Axiom of Choice

We begin by understanding why we need the Axiom of Choice at all, and why it is a natural thing to assume. Consider the following thought experiment: we have infinitely-many bins (let's say for simplicity that they're numbered $0, 1, 2, \ldots$) and in teach bin is a pair of shoes. Is there a way to select from each bin a shoe? Sure, we can just take every left shoe, for example. But what if instead of a pair of shoes in each bin, we have a pair of socks. What then? If we are to assume that the two socks in a pair are entirely undistinguishable, how can we be sure we can always pick one, since we can create no rule by which to pick a sock from each pair? This is where the Axiom of Choice comes in. It guarantees that we can make these infinitely-many choices.

More formally, the axiom of choice says that given any family of sets S in which every element of S is non-empty, then there exists a function $f : S \to \bigcup S$ such that $f(s) \in s$ for each $s \in S$. We call such a function a **choice function** for S, which is where the Axiom of Choice gets its name.

In the above example, the fact that there were infinitely-many bins is important, as if there were only finitely-many of them, we would not have needed to appeal to the Axiom of Choice (from here on out referred to by "AC "). The reason is because of what it means to say that a set is non-empty. By saying that a set s is non-empty, we are implicitly saying that there exists a set t such that $t \in s$. When we have only finitely-many sets, we can write down that each is non-empty, extract an element from each set using this, and use the remaining axioms of set theory to define our choice function. This result is called **finite choice**, and it follows from the rest of the axioms of set theory, showing that the infinite-case is why we need the stronger axiom of AC .

It is worth noting that AC is *independent* of the remaining axioms of Set Theory, the set of which is called ZF , or *Zermelo-Fraenkel Set Theory.* This result was established by Paul Cohen using a method called *Forcing.*

Note that there are few different ways we could try to understand what exactly the Axiom of Choice is saying, along with some slight discussion of variants.

Most are familiar with the notion of the Cartesian product, namely that $X \times Y = \{(x, y) \mid x \in X, y \in Y\}$. We can use this same construction to define the Cartesian product more generally for any finite collection of sets. When we think about taking the Cartesian product of infinitely-many sets, this gets a bit more difficult, however. One approach is to think of ordered pairs (x, y) as a function $\{X, Y\} \to X \cup Y$, i.e. as a choice function for $\{X, Y\}$. In this way, for any family of sets S, we could define the Cartesian product $\prod S$ as the set of all choice functions for S. (It is important to note that this definition isn't really that useful when we want to have duplicates, but in this setting we consider *indexed families of sets*, i.e. a function $f : I \to S$, from an indexing set I into a family of sets S, and then define the Cartesian product to be the set of choice functions for the graph of f, i.e. the set $\{(i, f(i)) \mid i \in I\}$.)

A variation on the above definition of the Axiom of Choice is in requiring that the elements of S be pairwise disjoint. Call this new axiom AC'. Clearly AC implies AC', since it is simply a special case, but AC' also implies AC . This is because, given any family of non-empty sets S, we can turn it into a family of pair-wise disjoint sets by replacing S with the set $S' = \{s \times \{s\} \mid s \in S\}$; in this case, we tag the elements in s with the set they came from. Then by AC', there exists a choice function $f' : S' \to \bigcup S'$. The graph of f' has elements of the form $(s, (a, s))$ for some $a \in s$ for each $s \in S$. Thus, we define a choice function $f : S \to \bigcup S$ by letting its graph contain the pairs of the form (s, a).

In order to develop and intuition and appreciation for the Axiom of Choice, we prove many results which are implied by or equivalent to AC .

The two most important of these, in that they are essential tools in proofs using AC is the **Well-Ordering Theorem** and **Zorn's Lemma**.

Section 1

Well-Ordering Theorem

We start with the first, the Well-Ordering Theorem. To state and prove it, we must first say what a *well-order* is. We say that an ordered pair $(W, <)$ is a **strict well-ordered set** if $<$ is a *strict total order* (i.e. it is irreflexive, asymmetric, transitive, and trichotomous) on W such that given any non-empty subset V of W, V has a minimum element. Closely related to the notion of a well-ordered set is that of an *ordinal.* These can be thought of as the "canonical strict well-ordered sets". Technically, we say that a set α is an **ordinal** if $(\alpha, \in)$ is a strict well-ordered set and if α is **transitive**, meaning that if $\beta \in \alpha$, then $\beta \subset \alpha$. In other words, being transitive means that if β is an element of α, so are the elements of β. We won't let ourselves get bogged down by this relatively complicated definition. Instead, we need to recognize the following facts concerning strict well-ordered sets and ordinals:

Theorem 1.1

(a) For every strict well-ordered set $(W, <)$, there exists an ordinal α order-isomorphic to it.

(b) (Hartog's Lemma) For every set X, there exists an ordinal γ such that there exists no injection from γ into X.

(c) Any set of ordinals is strict well-ordered by the $\in$-relation, e.g. given any two ordinals α, β, we have either $\alpha \in \beta$, $\beta \in \alpha$, or $\alpha = \beta$.

(d) An ordinal α is the set of all ordinals less than it.

(e) Transfinite Recursion: If α is an ordinal, X a set, and $G : \{(\beta, h : \beta \to X) \mid \beta \in \alpha\} \to X$ any function, then there exists a function $f : \alpha \to X$ such that $G(\beta, f|_\beta) = f(\beta)$.

Closely related is the notion of a *cardinal.* An ordinal α is said to be a **cardinal** if $\beta \in \alpha$ implies there is no bijection from β onto α. In this sense, cardinals are the *canonical representatives of size.*

With these results in mind, we state and prove the Well-Ordering Theorem:

Theorem 1.2: Well-Ordering Theorem

Given any set X, there exists a strict well-order $<$ on X.

Proof.
By Hartog's Lemma, we know that there is an ordinal γ such that there is no injection from γ into X. By the Axiom of Choice, we know that there exists a choice function for $\mathcal{P}(X) \setminus \{\varnothing\}$. We use this to define the well-order $<$ on X recursively, using the Transfinite Recursion Theorem.

Let $Y = \mathcal{P}(X) \setminus \{\varnothing\}$. We define $G : \{(\alpha, h : \alpha \to X) \mid \alpha \in \gamma\} \to X$ as follows:

$$G(\alpha, h : \alpha \to X) := \begin{cases} f(X \setminus h[\alpha]) & \text{if } h[\alpha] \neq X \\ f(X) & \text{otherwise} \end{cases}$$

Then transfinite-recursion says that we have a function $g : \gamma \to X$ such that $G(\alpha, g|_\alpha) = g(\alpha)$, i.e. that $g(\alpha) = f(X \setminus g[\alpha])$ if $g[\alpha] \neq X$ and $g(\alpha) = f(X)$ otherwise. If $g[\alpha] = X$ for some $\alpha \in \gamma$, then we choose β to be the least such element of γ for which $g(\beta) = X$. Then $g|_\beta$ is injective because if $\alpha_1 < \alpha_2 < \beta$, we have $g(\alpha_2) = f(X \setminus g[\alpha_2]) \in X \setminus g[\alpha_2]$. Since $g(\alpha_1) \in g[\alpha_2]$, we see that $g(\alpha_1) \neq g(\alpha_2)$. Thus, $g|_\beta$ gives us a bijection between β and X, and we can use this to embue X with a strict well-order.

Such a β must exist, because by using the same kind of argument, if such a β did not exist, then g would be an injection, contradicting our assumption that γ had no injection into X. □

This is an extremely powerful result, and admits the following immediate corollaries:

Corollary 1.3

Every set S is in bijection with some cardinal.

Proof.
We know that S can be strict well-ordered, so there is some ordinal α which is order-isomorphic to S under this ordering, i.e. is in bijection with S. There is then some cardinal κ in bijection with α, and so κ is in bijection with S. □

Given any set S, we use $|S|$ to denote the *unique* cardinal in bijection with S, called the **cardinality of** S.

Corollary 1.4: Trichotomy Theorem

Given any two sets X, Y, either X injects into Y or Y injects into X.

Proof.
We know that $|X| \subset |Y|$ or $|Y| \subset |X|$, so the inclusion functions give injections. Using the bijections from X onto $|X|$ and of Y onto $|Y|$ finishes it off. □

In effect, this says that all sizes are comparable, and that all sizes are represented by some cardinal.

Conversely, we can show that if instead of AC we assume the Well-Ordering Theorem, then we can prove AC :

Theorem 1.5

The Well-Ordering Theorem implies AC .

Proof.
Let S be a family of non-empty sets. Let $X = \bigcup S$. Then by the Well-Ordering Theorem, there exists a strict well-order $<$ on X. Then define $f : S \to X$ by $f(s) =$ least element of s. □

Section 2

Zorn's Lemma

Now we move on to Zorn's Lemma. Zorn's Lemma relates to the existence of maximal elements in a poset, and this applies to many problems. Recall that a **poset** is a pair $(P, \leq)$ where $\leq$ is a relation on P satisfying the properties of reflexitivity, anti-symmetry, and transitivity. A chain in a poset $(P, \leq)$ is a subset of P, which when $\leq$ is restricted to it becomes a total order. Then Zorn's Lemma says that if every chair in P has an upper bound, then there exists a maximal element of P. On the way to proving it, we begin with an intermediary result:

Theorem 2.1: Hausdorff Maximality Principle

Let $(P, \leq)$ be a poset. Then there exists a maximal chain, i.e. there exists a chain C in P such that if C' is another chain with $C \subset C'$, then $C = C'$.

Proof.
By the Well-Ordering Theorem there exists a strict well-order $<_w$ on P. Let α be the ordinal order-isomorphic to $(P, <_w)$ and $g : \alpha \to P$ the order-isomorphism witnessing this. We want to use this strict well-ordering to recursively define a maximal chain using Transfinite Recursion.

Let $p = g(\varnothing) = g(0)$ be the $<_w$-least element of P, and then define a recursion formula

$F : \{(\beta, h : \beta \to P) \mid \beta \in \alpha\}$ as follows:

$$F(\beta, h : \beta \to P) := \begin{cases} g(\beta) & \text{if } \{g(\beta)\} \cup h[\beta] \text{ is a chain in } P \\ p & p \text{ otherwise} \end{cases}$$

Then by Transfinite Recursion we know that there exists a function $f : \alpha \to P$ such that $F(\beta, f|_\beta) = f(\beta)$. Note that when $\beta = 0 = \varnothing$, we know that $F(\varnothing, f|_\varnothing) = p$ since $g(\varnothing) = p$ and $\{p\}$ is a chain in P. Thus, by construction we know that the image $f[\alpha]$ is a chain in P. Moreover, it is a maximal chain because if there were x such that $\{x\} \cup f[\alpha]$ is a chain in P, then $\{x\} \cup f[g^{-1}(x)]$ would have been a chain and thus $f(g^{-1}(x)) = x$. This shows that $f[\alpha]$ is maximal. □

Using this, we prove Zorn's Lemma:

Theorem 2.2: Zorn's Lemma

Let $(P, \leq)$ be a poset, and suppose that every chain in P has an upper bound. Then there exists a maximal element of P.

Proof.
Let $(P, \leq)$ be as described. By Hausdorff's Maximality Principle, we know that there exists a maximal chain C in P. By hypothesis, there exists an upper bound u for C. We claim that u is maximal. If not, then there is v such that $u < v$, so because $w \leq u$ for every $w \in C$, we see that $\{v\} \cup C$ is a chain, and so $\{v\} \cup C = C$. But this means that $v \in C$, so because u is an upper bound of C, we find that $v \leq u$, giving us a contradiction.

Thus, u is a maximal element. □

Zorn's Lemma has many applications to Order Theory and Algebra. We explore some of the most important results.

The first application is to vector spaces. Suppose that $\mathbb{F}$ is a field (e.g. $\mathbb{R}, \mathbb{C}, \mathbb{Q}$). Then a **$\mathbb{F}$-vector space** is a quadruple $(V, +, \mathbf{0}, \cdot)$ where the elements of V are called **vectors**, $+ : V \times V \to V$ is called **vector addition**, $\mathbf{0}$ is called the **zero vector**, $\cdot : \mathbb{F} \times V \to V$ is called **scalar multiplication**, and the elements of $\mathbb{F}$ called **scalars** satisfying the following axioms:

(i) $+$ is associative, commutative, $\mathbf{0}$ is an identity, and there exist inverse.

(ii) Compatibility with Unity: $1 \cdot \mathbf{v} = \mathbf{v}$ for every $\mathbf{v} \in V$, where 1 is the multiplicative identity of $\mathbb{F}$.

(iii) Compatibility with Field Multiplication: for any vector $\mathbf{v}$ and scalars λ, μ, we have $(\lambda\mu) \cdot \mathbf{v} = \lambda \cdot (\mu \cdot \mathbf{v})$.

(iv) Distributivity of Scalar Multiplication over Vector Addition: for any vectors $\mathbf{v}, \mathbf{w}$ and scalar λ, we have $\lambda \cdot (\mathbf{v} + \mathbf{w}) = (\lambda \cdot \mathbf{v}) + (\lambda \cdot \mathbf{w})$.

(v) Distributivity of Field Addition over Scalar Multiplication: for any vector $\mathbf{v}$ and scalars λ, μ, we have $(\lambda + \mu) \cdot \mathbf{v} = (\lambda \cdot \mathbf{v}) + (\mu \cdot \mathbf{v})$.

Vector spaces are generalizations of the familiar structures of $\mathbb{R}^n$ and $\mathbb{C}^n$, and also include the space $\mathbb{R}^{n \times m}$ of $n \times m$ matrices over $\mathbb{R}$, the space $\mathbb{R}[x]$ of single-variable polynomials over $\mathbb{R}$, and even $\mathbb{C}$ over $\mathbb{R}$ or $\mathbb{R}$ over $\mathbb{Q}$. In the former cases, there is a special invariant called the *dimension* of the vector space (which depends on the field we're working over). This is the size of a *basis*, which is a set of vectors which is *linearly independent* and which *spans* the vector space.

Given a set S of vectors in V, we say that it is **$\mathbb{F}$-linearly-independent** if $\sum_{i=1}^m \lambda_i \mathbf{v}_i = \mathbf{0}$ implies that $\lambda_i = 0$ for each i, where $\mathbf{v}_1, \ldots, \mathbf{v}_m \in S$. In other words, it means that no vector in S can be written as a *linear combination* of other vectors in S. We then define the **span** of S to be the set

$$\operatorname{span}(S) := \left\{ \sum_{i=1}^m \lambda_i \mathbf{v}_i \mid \mathbf{v}_1, \ldots, \mathbf{v}_m \in V, \lambda_1, \ldots, \lambda_m \in \mathbb{F}, m \geq 0 \right\}$$

This can be checked to be a $\mathbb{F}$-vector space as well. We say that S **spans** V if $\operatorname{span}(S) = V$. Finally, we say that S is a **basis** for V if S is linearly independent and spans V.

For the cases of the $\mathbb{R}$-vector space $\mathbb{R}^n$, such a basis is $(1,0,\dots,0),(0,1,0,\dots,0),\dots,(0,\dots,0,1)$. A basis for $\mathbb{R}^{n\times m}$ consists of those $n \times m$ matrices which have exactly one entry equal to 1 and the rest equal to 0. Finally, a basis for $\mathbb{R}[x]$ is given by $1, x, x^2, x^3, \dots$. But what of the $\mathbb{Q}$-vector space $\mathbb{R}$? It ends up that this question, and its generalization about whether every vector space has a basis, requires AC .

To prove that every vector space has a basis, we make use of an alternative definition of a basis, namely that S is a basis for V if it is a maximal linearly-independent set of vectors. To get some intuition for converting problems into maximality problems, we shall prove this. Suppose that S is a basis for V, and suppose for the sake of a contradiction that it is not a maximal linearly-independent set of vectors, i.e. there exists some v such that $S \cup \{v\}$ is linearly-independent. But this implies that $v \notin \operatorname{span}(S)$, contradicting the fact that S spans V. Conversely, suppose that S is a maximal linearly-independent set. We claim that $\operatorname{span}(S) = V$, as otherwise there is some $v \notin \operatorname{span}(S)$. But then $S \cup \{v\}$ is linearly-independent, contradicting the assumed maximality. Thus, S is a basis.

Theorem 2.3

Every $\mathbb{F}$-vector space V has a basis.

Proof.
To envoke Zorn's Lemma, we need a poset in which every chain has an upper bound. Since we want to find a *maximal* $\mathbb{F}$-linearly-independent set of vectors, the natural choice is to consider the set P of all linearly-independent subsets of V, ordered by $\subset$. Let C be a chain in P; we want to show that C has an upper bound in P. Indeed, we claim that $\bigcup C$ is an upper bound. To prove this, suppose that there are $\mathbf{v}_1, \dots, \mathbf{v}_m \in \bigcup C$ and scalar $\lambda_1, \dots, \lambda_m$ such that $\sum_{i=1}^m \lambda_i \mathbf{v}_i = \mathbf{0}$. Each of the vectors $\mathbf{v}_i$ originated from some element S_i in C. But because C is linearly-ordered by $\subset$, we know that $\bigcup_{i=1}^m S_i$ is equal to some S_j, since one of them is the largest (the finiteness of linear combinations is important here). But then this contradicts the fact that S_j is linearly-independent, giving us a contradiction.

Thus, every chain has an upper bound, so that by Zorn's Lemma there exists a maximal linearly-independent set S. This is our basis. □

We end by showing that Zorn's Lemma (and so Hausdorff's Maximality Principle) are equivalent to AC by showing that Zorn's Lemma implies the Axiom of Choice:

Theorem 2.4

Zorn's Lemma implies the Axiom of Choice.

Proof.
Suppose that S is a family of non-empty sets. Our approach will be to consider the poset $(P, \leq)$ whose elements are all the partially-defined choice functions for S (i.e. they are a choice function for a subset of S) ordered by saying that $f : T \to \bigcup T \leq g : U \to \bigcup U$ if $T \subset U$ and $f(s) = g(s)$ for every $s \in U$. In other words, g is an extension of f (equivalently, f is a restriction of g).

To see that every chain in P has an upper bound in X, suppose we have a chain C. The elements of C are of the form $f : T_f \to \bigcup T_f$, so define $g : \bigcup_{f\in C} T_f \to \bigcup\bigcup_{f\in C} T_f$ to have graph consisting of the unions of the graphs of the functions f. This is well-defined by the fact that C is totally-ordered by the extension relation, and is a choice function on $\bigcup_{f\in C} T_f$ because each of the functions f were choice functions on their domains.

Thus, P has a maximal element h. We claim that $h : T \to \bigcup T$ is a choice function for S, i.e. $T = S$. If not, then there is some $s \notin T$, and by the non-emptiness of s we know that there is an element $a \in s$. But then we can extend h by adding (s, a) to the graph, giving a contradiction. Thus, $T = S$, and h is the desired choice function. □

Section 3

More Implications

Above we have proven some of the most widely-used results which follow from AC . However, there is an extremely large number of such results. To get some idea of how many, we prove even more results making use of AC , the Well-Ordering Theorem, and Zorn's Lemma:

Theorem 3.1

Every surjective function has a right inverse.

Proof.
Suppose $f : X \to Y$ is surjection. Consider the family S of fibers $f^{-1}(y)$ for $y \in Y$. By AC there exists a choice function $g : S \to X$. Then let $h : Y \to S$ be the function sending y to the fiber $f^{-1}(y)$. We claim that $(g \circ h)$ is a right inverse of f. Indeed, given $y \in Y$, $(f \circ (g \circ h))(y) = (f \circ g)(f^{-1}(y))$, and $g(f^{-1}(y))$ is an element x such that $f(x) = y$, so $(f \circ (g \circ h))(y) = y$. □

Theorem 3.2

A countable union of countable sets is countable.

Proof.
Suppose T is a countable family of countable sets. Let $h : \mathbb{N} \to T$ be a surjection. Then let S be the set whose elements are the sets of surjections from $\mathbb{N}$ to an element in T. By AC , there exists a choice function $f : S \to \bigcup S$, so in particular for each element t of T, we can select a surjection $f_t : \mathbb{N} \to t$. Then define a surjection $g : \mathbb{N} \times \mathbb{N} \to T$ by sending (n, m) to $f_{h(n)}(m)$.

All that remains is to show that $\mathbb{N} \times \mathbb{N}$ is countable. To see this, we define an injection of $\mathbb{N} \times \mathbb{N}$ into $\mathbb{N}$ by sending (n, m) to $2^n 3^m$. □

Theorem 3.3

For every partition Π of a set S, there exists a canonical set of representatives for that partition.

Proof.
By AC there exists a choice function $f : \Pi \to S$. Its image is exactly the required canonical set of representatives. □

Theorem 3.4

Every Dedekind-Infinite set is infinite.

Proof.
A set S is **Dedekind-infinite** if there exists a bijection $f : S \to T$ for some proper subset T of S. To show that every Dedekind-infinite set is infinite, we show that every Dedekind-infinite set has a countable subset. By the AC , we know that every set is in bijection with some cardinal. Let $\kappa = |S|$. We need to show that κ is infinite, and to show that, we show that if κ is finite, then any proper subset of κ must have strictly smaller size.

We prove this by induction, with the claim that if $|X| = n$ and Y is a proper subset of X, then $|Y| < n$. This holds trivially for $n = 0$ because of the fact that there is no proper subset of $\varnothing$. Now suppose that the result is true for n, and suppose $|X| = n + 1$ and Y a proper subset of X. Suppose for the sake of a contradiction that $|Y| = n + 1$, and let $f : Y \to n + 1$ by a bijection realizing this. Let $x \notin Y$, and define $g : Y \cup \{x\} \to n + 2$ by $g(y) = f(y)$ for $y \in Y$ and $g(x) = n + 1$. $Y \cup \{x\} \subset X$, so

$n + 2 \leq |X|$, giving a contradiction. □

Theorem 3.5: Alexander Subbasis Theorem

Let X be a topological space with a subbasis B. If every subcollection of $\mathcal{B}$ that covers X has a finite subcover, then X is compact.

Proof.
Assume for the sake of a contradiction that every subcollection of $\mathcal{B}$ that covers X has a finite subcover, but X is not compact. If we let $\mathcal{P}$ be the collection of al open covers that do not have a finite subcover, then $\mathcal{P}$ is not empty. We partially order this by set inclusion.

Let $\mathcal{C}$ be a chain in $\mathcal{P}$, and let $C = \bigcup_{S \in \mathcal{C}} S$. Suppose that C has a fintie subcover, say $C_0 = \{U_1, U_2, \ldots, U_n\}$. For each i, $1 \leq i \leq n$, there is S_i such that $U_i \in S_i \in \mathcal{C}$. Since $\mathcal{C}$ is totally ordered, $\{S_i \mid 1 \leq i \leq n\}$ is also totally ordered, and since it is finite has a maximum element $C' = S_j$ for some j, $1 \leq j \leq n$. But then $C_0 \subset C' \in \mathcal{C}$, and C' has a finite subcover, contradicting the fact that $\mathcal{C}$ is a chain in $\mathcal{P}$. Thus, we see that C does not have a finite subcover, and C is an upper bound of $\mathcal{C}$.

We can thus use Zorn's Lemma to say that there is a maximal open cover C that does not have a finite subcover. Now we take $C \cap B$. If this covers X, then by hypothesis it has a finite subcover, and thus so does C. Thus, it does not cover X, so that there is $x \in C \cap B$. Since C is an open cover, there is $U \in C$ such that $x \in U$. Since B is a subbasis, there is $S_1, S_2, \ldots, S_m \in B$ with $x \in S_1 \cap S_2 \cap \cdots \cap S_m \subset U$. Since C is maximal and, for each i, $S_i \notin C$, it follows that $C \cup \{S_i\}$ has a finite subcover C_i. But then $U \cup \bigcup C_i$ is a finite subcover.

This leads us to a contradiction, and thus $\mathcal{P}$ is empty, meaning X is compact. □

This gives us a simple proof of the Tychonoff Theorem:

Theorem 3.6: Tychonoff Theorem

Given X_α compact for each $\alpha \in J$, $\prod_{\alpha \in J} X_\alpha$ is compact in the product topology.

Proof.
We use the Alexander Subbasis Theorem to give a short and simple proof. The subbasis of the product topology on $\prod_{\alpha \in J} X_\alpha$ is given by $B = \{\pi_\beta^{-1}(U) \mid \beta \in J \text{ and } U \subset X_\beta \text{ open}\}$.

Suppose for the sake of a contradiction that there is a subcollection C of B that covers $\prod_{\alpha \in J} X_\alpha$ that does not have a finite subcover. Take $C = \bigcup_{\alpha \in J} C_\alpha$ where $\pi_\alpha(C_\alpha)$ is a subcollection of open sets in X_α. For each $\beta \in J$, C_β has no a finite subcover, so that $\pi_\beta(C_\beta)$ has no finite subcover, as if it did, then C_β would as well, since C_β is a collection of open sets of X_β crossed with the remaining X_α for $\alpha \in J$ unequal to β. Since X_β is compact, this means that $\pi_\beta(C_\beta)$ must not be a open cover of X_β, and there is an element x_β not in $\pi_\beta(C_\beta)$. Using the Axiom of Choice, we choose for each $\alpha \in J$ such an x_α to give us an element $(x_\alpha)_{\alpha \in J}$ that is not covered by C, giving us a contradiction that C covered $\prod_{\alpha \in J} X_\alpha$.

Thus, every subcollection C o fB that covers $\prod_{\alpha \in J} X_\alpha$ has a finite subcover, so that by the Alexander Subbasis Theorem, we see that $\prod_{\alpha \in J} X_\alpha$ is compact. □

Theorem 3.7: Existence of Unmeasurable Sets

There exists a subset of $(0, 1]$ that is not Lebesgue measurable.

Proof.
We shall define an equivalence relation on $(0, 1]$ by saying that $x \simeq y$ if $x - y \in \mathbb{Q}$. There are necessarily uncountably-many equivalence classes. Using the axiom of choice, there exists a set E such that E

contains precisely one element from each equivalence class. Let $\mathbb{Q}^* := \mathbb{Q} \cap (0,1]$, and for each $q \in \mathbb{Q}^*$ we take $E + q = \{q + x \mod 1 \mid x \in E\}$. Note that $(E+q) \cap (E+r) = \varnothing$ if $q, r \in \mathbb{Q}^*$, because otherwise $x+q = y+r$, and thus $x-y = r-q$ is rational, meaning that $x \sim y$, contradicting the fact that E contains exactly one element from each equivalence class. Next, by construction we have $(0,1] = \bigcup_{q \in \mathbb{Q}^*} (E+q)$.

Our aim for a contradiction is to suppose that E is measurable. Then $E+q$ is also measurable and moreover by the (modulo 1 version of) translation invariance of the Lebesgue measure (which we will soon prove), we see that $\mu(E+q) = \mu(E)$. But then $\mu((0,1]) = 1 = \sum_{k=0}^{\infty} \mu(E+q_k) = \sum_{k=0}^{\infty} \mu(E)$. If E is measurable, then $\mu(E)$ is either 0 or a positive number; it cannot be 0 for then the sum is 0 and we have $1 = 0$, a contradiction, and it cannot be a positive number for then the sum is ∞ and we have $1 = \infty$. Thus, E cannot be measurable. □

Theorem 3.8: Hahn-Banach Theorem

Let X be a normed space, Y a subspace, and $g \in Y^*$. Then there is $f \in X^*$ such that $f|_Y = g$, and $\|f\| = \|g\|$.

Proof.
First assume that $\mathbb{F} = \mathbb{R}$, so X is a real normed space. Without loss of generality we can assume that $\|g\| = 1$. Now, if $Y = X$, then we are done. Otherwise, we can pick $x \in X \setminus Y$, and let $Y_1 = Y \oplus \text{span}\{x_1\}$. Then we define $f_1 : Y_1 \to \mathbb{R}$ by sending $f_1(y + \lambda x_1) = g(y) + \lambda\alpha$ where $y \in Y, \lambda \in \mathbb{R}$, and $\alpha \in \mathbb{R}$ is to be determined. It is clear that $f_1|_Y = g$, so we just need $\alpha \in \mathbb{R}$ so that $\|f_1\| \le 1$; it is automatically at least as large as 1 by the fact that $\|g\| = 1$. Then we want $|g(y) - \lambda\alpha| \le \|y + \lambda x_1\|$ for all $y \in Y$ and $\lambda \in \mathbb{R}$, which is equivalent to $|g(y) + \alpha| \le \|y + x_1\|$ for all $y \in Y$ (we divide both sides by $|\lambda|$), which is equivalent to $g(y) + \alpha \le \|y + x_1\|$ and $-(g(y)+\alpha) \le \|y + x_1\|$. This is equivalent to $-g(z) - \|z + x_1\| \le \alpha \le \|y + x_1\| - g(y)$ for all $y, z \in Y$. Such an α exists if and only if $-g(z) - \|z + x_1\| \le \|y + x_1\| - g(y)$ for all $y, z \in Y$. Then

$$-g(z) + g(y) = g(-z+y) \le \|-z+y\| = \|-z - x_1 + y + x_1\| \le \|z + x_1\| + \|y + x_1\|.$$

We consider $P = \{(Z,h) \mid Y \subset Z \subset X, h \in Z^*, h|_Y = g, \|h\| = 1\}$. We make this a partially ordered set by setting $(Z_1, h_1) \le (Z_2, h_2)$ if and only if $Z_1 \subset Z_2$ and $h_2|_{Z_1} = h_1$. P is non-empty because $(Y, g) \in P$, and given a chain $C = \{(Z_i, h_i) \mid i \in I\}$ in P this has an upper bound by taking $Z = \bigcup_{i \in I} Z_i$ and taking $h(z) = h_i(z)$ for $i \in I, z \in Z_i$, which is well-defined by the partial ordering on P. Then $(Z,h) \in P$ and $(Z,h) \ge (Z_i, h_i)$. Then by Zorn's Lemma, we find that P has a maximal element (W, f). Now, it must be that $W = X$ because otherwise we can extend W as in the first part of our proof to contradict the maximality of (W, f).

Now we approach the case where $\mathbb{F} = \mathbb{C}$. Let $X_{\mathbb{R}}$ be the space X viewed as a real normed space. For $f \in X^*$, we let $\Re(f)$ be defined by $x \mapsto \Re(f(x))$. This is $\mathbb{R}$-linear and

$$|\Re(f)(x)| = |\Re(f(x))| \le |f(x)| \le \|f\| \, \|x\|$$

so $\Re(f) \in (X_{\mathbb{R}})^*$ and $\|\Re(f)\| \le \|f\|$. We show that $f \mapsto \Re(f)$ is a $\mathbb{R}$-linear isometric isomorphism $X^* \to (X_{\mathbb{R}})^*$.

Given $x \in X$, we choose $\lambda \in \mathbb{C}$ such that $|\lambda| = 1$ and $|f(x)| = \lambda f(x)$. Then $|f(x)| = f(\lambda x) = \Re(f(\lambda x)) \le \|\Re(f)\| \cdot \|\lambda x\| = \|\Re(f)\| \, \|x\|$. So $\|\Re(f)\| = \|f\|$.

Given $h \in (X_{\mathbb{R}})^*$, suppose that $\Re(f) = h$ and let $k = \text{im}(f)$. Then $f = h + ik$, but also $-if(ix) = -ih(ix) + k(ix)$, and thus (using $\mathbb{C}$-linearity) $f(x) = h(x) - ih(ix)$. This shows that f is uniquely identified by $\Re(f)$, and thus we have injectivity. This also shows surjectivity.

[[finish?]]

□