The Axiom of Choice and Some Equivalences:

A Senior Exercise

Tate Deskins
Department of Mathematics
Kenyon College

November 29, 2012

Contents

1 Introduction and Motivation

1.1 Overview

Frequently in our undergraduate math courses a professor, while lecturing or helping a student through a proof, will give a sentence to this effect: "This step implicitly uses the Axiom of Choice." Some students nod knowingly, while others wonder, "What is the Axiom of Choice?" For those of us who fall into the latter category, hopefully this paper will answer that question and demonstrate some of the powerful consequences of the Axiom of Choice.

The plan is to first motivate why we should care about the Axiom of Choice and explain its statement. Following this we will show in a round-about, yet elegant way, the equivalence of the Axiom of Choice to the Well-Ordering Principle, Zorn's Lemma, and Tukey's Lemma. After each step of the proof we will discuss some of the consequences of the statements.

1.2 An Introduction to Set Theory

Axiomatic set theory is an attempt by mathematicians to describe mathematics starting with a primitive, the *set*, and a list of rules, the *axioms*, which govern how we are allowed to manipulate the primitive. We shall rely on our intuition from Foundations for what a *set* is. An *axiom* as stated dictates how we can manipulate sets. Here are a couple of examples of axioms for ZF[1] set theory[2]:

Axiom 1.1 (Extensionality): *If two sets have exactly the same members then they are equal. Formally stated:*

$$\forall X \forall Y (\forall a (a \in X \Leftrightarrow a \in Y) \Rightarrow X = Y).$$

Axiom 1.2 (Empty Set): *There exists a set with no members:*

$$\exists X \forall a (a \notin X).$$

Axiom 1.3 (Pairing): *For any sets X and Y there is a set containing only X and Y:*

$$\forall X \forall Y \exists Z \forall a (a \in Z \Leftrightarrow a = X \text{ or } a = Y).$$

[1] ZF stands for Zermelo-Fraenkel, the last names of the two mathematicians who are attributed with creating the axioms of this set theory.

[2] Some of the statements of these axioms are taken verbatim and others paraphrased from [1].

There are several more axioms in standard ZF set theory that allow us to take unions, make power sets, and construct subsets [1]. However, in order to avoid having to retitle this document "Basic Axiomatic Set Theory," we shall assume that you have a knowledge of the basic set theory that one would learn from Foundations.

1.3 Axiom of Choice

Given the basic axioms we can describe much of the mathematics that we know in set theoretic terms; however, there is another axiom that is needed to build some of the major theorems (or even what we may think of as obvious facts) of mathematics. This axiom is the *Axiom of Choice*. In essence, the Axiom of Choice allows us to select members from collections of nonempty sets, i.e. pull representatives from a partition, or create functions from arbitrary relations with the same domain.

These seem like very reasonable, and necessary, things to be able to do. Since axioms are something that we base on our intuition, we are tempted to accept the Axiom of Choice as an axiom. In general, Mathematicians find the Axiom of Choice too useful to ignore and thus include it as one of the Axioms of set theory.

Let us now give the statements of the Axiom of Choice and some of its equivalents:

Axiom of Choice 1 (Axiom of Choice)**:** *Every set has a choice function [1, 3, 4, 5, 6].*

Axiom of Choice 2 (Well-Ordering Principle)**:** *All sets can be well-ordered [1, 3, 4, 5, 6].*

Axiom of Choice 3 (Zorn's Lemma)**:** *If X is a partially ordered set where each chain has an upper bound, then X has a maximal element [2, 3, 5].*

Axiom of Choice 4 (Tukey's Lemma)**:** *If a collection of sets U is of finite character then U contains maximal sets under the ordering "$\subseteq$" [2, 3, 5].*

Before we dive in there is some terminology to discuss. We shall develop and define the mathematics necessary as the need arises.

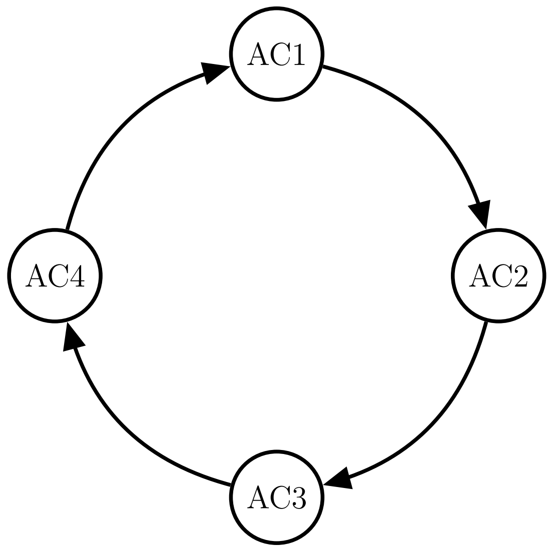

Figure 1: The path for our proof of some equivalent statements of the Axiom of Choice.

2 Axiom of Choice and Well-Ordering

Before we give the proof of the Well-Ordering Principle we have several mathematical concepts to develop.

2.1 Choice Functions, Orderings et al.

Colloquially, a *choice function* will pick an element out of a set. However since a choice function can not pick an element out of an empty set (there is nothing in there) it is useful at times to define for a set S, $\mathcal{F}(S) = \mathcal{P}(S) \setminus \{\emptyset\}$. Where $\mathcal{P}(S)$ is the power set of S.

Definition 2.1: *Let S be a set of nonempty sets. A function f is a* **choice function** *on S if*

$$f : S \to \bigcup_{T \in S} T,$$

such that $f(T) \in T$ for all $T \in S$ [5].

Notationally we shall use $\mathcal{D}om(f)$ and $\mathcal{R}an(f)$ to be the domain and range of a function f.

There are three types orderings that we may recall from Foundations[3]:

Definition 2.2 (Partial Ordering)**:** *Given a set S, a relation " $\sqsubseteq$" is a partial ordering if:*

- *for $s, t \in S$ $s \sqsubseteq t$ and $t \sqsubseteq s$ then $s = t$*
- *for $s, t, r \in S$, if $s \sqsubseteq t$ and $t \sqsubseteq r$ then $s \sqsubseteq r$*
- *for all $s \in S$ $s \sqsubseteq s$.*

The set and ordering in Figure 2 is a partial ordering.

Definition 2.3 (Total Ordering)**:** *Given a set S, a relation "$\trianglelefteq$" is a total ordering if:*

- *$(S, \trianglelefteq)$ is a partial ordering*
- *for all $s, t \in S$, $s \trianglelefteq t$ or $t \trianglelefteq s$.*

Definition 2.4 (Well-Ordering)**:** *Given a set S, a relation "$\preceq$" is a well-ordering if:*

- *$(S, \preceq)$ is a total ordering*
- *for all $T \subseteq S$, there is a $t \in T$ such that $t \preceq s$ for all $s \in T$.*

With choice functions and some orderings under our belt, let us talk about ordinals.

2.2 Ordinal Numbers

Let us define the successor of a set X as $\mathcal{S}(X) = X \cup \{X\}$ [4]. With this tool we will consider a different picture of the natural numbers:

$$\begin{aligned} 0 &= \emptyset \\ 1 &= \mathcal{S}(0) = \{0\} \\ 2 &= \mathcal{S}(1) = \{0, 1\} \\ 3 &= \mathcal{S}(2) = \{0, 1, 2\} \\ &\vdots \end{aligned}$$

[3]The following three definitions were paraphrased directly from [7].

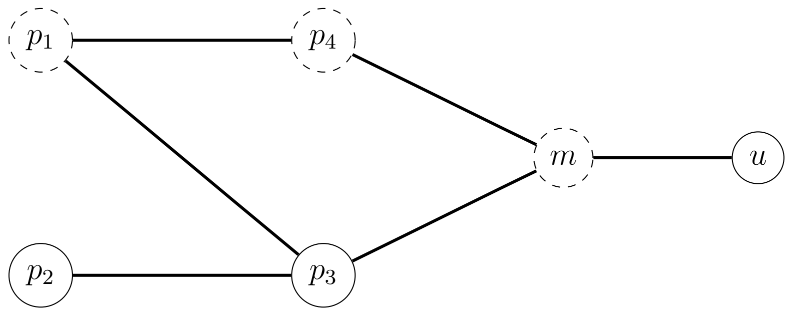

Figure 2: An example of a partially ordered set (by "rightness" on the page). The dashed nodes are part of a chain of which m is a maximal element and u an upper bound.

Thus for a non-zero natural number $n \in \mathbb{N}$ we define it recursively[4] as $n = \mathcal{S}(n-1)$. We can extend this definition of the natural numbers further by defining ω as the union of all natural numbers (finite ordinals), i.e. $\omega = \mathbb{N}$. We can then continue defining what we will call the ordinal numbers similarly:

$$\begin{aligned} \omega &= \mathbb{N} \\ \omega+1 &= \mathcal{S}(\omega) = \mathbb{N} \cup \{\omega\} \\ \omega+2 &= \mathcal{S}(\omega+1) = \mathbb{N} \cup \{\omega, \omega+1\} \\ \omega+3 &= \mathcal{S}(\omega+2) = \mathbb{N} \cup \{\omega, \omega+1, \omega+2\} \\ &\vdots \end{aligned}$$

As we continue creating ordinals we will get 2ω and $\omega \cdot \omega$ then eventually ω^ω. Now just as we stepped beyond the natural numbers by taking the union of all natural numbers and defining this as ω, we will define ω_1 to be the union of all the ordinals described by combinations of ω and the natural numbers. We can similarly define:

$$\omega_2, \omega_3, \ldots, \omega_\omega, \omega_{\omega+1}, \ldots, \omega_{\omega_1}, \ldots$$

So we have two different kinds of ordinal numbers: *limit ordinals*, like the list above, which are created by taking unions of all previous types of ordinals; and *successor ordinals*, such as $\omega + 9$, which are created using the successor function.

[4]We will discuss recursion in more detail in Section 2.4

It appears that the ordinal numbers are well-ordered under inclusion "$\subseteq$." This is in fact a theorem, but its proof is beyond the scope of the paper.

Let us now formalize this intuitive understanding of the ordinal numbers.

Definition 2.5: *A set α is an* **ordinal number** *if:*

- *every element of α is a subset of α*
- *α is strictly well-ordered by "$\in$" [4].*

Definition 2.6: *The ordinal α is a* **successor ordinal** *if there is an ordinal β such that $\mathcal{S}(\beta) = \alpha$. Otherwise α is a* **limit ordinal** *[4].*

Lemma 2.7: *Set inclusion "$\subseteq$" is a well-ordering on ordinal numbers.*

A proof of this lemma can be seen on page 109 of [4]. Another important fact about the ordinals is that they are not a set.

Lemma 2.8: *The collection of ordinal numbers is not a set.*

Proof: If the ordinals were a set, say $\mathbb{O}$, then we see that $\mathbb{O}$ is the union of all ordinals, (another limit ordinal). Thus it is natural for us to consider $\mathbb{O} + 1 = \mathcal{S}(\mathbb{O})$, which is an ordinal. So $\mathbb{O} + 1 \subseteq \mathbb{O}$, yet clearly $\mathbb{O} \subset \mathbb{O} + 1$. To avoid this paradox (the Burali-Forti paradox) we conclude that the ordinals are not a set [1]. ■

Accepting this fact, we shall slightly abuse notation and use $\mathbb{O}$ to represent the ordinals for the sake of succinctness. We will also use "$\leq$" to denote the well-ordering of the ordinals (which is equivalent to "$\subseteq$").

Now as the name suggests we would like to show that the ordinals and well-ordered sets are linked. So consider the following lemma:

Lemma 2.9: *There is a bijection $f : \beta \to S$ for an ordinal β and a set S if and only if there is a well-ordering of S.*

Proof: We shall just prove the forward direction of the lemma here. Let S be a set, β an ordinal, and $f : \beta \to S$ a bijection. We shall now define a well-ordering on S. Consider $r, t \in S$. Since f is a bijection we know that there exists unique $\gamma, \delta \in \beta$ such that $f(\gamma) = r$ and $f(\delta) = t$. Define "$\preceq$" such that $r \preceq t$ if and only if $\gamma \leq \delta$.

To show that this is a well-ordering let us show first that "$\preceq$" is a total ordering. To do this we need only show (i) antisymmetry, (ii) transitivity, and (iii) totality of "$\preceq$". Let $c, d, e \in S$. Thus we know that there exists $\gamma, \delta, \zeta \in \beta$ such that $f(\gamma) = c$, $f(\delta) = d$, and $f(\zeta) = e$.

(i) Let $c \preceq d$ and $d \preceq c$. Then $\gamma \leq \delta$ and $\delta \leq \gamma$. Hence $\gamma = \delta$. Therefore since f is a bijection $c = d$.

(ii) Let $c \preceq d$ and $d \preceq e$. Then $\gamma \leq \delta$ and $\delta \leq \zeta$. Hence $\gamma \leq \delta$. Therefore since f is a bijection $c \preceq e$.

(iii) Note that either $\gamma \leq \delta$ or $\delta \leq \gamma$. Thus $c \preceq d$ or $d \preceq c$.

Now we need to show that any subset of S has a least element. Let $T \subseteq S$. From Foundations we recall that the inverse image of $f^{-1}(T) \subseteq \beta$, and that $f : f^{-1}(T) \to T$, is a bijection. Thus since the ordinals are well-ordered we know that there exists a least element $\alpha \in f^{-1}(T)$. Since f is a bijection, $f(\alpha) = a$ is a unique element in T. Furthermore for $t \in T$ we know there exists a unique $\gamma \in f^{-1}(T)$ such that $f(\gamma) = t$. Since by definition $\alpha \leq \gamma$ we know that $a \preceq t$. Since t was arbitrary we know that a is the least element of T.

Thus we have shown that "$\preceq$" is a well-ordering on S. ■

2.3 Induction...

As mathematicians we have a tool in our pockets called induction. It is proved from the well-ordering of the natural numbers and works on natural numbers. Here is a formal statement of induction:

Theorem 2.10 (Complete Induction)**:** *Let $P(n)$ be a set of statements where $n \in \mathbb{N}$. If $P(0)$ is true and if $P(k)$ is true for $k < N \in \mathbb{N}$ implies $P(N)$, then $P(n)$ is true for all $n \in \mathbb{N}$.*

We shall extend this idea of induction to transfinite induction, just as we extended the natural numbers to the ordinals. The statement of transfinite induction will be

Theorem 2.11 (Transfinite Induction)**:** *Let $P(\alpha)$ be a collection of statements defined for all ordinals α. Suppose that whenever $P(\gamma)$ is true for all ordinals γ such that γ is less than some ordinal β, then $P(\beta)$ is true. Then P is true for all ordinals [1].*

Note that transfinite induction is always complete transfinite induction. Here is a proof for Transfinite Induction:

Proof: Let $P(\alpha)$ be a collection of statements defined for all ordinals α. Also suppose that whenever $P(\gamma)$ is true for all ordinals γ such that γ is less then some ordinal β then $P(\beta)$ is true. We shall proceed by contradiction. Thus we also assume that P is false for some ordinal δ. Define $F = \{\gamma \leq \delta : P(\gamma) \text{ is false}\}$[5]. F is a non-empty set of ordinals since $\delta \in F$. By the well-ordering of the ordinals, we know that there must be a least element of F which we will call β. This means for all ordinals γ such that $\gamma < \beta$, $P(\gamma)$ is true. Since $P(\gamma)$ is true for all $\gamma < \beta$ then by assumption $P(\beta)$ is true. Thus we have shown $P(\beta)$ is both true and false. Therefore we have reached a contradiction. Therefore transfinite induction is valid [1]. ■

2.4 Recursion

We have all previously used a tool called (finite) recursion in many proofs. It describes a way of constructing functions inductively.

Theorem 2.12 (Finite Recursion)**:** *If A is a set and $F : A \times \mathbb{N} \to A$ is a function, then there exists a unique function $f : \mathbb{N} \to A$ such that $f(0) = a$ for some $a \in A$. Furthermore, for all $n \in \mathbb{N}$, $f(n+1) = F(f(n), n)$ is unique*[6].

The validity of recursion as a method for constructing well-defined functions on the desired domain is provable by induction, as one would expect. To refresh our memories on how one might define a function inductively, consider the following recursive definition.

Given $F : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ where $F(a, b) = a \cdot b$ and $f(0) = 1$ we see that:

$$\begin{aligned} f(1) &= 1 \cdot f(0) = 1 \\ f(2) &= 2 \cdot f(1) = 2 \\ f(3) &= 3 \cdot f(2) = 6 \\ f(4) &= 4 \cdot f(3) = 24 \\ &\vdots \end{aligned}$$

We have recursively constructed the factorial function on natural numbers, $f(n) = n!$.

We can extend definition by recursion to the ordinals. When we do we call this *transfinite recursion*. With transfinite recursion we construct functions *transfinite-inductively*. We may naively write the theorem like this:

[5]Intuitively we want F to be the set of all ordinals for which P fails; however, this may not be a proper set. Thus we pick an arbitrary max element of F which is δ.

[6] It was tempting to state this theorem as "See Theorem 2.12."

Theorem 2.13 (Transfinite Recursion Take 1)**:** *Let X be a set. Let $F : X \times \mathbb{O} \to X$ be a function. There exists a unique function $f : \mathbb{O} \to X$ such that for a fixed $x \in X$, if $f(0) = x$, then for α and ordinal, $f(\alpha+1) = F(f(\alpha), \alpha)$ is unique.*

After reading this theorem we should be worried. The root of our issues is that we have treated the ordinal numbers as a set; however, as we proved in Lemma 2.8 we know that the ordinal numbers are not a set. Since they are not a set, the domains of our functions F and f are not sets, and therefore we can not claim that F and f are actually functions. This is a problem since the goal is to define a function.

We will need to reformulate this theorem to state it formally and avoid these issues. However, the intuition that one may get from *Transfinite Recursion Take 1* of building a function transfinite inductively is important to understanding any usage of transfinite recursion.

Let us first define what a *function-like formula* is.

Definition 2.14: *A formula P is function-like if for any f there is at most one y such that the formula $P(f, y)$ is true.*

We can now define transfinite recursion in a more formal way.

Theorem 2.15 (Transfinite Recursion)**:** *If S is a set, and P is a function-like formula, and $y_0 \in S$, then there exists a unique function-like formula F such that the following are true either for all ordinals β or for all ordinals β less than some ordinal γ:*

- *$F(0, y_0)$ is true and*
- *if $P(\{(\alpha, s') : \alpha < \beta$ and $F(\alpha, s')$ is true$\}, (\beta, y))$ holds then $F(\beta, y)$ holds as well.*

We will use the set $\{(\alpha, s') : \alpha < \beta \text{ and } F(\alpha, s') \text{ is true}\}$ frequently enough to give it a name O_β. After reading the statement for transfinite recursion we need to convince ourselves that O_β is a set. In order to show that it is a set, let us define two more bits of notation.

Definition 2.16: *Let O be a set of ordered pairs. We define the* **projection of O onto the left and right coordinates** *to be $\mathcal{L}(O) = \{a : (a, b) \in O\}$ and $\mathcal{R}(O) = \{b : (a, b) \in O\}$ respectively.*

If we consider O_β in the context of Theorem 2.15, we see that $\mathcal{L}(O_\beta) = \beta$. Since β is an ordinal number we know that O_β is a set. Therefore since F

is a function-like formula, there is a one-to-one correspondence between O_β and $\mathcal{L}(O_\beta)$. Thus O_β must be a set.

In Theorem 2.15 we use the phrase "either for all ordinals β or for all $\beta < \gamma$ for some ordinal γ." In the second case, where P only holds for $\beta < \gamma$, we say that the recursion has *terminated*.

Let us explore what this means. Assume the hypothesis of Theorem 2.15 and that there is some ordinal δ for which our transfinite recursion fails. We may assume without loss of generality that since the ordinal numbers are well-ordered, δ is the least such ordinal for which our recursion fails. Since for all $\beta < \delta$ there exists $s \in S$ such that $F(\beta, s)$ is true, we know that O_δ is a set. Since F is function-like, for all $\beta \in \mathcal{L}(O_\delta) = \delta$ there is a unique $s \in S$ such that $(\beta, s) \in O_\beta$. This means that O_β describes a function. Let us call this function $f : \delta \to S$. So that we are very clear, the function f is defined such that for all ordinals β such that $\beta \in \delta$, $F(\beta, f(\beta))$ is true.

As desired, terminating transfinite recursions do build functions transfinite inductively. We will not prove the validity of transfinite recursion in ZF set theory; however, the proof of the theorem relies on transfinite induction.

We need one more definition before we give an example in a proof of transfinite recursion.

Definition 2.17: *Let F be a function-like formula. We say F is one-to-one if there exists some γ such that if $F(\beta, \gamma)$ and $F(\alpha, \gamma)$ hold then $\alpha = \beta$.*

With the new tool of transfinite recursion we shall now prove the reverse direction of Lemma 2.9. This direction states that if a set S is well-ordered then there is a bijection $f : \beta \to S$ for some ordinal β.

Proof (Lemma 2.9 continued**):** Let S be a non-empty, well-ordered set under "$\preceq$". We want to define a function f through transfinite recursion thusly: Let s_0 be the least element of S,

- $f(0) = s_0$
- for α an ordinal
$$f(\alpha) = \min\left(S \setminus \bigcup_{\gamma\in\alpha} f(\gamma)\right).$$

However, we know that this is not a formal statement because we don't know that f is a proper function since its domain could be all of the ordinals (in theory). So using the formal statement of transfinite recursion, we define a formula P such that $P(A, s)$ holds if and only if α is an ordinal and A

is a set of ordered pairs given by $A = \{(\beta, s') : \beta < \alpha \text{ and } s' \in S\}$ and $s = \min(S \setminus \mathcal{R}(A))$. Since $\mathcal{R}(A)$ is unique given A, we know that P is a function-like formula.

Thus we can build a unique function-like formula F transfinite recursively where either for all ordinals β or for all $\beta < \gamma$ for some ordinal γ the following hold:

- $F(0, s_0)$ holds and
- and if $P(O_\beta, (\beta, s))$ holds then $F(\beta, s)$ also holds.

To re-state this in a more usable way, for an ordinal β, if $F(\beta, s)$ holds then $s = \min(S \setminus \mathcal{R}(O)_\beta)$.

Let us now show that F is one-to-one. We shall proceed by contradiction. Assume for ordinals α and β that $F(\alpha, s)$ and $F(\beta, s)$ hold and $\beta > \alpha$. Since $\beta > \alpha$ and $F(\alpha, s)$, we see that $s \in \mathcal{R}(O_\beta)$ because $(\alpha, s) \in O_\beta$. Thus $s \notin S \setminus \mathcal{R}(O_\beta)$ and therefore $s \neq \min(S \setminus \mathcal{R}(O)_\beta)$. Thus we conclude that $F(\beta, s)$ is not true. We have reached a contradiction, thus F is one-to-one.

Define $A_r = \{s : \text{ for an ordinal } \beta,\ F(\beta, s) \text{ holds}\}$. Note that A_r is a subset of S and thus a set. Furthermore, since F is one-to-one we know that there is a one-to-one correspondence between A_r and the collection $A_l = \{\beta : \text{ for } s \in S,\ F(\beta, s) \text{ holds}\}$. This means that the collection of ordinals A_l is a set. Since A_l is a set of ordinals there must be some ordinal δ such that $\delta \notin A_l$. Since the ordinals are well-ordered, we know that there is a least $\gamma \in \delta$ such that $\gamma \notin A_l$ Since $F(\gamma, s)$ is not true for any $s \in S$, we know that there are no elements in $S \setminus \mathcal{R}(O)_\gamma$. By construction, for all ordinals ζ such that $\zeta > \gamma$, $O_\gamma \subseteq O_\delta$. Therefore F doesn't hold for any $\zeta > \gamma$. Hence $A_l = \gamma$. Therefore our recursion has terminated and O_γ describes a function $f : \gamma \to S$. Since F does not hold for δ we know that every element of S is in the range of f, i.e. f is onto. Note that f is one-to-one since F is one-to-one. Therefore f (a function) is a bijection between an ordinal γ and S [4, 5]. ■

Note that with this lemma we can conduct transfinite recursion with any well-ordered set S instead of just the ordinals. Furthermore, such a recursion will always terminate since there is a bijection between a well-ordered set S and some ordinal number.

We could work every proof in which we use transfinite recursion with this formal setup, by constructing our formula-like functions. However in doing this we lose much of our intuition in the name of rigor. Thus we shall subsequently use transfinite recursion with the language of functions instead of function-like formulas. In this paper all of our recursions will terminate, meaning that if we used the formal setup we would be able to construct a proper function.

2.5 Hartog's Number

Definition 2.18 (Hartog's Number)**:** *Given a set S, Hartog's Number, $\mathcal{H}(S)$, for the set S is the least ordinal so that there is no one-to-one function $f : \mathcal{H}(S) \to S$ [8].*

Lemma 2.19 (Construction of Hartog's Number)**:** *For a set S,*

$$\mathcal{H}(S) = \{\alpha \in \mathbb{O} : \textit{ there is a bijection } f : \alpha \to T \textit{ for some } T \subseteq S\}.$$

The proof of this lemma is outside the scope of this paper, but here is a sketch of said proof.

Proof Sketch: Let S be a set. Consider:

$$\beta = \mathcal{H}(S) = \{\alpha \in \mathbb{O} : \text{ there is a bijection } f : \alpha \to T \text{ for some } T \subseteq S\}.$$

The difficulty of this proof is showing that β is a set and an ordinal number. These two things are provable using the Axiom of Replacement and the Schema of Separation from set theory. We shall assume this and show that β satisfies Definition 2.18.

Note that if there is a one-to-one function $f : \beta \to S$ then $\beta \in \beta$. However, since no ordinal number contains itself as an element, we have reached a contradiction. Thus there is no one-to-one function $f : \beta \to S$. All that is left to show is that β is the least such ordinal number. If $\gamma < \beta$ then $\gamma \in \beta$, since β is an ordinal. Thus there is a one-to-one function $f : \gamma \to S$ by the construction of β. Hence β is the least ordinal for which there is no one-to-one function $f : \beta \to S$ [1, 8]. □

With all of these tools and definitions we can now prove the Well-Ordering Principle from the Axiom of Choice.

2.6 Well-Ordering Principle

Theorem 2.20 (AC1→AC2)**:** *If for any set A, a collection of nonempty sets there is a choice function, then for any set S there exists a relation "$\preceq$" which is a well-ordering on S.*

Proof: Let S be a set. Note that if S is empty it is trivially well-ordered. Consider $\mathcal{F}(S)$.

We know that $\emptyset \notin \mathcal{F}(S)$, thus we know that there exists a choice function f on $\mathcal{F}(S)$. Consider the function h into S defined by transfinite recursion where $h(0) = f(S) \in S$ and

$$h(\alpha) = f\left(S \setminus \bigcup_{\beta \in \alpha} h(\beta)\right).$$

We see that this function is very similar to the function created in Lemma 2.9. The only change is that we are using a choice function f instead taking the minimum. Thus by a similar set of formula-like functions we could state this proof by the formal statement of transfinite recursion. However, we will stick with the more intuitive language of functions.

Note that $\mathcal{R}an(f) \subseteq S$ and thus $\mathcal{R}an(h) \subseteq S$.

Before we show that the recursion terminates, we will show that that the function is one-to-one. This is similar to when we showed F, the function-like formula in the proof of the backward direction of Lemma 2.9, was one-to-one. Let $\gamma, \delta \in \mathcal{D}om(h)$ be distinct. Thus without loss of generality, $\gamma < \delta$ (equivalently $\gamma \in \delta$). Hence by the recursive definition of h, $h(\gamma) \notin S \setminus \bigcup_{\beta \in \delta} h(\beta)$, so the choice function f can't choose $h(\gamma)$ for $h(\delta)$. Therefore $h(\delta) \neq h(\gamma)$. Thus we know that the function is one-to-one.

Now we need to show that the recursion terminates for some ordinal β. Note that by definition of Hartog's number $\mathcal{H}(S)$, there is no one-to-one function $g : \mathcal{H}(S) \to S$. If the recursion didn't terminate, then $h(\mathcal{H}(S)) \in S$. This would mean that we could define the subset $B = \{h(\delta) : \delta < \mathcal{H}(S)\}$ of S. Clearly $h : \mathcal{H}(S) \to B$ is a well defined one-to-one function. Thus $h : \mathcal{H}(S) \to S$ is one-to-one. This is a contradiction. Therefore for some $\tau < \mathcal{H}(S)$ the recursion must terminate, meaning that $\{h(\beta) : \beta \in \tau\} = S$.

Therefore h is a bijection from an ordinal number τ to S. Thus by Lemma 2.9 we conclude that S is well-ordered [1, 5]. ■

2.7 Implications?

If we assume the Axiom of Choice we get, or are forced to accept, that every set can be well-ordered. Consider the real numbers. No one to date has explicitly defined a well-ordering of the real numbers. We know that there must be a well-ordering by Theorem 2.20. Although the proof appears constructive with the recursive function, it is not actually a constructive argument because we never gave the form of the choice function.

The set $(\mathbb{Z}, \leq)$ is not well-ordered, but we can define several different well-orderings for the integers. Here are a couple of orderings pictorially shown:

$$0, 1, -1, 2, -2, 3, -3, \ldots$$

$$0, 1, 2, 3, 4, \ldots - 1, -2, -3, -4, \ldots$$

$$0, 2, 4, 6, \ldots - 2, -4, -6, \ldots, 1, 3, 5, \ldots - 1, -3, -5, \ldots$$

The visual layout here suggests (correctly) that the *order type* (the ordinal number for which the bijection onto $\mathbb{Z}$ is defined) is different depending on how we construct the sequence. The first has order type ω which is the same as the natural ordering of $\mathbb{N}$. The second ordering we gave has order type 2ω, and the third 4ω. The cardinality of a set S can be defined as the smallest ordinal for which there is a bijection onto S. It turns out that the cardinality is always a limit ordinal.

Accepting that every set, particularly the reals, can be well-ordered, is hard to stomach. However, let us note that without the Axiom of Choice, one can construct a model where there is a uncountable subset of the real line with no countable subset [5]. This is because the ability to compare cardinalities is dependent on our ability to well-order the sets (like we did with $\mathbb{Z}$) whose cardinalities we wish to compare. Without the Axiom of Choice, cardinalities are only partially ordered, and thus you can have incomparable cardinalities. So we are left with a choice: Do we say that we can well-order the reals, or do we give up on cardinal comparability?

3 Maximality

The next two alternate statements of the Axiom of Choice that we will prove (AC3 and AC4) describe maximal elements in a set. Before we start we have some more mathematics to refresh.

3.1 Maximally Basic

The following four definitions should clear up our understanding of Zorn's and Tukey's Lemmas.

Definition 3.1 (Chain)**:** *In a partially ordered set* $(S, \sqsubseteq)$*, a subset* $C \subseteq S$ *is a* **chain** *if* $(C, \sqsubseteq)$ *is totally ordered [2, 5].*

Intuitively, if we appeal to Figure 2 (copied again below) a chain will be the elements on a horizontal left-to-right path, like the elements denoted by dashed circles.

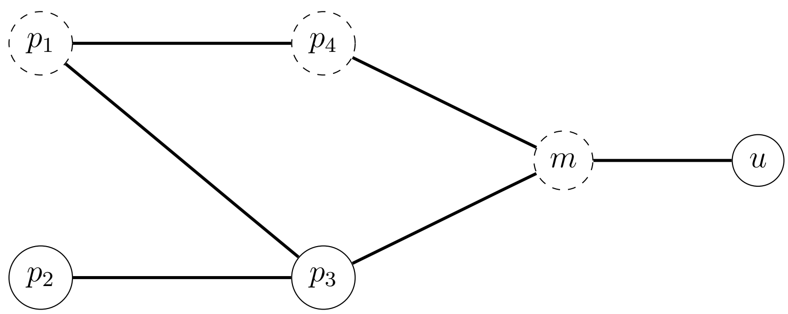

Figure 2: (Again) An example of a partially ordered set (by "rightness" on the page). The dashed nodes are part of a chain of which m is a maximal element and u an upper bound.

Definition 3.2 (Upper Bound)**:** *For a partially ordered set* $(S, \sqsubseteq)$*,* T*, a subset of* S*, has an* **upper bound** $u \in S$ *if for all* $t \in T$*,* $t \sqsubseteq u$ *[7].*

In Figure 2 u, m are upper bounds to the dashed chain.

Definition 3.3 (Maximal element)**:** *For a partially ordered set* $(S, \sqsubseteq)$*,* T*, a subset of* S*, has a* **maximal element** $m \in S$ *if for all* $t \in T$*,* $t \sqsubseteq m$ *and* $m \in T$ *[7].*

In Figure 2, m is a maximal element of the dashed chain.

Definition 3.4 (Finite Character)**:** *A set* S *has* **finite character** *if it has the following properties:*

- *for each* $A \in S$*, every finite subset* $B \subseteq A$ *is in* S
- *if every finite subset* B *of a set* A *is in* S*, then* $A \in S$ *[4].*

This definition is much harder to gain intuition about. Here is an example of a set with finite character: For a vector space V the set $I = \{B :$

B is a set of linearly independent vectors$\}$. Before we prove this, we need to remind ourselves of the definitions of linear dependence and independence for a vector space.

Definition 3.5 (Linear Dependence)**:** *A set of vectors A in a vector space V is linearly dependent if there is a finite subset of A for which there is a non-trivial linear combination which sums to zero.*

Definition 3.6 (Linear Independence)**:** *A set of vectors A in a vector space V is linearly independent if A is not linearly dependent.*

With this reminder of the definition of linear independence, let us show that I has finite character.

Lemma 3.7: *For a vector space V the set*

$$I = \{B : B \textit{ is a set of linearly independent vectors}\}$$

has finite character.

Proof: Consider the set I as defined above. Note that if $B \subset V$ is a set of linearly independent vectors, every subset of B is linearly independent and thus an element of I. Furthermore, if $A \subset V$ is a set where every finite subset is a set of linearly independent vectors, then every two vectors in A are linearly independent. Since every finite subset of A is linearly independent, this means that no finite subset of A is linearly dependent. Thus by definition A is not linearly dependent, i.e. A is a set of linearly independent vectors. Therefore $A \in I$. So we see that I has finite character. ■

This set I will be useful later in Section 3.4 in the proof that every vector space has a basis.

To fully solidify our understanding of finite character, let us now consider a set that does not have finite character.

Example: Let us consider the set

$$W = \{A : A \subseteq \mathbb{R} \text{ and } A \text{ has a least element under "}\leq\text{"}\}.$$

Note that $\mathbb{R} \subseteq \mathbb{R}$ and that every finite subset of the real numbers has a least element under the usual ordering. However, it is apparent that the real numbers don't have a least element under "$\leq$." Thus $\mathbb{R} \notin W$ and therefore W does not have finite character. □

With these definitions let us jump into proving Zorn's Lemma.

3.2 Well-Ordering to Zorn's Lemma

Theorem 3.8 (AC2→AC3)**:** *If for any set A there is a well-ordering, then for all partially ordered sets P where each chain has an upper bound in P, P has a maximal element.*

Proof: Let $(P, \sqsubseteq)$ be a partially ordered set where each chain has an upper bound. We know that P is well-ordered by some relation "$\preceq$" by the Well-Ordering Principle. Consider this transfinite recursive function $f : P \to \{0, 1\}$ given by:

$$f(p) = \begin{cases} 1 & \text{if for all } p' \prec p \text{ for which } f(p') = 1,\ p' \sqsubset p \\ 0 & \text{otherwise.} \end{cases}$$

Note that we are using Lemma 2.9 when we construct our transfinite function from the well-ordered set P to the set $\{0, 1\}$. Since P is a set, the discussion after the proof of the reverse direct of Lemma 2.9 tells us that the recursion terminates and thus f is a proper function.

Let $Q = \{p \in P : f(p) = 1\}$. Since $Q \subseteq P$, "$\sqsubseteq$" is a partial ordering and "$\preceq$" is a well-ordering on Q. Consider $p, q \in Q$. If $p \preceq q$, then by the definition of Q, $p \sqsubseteq q$. Similarly if $q \preceq p$, then by the definition of Q, $q \sqsubseteq p$. We see that "$\sqsubseteq$" is a total ordering on Q. Thus Q describes a chain in P. Therefore, by assumption, there exists an upper bound $m \in P$ such that for all $p \in Q$, $p \sqsubseteq m$.

To show that m is maximal, let us consider $u \in P$ such that $m \sqsubseteq u$ and show that $u = m$.

Let u be defined as hypothesized. Let $p \in Q$. Since m is an upper bound for Q, $p \sqsubseteq m$. By the transitivity of our relation, since $m \sqsubseteq u$ and $p \sqsubseteq m$, $p \sqsubseteq u$. Thus we just showed that for all p such that $f(p) = 1$ (which include the $p \prec u$ such that $f(p) = 1$), $p \sqsubseteq u$. Thus we conclude that $f(u) = 1$. So we know that $u \in Q$. Therefore by the definition of m, $u \sqsubseteq m$. We conclude by the anti-symmetry of "$\sqsubseteq$" that $u = m$. Hence m is a maximal element of P [1]. ■

3.3 Zorn's Lemma to Tukey's Lemma

Before we can prove Tukey's Lemma we first need the following lemma:

Lemma 3.9: *If F is a set with finite character and is partially ordered by "$\subseteq$" with chain C, then*

$$u = \bigcup_{c \in C} c$$

is an element of F.

Proof: Let F be a set with finite character partially ordered by "$\subseteq$". Let $C \subseteq F$ be a chain in F. Consider

$$u = \bigcup_{c \in C} c.$$

We will prove by induction that every finite subset of u is an element of F and thus, since F has finite character, $u \in F$.

Base Case: Consider $\{d\} \subseteq u$, or in other words $d \in u$. We know then by the construction of u that there is a $c \in C$ such that $\{d\} \subseteq c$. Since $c \in C$. then $c \in F$. Furthermore, we know that every subset of c is in F, i.e. $\{d\} \in F$.

Inductive Step: Assume for $n \in \mathbb{N}$ that for any set $\{d_i\}_{i=1}^{n} \subseteq C$ there is some $c \in C$, such that $\{d_i\}_{i=1}^{n} \subseteq c$ and thus $\{d_i\}_{i=1}^{n} \in F$. Consider the set $\{d_i\}_{i=1}^{n+1} \subseteq C$. Note that $\{d_i\}_{i=1}^{n+1} = \{d_i\}_{i=1}^{n} \cup \{d_{n+1}\}$. Note that from our base case there must be a $c_1 \in C$ such that $d_{n+1} \in c_1$, and from our induction hypothesis we know that there is a $c \in C$ such that $\{d_i\}_{i=1}^{n} \subseteq c$. Since C is a chain it is totally ordered by "$\subseteq$". Thus either $c \subseteq c_1$ or $c_1 \subseteq c$. Hence either $\{d_i\}_{i=1}^{n} \subseteq c_1$ or $\{d_{n+1}\} \subseteq c$. Therefore we conclude in either case that $\{d_i\}_{i=1}^{n+1}$ is a subset of some element of C (either c or c_1). Because F has finite character and $\{d_i\}_{i=1}^{n+1}$ is a finite subset of some element of C (and thus F), we know that $\{d_i\}_{i=1}^{n+1}$ is an element of F.

Therefore, since every finite subset of u is an element of F, and since F has finite character, we can conclude that $u \in F$. ■

With this lemma, we can now prove Tukey's Lemma.

Theorem 3.10 (AC3→AC4)**:** *If for all partially ordered sets P where each chain has an upper bound implies P has a maximal element, then all sets F of finite character have a maximal element under "$\subseteq$".*

Proof: Let P be a set with finite character. We know that P is partially ordered by "$\subseteq$", because by definition "$\subseteq$" is a partial ordering for sets and because every element of P is a set[7].

Let $C \subseteq P$ be a chain. Consider

$$u = \bigcup_{c \in C} c.$$

By Lemma 3.9 we know that $u \in P$. Let $c \in C$ and $a \in c$. Thus since u is the union of all the $c \in C$, we know that $a \in u$. Thus $c \subseteq u$. Since c is arbitrary, we know u is an upper bound for the chain C. Since C is an arbitrary chain in P, we may conclude that every chain in P has an upper bound. Therefore by our assumption (Zorn's Lemma), P has a maximal element [3, 5]. ■

3.4 Implications?

In Abstract Algebra we used Zorn's Lemma to prove the existence of the algebraic closure of a field. A sketch of the proof can be found on page 289 of [2].

One interesting consequence of Tukey's Lemma is that we are able to prove that every vector space has a basis. In this proof we will use the set I, which we defined in Section 3.1. For convenience it is redefined below:

$$I = \{B : B \text{ is a set of linearly independent vectors}\}.$$

We shall sketch the proof.

Theorem 3.11: *Tukey's Lemma is equivalent to the following statement: Every vector space has a basis [3, 5].*

Although the theorem is an equivalence, the reverse direction is outside the scope of this paper so we shall just prove the forward direction.

Proof Sketch: Let V be a vector space. By Lemma 3.7 I has finite character. Thus by Tukey's Lemma (AC4), I has a maximal element $\mathcal{B}$. Note that $\mathcal{B}$ is a set of linearly independent vectors. Furthermore, if $v \in V$ is linearly independent of the vectors of $\mathcal{B}$, then $\mathcal{B} \cup \{v\} \in I$. Since $\mathcal{B} \subset \mathcal{B} \cup \{v\}$, $\mathcal{B}$ can't be the maximal element of I. Therefore v can not be linearly independent of the vectors in $\mathcal{B}$. Hence $\mathcal{B}$ is a basis set for V [3, 5]. □

[7] This is because our set theory contains no individuals.

4 Equivalence

Now for a spot of elegance; if we prove that Tukey's Lemma implies the Axiom of Choice, we will have equivalence of all four statements of the Axiom of Choice (see Figure 1 for the sketch of our path). However, before we do this we must refresh our knowledge of functions and sets of functions.

4.1 Functions

Recall that in set theory a function from A to B is a subset of a relation, i.e. a function $f : A \to B$ is a subset of $A \times B$. This is because f can be completely characterized by the set of ordered pairs $\{(a, f(a)) : a \in A\}$.

Lemma 4.1 (Sets of Functions)**:** *Given sets A and B, the set*

$$F = \{f : f \textit{ is a function from some } A' \subseteq A \textit{ to } B\},$$

has finite character [5].

Proof: Let us consider $g \subseteq f$. Note that g is a function because for every $a \in \mathcal{D}om(g)$, $g(a) = f(a) = b$ where $b \in B$. Thus g is a well defined function.

Consider the set of all functions from any subset of A to B. Call this set F. We know that if $f \in F$, then $g \subseteq f$ is also an element of F since g is a function from a subset of A to B (see the above paragraph). Furthermore, if $h \notin F$, then h is not a function from a subset of A to B. Thus either there exists $(x, y) \in h$ such that $x \notin A$ or $y \notin B$, or there exists $a \in A$ such that $(a, b_1) \in h$ and $(a, b_2) \in h$ such that $b_1 \neq b_2$. In either case there is a finite subset of h, either $\{(x, y)\}$ or $\{(a, b_1), (a, b_2)\}$ respectively, that is not a function and thus not an element of F. Therefore we see that F has finite character. ■

4.2 Back to Choice

Theorem 4.2 (AC4→AC1)**:** *If for any set F of finite character there is a maximal element, then for all sets S there is a choice function on S.*

Proof: Let S be a collection of nonempty sets. If S is the empty set, then the null function is our choice function. Let us assume that S is nonempty. Consider the set of functions

$$F = \{f : f \text{ is a choice function on some } T \subseteq S\}.$$

By an argument parallel to that in Lemma 4.1, F has finite character. Thus by assumption F has a maximal element. Let us call this maximal element g. Consider $\mathcal{D}om(g)$. Note that if $\mathcal{D}om(g) \neq S$ then there exists $s \in S$, which is nonempty, such that $g(s)$ is not defined. Since s is nonempty, there exists $t \in s$. Let us extend g to g' such that $\mathcal{D}om(g') = \mathcal{D}om(g) \cup s$ and $g' = g \cup \{(s, t)\}$. Thus clearly g' is a choice function which strictly contains g. Thus g is not maximal. We have reached a contradiction, thus $\mathcal{D}om(g) = S$ and g is our choice function for S [5]. ■

5 Concluding Remarks

The journey we took was exciting, enlightening, and full of choice. Along the journey we saw the power of the Axiom of Choice. The power is in the theorems that we are able to prove with the axiom. That we need the Axiom of Choice to guarantee a basis for a vector space and compare cardinalities seems to suggest that the Axiom of Choice is a "True" and "Good" axiom; however, there is one great paradox that arrises when we assume the Axiom of Choice.

Enter the Banach-Tarski paradox. This paradox proves from the Axiom of Choice that if we are clever, we can cut a sphere into a finite number of pieces[8] and re-combine those pieces using only rigid rotations and translations into two solid spheres equal in volume to the first [3]. This result, which is so counterintuitive, is perhaps both the greatest failure and the greatest success of the Axiom of Choice. Entire branches of mathematics (measure theory for example) have been created to try and resolve this paradox and probe whether there are weaker versions of the Axiom of Choice that have easier to swallow consequences [1, 5].

Gödel proved that the Axiom of Choice was consistent with the other ZF axioms in 1939. In other words, he showed that the Axiom of Choice could not be *disproved* from the other ZF axioms. In 1963 Paul Cohen proved that the Axiom of Choice can not be *proved* from the ZF axioms either. Thus it is not determinable from the other set theory axioms whether or not the Axiom of Choice is true or false. It is then up to each mathematician to

[8] "Pieces" is a bit misleading because the pieces may just be a scattering of points which are not connected.

evaluate the usefulness and the oddness of the Axiom of Choice in order to determine its validity [1, 5].

Choice References

[1] Herbert B. Enderton. *Elements of Set Theory.* Academic Press, New York, 1977.

[2] John B. Fraleigh. *A First Course in Abstract Algebra.* Pearson, 7 edition, 2002.

[3] Horst Herrlich. *Axiom of Choice*, volume 1876. Springer, Berlin; New York, 2006. [Electronic Resource].

[4] Karel Hrbacek and Thomas J. Jech. *Introduction to Set Theory.* M. Dekker, New York, 3, rev. & expa. edition, 1999.

[5] Thomas J. Jech. *The Axiom of Choice.* Dover Publications, Mineola, N.Y., dover edition, 2008; 1973.

[6] Herman Rubin and Jean E. Rubin. *Equivalents of the Axiom of Choice.* North-Holland Publishing Company, 1963.

[7] Carol Schumacher. *Chapter Zero: Fundamental Notions of Abstract Mathematics.* Pearson, Reading, Mass., 2 edition, 2000.

[8] Wikipedia. Hartog's Number — Wikipedia, The Free Encyclopedia, October 2012. [Electronic Resource].