

MATH 355 LECTURE NOTES
CHAPTER 3: THE CUMULATIVE HIERARCHY

KAMERYN J. WILLIAMS

1. TROUBLE IN PARADISE

Way back in Chapter 0 we said that a set is a well-defined collection of objects. The trouble with this definition is that allows paradoxical objects—sets whose existence allows you to derive a contradiction. This is bad, because the logics we use in mathematics allow you to derive anything from a contradiction. We don't want this. (But there is a silver lining: it makes it really easy to prove the Riemann hypothesis, $P = NP$, $P \neq NP$ and other big open conjectures!¹)

The most famous of these paradoxical objects is due to the philosopher Bertrand Russell.

Theorem 1 (Russell's paradox). *If there is a set of all sets then you can derive a contradiction.*

Note that by the Chapter 0 definition, the collection $V = \{x : x \text{ is a set}\}$ is a set. So Russell demonstrates a problem with that definition.

Proof. Define $R \subseteq V$ as $R = \{x \in V : x \notin x\}$. This is a set by the Chapter 0 definition, since it's sensible to ask whether or not something is an element of R . For example, $\emptyset \in R$. You can also show that every ordinal is an element of R .

But is $R \in R$? If yes, then by the defining property of R we get that $R \notin R$. If no, then again by the defining property of R we get that $R \in R$. Either way, we get a contradiction. \square

Here's an alternate way to prove the result. Suppose there is a set V of all sets. Then $\mathcal{P}(V) \subseteq V$. So $|\mathcal{P}(V)| < |V|$, but Cantor's theorem tells us that $|V| < |\mathcal{P}(V)|$. Contradiction.

Similar arguments show that if there's a set of all ordinals or a set of all cardinals then you can derive a contradiction. Let me give the argument for ordinals, and leave the argument for cardinals to you.

Theorem 2 (Burali-Forti paradox). *If there is a set of all ordinals then you can derive a contradiction.*

Proof. Suppose that Ord , the collection of all ordinals, is a set. We know that Ord is well-ordered, so it would be an ordinal. So $\text{Ord} + 1$ would also be an ordinal. But then $\text{Ord} + 1 \in \text{Ord}$ and so $\text{Ord} + 1 < \text{Ord}$. But if α is any ordinal then $\alpha < \alpha + 1$, and we get $\text{Ord} < \text{Ord}$, a contradiction. \square

The Burali-Forti paradox is why in earlier chapters I was careful to only speak of the *collection* of all ordinals, and never call it a set. With this knowledge now in hand, let me remark that I made the same linguistic move elsewhere. Later in this chapter, after we've seen the basic axioms for set

Date: March 14, 2024.

¹A joke I once heard about an eminent set theorist: the reason he's been so prolific is that as a grad student he discovered a contradiction in the axioms, and he's spent his career hiding its use in complicated proofs by contradiction for his many fabulous theorems.

theory, we'll revisit the issue of what we mean when we talk about collections too big to be sets. For now, let me just say that none of the results in the earlier chapters are problematic.

Given these paradoxes we have a few options. One is to stick our heads in the sand and declare it doesn't matter and that we'll leave someone else to worry about it. While that option is popular among some, it is the coward's approach. Another approach is to throw out set theory entirely. But the elements of the discipline are ubiquitously useful in mathematics, so that's undesirable. The remaining option is to revise our flawed conception of set to produce something that doesn't fall prey to the paradoxes. At the same time, we don't want so restrictive a conception that we can't do anything with the resulting set theory. (Consider: you could trivially resolve the problem by declaring the only set is the empty set, but that's not useful for anything.) We want as much power as possible without crossing into paradox.

This was the option Russell himself took. His diagnosis of the problem was that the paradox was due to "vicious circles" like defining a set of all x so that $x \notin x$. Definitions like that are *impredicative*, meaning that the object being defined is among the objects being quantified over. Indeed, the contradiction arises precisely because we ask whether R itself falls under the definition. He worked to come up with a formal system for mathematics which would avoid vicious circles, in his epic three volume *Principia Mathematica* with Alfred Whitehead. This book marked the invention of *type theory*, where mathematics is stratified into a hierarchy of more and more complicated objects, with objects of larger type being built from objects of smaller type.

We know today that the two parts of Russell's response—avoiding impredicative definitions and having a hierarchy of objects—can be separated and that each one individually is sufficient to stave off Russell's paradox and its cousins. For example, W.V.O. Quine's formal set theory known as New Foundations takes the first response but not the second. And Zermelo–Fraenkel set theory, the dominant approach to set theory today, takes the second response but not the first. It is the ZF-style approach we take in this class.

Briefly, the big idea behind the ZF picture for sets is that the problem with the Russell set and other paradoxical objects is that they are too big to be sets. We will restrict sets to only be those collections which can be found by iterating the powerset operation along an ordinal. This will avoid the paradoxes like Russell's because the 'sets' involved are too big to be obtained in that way.

Next section contains an intuitive sketch of this view of what sets are, the *cumulative hierarchy view*. After that we'll jump into the formal details and understand how the basic axioms of set theory describe this picture of what sets are.

Exercises.

- (1) Show that if there is a set of all cardinals you can derive a contradiction.

2. THE CUMULATIVE HIERARCHY, INFORMALLY

Let me start with a caution sign: This is not a history of mathematics text. The history of the development of set theory is more complicated than what is sketched here. It took decades for these concepts to be understood and to crystallize into the forms we now use today. The point of this section is not to give you a nuanced understanding of history, but rather to convey an intuitive picture of the mathematical structure we will study in this chapter.²

Namely this structure is the *cumulative hierarchy of sets*. The trouble with the Chapter 0 definition of set is that it was too permissive; it allowed paradoxical objects to slip in. Rather than take a “top-down” approach where we legislate at once what all the sets are, we will take a “bottom-up” approach where we inductively build up more and more things to be part of the universe of sets.

To introduce the idea, let’s start with a bunch of mathematical objects that are not themselves sets—numbers and the like. Call these *urelements* and they form the stage 0 objects. Let’s not fuss about what we include with the urelements, as we’ll soon dispense with them entirely. Stage 1 introduces our first sets: any set of urelements. Phrased differently, stage 1 consists of all sets which are elements of $\mathcal{P}(U)$, where U is the set of all urelements. (We are justified in calling it a set because we added it in stage 1.) But is $\mathcal{P}(U)$ a set? We’d like it to be, but it’s not a stage 1 set. So let’s add it in stage 2. As a first attempt, say that stage 2 consists of all sets which are elements of $\mathcal{P}(\mathcal{P}(U))$.

This first try isn’t quite what we want. The trouble is it doesn’t allow sets of mixed type. We have the stage 1 sets, whose elements are urelements, and we have the stage 2 sets, whose elements are stage 1 sets. But we don’t have sets, like $\{7, \mathbb{N}\}$, whose elements can be either urelements or stage 1 sets. (These sets may be a bit weird, but if we’re going to have a general concept worth a damn we have to include some weird sets.) The fix is easy though: the stage 2 sets are the sets whose elements are either stage 1 sets or urelements. Let’s rephrase this to a simpler statement, one which is easier to generalize: the stage 2 sets are the sets whose elements are objects of stage < 2 .

In general, the inductive step goes like this. Having already constructed stages 0 through n , stage $n + 1$ gives us sets whose elements are objects of stage $< n + 1$.

This gives us all finite index stages, but we can keep going. We can have a stage after all the finite index stages, call it stage $\omega + 1$, to consist of sets whose elements are objects of stage n for some finite n . Then go again: stage $\omega + 2$ gives sets whose elements are objects of stage $< \omega + 2$. And so on to define stage $\omega + n$ for all finite n . And then we can go past another limit stage to get stage $\omega \cdot 2 + 1$ consisting of sets whose elements are objects of stage n or stage $\omega + n$ for some finite n .

Having the basic idea in mind, let me return to the remark that it doesn’t really matter what we take as the urelements. Suppose, for instance, that we had the real numbers among the urelements but not the complex numbers. Is this a problem? It turns not to be. We can build an isomorphic copy of \mathbb{C} from pairs of real numbers, so by going up a few stages in the iterative construction we get a copy of \mathbb{C} . (Technical detail: to do this, we need to be able to code ordered pairs as sets. There’s multiple ways to do this. Perhaps the best known is Kuratowski’s $(x, y) = \{\{x\}, \{x, y\}\}$, which we will see more of later this chapter.)

²This is not to disparage studying the history of set theory or, more broadly mathematics. I think it’s interesting and there’s some good sources to read on it. But that’s better as something to look at after you’ve studied the purely mathematical side of things.

What if we didn't start out with the real numbers in the urelements? We can construct \mathbb{R} as certain sets of rational numbers.³ And rational numbers can be constructed as equivalence classes of natural numbers. So if we have the natural numbers we can find a copy of \mathbb{R} as a set of sets of ... of sets of natural numbers, in some finite stage of our construction. Indeed, it turns out we don't need any non-set objects: as we'll explore later, any mathematical object can be coded by a set, possibly in a very complicated manner.

Because we can code any mathematical object as a set, it suffices to have sets as our only objects and to completely dispense with urelements. This is the standard approach among set theorists, because the parsimony makes some arguments simpler. And the cost is nil, because once you know you have a copy of a mathematical object you can just use it and not worry about where it came from. Like how after showing you can construct real numbers as Dedekind cuts you can just work with real numbers and not worry about the implementation details. But if you prefer to include non-set objects, there's no harm. It's known that the two approaches are equivalent, proving the same theorems.

With this simplified ontology, our starting stage 0 is to start with nothing, and build up from there. Let V_α denote the set of all sets which appear by stage α . (This is a set, because V_α appears by stage $\alpha + 1$.) We can then give the following definition by induction on Ord, which we have seen before and which amounts to a distillation of the above explanation:

- $V_0 = \emptyset$;
- $V_{\alpha+1} = \mathcal{P}(V_\alpha)$; and
- $V_\gamma = \bigcup_{\alpha < \gamma} V_\alpha$ if γ is a limit stage.
- Finally, the sets are exactly those objects which occur at some stage α of this iterative construction, where the stages are the ordinals.

Recall some properties of this hierarchy we proved back in Section 6 of Chapter 1. Each V_α is *transitive*, meaning that $x \in y \in V_\alpha$ implies $x \in V_\alpha$. And if $\alpha \leq \beta$ then $V_\alpha \subseteq V_\beta$. This is why it's sufficient to only take powersets in the definition; we get for free that previous stages are included and so don't have to worry about the mixed type problem.

Observe that at the limit stages we don't add any new sets, we just collect what we've seen so far. So new sets only appear at successor stages of the construction. (This is why, in the earlier gloss with urelements, we skipped limit ordinals for the indices.) Given a set x you can define its *rank* as the least ordinal α so that $x \in V_{\alpha+1}$ then the rank is always a successor.

In all, what the cumulative hierarchy gives us is a hierarchy for the universe of sets. This then gives natural sub-universes. For example, most of contemporary mathematics can be understood as appearing in $V_{\omega+\omega}$. But if you need sets of higher rank then you can go to a larger V_α . Experience tells us that any piece of mathematics appears in some, possibly very large, stage of this construction.

Why does the cumulative hierarchy definition of sets not fall prey to the paradoxes of size? Take Russell's paradox. To make it work we'd need that $R = \{x : x \notin x\}$ appears in some V_α . But you can check that $V_\alpha \notin V_\alpha$. So R would have to have rank $> \alpha$ for every α ; that is, R is not actually a set. What about Burali-Forti paradox? In chapter 1 we saw that $V_\alpha \cap \text{Ord} = \alpha$. So there's no stage in the construction at which we have all the ordinals. Thus we cannot get the paradox. In general, collections which are paradoxical because they are too large will never be sets, because their rank would have to be $> \alpha$ for every α .

³The two most popular options are via *Dedekind cuts* or via *Cauchy sequences*.

Does this mean we have an airtight proof that there are no other, more subtle paradoxes lurking? No. Gödel's incompleteness theorems teach us that we can never have complete confidence we have avoided paradoxes. But we do know that the known ones are avoided, with a uniform explanation for why. We also have nearly a century of work about this structure, and no one has found a contradiction in the basic axioms. Indeed, we can go much stronger than the basic axioms and still no one has found a contradiction, despite some very clever attempts.

3. THE AXIOMS OF SET THEORY

Now that we have idea what the subject matter of set theory is—the cumulative hierarchy—let’s isolate some basic axioms that describe this picture. After that we will see that these basic axioms suffice to do not just everything from the previous chapters, but most of mathematics as a whole.

We’ll take a parsimonious approach, where the only basic non-logical concepts are set and membership equality. Everything else will be defined in terms of these two concepts.⁴ Membership \in is a binary relation—whose properties will be explicated in the axioms—and sets are those objects which can be to the right of the membership relation. Indeed, the only objects we’ll have are sets, so sets could be defined as those objects x for which $x = x$.

Let’s see a few examples to get a taste for how to define other concepts in terms of these.

Example 3.

(1) $x \subseteq y$ means that for all objects $z \in x$ we have $z \in y$. Or in symbols:

$$x \subseteq y \quad \Leftrightarrow \quad \forall z \, z \in x \Rightarrow z \in y.$$

(2) $x \cap y = z$ means that for all objects w we have $w \in z$ if and only if $w \in x$ and $w \in y$.

(3) $x \cup y = z$ means that for all objects w we have $w \in z$ if and only if $w \in x$ or $w \in y$.

(4) $x \setminus y = z$ means that for all objects w we have $w \in z$ if and only if $w \in x$ but $w \notin y$.

(5) $\mathcal{P}(x) = y$ means that for all objects z we have $z \in y$ if and only if $z \subseteq x$. (As we saw earlier, we could write “ $z \subseteq x$ ” just using \in .)

We will see more examples as we go along.

Let’s start with some axioms that establish the basic properties of sets.

Definition 4. The *Extensionality* axiom defines membership. It asserts that two sets are equal if and only if they have the same members. In symbols

$$\text{(Extensionality)} \quad \forall x \forall y (x = y \Leftrightarrow \forall z \, z \in x \Leftrightarrow z \in y)$$

This axiom gives the meaning of equality for sets. It formally justifies the double inclusion method of proving two sets are equal: to prove $x = y$ prove both $x \subseteq y$ and $y \subseteq x$.

Definition 5. The *Pairing* axiom asserts that you can pair together any two objects into a set. That is, if x and y are any two objects there is a set $p = \{x, y\}$. In symbols

$$\text{(Pairing)} \quad \forall x \forall y \exists p \forall z (z \in p \Leftrightarrow [z = x \text{ or } z = y])$$

Note that “ $\forall z (z \in p \Leftrightarrow [z = x \text{ or } z = y])$ ” is how we expressed $p = \{x, y\}$ just using \in . By a similar process you can express $x = \{a\}$, $x = \{a, b, c\}$, and so on.

Definition 6. The *Union* axiom asserts that you can take a union of a set of sets. That is, if x is a set then so is $\cup x = \{y : y \in z \text{ for some } z \in x\}$. In symbols

$$\text{(Union)} \quad \forall x \exists u (y \in u \Leftrightarrow \exists z \in x \, y \in z)$$

Observe that using these axioms we can prove that $x \cup y$ is always a set for any x and y . Namely, Pairing tells us that $p = \{x, y\}$ is a set and Union tells us that $\cup p = x \cup y$ is also a set.

I hope that these three axioms are all completely uncontroversial. The next one needs a little bit of discussion, as it’s dangerously close to an inconsistent axiom.

⁴Using tools from logic, naturally. Usual practice is to treat equality as a logical concept, so we have that too.

Speaking informally the Separation axiom asserts that you can form subsets of a given set by separating out its elements satisfying some property. For example, if you have \mathbb{N} is a set you could define the set $\{n \in \mathbb{N} : n \text{ is a perfect square}\}$.

It will also be convenient to introduce some new terminology before we officially state this axiom. We know that some collections are too big to be sets. Since the only objects are sets, these can't be actual objects. But talk of collections such as Ord —the collection of all ordinals—is convenient. We treat a class as a short-hand for a definition we express just in terms of \in . For example, we will later define “ x is an ordinal” by

$$\forall y, y' \in x (y \in y' \text{ or } y = y' \vee y' \in y) \text{ and } \forall y \in x \forall z \in y z \in x.$$

We will treat “ $\alpha \in \text{Ord}$ ” as an abbreviation for that definition, so that we don't have to write it out every time. Similarly we can make sense of expressions like $x \cap \text{Ord}$ as referring to the set $\{y \in x : y \in \text{Ord}\}$, where this use of set-builder notation is shorthand for a longer definition just using \in .

In general, a *class* is an abbreviation for a definable property. Some classes are small enough that they are *co-extensive* with sets—meaning they have the same elements. For example, the empty class abbreviates the property $x \neq x$. It has the same elements as the empty set. Or given any set a there's a class for the property $x \in a$. If a class is not co-extensive with a set we call it a *proper class*. For example, Ord is a proper class, by the Burali-Forti paradox, and the class V of all sets is a proper class.

Remark 7. It's very convenient to think about classes as collections. This is fine to do, provided one keeps in mind a couple rules.

- (1) Classes cannot occur to the left of \in . It wouldn't make sense, for example, to write $\text{Ord} \in x$.
- (2) You cannot quantify over classes. It's fine to define things by quantifying over sets, but quantifying over classes would amount to quantifying over definitions, which is logically problematic.

With this in mind, let's officially state Separation.

Definition 8. The *Separation* axiom asserts that for any set x and any class A we have that $x \cap A = \{y \in x : y \in A\}$ is a set. In symbols:

$$\text{(Separation)} \quad \forall x \exists y y = x \cap A.$$

Remark 9. Didn't I just say that you can't quantify over classes, so why does this axiom seem to quantify over classes? To be precise, this is not a single axiom, but rather an *axiom schema* giving infinitely many axioms: for each definition $\varphi(x)$ giving a class A we have a different axiom. For more advanced uses, that Separation is infinitely many axioms matters. But for our purposes it's fine to think of as a single axiom.

This axiom justifies a use of set-builder notation to define sets. When you define, for example,

$$E = \{n \in \mathbb{N} : \exists k n = 2k\}$$

it's Separation that says it's allowed to separate \mathbb{N} into the smaller subset of evens. Compare, on the other hand, the definition

$$R = \{x : x \notin x\}.$$

This definition is not allowed under Separation, because we're not separating out elements from an extant set, we're trying to collect across the entire universe of sets. And it's good this definition isn't allowed, because it leads to a contradiction!

Separation also justifies the ideas that proper classes are those which are too big to be a set. Namely, if a class is contained in a set then that class is the same as a set.

Proposition 10. *Suppose a class A is a subset of a set x . Then A is co-extensive with a set.*

Thinking of this without the abbreviation, we're saying that if a property φ has the property that there is a set x so that $\varphi(y)$ implies $y \in x$ then $\{y : \varphi(y)\}$ is a set.

Proof. By Separation $x \cap A = \{y \in x : \varphi(y)\}$ is a set. Done. \square

The Separation axiom is quickly seen to be a consequence of the cumulative hierarchy view.

Proposition 11. *If $x \in V_\alpha$ and A is any class then $x \cap A \in V_{\alpha+1}$.*

Proof. We proved back in Chapter 1 that V_α is transitive—any element is a subset. So we get that $x \cap A \subseteq x \subseteq V_\alpha$ and so $x \cap A \in \mathcal{P}(V_\alpha) = V_{\alpha+1}$. \square

Separation also justifies the use of other basic set theoretic operations.

Proposition 12. *Let x and y be sets. Then $x \cap y$ and $x \setminus y$ are also sets.*

Proof. Because $x \cap y = \{z \in x : z \in y\}$ and $x \setminus y = \{z \in x : z \notin y\}$ are sets by Separation. \square

The next axioms we look at are about the height of the universe of sets. The idea was, remember, that we generated the universe by iterating the powerset operation along all the ordinals. But what is *all the ordinals*? How do we say how many ordinals there are?

First off, for this idea to even make sense in the first place we need that we can take powersets. The next axiom says we can do just that.

Definition 13. The *Powerset* axiom asserts that every set has a powerset. In symbols:

$$\text{(Powerset)} \quad \forall x \exists y \forall z (z \in y \Leftrightarrow z \subseteq x)$$

This y is the *powerset* of x , and we call it $\mathcal{P}(x)$.

Remark 14. The Extensionality axiom implies that the powerset of x is unique. If there are y and y' both satisfying that property then it must be that $y = y'$. So we are justified in talking about *the* powerset.

Like Separation, Powerset is readily seen to be true.

Proposition 15. *If $x \in V_\alpha$ then $\mathcal{P}(x) \in V_{\alpha+2}$.*

Proof. If $y \subseteq x$ then $y \subseteq V_\alpha$ because $x \subseteq V_\alpha$. Thus, $\mathcal{P}(x) \subseteq V_{\alpha+1}$ and so $\mathcal{P}(x) \in V_{\alpha+2}$. \square

The Powerset axiom implies we can always build the cumulative hierarchy one more step. If we have V_α then we also have $V_{\alpha+1} = \mathcal{P}(V_\alpha)$.

But it's not enough to be able to go +1. We want to go further, infinitely further. At minimum, there should be an infinite step in the process.

Proposition 16. *Every element of V_ω is finite.*

Proof. By induction prove that $|V_n| = 2^{2^{\dots^2}}$ (with n many 2s). And any element of V_ω is a subset of some V_n , and any subset of a finite set is finite. \square

Thus, if we want infinite objects such as \mathbb{N} or ω_1 —and we do!—then we need transfinite steps in the cumulative hierarchy.

The next axiom says that there's a set that's a copy of \mathbb{N} , and hence an infinite ordinal. And once we have ω we can recursively build up V_ω and then beyond. But we have to be careful how to phrase this idea; how do you say that something is infinite without already having \mathbb{N} as an object?

Definition 17. The *adjunction* of a set x with itself is the set $x \cup \{x\}$. Call a set I *adjunction inductive* if $\emptyset \in I$ and for every $x \in I$ we have that $x \cup \{x\} \in I$.⁵

Note that the Pairing and Union axioms imply that the adjunction of x with itself always exists. (Why?) Think of this as a construction of \mathbb{N} by iterating the successor operation starting from 0. We need a set to be zero; we pick the empty set. Then we need to say what $n + 1$ is; we say it's $n \cup \{n\}$. This way of formalizing $+1$ with just sets should look familiar if you think back to Chapter 1.

Definition 18. The *Infinity* axiom asserts that there exists an adjunction inductive set. In symbols:

$$\text{(Infinity)} \quad \exists x (\emptyset \in x \text{ and } \forall y \in x \ y \cup \{y\} \in x)$$

Three comments. First, we're really starting to see the advantage of using defined relations and not breaking everything down to just \in . You could write the Infinity axiom just using \in , but you'd need a lot more quantifiers and it'd be awful to read.

Second, this is the first and only axiom of set theory which begins with an existential quantifier. Every other one begins with a universal quantifier. So this axiom also plays the role of saying there exist any sets at all!

Proposition 19. *Assuming Infinity + Separation, the empty set exists.*⁶

Proof. By Infinity there exists a set, call it y . So $\{x \in y : x \neq x\} = \emptyset$ is also a set by Separation. \square

Third, once we have that there's any adjunction inductive set we can prove there's a smallest one.

Proposition 20. *There is a smallest adjunction inductive set. That is, there is an adjunction inductive set I which is a subset of every adjunction inductive set.*

Proof. Let J be an adjunction inductive set, which exists by the Infinity axiom. Define

$$I = \{x \in J : x \text{ is in every adjunction inductive set}\},$$

which is a set by Separation. It's clear by construction that I is a subset of any adjunction inductive set, so we have only to prove that I is adjunction inductive. First, note that $\emptyset \in I$ because \emptyset is an element of every adjunction inductive set. Now assume $x \in I$. Let K be any adjunction inductive set. Then $x \in K$ so $x \cup \{x\} \in K$. Since this holds for any K , we get $x \cup \{x\} \in I$. \square

One of the problems in the Chapter 1 problem set was to prove that any linear order with a successor operation and satisfying induction was isomorphic to \mathbb{N} . This smallest adjunction inductive set I is such a linear order, so it's isomorphic to \mathbb{N} . (Okay, we didn't define I as a linear order. But you can make it one: set $x < y$ if y is obtained from x by iterating adjunction.) Thus we have that \mathbb{N} is a set. Indeed, this I is exactly the ordinal ω . (But let's put off to talking about how we get ordinals in general until after we've seen more axioms.)

⁵In most texts, such I is simply called *inductive*. I use the slightly nonstandard terminology to emphasize that it's a specific implementation of $+1$ we're looking at.

⁶Being really careful, calling it *the* empty set requires Extensionality.

Definition 21. The axioms listed so far—Extensionality, Pairing, Union, Separation, Powerset, and Infinity—make up the theory known as *Zermelo set theory*, denoted Z .

The next axiom we’ll look at says there’s even more stages to the construction of the universe than just ω many. When you first see it, it’s probably not clear why that’s something it implies. To motivate it, let’s look at a toy model.

Theorem 22. *Every axiom of Z is true in $V_{\omega+\omega}$.*

Informal Proof. A completely formal proof of this fact relies on a careful definition of what it means for an axiom of set theory to be true in a set. This definition (or rather, a generalization to axioms and structures of any kind) is a building block of the area of mathematical logic known as *model theory*. But we will proceed with a naive notion of truth.

Extensionality is true because it’s true for all sets, not just those in $V_{\omega+\omega}$. Pairing is true because if $x, y \in V_\alpha$ then $\{x, y\} \in V_{\alpha+1}$. (Why?) Union is true because if $x \in V_\alpha$ then $\cup x \in V_\alpha$. (Why?) Separation is true because if $x \in V_\alpha$ then $y = \{z \in x : \varphi(z)\}$ is a subset of V_α and hence in $V_{\alpha+1}$. Powerset is true because if $x \in V_\alpha$ then $\mathcal{P}(x) \in V_{\alpha+1}$. And Infinity is true because ω is adjunction inductive. \square

Remark 23. There is a subtlety which must be addressed. When we say an axiom is true in $V_{\omega+\omega}$, we mean that if you restrict the quantifiers to only quantify over $V_{\omega+\omega}$ then the statement is true. But when you restrict quantifiers you might change what’s happening. With Separation, the set defined by $\{n \in \omega : \varphi(n)\}$ can change based on whether φ quantifies over all sets, or just over the sets in $V_{\omega+\omega}$. For example, consider $\varphi(n)$ to be “ $\omega \cdot n$ exists”. Then $\varphi(n)$ is true for every $n \in \omega$ if we quantify over the whole universe, but only $\varphi(0)$ and $\varphi(1)$ are true if we quantify just over $V_{\omega+\omega}$.

This subtlety doesn’t matter for our proof. We still get Separation is true anyway, because whatever subset of $x \in V_{\omega+\omega}$ is defined by a property φ , it’s still a subset of x so it’s still in $V_{\omega+\omega}$. But if you look at less thick structures then the difference does matter.

As an extreme example of this: the property “ x is uncountable” can change from true to false if you restrict where you quantify.

The upshot of this theorem is that Z cannot prove that $V_{\omega+\omega}$ exists. After all, every axiom of Z is true in $V_{\omega+\omega}$ and yet $V_{\omega+\omega}$ is not an element of $V_{\omega+\omega}$. If we want to guarantee the universe grows beyond $\omega + \omega$ many stages we need stronger axioms. The Replacement axiom gives us this.

Replacement is about class functions. Remember that we are thinking of classes as abbreviations for definable properties. So a class function is an abbreviation for a property $\varphi(x, y)$ that defines a function—for each x there is at most one y so that $\varphi(x, y)$ is true.⁷ It’s just more convenient to write $F(x) = y$ rather than talk about definitions.

Definition 24. “ F is a class function” is an abbreviation for $F(x) = y$ is defined by a formula $\varphi(x, y)$, and for every x there is at most one y so that $\varphi(x, y)$ is true. We can similarly talk about class functions of more than one input.

For example, the function $x \mapsto |x|$ is a class function. For another example, $\alpha \mapsto V_\alpha$ is a class function. The arithmetic operations on cardinals and ordinals are all class functions.⁸

⁷If there were multiple y for one x then there wouldn’t be a unique choice for $F(x)$ —it wouldn’t be a function.

⁸There’s something to be said here. Many of these functions were defined by recursion, so are they really class functions—i.e. are they really definable? It turns out that a consequence of Replacement is that definitions of class functions by transfinite recursion are valid.

Definition 25. The *Replacement* axiom asserts that the image of a set under a class function is a set. In symbols, with F being a class function:

$$\text{(Replacement)} \quad \forall a \exists b \forall x \in a \exists y \in b F(x) = y.$$

Remark 26. Again, we aren't actually quantifying over classes. Technically this is infinitely many axioms given by a single schema.

Remark 27. The set b Replacement says exists isn't guaranteed to be *exactly* the image of a under F . There could be other elements of b . But we can use Separation to cut this set down to just the image of a under F .

The point is, it's enough to say that there's a set which contains the image of a , so we use this simpler thing to be the official axiom.

Replacement justifies one of the variants of set-builder notation. You might define, for example, a set

$$\{V_{\omega+n} : n < \omega\}.$$

This set is the image of the function $n \mapsto V_{\omega+n}$. Replacement exactly says that it really is a set.

Let's see an example of how Replacement gets applied.

Proposition 28. *Let (X, \leq) be a well-ordered set, with ξ its order type. Then V_ξ exists.*

Proof. By doing transfinite recursion along X we can define the sequence $\langle V_\alpha : \alpha < \xi \rangle$. Sequences are, remember, just functions, so we have a class function. Replacement tells us that its image $I = \{V_\alpha : \alpha < \xi\}$ exists. Now we split into cases. (Zero) If $\xi = 0$ then $V_\xi = \emptyset$ exists. (Successor) If $\xi = \alpha + 1$, then we know V_α exists because it's an element of I . So $V_\xi = \mathcal{P}(V_\alpha)$ exists by Powerset. (Limit) If ξ is limit then $V_\xi = \cup I$ exists by Union. \square

As a specific example, using Replacement we can prove that $V_{\omega+\omega}$ exists. Using the axioms of Z we can get a well-order of order type $\omega + \omega$: order the natural numbers as evens followed by odds. The proposition then implies that $V_{\omega+\omega}$ exists. In general, Replacement implies that if you can define a set of ordertype ξ then you can build the V_α hierarchy up through ξ .

Contrast with the following fact.

Observation 29. *Z cannot prove that $V_{\omega+\omega}$ exists.*

Proof. The reason for this is, we saw earlier that every axiom of Z is true in $V_{\omega+\omega}$. But if Z could prove that $V_{\omega+\omega}$ exists, then that would be true in every structure satisfying all of Z . But $V_{\omega+\omega}$ is not an element of itself. \square

The upshot is, Replacement guarantees that that the iterative hierarchy goes as high as possible. It does a lot more than that, and there's still much to say, but let's put that off for now.

What Replacement doesn't tell us is that the universe of sets consists *only* of the sets in the V_α hierarchy. To say there no sets outside we need a different axiom. This is given by the Foundation axiom.

When we talked about well-orders, their defining property was well-foundedness. But you don't need a relation to be a linear order for well-foundedness to make sense. You can ask if any binary relation is well-founded.

Definition 30. Let \vdash be a binary relation on a set X . (That is, $\vdash \subseteq X^2$.) Then \vdash is *well-founded* if it satisfies: for any nonempty $Y \subseteq X$ there is $m \in Y$ so that no $y \in Y$ has $y \vdash m$.

To match with our definition for linear orders, saying that m is the minimum of Y is saying that there is no $y \in Y$ with $y < m$.

Definition 31. The *Foundation* axiom says that the membership relation \in is well-founded. That is, if x is any nonempty set then there is $m \in x$ so that no $y \in x$ has $y \in m$. In symbols:

(Foundation) $\forall x x \neq \emptyset \Rightarrow \exists m \in x \forall y \in x y \notin m$.

Definition 32. *Zermelo–Fraenkel set theory* ZF is the set theory with axioms Extensionality, Pairing, Union, Separation, Powerset, Infinity, Replacement, and Foundation. That is, ZF is Z plus Replacement and Foundation.

We used well-foundedness to see that you can do induction along well-orders. For this to work, you don't need the linear order part, you only need well-foundedness. Of course, if your relation is not a linear order then it's not going to be ordered like a line. Instead, the induction will be along a more complicated pattern, such as a tree. That's fine, we can do that.

Theorem 33. *Let \vdash be a well-founded relation on a set X . Then, induction is valid along \vdash . That is, suppose $Y \subseteq X$ has the property that for all $y \in Y$ if every $x \vdash y$ has $x \in Y$ then $y \in Y$. Then, $Y = X$.*

Proof. This is the same as what we saw before with well-orders. Suppose otherwise toward a contradiction. By well-foundedness there is a minimal element m of $X \setminus Y$. That is, $m \notin Y$ but $x \in Y$ for every $x \vdash m$. But then we'd get that $m \in Y$, a contradiction. \square

This fact also holds for well-founded class relations, provided that the relation is “locally” a set.

Theorem 34. *Let \vdash be a well-founded class relation on a class X with the property that for every $x \in X$ we have $\{y : y \vdash x\}$ is a set. Then induction is valid along \vdash . That is, suppose $Y \subseteq X$ has the property that for all $y \in Y$ if every $x \vdash y$ has $x \in Y$ then $y \in Y$. Then, $Y = X$.*

The most useful instance of this theorem is when the relation is the order relation on Ord.

Remark 35. Again, we seem to be quantifying over classes. We avoid this similar to how we avoided it with the Separation and Replacement axioms. Technically, this isn't a single theorem, but rather a *theorem schema* giving infinitely many theorems. For any definitions $\varphi(x, y), \chi(x), v(x)$ giving appropriate \vdash, X , and Y we have that $\chi(x)$ iff $v(x)$.

Proof. Suppose $Y \neq X$ and pick $x \in X \setminus Y$. Then \vdash restricted to $X_0 = \{y : y \vdash x\} \cup \{x\}$ is a well-founded relation on a set. So by induction on sets we get that $Y \cap X_0 = X_0$ and so $x \in Y$, a contradiction. \square

Definition 36. Let V denote the class of all sets. Since our only objects are sets, we may formally define $V = \{x : x = x\}$.

Theorem 37 (In ZF). *Every set is in V_α for some α . That is,*

$$V = \bigcup_{\alpha \in \text{Ord}} V_\alpha.$$

Proof. We prove this by induction on \in . To that end, fix a set x and assume that for every $y \in x$ there is an ordinal α_y so that $y \in V_{\alpha_y}$. Without loss of generality we assume α_y is the smallest ordinal for which this happens. Then $y \mapsto \alpha_y$ is a class function, so Replacement tells us that the set $\{\alpha_y : y \in x\}$ exists. Let $\alpha = \sup_{y \in x} \alpha_y$. Then $y \in V_\alpha$ for all $y \in x$. That is, $x \subseteq V_\alpha$ and so $x \in V_{\alpha+1}$.

By induction, we've proved this for every set x . Done. \square

By a similar argument we can show that there is a *rank function* on the ordinals, so that elements of a set must have a smaller rank.

Definition 38. Define the *rank function* $\rho : V \rightarrow \text{Ord}$ on sets as:

$$\rho(x) = \sup\{\rho(y) + 1 : y \in x\},$$

where we stipulate that $\sup \emptyset = 0$ (and so $\rho(\emptyset) = 0$).

This rank function more-or-less tells you which V_α first has x appear within. Let me state this result without proof.

Fact 39. $\rho(x)$ is the least α so that $x \in V_{\alpha+1}$.

There's still one more axiom to go, Zermelo's axiom of choice. But before we see it, let's take some time to see what mathematics we can do with what we have so far.

Exercises.

- (1) Write out as many axioms as you can using just \in and $=$, with no abbreviations or defined concepts.

4. WORKING WITH THE AXIOMS

In the previous section we saw the axioms of **ZF** and talked about why they are true in the cumulative hierarchy. What we do in this section is see that they are sufficient to do a lot of mathematics.

So far what we have seen is that the axiom of Infinity guarantees we have a copy of \mathbb{N} in the cumulative hierarchy. But we need a lot more than just that to do mathematics.

A lot of mathematics is built on top of the basic notion of an ordered pair. Functions and relations both built from ordered pairs. And these are used to define many notions in mathematics: groups, isomorphisms, orders, equivalence classes, and many, many more. For example, once we have ordered pairs we can define the cartesian product of a set using ordered pairs. A relation on X is then a subset of $X \times X$ and a function $X \rightarrow Y$ is a subset of $X \times Y$ satisfying certain properties. We then define linear orders as relations satisfying appropriate axioms, and an isomorphism between linear orders is a function so that [blah blah].

But how do we define an ordered pair using just sets? After all, the whole point of an ordered pair is to be ordered, while sets are not ordered. If we want a way to code ordered pairs as sets, we need a way to code the first element of the pair versus the second. There's a few ways to do this. Kuratowski's is especially nice.

Definition 40 (Kuratowski). The *Kuratowski ordered pair* of x and y is the set

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

An object is the first element of $(x, y)_K$ if it is an element of every element of $(x, y)_K$ and it is the second element if it is in exactly one element of $(x, y)_K$.

Usually we will just write (x, y) .

For example, the pair (x, x) is

$$\{\{x\}, \{x, x\}\} = \{\{x\}, \{x\}\} = \{\{x\}\}.$$

Then x is the first element of that pair because it's an element of all the sets in the pair, and x is also the second element of that pair because it's an element of precisely one set in the pair.

Remark 41. You could alternatively formalize set theory with ordered pairs as a primitive notion, similar to how 'set' and 'membership' were taken as primitive notions. What Kuratowski's definition shows is that these two approaches are equivalent.

If they are equivalent, why go with the formalization where ordered pairs aren't a primitive notion? One reason is parsimony; the fewer basic concepts the fewer things to check when reasoning about your structure. Another reason is it enables such a clear picture of the cumulative hierarchy; if pairs were different kinds of objects than sets we'd have a more complicated definition than iterating the powerset operation.

A third reason is that this conceptual simplicity helps to avoid accidental problems. For example, in Bourbaki's 1949 axiomatization of set theory, they have ordered pairs as basic objects, with a corresponding axiom asserting that for any two objects x, y the ordered pair (x, y) exists. But they don't assert that the unordered pair $\{x, y\}$ exists, and indeed their system cannot prove it does!⁹

Lemma 42. *If x and y are sets then their Cartesian product $x \times y = \{(a, b) : a \in x \text{ and } b \in y\}$ is a set.*

⁹A. R. D. Mathias has a short but not particularly readable paper where he proves this: <https://www.dpmms.cam.ac.uk/~ardm/uprbou49pub.pdf>.

Proof. Note that if $a \in x$ and $b \in y$ then $\{a\}, \{a, b\}$ are elements of $\mathcal{P}(x \cup y)$, which exists by Pairing, Union, and Powerset. So (a, b) is an element of $\mathcal{P}(\mathcal{P}(x \cup y))$. Applying Separation to cut down $\mathcal{P}(\mathcal{P}(x \cup y))$ we get that $x \times y$ exists. \square

Remark 43. It maybe seems a tad silly to prove that cartesian products exist by using powerset a couple times. For example, $\omega \times \omega$ is countable but $|\mathcal{P}(\mathcal{P}(\omega))| = \beth_2$ is pretty big. You could give a more economical argument which doesn't use powersets. But it's quicker to just use a big hammer.

With ordered pairs now can build up abstract objects in mathematics using functions, relations, sets, etc. But we don't just want abstract groups or topological spaces, we also want the familiar concrete objects of mathematics— \mathbb{R} and so on.

We already saw that the Infinity axiom gives us a copy of \mathbb{N} in the sets. We will just call that copy \mathbb{N} . (After all, when working with \mathbb{N} it doesn't matter what the elements of \mathbb{N} are, it's only their structure that matters.) We can construct more number systems using the idea of equivalence relations. Let's first go from \mathbb{N} to \mathbb{Q}^+ , the non-negative rationals, to illustrate the concept.

The idea is, \mathbb{Q}^+ consists of fractions $\frac{n}{d}$, where n and $d \neq 0$ are natural numbers. This is just a pair (n, d) written differently. And we need to say when two fractions are the same. This is done by the familiar algebra trick of cross multiplying:

$$\frac{n}{d} = \frac{m}{e} \quad \Leftrightarrow \quad ne = md.$$

So we define that two pairs (n, d) and (m, e) represent the same rational when $ne = md$. It's straightforward enough to check that this is an equivalence relation on the pairs. And then we can define the arithmetic operations using the usual rules for fractions. For example,

$$\frac{n}{d} + \frac{m}{e} = \frac{ne + md}{de},$$

so we say that $(n, d) + (m, e) = (ne + md, de)$. You can check that this operation is well-defined on equivalence classes: two ways of writing the same fraction give the same sum at the end. Multiplication is defined similarly, as are subtraction and division.

Carry out this work I just hand waved over, and what you've done is built a copy of \mathbb{Q}^+ out of equivalence classes on $\omega \times (\omega \setminus \{0\})$. We know ZF proves cartesian products exist, and we have Separation, so in all this gives a copy of \mathbb{Q}^+ in the cumulative hierarchy.

You then go from \mathbb{Q}^+ to all of \mathbb{Q} by a similar process to add negatives. You can think of an arbitrary fraction as given by $\frac{n}{d} - \frac{m}{e}$ for two non-negative fractions $\frac{n}{d}$ and $\frac{m}{e}$. So we represent the whole \mathbb{Q} as pairs of fractions (which are themselves pairs of natural numbers, which are themselves finite von Neumann ordinals), with an appropriate equivalence relation. And then you can define the arithmetic operations on these equivalence relations. All in all, we find a copy of \mathbb{Q} in the universe of pure sets.

Alternatively, you could define \mathbb{Z} from \mathbb{N} by the same idea, and then go from \mathbb{Z} to \mathbb{Q} . That also works, but I went to fractions first because I think it's more intuitive.

Going from \mathbb{Q} to \mathbb{R} requires a new idea. (This makes sense— \mathbb{R} is countable so we can't build it from a countable object like equivalence classes over \mathbb{Q} .) The most well known constructions of \mathbb{R} from \mathbb{Q} are Dedekind's, using *Dedekind cuts*, and Cantor's, using equivalence classes of Cauchy sequences of rational numbers. Briefly: Dedekind's idea is that a real x can be represented by the set $\{q \in \mathbb{Q} : q \leq x\}$ and Cantor's idea is that a real x can be represented by a series

$$\sum_{n=0}^{\infty} q_n$$

of rationals q_n which converges quickly to x .

Exploring one (or both) of these constructions is a topic for an analysis class, or as a project for this class. So I will leave it at that quick gloss, which hopefully suffices to convince you that we can find a copy of \mathbb{R} in the pure sets. And once we have \mathbb{R} we of course get \mathbb{C} by representing $a + bi$ as a pair (a, b) of reals. So we have all our familiar number systems, and can build upward on them to get continuous functions, derivatives, \mathbb{C} -vector spaces, Banach spaces, and other objects of mathematics.

This isn't enough to do all of mathematics, of course. Some areas of mathematics aren't concerned with concrete structures like \mathbb{R} but concern e.g. abstract groups or topological spaces. These can also be coded in the pure sets as well. After all, with a group you don't care what its elements are, you only care about the structure. So building a copy of S_3 (the permutation group on 3 elements) can be done just fine in the pure sets—we don't care which 3 element set we build the permutation group on!

All in all, (a copy of) any object of non-set theoretic mathematics can be found somewhere in some V_α . Usually this is some $\alpha < \omega + \omega$, but sometimes you have to go further. And often this coding of X es into sets is not interesting to specialists in X , but you can nevertheless do it. This is why it's standard practice to take ZFC—ZF plus the axiom of choice, coming soon—as the standard foundation for mathematics. Whatever your preferred branch of mathematics, you can think of it as happening within set theory.¹⁰

Knowing now that we can do non-set theory in ZF, let's look at set theory in particular.

Our two important basic objects are ordinals and cardinals. The latter we defined as certain ordinals,¹¹ so what we need is to make sure we can talk about ordinals in the ZF framework. There's a few ways to implement a definition of ordinals in the cumulative hierarchy, the most convenient is due to von Neumann.

Definition 44 (von Neumann). An ordinal is a transitive set which is linearly ordered by \in . That is, α is an ordinal if

- (Transitive) $\gamma \in \beta \in \alpha$ implies $\gamma \in \alpha$; and
- (Trichotomy) If $\beta, \gamma \in \alpha$ then exactly one of the following holds: $\beta \in \gamma$, $\beta = \gamma$, or $\gamma \in \beta$.

If α and β are ordinals then $\alpha < \beta$ if $\alpha \in \beta$. And as usual $\alpha \leq \beta$ if and only if $\alpha < \beta$ or $\alpha = \beta$.

Remark 45. Anti-symmetry—that you can't have $\alpha < \beta < \alpha$ —and irreflexivity for the strict order—that $\alpha \not< \alpha$ —both are free. This is because Foundation says that \in is well founded. Having $\alpha \in \beta \in \alpha$ or $\alpha \in \alpha$ would let you construct an infinite descending in the membership relation:

$$\alpha \ni \alpha \ni \alpha \ni \alpha \ni \alpha \ni \alpha \ni \dots$$

Let's see why this definition exactly matches the recursive definition that an ordinal α is the set of all ordinals $< \alpha$. That the definition of the orders match is immediate: according to either version, $\alpha < \beta$ if and only if $\alpha \in \beta$. Back in Chapter 1 we saw that ordinals (by the previous definition) were always transitive. Finally, the reason ordinals by von Neumann's definition are

¹⁰I must footnote this with a caveat. This can really only be a statement about mathematical objects that have already been invented. How could we possibly know whether this is some new object that can't be fit into the universe of sets?

We can't, of course. It's always possible that in the future we'll discover some new thing that overturns our current understanding. But based on the fact that any mathematical object can be coded in the pure sets, it seems reasonable to assume that will continue to happen.

¹¹This uses the axiom of choice! But we will get to that soon.

well-ordered is that, by Foundation, \in is well-founded. So if a set is linearly ordered by \in then it is well-ordered by \in .

One nice fact about von Neumann ordinals is that there is a particularly canonical isomorphism from a well-order to its corresponding ordinal.

Theorem 46 (Mostowski collapse theorem for well-orders). *Let (A, \leq) be a well-order. Then the recursive definition along A given by $\pi(a) = \{\pi(b) : b < a\}$ gives an isomorphism from A to a von Neumann ordinal α .*

Proof. π is defined by transfinite recursion along a well-order, so this gives a valid definition of a function. Let α be the range of π . We need to see that α is an ordinal and that (A, \leq) is isomorphic to (α, \in) . First, note that by construction $b < a$ iff $\pi(b) \in \pi(a)$. That α is transitive is then clear: if $\pi(c) \in \pi(b) \in \pi(a)$ then $c < b < a$ so $c < a$ so $\pi(c) \in \pi(a)$. And α is linearly ordered by \in because $<$ was a linear order. \square

There's a more general version of the Mostowski collapse theorem, not just about linear orders. One of the problems in the problem set is to prove it.

With the von Neumann definition of the ordinals in hand, we almost have everything we need to get that everything in Chapter 1 can be done in ZF. The one missing piece, that ZF allows transfinite recursion, will be seen next. And once we see the axiom of Choice we'll then we'll also have everything from Chapter 2.

So why are definitions by transfinite recursion, like the one for the Mostowski collapse theorem or the definition of V_α , valid in ZF?

The point is, it's Replacement. For example, suppose you want to define $V_{\omega+\omega}$, and it's natural to do this by recursion. This amounts to defining a function with domain $\omega + \omega$, and it's Replacement which says the image $\{V_\alpha : \alpha < \omega + \omega\}$ of this function is a set, whence you can get $V_{\omega+\omega}$. Without Replacement, we can't guarantee we can do this, like we saw earlier.

Theorem 47 (Recursion theorem on Ord). *Let $G(\alpha, f)$ be a class function whose inputs are an ordinal α and a function f with domain α . Then there is a class function $F(\alpha)$ whose input is an ordinal α so that, for all $\alpha \in \text{Ord}$,*

$$F(\alpha) = G(\alpha, F \upharpoonright \alpha).$$

A few comments. First, this theorem implies that you can do recursion along a fixed ordinal β —just define G to do nothing for inputs $\geq \beta$.

Second, recall that class functions are just abbreviations for definable mappings. So what this says is that if I give you a definition which gives a function G (even if it's too big to be a set), then you can turn that definition into a new, more complicated definition for F . Note that by $F \upharpoonright \alpha$ we mean the set which is co-extensive with the class function that is the restriction of F to α . We can't use the class function itself as an input, because inputs have to be sets.

Third, again this is a theorem schema giving infinitely many theorems.

Proof. We prove this by doing the same thing we did back in Chapter 1, but with a little care to see that ZF is enough to do the required work. To that end, let's repeat the argument, with annotations as footnotes.

Say that a function f with domain β is a *partial solution below β* if $f(\gamma) = G(\gamma, f \upharpoonright \gamma)$ for all $\gamma < \beta$.

Claim: Any two partial solutions agree on their shared domain. In particular, for each β there is at most one partial solution below β .

Suppose f and f' are two partial solutions with a shared domain β . Use induction to prove $f(\gamma) = f'(\gamma)$ for every $\gamma < \beta$. Namely, if f and f' agree below γ then $f \upharpoonright \gamma = f' \upharpoonright \gamma$. Then $f(\gamma) = G(\gamma, f \upharpoonright \gamma)$ and $f'(\gamma) = G(\gamma, f' \upharpoonright \gamma)$ are the same.

Claim: For every β there exists a partial solution below β .

By induction. The empty function is a partial solution below 0, establishing the zero case. For the limit case, let f_γ be the partial solution up to γ for each $\gamma < \beta$. By the previous claim we know the f_γ 's agree on their common domain, so $f_\beta = \bigcup_{\gamma < \beta} f_\gamma$ is a function.¹² Then f_β is a partial solution below β .

For the successor case, let f_β be the partial solution below β . To extend to a partial solution below $\beta + 1$ set $f_{\beta+1} = f_\beta \cup \{(\beta, G(\beta, f_\beta))\}$.

We can now define the full solution F . Namely, set $F(\beta) = x$ if any partial solution f with β in its domain has $f(\beta) = x$. This defines a function¹³ because all partial solutions agree, and its domain is Ord because there is a partial solution below every β . \square

It turns out that Replacement is actually equivalent to transfinite recursion. If you assume transfinite recursion—the theorem just proved—and the other axioms of ZFC, then you can prove Replacement. This is one of the problems for this chapter.

And just like you can have induction for well-founded relations in general, not just well-orders, you can have recursion as well. Let me state without proof two forms of this.

Theorem 48 (Transfinite recursion along a well-founded relation). *Fix (X, E) , where E is a well-founded relation on the set X . Let $G(x, f)$ be a class function whose inputs are any $x \in X$ and a function f with domain $E \downarrow x = \{y \in X : y E x\}$. Then there is a class function $F(x)$ whose input is $x \in X$ so that, for all $x \in X$*

$$F(x) = G(x, F \upharpoonright (E \downarrow x)).$$

Theorem 49 (Transfinite recursion along \in). *Let $G(x, f)$ be a class function whose inputs are any set x and a function f with domain x . Then there is a class function $F(x)$ whose input is a set x so that, for all x*

$$F(x) = G(x, F \upharpoonright x).$$

For example, when we defined the rank function on sets by recursion on \in it was this result that was being used in the background.

Let me close this section by mentioning a small fact proved by transfinite recursion.

Lemma 50. *Every set x is contained inside a transitive set t .*

Proof. The idea is, we turn x into a transitive set by adding in the elements of elements of x , the elements of elements of elements of x , and so on. This is formally done by recursion.

Define x_n for $n < \omega$ by recursion: start with $x_0 = x$ and set $x_{n+1} = x_n \cup \bigcup_{y \in x_n} y$. Then $t = \bigcup_{n < \omega} x_n$ is transitive. This is because if $y \in t$ then $y \in x_n$ for some n . But then any $z \in y$ is in $x_{n+1} \subseteq t$. \square

¹²This is where we use Replacement (and Union, I suppose), and it is the *only* place. To know that we can take the union of f_γ to get a set we have to know that $\{f_\gamma : \gamma < \beta\}$ is a set.

¹³Rather, a class function.

Exercises.

- (1) Check that saying (n, d) and (m, e) represent the same fraction if $ne = md$ is an equivalence relation.
- (2) Write down explicit definitions for $-$, \cdot , and $/$ for pairs representing rationals. Check that they, and the $+$ from the text, are well-defined on equivalence classes.

5. THE AXIOM OF CHOICE

Some of the axioms of ZF, such as Extensionality or Union, describe basic facts or operations about sets. Others, such as Infinity or Replacement, were motivated by the cumulative hierarchy perspective. In this section we see one last axiom, which is neither an elementary fact nor a distillation of a property of the cumulative hierarchy. But it is necessary for a lot of work in math, so we want to have it.

This axiom is Zermelo’s axiom of choice.

Definition 51. A *choice function* on a set x is a function f whose inputs are nonempty sets $y \in x$ so that $f(y) \in y$ for every nonempty $y \in x$.

Definition 52 (Zermelo, 1904). The *axiom of Choice* (AC) asserts that for any set x there is a choice function on x .

Definition 53. The set theory ZFC (Zermelo–Fraenkel set theory with the axiom of choice) is axiomatized by the axioms of ZF and the axiom of Choice.

Informally, what AC says is that if you need to make infinitely many choices in a proof, you can do so. Indeed, you already did this in one of the problems for Chapter 0.

Example 54. To prove that a countable union of countable sets is countable, you *chose* a bijection $A_n \rightarrow \mathbb{N}$ for each of your countable sets A_n . It was the axiom of choice that allowed these infinitely many choices. Namely, for each n let B_n be the set of bijections $A_n \rightarrow \mathbb{N}$ and let $b = \{B_n : n \in \mathbb{N}\}$. Then AC says that there is a choice function on b , and $b(B_n)$ gives us a bijection $A_n \rightarrow \mathbb{N}$.

Naturally, when using AC in proofs no one explicitly writes down all that. They just say something like “pick a foo for each bar”.

AC generalizes a consequence of Existential Instantiation, a logical principle. Remember that Existential Instantiation says that if $\exists x \varphi(x)$ is true then you may infer that $\varphi(a)$ is true for some newly introduced variable a . One common way this is applied is for the formula $\exists x x \in A$ (namely, that A is nonempty) to get an element of A . Repeating this you can get elements for finitely many sets A_0, A_1, \dots, A_n . But you can’t do it for infinitely many sets, as that would require an infinitely long deduction, and proofs are finite. AC says that you can make all your infinitely many choices simultaneously.

The axiom of choice was introduced by Zermelo in an article about his well-ordering theorem. His earlier proof of the theorem had faced criticism, so in response he wrote a new article carefully laying out just what assumptions he used to prove the theorem. AC was one of those axioms. Indeed, he proved that AC is equivalent to that theorem. This article is where Zermelo set theory Z was first set forth. It was later supplemented with Replacement and Foundation to get ZF, and adding AC gives ZFC.

Theorem 55 (Well-ordering theorem). *Every set can be well-ordered.*

Proof. The idea is, having a well-order of a set is equivalent to having a sequence indexed by an ordinal which enumerates every element of the set. Then we can set $x < y$ if the first index where x appears is before the first index where y appears. We construct such a sequence by using the axiom of choice to always pick the next element, going until we run out of space.

In more detail: fix a set X and let $\aleph(X)$, the smallest ordinal which doesn’t inject into X .¹⁴ Fix a choice function f for $\mathcal{P}(X)$. We construct an enumeration $\langle x_\xi : \xi < \beta \rangle$ of length some $\beta < \aleph(X)$

¹⁴It was a Chapter 1 problem to prove $\aleph(X)$ has that property.

by transfinite recursion along α . Namely, we set

$$x_\xi = f(X \setminus \{x_\eta : \eta < \xi\})$$

if $X \setminus \{x_\eta : \eta < \xi\}$ is nonempty, and otherwise x_ξ is undefined. Observe that if $\eta < \xi$ then $x_\eta \neq x_\xi$. And observe that if x_ξ is undefined then so is x_η for all $\eta > \xi$. Thus, the indices ξ for which x_ξ exists form an initial segment of α , say all $\xi < \beta$.

Note now that if there is any $\xi < \alpha$ for which x_ξ is undefined, it's because our sequence enumerated all of X and we had no more space to keep going. We are done if we can prove this happened. Towards a contradiction, assume that x_ξ is defined for all $\xi < \alpha$. Then, $\xi \mapsto x_\xi$ is a bijection $\alpha \rightarrow X$, contradicting that there's no injection $\alpha \rightarrow X$. Done. \square

Theorem 56. *Working in ZF, the well-ordering theorem implies AC.*

Proof. Let x be a set. Set $y = \cup x$. Note that if $a \in x$ then $a \subseteq y$. By the well-ordering theorem there is a well-order \leq on y . Now define a choice function f on x by setting $f(a)$ to be the least element of a according to the well-order \leq . Done. \square

Corollary 57. *Over ZF, AC and the well-ordering theorem are equivalent.* \square

Now that we have the well-ordering theorem, we get that everything from Chapter 2 can be done in ZFC.

There are many other equivalent forms of AC. I list a few here, without proof. (Some of the proofs are problems for you.) For those equivalences which are substantive theorems I've provided the names of who first proved them.

Theorem 58. *Over ZF, the following are all equivalent.*

- (1) *The axiom of choice;*
- (2) *(Ernst Zermelo) The well-ordering theorem;*
- (3) *(Friedrich Hartogs) Cardinal trichotomy, the assertion that for any two sets there's an injection from one to the other;*
- (4) *(Dorothy Wrinch) There are no sets A, B so that $\kappa < |B|$ whenever $\kappa < |A|$ but A and B are incomparable in cardinality;*
- (5) *(Alfred Tarski) For any infinite set A there's a bijection $A \rightarrow A^2$;*
- (6) *König's theorem (Theorem 49 from Chapter 2).*
- (7) *For any sets A, B there is an injection $A \rightarrow B$ if and only if there is a surjection $B \rightarrow A$.*
- (8) *For any surjection $f : A \rightarrow B$ there is a function $g : B \rightarrow A$ so that $f \circ g$ is the identity function.¹⁵*
- (9) *(Kazimierz Kuratowski) Zorn's lemma. (See the problem set for a statement.)*

There are many important theorems in mathematics which are provable in ZFC but not in ZF. Sometimes they need the full power of AC—if you assume the theorem you can prove AC—and other times they are weaker than full AC. We've seen a few of these theorems in set theory. Here's some examples outside of set theory.

Example 59. The following theorems are provable in ZFC but not ZF:

- (1) Every vector space has a basis.
- (2) The Hahn–Banach theorem in analysis.
- (3) Krull's theorem that rings have maximal ideals.

¹⁵For the category theoretically inclined: this is saying that every epimorphism splits in the category of sets.

- (4) Tychonoff's theorem for products of compact topological spaces.
- (5) The De Bruijn–Erdős theorem about graph colorings.
- (6) The Banach–Tarski theorem about paradoxical decompositions of the sphere.

There is a hierarchy of *choice principles*, principles which aren't provable in ZF but are provable in ZFC, which measure the strength of different theorems needing the axiom of choice. This is a possible topic for Chapter ω .

Exercises.

- (1) Check that AC is equivalent to the assertion that any cartesian product of nonempty sets is nonempty. [Hint: show that an element of $\prod_{i \in I} X_i$ is, up to some shuffling of indices, a choice function on $\{X_i : i \in I\}$.]

6. THE LIMITS OF THE AXIOMS AND INDEPENDENCE

Back in Chapter 2 I told you that the continuum hypothesis (CH) is independent of the basic axioms of set theory. Now that we know what those axioms are, let's revisit that topic.

First off, let's make clear just what it means to say something is independent of ZFC.

Definition 60. A statement φ is *independent* of ZFC if one of the following two equivalent conditions holds:

- (1) There is no proof of φ nor $\neg\varphi$ starting from the axioms of ZFC; or
- (2) There are structures Y and N which satisfy all the axioms of ZFC and Y satisfies φ while N satisfies $\neg\varphi$.

You can similarly define what it means for φ to be independent of some other set of axioms.

That (1) and (2) are equivalent is Gödel's *completeness theorem*, one of the foundational results in logic. (Not to be confused with his incompleteness theorem!) The second formulation of independence is often the nicest to think about, since it's not so clear how you go about showing there cannot be a proof of such-and-such.

Briefly, here's what Gödel and Cohen did to show that CH is independent

Gödel proved that you could thin out V to a thinnest possible universe of sets, which he called L . Like V this universe is built up iteratively. But rather than add in the full powerset at each stage, L adds in only a little bit more. He proved that if you assume ZF is true, then ZFC + GCH is true in L .

Cohen worked in the opposite direction. He proved that starting with V you could add in a "generic" object G and define an even larger universe $V[G]$, in which ZFC is true and CH is false. So L and $V[G]$ give structures witnessing that CH is independent of ZFC. (Indeed, you could pick a different generic object H and get that $V[H]$ satisfies ZFC + CH. So you can use Cohen's method to prove both parts of the independence.)

There's a lot of work to be done to understand the details of what they did, not the least how to actually make sense of what it means to add a new object outside the universe. Sadly we don't have the time to do so in this class.

Gödel's work is the start of *inner model theory* and Cohen invented *forcing*. These are two central topics in modern set theory, and they have been pushed to do a lot more than prove the independence of CH. We very much are still living in their shadows.

Gödel and Cohen both modified the "width" of the universe. Gödel thinned it down while Cohen fattened it out. But neither one changed the "height" of the universe—you neither add nor remove ordinals. You can also talk about changing the height of the universe. You can cut down V to some V_α , or you could imagine building even higher, using well-orders too long to exist in the current universe.

Let's see an example of this.

Definition 61. An uncountable cardinal κ is *inaccessible* if κ is regular and $2^\lambda < \kappa$ for all $\lambda < \kappa$. Recall that this last property is what we termed being a strong limit.

In Chapter 2 I said that the existence of a regular strong limit is independent of the basic axioms of set theory. Let's see why now.

Theorem 62. *All of the axioms of ZFC are true in V_κ whenever κ is inaccessible. Thus, "there is an inaccessible cardinal" is independent of ZFC.*

We get the “thus” because if κ is the smallest inaccessible cardinal then V_κ is a structure in which ZFC + “there are no inaccessible cardinals” is true.

Proof Sketch. κ is a limit ordinal, and in a problem for this chapter you prove that most of the axioms of ZFC are true in V_α whenever α is limit. The sticking point is Replacement.

The idea is, suppose you have a class function $F(x)$ over V_κ . Thinking externally: a class function over V_κ matches up with a function whose domain is a subset of V_κ . Composing F with the rank function we may assume that $F : \kappa \rightarrow \kappa$. So if $a \in V_\kappa$ then the image of a under F must be bounded by an ordinal β , by regularity. (Else there’d be a cofinal function from $\alpha = \sup a$ to κ .) So β gives you a set containing the image of a under F . So Replacement for F holds. \square

Remark 63. With some more technology (not covered in this class) you can also see why regular limit cardinals are independent of ZFC. Namely, if κ is a regular limit then you can prove that in L that κ is inaccessible. (A key point is that GCH is true in L and so being a strong limit cardinal is equivalent to being a limit cardinal.) Then run the same argument.

Inaccessible cardinals are the beginning of *large cardinals*, those cardinal numbers which are so large their existence is independent of ZFC. These give a different kind of independence than we had with CH. Getting the independence of large cardinals is about shorter versus taller universes. In effect, we get large cardinals by growing the cumulative hierarchy higher, and we lose them by cutting off at some V_κ .

On the other hand, note that CH can be cast as a statement about what’s true in $V_{\omega+\omega}$ —it asserts that there is no subset of \mathbb{R} which is neither countable nor in bijection with all of \mathbb{R} . So the truth or falsity of CH cannot be determined by which V_κ we’re in (so long as $\kappa > \omega + \omega$). Instead, it’s about having a thinner or wider universe.

Large cardinals are another central theme in modern set theory. Indeed, the three themes mentioned—inner models and forcing being the other two—enjoy a lot of interplay; much fruitful work has been done by seeing how the different concepts relate.

Large cardinals are a possible topic for Chapter ω .

Believing the axioms. Mentioning inaccessible cardinals raises a question: if we can’t prove their existence in ZFC, why should we think they even exist? We might as well go a step further and ask, why should we believe the axioms of ZFC in the first place.

The short answer is, most of the axioms describe a specific structure, the cumulative hierarchy. If you think that structure makes sense to talk about, then you should believe those axioms are true in it.

This doesn’t explain why you should believe AC. Some don’t. But whatever you decide, it cannot be denied that AC is essential to a lot of interesting work in mathematics.

The long answer belongs to philosophers of mathematics, and so falls outside the scope of this class. What I will say is, if you are interested in this question, then Penelope Maddy’s seminal paper “Believing the axioms” is the best starting place for you. Part 1 concerns the axioms of ZFC itself, while part 2 looks at large cardinal axioms. (Warning! Like most philosophy of math which looks at set theory, Maddy’s paper very quickly dives into technical mathematics.)

Exercises.

- (1) Check that if GCH is true—that is, if $2^\kappa = \kappa^+$ for all infinite κ , then a cardinal is a limit cardinal if and only if it is a strong limit cardinal.

(Kameryn J. Williams) BARD COLLEGE AT SIMON'S ROCK, 84 ALFORD RD, GREAT BARRINGTON, MA 01230
E-mail address: `kwilliams@simons-rock.edu`
URL: `http://kamerynjw.net`