_{\mathrm{ST}}$-formula φ and a set A, there are two cases to consider. The first case is when $(\forall x \in A)(\neg\varphi(x))$, in this case we have that $\{x \in A \big| \varphi(x)\} = \varnothing$, which is assumed to exist as an axiom. In the remaining case, pick some fixed $z \in A$ satisfying $\varphi(z)$, and define a class function $\mathbf{F} : A \longrightarrow \mathbf{V}$ by

$$\mathbf{F}(x) = \begin{cases} x; & \text{if } \varphi(x), \\ z; & \text{otherwise.} \end{cases}$$

Then by replacement, the following class is a set:

$$\mathbf{F}[A] = \{\mathbf{F}(x) \big| x \in A\} = \{x \big| x \in A \wedge \varphi(x)\} \cup \{z \big| z \in A \wedge \neg\varphi(x)\} = \{x \in A \big| \varphi(x)\} \cup \{z\} = \{x \in A \big| \varphi(x)\}.$$

Proof. Let $\pi(x, y)$ be the following $\mathscr{L}_{\mathrm{ST}}$-formula:

$$x \in \mathbf{Ord} \land y \text{ is a function} \land \mathrm{dom}(y) = x \land (\forall z \in x)(\varphi(y \upharpoonright z, y(z)))$$

[in other words, the formula $\pi(\alpha, h)$ states that $h : \alpha \longrightarrow \mathbf{V}$ approximates the desired $\mathbf{F}$ up to α, in the sense that $(\forall\beta < \alpha)(h(\beta) = \mathbf{G}(h \upharpoonright \beta))$].

Note that, for each α, if there exists an h such that $\pi(\alpha, h)$ then such h is unique (this can be proved very straightforwardly by transfinite induction). Also, whenever $\pi(\alpha, h)$ and $\beta < \alpha$ we have that $\pi(\beta, h \upharpoonright \beta)$ holds (again by uniqueness), so the "family of all h" is coherent. Hence we can let $\psi(x, y)$ be the $\mathscr{L}_{\mathrm{ST}}$ formula

$$(x \notin \mathbf{Ord} \land y = 0) \lor (x \in \mathbf{Ord} \land (\exists\gamma \in \mathbf{Ord})(\exists h)(x < \gamma \land \pi(\gamma, h) \land h(x) = y)).$$

Once again, for each $\alpha \in \mathbf{Ord}$, if $\psi(\alpha, y)$ for some y then this y has to be unique. Thus we only need to show that for every $\alpha \in \mathbf{Ord}$, there exists at least one y such that $\psi(\alpha, y)$. In fact, it suffices to show that for all ordinals α, there is one h such that $\pi(\alpha, h)$ (if we manage to prove this, then we see that, for all ordinals α, if $\pi(S(\alpha), h)$ then $\alpha \in \mathrm{dom}(h)$ and so $\psi(\alpha, h)$). This is done by transfinite induction:

1. If $\alpha = 0$ then it suffices to let $h = \varnothing$, the empty function.

2. Suppose that $\pi(\alpha, h)$, and let $h' = h \cup \{(\alpha, \mathbf{G}(h))\} : S(\alpha) \longrightarrow \mathbf{V}$. Then $\pi(S(\alpha), h)$ holds.

3. If for every $\xi < \alpha$ there is a h_ξ (which must necessarily be unique, so in fact we have a class function : $\alpha \longrightarrow \mathbf{V}$, $\xi \longmapsto h_\xi$), then we can just let $h = \bigcup_{\xi<\alpha} h_\xi$ (note that the Axiom of Replacement is exactly what guarantees that the family that we are here taking the union of is a set), and note that $\pi(\alpha, h)$ holds.

□

We proceed to provide an example of a construction that utilizes the Transfinite Recursion Theorem Scheme. Immediately after, we provide an example of a proof that deals with this construction, in order to also provide an example of the usage of the Transfinite Induction Theorem Scheme.

Definition 111. The following defines a class function : $\mathbf{Ord} \longrightarrow \mathbf{V}$, which is known as *Zermelo's cumulative hierarchy.* Using the Transfinite Recursion Theorem Scheme, we define

1. $V_0 = \varnothing$,

2. $V_{S(\alpha)} = \mathfrak{P}(V_\alpha)$,

3. $V_\alpha = \bigcup_{\xi<\alpha} V_\xi$ if $\alpha = \bigcup \alpha$.

Definition 111 uniquely determines a set V_α for all ordinal numbers α. It is possible to talk about the class $V_\infty = \bigcup_{\alpha\in\mathbf{Ord}} V_\alpha$, by stipulating that $x \in V_\infty$ is just a shorthand for $(\exists\alpha)(\alpha \text{ is an ordinal} \land x \in V_\alpha)$. Like we said before, we now proceed to prove an interesting fact about the V_α. The proof is done by transfinite induction.

Proposition 112. *For every ordinal number α, the set V_α is a transitive set.*

Proof. The proof is by transfinite induction on $\alpha \in \mathbf{Ord}$, and so it includes three parts (base case, succesor step, and limit step).

1. $V_0 = \varnothing$ is vacuously transitive.

2. Assuming that V_α is transitive, then $V_{S(\alpha)} = \mathfrak{P}(V_\alpha)$ is transitive too (you guys proved in a previous assignment that x is transitive iff so is $\mathfrak{P}(x)$).

3. Assuming that $\alpha = \bigcup \alpha$ and V_ξ is transitive for each $\xi < \alpha$, then so is $V_\alpha = \bigcup_{\xi<\alpha} V_\xi$ (you guys proved in a previous assignment that the union of a family of transitive sets is also transitive).

By the Transfinite Induction Theorem Scheme, it follows that $(\forall\alpha \in \mathbf{Ord})(V_\alpha \text{ is transitive})$. □

4.5 Ordinal numbers as representatives of well-ordered sets

Now we are going to use the Axiom of Replacement to show that the ordinal numbers are indeed a complete class of representatives for well-ordered sets modulo isomorphism. This consists of two steps: we first need to show that no two different ordinals can be isomorphic, and after that we need to show that every well-ordered set is isomorphic to some ordinal. The following is the first step.

Theorem 113. *Suppose that $\alpha, \beta \in \mathbf{Ord}$, and $f : \alpha \longrightarrow \beta$ is an (order) isomorphism. Then $\alpha = \beta$ and $f = \mathrm{id} \upharpoonright \alpha$.*

Proof. Let us prove, by induction on $\xi \in \alpha$, that $f(\xi) = \xi$. Succeeding in this task will immediately imply that $f = \mathrm{id} \upharpoonright \alpha$, which, together with the assumption that f is onto β, allows us to conclude that $\alpha = \beta$. So let $\xi \in \alpha$, and suppose that $f(\mu) = \mu$ for all $\mu < \xi$. Then we have that

$$\begin{aligned} f(\xi) &= \{\eta \in \beta \big| \eta < f(\xi)\} \ \text{(because } < \text{ is synonimous of } \in \text{ among ordinals, and } f(\xi) \in \beta) \\ &= \{f(\mu) \big| \mu \in \alpha \wedge f(\mu) < f(\xi)\} \ \text{(because } f \text{ is onto } \beta) \\ &= \{f(\mu) \big| \mu < \xi\} \ \text{(because } f \text{ is an order isomorphism)} \\ &= \{\mu \big| \mu < \xi\} \ \text{(because by inductive hypothesis, } f(\mu) = \mu \text{ for all } \mu < \xi) \\ &= \xi. \end{aligned}$$

□

The following theorem is an extremely important application of the Axiom of Replacement, and finalizes the proof that ordinals numbers constitute a complete class of representatives of well-ordered sets modulo the isomorphism relation.

Theorem 114. *Let $(X, \leq)$ be a well-ordered set. Then there exists a unique $\alpha \in \mathbf{Ord}$ such that $(X, \leq) \cong (\alpha, \in)$.*

Proof. Start by choosing some element $u \notin X$ (u will be used as a "garbage value", or a value that signals that the function is "undefined"). Now, using the Transfinite Recursion Theorem Scheme, define a class function $\mathbf{F} : \mathbf{Ord} \longrightarrow X \cup \{u\}$ by

$$\mathbf{F}(\alpha) = \begin{cases} \min(X \setminus \mathrm{ran}(\mathbf{F} \upharpoonright \alpha)); & \text{if } X \setminus \mathrm{ran}(\mathbf{F} \upharpoonright \alpha) \neq \varnothing, \\ u; & \text{otherwise.} \end{cases}$$

Observe that "once we hit the garbage value, we never leave it" (or "once the function is undefined, it remains undefined"). That is, if $\mathbf{F}(\alpha) = u$ and $\beta \geq \alpha$, then $\mathbf{F}(\beta) = u$. Observe also that "before hitting the garbage value, $\mathbf{F}$ is injective" (or "as long as $\mathbf{F}$ is not undefined, it is injective"). That is, if $(\forall \xi < \alpha)(\mathbf{F}(\xi) \neq u)$ then $\mathbf{F} \upharpoonright \alpha$ is an injective (class) function; in fact it is order-preserving.

We now claim that "eventually we hit the garbage value", or "eventually $\mathbf{F}$ will be undefined". That is, there is an $\alpha \in \mathbf{Ord}$ such that $\mathbf{F}(\alpha) = u$. Otherwise, we would have that the class function $\mathbf{F} : \mathbf{Ord} \longrightarrow X$ is injective. Then an instance of the Axiom of Comprehension would yield the existence of the set $Y = \mathrm{ran}(\mathbf{F}) = \{x \in X \big| (\exists \alpha \in \mathbf{Ord})(\mathbf{F}(\alpha) = x)\}$, and we would have a bijective class function $\mathbf{F} : \mathbf{Ord} \longrightarrow Y$. This means that we can invoke the inverse of this class function, call it $\mathbf{G} : Y \longrightarrow \mathbf{Ord}$ (described by the formula $\varphi(x, y)$ given by $x = \mathbf{F}(y)$); now the Axiom of Replacement would imply that $\mathbf{Ord} = \mathbf{G}[Y]$ is a set, contradicting the Burali–Forti paradox. The conclusion is that there must be an ordinal number α such that $\mathbf{F}(\alpha) = u$.

We let α be the least ordinal such that $\mathbf{F} = u$. Then $\mathbf{F}[\alpha] \subseteq X$ is a set by the Axiom of Replacement, hence the following class

$$f = \{(x, y) \in \alpha \times \mathbf{F}[\alpha] \big| y = \mathbf{F}(x)\} = \mathbf{F} \upharpoonright \alpha$$

is a set, which happens to be an injective function $f : \alpha \longrightarrow X$. Furthermore, the fact that $\mathbf{F}(\alpha) = u$ implies that $X \setminus \mathrm{ran}(\mathbf{F} \upharpoonright \alpha) = \varnothing$, in other words, $\mathrm{ran}(f) = \mathrm{ran}(\mathbf{F} \upharpoonright \alpha) = X$, so that f is also surjective. Recall that this f must also be order-preserving. The conclusion is that the function $f : \alpha \longrightarrow X$ is an order-isomorphism between the well-ordered sets $(\alpha, \in)$ and $(X, \leq)$. □

4.6 Sets and their ranks

Recall that the sequence of sets V_α was defined by transfinite recursion on α by the recursive rules $V_0 = \varnothing$, $V_{S(\alpha)} = \mathfrak{P}(V_\alpha)$ and $V_\alpha = \bigcup_{\xi < \alpha} V_\xi$ for limit α. We already proved that each V_α is a transitive set. Furthermore, since V_α is transitive, we have that $V_\alpha \subseteq \mathfrak{P}(V_\alpha) = V_{S(\alpha)}$. From this observation, it is easy to prove by induction on $\beta > \alpha$ that $V_\alpha \subseteq V_\beta$. It is for this reason that the V_α are said to be a *cumulative hierarchy.* We will now see that most of the sets that are used as everyday mathematical objects belong to some V_α. From now on, we introduce the symbol V_∞ to denote the class $\bigcup_{\alpha \in \mathbf{Ord}} V_\alpha$, in other words, $x \in V_\infty$ is taken to be an abbreviation of the $\mathscr{L}_{\mathrm{ST}}$-formula $(\exists \alpha \in \mathbf{Ord})(x \in V_\alpha)$. We will prove very soon that $\mathbf{Ord} \subseteq V_\infty$, in particular, V_∞ is a proper class.

Definition 115. We define a class function rank : $V_\infty \longrightarrow \mathbf{Ord}$, given as follows: for each $x \in V_\infty$, we let $\mathrm{rank}(x) = \min\{\alpha \in \mathbf{Ord} \big| x \in V_{S(\alpha)}\}$.

It is not hard to see that, given $x \in V_\infty$, the least α for which $x \in V_\alpha$ must be a successor ordinal, hence the definition above is coherent. Note also that $\mathrm{rank}(x)$ happens to be the unique ordinal number α such that $x \in V_{S(\alpha)} \setminus V_\alpha$.

Proposition 116. *Each $\alpha \in \mathbf{Ord}$ satisfies $\alpha \in V_\infty$, and moreover $\mathrm{rank}(\alpha) = \alpha$.*

Proof. We prove this by transfinite induction on $\alpha \in \mathbf{Ord}$. Hence, there are three steps to this proof.

1. Certainly $\varnothing \notin V_0 = \varnothing$, whilst $\varnothing \in \{\varnothing\} = V_1$, so $\mathrm{rank}(0) = 0$.
2. Suppose that $\mathrm{rank}(\alpha) = \alpha$. This means that $\alpha \in V_{S(\alpha)}$ and $\alpha \notin V_\alpha$. Now, since $V_{S(\alpha)}$ is transitive, we also have $\alpha \subseteq V_{S(\alpha)}$, thus $S(\alpha) = \alpha \cup \{\alpha\} \subseteq V_{S(\alpha)}$. Hence $S(\alpha) \in \mathfrak{P}(V_{S(\alpha)}) = V_{S(S(\alpha))}$. Moreover, it can't be the case that $S(\alpha) \in V_{S(\alpha)}$, because this would imply (again by transitivity of $V_{S(\alpha)}$) that $\alpha \cup \{\alpha\} = S(\alpha) \subseteq V_{S(\alpha)}$ and, in particular, $\alpha \in V_\alpha$, contradicting the inductive hypothesis. Thus $S(\alpha) \in V_{S(S(\alpha))} \setminus V_{S(\alpha)}$.
3. If $\alpha = \bigcup \alpha$ and $\mathrm{rank}(\xi) = \xi$ for all $\xi < \alpha$, then in the first place we have, for all $\xi < \alpha$, that (since $S(\xi) < \alpha$) $\xi \in V_{S(\xi)} \subseteq \bigcup_{\zeta<\alpha} V_\zeta = V_\alpha$. Thus $\alpha \subseteq V_\alpha$ and so $\alpha \in \mathfrak{P}(V_\alpha) = V_{S(\alpha)}$. We now argue that it cannot be the case that $\alpha \in V_\alpha$: otherwise, since $V_\alpha = \bigcup_{\xi<\alpha} V_\xi$, we would have that $\alpha \in V_\xi$ for some $\xi < \alpha$. Since V_ξ is transitive, we conclude that $\xi \in \alpha \subseteq V_\xi$. This contradicts that $\xi \notin V_\xi$ (because our inductive assumption is that $\mathrm{rank}(\xi) = \xi$). Therefore $\alpha \notin V_\alpha$, thus $\alpha \in V_{S(\alpha)} \setminus V_\alpha$, and we are done.

□

Note that, as it transpired from the previous proof, another equivalent definition of the rank function is that $\mathrm{rank}(x)$ is the least ordinal number α such that $x \subseteq V_\alpha$.

Let us compute the rank of some familiar objects:

- For each $n \in \omega$, $\mathrm{rank}(n) = n$; and $\mathrm{rank}(\omega) = \omega$.
- Suppose $\mathrm{rank}(a) = \alpha$ and $\mathrm{rank}(b) = \beta$. Then $\{a\} \subseteq V_{S(\alpha)}$ and $\{a, b\} \subseteq V_{S(\max\{\alpha,\beta\})}$. Consequently $\mathrm{rank}(\{a\}) = S(\alpha)$ and $\mathrm{rank}(\{a, b\}) = S(\max\{\alpha, \beta\})$. With a similar reasoning, since $(a, b) = \{\{a\}, \{a, b\}\}$, we can conclude that $\mathrm{rank}((a, b)) = S(S(\max\{\alpha, \beta\}))$.
- As per the above, whenever $n, m \in \mathbb{N}$, we have $\mathrm{rank}((n, m)) = \max\{n, m\} + 2$. So each element of $\mathbb{N} \times \mathbb{N}$ has some finite rank, and we can find elements of $\mathbb{N}\times\mathbb{N}$ of arbitrarily high rank. Hence $\mathbb{N}\times\mathbb{N} \subseteq V_\omega$ and so $\mathrm{rank}(\mathbb{N}\times\mathbb{N}) = \omega$. Also each integer is a subset of $\mathbb{N} \times \mathbb{N}$ with elements of arbitrarily high finite rank, thus each integer has rank ω, meaning that each integer occurs for the first time at $V_{S(\omega)}$, and consequently $\mathbb{Z} \subseteq V_{S(\omega)}$, showing that $\mathrm{rank}(\mathbb{Z}) = S(\omega)$.
- Since each element of $\mathbb{Z}$ has rank ω, then each element of $\mathbb{Z}\times\mathbb{Z}$ has rank $S(S(\omega))$ and thus $\mathrm{rank}(\mathbb{Z}\times\mathbb{Z}) = S(S(S(\omega)))$. Similarly each rational number has rank $S(S(S(\omega)))$, and thus $\mathrm{rank}(\mathbb{Q}) = S(S(S(S(\omega))))$.
- Each ordered pair (n, q) with $n \in \mathbb{N}$ and $q \in \mathbb{Q}$ has rank $S^6(\omega)$. Hence any set of such ordered pairs, e.g. any sequence, will have rank $S^7(\omega)$. Any set of such sequences, e.g. any equivalence class of Cauchy sequences, will have rank $S^8(\omega)$ and therefore $\mathrm{rank}(\mathbb{R}) = S^9(\omega)$.
- Finally, every ordered pair of real numbers will have rank $S^{11}(\omega)$ and so the set of all such pairs has rank $\mathrm{rank}(\mathbb{C}) = S^{12}(\omega)$.

Note that the exact rank of these sets depends on how we choose to implement them. That is, if we define the real numbers as Dedekind cuts rather than equivalence classes of Cauchy sequences, then the rank of the set $\mathbb{R}$ will be different; similarly for each of the constructions where we have two or more alternatives that produce isomorphic results (the reader should feel free to compute these alternative ranks). What seems to be true, however, is that even with the most complicated definitions available, these sets have rank less than $\omega + \omega$, and the same can be said of most of the constructions that are used throughout most of mathematics. It could be said that ordinary (i.e. non-set-theoretical) mathematics takes place in $V_{\omega+\omega}$, in fact, it is probably fair to say that it all takes place in $V_{\omega+n}$ for some sufficiently large n (certainly $n = 100$ should suffice).

4.7 Ordinal arithmetic

It is possible to extend the arithmetic operations (addition, multiplication, and exponentiation), as they are defined on ω, to all ordinal numbers, by means of the Transfinite Recursion Theorem Scheme. In order to appropriately learn this topic, the reader is encouraged to look at the worksheet that appears as Appendix D in this document.

Chapter 5

Cardinality

We finally are in good shape to dip our toes in the exciting topic of cardinality. In this chapter we explore as much as it is possible without utilizing the Axiom of Choice, and in the next chapter, after introducing this axiom, we will explore a little bit more about cardinality.

5.1 Equipotence, countable and uncountable sets

After spending decades working with ordinal numbers (which generalize the idea of "counting the elements of a set one by one"), Cantor slowly abstracted a new idea, namely, that it is possible to compare the number of elements of two sets not by counting each of the sets separately, but by arranging the elements of both sets into a one-to-one correspondence. This is how the idea of "cardinality of a set" was born.

Definition 117. Given two sets A, B, we say that A is **equipotent** to B, in symbols $|A| = |B|$, if there exists a bijection $f : A \longrightarrow B$.

Note that the relation "is equipotent to", even though it is a class rather than a set relation, behaves like an equivalence relation:

- $\mathrm{id} \upharpoonright A : A \longrightarrow A$ is a bijection (thus our class relation is reflexive),
- if $f : A \longrightarrow B$ is a bijection then $f^{-1} : B \longrightarrow A$ is a bijection (thus our class relation is symmetric),
- if $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are bijections, then so is $(g \circ f) : A \longrightarrow C$ (thus our class relation is transitive).

Let us quickly prove a lemma, so that we have some examples available.

Lemma 118. *Let $n < \omega$. If $f : n \longrightarrow n$ is injective, then it is surjective.*

Proof. The proof is done by induction on $n < \omega$. If $n = 0$, then this is clearly true (there is only one function $: \varnothing \longrightarrow \varnothing$, namely the empty function, which is surjective). Suppose that the lemma holds for some $n < \omega$, and let $f : n+1 \longrightarrow n+1$ be an injective function. There are two cases to consider:

1. $\mathrm{ran}(f \upharpoonright n) \subseteq n$. In this case, by inductive hypothesis the injective function $f \upharpoonright n$ must be onto n. Since f is injective, we must have that $f(n) = n$. Hence $f = (f \upharpoonright n) \cup \{(n, n)\}$ is surjective too.
2. $\mathrm{ran}(f \upharpoonright n) \nsubseteq n$. In this case, there must be some $k < n$ such that $f(k) = n$; and since f is injective, it must be the case that $f(n) < n$. So we can let $g = ((f \upharpoonright n) \setminus \{(k, n)\}) \cup \{(k, f(n))\}$. It is not hard to see that $g : n \longrightarrow n$ is injective, hence by induction hypothesis it must be onto n. This means that $f = (g \setminus \{(k, f(n))\}) \cup \{(k, n), (n, f(n))\}$ must be onto $n + 1$, and we are done.

□

Example 119.

1. If $n, m < \omega$ and $n \neq m$ then $|n| \neq |m|$. For if we assume (without loss of generality) that $m < n$, then every injective function $f : n \longrightarrow m \subseteq n$ must be onto n, by Lemma 118. This means that there cannot be a bijection $f : n \longrightarrow m$, and we are done.

2. it is an old observation of Galileo that $|\mathbb{N}| = |\{n^2 \big| n \in \mathbb{N}\}|$, as witnessed by the mapping $n \longmapsto n^2$. This is the first example of an interesting phenomenon among infinite sets: they can be in bijection with proper subsets of themselves[1].

3. $|\mathbb{N}| = |\{2n \big| n \in \mathbb{N}\}|$, witnessed by $n \longmapsto 2n$.

4. $|\mathbb{N}| = |\omega|$, this is witnessed by the mapping $n \longmapsto n - 1$.

5. For each $n < \omega$, we have that $|\omega + n| = |\omega|$. We can see this by considering the mapping $\omega + i \longmapsto i$ for $i < n$, and $m \longmapsto n + m$ for $m < \omega$.

6. **Hilbert's hotel**: in a hotel with infinitely many rooms, with all of these rooms full, a new guest arrives. Even though the hotel is full, by moving each guest from room n to room $n + 1$, suddenly we have freed up room 0, and made some space for the new guest. If it is m new guests, rather than just 1, we still have enough space to accommodate them! Just move each guest in room n to room $n + m$, and now rooms 0 through $m - 1$ are free.

7. **More Hilbert's hotel**: in fact, even if infinitely many guests were to arrive, we could still make room for them! Just move the person in room n to room $2n$, and now all of the odd-numbered rooms are free for the new guests to use.

8. $|\mathbb{Z}| = |\mathbb{N}|$; this is witnessed by the mapping $-n \longmapsto 2n$ for $n \in \mathbb{N}$ and $n \longmapsto 2n + 1$ for $n \geq 0$.

9. $|\mathbb{N} \times \mathbb{N}| = |\mathbb{N}|$, where the bijection that witnesses this fact corresponds to "traversing $\mathbb{N} \times \mathbb{N}$ in a diagonal fashion", and which is given by $(a, b) \longmapsto \frac{1}{2}(a + b - 1)(a + b - 2) + b$.

10. For every X, $|\mathfrak{P}(X)| = |2^X|$. The bijection is the function mapping each set $A \subseteq X$ to its characteristic function χ_A, given by $\chi_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A. \end{cases}$ The inverse of this mapping is given by $(f : X \longrightarrow 2) \longmapsto \{x \in X \big| f(x) = 1\}$.

11. $|[\omega]^{<\omega}| = |\omega|$, where $[\omega]^{<\omega}$ denotes the set of all finite subsets of ω. Since a finite subset of ω has a characteristic function with only finitely many ones, we can map each such subset to the finite ordinal whose binary expansion coincides with such characteristic function. In other words, the mapping is given by $A \longmapsto \sum_{n \in A} 2^n$, and it is a bijection.

12. $|(0, 1)| = |\mathbb{R}|$, as witnessed by the mapping $x \longmapsto \frac{1}{\pi}\arctan(x) + \frac{1}{2}$.

13. $|\mathbb{N}| = |\mathbb{Q}|$. In order to prove this, start with a bijection $\varphi : \mathbb{N} \longrightarrow \mathbb{N} \times \mathbb{N}$, and then use the recursion theorem to define $f : \mathbb{N} \longrightarrow \mathbb{Q}^+ = \{q \in \mathbb{Q} \big| q > 0\}$ by[2] $f(1) = [\varphi(1)]_\sim$, and

$$f(n + 1) = [\varphi(\min\{k \in \mathbb{N} \big| [\varphi(k)]_\sim \notin \operatorname{ran}(f \restriction \{1, \dots, f(n)\})\})]_\sim;$$

this mapping is a bijection (which corresponds to traversing the diagonal of the $\mathbb{N} \times \mathbb{N}$-matrix whose (n, m)-th entry is the rational number n/m, while skipping those rational numbers that have already been counted by virtue of their being equivalent to some other fraction that was counted before). To finish the proof, just observe that $|\mathbb{Q}^-| = |\mathbb{Q}^+| = |\mathbb{N}|$ (where $\mathbb{Q}^- = \{q \in \mathbb{Q} \big| q < 0\}$, and the equipotence claimed is witnessed by the bijection $q \longmapsto -q$), and thus a bijection witnessing $|\mathbb{N}| = |\mathbb{Q}^+ \cup \mathbb{Q}^- \cup \{0\}| = |\mathbb{Q}|$ can be constructed by using the same trick that was used to show that $|\mathbb{N}| = |\mathbb{Z}|$.

Definition 120. Let X be a set.

1. X is **finite** if $|X| = |n|$ for some $n < \omega$.

2. X is **infinite** if it is not finite.

3. X is **countably infinite** (in ancient times, people would say *denumerable*) if $|X| = |\mathbb{N}|$.

4. X is **countable** if it is either finite or countably infinite, in which case it absolutely makes sense to write $|X| \leq \aleph_0$ or $|X| \leq \omega$.

5. X is **uncountable** if it is not countable.

[1] In Jorge Luis Borges's short story *El Aleph*, we see the main character stating, about the letter $\aleph$, that: "para la *Mengenlehre*, es el símbolo de los números transfinitos, donde el todo no es mayor que alguna de las partes".

[2] Recall that rational numbers are equivalence classes of ordered pairs in $\mathbb{N} \times \mathbb{N}$. Hence $[(n, m)]_\sim$ is just denoting the rational number $\frac{n}{m}$.

Given the examples above, it would seem as though infinite sets can always be put into correspondence with one another. More specifically, most of the infinite sets mentioned in Example 119 happen to be countable. One might ask, very naturally, whether there is some set out there which is uncountable. The answer to this question is given by the following theorem, which constituted an extremely important event in the history of set theory. In fact, many authors identify the moment where Cantor proved this theorem (which was a cold December, in 1873) as the moment in which set theory itself was born.

Theorem 121 (Cantor)**.** *The set $\mathbb{R}$ of real numbers is uncountable. That is, $|\mathbb{N}| \neq |\mathbb{R}|$.*

Proof. Let $f : \mathbb{N} \longrightarrow \mathbb{R}$ be any function. We will prove that f cannot be a bijection by explicitly exhibiting a real number $x \in \mathbb{R} \setminus \operatorname{ran}(f)$. In order to do this, recursively define $\varphi : \mathbb{N} \longrightarrow \mathscr{I}$, where $\mathscr{I}$ is the set of all closed intervals in $\mathbb{R}$, by

$$\varphi(1) = \begin{cases} [0,1] \text{ if } f(1) \notin [0,1], \\ [2,3] \text{ otherwise,} \end{cases}$$

and, assuming that $\varphi(n) = [a, b]$ then we let

$$\varphi(n+1) = \begin{cases} \left[a, a + \frac{b-a}{3}\right] & \text{if } f(n+1) \notin \left[a, a + \frac{b-a}{3}\right], \\ \left[a + \frac{2(b-a)}{3}, b\right] & \text{otherwise.} \end{cases}$$

Note that $\varphi(1) \supseteq \varphi(2) \supseteq \cdots \supseteq \varphi(n) \supseteq \cdots$, with each $\varphi(n)$ a closed interval of length $\frac{1}{3^{n-1}}$ and such that $f(n) \notin \varphi(n)$. Hence (by the completeness of $\mathbb{R}$, as you guys surely learned in calculus/analysis) we know that

$$\bigcap_{n \in \mathbb{N}} \varphi(n)$$

is nonempty (it is, in fact, a singleton, due to how the lengths of the $\varphi(n)$ converge to 0), and if $x \in \bigcap_{n \in \mathbb{N}} \varphi(n)$, then since $(\forall n \in \mathbb{N})(f(n) \notin \varphi(n))$, it must be that $(\forall n \in \mathbb{N})(x \neq f(n))$. This means that $x \notin \operatorname{ran}(f)$, and so f is not surjective. Hence there are no bijections $: \mathbb{N} \longrightarrow \mathbb{R}$, and we are done. □

5.2 A partial order among cardinalities

We now know how to determine whether or not two sets (regardless of whether or not they are infinite) have the same number of elements. Now we proceed to define ways of comparing the number of elements of two sets in a somewhat more detailed fashion.

Definition 122. Let A and B be two sets.

1. We will say that $|A| \leq |B|$ if there exists an injection $f : A \longrightarrow B$. Equivalently, if $|A| = |C|$ for some $C \subseteq B$.

2. We will say that $|A| < |B|$ if $|A| \leq |B|$ and $\neg(|A| = |B|)$.

Remark 123.

1. The relation $|A| \leq |B|$ is "well-defined" with respect to equipotence, in the sense that if $|A| = |X|$ and $|B| = |Y|$ then $|A| \leq |B| \iff |X| \leq |Y|$ (given an injection from A to B and bijections from X to A and from B to Y, composing those three functions in the appropriate order yields an injection from X to Y).

2. Since $\operatorname{id} \upharpoonright A : A \longrightarrow A$ is injective, we have that $|A| \leq |A|$. Hence this relation is a reflexive (class) relation.

3. If $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are both injections, then so is $(g \circ f) : A \longrightarrow C$; thus the class relation $|A| \leq |B|$ is also transitive.

If the relation $|A| \leq |B|$ was antisymmetric, then it would be a partial order. This is not quite the case, however, the following deep theorem shows that this relation behaves like an antisymmetric relation modulo the relation of equipotence. This, furthermore, justifies our definition of the corresponding "strict" relation $|A| < |B|$.

Theorem 124 (Cantor–Schröder–Bernstein)**.** *If $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$. In other words, if there are injective functions $f : A \longrightarrow B$ and $g : B \longrightarrow A$, then there is a bijective function $h : A \longrightarrow B$.*

Proof. Suppose that $f : A \longrightarrow B$ and $g : B \longrightarrow A$ are two injections. Recursively define $C_0 = A \setminus \operatorname{ran}(g)$, $D_0 = f[C_0]$ and $C_{n+1} = g[D_n]$, $D_{n+1} = f[C_{n+1}]$.

Notice the following facts:

- The sets $C_0, C_1, \dots, C_n, \dots$ are pairwise disjoint.
- The sets $D_0, D_1, \dots, D_n, \dots$ are pairwise disjoint as well.
- For each $n < \omega$, $f \upharpoonright C_n$ is a bijection onto D_n.
- For each $n < \omega$, $g \upharpoonright D_n$ is a bijection onto C_{n+1}.

If we let $X = A \setminus \bigcup_{n<\omega} C_n$ and $Y = B \setminus \bigcup_{n<\omega} D_n$, then $g \upharpoonright Y$ maps every element in Y to an element in X (since the only x such that $g(x) \in C_n$ are those $x \in D_{n-1}$), furthermore, this function is onto X (since $X \subseteq \operatorname{ran}(g)$, and every $x \in D_n$ is such that $g(x) \in C_{n+1}$, hence given an arbitrary $x \in X$ there must be a $y \in Y$ such that $g(y) = x$).

This means that we can let $h : A \longrightarrow B$ be given by

$$h(x) = \begin{cases} f(x) \text{ if } x \in \bigcup_{n\in\omega} C_n, \\ g^{-1}(x) \text{ otherwise,} \end{cases}$$

and h will be a bijection. □

We proceed to mention a couple of applications of the Cantor–Schröder–Bernstein theorem.

1. $(0,1) \subseteq [0,1] \subseteq \mathbb{R}$, so the corresponding inclusion mappings witness that $|\mathbb{R}| = |(0,1)| \leq |[0,1]| \leq |\mathbb{R}|$. Therefore $|\mathbb{R}| = |[0,1]| = |(0,1)|$.
2. To show that $|\mathfrak{P}(\omega)| = |2^\omega| = |(0,1)|$, we will show two injections witnessing $|(0,1)| \leq |2^\omega| \leq |(0,1)|$. The first injection is the one mapping every $x \in (0,1)$, which can be expressed as a sequence of binary digits, $x = 0.d_0d_1d_2 \cdots d_n \cdots$ (assumed not to end with a sequence of infinitely many 1s), to the sequence $\langle d_n \big| n < \omega \rangle$. The second injection is the one mapping each $A \subseteq \omega$ to the number in $(0,1)$ whose ternary expansion equals the characteristic function of A, in other words, we are mapping $A \longmapsto \sum_{n\in A} \frac{1}{3^n}$. Therefore, by the Cantor–Schröder–Bernstein theorem, $|\mathbb{R}| = |(0,1)| = |2^\omega| = |\mathfrak{P}(\omega)|$.

Since we have seen in the past that $\mathbb{N} \times \mathbb{N}$ is equipotent with $\mathbb{N}$, we can appropriately arrange bijections[3] to prove that $\mathbb{N}$ is equipotent to $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$, and, continuing by induction, we get that $\mathbb{N}^n$ is countable for all $n < \omega$. Similarly, we will be able to conclude that, for all n, the set $\mathbb{R}^n$ is equipotent to $\mathbb{R}$, as soon as we provide a bijection between $\mathbb{R} \times \mathbb{R}$ and $\mathbb{R}$. To do this, it will suffice[4] to exhibit a bijection between $2^\omega \times 2^\omega$ and 2^ω. Such a bijection can be obtained by "interleaving two sequences", that is, by mapping the ordered pair $(\langle x_n \big| n < \omega \rangle, \langle y_n \big| n < \omega \rangle)$ to the single sequence $\langle x_0, y_0, x_1, y_1, \dots, x_n, y_n, x_{n+1}, y_{n+1}, \dots \rangle$ (formally speaking, that ordered pair is mapped to the sequence $\langle z_n \big| n < \omega \rangle$ which is defined by $z_{2n} = x_n$ and $z_{2n+1} = y_n$).

5.3 The continuum, the continuum hypothesis, the $\aleph$ sequence

We now have plenty of different sets that are countable, as well of plenty of different sets (all of them defined naturally) which are uncountable, and they all have the same cardinality: $|(0,1)| = |2^\omega| = |\mathbb{R}| = |\mathfrak{P}(\omega)| = |\mathfrak{P}(\mathbb{N})| = |\mathbb{R}^n|$. Thus it makes sense to introduce special symbols for these two cardinalities.

Definition 125. We introduce the following abbreviations:

1. the symbol $\aleph_0$ will abbreviate $|\omega| = |\mathbb{N}|$,
2. the symbol $\mathfrak{c}$ ("the continuum") will abbreviate $|\mathbb{R}| = \mathfrak{P}(\mathbb{N})|$.

Let us remember that we still have not properly defined what the cardinality of a set is, in other words, the symbol $|A|$ still does not have a meaning. What has a meaning are statements containing comparisons, such as $|A| = |B|$ or $|A| \leq |B|$. It is in this context that, every time we would write $|\omega|$ (respectively $|\mathbb{R}|$), we are now allowed to replace that with the symbol $\aleph_0$ (respectively $\mathfrak{c}$).

By Cantor's theorem, we know that $\aleph_0 \neq \mathfrak{c}$. In fact, it is the case that $\aleph_0 < \mathfrak{c}$ (this is because $\aleph_0 \leq \mathfrak{c}$, as witnessed by the inclusion map : $\mathbb{N} \longrightarrow \mathbb{R}$). This, along with the absence of naturally defined examples of sets with cardinalities other than $\aleph_0$ or $\mathfrak{c}$, led Cantor to the following very natural question.

Question 1 (Cantor)**.** *Is there a set X such that $\aleph_0 < |X| < \mathfrak{c}$?*

[3] Concretely, let $f : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ be a bijection. We now define the mapping $g : \mathbb{N} \times \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N} \times \mathbb{N}$ by $g(n, m, k) = (f(n, m), k)$. This is a bijection, and can be composed with f to obtain a bijection between $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$ and $\mathbb{N}$.

[4] Concretely, if $f : \mathbb{R} \longrightarrow 2^\omega$ and $g : 2^\omega \times 2^\omega \longrightarrow 2^\omega$ are two bijections, then the mapping $h : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$ given by $h(x, y) = f^{-1}(g(f(x), f(x)))$ will be a bijection.

The **Continuum Hypothesis**, abbreviated CH, is the statement that the question above has a negative answer. In other words, the symbols CH are an abbreviation for the following $\mathscr{L}_{\mathrm{ST}}$-formula:

$$\neg(\exists X)(|\mathbb{N}| < |X| < |\mathfrak{P}(\mathbb{N})|)$$

Although Cantor suffered quite a bit and never got to prove the Continuum Hypothesis (while suffering disparaging remarks by Kronecker and other members of the establishment of the time), eventually David Hilbert provided support to the idea that this problem is worth thinking about. In fact, the Continuum Hypothesis appeared as the first problem in Hilbert's famous list of problems for the (twentieth) century. The solution to this problem is provided by the following two theorems (whose proofs are quite standard, and covered in MATH 682).

Theorem 126 (Gödel, 1939)**.** *There exists a proper class* $\mathbf{L}$ *(known as Gödel's constructible universe) such that*

$$\mathbf{L} \vDash \mathsf{ZFC} + \mathsf{CH},$$

in particular, $\mathsf{ZFC} \nvdash \neg\mathsf{CH}$ *(we say that* CH *is consistent with* ZFC*).*

Theorem 127 (Cohen, 1960)**.** *(Roughly) out of a set* V *such that* $V \vDash \mathsf{ZFC}$*, it is possible to obtain a set* $V[G] \supseteq V$ *such that* $V[G] \vDash \mathsf{ZFC} + \neg\mathsf{CH}$*. In particular,* $\mathsf{ZFC} \nvdash \mathsf{CH}$ *(we say that* $\neg\mathsf{CH}$ *is consistent with* ZFC*, or that* CH *is independent from* ZFC*).*

Thus, in the end the CH turned out to be *undecidable* from the ZFC axioms. Cohen received a Fields Medal in 1964 for his theorem (which gave birth to what is now known as the *forcing technique*), which to this day remains the only Fields Medal that has ever been awarded to someone for work in mathematical logic. Nowadays, there is a lively topic of research within set theory, namely the study of the *cardinal characteristics of the continuum*, which investigates the extent to which the ZFC axioms allow the CH to fail.

Cantor's theorem that $\aleph_0 < \mathfrak{c}$ (in other words, $|\mathbb{N}| < |\mathfrak{P}(\mathbb{N})|$) has a more general version, also due to Cantor.

Theorem 128 (Cantor)**.** *For every set* X*,* $|X| < |\mathfrak{P}(X)|$*.*

Proof. The function $x \longmapsto \{x\}$ is an injection $: X \longrightarrow \mathfrak{P}(X)$, so $|X| \leq |\mathfrak{P}(X)|$. Now to prove that the inequality is strict, let $f : X \longrightarrow \mathfrak{P}(X)$ be any function; we will show that f cannot be surjective (and hence there is no bijection between X and $\mathfrak{P}(X)$). To see this, just notice that

$$W = \{x \in X \big| x \notin f(x)\} \notin \operatorname{ran}(f),$$

because otherwise we would have $W = f(x)$ for some x, and then we would have that $x \in W \iff x \notin f(x) = W$, which is a contradiction. □

The above argument is sometimes known as the *diagonal argument.* The reason for this is that we can think of the above proof as follows: any function $f : X \longrightarrow \mathfrak{P}(X)$ can be represented by an $X \times X$-matrix with entries in the set $\{0, 1\}$: simply think of the x-th row as containing the characteristic function of $f(x)$. Then the set W defined above corresponds exactly to defining a new row that cannot be found among the ones in the matrix, by making sure that its x-th entry differs from the (x, x)-th entry of the matrix[5].

Corollary 129 (Cantor's Paradox)**.** *If* V *was a set, then (since it would necessarily follow that* $\mathfrak{P}(V) = V$*) we would have that* $|V| = |\mathfrak{P}(V)|$*, contradicting Cantor's theorem.*

Question 2 (Cantor)**.** *Is it the case that, for some set* X*, there exists a* Y *such that* $|X| < |Y| < |\mathfrak{P}(X)|$*?*

A negative answer to the above question is known as the **Generalized Continuum Hypothesis**, abbreviated GCH. In other words, GCH is an abbreviation for the following $\mathscr{L}_{\mathrm{ST}}$-formula:

$$\neg(\exists X)(\exists Y)(|X| < |Y| < |\mathfrak{P}(X)|).$$

Gödel also provided half of the solution for this problem in his 1939 work.

Theorem 130 (Gödel, 1939)**.** $\mathbf{L} \vDash \mathsf{ZFC} + \mathsf{GCH}$*, in particular* $\mathsf{ZFC} \nvdash \neg\mathsf{GCH}$*, i.e. the* GCH *is consistent with* ZFC*.*

And clearly Cohen's theorem also provides the second half of the answer to the GCH question, since $\mathsf{GCH} \Rightarrow \mathsf{CH}$. Hence, in Cohen's model not just $\neg\mathsf{CH}$, but in fact $\neg\mathsf{GCH}$, holds.

Cantor's theorem also allows us to introduce the sequence of "beth numbers".

[5] The reader is encouraged to draw such a matrix to make this explanation clearer.

Definition 131. Using the recursion theorem, we define the following sequence of sets:

1. $X_0 = \mathbb{N}$,
2. $X_{\alpha+1} = \mathfrak{P}(X_\alpha)$,
3. $X_\alpha = \bigcup_{\xi<\alpha} X_\xi$, if $\alpha = \bigcup \alpha$.

We introduce the symbol $\beth_\alpha$ to be an abbreviation (with the same caveats as in Definition 125) of $|X_\alpha|$.

Thus $\beth_0 = \aleph_0$ and $\beth_1 = \mathfrak{c}$. Much more famous than the sequence of beth numbers is the sequence of aleph numbers. These are cardinalities of certain specific ordinal numbers, so in order to introduce this new sequence, we need to prove the following theorem. This theorem provides us with a way of obtaining larger and larger ordinals, in the same way that Cantor's theorem gave us a way of obtaining larger and larger sets.

Theorem 132 (Hartogs)**.** *For every set A, there exists an ordinal α such that $|\alpha| \nleq |A|$ (that is, there is an $\alpha \in$ **Ord** such that there is no injection $: \alpha \longrightarrow A$).*

Proof. Let

$$W = \{(B, \leq) \in \mathfrak{P}(A) \times \mathfrak{P}(A \times A) \big| B \subseteq A \wedge \leq \text{ is a well-order}\}.$$

By replacement, $X = \{\text{otp}(B, \leq) \big| (B, \leq) \in W\}$ exists (is a set). Note that this set is none other than $X = \{\beta \in \textbf{Ord} \big| |\beta| \leq |A|\}$ (if there is an injection $f : \beta \longrightarrow A$, then $\text{ran}(f)$ is a well-ordered subset of A with order-type β, when ordered with the order inherited from β and f; conversely, if $(B, \leq) \in W$, and $f : B \longrightarrow \beta$ is the isomorphism from B onto its ordertype β, then clearly f^{-1} injects β into A). Note that the set $X \subseteq \textbf{Ord}$ is transitive: if $\gamma \in \beta \in X$ then we have an injection $f : \beta \longrightarrow A$, which means that $f \upharpoonright \gamma : \gamma \longrightarrow A$ is another injection, thus $\gamma \in X$. This means that $X = \alpha$ is an ordinal (we just saw that it is transitive and, it being a set of ordinals, it is automatically well-ordered by $\in$), and clearly $|\alpha| \nleq |A|$ because otherwise we would have that $\alpha \in X = \alpha$. Moreover, since $(\forall \beta < \alpha)(|\beta| \leq |A|)$, we have that α is the least ordinal that does not inject into A. □

Given a set A, the (least) ordinal α whose existence is ensured by the previous theorem is known as the **Hartogs number** of A, usually denoted by A^+. Notice that, if α is an ordinal, then α is comparable with α^+ and, since $|\alpha^+| \nleq |\alpha|$, then it must be the case that $\alpha < \alpha^+$. Moreover, every β satisfying $\alpha \leq \beta < \alpha^+$ must be such that $|\beta| \leq |\alpha|$ (by minimality of $|\alpha^+|$), and since we also have that $|\alpha| \leq |\beta|$ (as witnessed by the inclusion mapping), the Cantor–Schröder–Bernstein theorem ensures that $|\alpha| = |\beta|$. The conclusion of all this is that $|\alpha^+|$ is the least ordinal $\geq \alpha$ which is not equipotent to α.

With this, we are ready to introduce the sequence of aleph numbers.

Definition 133. Using the recursion theorem on ordinals, we define the following sequence of ordinal numbers:

1. $\omega_0 = \omega$,
2. $\omega_{\alpha+1} = \omega_\alpha^+$,
3. $\omega_\alpha = \sup\{\omega_\xi \big| \xi < \alpha\}$ for $\alpha = \bigcup \alpha$,

and we declare the symbol $\aleph_\alpha$ to be an abbreviation (with the same caveats as in Definitions 125 and 131) of $|\omega_\alpha|$.

This sequence is strictly increasing in cardinality. For example, ω_1 is just the least ordinal without a bijection with ω (the least uncountable ordinal). Since there are bijections between ω and $\omega + n$, or even $\omega + \omega$ (or even $\omega \cdot \omega$), then we know that ω_1 must be larger than all of these ordinals. The ordinal ω_1 is characterized by the properties that it is uncountable, and every strictly smaller ordinal is countable.

I will let the reader ponder the extremely interesting question of whether $\mathbb{R}$ is equipotent to some ω_α (i.e. if $\mathfrak{c} = \aleph_\alpha$ for some $\alpha \in \textbf{Ord}$). In case the answer is affirmative, then we know by Cantor's theorem that $\alpha \neq 0$. Could it be that $\alpha = 1$?

5.4 Cardinal Arithmetic

It is possible to define arithmetic operations (addition, multiplication and exponentiation) on cardinal numbers (modulo appropriate caveats to deal with the fact that the cardinality of a set has still not been defined). The definitions, as well as proofs for their basic properties, can be worked out by the reader, should she wish to do so, by means of going through the exercises on the corresponding worksheet, which in this document is Appendix E.

Chapter 6

The Axiom of Choice

The Axiom of Choice is a somewhat special axiom, due to its non-constructive nature. It provides us with a way of making infinitely many choices at once, without the need to explicitly describe those choices. This axiom was controversial at some point in history, with numerous people hesitating to use it. Eventually, mathematicians have slowly discovered that this axiom has so many desirable consequences, that most have dropped their hesitations by now. The reader should begin by working through the worksheet on the Axiom of Choice (Appendix F), to get comfortable with detecting when does a proof require the usage of this axiom to be carried out.

6.1 Equivalences of the Axiom of Choice

There are many statements that have been studied over the years, that have been eventually discovered to be logically equivalent to the Axiom of Choice. A relatively extensive (for the elementary level of this course) list of statements that can be proved to be equivalent to this axiom can be found in Appendix G. In what follows, we proceed to prove most of those equivalences[1].

Theorem 134. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{AC}_2$

Proof.

$\Leftarrow$: Given a pairwise disjoint nonempty family of nonempty subsets X, let $f : X \longrightarrow \bigcup X$ be a choice function (i.e. $(\forall x \in X)(f(x) \in x)$) and let $A = \operatorname{ran}(f) = \{f(x) \big| x \in X\}$. Then for every $x \in X$, it is the case that $A \cap x = \{f(x)\}$, which is a singleton.

$\Rightarrow$: Let X be a nonempty family of nonempty sets. Note that if X was pairwise disjoint then this would be easy, for if A is a selector then we would be able to define $f : X \longrightarrow \bigcup X$ by $f(x) = \bigcup(A \cap x)$.

So for arbitrary X, we just need to "disjointify" it first. To this effect, we let $Y = \{x \times \{x\} \big| x \in X\}$ (it is immediate that Y is a set by replacement; one can also prove this without using replacement by extracting Y as a subset of $\mathfrak{P}(\mathfrak{P}(\mathfrak{P}((\bigcup X) \cup X))))$. This way, if $x \neq y$ then we have that $x \times \{x\} \cap y \times \{y\} = \varnothing$. By AC, we can now choose a selector A for Y, and define $f : X \longrightarrow \bigcup X$ by letting $f(x)$ be the first coordinate of the ordered pair $\bigcup(A \cap (x \cup \{x\}))$, which must be an element of x.

□

From now on, we will drop the subscript from the statement AC_2. Each time we use either AC or AC_2, we will simply mention AC and let the reader utilize the context to determine whether we are using the "original" AC (the version that provides a selector for a pairwise disjoint family) or AC_2 (the version that provides a choice function).

Enderton's way of stating the Axiom of Choice (as done in his book) is not very standard, but it is fairly straightforward to prove that it is equivalent to either of the other two formulations (and the reader is encouraged to provide such a proof by herself). We denote this particular statement by AC_3 in Appendix G.

We know that a function is injective if and only if it has a left inverse; we also know that functions with a right inverse must be surjective. To have full symmetry and duality between these two statements, it would be nice if we could say that every surjective function must have a right inverse; it turns out that this statement requires the Axiom of Choice to be proved, and it is in fact equivalent to it.

[1] In class, there is clearly no time to go through all of these equivalences. What I typically do is prove the equivalence between AC and AC_2, and then between these and ES. Then I do the proof that "the big three" (AC, WO and ZL) are equivalent, I mention that AC implies that every vector space has a basis (as an application of Zorn's Lemma), and close the topic with the proof that AC is equivalent to IC.

Theorem 135. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{ES}$.

Proof.

⇒: Suppose that $f : A \longrightarrow B$ is surjective, and consider the family of sets

$$X = \{\{a \in A \big| f(a) = b\} \big| b \in B\} = \{f^{-1}[\{b\}] \big| b \in B\}.$$

The fact that f is onto B ensures that each element of this family is nonempty. Let $g : B \longrightarrow \bigcup X = A$ be a choice function. This means that, for every $b \in B$, $g(b) \in f^{-1}[\{b\}]$; in other words, $f(g(b)) = b$. That is, $f \circ b = \mathrm{id} \restriction B$.

⇐: Let $X \neq \varnothing$ be a pairwise disjoint family, each of whose elements are nonempty. Define $f : \bigcup X \longrightarrow X$ by

$$f(a) = \bigcup \{x \in X \big| a \in x\},$$

in other words, we are mapping each $a \in \bigcup X$ to the (unique, since we are assuming X to be pairwise disjoint) $x \in X$ such that $a \in x$. Each x is nonempty, so there is an $a \in x$ and this means that $f(a) = x$. Thus f is surjective. Therefore we can grab a $g : X \longrightarrow \bigcup X$ such that $f \circ g = \mathrm{id} \restriction X$. This means that, for each $x \in X$, $f(g(x)) = x$, where $f(g(x))$ is by definition the unique $z \in X$ with $g(x) \in z$; in other words, $g(x) \in x$. This means that (since g is a choice function on X) $\mathrm{ran}(g)$ is a selector for X.

□

We now address the so-called *Well-Ordering Principle*, which states that every set can be equipped with a well-ordering. The Well-Ordering Principle has an interesting history. In the beginnings of set theory, when Cantor was thinking strongly about how infinite sets should be thought of as not being very different from finite sets (which in practice consisted in his considering all sets as being well-ordered), he explicitly wrote that this principle was self-evident. I have read claims from other authors (although not from Cantor himself, but I have by no means extensively read Cantor's work) that he later changed his mind regarding this issue. Afterwards, however, Zermelo "proved" the Well-Ordering Principle: the publication where he first stated his axiomatization of set theory (what we now know as ZC^- Set Theory), was precisely the one where he proves the Well-Ordering Principle. In fact, the main point of Zermelo's axiomatization of set theory was to explicitly point out the assumptions that seemed necessary to secure this theorem. So in a sense, the Well-Ordering Principle is directly responsible for the fact that we nowadays have the specific axiomatization ZFC as our framework for set theory. The Well-Ordering Principle turns out to be equivalent to the Axiom of Choice, and this is spelled out in the following theorem.

Theorem 136. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{WOP}$.

Proof.

⇐: Let X be a nonempty family of nonempty sets, and let $\leq$ be a well-order for $\bigcup X$. We obtain a choice function for X by letting $f : X \longrightarrow \bigcup X$ be given by $f(x) = \min_{\leq}(x)$.

⇒: This proof bears an outstanding similarity to the proof that ordinal numbers form a complete class of representatives for well-ordered sets. Let A be a (nonempty, since if A is empty then the theorem is obvious) set, and pick a choice function $f : \{X \subseteq A \big| X \neq \varnothing\} \longrightarrow \bigcup \{X \subseteq A \big| X \neq \varnothing\} = A$. Pick some $u \notin A$ (u will just be a "garbage value", representing that a function that we want to define will be "undefined") and recursively define a class function $F : \mathbf{Ord} \longrightarrow A \cup \{u\}$ by letting

$$F(\alpha) = \begin{cases} f(A \setminus \{F(\xi) \big| \xi < \alpha\}); \text{ if } A \setminus \{F(\xi) \big| \xi < \alpha\} \neq \varnothing, \\ u; \text{ otherwise.} \end{cases}$$

Note that, if α is such that $F(\alpha) = u$, then $F(\beta) = u$ for all $\beta \geq \alpha$ as well ("once F is undefined, it remains undefined"). This means that the α for which $F(\alpha) \neq u$ (the α for which F "is defined") form an initial segment of the class $\mathbf{Ord}$ (and hence the class of such α is either an ordinal number, or the whole class $\mathbf{Ord}$). Notice also that, by construction, if α is such that $(\forall \xi < \alpha)(F(\xi) \neq u)$, then $F \restriction \alpha : \alpha \longrightarrow A$ is an injective function. Hence by Hartogs's theorem[2], there must be one α such that $F(\alpha) = u$, and if we take the least such α, we obtain a bijection $g = F \restriction \alpha : \alpha \longrightarrow A = \{(\xi, a) \in \alpha \times A \big| a = F(\xi)\}$, which is a set (as opposed to a proper class) by the Axiom of Comprehension. Given this bijection, a well-ordering for A arises from "copying" the one on α via the bijection g, in other words,

$$\leq = \{(a, b) \in A \times A \big| (\exists \gamma \leq \beta \in \mathbf{Ord})(F(\beta) = a \wedge F(\gamma) = b)\}$$

is a well-order relation on A, and we are done.

[2] Alternatively, notice that if $u \notin \mathrm{ran}(F)$, then $F : \mathbf{Ord} \longrightarrow A$ is injective, $\mathrm{ran}(F) \subseteq A$ is a set, and we have the class function $F^{-1} : \mathrm{ran}(F) \longrightarrow \mathbf{Ord}$. By the Axiom of Replacement, this would imply that $F^{-1}[\mathrm{ran}(F)] = \mathbf{Ord}$ is a set, a contradiction.

□

Arguably, the two most classical equivalences of the Axiom of Choice are the Well-Ordering Principle and Zorn's Lemma. We have already explained the former; the latter is used fairly frequently outside of set theory. In fact, it is essentially designed to provide non-set theorists with a way of performing constructions without needing to know anything about ordinal numbers, or about the transfinite recursion theorem. The proof of this equivalence follows.

Theorem 137. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{ZL}$.

Proof.

⇒: Let $(\mathbb{P}, \leq)$ be a preorder such that every linearly ordered $X \subseteq \mathbb{P}$ has a lower bound, and let $f : \{X \subseteq \mathbb{P} \big| X \neq \varnothing\} \longrightarrow \mathbb{P}$ be a choice function. Once again pick a garbage "undefined" value $u \notin \mathbb{P}$ and recursively define a class function $G : \mathbf{Ord} \longrightarrow \mathbb{P} \cup \{u\}$ by

$$G(\alpha) = \begin{cases} f(\{x \in \mathbb{P} \big| (\forall \xi < \alpha)(x < G(\xi))\}); & \text{if } \{x \in \mathbb{P} \big| (\forall \xi < \alpha)(x < G(\xi))\} \neq \varnothing, \\ u; & \text{otherwise} \end{cases}$$

(with the convention that every $x \in \mathbb{P}$ satisfies $x < u$). In other words, $G(\alpha)$ is some choice of a strict lower bound for the set $\{G(\xi) \big| \xi < \alpha\}$, if there is at least one such strict lower bound; or else $G(\alpha)$ is just undefined.

As is the case in these proofs (of which we have already seen at least two instances), this is a function that once undefined, remains undefined; and as long as it is not undefined it remains injective (because the lower bounds that we choose are strict, hence every $G(\alpha)$ which is not undefined is distinct from all of the $G(\xi)$ for $\xi < \alpha$). Hence Hartogs's theorem guarantees that the function G is eventually undefined, that is, $u \in \operatorname{ran}(G)$. We let α be the least ordinal such that $G(\alpha) = u$. Note then that the set $\{G(\xi) \big| \xi < \alpha\}$ is a linearly ordered subset of $\mathbb{P}$ (since for every $\xi < \eta < \alpha$, the definition of G guarantees that $G(\eta) < G(\xi)$). By our assumption on $\mathbb{P}$, it must then be the case that there exists lower bound x for $\{G(\xi) \big| \xi < \alpha\}$; the fact that $G(\alpha)$ is undefined implies that this x cannot be a strict lower bound for that set. Furthermore, every $y \in \mathbb{P}$ satisfying $y < x$ would be a strict lower bound for the aforementioned set, contradicting that $G(\alpha)$ is undefined. Thus it must be the case that $(\forall y \in \mathbb{P})(y \leq x \Rightarrow y = x)$, and therefore x is a minimal element[3] of the partially ordered set $\mathbb{P}$.

⇐: Let X be a nonempty family of nonempty pairwise disjoint sets, and define

$$\mathbb{P} = \{A \subseteq \bigcup X \big| (\forall x \in X)(|A \cap x| \leq 1\},$$

ordered by $A \leq B$ iff $B \subseteq A$. Note that, if $\mathscr{C} \subseteq \mathbb{P}$ is a linearly ordered subset, then $A = \bigcup \mathscr{C} \in \mathbb{P}$. To see this, note first that $A \subseteq \bigcup X$, and if $x \in X$ is such that $y, z \in A \cap x$ then for some $B, C \in \mathscr{C}$ we have that $y \in B$ and $z \in C$. However $\mathscr{C}$ is linearly ordered, so either $B \subseteq C$ or $C \subseteq B$, suppose without loss of generality that $B \subseteq C$. Then both $y, z \in C$ and since $C \in \mathbb{P}$ and $y, z \in C \cap x$, it follows that $y = z$. Hence $|A \cap x| \leq 1$ and so $A \in \mathbb{P}$, therefore A is a lower bound for $\mathscr{C}$.

So $\mathbb{P}$ satisfies the hypotheses of Zorn's Lemma and it therefore has a minimal element A. So for every $x \in X$, there is at most one element in $A \cap x$. Suppose that for some $x \in X$ we had $A \cap x = \varnothing$, then by picking some $y \in x$ and taking $A \cup \{y\}$ we obtain another element of $\mathbb{P}$ strictly below A, contradicting minimality of A. Therefore $A \cap x$ is a singleton for all $x \in X$, so A is a selector for X, and we are done.

□

The proof of the reverse implication in the previous theorem constitutes an example of the application of Zorn's Lemma to prove non-constructive existential statements. For a further example, let us have a look at the following proof, utilizing Zorn's Lemma, that every vector space has a basis.

Theorem 138. $\mathsf{ZF} \vdash \mathsf{AC} \Rightarrow$ *every vector space has a basis.*

(This is actually an equivalence, rather than a one-way conditional. The converse of this theorem was proved by Andreas Blass (in *Existence of bases implies the Axiom of Choice*, Contemporary Mathematics **31** (1984), 31–33), and the proof is highly nontrivial[4].)

[3] Furthermore, the fact that x is not a strict lower bound for $\{G(\xi) \big| \xi < \alpha\}$ implies that $x = G(\xi)$ for some $\xi < \alpha$, and the fact that G is strictly decreasing allows us to conclude then that $\xi = \max(\alpha)$. In other words, α is a successor ordinal, and if $\alpha = \xi + 1$ then $G(\xi)$ is a minimal element for $\mathbb{P}$.

[4] In particular, what Blass proves is that the existence of bases on every vector space (or at least on all vector spaces over every field of a given characteristic) implies the so-called *principle of multiple choice*, which is the statement that for every indexed family $\{X_i \big| i \in I\}$ of nonempty sets, there exists a family $\{Y_i \big| i \in I\}$ such that $(\forall i \in I)(Y_i \subseteq X_i \wedge Y_i \text{ is finite})$ (that is, out of a family of sets we can always choose finitely many elements from each member of the family). That the principle of multiple choice implies the Axiom of Choice is, once again, a statement with a highly nontrivial proof, and this proof utilizes the Axiom of Foundation in a very strong way.

Proof. This is a nice application of Zorn's Lemma (although it is infinitely more natural, for the mathematician trained in the usage of ordinal numbers, to carry out this proof by transfinite recursion). Let V be a vector space and consider the forcing notion

$$\mathbb{P} = \{X \subseteq V \big| X \text{ is linearly independent}\},$$

ordered by $X \leq Y$ iff $Y \subseteq X$.

We will show that, if $\mathscr{C} \subseteq \mathbb{P}$, then $X = \bigcup \mathscr{C} \in \mathbb{P}$. It is certainly the case that $X \subseteq V$, and if $a_1, \dots, a_k$ are scalars with $x_1, \dots, x_k \in X$ satisfying $a_1 x_1 + \cdots + a_n x_n = 0$, then for each i we have some $X_i \in \mathscr{C}$ such that $x_i \in X_i$. But since $\mathscr{C}$ is linearly ordered, there is some X_i such that $X_j \subseteq X_i$ for all j (i.e. X_i is a minimum element in the family $\{X_1, \dots, X_k\}$). Since X_i is linearly independent and $x_1, \dots, x_n \in X_i$, we can conclude that $a_1 = \cdots = a_n = 0$. This shows that X is linearly independent, so $X \in \mathbb{P}$ and clearly X is a lower bound for $\mathscr{C}$.

Since $\mathbb{P}$ satisfies the hypotheses of Zorn's Lemma, it must have a minimal element $X \in \mathbb{P}$. Then X is a linearly independent subset of V; in order to show that it is a basis, we will show that X spans all of V, so suppose not. Then there is some vector $v \in V$ such that no linear combination of elements of X equals v. This means that $X \cup \{v\}$ is linearly independent, and so it constitutes an element of $\mathbb{P}$ strictly below X, contradicting the minimality of X. Thus X spans V and we are done. □

In order to exhibit a couple more uses of Zorn's Lemma, we will show the proof of a couple more equivalences of the Axiom of Choice. Both of these equivalences are examples of *maximality principles*, statements that allow us to prove non-constructive existential statements by guaranteeing that certain maximal (with respect to some meaningful notion of ordering, typically $\subseteq$) objects exist. Zorn's Lemma itself is the first and best known example of a maximality principle (think about flipping the order in the statement of Zorn's Lemma, so that you get a maximal element instead of a minimal one).

Theorem 139. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{HM}$.

Proof.

⇒: For this implication, we will use ZL. Let $(X, \leq)$ be a partially ordered set, and suppose that $Y \subseteq X$ is a linearly ordered subset. We let $\mathbb{P} = \{Z \subseteq X \big| Y \subseteq Z \wedge Z \text{ is linearly ordered}\}$, and we order it by declaring $Z \preceq W$ iff $Z \supseteq W$. Then $(\mathbb{P}, \preceq)$ is a partially ordered set. Moreover, if $\mathscr{C} \subseteq \mathbb{P}$ is linearly ordered, then $\bigcup \mathscr{C} \in \mathbb{P}$. To see this, note that if $x, y \in \bigcup \mathscr{C}$, then $x \in Z$ and $y \in W$ for some $Z, W \in \mathscr{C}$; since $\mathscr{C}$ is linearly ordered, we must have that either $Z \subseteq W$ or $W \subseteq Z$. Assume without loss of generality that $W \subseteq Z$, then $x, y \in Z$ and since $Z \in \mathbb{P}$, it is linearly ordered, which implies that either $x \leq y$ or $y \leq x$. This proves that the set $\bigcup \mathscr{C}$ is linearly ordered, hence an element of $\mathbb{P}$, and it clearly is a lower bound.

Thus the partially ordered set $(\mathbb{P}, \preceq)$ satisfies the hypothesis of Zorn's Lemma, and so it must have a minimal element $Z \in \mathbb{P}$. It is straightforward to verify that such a minimal element Z must be a $\subseteq$-maximal linearly orderded subset of X.

⇐: Let X be a family of nonempty sets. Consider the set $\mathbb{P} = \{f \subseteq X \times (\bigcup X) \,\big|\, f \text{ is a function} \wedge (\forall x \in \text{dom}(f))(f(x) \in x)\}$, ordered by $f \leq g$ if and only if $f \subseteq g$ (thus $\mathbb{P}$ consists of partial choice functions, that get bigger the more elements of X they choose from). Then $\{\varnothing\}$ is a linearly ordered subset of $\mathbb{P}$, so by HM there must be a maximal linearly ordered $\mathscr{C} \subseteq \mathbb{P}$ extending $\{\varnothing\}$. Notice that, if we let $f = \bigcup \mathscr{C}$, then in fact $f \in \mathbb{P}$. To see this, note first that f will be a binary relation, with $\text{dom}(f) \subseteq X$. Furthermore, if $(x, a), (x, b) \in f$, then $(x, a) \in g$ and $(y, b) \in h$ for some $g, h \in \mathscr{C}$. Since $\mathscr{C}$ is linearly ordered, either $g \subseteq h$ or $h \subseteq g$; assume without loss of generality that $g \subseteq h$. Then $(x, a), (x, b) \in h$, and h is a function, which implies that $a = b$ and moreover (since $h \in \mathbb{P}$) $a = b = h(x) \in x$. Thus f satisfies that $(\forall x \in \text{dom}(f))(f(x) \in x)$; so we are done justifying that $f \in \mathbb{P}$. Moreover, if there was an $x \in X$ with $x \notin \text{dom}(f)$, then by picking an $a \in x$ we would be able to obtain the element $f \cup \{(x, a)\} \in \mathbb{P}$, which would be a strict upper bound for $\mathscr{C}$. This would mean that $\mathscr{C} \cup \{f \cup \{(x, a)\}\}$ is a linearly ordered subset of $\mathbb{P}$ properly containing $\mathscr{C}$, which was assumed to be $\subseteq$-maximal, a contradiction. Therefore we must conclude that $\text{dom}(f) = X$, and so f is indeed a choice function on X.

□

Theorem 140. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{TL}$.

Proof.

⇒: For this implication, we will once again use ZL. Suppose that we have a family $\mathscr{X}$ of finite character. We partially order $\mathscr{X}$ by $X \leq Y$ iff $X \supseteq Y$. We claim that $(\mathscr{X}, \leq)$ satisfies the hypothesis of Zorn's Lemma. To see this, suppose that $\mathscr{C} \subseteq \mathscr{X}$ is a linearly ordered subset. If we could show that $\bigcup \mathscr{C} \in \mathscr{X}$, then it would follow immediately that $\bigcup \mathscr{C}$ is a lower bound for $\mathscr{C}$, so in what follows our efforts will be focused in showing that $\bigcup \mathscr{C} \in \mathscr{X}$. For this we will use the finite character of the family $\mathscr{X}$: it suffices to show that, for every finite $F \subseteq \bigcup \mathscr{C}$, it is the case that $F \in \mathscr{X}$. So suppose that $F \subseteq \bigcup \mathscr{C}$ is finite. For each $x \in F$, there is some[5] $X_x \in \mathscr{C}$ such that $x \in X_x$. Since $\{X_x | x \in F\}$ is finite and linearly ordered, it must have a minimum, say X_z. In other words, $(\forall x \in F)(X_x \subseteq X_z)$. Hence every $x \in F$ belongs to X_z, that is, $F \subseteq X_z$. Since $X_z \in \mathscr{X}$, with $\mathscr{X}$ of finite character, and F a finite subset of X_z, we can conclude that $F \in \mathscr{X}$. Hence every finite $F \subseteq \bigcup \mathscr{C}$ must belong to $\mathscr{X}$; by the finite character of $\mathscr{X}$, we can conclude that $\bigcup \mathscr{C} \in \mathscr{X}$.

Therefore, by Zorn's Lemma, we can pick a minimal element $X \in \mathscr{X}$; this is clearly a $\subseteq$-maximal element of $\mathscr{X}$.

⇐: The proof of this implication mirrors the proof that ZL implies AC. Let X be a nonempty, pairwise disjoint family of nonempty sets. Let $\mathscr{X} = \{A \subseteq \bigcup X \big| (\forall x \in X)(|A \cap x| \leq 1)\}$, the set of partial selectors for the family X. We will see that $\mathscr{X}$ is of finite character. First, note that if $A \in \mathscr{X}$ then every subset of A is an element of $\mathscr{X}$ as well; in particular this holds for finite subsets of A and so $[A]^{<\omega} \subseteq \mathscr{X}$. Conversely, suppose that A is a set, all of whose finite subsets belong to $\mathscr{X}$. Note that, for each $x \in A$, we have $\{x\} \in \mathscr{X}$, which by definition implies that $\{x\} \subseteq \bigcup X$, or in other words, $x \in \bigcup X$; this shows that $A \subseteq \bigcup X$. Now to see that A is a partial selector for X, suppose that there is an $x \in X$ and y, z such that $y, z \in A \cap x$. Then $\{y, z\}$ is a finite subset of A, so by our assumption we must have that $\{y, z\} \in \mathscr{X}$, meaning that $\{y, z\}$ is a partial selector and so $\{y, z\} \cap x$ must be either empty or a singleton; the obvious conclusion is that $y = z$. Hence A was indeed a partial selector for $\mathscr{X}$, and therefore $A \in \mathscr{X}$.

Thus $\mathscr{X}$ is a family of finite character, and so by TL it must have a $\subseteq$-maximal element A. It is straightforward to check that A must be a selector for the family X (if not, we would have $A \cap x = \varnothing$ for some $x \in X$, now picking a $z \in x$ we would get that $A \subsetneq A \cup \{z\} \in \mathscr{X}$, contradicting $\subseteq$-maximality of A).

□

We will now turn our attention to some equivalences that directly relate the Axiom of Choice with some other statement of interest in an area of mathematics that is not set theory. We already saw an example of that above, when we mentioned that the Axiom of Choice is equivalent to the statement that every vector space has a basis. The next two equivalences correspond to statements in algebra and topology, respectively, that happen to be closely intertwined with the Axiom of Choice. The first of these is just the statement that we can always guarantee the existence of maximal ideals.

Theorem 141. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{MIP}$.

Proof.

⇒: Let R be a ring, and let $\mathbb{P}$ be the collection of all ideals on R, partially ordered by $\supseteq$. Recall that ideals cannot contain 1, or else they would be the whole ring R. The union of ideals is in general not an ideal, however, it turns out that the union of a linearly ordered family of ideals is an ideal: for if $\mathscr{I}$ is a linearly ordered family of ideals and $x, y \in \bigcup \mathscr{I}$, then $x \in I$ and $y \in J$ for some $I, J \in \mathscr{I}$; now we have either $I \subseteq J$ and $J \subseteq I$, so assume without loss of generality that $I \subseteq J$. This means that $x, y \in J$, which is an ideal, so $x + y \in J \subseteq \bigcup \mathscr{I}$. Furthermore, if $z \in R$ is an arbitrary element, then $zx, xz \in J \subseteq \bigcup \mathscr{I}$ because J is and ideal. Finally, since $1 \notin I$ for all $I \in \mathscr{I}$, we can conclude that $1 \notin \bigcup \mathscr{I}$ either, and so $\bigcup \mathscr{I}$ is an ideal; since it includes each element of $\mathscr{I}$, and our $\mathbb{P}$ is partially ordered by $\supseteq$, the conclusion is that every linearly ordered subset of $\mathbb{P}$ has a lower bound. Hence by Zorn's Lemma, $\mathbb{P}$ must have a minimal element I; since the partial ordering was $\supseteq$, this means that I must be a maximal ideal[6].

⇐: Suppose that every ring has a maximal ideal, and let X be a pairwise disjoint family of nonempty sets. Consider the ring $R = \mathbb{Q}[\bigcup X]$, the polynomial ring with coefficients in $\mathbb{Q}$ and one indeterminate per each element of $\bigcup X$. Let P be the collection of all partial selectors for X (that is, each $p \in P$ is a subset $p \subseteq \bigcup X$ such that $(\forall x \in X)(|p \cap x| \leq 1)$), and for each $p \in P$ let I_p be the ideal of R generated by p. It is not hard to see that each of the I_p is a prime ideal, and therefore we can localize the ring R outside of these I_p, in other words, we can move to the larger ring RU^{-1} that contains a multiplicative inverse for each element of $R \setminus \left(\bigcup_{p \in P} I_p\right)$. Let I be a maximal ideal in the ring RU^{-1}, and let $J = I \cap R$; then J is an ideal of R satisfying $J \subseteq \bigcup_{p \in P} I_p$ (because every $r \in R \setminus \left(\bigcup_{p \in P} I_p\right)$ is a unit in RU^{-1}, so if such an $r \in J \subseteq I$ we would also have $1 = r^{-1}r \in I$ and so $I = RU^{-1}$, a contradiction).

We now let $S = J \cap (\bigcup X)$, and we claim that S is a selector for X. First assume that, for some $x \in X$, we have $y, z \in x \cap S$. Then $y, z \in J$, and so $y + z \in J \subseteq \bigcup_{p \in P} I_p$, so for some $p \in P$ we have $y + z \in I_p$, which is the ideal generated by the elements of p. This can easily be seen to imply that $y, z \in p$, with p a partial selector for X and $y, z \in x \in X$; the conclusion is that we must have $y = z$. Thus $|S \cap x| \leq 1$, for all $x \in X$. To see that actually $|S \cap x| = 1$, suppose not, that is, assume that for some $x \in X$, $S \cap x = \varnothing$. let $z \in x$ and consider the ideal J' of R

[5] Note that we do not need to use the Axiom of Choice for this part of the proof, since there are only finitely many $x \in F$.

[6] It is also possible to prove the stronger-looking statement that every ideal I of R can be extended to a maximal ideal containing it, without any additional difficulty: just find a maximal ideal on the quotient ring R/I, and its preimage under the canonical projection will be a maximal ideal of R containing I, by the correspondence theorem.

generated by J and z. Note that maximality of I implies that J must be maximal among all ideals of R that are included in $\bigcup_{p\in P} I_p$; we claim that J' is another such ideal. To see this, note that each element of J is of the form $r+az$, with $r\in J$ and $a\in R$. Then if $r\in I_p$, with $p\in P$, we can assume that $x\cap p=\varnothing$, and so we will have that $r+az\in I_{p\cup\{z\}}$, with $p\cup\{z\}\in P$. Thus $J'\subseteq\bigcup_{p\in P} I_p$, now $z\in J'\setminus J$ implies that $J\subsetneq J'$. This contradicts the maximality of J (because it contradicts the maximality of I), which shows that S is indeed a selector, and the proof is complete.

□

Our next equivalence, known as Tychonoff's theorem, relates the Axiom of Choice with topology, and it concerns products of topological spaces. It is the statement that every product of a family of compact spaces is compact.

Theorem 142. $\mathsf{ZF}\vdash\mathsf{AC}\iff\mathsf{TY}$.

Proof.

⇒: This proof really belongs in a course on General Topology, so over here we will only sketch it. Suppose that $\{X_i\big|i\in I\}$ is a bunch of compact topological spaces, and let $\mathscr{F}$ be a filter on $X=\prod_{i\in I} X_i$. Extend $\mathscr{F}$ to an ultrafilter u (this *partially requires* the Axiom of Choice, that is, we need to apply the Boolean Prime Ideal theorem in order to get that every filter can be extended to an ultrafilter), and consider the ultrafilters $\pi_i(u)$ (where $\pi_i:\prod_{i\in I} X_i\longrightarrow X_i$ is the i-th canonical projection), the Rudin–Keisler images of the ultrafilter u under the mappings π_i. Since each X_i is compact, each of the $\pi_i(u)$ converges to a point, choose such a point[7] x_i. Define $x=\langle x_i\big|i\in I\rangle\in X$, it is easy to see that x is an accumulation point of the filter $\mathscr{F}$, which proves that X is a compact topological space.

⇐: Let $\{A_\alpha\big|\alpha\in\Lambda\}$ be a family of nonempty sets, and pick a $z\notin\bigcup_{\alpha\in\Lambda} A_\alpha$. For each $\alpha\in\Lambda$, define $X_\alpha=A_\alpha\cup\{z\}$ and endow each X_α with the topology $\tau_\alpha=\{\varnothing,\{z\},X_\alpha\}$. Then each X_α is compact (though not Hausdorff; which is a remark that interacts well with the fact, noted in the previous footnote, that Tychonoff's theorem restricted to Hausdorff spaces is indeed strictly weaker than the Axiom of Choice), so by assumption the product $\prod_{\alpha\in\Lambda} X_\alpha$ is compact (this product is also nonempty, since the function constantly z lies there).

Now for each $\beta\in\Lambda$ define $F_\beta=\{f\in\prod_{\alpha\in\Lambda} X_\alpha\big|f(\beta)\neq z\}$. This set is closed and nonempty (since each $A_\beta\neq\varnothing$), and moreover the family $\{F_\beta\big|\beta\in\Lambda\}$ has the finite intersection property, hence by compactness there must be some element $f\in\bigcap_{\alpha\in\Lambda} F_\beta=\prod_{\alpha\in\Lambda} A_\alpha$.

□

We now proceed to prove that what we decided to call the *Disjointification Principle*, that basically states that every union can be thought of as a disjoint union, is equivalent to the Axiom of Choice.

Theorem 143. $\mathsf{ZF}\vdash\mathsf{AC}\iff\mathsf{DIS}$.

Proof.

⇒: Let $\{A_\alpha\big|\alpha\in\Lambda\}$ be an indexed family. The idea is that each $x\in\bigcup_{\alpha\in\Lambda} A_\alpha$ belongs to one, but possibly several, of the A_α, and so what we need to do is to choose one such α per each such x. To make this idea formal, define the indexed family $\{I_x\big|x\in\bigcup_{\alpha\in\Lambda} A_\alpha\}$ by

$$I_x=\{\alpha\in\Lambda\big|x\in A_\alpha\},$$

which is always nonempty because $x\in\bigcup_{\alpha\in\Lambda} A_\alpha$. Now appeal to the Axiom of Choice to obtain a choice function $f:\bigcup_{\alpha\in\Lambda} A_\alpha\longrightarrow\bigcup_{x\in\bigcup_{\alpha\in\Lambda} A_\alpha} I_x=\Lambda$, that is, f satisfies that $f(x)\in I_x$, or in other words, $x\in A_{f(x)}$, for all $x\in\bigcup_{\alpha\in\Lambda} A_\alpha$. Now for each $\alpha\in\Lambda$, define

$$B_\alpha=\{x\in A_\alpha\big|f(x)=\alpha\}.$$

First of all, we claim that the indexed family $\{B_\alpha\big|\alpha\in\Lambda\}$ is pairwise disjoint: if $\alpha,\beta\in\Lambda$ and $x\in B_\alpha\cap B_\beta$, it means that $\alpha=f(x)=\beta$. Also, since for each $\alpha\in\Lambda$ we have $B_\alpha\subseteq A_\alpha$, it must be the case that $\bigcup_{\alpha\in\Lambda} B_\alpha\subseteq\bigcup_{\alpha\in\Lambda} A_\alpha$; to show the other inclusion, let $x\in\bigcup_{\alpha\in\Lambda} A_\alpha$ and notice that $f(x)\in I_x=\{\alpha\in\Lambda\big|x\in A_\alpha\}$, which means that $x\in B_{f(x)}$ by definition, and so $x\in\bigcup_{\alpha\in\Lambda} B_\alpha$. Therefore $\bigcup_{\alpha\in\Lambda} A_\alpha=\bigcup_{\alpha\in\Lambda} B_\alpha$, and we are done.

[7] This choice of a point of convergence is the place where we are actually using the full Axiom of Choice. If we had assumed that each X_i was a Hausdorff space, the point of convergence x_i would have been unique, and therefore we would not have needed to use the Axiom of Choice at this point. Hence the version of Tychonoff's theorem that is restricted to Hausdorff topological spaces is no longer equivalent to the full Axiom of Choice; in fact, such a restricted version is only equivalent to the Boolean Prime Ideal theorem.

$\Leftarrow$**:** Let X be a pairwise disjoint family of nonempty sets. Define an indexed family $\{A_z \big| z \in \bigcup X\}$ by

$$A_z = \{x \in X \big| z \in x\},$$

in other words, A_z will be the singleton containing the unique $x \in X$ such that $z \in x$ (this element is unique because we are assuming the family X to be pairwise disjoint). Using the disjointification principle, let $\{B_z \big| z \in \bigcup X\}$ be an indexed family satisfying $(\forall z \in \bigcup X)(B_z \subseteq A_z)$ (thus each B_z is either empty, or a singleton) and $\bigcup_{z\in\bigcup X} B_z = \bigcup_{z\in\bigcup X} A_z$. Define

$$S = \{z \in \bigcup X \big| B_z \neq \varnothing\}.$$

We claim that S is a selector for the family X. To see this, take an arbitrary $x \in X$, and let us show that $S \cap x$ consists of exactly one element. Suppose first that $y, z \in S \cap x$. This means that $B_y \neq \varnothing \neq B_z$; so that B_y and B_z are both singletons and, furthermore, since $y, z \in x$, we may conclude that $B_y = \{x\} = B_z$. Since the Bs are pairwise disjoint, the conclusion must be that $y = z$, and so $S \cap x$ has at most one element. Now to see that $S \cap x$ has at least one element, note that, since x is nonempty, there is at least one $y \in x$, which means that $x \in \{x\} = A_y \subseteq \bigcup_{z\in\bigcup X} A_z = \bigcup_{z\in\bigcup X} B_z$. Therefore $x \in B_z$ for some z, which implies that $B_z \neq \varnothing$, so $z \in S$ and B_z is a singleton; since $x \in B_z \subseteq A_z$, we may conclude that $A_z = \{x\}$ and so $z \in x$. Thus $x \cap S = \{z\}$; the conclusion now is that S is a selector for the family X, and we are done[8].

□

6.2 Cardinal numbers after the Axiom of Choice (two more equivalences)

We have seen above that, assuming the Axiom of Choice (equivalent to WOP), every set can be endowed with a well-order relation. So if X is an arbitrary set, and we pick a well-order relation $\leq$ on X, then there will be an $\alpha \in \mathbf{Ord}$ such that $(X, \leq) \cong (\alpha, \in)$. The witnessing isomorphism is, in particular, a bijection between X and α; we have thus proved that under the Axiom of Choice every set can be put in bijection with (is equipotent to) some ordinal. This motivates the following definition.

Definition 144.

1. For every set X, we define $|X| = \min\{\alpha \in \mathbf{Ord} \big| X \text{ is equipotent to } \alpha\}$.
2. An ordinal κ will be said to be a **cardinal number**, or an **initial ordinal**, if there exists some X such that $\kappa = |X|$.

The term "initial ordinal" is motivated by the fact that an ordinal number κ is a cardinal if and only if it cannot be put in bijection with any ordinal $\alpha < \kappa$. For if κ is not in bijection with any $\alpha < \kappa$, then clearly $\kappa = \min\{\xi \in \mathbf{Ord} \big| \kappa \text{ is equipotent to } \xi\} = |\kappa|$; conversely, if $\kappa = |X|$ then it is not possible to have some $\alpha < \kappa$ be equipotent with κ, since this would mean that X is equipotent to α, contradicting the minimality of κ.

Recall that sequence of omegas was defined recursively by letting $\omega_0 = \omega$, declaring $\omega_{\alpha+1}$ to be the least ordinal $\geq \omega_\alpha$ that is not equipotent to ω_α (equivalently, the least ordinal that does not inject into ω_α, which exists because of Hartogs's theorem), and finally defining $\omega_\alpha = \sup_{\xi<\alpha} \omega_\xi$ for limit ordinals α. It turns out that this sequence consists of cardinal numbers and, moreover, it exhausts most of the cardinals that exist on the ordinal number line.

Proposition 145. *An ordinal κ is a cardinal number if and only if it is either finite, or equal to ω_α for some α.*

Proof.

$\Leftarrow$**:** We already know that, if $n < \omega$ and $m < n$, it is not possible to have a bijection between m and n. We also know that there is no bijection between ω and any finite ordinal $n < \omega$; this shows that all finite ordinals, as well as ω_0, are cardinal numbers. Now consider an element of the omega sequence with a successor index, $\omega_{\alpha+1}$. If there was a bijection between $\omega_{\alpha+1}$ and some $\xi < \omega_{\alpha+1}$, composing this bijection with an injection $: \xi \longrightarrow \omega_\alpha$ (which must exist by the definition of $\omega_{\alpha+1}$) we would obtain an injection from $\omega_{\alpha+1}$ into ω_α, a contradiction.

Finally, consider an element of the omega sequence with a limit index, ω_α, $\alpha = \bigcup \alpha$. Suppose that ω_α was equipotent to some $\beta < \omega_\alpha = \sup_{\xi<\alpha} \omega_\xi$, then we would have $\beta < \omega_\xi$ for some $\xi < \alpha$. Composing this bijection between ω_α and β with the inclusion mapping $: \beta \longrightarrow \omega_\xi$ yields an injection from ω_α to ω_ξ, which would imply that $\omega_\alpha < \omega_{\xi+1} < \omega_\alpha$, a contradiction. Therefore we have now proved that each of the ω_α is a cardinal number, regardless of the nature of the index α.

[8] Another way of doing this proof, obtaining a choice function rather than a selector, is as follows: let X be a family of nonempty sets. For each $z \in \bigcup X$ let $A_z = \{x \in X \big| z \in x\}$; this need no longer be a singleton, because we are no longer assuming that the family X is pairwise disjoint. Now use the Disjointification Principle to get pairwise disjoint $B_z \subseteq A_z$ such that $\bigcup_{z\in\bigcup X} B_z = \bigcup_{z\in\bigcup X} A_z = X$; and for each $x \in X$, define $f(x)$ to be the unique $z \in \bigcup X$ such that $x \in B_z$ (such a z is unique because the B_z are pairwise disjoint). Then we have $f : X \longrightarrow \bigcup X$, and it is easy to check that f is a choice function (because since $x \in B_{f(x)} \subseteq A_{f(x)}$, it is the case that $f(x) \in x$).

⇒: Let $\kappa \in \mathbf{Ord}$ be a cardinal number, and assume that κ is not finite, that is, $\omega \leq \kappa$. As a side remark, note that for each $\alpha \in \mathbf{Ord}$ it is the case that $\omega_\alpha \geq \alpha$ (this can be proved easily by transfinite induction on α), and so in particular we have that $\kappa \leq \omega_\kappa < \omega_{\kappa+1}$. This shows that $\kappa < \omega_\alpha$ for some $\alpha \in \mathbf{Ord}$, and we can choose the least such α. We claim that α is a successor ordinal; otherwise we would have that $\kappa \geq \omega_\xi$ for all $\xi < \alpha$ and this would imply that $\kappa \geq \sup_{\xi<\alpha} \omega_\xi = \omega_\alpha$; a contradiction. Thus we may write $\alpha = \xi + 1$, and by our choice of $\alpha = \xi + 1$ we know that $\omega_\xi \leq \kappa < \omega_{\xi+1}$. By definition of $\omega_{\xi+1}$, this means that κ is equipotent to ω_ξ; the fact that κ is an initial ordinal prevents $\omega_\xi < \kappa$ from happening, and so the only possibility left is that $\kappa = \omega_\xi$. This finishes the proof.

□

Thus, (under the Axiom of Choice) the cardinality of every infinite set is equal to some ω_α. Remembering that we introduced the symbols $\aleph_\alpha$ to be "abbreviations" of $|\omega_\alpha|$, we now see that the current situation is that $\omega_\alpha = \aleph_\alpha$ for all $\alpha \in \mathbf{Ord}$. Hence we have two different symbols to denote the same mathematical entity; what some people do in this situation is to write ω_α when "they are thinking of the entity as an ordinal" and to correspondingly write $\aleph_\alpha$ when "they are thinking of the entity as a cardinal". In my humble opinion this is, of course, none other than a perfectly fine sample of the most exquisite nonsense, and therefore in this course I will mostly stick to writing only ω_α, and not using the symbol $\aleph$ unless it is within an expression where the consistent usage of ω would lead to unbearably cumbersome notational repetitiveness[9].

After this discussion of the omega sequence and how it contains representatives for all infinite cardinalities, we now know that cardinal numbers are linearly ordered (in fact, well-ordered) under the Axiom of Choice[10]. A very surprising fact is that linear orderedness of cardinal numbers is not only a consequence of, but in fact, equivalent to, the Axiom of Choice.

Theorem 146. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{IC}$.

Proof.

⇒ Under AC, given any two sets A, B, both are equipotent to some initial ordinal, say $|A| = \kappa$ and $|B| = \lambda$. Since the ordinal line is linearly ordered, either $\kappa \leq \lambda$ or $\lambda \leq \kappa$; either way, a three-fold composition of the two given bijections with the corresponding inclusion mapping yields an injection, either from X to Y or from Y to X.

⇐ Given any set A, use Hartogs's theorem to obtain an $\alpha \in \mathbf{Ord}$ that cannot inject into A. By our assumption, either A injects into α, or α injects into A; our choice of α prevents the second case from holding any water, and therefore there must be an injection from A into α. In particular, we can copy into A the well-order that the range of such injection inherits from α. This provides us with a well-ordering for the (arbitrary) set A, so we have proved that WOP holds. Hence the Axiom of Choice holds.

□

We will now start to think about a "dual" way of describing cardinal inequalities. Our official definition stated originally that $|A| \leq |B|$ means there is an injection $f : A \longrightarrow B$. We could have defined $|A| \leq |B|$ by saying instead that there is a surjection $g : B \longrightarrow A$; interestingly, these two definitions need you[11] to assume the Axiom of Choice in order for them to be equivalent.

Theorem 147. *If A and B are two sets, then $|A| \leq |B|$ iff there exists a surjective $g : B \longrightarrow A$.*

Proof.

⇒: If $f : A \longrightarrow B$ is injective, we know we can find a left inverse $g : B \longrightarrow A$ for it. That is, $g \circ f = \mathrm{id} \restriction A$. This implies that g is surjective, for if $a \in A$, then $a = g(f(a)) \in \mathrm{ran}(g)$.

⇐: If $g : B \longrightarrow A$ is surjective, the Axiom of Choice (since it implies ES) provides us with a right inverse $f : A \longrightarrow B$. That is, $g \circ f = \mathrm{id} \restriction A$, which implies that f is injective, since every time we have $a, b \in A$ with $f(a) = f(b)$, applying g on both sides of this equation yields $a = g(f(a)) = g(f(b)) = b$. Thus $|A| \leq |B|$.

□

[9] For example, it is possible (though not very likely) that $\aleph_{\omega_\omega}$ looks somewhat more aesthetically appealing than ω_{ω_ω}.

[10] In fact, assuming the Axiom of Choice we can avoid the need for proving the Cantor–Schroeder–Bernstein theorem in the complicated way in which we did this in Theorem 124. For, if we assume the Axiom of Choice, then every two sets X, Y must satisfy that $|X| = \omega_\alpha$ and $|Y| = \omega_\beta$ for some $\alpha, \beta \in \mathbf{Ord}$. Thus if $|X| \leq |Y|$ and $|Y| \leq |X|$, it means that $\omega_\alpha \leq \omega_\beta$ and $\omega_\beta \leq \omega_\alpha$; thus $\omega_\alpha = \omega_\beta$ and so $|X| = \omega_\alpha = \omega_\beta = |Y|$. This is Gödel's proof of the Cantor–Schroeder–Bernstein theorem, and one of my favourite ones in that category.

[11] Yes, YOU, the reader!

The principle of Injective Comparability, which appeared earlier, has that name because it states that every two cardinals are comparable according to the definition of cardinal inequality that uses injections. Dually, there should be a principle of Surjective Comparability, stating that any two cardinals are comparable according to the definition of cardinal inequality that uses surjections. Since the equivalence of both definitions of cardinal inequality requires the Axiom of Choice, it is not clear a priory that Injective Comparability should be equivalent to Surjective Comparability over ZF. Below we see that, nevertheless, these two principles are in fact equivalent (and also equivalent to the Axiom of Choice).

Theorem 148. $\mathsf{ZF} \vdash \mathsf{AC} \iff \mathsf{SC}$.

Proof.

$\Rightarrow$**:** Under AC, given any two sets A, B, we have that either $|A| \leq |B|$ or $|B| \leq |A|$. Also assuming AC, the two definitions for when a cardinal is less than or equal to the other coincide, and therefore what we said in the previous sentence actually means that either there is a surjection $g : B \longrightarrow A$, or there is a surjection $g : A \longrightarrow B$. This finishes the proof.

$\Leftarrow$**:** Before we can actually carry out the proof of this implication, we need a little lemma: we need to prove the surjective version of Hartogs's theorem in ZF.

Lemma 149. *Working in the axiom system* ZF*, for every set* X *there exists an ordinal* α *such that there is no surjection* $g : X \longrightarrow \alpha$ *(that is, there is* α *such that it is not the case that* $|\alpha| \leq |X|$ *according to the surjective definition).*

Proof. Let X be an arbitrary set, and let

$$W = \{(P, \leq) \in \mathfrak{P}(\mathfrak{P}(X)) \times \mathfrak{P}(\mathfrak{P}(X) \times \mathfrak{P}(X)) \big| P \text{ is a partition of } X \wedge \leq \text{ well-orders } P\},$$

which is a set by the Axiom of Comprehension. Letting $Z = \{\operatorname{otp}(P, \leq) \big| (P, \leq) \in W\}$, which is a set by the Axiom of Replacement, we claim that in fact $Z = \{\alpha \in \mathbf{Ord} \big| (\exists g : X \longrightarrow \alpha)(g \text{ is surjective})\}$. This is because every surjection $g : X \longrightarrow \alpha$ determines a bijection $\hat{g} : \{g^{-1}[\{\xi\}] \big| \xi < \alpha\} \longrightarrow \alpha$ (given by $\hat{g}(g^{-1}[\{\xi\}]) = \xi$), where $\{g^{-1}[\{\xi\}] \big| \xi < \alpha\}$ is always a partition of X. Hence every surjection $: X \longrightarrow \alpha$ allows us to well-order the corresponding partition of X that arises (by copying the well-ordered structure of α into said partition); conversely, every well-ordered structure on a partition P of X has some order-type α, and this induces a corresponding obvious surjection $: X \longrightarrow \alpha$ (just map each $x \in X$ to the image of p under the order-type isomorphism, where p is the unique member of P satisfying $x \in p$).

Thus, since Z is a set, it cannot contain the whole class $\mathbf{Ord}$, so we can find an ordinal number $\alpha \notin Z$, that is, an α onto which X cannot surject[12]. $\square_{\text{Lemma}}$

With this lemma under our belt, let X be an arbitrary set, and pick an ordinal number α such that X cannot surject onto α. Since we are assuming surjective comparability, it must be the case that α surjects onto X, so let $g : \alpha \longrightarrow X$ be such a surjection. Now let $f : X \longrightarrow \alpha$ be defined by

$$f(x) = \min\{\xi < \alpha \big| g(\xi) = x\},$$

which is defined for all $x \in X$ because g is surjective. Notice that, by definition, we must have $g(f(x)) = x$ for all $x \in X$, which implies that f is injective (if $f(x) = f(y)$ then $x = g(f(x)) = g(f(y)) = y$). Thus we have injected our set X into some ordinal number, so, in the same way as described in the proof of Theorem 146, this implies that we can endow X with a well-ordering (induced by f and the well-ordering of $\operatorname{ran}(f) \subseteq \alpha$). We have thus proved that every set can be well-ordered; since we know that $\mathsf{WOP} \Rightarrow \mathsf{AC}$, this means that the Axiom of Choice holds.

$\square$

6.3 The Banach-Tarski paradox

After encountering some very pleasant consequences of the Axiom of Choice (for example, cardinalities are linearly ordered, and every epi is a split epi in the category of sets), we feel the moral obligation to point out that some people are not entirely happy with the Axiom of Choice, due to some other consequences that this axiom has, which might be perceived as "counterintuitive". We will devote this section to explain the statement of one such counterintuitive consequence, the

[12] In fact, it can be proved that $\xi < \alpha \in Z \Rightarrow \xi \in Z$, and so Z itself is an ordinal number, which happens to be the least ordinal onto which X cannot surject. But we do not need to spell this extra-information out in detail to have our theorem.

so-called Banach-Tarski theorem. There are no proofs whatsoever in this section, as we are only aiming to state the theorem for the reader's contemplative recreation.

The first step for understanding the statement of Banach–Tarski is to know about Lebesque measure in three-dimensional space. In order to build towards that goal, we beging by explaining the one-dimensional Lebesque measure, which can be viewed as a generalization of the concept of *length.* That is, we want to be able to measure the length of an arbitrary subset of $\mathbb{R}$. We know what the length should be if our set was an interval, so we start by declaring that

$$\mu((a,b)) = \mu([a,b]) = b - a,$$

(we will denote such "generalized length" with the symbol μ). We also note that it would be desirable to have

$$\mu(\varnothing) = 0.$$

If A and B are disjoint, then we expect that $\mu(A \cup B) = \mu(A) + \mu(B)$; in fact, we would like that, if we have a pairwise disjoint countable collection of sets A_n, $n < \omega$, then

$$\mu\left(\bigcup_{n\in\omega} A_n\right) = \sum_{n=0}^{\infty} \mu(A_n).$$

Consider the following examples:

1. $\mu([0,1] \cup (2,3)) = 2.$
2. $\mu\left(\bigcup_{n\in\omega}[n, n + \frac{1}{2^n})\right) = \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{1}{1-\frac{1}{2}} = \frac{1}{\frac{2-1}{2}} = 2.$
3. On the other hand, $\mu\left(\bigcup_{n\in\omega}[n, n + \frac{1}{2})\right) = \sum_{n=0}^{\infty} \frac{1}{2} = \infty.$

How can we possibly define such a "generalized length" (or measure) on an arbitrary $X \subseteq \mathbb{R}$? The standard move is to (concentrate on $[0,1]$ and), given an X, define

$$\mu^*(X) = \inf\left\{\sum_{n\in\omega} \mu(I_n) \middle| (\forall n < \omega)(I_n \text{ is an interval}) \wedge X \subseteq \bigcup_{n\in\omega} I_n\right\},$$

(this is known as the *exterior measure* of the set X), and let $\mu_*(X) = 1 - \mu^*([0,1] \setminus X)$ (the *interior measure* of X). It turns out that for many, many sets X (all closed sets, all open sets, and more –technically, at least for all *Borel* sets–), we can get lucky, and it will be the case that $\mu^*(X) = \mu_*(X)$, in which case we define $\mu(X) = \mu^*(X) = \mu_*(X)$ and we will say that X is **measurable**. The measure μ has all the desirable properties that we mentioned above, as long as we restrict ourselves to measurable sets. It turns out that the Axiom of Choice allows us to prove the following theorem, stating that this restriction is indeed a nontrivial one:

Theorem 150 (Vitali)**.** *There are non-measurable sets.*

The previous theorem is known to actually require the Axiom of Choice to be proved (although it is not equivalent to the Axiom of Choice, but strictly weaker). This was proved by Solovay.

Theorem 151 (Solovay)**.** *(If there exists an inaccessible cardinal then) there exists a model of* ZF *where every set is measurable.*

(If we Lévy-collapse an inaccessible cardinal, the class **HOD** computed in the generic extension satisfies ZF and DC, and it also satisfies that every set of reals is Lebesque measurable, has the perfect set property, the Baire property, and basically any other regularity property currently in existence. The reader that is not familiar with any of these terms is strongly encouraged to summarily ignore this parenthetical remark.)

Lebesgue measure can also be defined in the two-dimensional space $\mathbb{R}^2$ (in which case we think of it as a "generalized area"), as well as in three-dimensional space $\mathbb{R}^3$ (in which case it would be more of a "generalized volume"). The definitions of exterior and interior measure, of measurable, and of the measure of a measurable set, are exactly as in the one-dimensional case. The only difference can be found in the initial step of the construction, where our basic subsets are, in the two-dimensional case, rectangles $[a,b] \times [c,d]$ instead of intervals, and we declare that

$$\mu([a,b] \times [c,d]) = (b-a)(d-c);$$

and in the three-dimensional case we start by considering parallelepipeds $[a,b] \times [c,d] \times [e,f]$ instead of intervals or rectangles, stipulating that

$$\mu([a,b] \times [c,d] \times [e,f]) = (b-a)(d-c)(f-e).$$

Remark 152. Lebesque measure is geometrically well-behaved, in the sense that, if $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ is an isometry, and $X \subseteq \mathbb{R}^3$ is a measurable set, then so is $f[X]$, and moreover $\mu(f[X]) = \mu(X)$.

Theorem 153 (Banach–Tarski)**.** *Let $B = \{\vec{x} \in \mathbb{R}^3 \big| \|\vec{x}\| \leq 1\}$ be the unit ball in $\mathbb{R}^3$. Then there is a partition $\{X_1, \dots, X_5\}$ (i.e. the X_i are disjoint and $B = X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5$) and isometries $f_i : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$, for $i \in \{1, 2, 3, 4, 5\}$, such that*

$$B = f_1[X_1] \cup f_2[X_2] \cup f_3[X_3]$$

(with $f_1[X_1], f_2[X_2], f_3[X_3]$ pairwise disjoint), and

$$B = f_4[X_4] \cup f_5[X_5]$$

(with $f_4[X_4], f_5[X_5]$ disjoint).

Think very carefully about the theorem above. Recalling the formula for the volume of a sphere, we must have that

$$\begin{aligned} \frac{4}{3}\pi = \mu(B) &= \mu\left(f_1[X_1] \cup f_2[X_2] \cup f_3[X_3]\right) \\ &= \mu(f_1[X_1]) + \mu(f_2[X_2]) + \mu(f_3[X_3]) = \mu(X_1) + \mu(X_2) + \mu(X_3); \end{aligned}$$

on the other hand,

$$\begin{aligned} \frac{4}{3}\pi = \mu(B) &= \mu\left(f_4[X_4] \cup f_5[X_5]\right) \\ &= \mu(f_4[X_4]) + \mu(f_5[X_5]) = \mu(X_4) + \mu(X_5). \end{aligned}$$

Putting together these two equations, we see that

$$\begin{aligned} \frac{4}{3}\pi &= \mu(B) = \mu\left(X_1 \cup X_2 \cup X_3 \cup X_4 \cup X_5\right) \\ &= \mu(X_1) + \mu(X_2) + \mu(X_3) + \mu(X_4) + \mu(X_5) \\ &= (\mu(X_1) + \mu(X_2) + \mu(X_3)) + (\mu(X_4) + \mu(X_5)) = \frac{4}{3}\pi + \frac{4}{3}\pi \\ &= \frac{8}{3}\pi, \end{aligned}$$

so that $\frac{8}{3}\pi = \frac{4}{3}\pi$, and therefore $2 = 1$. What's going on?

(Seriously, reader: spend some time staring at this.)

What is going on is that we assumed that each of the X_i is Lebesque measurable. In fact, this is not the case, which explains why the Banach–Tarski "paradox" is not actually a contradiction in ZF. In particular, this theorem requires the existence of non-measurable sets, which tells us that the Axiom of Choice is instrumental for the proof of the theorem. That is why some people argue that the Banach–Tarski paradox shows how counterintuitive the Axiom of Choice can be.

6.4 Dedekind-finite sets

In order to provide some balance against the previous diatribe, we will now devote a section to explaining a few rather unpleasant consequences of not assuming the Axiom of Choice.

An exercise from the Axiom of Choice worksheet consists in proving countable unions of countable sets are countable, the punchline being that the Axiom of Choice is needed in order to carry out this proof. The reader might find this a bit unsettling. Even more unsettling is the fact that it is possible, in ZF (that is, provided we do not assume the Axiom of Choice), to have that $\mathbb{R}$ is the countable union of countable sets. Another frightening possibility, should we reject the Axiom of Choice, is that we can partition the real line $\mathbb{R}$ into a number of pieces which is not less than or equal to the cardinality of the continuum. These are all statements known to be consistent with ZF, so they necessarily remain very real possibilities unless we assume the Axiom of Choice. We will utilize most of the section to mention another extremely unpleasant consequence of dropping the Axiom of Choice: namely, that we might not even have a proper definition of what an infinite set should be.

Recall that we defined a set X to be finite if $|X| = n$ for some $n < \omega$, and infinite otherwise. We will begin to explore Dedekind's definition of infinite set, which differs from the one that we just recalled.

Definition 154. A set X is said to be **Dedekind-infinite** if there exists a proper subset $Y \subsetneq X$ which is equipotent with X. Equivalently, X is Dedekind-infinite if there exists an $f : X \longrightarrow X$ which is injective but not surjective.

This reflects the old observation of Galileo that some infinite sets are equinumerous with proper subsets (case in point: $\mathbb{N}$ can be put in bijection with the set of perfect squares; for another example, note that the exponential function constitutes a bijection from $\mathbb{R}$ to $(0, \infty)$). It also vindicates Jorge Luis Borges, who wrote (in "El Aleph"), about the symbol Aleph, that

> para la *Mengenlehre*[13], es el símbolo de los números transfinitos, en los que el todo no es mayor que alguna de sus partes.

It is clear that every Dedekind-infinite set must be infinite (the latter occurence of the word "infinite" refers to the first definition of infinity, namely, not equipotent with any $n \in \omega$), since we have already proved that for every $n < \omega$ it is the case that every injection $f : n \longrightarrow n$ must be a surjection; note that both the "usual" notion of finite/infinite, as well as Dedekind's one, are invariant under equipotence. Analysing the converse of this statement requires us to be extremely careful.

Theorem 155. *In* ZF*, the following are equivalent for every set X:*

1. *X is Dedekind-infinite,*
2. *there is an injection $f : \mathbb{N} \longrightarrow X$ (i.e. $\aleph_0 \leq |X|$),*
3. *there is a subset $A \subseteq X$ such that $A = \aleph_0$.*

Proof.

(2) $\iff$ (3) Obvious.

(2,3)$\Rightarrow$(1) If $A = \{a_n \big| n \in \omega\} \subseteq X$, then we can define $f : X \longrightarrow X$ by

$$f(x) = \begin{cases} x \text{ if } x \notin A, \\ a_{2n} \text{ if } x = a_n \in A. \end{cases}$$

Clearly f is injective but not surjective.

(1)$\Rightarrow$(2,3) Let $f : X \longrightarrow X$ be injective but not surjective. Using the recursion theorem, define $g : \omega \longrightarrow X$ by letting $g(0) \in X \setminus \operatorname{ran}(f)$ and $g(n+1) = f(g(n))$. The fact that $g(0) \notin \operatorname{ran}(f)$ and f is injective implies that g is an injective function. This means that $A = \operatorname{ran}(g) = \{g(n) \big| n \in \mathbb{N}\}$ is a countably infinite subset of X.

□

Notice that, under ZF, every infinite well-ordered set is Dedekind infinite (since every such set is in bijection with some ordinal $\alpha \geq \omega$, and the image of ω under such bijection constitutes a countably infinite subset of the given set). However, there are models of ZF where there are sets X that are infinite, yet Dedekind-finite. That is, X has no countable subsets, even though it cannot be put in bijection with any $n < \omega$. This can be reinterpreted as follows: an infinite Dedekind-finite set X satisfies that, even though $|X| > n$ for all $n \in \omega$ (infinite), it is not the case that $\aleph_0 \leq |X|$ (Dedekind-finite). In other words, we have a cardinality that is not comparable with $\aleph_0$. We have seen before that the Axiom of Choice is equivalent to the linear orderedness of cardinalities; now we get to see a concrete example of how the latter might fail, should we refrain from assuming the former.

We close this section by cleaning up our mess. To do this, we will show that, once we assume the Axiom of Choice, everything finally falls back into its right place, and infinite sets behave nicely once again.

Theorem 156. *In* ZFC *(in fact, in* $\mathsf{ZF} + \mathsf{AC}(\omega)$*), every infinite set is Dedekind infinite.*

Proof. The use of the Axiom of Choice pretty much trivializes the issue: Let X be infinite, let f be a choice function on $\mathfrak{P}(X) \setminus \{\varnothing\}$, and use the recursion theorem to define a function g given by $g(n) = f(X \setminus \operatorname{ran}(g \restriction n))$. This construction[14] can be carried out because, by definition, $g \restriction n$ is injective for each $n < \omega$; and therefore $X \setminus \operatorname{ran}(g \restriction n)$ is always nonempty (otherwise X would be finite). This way we obtain an injective function $g : \omega \longrightarrow X$, and we are done. □

[13] Which means "Set Theory" in German. Why did he wrote the German word instead of the Spanish "Teoría de Conjuntos", is one of the most mysterious facts of this Universe (there is probably only one other fact even more mysterious, namely, that there is something rather than nothing).

[14] Intuitively, this construction just amounts to choosing elements of X one by one, each choice justified by the fact that we have only made finitely many choices so far, and so we still have stuff to choose from because X is assumed to be infinite.

6.5 Some more cardinal arithmetic

We are now back to the world where we shamelessly assume the Axiom of Choice. It turns out that there is still much more that can be proved about cardinal numbers and their arithmetic, so we proceed to do just that. Recall that, once we assume the Axiom of Choice, initial ordinals constitute a complete class of representatives for all cardinalities. In what follows, we use letters from the middle of the Greek alphabet (such as κ, λ, or μ) to denote cardinal numbers (which are now initial ordinals). When doing cardinal arithmetic, we will occasionally use the $\aleph$ notation (not so much because we wish to do so, but rather for the benefit of the reader).

Theorem 157. *If κ is an infinite cardinal, then $\kappa + \aleph_0 = \kappa$.*

Proof. Since κ is infinite, we have $\omega \leq \kappa$. Let $Y = \{y_n \big| n < \omega\}$ be a countable set disjoint from κ. It is clear that the function $f : \kappa \cup Y \longrightarrow \kappa$ given by

$$f(x) = \begin{cases} 2n+1 \text{ if } x = y_n \in Y, \\ 2n \text{ if } x = n < \omega, \\ x \text{ if } x \in \kappa \setminus \omega \end{cases}$$

is a bijection, witnessing that $\kappa + \aleph_0 = |\kappa \cup Y| = \kappa$. □

We can use this theorem to prove that there are as many irrational numbers as there are real numbers.

Corollary 158. $|\mathbb{R} \setminus \mathbb{Q}| = \mathfrak{c}$*, in particular, there is at least one irrational number.*

Proof. Notice that $\mathfrak{c} = |\mathbb{R}| = |\mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q})| = |\mathbb{Q}| + |\mathbb{R} \setminus \mathbb{Q}| = \aleph_0 + |\mathbb{R} \setminus \mathbb{Q}| = |\mathbb{R} \setminus \mathbb{Q}|$ (for the last equality, note that $|\mathbb{R} \setminus \mathbb{Q}|$ cannot be finite, for otherwise the second to last link of this chain of equations would equal $\aleph_0$, a contradiction; hence $|\mathbb{R} \setminus \mathbb{Q}|$ must be infinite and so we can apply Theorem 157), and we are done. □

We would like to generalize this result so that we know what the result of adding two infinite cardinals is, regardless of whether one of them equals $\aleph_0$ or not. In order to do that, it will turn out that the following really interesting well-order relation is quite useful.

6.6 The canonical well-ordering of Ord × Ord and its consequences

Definition 159. We proceed to define the following (binary relation, which can easily be checked to be a) strict partial order in the class **Ord** × **Ord**:

$(\xi_1, \xi_2) \lhd (\eta_1, \eta_2)$ iff $\max\{\xi_1, \xi_2\} < \max\{\eta_1, \eta_2\}$, or $\max\{\xi_1, \xi_2\} = \max\{\eta_1, \eta_2\}$ and $\xi_1 < \eta_1$, or $\max\{\xi_1, \xi_2\} = \max\{\eta_1, \eta_2\}$ and $\xi_1 = \eta_1$ and $\xi_2 < \eta_2$.

Remark 160. The relation $\lhd$ is in fact a well-ordering: if $X \subseteq \mathbf{Ord} \times \mathbf{Ord}$, we can let $\alpha = \min\{\max\{\xi_1, \xi_2\} \big| (\xi_1, \xi_2) \in X\}$, $\beta = \min\{\xi \big| (\exists \xi_2)(\xi, \xi_2) \in X \wedge \max\{\xi, \xi_2\} = \alpha)\}$, and $\gamma = \min\{\xi \big| (\beta, \xi) \in X \wedge \max\{\beta, \xi\} = \alpha\}$. It is fairly straightforward to see that (β, γ) is the $\lhd$-minimum of X.

Recall that an initial ordinal κ is characterized by the fact that every $\xi < \kappa$ satisfies $|\xi| < \kappa$. Recall also that every infinite initial ordinal is equal to ω_α, for some α. This will allow us to prove the following theorem.

Theorem 161. *If κ is an infinite cardinal, then $\kappa\kappa = \kappa$ (that is, $|\kappa \times \kappa| = \kappa$).*

Proof. Let $\kappa = \omega_\alpha$. The proof will be done by induction on α. Well-order $\omega_\alpha \times \omega_\alpha$ by letting it inherit the well-order relation $\lhd$ from **Ord** × **Ord**, and let $\gamma = \text{otp}_\lhd(\omega_\alpha \times \omega_\alpha)$. Since clearly $|\omega_\alpha| \leq |\omega_\alpha \times \omega_\alpha|$, it follows that $\omega_\alpha \leq \gamma$. We will now prove that every initial segment of $\omega_\alpha \times \omega_\alpha$ has cardinality $< \omega_\alpha$, which will evidence that γ must actually equal $\gamma = \omega_\alpha$. This will show that, in particular, there is a bijection between $\omega_\alpha \times \omega_\alpha$ and ω_α.

So let $(\xi_1, \xi_2) \in \omega_\alpha \times \omega_\alpha$, let $\xi = \max\{\xi_1, \xi_2\}$, and let $S = \text{seg}_\lhd(\xi_1, \xi_2)$. Then $\xi + 1 < \omega_\alpha$, and the ordering $\lhd$ is such that $S \subseteq (\xi + 1) \times (\xi + 1)$. Now we actually run the induction:

- In the base case, if $\alpha = 0$, then $\xi + 1$ is finite and so $S \subseteq (\xi + 1) \times (\xi + 1)$ is a finite set, in particular, $|S| < \omega$.
- In the inductive case, since $\xi + 1 < \omega_\alpha$ it follows that $|\xi + 1| = \omega_\beta$ for some $\beta < \alpha$. Then, by induction hypothesis, we have $|(\xi + 1) \times (\xi + 1)| = |\omega_\beta \times \omega_\beta| = \omega_\beta$, and so $|S| \leq \omega_\beta < \omega_\alpha$.

Hence, as was explained above, $\gamma = \omega_\alpha$, and our proof is complete. □

Corollary 162. *If κ and λ are cardinals, at least one of them infinite, then*

$$\kappa + \lambda = \kappa\lambda = \max\{\kappa, \lambda\}.$$

Proof. $\max\{\kappa,\lambda\} \leq \kappa+\lambda \leq \kappa\lambda \leq \max\{\kappa,\lambda\}\max\{\kappa,\lambda\} = \max\{\kappa,\lambda\}$. □

This greatly simplifies the computations involved in cardinal arithmetic. For example,

Example 163. Perform the following operations:

1. $\omega_{17}+\omega_{23}$
2. $\omega_{28}\omega_{54}$
3. $\omega_{40}+\omega_{2018}$
4. $\omega_3\omega_{10^6}$
5. $\mathfrak{c}+\omega_1$

Another consequence of Theorem 161 (I would almost make it a corollary, but it's too simple) is that for every infinite set X and every finite nonzero $n \in \mathbb{N}$, we have $|X^n| = |X|$ (this is proved by induction, using the fact that there is a bijection between $X \times X$ and X). In fact, just assuming this statement for $n = 2$ is already quite strong, as such a statement turns out to be equivalent to the Axiom of Choice. This result, which we now proceed to prove, is known as Tarski's Theorem. When Tarski sent it to publication to *Comptes Rendus de l'Academie des Sciences*, Fréchet argued that an implication between two well-known propositions is not a new result, whereas Lebesque argued that an implication between two false propositions is of no interest.

Theorem 164 (Tarski)**.** $\mathsf{ZF} \vdash \mathsf{AC} \iff (\forall X)(X \textit{ is infinite} \Rightarrow |X \times X| = |X|)$.

Proof.

⇒: This is just Theorem 161.

⇐: Let X be an infinite set and let $\alpha = X^+$ be the Hartogs number of X. We will inject X into α, showing that X is well-orderable. Assume without loss of generality that $X \cap \alpha = \varnothing$. Our hypothesis is that there exists a bijection $f : X \cup \alpha \longrightarrow (X \cup \alpha) \times (X \cup \alpha)$.

In order to define an injection from X into α, we let $Y_x = \{\xi < \alpha \big| f(\xi) \in \alpha \times \{x\}\}$ and define $g : X \longrightarrow \alpha$ by $g(x) = \min(Y_x)$. If this function is well-defined (meaning, if $Y_x \neq \varnothing$ for all $x \in X$), then it should clearly be injective, and we will be done. So fix $x \in X$, and suppose that $Y_x = \varnothing$. This means that for every $\xi < \alpha$, either the first coordinate of $f(\xi)$ belongs to X, or the second coordinate of $f(\xi)$ is not equal to x, i.e. $f[\alpha] \cap \alpha \times \{x\} = \varnothing$. Since f is surjective, each element of $\alpha \times \{x\}$ must be the f-image of some element of $X \cup \alpha$, and since we know that such elements cannot belong to α, the conclusion is that for every $\xi < \alpha$ there exists a $y \in X$ such that $f(y) = (\xi, x)$. Since such a y must be unique, we have just defined an injective function : $\alpha \longrightarrow X$, contradicting our choice of α. Hence Y_x was nonempty after all, and we are done.

□

6.7 The last few equivalences

In order to generalize Tarski's theorem to prove the equivalence between AC and FPP, we introduce generalized summations and products of cardinals.

Definition 165. Let $\langle \kappa_i \big| i \in I\rangle$ be a family of cardinal numbers. We define

1. $\displaystyle\sum_{i\in I}\kappa_i = \left|\bigcup_{i\in I}\kappa_i \times \{i\}\right|$, and
2. $\displaystyle\prod_{i\in I}\kappa_i = \left|\prod_{i\in I}\kappa_i\right|$.

Assuming the Axiom of Choice we have that, as long as either I or one of the κ_i is infinite, it will be the case that

$$\sup(\{\kappa_i \big| i \in I\} \cup \{I\}) \leq \sum_{i\in I}\kappa_i = \left|\bigcup_{i\in I}\kappa_i \times \{i\}\right| \leq \left|\prod_{i\in I}\kappa_i\right| \leq \left(\sup(\{\kappa_i \big| i \in I\} \cup \{I\})\right)^2 = \sup(\{\kappa_i \big| i \in I\} \cup \{I\}),$$

the middle equation being justified by the fact that the set $\prod_{i\in I}\kappa_i$ contains pairwise disjoint subsets equipotent to each of the κ_i (this is something that you proved on the last assignment), and the next inequality is justified similarly.

As a corollary of the above, we have that, if X is an infinite set, then the cardinality of the set of finite sequences of elements of X is given by $|X^{<\omega}| = \bigcup_{n<\omega} |X|^n = \sup\{|X|, \omega\} = |X|$. Similarly, we have that the cardinality of the *finite powerset* of X is given by $|[X]^n| = |X|$ (to see this, note that clearly $|X| \leq |[X]^{<\omega}|$ (just map each $x \in X$ to $\{x\}$), and on the other hand, $|[X]^{<\omega}| \leq |X^{<\omega}| = |X|$ (this is witnessed by the surjective mapping that sends each sequence in $X^{<\omega}$ to its range)). In fact, this turns out to provide yet another bunch of equivalences with the Axiom of Choice, thus generalizing Tarski's theorem.

Theorem 166. *In* ZF*, the following are equivalent:*

1. AC*,*

2. FP*,*

3. AG*,*

4. GS*.*

Proof.

AC ⇒ FP**:** This is just the observation before the statement of this theorem.

FP ⇒ AG**:** Given X, look at $([X]^{<\omega}, \triangle)$, which is a Boolean group. Now if $[X]^{<\omega}$ is equipotent with X, we can copy that abelian group structure onto X via the corresponding bijection, and we are done.

AG ⇒ GS**:** Obvious.

GS ⇒ AC**:** Given any set X, use Hartogs's theorem to find an ordinal number α that cannot be injected into X. Now equip $G = X \cup \alpha$ with a group operation $*$. Given $x \in X$, we have a function $f : \alpha \longrightarrow G$ given by $f(\xi) = x * \xi$. Note that f is injective, so it can't be the case that $\mathrm{ran}(f) \subseteq X$ (otherwise we would have that f injects α into X, contradicting our choice of α). This means that $\mathrm{ran}(f) \cap \alpha \neq \varnothing$, so there is some $\xi < \alpha$ such that $x * \xi \in \alpha$.

What the previous paragraph proves is the following: for every $x \in X$ there exists an $\xi < \alpha$ such that $x * \xi \in \alpha$. For each x, we let ξ_x be the least such, and this allows us to define a function $g : X \longrightarrow \alpha \times \alpha$ by letting

$$g(x) = (\xi_x, x * \xi_x).$$

Note that g is an injective function (if x and y are such that $\xi_x = \xi_y = \xi$ and $\xi * x = \xi_x * x = \xi_y * y = \xi * y$, then, cancelling ξ, it must be the case that $x = y$). Appealing to the canonical well-ordering of $\mathbf{Ord} \times \mathbf{Ord}$, we see that $\alpha \times \alpha$ is well-orderable, so it follows that X is well-orderable as well. Hence we have proved WOP, which, as we now know, implies AC.

□

We finalize this chapter with a corollary that is originally due to Cantor, who faced some backlash for solving what many considered to be a difficult problem of the time (to prove that there are transcendental numbers) by simply comparing cardinalities.

Corollary 167. *There are continuum many transcendental numbers.*

Proof. Each polynomial can be identified with the finite sequence of its coefficients. Hence the ring of polynomials with integer coefficients $\mathbb{Z}[X]$ has cardinality $|\mathbb{Z}^{<\omega}| = \omega$. Each of these polynomials has finitely many roots, and therefore the set of algebraic numbers (all roots of polynomials in $\mathbb{Z}[X]$) has cardinality $\leq \omega|\mathbb{Z}[X]| = \omega$. Hence there are only countably many algebraic numbers; since $\mathbb{R}$ is the disjoint union of the set of algebraic numbers (countable) and its complement, it must be the case that there are $|\mathbb{R}| = \mathfrak{c}$ many transcendental ones. □

Although cardinal addition and multiplication have very simple formulas once we assume the Axiom of Choice, cardinal exponentiation is much more complicated. In the realm of exponenciation, there tends to be many statements that actually are independent from the ZFC axioms. However, there are some results that can be proved in ZFC; a notable example of which is the following (apparently known as the "why-the-hell-is-it-four-theorem").

Theorem 168 (Shelah)**.** $\aleph_\omega^{\aleph_0} \leq \max\{\mathfrak{c}, \aleph_{\omega_4}\}$.

Chapter 7

Well-founded relations and the Axiom of Foundation

We will now address the systematic study of well-founded relations, which are relations that in a sense generalize well-orders, in that we can use these relations to perform inductive constructions. Immediately afterwards, we will utilize our newly gained knowledge about these relations to further discuss the Axiom of Foundation in more depth.

7.1 Well-founded relations

Well-founded relations are a truly wonderful product of human cultural activity. They are the ultimate heirs of a 2500-years-old tradition that starts with Euclid, who first wrote a proof using the principle that there is no infinite decreasing sequence of natural numbers, continuing with Dedekind's formalization, and includes also more recent generalizations, such as induction over well-orders or over the class of ordinal numbers. As such, we treat it with utmost respect by stating the appropriate definitions and properties below.

Definition 169. Let R be a (set or class) relation. We say that R is **well-founded** if

$$(\forall X)(X \neq \varnothing \Rightarrow (\exists x \in X)(\forall y \in X)\neg(y\ R\ x)).$$

An $x \in X$ with the property that $(\forall y \in X)\neg(y\ R\ x)$ is usually known as an **R-minimal[1] element for X**.

Example 170. The following are examples of well-founded relations. The reader is strongly encouraged to provide her own proofs that each of these relations is indeed well-founded.

1. $R = \{(f,g) \in \mathbb{R}[X] \times \mathbb{R}[X] \big| f = g' \wedge g \neq 0\}$,

2. $R = \{(n,m) \in \mathbb{N} \times \mathbb{N} \big| n \mid m \wedge m \neq n\}$,

3. if R is a Noetherian ring, then $\supsetneq\upharpoonright \{I \subseteq R \big| I \text{ is an ideal}\}$ (in fact, R is by definition a Noetherian ring if and only if the relation $\supseteq$ on its collection of ideals is well-founded),

4. in algebraic geometry, the relation "being a proper subvariety",

5. if $X = \{V \subseteq \mathbb{R}^n \big| V \text{ is a subspace}\}$ and $R = \{(V,W) \in X \times X \big| (\exists v \in W \setminus V)(\mathrm{span}(V,v) = W)\}$.

Remark 171.

1. Note that if R is well-founded, then it is irreflexive. For if some x satisfied $x\ R\ x$, then the set $\{x\}$ would not have any R-minimal elements.

2. Also note that every well-founded strict linear order is a well-order. For if $<$ is a strict linear order, then $x \in \mathrm{dom}(<)$ is an $<$-minimal element for some set X if and only if $x = \min(X)$.

3. Additionally, note that if R is well-founded then it is impossible to have "a cycle", that is, there are no elements $x_0, \dots, x_n$ satisfying that $x_0\ R\ x_1\ R\ \cdots\ R\ x_n\ R\ x_0$ (otherwise the set $\{x_0, \dots, x_n\}$ would not have an R-minimal element).

[1] In fact, if R was a strict partial order then such an x would be an actual minimal element, according to the definitions that we have stated earlier in this course. Right now, the only difference is that we are not necessarily assuming that the relation R is transitive.

Regarding the last point in the above remark, we can say even more. Not only do well-founded relations have no "cycles", they in fact do not have infinite decreasing sequences, as the next theorem shows.

Theorem 172. *A binary (set or class) relation R is well-founded if and only if there does not exist a sequence $\langle x_n \big| n < \omega\rangle$ satisfying that $(\forall n < \omega)(x_{n+1}\ R\ x_n)$.*

Proof.

⇒: If there is a sequence as described in the right side of the biconditional in the statement of the theorem, the range of that sequence, $\{x_n \big| n < \omega\}$ would not have an R-minimal element, and so R would not be well-founded.

⇐: Suppose that R is not well-founded. Then there exists a set X such that $(\forall x \in X)(\exists y \in X)(y\ R\ x)$. This means that each element in the family $\{\{y \in X \big| y\ R\ x\} \big| x \in X\}$ is nonempty, and so we can let f be a choice function for that family (defined on the index set X, that is, $(\forall x \in X)(f(x) \in \{y \in X \big| y\ R\, x\})$, or in other words, $(\forall x \in X)(f(x)\ R\, x)$). Thus we can (upon choosing some fixed $x \in X$) recursively define $g : \omega \longrightarrow X$ by $g(0) = x$ and $g(n+1) = f(g(n))$, and then let g constitute the sequence $\langle x_n \big| n < \omega\rangle$ (that is, define $x_n = g(n)$). This sequence satisfies that for every $n < \omega$, $x_{n+1} = g(n+1) = f(g(n))\ R\ g(n) = x_n$, and the proof is finished.

□

Definition 173. Let R be a binary (set or class) relation (i.e. $R \subseteq \mathbf{V} \times \mathbf{V}$). We say that R is **set-like** if for every x, the class $\{y \big| y\ R\ x\}$ is a set.

Lemma 174. *If R is a set-like well-founded relation, then every nonempty class X has an R-minimal element.*

Proof. Pick some $x \in X$. By the transfinite recursion theorem scheme for ordinal numbers, restricted to ordinals less that ω, define a sequence of sets by $X_0 = \{x\}$ and $X_{n+1} = \bigcup_{y \in X_n}\{z \in X \big| z\ R\ y\} = \{z \in X \big| (\exists y \in X_n)(z\ R\ y)\}$ (which is a set because of the Replacement Axiom and because, R being set-like, each of the classes $\{z \big| z\ R\ y\}$ is a set). Then we let $Y = \bigcup_{n<\omega} X_n$. Since Y is a set, there is some $y \in Y$ that is R-minimal for Y. We claim that y is R-minimal for X. This is because, since $y \in Y = \bigcup_{n<\omega} X_n$, we must have $y \in X_n$ for some $n \in \omega$. This means that any $z \in X$ such that $z\ R\ y$ would satisfy $z \in X_{n+1} \subseteq Y$; our y was assumed to be R-minimal for Y, meaning that there is no such z. This means that for no $z \in X$ can it be the case that $z\ R\ y$, and so y is R-minimal for X, and we are done. □

We are now ready to state the most general version of an induction theorem that we will ever lay our hands on. Immediately after, we will also state the corresponding version of a recursion theorem.

Theorem 175. *Let R be a well-founded, set-like relation on a class X. Suppose we have an $\mathscr{L}_{\mathrm{ST}}$-formula $\varphi(x)$. If for every x it is the case that*

$$(\forall y)(y\ R\ x \Rightarrow \varphi(y)) \Rightarrow \varphi(x)$$

then it is the case that $(\forall x)(\varphi(x))$.

Proof. If there was an x such that $\neg\varphi(x)$, by taking an R-minimal element for the (nonempty) class $\{x \big| \neg\varphi(x)\}$, we obtain that $\neg\varphi(x)$ yet $(\forall y)(y\ R\ x \Rightarrow \varphi(y))$. This is a contradiction, and we are done. □

As announced before, here is the most general version of the recursion theorem that we will ever lay our hands on.

Theorem 176. *Let $\mathbf{G} : \mathbf{V} \longrightarrow \mathbf{V}$ be a (class) function, and let R be a set-like well-founded relation. Then there is an $\mathbf{F} : \mathbf{V} \longrightarrow \mathbf{V}$ such that, for all x,*

$$\mathbf{F}(x) = \mathbf{G}(\mathbf{F} \upharpoonright \{y \big| y\ R\ x\}).$$

Proof. Let $\pi(X, h)$ be the $\mathscr{L}_{\mathrm{ST}}$-formula

$$h \text{ is a function} \wedge \mathrm{dom}(h) = X \wedge (\forall x \in X)((\forall y)(y\ R\ x \Rightarrow y \in X) \wedge h(x) = \mathbf{G}(h \upharpoonright \{y \in X \big| y\ R\ x\})),$$

then we let $\psi(x, y)$ be the $\mathscr{L}_{\mathrm{ST}}$-formula

$$(\exists X)(x \in X \wedge (\exists h)(\pi(X, h) \wedge h(x) = y)),$$

and it is readily checked that ψ defines the desired class-function F (the fact that for every x there is an X and an h such that $\pi(X, h)$ holds is proved by well-founded induction on x). □

7.2 The Axiom of Foundation

Recall that the Axiom of Foundation is the following $\mathscr{L}_{\mathrm{ST}}$-formula

$$(\forall x)(x \neq \varnothing \Rightarrow (\exists y \in x)(y \cap x = \varnothing)),$$

noting that $y \cap x = \varnothing$ can be rewritten as $(\forall z \in x)\neg(z \in y)$, we see that the Axiom of Foundation is just the statement that every nonempty set x has an $\in$-minimal element. In other words, the Axiom of Foundation is simply stating that the membership relation $\in$ is well-founded.

Thus, assuming the Axiom of Foundation allows us to utilize Theorems 175 and 176 applied to the relation $\in$, since we are assuming that relation to be well-founded (note that this relation is always set-like, pretty much by definition: if x is a set, then $\{y \big| y \in x\} = x$ is "also" a set). The following is an example of how to utilize recursion over the well-founded relation $\in$ to define a class function, and then use induction over this same well-founded relation to prove statements about this class function.

Example 177. We define the **transitive closure** of a set x by $\in$-recursion, by means of the following formula:

$$\mathrm{trcl}(x) = x \cup \left(\bigcup_{y \in x} \mathrm{trcl}(y) \right).$$

We can now proceed to prove some properties of the transitive closure by $\in$-induction.

Theorem 178. *Let x be an arbitrary set.*

1. $\mathrm{trcl}(x)$ *is a transitive set.*
2. $\mathrm{trcl}(x)$ *is "the smallest transitive set containing x as a subset". That is, if $x \subseteq X$ and X is transitive, then* $\mathrm{trcl}(x) \subseteq X$.

Proof.

1. Let $z \in y \in \mathrm{trcl}(x) = x \cup \left(\bigcup_{w \in x} \mathrm{trcl}(w)\right)$. Then either $y \in \mathrm{trcl}(w)$ for some $w \in x$, or $y \in x$. If $y \in \mathrm{trcl}(w)$ for some $w \in x$, then by inductive hypothesis $\mathrm{trcl}(w)$ is transitive, so from $z \in y \in \mathrm{trcl}(w)$ we conclude $z \in \mathrm{trcl}(w) \subseteq \bigcup_{w \in x} \mathrm{trcl}(w) \subseteq \mathrm{trcl}(x)$. Otherwise, if $y \in x$, then $z \in y \subseteq y \cup \left(\bigcup_{v \in y} \mathrm{trcl}(v)\right) = \mathrm{trcl}(y) \subseteq \mathrm{trcl}(x)$, and we are done.
2. Suppose that $x \subseteq X$ and X is transitive; and let $z \in \mathrm{trcl}(x) = x \cup \left(\bigcup_{y \in x} \mathrm{trcl}(y)\right)$. Then either $z \in x$, or $z \in \mathrm{trcl}(y)$ for some $y \in x$. If $z \in x$, since $x \subseteq X$ then $z \in X$ and we are done. Otherwise, if $z \in \mathrm{trcl}(y)$ for some $y \in x$, note that $x \subseteq X$ and so $y \in X$. The assumption that X is transitive means that $y \subseteq X$. By induction hypothesis, $\mathrm{trcl}(y)$ is the smallest transitive set including y as a subset, so $\mathrm{trcl}(y) \subseteq X$. Since $z \in \mathrm{trcl}(y)$, this means $z \in X$. In either case we concluded that $z \in X$, and we are done.

□

The Axiom of Foundation has another outstanding consequence, in addition to allowing us to do $\in$-induction and $\in$-recursion. Recall that Zermelo's cumulative hierarchy is defined transfinitely by $V_0 = \varnothing$, $V_{\alpha+1} = \mathfrak{P}(V_\alpha)$ and $V_\alpha = \bigcup_{\xi<\alpha} V_\xi$ if α is limit. We have discussed the class $\bigcup_{\alpha \in \mathbf{Ord}} V_\alpha$: sets that are members of this class have a very clear structure, in terms of being built by a process that starts with $\varnothing$ and proceeds by enclosing this between brackets in different combinations. We have seen that most everyday mathematical objects can be implemented not just as sets, but as sets that belong to this proper class. Now we will see that the Axiom of Foundation is nothin but the assumption that every set belongs to this proper class.

Theorem 179. $\mathsf{ZF}^- \vdash$ *Axiom of Foundation* $\iff V = \bigcup_{\alpha \in \mathbf{Ord}} V_\alpha$.

Proof.

⇒: If we assume Foundation, we can do proofs by $\in$-induction. So we will prove, by $\in$-induction on $x \in V$, the statement $(\exists \alpha \in \mathbf{Ord})(x \in V_\alpha)$. Suppose it is indeed the case that for all $y \in x$, $(\exists \xi \in \mathbf{Ord})(y \in V_\xi)$, and let[2] $\alpha = \sup\{\mathrm{rank}(y) + 1 \big| y \in x\}$. Then for all $y \in x$, $y \in V_{\mathrm{rank}(y)+1} \subseteq V_\alpha$. Hence $x \subseteq V_\alpha$, which means that $x \in \mathfrak{P}(V_\alpha) = V_{\alpha+1} \subseteq \bigcup_{\alpha \in \mathbf{Ord}} V_\alpha$, and we are done.

[2] Recall that, if $x \in \bigcup_{\alpha \in \mathbf{Ord}} V_\alpha$, we defined $\mathrm{rank}(x)$ as the least ordinal α such that $x \in V_{\alpha+1}$; equivalently, $\mathrm{rank}(x)$ is the unique α such that $x \in V_{\alpha+1} \setminus V_\alpha$, or, also equivalently, $\mathrm{rank}(x)$ is the least ordinal α such that $x \subseteq V_\alpha$.

$\Leftarrow$: Using once again the definition of rank, we note that our assumption here is that it makes sense to mention $\text{rank}(x)$ for every set x. Note that, if $y \in x$, then $\text{rank}(y) < \text{rank}(x)$, for if $\alpha = \text{rank}(x)$, meaning that α is the least ordinal such that $x \subseteq V_\alpha$, then we have that $y \in V_\alpha$. If α is limit this implies that $y \in V_\xi$ for some $\xi < \alpha$, and so $y \subseteq V_\xi$ because V_ξ is transitive; otherwise α is a successor, say $\alpha = \xi + 1$, and so from $y \in V_\alpha = V_{\xi+1} = \mathfrak{P}(V_\xi)$ we infer that $y \subseteq V_\xi$. In both the successor and the limit case we obtained that $y \subseteq V_\xi$ for some $\xi < \alpha$, showing that $\text{rank}(y) < \alpha = \text{rank}(x)$.

Having that property at our disposal, let us now prove that the relation $\in$ is well-founded. So let $X \neq \varnothing$, let $\alpha = \min\{\text{rank}(x) \big| x \in X\}$, and let $x \in X$ be such that $\text{rank}(x) = \alpha$. Then every $y \in x$ must satisfy that $\text{rank}(y) < \text{rank}(x) = \alpha$, and since α is the least possible rank for elements of X, we conclude that $y \notin X$. This means that x is $\in$-minimal for X, and we are done.

□

Appendix A

Axioms of Set Theory

0. **Axiom of Existence:** $(\exists x)(x = x)$

   Informally: *there exists a set.*

1. **Axiom of Extensionality:** $(\forall x)(\forall y)((\forall z)(z \in x \iff z \in y) \Rightarrow x = y)$

   Informally: *a set is determined by its elements.*

2. **Axiom Schema of Comprehension:** For every formula φ of the Language of Set Theory with one free variable,

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

   Informally: *for every set x, the set $y = \{z \in x \big| \varphi(z)\}$ exists.*

3. **Axiom of Pairing:** $(\forall x)(\forall y)(\exists z)(\forall w)(w \in z \iff (w = x \vee w = y))$

   Informally: *for every two sets x and y, the set $z = \{x, y\}$ exists.*

4. **Axiom of Union:** $(\forall x)(\exists y)(\forall z)(z \in y \iff (\exists w)(w \in x \wedge z \in w))$

   Informally: *for every set x, the set $y = \bigcup\limits_{w \in x} w$ exists.*

5. **Axiom of Powerset:** $(\forall x)(\exists y)(\forall z)(z \in y \iff z \subseteq x)$

   Informally: *for every set x, the powerset $y = \mathfrak{P}(x) = \{z \big| z \subseteq x\}$ exists.*

6. **Axiom of Infinity:** $(\exists x)(\varnothing \in x \wedge (\forall y)(y \in x \Rightarrow y \cup \{y\} \in x))$

   Informally: *there exists an infinite set.*

7. **Axiom of Foundation:** $(\forall x)((\exists y)(y \in x) \Rightarrow (\exists y)(y \in x \wedge \neg(\exists z)(z \in x \wedge z \in y)))$

   Informally: *for every nonempty set x there exists a $y \in x$ such that y and x are disjoint.*

   In fact, a better way of informally phrasing this axiom would be: *the membership relation $\in$ is well-founded*; unfortunately, this will only make sense towards the final stages of the course.

8. **Axiom Schema of Replacement:** For every formula φ of the Language of Set Theory with two free variables,

   $$(\forall z)((\forall x)(x \in z \Rightarrow (\exists! y)\varphi(x, y)) \Rightarrow (\exists w)(\forall x)(x \in z \Rightarrow (\exists y)(y \in w \wedge \varphi(x, y))))$$

   Informally: *for every set x, if what φ describes behaves like a function with domain x, then the range of that function (the set of images of elements of x under the function) exists.*

9. **Axiom of Choice:**

   $$(\forall x)((\varnothing \notin x \wedge (\forall y)(\forall z)((y \in x \wedge z \in x \wedge y \neq z) \Rightarrow y \cap z = \varnothing)) \Rightarrow (\exists w)(\forall y)(y \in x \Rightarrow (\exists! z)(z \in y \cap w)))$$

   Informally: *for every set x whose elements are pairwise disjoint and nonempty, there exists a set w that contains exactly one element from each $y \in x$.*

ZFC stands for all axioms 0 through 9, whereas ZF stands for axioms 0 through 8. Z is axioms 0 through 7, and ZC consists of axioms 0 through 7 plus axiom 9. So, intuitively speaking, the letter F stands for the Axiom Schema of Replacement, whereas the letter C stands for the Axiom of Choice. Similarly, a minus sign as a superindex denotes the deletion of the Foundation axiom (thus e.g. Z^- consists of axioms 0 through 6, whereas ZFC^- consists of axioms 0 through 6 together with axioms 8 and 9). There are, of course, many other commonly used notations for other fragments of the ZFC axiom system, which we will not mention here.

Appendix B

Worksheet 1: Constructing $\mathbb{Z}$

The main idea is that we would like to represent integers as differences of natural numbers. Thus, we would like for an element of $\mathbb{Z}$ to be "something of the form $n - m$ where $n, m \in \mathbb{N}$". Obviously, two things of the form $n - m$ and $p - q$ are actually supposed to be the same thing if $n + q = p + m$ (and notice that the latter already makes sense, since now we only have elements of $\mathbb{N}$ and the usual addition of $\mathbb{N}$). In order to make this formal, we do the following:

1. Define the relation $\sim$ on $\mathbb{N} \times \mathbb{N}$ by letting

$$(a, b) \sim (c, d) \iff a + d = c + b.$$

Prove that $\sim$ is an equivalence relation on $\mathbb{N} \times \mathbb{N}$.

Definition 180.

- We define the set $\mathbb{Z} = (\mathbb{N} \times \mathbb{N})/\sim$.
- We define the binary operation $+ : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}$ by stipulating that $[(a, b)]_\sim + [(c, d)]_\sim = [(a + c, b + d)]_\sim$.
- We define the binary operation $\cdot : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}$ by stipulating that $[(a, b)]_\sim \cdot [(c, d)]_\sim = [(ac + bd, bc + ad)]_\sim$.
- We define the binary relation $\leq \subseteq \mathbb{Z} \times \mathbb{Z}$ by stipulating that $[(a, b)]_\sim \leq [(c, d)]_\sim$ if and only if $a + d \leq c + b$.

We will now prove that the set $\mathbb{Z}$, with the operations thus defined, satisfies what we would expect it to satisfy (which in technical terms reads: "the structure $(\mathbb{Z}, +, \cdot, \leq)$ is an ordered commutative ring with identity").

2. Prove that $+$ is well-defined.

3. Prove that $+$ is commutative and associative.

4. Prove that there exists an element of $\mathbb{Z}$ (from now on we will denote this element by 0) which is an identity element for $+$, that is, $(\forall n \in \mathbb{Z})(n + 0 = 0 + n = n)$. Can you explicitly describe all of the elements of 0?

5. Prove that every element of $\mathbb{Z}$ has an inverse with respect to $+$, that is, $(\forall n \in \mathbb{Z})(\exists m \in \mathbb{Z})(n + m = m + n = 0)$.

Thus $(\mathbb{Z}, +)$ is an abelian group.

6. Prove that $\cdot$ is well-defined.

7. Prove that $\cdot$ is commutative and associative.

8. Prove that $\cdot$ is distributive over $+$, that is, $(\forall m, n, k \in \mathbb{Z})(m \cdot (n + k) = m \cdot n + m \cdot k)$.

9. Prove that there exists an element of $\mathbb{Z}$, distinct from 0 (and this one will from now on be denoted by 1) which is an identity element for $\cdot$, that is, $(\forall n \in \mathbb{Z})(n \cdot 1 = 1 \cdot n = n)$. Can you explicitly describe all of the elements of 1?

Thus the structure $(\mathbb{Z}, +, \cdot)$ is a commutative ring with identity. It can be shown (but we won't do it here) that it is in fact an integral domain (that is, there are no zero divisors in $\mathbb{Z}$).

10. Prove that $\leq$ is well-defined.

11. Prove that $\leq$ is a linear order.

12. Prove that $\leq$ respects the addition operation, that is, $(\forall n, m, k \in \mathbb{Z})(n \leq m \Rightarrow n + k \leq m + k)$.

13. Prove that, for all $n, m \in \mathbb{Z}$ and all $k \in \mathbb{Z}$ such that $0 < k$, it is the case that $n \leq m \Rightarrow n \cdot k \leq m \cdot k$.

Thus $(\mathbb{Z}, +, \cdot, \leq)$ is a commutative ordered ring with identity. We now only need to show that this ring in fact "extends" the structure given by $(\mathbb{N}, +, \cdot, \leq)$. So consider the mapping $E : \mathbb{N} \longrightarrow \mathbb{Z}$ given by $E(n) = [(n+1, 1)]_{\sim}$.

14. Prove that $E(m+n) = E(m) + E(n)$, $E(m \cdot n) = E(m) \cdot E(n)$, and $m < n \Rightarrow E(m) < E(n)$ for all $m, n \in \mathbb{N}$ (the latter implies, in particular, that E is injective, thus we have that E is an *embedding*).

15. Notice that, for all $m, n \in \mathbb{N}$, it is indeed the case that $E(m) - E(n) = [(m, n)]_{\sim}$.

Appendix C

Worksheet 2: Constructing $\mathbb{Q}$

We now want to represent rationals as quotients of integers, that is, an element of $\mathbb{Q}$ would be "something of the form $\frac{n}{m}$ where $n, m \in \mathbb{Z}$ (and $m \neq 0$)". Obviously, two things of the form $\frac{n}{m}$ and $\frac{p}{q}$ are actually supposed to be the same thing if $nq = pm$. So we proceed to formalize this below:

1. Define the relation $\sim$ on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ by letting

$$(a, b) \sim (c, d) \iff ad = cb.$$

   Prove that $\sim$ is an equivalence relation on $\mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$.

Definition 181.

- We define the set $\mathbb{Q} = (\mathbb{Z} \times (\mathbb{Z} \setminus \{0\}))/\sim$.
- We define the binary operation $+ : \mathbb{Q} \times \mathbb{Q} \longrightarrow \mathbb{Q}$ by stipulating that $[(a, b)]_\sim + [(c, d)]_\sim = [(ad + bc, bd)]_\sim$.
- We define the binary operation $\cdot : \mathbb{Q} \times \mathbb{Q} \longrightarrow \mathbb{Q}$ by stipulating that $[(a, b)]_\sim \cdot [(c, d)]_\sim = [(ac, bd)]_\sim$.
- Note that, for every equivalence class $q \in \mathbb{Q}$, there exist a *positive* b such that $(a, b) \in q$. This allows us to define the binary relation $\leq \subseteq \mathbb{Q} \times \mathbb{Q}$ by stipulating that $[(a, b)]_\sim \leq [(c, d)]_\sim$, for $b, d > 0$, if and only if $ad \leq cb$.

We will now prove that the set $\mathbb{Q}$, with the operations thus defined, satisfies what we would expect it to satisfy (which in technical terms reads: "the structure $(\mathbb{Q}, +, \cdot, \leq)$ is an ordered commutative ring with identity").

2. Prove that $+$ is well-defined.
3. Prove that $+$ is commutative and associative.
4. Prove that there exists an element of $\mathbb{Q}$ (from now on we will denote this element by 0) which is an identity element for $+$, that is, $(\forall q \in \mathbb{Q})(q + 0 = 0 + q = q)$. Can you explicitly describe all of the elements of 0?
5. Prove that every element of $\mathbb{Q}$ has an inverse with respect to $+$, that is, $(\forall q \in \mathbb{Q})(\exists r \in \mathbb{Q})(q + r = r + q = 0)$.

Thus $(\mathbb{Q}, +)$ is an abelian group.

6. Prove that $\cdot$ is well-defined.
7. Prove that $\cdot$ is commutative and associative.
8. Prove that $\cdot$ is distributive over $+$, that is, $(\forall q, r, s \in \mathbb{Q})(q \cdot (r + s) = q \cdot r + q \cdot s)$.
9. Prove that there exists an element of $\mathbb{Q}$, distinct from 0 (and this one will from now on be denoted by 1) which is an identity element for $\cdot$, that is, $(\forall q \in \mathbb{Q})(q \cdot 1 = 1 \cdot q = q)$. Can you explicitly describe all of the elements of 1?
10. Prove that every non-zero $q \in \mathbb{Q}$ has a multiplicative inverse, that is, $(\forall q \in \mathbb{Q})(q \neq 0 \Rightarrow (\exists r \in \mathbb{Q})(q \cdot r = r \cdot q = 1))$.

Thus the structure $(\mathbb{Q}, +, \cdot)$ is a commutative field.

11. Prove that $\leq$ is well-defined.
12. Prove that $\leq$ is a linear order.

13. Prove that $\leq$ respects the addition operation, that is, $(\forall q, r, s \in \mathbb{Q})(q \leq r \Rightarrow q + s \leq r + s)$.

14. Prove that, for all $q, r \in \mathbb{Q}$ and all $s \in \mathbb{Q}$ such that $0 < s$, it is the case that $q \leq r \Rightarrow q \cdot s \leq r \cdot s$.

Thus $(\mathbb{Q}, +, \cdot, \leq)$ is an ordered field. We would now like to show that this field in fact "extends" the ring $(\mathbb{Z}, +, \cdot, \leq)$. So consider the mapping $E : \mathbb{Z} \longrightarrow \mathbb{Q}$ given by $E(n) = [(n, 1)]_{\sim}$.

15. Prove that $E(m + n) = E(m) + E(n)$, $E(m \cdot n) = E(m) \cdot E(n)$, and $m < n \Rightarrow E(m) < E(n)$ for all $m, n \in \mathbb{Z}$. Moreover, $E(0) = 0$ and $E(1) = 1$.

16. Prove that E is injective, so that it is an embedding.

17. Notice that, for all $m, n \in \mathbb{Z}$, it is indeed the case that $E(m) \cdot (E(n))^{-1} = [(m, n)]_{\sim}$.

Appendix D

Worksheet 3: Ordinal arithmetic

Thanks to the Ordinal Recursion Theorem Scheme, we can now define arithmetic operations on the ordinal numbers, in a way that is reminiscent of what is done with natural numbers.

Definition 182. Given a fixed ordinal α, we define the **ordinal addition** $\alpha\dot{+}\beta$ by means of the following transfinite recursion on β:

- $\alpha\dot{+}0 = \alpha$,
- $\alpha\dot{+}S(\beta) = S(\alpha\dot{+}\beta)$,
- $\alpha\dot{+}\beta = \sup\{\alpha\dot{+}\gamma\big|\gamma < \beta\}$ if $\beta = \bigcup\beta$.

Let us now proceed to prove a few properties of the ordinal addition operation $\dot{+}$.

1. Prove that, if γ is a limit ordinal, then so is $\alpha\dot{+}\gamma$ (*hint*: do it by contradiction!).
2. Prove that, for all ordinals α, β, γ, $\alpha\dot{+}(\beta\dot{+}\gamma) = (\alpha\dot{+}\beta)\dot{+}\gamma$.
3. Is ordinal addition commutative? Prove or provide a counterexample.
4. Prove that, for all ordinals α, β, it is the case that $\beta \leq \alpha\dot{+}\beta$, and if $\beta > 0$ then actually $\alpha < \alpha\dot{+}\beta$.
5. Prove that, for every ordinal number α, we have that $0\dot{+}\alpha = \alpha$.
6. Prove that ordinal addition satisfies the cancellative property *from the left*: For all ordinals α, β, γ, $\beta\dot{+}\alpha = \beta\dot{+}\gamma$ implies $\alpha = \gamma$.
7. Does ordinal addition satisfies the cancellative property from the right? Prove or provide a counterexample.

Definition 183. We now define **ordinal multiplication** $\alpha\cdot\beta$, for each fixed ordinal number α, by the following transfinite recursion on β:

- $\alpha \cdot 0 = 0$,
- $\alpha \cdot S(\beta) = \alpha \cdot \beta\dot{+}\alpha$,
- $\alpha \cdot \beta = \sup\{\alpha \cdot \gamma\big|\gamma < \beta\}$ if $\beta = \bigcup\beta$.

We now proceed to prove properties of the ordinal multiplication operation $\cdot$:

8. Prove that, if γ is a limit ordinal, then so is $\alpha \cdot \gamma$ (*hint*: do it by contradiction!).
9. Prove that, for all ordinals α, β, γ, $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$.
10. Is ordinal multiplication commutative? Prove or provide a counterexample.
11. Prove that, if α, β are ordinals and $\alpha \geq 1$, it is the case that $\beta \leq \alpha \cdot \beta$; and furthermore if $\beta > 1$ then actually $\alpha < \alpha \cdot \beta$.
12. Prove that, for every ordinal number α, we have that $0 \cdot \alpha = 0$ and $1 \cdot \alpha = \alpha$.
13. Prove that ordinal multiplication satisfies the cancellative property *from the left*: For all ordinals α, β, γ, $\beta \cdot \alpha = \beta \cdot \gamma$ implies $\alpha = \gamma$.
14. Does ordinal multiplication satisfies the cancellative property from the right? Prove or provide a counterexample.

Bonus problem: Use the Ordinal Recursion Theorem Scheme, and your appropriate intuition for analogies, to define an ordinal exponenciation operation $\alpha^{\cdot\beta}$. Which of the properties that you expect this operation to have actually hold true?

Appendix E

Worksheet 4: Naïve cardinal arithmetic

The cardinal arithmetic that we will develop here is naïve in the sense that we still have not properly defined cardinality. That is, we still have not defined a class function $|\cdot| : \mathbf{V} \longrightarrow \mathbf{V}$ such that for any two sets X, Y, it is the case that $|X| = |Y|$ if and only if X is equipotent to Y. So we either assume, for the time being, that we have such a function (so that it makes sense to speak about "the cardinality of the set X" as an object, and of elements $|X| \in \operatorname{ran}(F)$ as "cardinals"), or else we think of every statement here as being an abbreviation of an appropriate statement that can be uttered in the metatheory.

Definition 184. Given two cardinals $|A|, |B|$, we define

- $|A| + |B| = |(A \times \{0\}) \cup (B \times \{1\})|$
- $|A| \cdot |B| = |A \times B|$
- $|A|^{|B|} = |A^B|$

In the following exercises, don't bother with checking that the functions you define are bijections (in general, any function that is reasonably defined from a bunch of bijections will still be a bijection), just explicitly write what the bijection should be.

1. Convince yourself that if the sets A, B are both finite, then the results of performing cardinal operations coincide with what you would expect.

2. Prove that cardinal addition, cardinal multiplication and cardinal exponentiation are well-defined, that is, if $|X| = |A|$ and $|Y| = |B|$, then $|A| + |B| = |X| + |Y|$, $|A| \cdot |B| = |X| \cdot |Y|$ and $|A|^{|B|} = |X|^{|Y|}$.

3. Prove that, for every finite n, $n + \aleph_0 = \aleph_0$ and, if $n \neq 0$, then $n \cdot \aleph_0 = \aleph_0$. Prove also that $\aleph_0 \cdot \aleph_0 = \aleph_0$ (in particular, $+$ and $\cdot$ are not cancellative).

4. Prove that, for all sets A, B, C:

   (a) $|A| + |B| = |B| + |A|$,

   (b) $|A| \cdot |B| = |B| \cdot |A|$,

   (c) $(|A| + |B|) + |C| = |A| + (|B| + |C|)$,

   (d) $(|A| \cdot |B|) \cdot |C| = |A| \cdot (|B| \cdot |C|)$,

   (e) $|A| \cdot (|B| + |C|) = |A| \cdot |B| + |A| \cdot |C|$,

   (f) $|A| + 0 = |A|$,

   (g) $|A| \cdot 0 = 0$,

   (h) $|A| \cdot 1 = |A|$,

   (i) $|A|^0 = 1$,

   (j) $0^{|A|} = \begin{cases} 0 \text{ if } A \neq \varnothing, \\ 1 \text{ if } A = \varnothing \end{cases}$

   (k) $2^{|A|} = |\mathfrak{P}(A)|$.

5. Prove the following facts about exponenciation for all sets A, B, C:

(a) $(|A|^{|B|})^{|C|} = |A|^{|B| \cdot |C|}$,

(b) $|A|^{|B|} \cdot |A|^{|C|} = |A|^{|B|+|C|}$,

(c) $(|A| \cdot |B|)^{|C|} = |A|^{|C|} \cdot |B|^{|C|}$.

(d) Compute how many sequences of real numbers there are (that is, compute $|\mathbb{R}^{\mathbb{N}}| = \mathfrak{c}^{\aleph_0}$).

(e) Let $\mathscr{C}(\mathbb{R}, \mathbb{R})$ be the set of all continuous functions $: \mathbb{R} \longrightarrow \mathbb{R}$. What is $|\mathscr{C}(\mathbb{R}, \mathbb{R})|$?

Appendix F

Worksheet 5: The Axiom of Choice

Choose (see what I did there?) one of the problems, whichever looks most appealing to you, and work through it. Assume that you're not in a Set Theory class, but rather on an "ordinary" math class.

1. Let $f : A \longrightarrow B$ be a function.
   (a) Prove that f is injective if and only if there exists a $g : B \longrightarrow A$ such that $g \circ f = \mathrm{id}_A$ (that is, a function is injective if and only if it has a left inverse).
   (b) Try to prove that f is surjective if and only if it has a right inverse.
   (c) Does this become any easier for bijective f?
2. (a) Let's do some Real Analysis[1]: given any $A \subseteq \mathbb{R}$, prove that for every $x \in \mathbb{R}$, $x \in \bar{A}$ if and only if there exists a sequence $\langle x_n \big| n \in \mathbb{N}\rangle \in A^{\mathbb{N}}$ converging to x.
   (b) Does this become any easier if we add the assumption that A is countable?
   (c) How about an $x \in A$? Is it easier to find a sequence $\langle x_n \big| n \in \mathbb{N}\rangle \in A^{\mathbb{N}}$ converging to x?
3. Prove that a countable union of countable sets is countable (here "countable" should be taken to mean "countably infinite").
4. Suppose we have a collection $\mathscr{I}$ of intervals in $\mathbb{R}$ (that is, every element of $\mathscr{I}$ is of the form $I = (a, b)$ for $a, b \in \mathbb{R}$, $a < b$) that are pairwise disjoint. Prove that $\mathscr{I}$ is countable.
5. Let A be an arbitrary set and let $\sim$ be an equivalence relation on A. Recall that a *complete set of representatives*, or a *transversal*, for $\sim$, is just a subset $X \subseteq A$ such that $A/\sim = \{[x]_\sim \big| x \in X\}$ and for any two distinct $x, y \in X$, $x \not\sim y$ (that is, a complete set of representatives is a set containing exactly one element from each equivalence class).
   (a) Exhibit a complete set of representatives for the equivalence relation "congruent modulo 7" in $\mathbb{Z}$.
   (b) Exhibit a complete set of representatives for the equivalence relation $\sim$ on $\mathbb{R}$ defined by $x \sim y$ if and only if $y - x \in \mathbb{Q}$.
   (c) Now let A be an arbitrary set, and let $\sim$ be an arbitrary equivalence relation on A. Is it hard to come up with a complete set of representatives?
6. For those of you who like Modern Algebra: recall that a ring R (commutative and with identity) is said to be *Noetherian* if every ideal $I \subseteq R$ is finitely generated. Prove that if a ring R satisfies that every ascending chain $I_0 \subseteq I_1 \subseteq \cdots \subseteq I_n \subseteq \cdots$ of ideals in R eventually halts (that is, there is an $n \in \mathbb{N}$ such that $I_n = I_{n+1} = I_{n+2} = \cdots$), then R is Noetherian.
7. Given a family of sets, $\mathscr{X} = \{X_i \big| i \in I\}$, we define the **cartesian product** of the X_i to be the set

$$\prod_{i \in I} X_i = \{f : I \longrightarrow \bigcup_{i \in I} X_i \big| (\forall i \in I)(f(i) \in X_i)\}.$$

   Why does it make sense to define the cartesian product like that?

[1] What I take to be the definition of the closure $\bar{A}$ of the set $A \subseteq \mathbb{R}$ is, of course, the right definition:

$$\bar{A} = \{x \in \mathbb{R} \big| (\forall \varepsilon > 0)(A \cap (x - \varepsilon, x + \varepsilon) \neq \varnothing)\}.$$

(a) More importantly, prove that if each of the X_i is nonempty, then $\prod_{i \in I} X_i$ is nonempty as well.

(b) In the described situation, let $I = \mathbb{N}$ and for each $i \in I$, we let $X_i = \{n \in \mathbb{N} \big| n \geq i\}$. Exhibit a specific member of $\prod_{i \in \mathbb{N}} X_i$.

(c) Now let X, I be arbitrary nonempty. If $X_i = X$ for all $i \in I$, exhibit a particular element of $\prod_{i \in I} X$.

Appendix G

Statements equivalent to the Axiom of Choice

Axiom of Choice (AC):

If $X \neq \varnothing$ is such that every $x \in X$ is nonempty, and every two distinct $x, y \in X$ are disjoint, then there exists a set A such that for every $x \in X$, $|A \cap x| = 1$ (A is sometimes called a *selector* for X).

Axiom of Choice, version 2 (AC_2):

For every $X \neq \varnothing$ such that every $x \in X$ is nonempty, there exists a function $f : X \longrightarrow \bigcup X$ such that for every $x \in X$, $f(x) \in x$ (this f is known as a *choice function* for X).

Axiom of Choice, Enderton style (AC_3):

For every relation R there exists a function $F \subseteq R$ with $\mathrm{dom}(F) = \mathrm{dom}(R)$.

All Epis Split (ES):

Whenever $f : A \longrightarrow B$ is surjective, there exists a $g : B \longrightarrow A$ such that $f \circ g = \mathrm{id}_B$ (we say that g is a *right inverse* for f, and that f *splits*).

Well-Ordering Principle (WO):

For every set A, there exists a $\leq \subseteq A \times A$ which is a well-order relation on A (i.e. *every set is well-orderable*).

Zorn's Lemma (ZL):

Every (nonempty) partially ordered set $\langle \mathbb{P}, \leq \rangle$ such that every (nonempty) totally ordered $X \subseteq \mathbb{P}$ has a lower (respectively upper) bound, must have minimal (respectively maximal) elements (this would be the first example of a *forcing axiom*, since partially ordered sets are sometimes called *forcing notions*).

Every Vector Space has a Basis (VB)

Hausdorff's Maximal Principle (HM):

Every totally ordered subset of a partially ordered set can be extended to a $\subseteq$-maximal totally ordered subset.

Teichmüller-Tukey Lemma (TL):

If $\mathscr{X}$ is a family of sets "of finite character" (this is, for every X we have $X \in \mathscr{X} \iff [X]^{<\aleph_0} \subseteq \mathscr{X}$), then $\mathscr{X}$ has a $\subseteq$-maximal element.

Maximal Ideal Principle (MIP):

Every ring (with identity) has a maximal ideal.

Tychonoff's Theorem (TY):

The topological product of any family of compact topological spaces is compact.

Disjointification principle (DIS):

For every indexed family $\{A_\alpha \big| \alpha \in \Lambda\}$ there exists another indexed family $\{B_\alpha \big| \alpha \in \Lambda\}$ such that $(\forall \alpha \in \Lambda)(B_\alpha \subseteq A\alpha)$ and $\bigcup_{\alpha \in \Lambda} A_\alpha = \bigcup_{\alpha \in \Lambda} B_\alpha$.

Injective Comparability (IC)

Given any two sets A, B, either there is an injection from A into B, or there is an injection from B into A (i.e. either $|A| \leq |B|$ or $|B| \leq |A|$).

Surjective Comparability (SC)

Given any two sets A, B, either there is a surjection from A onto B, or there is a surjection from B onto A.

Finite Powerset Principle (FP)**:**

Every infinite set X is equipotent to its *finite powerset* $[X]^{<\aleph_0}$.

Every set can be equipped with a group structure (GS)

Every set can be equipped with an abelian group structure (AG)