Informal Introduction to Set Theory

Kenneth A. Ross

These notes are based on ten lectures given in the Fall of 1988 at the University of Oregon. They are intended to give an introduction to Zorn's Lemma and its equivalents. Some results about cardinal and ordinal numbers are also given. The primary reference is sections 3 and 4 of Hewitt and Stromberg *Real and Abstract Analysis*. I thank Peter Horn at Northern Arizona University who shared some notes from a version of these lectures that I gave back in the 1960's.

If your understanding of basic logic and its use in mathematics is rusty, or if you never really understood the ground rules for proofs in mathematics, I highly recommend the readable and lively book by Robert S. Wolf, *Proof, Logic, and Conjecture: the mathematician's toolbox*, W. H. Freeman and Company, 1998.

There are several equivalent statements that cannot be proved from the usual set-theoretic axioms. These include Zorn's Lemma, the Axiom of Choice, and the Well Ordering Principle. I will present each of them and prove their equivalence later.

The most intuitive is the Axiom of Choice. For any indexed family of sets $\{A_\lambda\}_{\lambda\in\Lambda}$, the product of these sets is $\prod_{\lambda\in\Lambda} A_\lambda$. This is the set of all functions f from Λ into $\cup_{\lambda\in\Lambda} A_\lambda$ such that $f(\lambda)$ is in A_λ for all λ.

Axiom of Choice. If each A_λ is nonempty, then $\prod_{\lambda\in\Lambda} A_\lambda$ is nonempty.

I.e., there exists a choice function f on Λ that chooses an element $f(\lambda)$ in each A_λ. If each $A_\lambda = \{0\}$, then this is obvious: Just set $f(\lambda) = 0$ for all λ. If each $A_\lambda = \{0, 1\}$, then the same choice function f works. But if Λ is huge and all you know is that each A_λ has two points, then one cannot specify such an f. The Axiom of Choice is needed.

Our equivalent axioms will be less intuitive than the Axiom of Choice, but they are logically equivalent to it. You either accept all of them or none of them.

We need to discuss some orderings.

Definition. A *partial order* on a set X is a relation $\leq$ such that

(R) $x \leq x$ for all x in X [reflexive]

(AS) $x \leq y$ and $y \leq x$ imply $x = y$ [anti-symmetric]

(T) $x \leq y$ and $y \leq z$ imply $x \leq z$ [transitive].

Remarks.

(a) (T) makes it legal to write $x \leq y \leq z$ unambiguously.

(b) A relation like $\leq$ on X is "really" a subset R of $X \times X$, where we agree that (x, y) is in R if and only if $x \leq y$.

We need two more definitions; then we'll give some examples.

Definitions. A *linear* [total, complete, or simple] ordering is a partial ordering such that every two elements are comparable:

(L) given $x, y \in X$ either $x \leq y$ or $y \leq x$ (or both).

Linearly ordered sets are sometimes called *chains*.

A set is *well ordered* if it is linearly ordered and

(W) every nonempty subset A of X contains a least element, i.e., an element $\ell \in A$ so that $\ell \leq a$ for all $a \in A$.

Examples.

(a) Consider the set $\mathbf{R}$ of all real numbers with its usual ordering $\leq$. This set is linearly ordered, but not well ordered. The same comments apply to $[0, \infty)$, the set $\mathbf{Q}$ of rationals, and $\mathbf{Q}^+ = \{r \in \mathbf{Q} : r \geq 0\}$.

(b) Consider the set $\mathbf{Z}$ of all integers with the usual ordering. then $\mathbf{Z}$ is linearly ordered, but not well ordered. All finite linearly ordered sets are well ordered. The simplest infinite well ordered set is $\mathbf{N} = \{1, 2, 3, \ldots\}$.

(c) Here is a more complicated well ordered subset of $\mathbf{R}$: $\{m - \frac{1}{n} : n, m \in \mathbf{N}\}$. To see why this is well ordered, draw the real line and mark some of these numbers on it.

(d) Here is a well ordering of $\mathbf{N} \times \mathbf{N}$:

$$(m, n) \leq (m', n') \quad \text{if} \quad m < m' \quad \text{or if} \quad m = m' \text{ and } n \leq n'.$$

Thus $(1,1) \leq (1,2) \leq (1,3) \leq \cdots \leq (2,1) \leq (2,2) \leq (2,3) \leq \cdots$ etc.

This example is *order-isomorphic* to the example in (c) via the map

$$f(m, n) = m - \frac{1}{n}.$$

Definition. Let $f\colon X \to Y$ where $(X, \leq_X)$ and $(Y, \leq_Y)$ are partially ordered sets. f is *order-preserving* if

$$f(x_1) \leq_Y f(x_2) \quad \text{whenever} \quad x_1 \leq_X x_2.$$

If f is one-to-one and onto and both f and f^{-1} are order-preserving, then f is an *order isomorphism.*

More examples.

(e) Here's a nice partial order on $\mathbf{N}$: $m|n$ if m divides n, i.e., $\frac{n}{m}$ is an integer. For transitivity, note

$$m|n \text{ and } n|p \quad \Longrightarrow \quad \frac{n}{m} \text{ and } \frac{p}{n} \text{ are integers} \quad \Longrightarrow \quad \frac{n}{m}\frac{p}{n} = \frac{p}{m} \text{ is an integer} \quad \Longrightarrow \quad m|p.$$

This ordering is not a linear order: given 3 and 7, neither divides the other.

Here's a chain: $3|27|54|108|756|\cdots$. Here's the beginning of a maximal chain: $2|4|8|16|32|64|\cdots$.

(f) Let $\mathcal{P}(X)$ be the set of all subsets of some set X. Then inclusion $\subseteq$ is a partial order for $\mathcal{P}(X)$. If X has two distinct elements a and b, then $\subseteq$ is not a linear order on $\mathcal{P}(X)$: $\{a\}$ and $\{b\}$ are not comparable, i.e., neither is a subset of the other.

Of course, $\mathcal{P}(X)$ has chains: $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$. Here is a maximal chain in $\mathcal{P}(\mathbf{N})$: $\emptyset, \{1\}, \{1,2\}, \{1,2,3\}, \{1,2,3,4\}, \ldots$.

(g) Let G be a group. The set of subgroups of G is a partially ordered set under inclusion $\subseteq$.

(h) Let $C([0,1])$ be the set of all continuous real-valued functions on $[0,1]$ and write $f \leq g$ if $f(x) \leq g(x)$ for all x in $[0,1]$. This is a fine partial ordering, but not a linear ordering. For example, the functions $f(x) = x$ and $g(x) = 1 - x$ are not comparable.

It is easy to find chains $f_1 \leq f_2 \leq f_3 \leq \cdots$ in $C([0,1])$. Maximal chains in this partially ordered set are not particularly useful and tend to be weird.

(i) Here is an **important**, more sophisticated partially ordered set. Let $\mathcal{F}$ be a set of functions whose domains are a subset of a fixed set X, and whose ranges are in some set Y. Define $f \leq g$ if g *extends* f, i.e.,

$$f \leq g \quad \text{if} \quad \text{dom}(f) \subseteq \text{dom}(g) \quad \text{and} \quad f(x) = g(x) \quad \text{for all} \quad x \in \text{dom}(f).$$

Note. $\leq$ is simply inclusion if we view functions as relations, i.e., as subsets of $X \times Y$. To see this, think what $f \subseteq g$ would have to mean.

Proposition. Let $\mathcal{F}$ be as above, and let $\mathcal{C}$ be a chain in $\mathcal{F}$. Then $h = \cup_{f \in \mathcal{C}} f$ is a function whose domain is the union of all the domains $\text{dom}(f)$, $f \in \mathcal{C}$, and whose range is the union of all the ranges $\text{range}(f)$, $f \in \mathcal{C}$.

Proof. To see that h is a function, suppose (x, y_1) and (x, y_2) are in h. We need to show that $y_1 = y_2$. By the definition of h, there exist f_1, f_2 in $\mathcal{C}$ so that $(x, y_1) \in f_1$ and $(x, y_2) \in f_2$. Either $f_1 \subseteq f_2$ or $f_2 \subseteq f_1$, say $f_1 \subseteq f_2$. Then (x, y_1) and (x, y_2) both belong to f_2. Since f_2 is a function, we conclude that $y_1 = y_2$.

Clearly $\text{dom}(f) \subseteq \text{dom}(h)$ for all f in $\mathcal{C}$, so $\cup_{f \in \mathcal{C}} \text{dom}(f) \subseteq \text{dom}(h)$. Now suppose that x is in $\text{dom}(h)$. Then (x, y) is in h for some $y \in Y$. Hence (x, y) is in f for some f in $\mathcal{C}$; hence x belongs to $\text{dom}(f)$. This shows that $\text{dom}(h) \subseteq \cup_{f \in \mathcal{C}} \text{dom}(f)$. A similar argument shows that $\text{range}(h) = \cup_{f \in \mathcal{C}} \text{range}(f)$. ♣

Hausdorff Maximality Principle (1915). Every nonempty partially ordered set has a maximal chain in it.

This principle is not used very much any more, because it is very closely related to Zorn's Lemma. Note that this principle and the last proposition together show that, given nonempty sets X and Y, there exists a function $f: X \to Y$.

Definitions. Let X be a partially ordered set. For a subset A of X and x in X, x is an *upper bound* for A if $a \leq x$ for all a in A. x_0 is a *maximal element* for X if $x_0 \leq x$ implies $x_0 = x$, i.e., nothing in X is bigger than x_0.

Examples. (f) again. In $\mathcal{P}(X)$, if $\mathcal{A} \subseteq \mathcal{P}(X)$ then $\cup_{A \in \mathcal{A}} A$ is an upper bound for $\mathcal{A}$. The whole space X is the unique maximal element in $\mathcal{P}(X)$.

(i) again. In $\mathcal{F}$, any function with domain X is a maximal element. So $\mathcal{F}$ might have many maximal elements. In the proposition following (i), h is an upper bound for $\mathcal{C}$.

Zorn's Lemma (1930). Let $(X, \leq)$ be a partially ordered set. Suppose
(i) $X \neq \emptyset$,
(ii) every nonempty chain C in X has an upper bound (in X).
Then X has a maximal element.

Proof from Maximality Principle. Select a maximal chain C. Let u be an upper bound in X for C: u is in X and $u \geq c$ for all c in C. Now u is a maximal element for X, since otherwise there exists an x in X such that $x \geq u$ and $x \neq u$, and then $C \cup \{x\}$ would be a bigger chain in X. ♣

Examples.

1) Let R be a commutative ring with unit 1. Then every ideal $I \neq R$ is contained in a maximal proper ideal M of R.

Proof. Let $\mathcal{X}$ consist of all proper ideals $J \supseteq I$, ordered by inclusion. $\mathcal{X} \neq \emptyset$ because I is in $\mathcal{X}$.

Let $\mathcal{C}$ be a chain of proper ideals in $\mathcal{X}$. Let $J_0 = \cup_{J \in \mathcal{C}} J$. Then J_0 is easily seen to be an ideal. [For example, suppose a_1 and a_2 are in J_0. Then there exist ideals J_1 and J_2 in $\mathcal{C}$ such that $a_1 \in J_1$ and $a_2 \in J_2$. Now $J_1 \subseteq J_2$ or $J_2 \subseteq J_1$, say $J_1 \subseteq J_2$. Then a_1 and a_2 are both in J_2, so the sum $a_1 + a_2$ is in $J_2 \subseteq J_0$.] Since $1 \notin J$ for all J in $\mathcal{C}$, it follows that $1 \notin J_0$. Hence J_0 is proper. I.e., J_0 is in $\mathcal{X}$ and J_0 is an upper bound for $\mathcal{C}$. By Zorn's Lemma, $\mathcal{X}$ has a maximal element M. ♣

2) Let G be a group and H an abelian subgroup. Then there exists a maximal abelian subgroup J of G so that $H \subseteq J$.

Proof. Let $\mathcal{X}$ consist of all abelian subgroups J of G such that $H \subseteq J$, ordered by inclusion. $\mathcal{X} \neq \emptyset$ because H is in $\mathcal{X}$.

Let $\mathcal{C}$ be a chain in $\mathcal{X}$. Let $J_0 = \cup_{J \in \mathcal{C}} J$. Easily J_0 is an abelian subgroup of G. [For example, given a_1 and a_2 in J_0, as before there exists J_2 in $\mathcal{C}$ containing both a_1 and a_2. Since J_2 is abelian, we conclude that $a_1 a_2 = a_2 a_1$.] So J_0 is in $\mathcal{X}$ and J_0 is an upper bound for $\mathcal{C}$.

By Zorn's Lemma, $\mathcal{X}$ has a maximal element. ♣

3) Let X be a metric space with metric d, and consider a subset S so that

$$d(x, y) \geq 1 \quad \text{whenever} \quad x \quad \text{and} \quad y \quad \text{are in} \quad S \quad \text{and} \quad x \neq y.$$

Then there exists a maximal set $S_0 \supseteq S$ so that

$$d(x, y) \geq 1 \quad \text{whenever} \quad x \quad \text{and} \quad y \quad \text{are in} \quad S_0 \quad \text{and} \quad x \neq y.$$

Proof. Let $\mathcal{X}$ be the collection of all sets $S' \supseteq S$ such that

$$d(x, y) \geq 1 \quad \text{whenever} \quad x \quad \text{and} \quad y \quad \text{are in} \quad S' \quad \text{and} \quad x \neq y.$$

Again $\mathcal{X}$ is ordered by inclusion. Clearly $\mathcal{X} \neq \emptyset$ since S is in $\mathcal{X}$.

Easily every chain in $\mathcal{X}$ has an upper bound in $\mathcal{X}$; again it's the union. So by Zorn's Lemma, there exists a maximal element in $\mathcal{X}$. ♣

4) Let H be a subgroup of an abelian group $(G, +)$. If $h\colon H \to (\mathbf{R}, +)$ is a group homomorphism, then h extends to a group homomorphism of G into $\mathbf{R}$.

Proof. For such results there are typically two steps:

(i) get a maximal extension using Zorn's Lemma [the easy step];

(ii) extend it further unless the maximal extension is the desired extension.

(i) Let $\mathcal{F}$ be the set of all extensions g of h such that $\mathrm{dom}(g)$ is a subgroup of G and such that $g\colon \mathrm{dom}(g) \to \mathbf{R}$ is a group homomorphism. $\mathcal{F}$ is partially ordered as in Example (i) on page 3. Since h itself belongs to $\mathcal{F}$, we see that $\mathcal{F} \neq \emptyset$.

Consider a chain $\mathcal{C}$ in $\mathcal{F}$. Let $g_0 = \cup_{g \in \mathcal{C}} g$. As noted in the proposition on page 3, g_0 is a function whose domain is the union of the domains $\mathrm{dom}(g)$, $g \in \mathcal{C}$. Since each $\mathrm{dom}(g)$ is a subgroup of G, it follows easily that $\mathrm{dom}(g_0)$ is a subgroup of G. To see that g_0 is a homomorphism, consider x_1 and

x_2 in $\mathrm{dom}(g_0)$. Then there exist g_1 and g_2 in $\mathcal{C}$ so that x_1 is in $\mathrm{dom}(g_1)$ and x_2 is in $\mathrm{dom}(g_2)$. Either $g_1 \le g_2$ or $g_2 \le g_1$, say $g_1 \le g_2$. Then both x_1 and x_2 are in $\mathrm{dom}(g_2)$, so $g_2(x_1-x_2) = g_2(x_1)-g_2(x_2)$. Hence $g_0(x_1 - x_2) = g_0(x_1) - g_0(x_2)$. So g_0 is a group homomorphism. Hence g_0 is in $\mathcal{F}$, and g_0 is an upper bound for $\mathcal{C}$.

Now by Zorn's Lemma, $\mathcal{F}$ has a maximal element g_∞. If $\mathrm{dom}(g_\infty) = G$, then we're done.

(ii) Claim. Otherwise g_∞ extends to a bigger element in $\mathcal{F}$ (a contradiction).

Prf. Let $H_\infty = \mathrm{dom}(g_\infty)$ and select y in $G \setminus H_\infty$. Let H^* be the group generated by $\{H_\infty, y\}$, i.e.,

$$H^* = \{x + ky : x \in H_\infty \quad \text{and} \quad k \in \mathbf{Z}\}.$$

Case 1. $ky \notin H_\infty$ for all integers $k \neq 0$.

Then we define $g^*(x + ky) = g_\infty(x)$ for all $x + ky$ in H^*. To see that this is well defined, observe that

$$x+ky = x_1+\ell y \implies (k-\ell)y = x_1-x \text{ is in } H_\infty \implies k-\ell = 0 \implies x = x_1 \implies g_\infty(x) = g_\infty(x_1).$$

It is now easy to check that g^* is a group homomorphism:

$$g^*(x+ky-(x_1+\ell y)) = g^*(x-x_1+(k-\ell)y) = g_\infty(x-x_1) = g_\infty(x)-g_\infty(x_1) = g^*(x+ky)-g^*(x_1+\ell y).$$

So g^* is in $\mathcal{F}$ and is bigger than g_∞, a contradiction.

Case 2. $ky \in H_\infty$ for some $k \neq 0$.

Since ky is in H_∞ if and only if $-ky$ is, there exists $k > 0$ with $ky \in H_\infty$. Select a minimal $k_0 > 0$ so that $k_0 y$ is in H_∞. Let $\alpha = \frac{1}{k_0} g_\infty(k_0 y)$ and define

$$g^*(x + ky) = g_\infty(x) + k\alpha \quad \text{for all} \quad x + ky \quad \text{in} \quad H^*.$$

We must first check that this definition is well defined. We begin by showing

$$ky \in H_\infty \quad \text{implies} \quad k_0 \quad \text{divides} \quad k. \tag{1}$$

Suppose that $ky \in H_\infty$. We can write k as $ak_0 + r$ where a and r are integers and $0 \le r < k_0$. Then the element $ry = -ak_0 y + ky$ belongs to H_∞, so $r = 0$ by the minimality of k_0. That is, k_0 divides k. Thus (1) holds. To check that g^* is well defined, consider $x + ky$ and $x_1 + \ell y$. Then $(k - \ell)y = x_1 - x$ belongs to H_∞, so by (1) we have $k - \ell = ak_0$ for some $a \in \mathbf{Z}$. Therefore

$$\begin{aligned} g_\infty(x) + k\alpha &= g_\infty(x) + [ak_0 + \ell]\alpha = g_\infty(x) + a g_\infty(k_0 y) + \ell\alpha \\ &= g_\infty(x + ak_0 y) + \ell\alpha = g_\infty(x + (k - \ell)y) + \ell\alpha = g_\infty(x_1) + \ell\alpha. \end{aligned}$$

Thus g^* is well defined. Now easily g^* is a homomorphism:

$$\begin{aligned} g^*(x + ky - (x_1 + \ell y)) &= g_\infty(x - x_1) + (k - \ell)\alpha = g_\infty(x) + k\alpha - g_\infty(x_1) - \ell\alpha \\ &= g^*(x + ky) - g^*(x_1 + \ell y). \end{aligned}$$

Again g^* belongs to $\mathcal{F}$ and is bigger than g_∞, a contradiction. ♣

The Hahn-Banach Theorem in functional analysis is similar to example 4) above, but the details are even more delicate.

5) See the appendix for an elegant proof of Tychonoff's Theorem [in topology] that uses nets.

6) Let X be a vector space over a field $\mathbf{F}$. There exist *Hamel bases*, i.e., maximal independent subsets of X. The classical case is the vector space $\mathbf{R}$ of real numbers over the field $\mathbf{Q}$ of rationals. [A subset E of X is *independent over* $\mathbf{F}$ if, given distinct $e_1, e_2, \ldots, e_n$ in E and given $\alpha_1, \alpha_2, \ldots, \alpha_n$ in $\mathbf{F}$, the equality $\sum_{k=1}^{n} \alpha_k e_k = 0$ implies $\alpha_k = 0$ for $k = 1, 2, \ldots, n$.]

Proof. The proof is easy. Let $\mathcal{P}$ be the set of all nonempty independent subsets of X, ordered by inclusion. Given $x \neq 0$, the one-element set $\{x\}$ belongs to $\mathcal{P}$ since $\alpha x = 0$ and $\alpha \neq 0$ imply $x = \alpha^{-1}\alpha x = 0$. So $\mathcal{P} \neq \emptyset$.

Consider a chain $\mathcal{C}$ in $\mathcal{P}$. Let $E_0 = \cup_{E \in \mathcal{C}} E$. We claim that E_0 is independent over $\mathbf{F}$. So consider distinct $e_1, e_2, \ldots, e_n$ in E and $\alpha_1, \alpha_2, \ldots, \alpha_n$ in $\mathbf{F}$ where $\sum_{k=1}^{n} \alpha_k e_k = 0$. Each e_k belongs to E_k for some E_k in $\mathcal{C}$. Since $\mathcal{C}$ is a chain, one of the sets E_k is biggest, say $E_n \supseteq E_k$ for all $k = 1, 2, \ldots, n$. Then $e_1, e_2, \ldots, e_n$ are in E_n. Since E_n is independent, $\alpha_k = 0$ for $k = 1, 2, \ldots, n$. Thus E_0 is independent.

By Zorn's Lemma, $\mathcal{P}$ has a maximal element, i.e., there is a Hamel basis. ♣

Note. All that is vital in the last proof is that a set E is independent over $\mathbf{F}$ if and only if each of its finite subsets is independent over $\mathbf{F}$. This feature comes up so often that some people, including me, like the following assertion which is equivalent to Zorn's Lemma, etc.

Tukey's Lemma. Consider a set X. Let $\mathcal{F}$ be a nonempty family of subsets of X of *finite character*: E belongs to $\mathcal{F}$ if and only if every finite subset of E belongs to $\mathcal{F}$. Then $\mathcal{F}$ has a maximal member.

Proof from Zorn's Lemma. Just imitate the argument in example 6).

Examples.

6) again. Vector spaces have Hamel bases.

Proof. The family of independent subsets is of finite character. So apply Tukey's Lemma. ♣

7) Every Hilbert space has an orthonormal basis.

Proof. Apply Tukey's Lemma to the family of orthonormal sets. ♣

8) Let $\mathcal{A}$ be a family of subsets of some set X. There exist maximal pairwise disjoint subfamilies of $\mathcal{A}$.

Proof. Tukey's Lemma. ♣

9) Let A be a subset of the plane $\mathbf{R}^2$. There exists a maximal subset of A such that no three points are collinear.

Proof. Tukey. ♣

We have one more assertion that is equivalent to Zorn's Lemma. It is the least intuitive of the bunch.

Well Ordering Principle. Every set can be well ordered, i.e., there is some ordering $\leq$ on the set that is well ordered.

Example. So, if you believe this, the set $\mathbf{R}$ of real numbers can be well ordered with some weird ordering. *No one has ever seen* such an ordering.

In the olden days, the Well Ordering Principle was used in the form:

Transfinite Induction. Let $(W, \leq)$ be a nonempty well ordered set. Suppose that $A \subseteq W$ satisfies

(i) the least element 1 of W is in A;

(ii) whenever x is in W and $\{y \in W : y < x\} \subseteq A$, then x is in A.

Then $A = W$.

Proof. Suppose $A \neq W$. Then $B = W \setminus A$ is nonempty. By (i), $1 \notin B$. So the least element x_0 of B is bigger than 1. Clearly $\{y \in W : y < x_0\} \subseteq A$; so by (ii) x_0 belongs to A. Thus $A \cap B$ is nonempty, a contradiction. ♣

Here is the promised big theorem.

Theorem. The following are equivalent:

(A) Axiom of Choice.

(M) Hausdorff Maximality Principle.

(Z) Zorn's Lemma.

(T) Tukey's Lemma.

(W) Well Ordering Principle.

Proof. We've already shown the implications $\rightarrow$ indicated below.

$$\begin{array}{ccccc} \mathrm{A} & \Longrightarrow & \mathrm{M} & \Longleftarrow & \mathrm{T} \\ \Uparrow & & \downarrow & & \\ \mathrm{W} & \Longleftarrow & \mathrm{Z} & \nearrow & \end{array}$$

It suffices to prove the four implications marked $\Longrightarrow$.

Tukey $\Longrightarrow$ Maximality. The chains in a partially ordered set are sets of finite character.

Zorn $\Longrightarrow$ Well Ordering. Let S be the set we wish to well order. Let $\mathcal{P}$ be the set of all well orderings W of subsets of S. We regard each W as a subset of $S \times S$ and write $\mathrm{dom}(W)$ for $\{x \in S : (x, y) \text{ is in } W \text{ for some } y \in W\}$. We partially order $\mathcal{P}$ by $W_1 \leq W_2$ if and only if $W_1 \subseteq W_2$ *and* $\mathrm{dom}(W_1)$ is an initial segment of $\mathrm{dom}(W_2)$. It is easy to check that this is a partial order. Also $\mathcal{P} \neq \emptyset$, since $\{(x, x)\}$ is in $\mathcal{P}$ for each x in S.

Given a chain $\mathcal{C}$ in $\mathcal{P}$, we show that $W_0 = \cup_{W \in \mathcal{C}} W$ is in $\mathcal{P}$. We easily check that W_0 is a linear order: If x_1, x_2 are in $\mathrm{dom}(W_0)$, then x_1, x_2 are in $\mathrm{dom}(W)$ for some W in $\mathcal{C}$ and hence $(x_1, x_2) \in W \subseteq W_0$ or else $(x_2, x_1) \in W \subseteq W_0$. To see that W_0 is a well ordering, consider a nonempty subset A of $\mathrm{dom}(W_0)$. Then $A \cap \mathrm{dom}(W) \neq \emptyset$ for some W in $\mathcal{C}$. The least element in $A \cap \mathrm{dom}(W)$ will also be the least element of A in W_0. Thus W_0 is in $\mathcal{P}$, and it is easy to check that W_0 is an upper bound to $\mathcal{C}$.

Now by Zorn's Lemma, $\mathcal{P}$ has a maximal element W^*. If $\mathrm{dom}(W^*) = S$, we're done. Otherwise, select x_0 in $S \setminus \mathrm{dom}(W^*)$ and extend the order to $\mathrm{dom}(W^*) \cup \{x_0\}$ by putting x_0 at the top. Since the new well ordered set would be bigger than W^*, we have a contradiction.

Well Ordering $\Longrightarrow$ Axiom of Choice. Consider a family of nonempty sets $\{A_\lambda\}$. Well order the union $\cup_\lambda A_\lambda$. Define $f(\lambda) =$ least member of A_λ for each λ. Then f belongs to $\prod_\lambda A_\lambda$, so this product set is nonempty.

Axiom of Choice $\Longrightarrow$ Maximality Principle. This is the hard implication. Assume that the Axiom of Choice holds, but that the Maximality Principle fails. Then there exists a partially ordered set $(X, \leq)$ with no maximal chain. Let $\mathcal{A}$ be the set of all chains in X. For each C in $\mathcal{A}$, the set $\mathcal{A}_C = \{C' \in \mathcal{A} : C \subset C'\}$ is nonempty. [$C \subset C'$ means that C is a *proper* subset of C'.] By the Axiom of Choice there is a choice function f from $\mathcal{A}$ into $\cup_{C\in\mathcal{A}}\mathcal{A}_C$ so that $f(C)$ is in $\mathcal{A}_C$ for each C in $\mathcal{A}$. In other words, f maps $\mathcal{A}$ into $\mathcal{A}$ and satisfies

$$C \subset f(C) \quad \text{for all} \quad C \quad \text{in} \quad \mathcal{A}. \tag{2}$$

Note that every chain $\mathcal{C}$ in $\mathcal{A}$ has a least upper bound (*lub*) in $\mathcal{A}$, namely $\cup_{C\in\mathcal{C}}C$. The existence of such an $\mathcal{A}$ and f contradicts:

Bourbaki's Fixed Point Theorem. Let X be a nonempty partially ordered set in which every nonempty chain in X has a lub in X. If $f\colon X \to X$ satisfies $f(x) \geq x$ for all $x \in X$, then there exists an x_0 in X with $f(x_0) = x_0$.

Proof. Fix a in X. A subset A of X is *good* if
(a) a is in A,
(b) $f(A) \subseteq A$,
(c) whenever C is a nonempty chain in A, then $\operatorname{lub} C$ is in A.
Clearly X itself is good.

Claim 1. Without loss of generality, X is the only good subset of X, so that if A satisfies (a)-(c) then $A = X$.

Prf. Let M be the intersection of all good subsets of X. Easily M is also good. So replace X by M.

The plan. We will show that X must itself be a chain. Then $x_0 = \operatorname{lub} X$ must belong to X and satisfy $f(x_0) \geq x_0$. Therefore $f(x_0) = x_0$.

Claim 2. a is the least element of X.

Prf. It suffices to show that $A = \{x \in X : x \geq a\}$ is good, since this set would then have to be equal to all of X.

(a) Clearly a is in A.

(b) Given x in A we need to show that $f(x)$ is in A. But $f(x) \geq x \geq a$.

(c) Let C be a nonempty chain in A, and let $w = \operatorname{lub} C$. For any x in C we have $w \geq x \geq a$, so w is in A.

This completes Claim 2.

Now we say that x in X has *property* $\mathrm{P}(x)$ if $y < x$ implies $f(y) \leq x$.

Claim 3. If $\mathrm{P}(x)$ holds, then

$$\text{for each} \quad z \quad \text{in} \quad X \quad \text{either} \quad z \leq x \quad \text{or} \quad z \geq f(x). \tag{3}$$

Prf. It suffices to show that $A = \{z \in X : z \leq x \text{ or } z \geq f(x)\}$ is good.

(a) $a \le x$ by Claim 2, so a is in A.

(b) Given z in A we need $f(z) \in A$. Either $z \le x$ or $z \ge f(x)$. If $z < x$, then $f(z) \le x$ by property P(x). If $z = x$, then obviously $f(z) \ge f(x)$. Finally, if $z \ge f(x)$, then $f(z) \ge z \ge f(x)$. So in all three cases, $f(z)$ belongs to A.

(c) Again consider a nonempty chain C in A and $w = \operatorname{lub} C$. Now either $z \le x$ for all $z \in C$ [in which case $w \le x$] or some z in C satisfies $z \ge f(x)$ [in which case $w \ge f(x)$]. Thus w is in A.

So A is good and $A = X$.

Claim 4. Every x in X satisfies P(x), so X is a chain and we're done.

Prf. X will be a chain because for all x and z either $z \le x$ or $z \ge f(x) \ge x$. So it suffices to show that $A = \{x \in X : \text{P}(x) \text{ holds}\}$ is good.

(a) a is in A because P(a) holds vacuously.

(b) Assume that P(x) holds; we need to show that P($f(x)$) holds. It suffices to show that $y < f(x)$ implies $f(y) \le f(x)$. By Claim 3 for x, either $y \le x$ or $y \ge f(x)$, but the latter cannot hold since $y < f(x)$. So $y \le x$. If $y = x$, then $f(y) \le f(x)$ is obvious, while if $y < x$, then $f(y) \le x$ by property P(x), so $f(y) \le x \le f(x)$.

(c) Again consider a nonempty chain C in A and $w = \operatorname{lub} C$. We need to verify P(w): $y < w$ implies $f(y) \le w$. We suppose that $y < w$ and we first show

Subclaim. There exists x in C so that $y < x$.

Prf. Suppose not. Each x in C is in A and satisfies P(x), so by (3) either $y < x$ [which we've disallowed], $y = x$ or $y \ge f(x) \ge x$. Thus y would be an upper bound for C, a contradiction, since $y < \operatorname{lub} C$. This completes the subclaim.

By the subclaim, we have $y < x$ for some x in C. By property P(x), we have $f(y) \le x \le w$. So (c) holds.

Finally, then, Claim 4 holds, and the proof is completed. ♣

Cardinal Numbers

Let $\mathcal{S}$ be a set of sets. We say $A \sim B$ or that A and B have the *same size* or *cardinality* if there is a one-to-one correspondence mapping A onto B. This gives an equivalence relation. A *cardinal number* [or *cardinal*] will be a symbol attached to each equivalence class. For a set A, card(A) will denote the cardinal number of its equivalence class. Finite sets have the same cardinal number if and only if they are the same size.

Some standard cardinals. $0 = \text{card}(\emptyset)$. For n in $\mathbf{N}$, $n = \text{card}\{1, 2, \ldots, n\}$. $\text{card}(\mathbf{N}) = \aleph_0$. $\text{card}(\mathbf{R}) = \mathfrak{c}$. The use of the different symbols, $\aleph_0$ and $\mathfrak{c}$, will be justified later.

Definition. A set is *countable* if it is finite or has cardinality $\aleph_0$. Otherwise it is *uncountable*.

If $\mathfrak{a}$ and $\mathfrak{b}$ are cardinals, we say that $\mathfrak{a} \le \mathfrak{b}$ if $\text{card}(A) = \mathfrak{a}$ and $\text{card}(B) = \mathfrak{b}$ imply that there is a one-to-one mapping of A into B.

Clearly $0 \le 1 \le 2 \le \cdots \le \aleph_0 \le \mathfrak{c}$.

Theorem 1. Any set of cardinal numbers is linearly ordered. I.e.,

(R) $\mathfrak{a} \le \mathfrak{a}$ for all $\mathfrak{a}$,

(AS) $\mathfrak{a} \le \mathfrak{b}$ and $\mathfrak{b} \le \mathfrak{a}$ imply $\mathfrak{a} = \mathfrak{b}$,

(T) $\mathfrak{a} \le \mathfrak{b}$ and $\mathfrak{b} \le \mathfrak{c}$ imply $\mathfrak{a} \le \mathfrak{c}$,

(L) given $\mathfrak{a}$ and $\mathfrak{b}$, either $\mathfrak{a} \le \mathfrak{b}$ or $\mathfrak{b} \le \mathfrak{a}$.

Proof. (R) is obvious and (T) is simple.

(L) We are given $\mathfrak{a} = \text{card}(A)$ and $\mathfrak{b} = \text{card}(B)$. Let $\mathcal{F}$ be the set of all one-to-one functions f with $\text{dom}(f) \subseteq A$ and $\text{range}(f) \subseteq B$. Viewing functions as sets of ordered pairs, we see that $\mathcal{F}$ is a set of finite character, so by Tukey's Lemma $\mathcal{F}$ has a maximal member h. If $\text{dom}(h) = A$, then $\mathfrak{a} \le \mathfrak{b}$ and we're done. If $\text{range}(h) = B$, then h^{-1} is a one-to-one map of B into A so that $\mathfrak{b} \le \mathfrak{a}$ and again we're done.

Otherwise, we can choose x in $A \setminus \text{dom}(h)$ and y in $B \setminus \text{range}(h)$. Then $h^* = h \cup \{(x, y)\}$ is in $\mathcal{F}$, contradicting the maximality of h.

To prove (AS) we need an easy

Fixed Point Lemma. Let X be a partially ordered set in which every nonempty set has a least upper bound (lub) and greatest lower bound (glb). If $f\colon X \to X$ is order-preserving [$x \le y$ implies $f(x) \le f(y)$], then f has a fixed point.

Proof. Let $0 = \text{glb}\, X$. Then $f(0) \ge 0$. Let $A = \{x \in X : x \le f(x)\}$. Since 0 is in A, A is nonempty. Hence $x_0 = \text{lub}\, A$ exists. For x in A we have $x \le f(x) \le f(x_0)$, so $f(x_0)$ is an upper bound for A. Thus $x_0 \le f(x_0)$. This implies that $f(x_0) \le f(f(x_0))$, so $f(x_0)$ is in A. Hence $f(x_0) \le x_0$, and we conclude that $x_0 = f(x_0)$. ♣

Here is property (AS).

Schröder-Bernstein Theorem. If $\mathfrak{a} \le \mathfrak{b}$ and $\mathfrak{b} \le \mathfrak{a}$, then $\mathfrak{a} = \mathfrak{b}$.

Proof. Let $\mathfrak{a} = \text{card}(A)$ and $\mathfrak{b} = \text{card}(B)$. By hypothesis, there exist one-to-one functions $f\colon A \to B$ and $g\colon B \to A$. We apply the Fixed Point Lemma to the partially ordered set $\mathcal{P}(A)$ of all subsets of A and the function $\phi(E) = g(f(E)^c)^c$. Here the exponent c signifies set complementation. Now ϕ is order-preserving since

$$E \subseteq F \implies f(E) \subseteq f(F) \implies f(E)^c \supseteq f(F)^c \implies g(f(E)^c) \supseteq g(f(F)^c) \implies \phi(E) \subseteq \phi(F).$$

So by the lemma, there is a set $D \subseteq A$ so that $\phi(D) = D$. Then $g(f(D)^c)^c = D$ or $D^c = g(f(D)^c)$. So if we define $h(x) = f(x)$ for $x \in D$ and $h(x) = g^{-1}(x)$ for $x \in D^c$, we obtain a one-to-one function that carries D onto $f(D)$ and D^c onto $f(D)^c$. Thus h is one-to-one and maps A onto B. This shows that $\mathfrak{a} = \mathfrak{b}$. ♣

Definitions. Let $\mathfrak{a}$ and $\mathfrak{b}$ be cardinal numbers, $\text{card}(A) = \mathfrak{a}$ and $\text{card}(B) = \mathfrak{b}$. We define $\mathfrak{a}\mathfrak{b} = \text{card}(A \times B)$, $\mathfrak{a}^{\mathfrak{b}} = \text{card}(A^B)$ and, if $A \cap B = \emptyset$, $\mathfrak{a} + \mathfrak{b} = \text{card}(A \cup B)$. [As usual, A^B is the set of all functions from B into A.]

One can easily check that these definitions are well defined; for example, if $\text{card}(A) = \text{card}(A')$ and $\text{card}(B) = \text{card}(B')$, then $\text{card}(A \times B) = \text{card}(A' \times B')$.

For finite cardinals, these definitions agree with ordinary arithmetic.

There is no largest cardinal. Before we prove this, we observe that for any set A, with cardinality $\mathfrak{a}$,

$$\mathrm{card}(\mathcal{P}(A)) = 2^{\mathfrak{a}}.$$

[For $f\colon A \to \{0,1\}$, let $\phi(f) = \{x \in A : f(x) = 1\}$. then $\phi\colon \{0,1\}^A \to \mathcal{P}(A)$ is one-to-one and onto.]

Theorem 2. $\mathfrak{a} < 2^{\mathfrak{a}}$ for all cardinals.

Proof. Obviously $\mathfrak{a} \le 2^{\mathfrak{a}}$, since $x \to \{x\}$ is a one-to-one mapping of A *into* $\mathcal{P}(A)$.

Assume that $\mathfrak{a} = 2^{\mathfrak{a}}$. Then there is a one-to-one mapping $h\colon A \to \mathcal{P}(A)$ that is *onto*. But consider the set

$$E = \{x \in A : x \notin h(x)\}.$$

E must equal $h(x_0)$ for some x_0 in A. Is x_0 in E? If yes, $x_0 \notin h(x_0) = E$, hence no. If no, $x_0 \in h(x_0) = E$, hence yes. Either way, we have a contradiction. ♣

Theorem 3. $\aleph_0 \le \mathfrak{a}$ for all infinite cardinals $\mathfrak{a}$.

Proof. In other words, every infinite set A contains a countably infinite subset. By induction A contains subsets A_n with exactly n elements. To get a countably infinite subset, first let

$$B_n = A_{2^n} \setminus \bigcup_{k=0}^{n-1} A_{2^k}.$$

The B_n's are disjoint and nonempty, so we can select one element from each B_n to obtain a countably infinite subset of A. ♣

Corollary. Subsets of countable infinite sets are countable.

It is also easy to give a direct proof of the corollary.

Theorem 4. $\aleph_0\aleph_0 = \aleph_0$.

Proof. In other words, $\mathbf{N} \times \mathbf{N} \sim \mathbf{N}$. Here is a one-to-one map of $\mathbf{N} \times \mathbf{N}$ onto $\mathbf{N}$: $h(m,n) = 2^{m-1}(2n-1)$. ♣

Theorem 5. The countable union of countable sets is countable.

Proof. We want to show that if $A_1, A_2, \ldots$ are countable, so is their union $A = \cup_n A_n$. Without loss of generality, we may assume that the A_n's are pairwise disjoint. By making them bigger, if necessary, we may also assume that $\mathrm{card}(A_n) = \aleph_0$ for all n. Thus

$$A_n = \{a_{nk} : k = 1, 2, \ldots\}.$$

Then $h(n,k) = a_{nk}$ defines a one-to-one mapping h from $\mathbf{N} \times \mathbf{N}$ onto A. $\mathbf{N} \times \mathbf{N}$ is countable by Theorem 4, so A is also countable. ♣

Corollary. The sets $\mathbf{Z}$ and $\mathbf{Q}$ are countable.

Proof. $\mathbf{Z} = \mathbf{N} \cup \{0\} \cup \{-n : n \in \mathbf{N}\}$ and $\mathbf{Q} = \{\frac{m}{n} : m \in \mathbf{Z}, n \in \mathbf{N}\}$. ♣

Example. The sets $\mathbf{R}$, $(0,1)$, $(0,1]$, and $[0,1]$ all have cardinality $\mathfrak{c}$.

Proof. This is all clear from the Schröder-Bernstein theorem, provided we can exhibit a one-to-one function from $\mathbf{R}$ to $(0,1)$. To do this, one can use

$$h(x) = \frac{x}{1+|x|} \quad \text{or} \quad h(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}. \quad ♣$$

Theorem 6. $2^{\aleph_0} = \mathfrak{c}$.

Proof. $2^{\aleph_0} = \text{card}(\{0,1\}^{\mathbf{N}})$. For a sequence $\{\epsilon_n\}$ in $\{0,1\}^{\mathbf{N}}$, we define

$$f(\{\epsilon_n\}) = \sum_{n=1}^{\infty} \frac{\epsilon_n}{3^n}.$$

Then f is a one-to-one mapping of $\{0,1\}^{\mathbf{N}}$ into $\mathbf{R}$, so $2^{\aleph_0} \leq \mathfrak{c}$.

On the other hand, each x in $[0,1)$ has a unique binary expansion $.\epsilon_1\epsilon_2\epsilon_3\cdots$ where $\epsilon_n = 0$ or 1 and $\epsilon_n = 0$ for infinitely many n. [So if x has two binary expansions like $\frac{1}{2} = .1000\cdots = .0111\cdots$, the second expansion is not allowed.] This gives a one-to-one map g of $[0,1)$ into $\{0,1\}^{\mathbf{N}}$, so $\mathfrak{c} \leq 2^{\aleph_0}$.

Thus $2^{\aleph_0} \leq \mathfrak{c}$ and $\mathfrak{c} \leq 2^{\aleph_0}$, and we can apply the Schröder-Bernstein theorem. ♣

Theorem 7 [cardinal arithmetic]. For cardinals $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$ we have

(i)	$\mathfrak{a} + (\mathfrak{b} + \mathfrak{c}) = (\mathfrak{a} + \mathfrak{b}) + \mathfrak{c}$	(vii)	$\mathfrak{a}^{\mathfrak{c}}\mathfrak{b}^{\mathfrak{c}} = (\mathfrak{a}\mathfrak{b})^{\mathfrak{c}}$
(ii)	$\mathfrak{a} + \mathfrak{b} = \mathfrak{b} + \mathfrak{a}$	(viii)	$(\mathfrak{a}^{\mathfrak{b}})^{\mathfrak{c}} = \mathfrak{a}^{\mathfrak{b}\mathfrak{c}}$
(iii)	$\mathfrak{a}(\mathfrak{b} + \mathfrak{c}) = \mathfrak{a}\mathfrak{b} + \mathfrak{a}\mathfrak{c}$	(ix)	$\mathfrak{a} \leq \mathfrak{b}$ implies $\mathfrak{a} + \mathfrak{c} \leq \mathfrak{b} + \mathfrak{c}$
(iv)	$\mathfrak{a}(\mathfrak{b}\mathfrak{c}) = (\mathfrak{a}\mathfrak{b})\mathfrak{c}$	(x)	$\mathfrak{a} \leq \mathfrak{b}$ implies $\mathfrak{a}\mathfrak{c} \leq \mathfrak{b}\mathfrak{c}$
(v)	$\mathfrak{a}\mathfrak{b} = \mathfrak{b}\mathfrak{a}$	(xi)	$\mathfrak{a} \leq \mathfrak{b}$ implies $\mathfrak{a}^{\mathfrak{c}} \leq \mathfrak{b}^{\mathfrak{c}}$
(vi)	$\mathfrak{a}^{\mathfrak{b}}\mathfrak{a}^{\mathfrak{c}} = \mathfrak{a}^{\mathfrak{b}+\mathfrak{c}}$	(xii)	$\mathfrak{a} \leq \mathfrak{b}$ implies $\mathfrak{c}^{\mathfrak{a}} \leq \mathfrak{c}^{\mathfrak{b}}$

Proof. Most of these are easy. We outline a few of the proofs. In each case, A, B and C are disjoint sets such that $\text{card}(A) = \mathfrak{a}$, $\text{card}(B) = \mathfrak{b}$ and $\text{card}(C) = \mathfrak{c}$.

(iii) We need a one-to-one correspondence between the sets $A \times (B \cup C)$ and $(A \times B) \cup (A \times C)$. But these sets are equal.

(vi) We need a one-to-one correspondence between $A^B \times A^C$ and $A^{B \cup C}$. Given ϕ in $A^{B \cup C}$, i.e., a function from $B \cup C$ into A, we let $\tau(\phi)$ be the ordered pair $(\phi|_B, \phi|_C)$ in $A^B \times A^C$. Then τ is a one-to-one mapping of $A^{B \cup C}$ onto $A^B \times A^C$. [The notation $\phi|_B$, for example, denotes the restriction of ϕ to B.]

(vii) We need a one-to-one correspondence between $A^C \times B^C$ and $(A \times B)^C$. Given (ϕ, ψ) in $A^C \times B^C$, let $\tau(\phi,\psi)(c) = (\phi(c), \psi(c))$ for all c in C. Then τ is a one-to-one mapping of $A^C \times B^C$ onto $(A \times B)^C$.

(viii) This is the trickiest. We need a one-to-one correspondence between $(A^B)^C$ and $A^{B \times C}$. Given ϕ in $(A^B)^C$, we define $\tau(\phi)$ via $\tau(\phi)(b,c) = \phi(c)(b)$ for all (b,c) in $B \times C$. Note that each $\phi(c)$ is a function with domain B.

Now τ maps *onto* $A^{B \times C}$ because given ψ in $A^{B \times C}$, for each c in C we can define $\phi(c)$ to be that function on B so that $\phi(c)(b) = \psi(b,c)$. Then $\tau(\phi) = \psi$.

To see that τ is *one-to-one*, consider distinct ϕ_1 and ϕ_2 in $(A^B)^C$. Then there exists c_0 in C so that $\phi_1(c_0) \neq \phi_2(c_0)$. So there must exist b_0 in B so that $\phi_1(c_0)(b_0) \neq \phi_2(c_0)(b_0)$. But then $\tau(\phi_1)(b_0, c_0) \neq \tau(\phi_2)(b_0, c_0)$. This shows that $\tau(\phi_1) \neq \tau(\phi_2)$. ♣

Lemma. If $\mathfrak{a}$ is an infinite cardinal and n is finite, then $\mathfrak{a} + n = \mathfrak{a}$.

Proof. Consider A where $\text{card}(A) = \mathfrak{a}$ and $B = \{1, 2, \ldots, n\}$, so that A and B are disjoint. A has a countably infinite subset $C = \{c_1, c_2, c_3, \ldots\}$. We define ϕ from A to $A \cup B$ as follows: $\phi(a) = a$ for $a \notin C$; $\phi(c_k) = k$ for $1 \leq k \leq n$; and $\phi(c_k) = c_{k-n}$ for $k \geq n+1$. Then ϕ is a one-to-one correspondence between A and $A \cup B$. ♣

Theorem 8. $\mathfrak{a} + \mathfrak{a} = \mathfrak{a}$ for infinite cardinals.

In other symbols, $2\mathfrak{a} = \mathfrak{a}$. In fact, $n\mathfrak{a} = \mathfrak{a}$ for $n \in \mathbf{N}$ and infinite cardinals $\mathfrak{a}$. This all follows from the next theorem, since $n\mathfrak{a} \leq \aleph_0\mathfrak{a}$.

Theorem 9. $\aleph_0\mathfrak{a} = \mathfrak{a}$ for infinite cardinals $\mathfrak{a}$.

Proof. Consider a set A where $\text{card}(A) = \mathfrak{a}$ and set $B = A \times \mathbf{N}$. Our task is to show that $A \sim B$. We will use Zorn's Lemma. Let $\mathcal{F}$ consist of all one-to-one functions f where

$$\text{dom}(f) \subseteq A \quad \text{and} \quad \text{range}(f) = \text{dom}(f) \times \mathbf{N}.$$

Since A is infinite, A has a countably infinite subset C by Theorem 3. By Theorem 4, there exists a one-to-one function f from C onto $C \times \mathbf{N}$. Hence $\mathcal{F}$ is nonempty.

Consider a chain $\mathcal{C}$ in $\mathcal{F}$ and let $g = \cup\{f : f \in \mathcal{C}\}$. Then g is certainly one-to-one, and we claim that $\text{range}(g) = \text{dom}(g) \times \mathbf{N}$, so that $g \in \mathcal{F}$. If x is in the domain of g, then x is in the domain of f for some f in $\mathcal{C}$, so

$$g(x) \in \text{dom}(f) \times \mathbf{N} \subseteq \text{dom}(g) \times \mathbf{N}.$$

So $\text{range}(g) \subseteq \text{dom}(g) \times \mathbf{N}$. Given (y, n) in $\text{dom}(g) \times \mathbf{N}$, there is some f in $\mathcal{C}$ so that $y \in \text{dom}(f)$. But then (y, n) is in $\text{dom}(f) \times \mathbf{N}$, so (y, n) is in $\text{range}(f) \subseteq \text{range}(g)$. Thus $\text{dom}(g) \times \mathbf{N} = \text{range}(g)$.

By Zorn's Lemma, $\mathcal{F}$ has a maximal element h. It suffices to show that $\text{card}(\text{dom}(h)) = \mathfrak{a}$. By the lemma, it suffices to show that $A \setminus \text{dom}(h)$ is finite. Otherwise $A \setminus \text{dom}(h)$ has a countably infinite subset D and there exists a one-to-one mapping g from D onto $D \times \mathbf{N}$. Then, if we define $h_1(x) = h(x)$ for $x \in \text{dom}(h)$ and $h_1(x) = g(x)$ for $x \in D$, we obtain a function in $\mathcal{F}$, contradicting the maximality of h. ♣

Corollary to Theorem 8. If $\mathfrak{a}$ is an infinite cardinal and $\mathfrak{b} \leq \mathfrak{a}$, then we have $\mathfrak{a} + \mathfrak{b} = \mathfrak{a}$.

Proof. $\mathfrak{a} \leq \mathfrak{a} + \mathfrak{b} \leq \mathfrak{a} + \mathfrak{a} = \mathfrak{a}$. ♣

Theorem 10. $\mathfrak{a}^2 = \mathfrak{a}\mathfrak{a} = \mathfrak{a}$ for infinite cardinals $\mathfrak{a}$.

Proof. Let A be a set where $\text{card}(A) = \mathfrak{a}$. Let $\mathcal{F}$ consist of all one-to-one functions f where

$$\text{dom}(f) \subseteq A \quad \text{and} \quad \text{range}(f) = (\text{dom}\, f) \times (\text{dom}\, f).$$

Since A contains countably infinite subsets, and since $\aleph_0\aleph_0 = \aleph_0$ by Theorem 9, $\mathcal{F}$ is nonempty. Just as in the proof of Theorem 9, $\mathcal{F}$ has a maximal element h, and it suffices to show that $\text{card}(\text{dom}\, h) = \mathfrak{a}$.

Let $D = \text{dom}(h)$ and assume that $\mathfrak{b} = \text{card}(D) < \mathfrak{a}$. Then by the Corollary to Theorem 8, $\text{card}(A \setminus D) = \mathfrak{a}$; otherwise $\text{card}(A \setminus D) = \mathfrak{d} < \mathfrak{a}$, and we would have

$$\mathfrak{a} = \text{card}(A) = \text{card}(D) + \text{card}(A \setminus D) = \mathfrak{b} + \mathfrak{d} = \max\{\mathfrak{b}, \mathfrak{d}\} < \mathfrak{a}.$$

Hence $A \setminus D$ contains a set E where $\text{card}(E) = \text{card}(D) = \mathfrak{b}$. We know $\mathfrak{b}^2 = \mathfrak{b}$. So

$$\begin{aligned} \text{card}((D \times E) \cup (E \times D) \cup (E \times E)) &= \text{card}(D \times E) + \text{card}(E \times D) + \text{card}(E \times E) \\ &= \mathfrak{b}^2 + \mathfrak{b}^2 + \mathfrak{b}^2 = \mathfrak{b} + \mathfrak{b} + \mathfrak{b} = \mathfrak{b}, \end{aligned}$$

by Theorem 8. Since $\text{card}(E) = \mathfrak{b}$, there is a one-to-one correspondence

$$g\colon E \to (D \times E) \cup (E \times D) \cup (E \times E).$$

In the picture below, g maps E onto the union of the three white squares, while h maps D onto the darkened square. If we define $h_0(x) = h(x)$ for x in D and $h_0(x) = g(x)$ for x in E, we obtain a one-to-one mapping of $D \cup E$ onto $(D \cup E) \times (D \cup E)$. So h_0 belongs to $\mathcal{F}$, contradicting the maximality of h.

Hence $\text{card}(\text{dom}\, h) = \mathfrak{a}$ and so $\mathfrak{a}^2 = \mathfrak{a}$ as claimed. ♣

Corollary If $\mathfrak{a}$ is an infinite cardinal and $0 < \mathfrak{b} \leq \mathfrak{a}$, then $\mathfrak{a}\mathfrak{b} = \mathfrak{a}$.

Proof. $\mathfrak{a} \leq \mathfrak{a}\mathfrak{b} \leq \mathfrak{a}\mathfrak{a} = \mathfrak{a}$. ♣

Applications.

1) If $2 \leq \mathfrak{a} \leq \mathfrak{c} = 2^{\aleph_0}$, then $\mathfrak{a}^{\aleph_0} = \mathfrak{c}$ and $\mathfrak{a}^{\mathfrak{c}} = 2^{\mathfrak{c}}$.

Proof. We have

$$\mathfrak{c} = 2^{\aleph_0} \leq \mathfrak{a}^{\aleph_0} \leq \mathfrak{c}^{\aleph_0} = (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \aleph_0} = 2^{\aleph_0} = \mathfrak{c}$$

and $2^{\mathfrak{c}} \leq \mathfrak{a}^{\mathfrak{c}} \leq \mathfrak{c}^{\mathfrak{c}} = (2^{\aleph_0})^{\mathfrak{c}} = 2^{\aleph_0 \mathfrak{c}} = 2^{\mathfrak{c}}$. ♣

Here is a very easy result that we'll use in the next application.

Theorem 11. If $f: B \to A$ maps B *onto* A, then $\text{card}(A) \leq \text{card}(B)$.

Proof. For each x in A, let $g(x)$ be any element in $f^{-1}(x)$. Then $g: A \to B$ is one-to-one. Note the use of the Axiom of Choice here. ♣

2) Let $\mathcal{F}$ be the family of all finite subsets of a set A with $\mathfrak{a} = \text{card}(A) \geq \aleph_0$. Then $\text{card}(\mathcal{F}) = \mathfrak{a}$.

Proof. Let B be the union of the sets $A^{\{1,2,\ldots,n\}}$. Each $A^{\{1,2,\ldots,n\}}$ has cardinality $\mathfrak{a}$, so $\text{card}(B) = \mathfrak{a}\aleph_0 = \mathfrak{a}$. Now $f \to \text{range}(f)$ maps B onto $\mathcal{F}$, so $\text{card}(\mathcal{F}) \leq \mathfrak{a}$ by Theorem 11. Obviously $\mathfrak{a} \leq \text{card}(\mathcal{F})$, so we are done. ♣

3) Let $\mathcal{C}$ be the set of all countable subsets of a set A with $\mathfrak{a} = \text{card}(A) \geq \aleph_0$. Then $\text{card}(\mathcal{C}) = \mathfrak{a}^{\aleph_0}$. **Notes.** By 1) this equals $\mathfrak{c}$ if $2 \leq \mathfrak{a} \leq \mathfrak{c}$. Also $\mathfrak{a}^{\aleph_0} = \mathfrak{a}$ if $\mathfrak{a}$ has the form $2^{\mathfrak{b}}$.

Proof. $f \to \text{range}(f)$ maps $A^{\mathbf{N}}$ onto $\mathcal{C}$, so by Theorem 11, $\text{card}(\mathcal{C}) \leq \mathfrak{a}^{\aleph_0}$.

Now each f in $A^{\mathbf{N}}$ "is" its graph $G_f \subseteq \mathbf{N} \times A$. So $f \to G_f$ is a one-to-one map of $A^{\mathbf{N}}$ into the family $\mathcal{C}(\mathbf{N} \times A)$ of countable subsets of $\mathbf{N} \times A$. Hence $\mathfrak{a}^{\aleph_0} \leq \text{card}(\mathcal{C}(\mathbf{N} \times A)) = \text{card}(\mathcal{C})$. ♣

Ordinal numbers

Definitions. If A and B are ordered sets, an *order isomorphism* is a one-to-one mapping f of A onto B such that

$$x \leq y \quad \text{in} \quad A \quad \text{if and only if} \quad f(x) \leq f(y) \quad \text{in} \quad B.$$

This is an equivalence relation on sets of ordered sets. We are interested in the well ordered sets! An *ordinal number* [or *ordinal*] will be a symbol attached to an equivalence class of well ordered sets. For a well ordered set W, let's write $\text{ord}(W)$ for this symbol.

Examples. We define

$0 = \text{ord}(\emptyset)$,

$n = \text{ord}\{0 < 1 < \cdots < n-1\}$ for $n \in \mathbf{N}$,

$\omega = \text{ord}(\mathbf{N})$,

Ω to be the (unambiguously defined, as we'll see) smallest uncountable ordinal.

To get more examples, let's do a tiny bit of ordinal arithmetic. Say $\alpha = \text{ord}(A)$ and $\beta = \text{ord}(B)$, where A and B are well ordered sets. Then $\alpha + \beta$ is defined to be the ordinal for the disjoint union $A \cup B$ where A and B have their given orders and everything in A precedes everything in B. For example, $1 + \omega$ is the ordinal of the set obtained by putting one more element at the *beginning* of $\mathbf{N}$; this set is order isomorphic to $\mathbf{N}$, so $1 + \omega = \omega$. On the other hand, $\omega + 1$ is the ordinal of the set obtained by putting one extra element at the *end* of $\mathbf{N}$. This set has a last element, so $\omega + 1 \neq \omega = 1 + \omega$. Addition is not even commutative! Note that

$$\omega + 1 = \text{ord}(\{0 < 1 < 2 < \cdots < \omega\}).$$

Also let $\alpha\beta$ be the ordinal of the set obtained by replacing each b in B by a copy of A. In other words, this is $A \times B$ with the reverse lexicographic ordering:

$$(a, b) < (a', b') \quad \text{if and only if} \quad b < b' \quad \text{or} \quad b = b' \text{ and } a < a'.$$

Multiplication of ordinals isn't commutative either. For example, $\omega 2$ is the ordinal number for the well ordered set obtained by placing one copy of $\mathbf{N}$ above another copy of $\mathbf{N}$. Thus

$$\omega 2 = \omega + \omega = \text{ord}\{0 < 1 < 2 < \cdots < \omega < \omega + 1 < \omega + 2 < \cdots\}.$$

Note that ω, and each element beyond ω, has an infinite number of predecessors. In contrast, 2ω is the ordinal number of the well ordered set obtained by replacing each integer in $\mathbf{N}$ by a copy of the two-element ordered set $\{0 < 1\}$. That is,

$$2\omega = \text{ord}\{0 < 1 < 0' < 1' < 0'' < 1'' < 0''' < 1''' < \cdots\} = \omega.$$

In general, for $n \in \mathbf{N}$, $n\omega = \omega$ but ωn has $n - 1$ limit ordinals, i.e., elements with no immediate predecessors. Note also that

$$\omega^2 = \omega\omega = \text{ord}\left\{m - \frac{1}{n} : n, m \in \mathbf{N}\right\}.$$

This ordered set was briefly discussed on page 2.

Definition. Given a well ordered set W and x in W, the *initial segment determined by* x in W is the set $W_x = \{y \in W : y < x\}$. If $\alpha = \text{ord}(A)$ and $\beta = \text{ord}(B)$, we write $\alpha < \beta$ if A is order isomorphic to some initial segment of B determined by some x in B. $\alpha \leq \beta$ means that $\alpha < \beta$ or $\alpha = \beta$.

It is easy to check that the definition of $<$ is well defined, i.e., is independent of the representatives A and B. Also, $\leq$ is a partial ordering on any set of ordinal numbers. Reflexivity is clear, and transitivity is easy to verify. We will show anti-symmetry in the Corollary to Theorem 12.

Lemma. If W is well ordered and $f: W \to W$ is an order isomorphism of W into W, then $x \leq f(x)$ for all $x \in W$.

Proof. If not, $A = \{x \in W : f(x) < x\}$ is nonempty and has a least element a. Then $f(a) < a$, so $f(f(a)) < f(a)$. I.e., $f(a)$ is in A, contradicting the fact that a is the least element of A. ♣

Corollary. If $\text{ord}(W_1) = \text{ord}(W_2)$ for well ordered sets W_1 and W_2, then there is exactly one order isomorphism of W_1 onto W_2.

Proof. Consider order isomorphisms f and g of W_1 onto W_2. Then $f^{-1} \circ g$ is an order isomorphism of W_1 onto W_1. So $x \leq f^{-1}(g(x))$ for all x in W_1 by the Lemma. Hence $f(x) \leq g(x)$ for all x in W_1. Similarly, $g(x) \leq f(x)$ for all x in W_1. Therefore we have $f = g$. ♣

Theorem 12. Let W be a well ordered set. Then

(a) W is order isomorphic to no initial segment of itself;

(b) if $\text{ord}(W_x) = \text{ord}(W_y)$ for x and y in W, then $x = y$.

Proof. (a) Otherwise there is an order isomorphism $f: W \to W_x$. Since $f(x)$ is in W_x, we must have $f(x) < x$, contradicting the lemma.

(b) Assume that there's an order isomorphism of W_x onto W_y, with $x \neq y$. We may assume that $y < x$. But then W_y is an initial segment of W_x and we have violated part (a). ♣

Corollary. On any set of ordinal numbers, $\leq$ is anti-symmetric.

Proof. Assume that $\alpha < \beta$ and that $\beta < \alpha$. Then there exist order isomorphisms $f: A \to B$ and $g: B \to A$ whose ranges are initial segments. Then $g \circ f$ would be an order isomorphism of A onto an initial segment of A, contradicting Theorem 12(a). ♣

Theorem 13. Any set of ordinal numbers is linearly ordered. That is, given ordinals α and β, exactly one of the following holds: $\alpha < \beta$, $\alpha = \beta$, $\beta < \alpha$.

Proof. The Corollary to Theorem 12 shows that at most one of these holds.

To show that at least one holds, we will use a now familiar Zorn Lemma argument. Let A and B be well ordered sets so that $\text{ord}(A) = \alpha$ and $\text{ord}(B) = \beta$. Let $\mathcal{F}$ be the family of all order isomorphisms f where

$\text{dom}(f)$ is an initial segment of A or is A itself,
$\text{range}(f)$ is an initial segment of B or is B itself.

$\mathcal{F}$ is nonempty because we can map the least element of A onto the least element of B. By Zorn's Lemma, $\mathcal{F}$ has a maximal element h. [You should check that Zorn's Lemma applies.] Either

(a) $\text{dom}(h) = A$ but $\text{range}(h) \neq B$,

(b) $\text{range}(h) = B$ but $\text{dom}(h) \neq A$, or

(c) $\text{dom}(h) = A$ and $\text{range}(h) = B$.

[To see this, assume that $A \setminus \mathrm{dom}(h) \neq \emptyset$ and $B \setminus \mathrm{range}(h) \neq \emptyset$. Let a and b be the least elements of these sets. If we extend h to $\mathrm{dom}(h) \cup \{a\}$ by defining $h^*(a) = b$, we obtain a bigger member h^* of $\mathcal{F}$, a contradiction.] In case (a), $\alpha < \beta$; in case (b), we have $\beta < \alpha$ [using h^{-1}]; and in case (c), $\alpha = \beta$. ♣

More is true: any set of ordinal numbers is well ordered. This follows from

Theorem 14. Let α be an ordinal number, and let P_α be the set of all ordinal numbers preceding α. Then P_α is well ordered and $\mathrm{ord}(P_\alpha) = \alpha$.

Proof. We know that there exists some well ordered set with $\mathrm{ord}(A) = \alpha$. We need to show that A and P_α are order isomorphic. Consider β in P_α. Since $\beta < \alpha$, there exists x in A so that $\mathrm{ord}(A_x) = \beta$. By Theorem 12(b), x is unique. So we can define $\phi\colon P_\alpha \to A$ unambiguously so that $\mathrm{ord}(A_{\phi(\beta)}) = \beta$ for all β in P_α. It is now easy to check that ϕ is an order isomorphism of P_α onto A. Hence $\mathrm{ord}(P_\alpha) = \alpha$. ♣

You might find it worthwhile to look back at the examples of ordinal numbers we've given and observe that most of them were described using sets of predecessors.

Theorem 15. For each cardinal number $\mathfrak{a}$, there exists a least ordinal number $\alpha_\mathfrak{a}$ so that $\mathrm{card}(P_{\alpha_\mathfrak{a}}) = \mathfrak{a}$.

Proof. Let A be any set with $\mathrm{card}(A) = \mathfrak{a}$. By the Well Ordering Principle, A can be well ordered. Let $\alpha = \mathrm{ord}(A)$. Then A and P_α are order isomorphic. If $\mathrm{card}(\{\beta \in P_\alpha : \beta < \gamma\}) < \mathfrak{a}$ for all γ in P_α, then α works. Otherwise, let $\alpha_\mathfrak{a}$ be the least ordinal α_0 in P_α with the property that $\mathrm{card}(\{\beta \in P_\alpha : \beta < \alpha_0\}) = \mathfrak{a}$. ♣

Corollary. Any set of cardinal numbers is well ordered.

Proof. If A is any set of cardinals, the corresponding set $\{\alpha_\mathfrak{a} : \mathfrak{a} \in A\}$ of ordinals has a least element $\alpha_{\mathfrak{a}_0}$. Then $\mathfrak{a}_0$ is the least element of A. ♣

Now let Ω be *the* least uncountable ordinal, i.e., the least ordinal α such that P_α is uncountable. Then $\mathrm{ord}(P_\Omega) = \Omega$ and people often write Ω for P_Ω. The following theorem is now clear from the foregoing theory.

Theorem 16.

(a) P_Ω is well ordered.
(b) P_Ω is uncountable.
(c) For each α in P_Ω, the set $\{\beta \in P_\Omega : \beta \leq \alpha\}$ is countable.
(d) If C is a countable subset of P_Ω, then there is an α in P_Ω so that $\beta \leq \alpha$ for all $\beta \in C$.

Corollary. There are uncountably many distinct ways to well order $\mathbf{N}$.

Proof. For each infinite ordinal α in P_Ω, the set $\{\beta \in P_\Omega : \beta \leq \alpha\}$ is countably infinite, so it has the same cardinality as $\mathbf{N}$. Moreover, different α's in P_Ω correspond to different orderings [i.e., not order isomorphic] of $\mathbf{N}$ in view of Theorem 12. ♣

Applications to topology. The ordinal space P_Ω, with the order topology, is an interesting example in general topology. For any linearly ordered set, the open intervals [plus half-open intervals at the top and bottom] form a basis for the *order topology*. It turns out that such spaces are always regular [Hausdorff] spaces.

Any closed interval $[\alpha, \beta]$ consisting of ordinals is compact. To see this, assume not. Then β belongs to the set $A = \{\gamma \geq \alpha : [\alpha, \gamma] \text{ is not compact}\}$, so A is nonempty. Let γ_0 be the least member of A. We obtain a contradiction by showing that $[\alpha, \gamma_0]$ is, in fact, compact. Given an open cover $\mathcal{W}$ of this set, one of the sets W_0 in the cover contains an interval $(\alpha_0, \gamma_0]$ where $\alpha \leq \alpha_0 < \gamma_0$. Now a finite number of sets in $\mathcal{W}$ covers $[\alpha, \alpha_0]$, since it is compact. Then these sets plus W_0 cover $[\alpha, \gamma_0]$.

The ordinal space $P_\Omega = [0, \Omega)$ is locally compact, non-compact but sequentially compact. Local compactness follows from the previous paragraph. It is not compact, since the sets $\{[0, \alpha) : \alpha < \Omega\}$ form an open cover of $[0, \Omega)$ having no finite subcover. Each of its closed subintervals is metrizable by Urysohn's Metrization Theorem, which asserts that a regular space with a countable base is metrizable. [It's hard to imagine a direct proof that all closed subintervals of $[0, \Omega)$ are metrizable.] To show that $[0, \Omega)$ is sequentially compact, consider any sequence $\{\alpha_n\}$ in $[0, \Omega)$. By Theorem 16(d), there is $\alpha < \Omega$ so that $\alpha_n \leq \alpha$ for all n. Thus $\{\alpha_n\}$ is a sequence in the compact metric space $[0, \alpha]$, so it has a convergence subsequence.

The space $[0, \Omega)$ itself is not metrizable. One reason is that it has the following property:

(P) every continuous function on the space is bounded,

and the only metric spaces with this property are compact. To check property (P), assume that f is an unbounded continuous function on $[0, \Omega)$. For each n in $\mathbf{N}$, there is α_n in $[0, \Omega)$ so that $|f(\alpha_n)| > n$. By Theorem 16(d) again, there is an α in $[0, \Omega)$ so that $\alpha_n \leq \alpha$ for all n. It follows that the [continuous] restriction of f to the compact set $[0, \alpha]$ also is unbounded, an impossibility.

A related space is $[0, \Omega]$. This is the one-point compactification of $[0, \Omega)$, and it also turns out to be its Stone-Čech compactification. The space $[0, \Omega]$ isn't metrizable, since its subspace $[0, \Omega)$ isn't.

Inductive Constructions

Consider the statements

(A) For each n in $\mathbf{N}$ there exists a sequence of objects $O_1, \ldots, O_n$ so that property $P_n(O_1, \ldots, O_n)$ holds.

(B) There exists an infinite sequence of objects $O_1, O_2, \ldots$ so that

$$\text{for each} \quad n \quad \text{in} \quad \mathbf{N}, \quad \text{property} \quad P_n(O_1, \ldots, O_n) \quad \text{holds.}$$

Statements (A) and (B) might look equivalent, but they are not. Confusing them can lead to unfortunate errors.

Trivial, but illustrative, example. Let X be an infinite subset of $\mathbf{R}$, and let $P_n(x_1, \ldots, x_n)$ be the property "$x_1, \ldots, x_n$ are in X and $x_1 < x_2 < \cdots < x_n$."

Claim. (A) holds. I.e., for each n, there exist numbers $x_1, \ldots, x_n$ so that $x_1 < x_2 < \cdots < x_n$.

This is pretty obvious, since we could simply select n numbers from X and note that they can be ordered. But this wouldn't be a very constructive proof.

Proof by induction. Since the claim is trivial for $n = 1$, we assume that the claim is true for n. Thus there exist $x_1, \ldots, x_n$ in X such that $x_1 < x_2 < \cdots < x_n$. Since X is infinite, there exists a number y in $X \setminus \{x_1, \ldots, x_n\}$. If $y > x_n$, let $x_{n+1} = y$ and observe that $P_{n+1}(x_1, \ldots, x_{n+1})$ holds. If $y < x_n$, then there is a smallest k so that $y < x_k$. If we define $x'_i = x_i$ for $i < k$, $x'_k = y$, and $x'_i = x_{i-1}$ for $k < i \leq n + 1$, then clearly $P_{n+1}(x'_1, x'_2, \ldots, x'_{n+1})$ holds. ♣

So (A) holds. But (B) can fail. For example, (B) fails if X is the set of negative integers. The trouble is that in our inductive proof of (A) we often changed earlier terms along the way. So our constructive proof by induction doesn't qualify as an *inductive construction.*

Induction Construction. Assume

(i) There exists an object O_1 so that property $P_1(O_1)$ holds.

(ii) Given objects $O_1, \ldots, O_n$ so that property $P_n(O_1, \ldots, O_n)$ holds, there exists an object O_{n+1} so that property $P_{n+1}(O_1, \ldots, O_{n+1})$ holds.

Then

(iii) There exists an infinite sequence of objects $O_1, O_2, O_3, \ldots$ so that

$$\text{for each} \quad n, \quad \text{property} \quad P_n(O_1, \ldots, O_n) \quad \text{holds.}$$

Remark. This does not follow directly from Peano's axiom that $\mathbf{N}$ is well ordered. If it did, we'd assume not, consider some set $A \subseteq \mathbf{N}$ consisting of integers with some property, and show that $A = \mathbf{N}$. It's hard to imagine A in this situation. Indeed, it is not clear how to formulate "assume not," i.e., how to deny the conclusion (iii) in a useful way. We seem to need Zorn's Lemma.

Proof. Let $\mathcal{F}$ consist of all functions f such that

(1) $\text{dom}(f)$ is a nonempty initial segment of $\mathbf{N}$, i.e., $n \in \text{dom}(f)$ and $m < n$ imply $m \in \text{dom}(f)$,

(2) $\text{range}(f)$ consists of objects,

(3) for each n in $\text{dom}(f)$, property $P_n(f(1), \ldots, f(n))$ holds.

It suffices to show that $\mathcal{F}$ contains a function with domain $\mathbf{N}$.

By hypothesis (i), $\mathcal{F}$ is nonempty. It is easy to check that if $\mathcal{C}$ is a chain in $\mathcal{F}$, then the function $h = \cup_{f \in \mathcal{C}} f$ also is in $\mathcal{F}$. So by Zorn's Lemma, $\mathcal{F}$ has a maximal element h_0.

We claim that $\text{dom}(h_0) = \mathbf{N}$, completing the proof. Otherwise $\mathbf{N} \setminus \text{dom}(h_0)$ has some elements in it. So it has a least element [which I will call N], since $\mathbf{N}$ is well ordered by Peano's axiom. Note that $N > 1$, since 1 must be in $\text{dom}(h_0)$. In view of (1), we have $\text{dom}(h_0) = \{1, 2, \ldots, N-1\}$. Since property $P_{N-1}(h_0(1), \ldots, h_0(N-1))$ holds, hypothesis (ii) shows that there is an object O_N so that $P_N(h_0(1), \ldots, h_0(N-1), O_N)$ holds. Now define h^* on $\{1, 2, \ldots, N\}$ so that $h^*(k) = h_0(k)$ for $k < N$ and $h^*(N) = O_N$. Then h^* belongs to $\mathcal{F}$ and is bigger than [i.e., an extension of] h_0, contradicting the maximality of h_0. ♣

Essentially the same proof yields

Transfinite Inductive Construction. Let W be a well ordered set with a least element 1 and no greatest element. For each $\alpha \in W$ and objects $\{O_\beta : \beta < \alpha\}$ indexed by $\{\beta \in W : \beta < \alpha\}$, let $P_\alpha(\{O_\beta : \beta < \alpha\})$ be a property about the set $\{O_\beta : \beta < \alpha\}$. Assume

(i) There is an object O_1 so that property $P_2(O_1)$ holds.

(ii) Given objects $\{O_\beta : \beta < \alpha\}$ so that property $P_\alpha(\{O_\beta : \beta < \alpha\})$ holds, there exists an object O_α so that property $P_{\alpha+1}(\{O_\beta : \beta < \alpha + 1\})$ holds. [$\alpha + 1$ is the successor α, i.e., the least element bigger than α.]

Then

(iii) There exist objects $\{O_\beta : \beta \in W\}$ indexed by W so that

$$\text{for each} \quad \alpha \quad \text{in} \quad W, \quad \text{property} \quad P_\alpha(\{O_\beta : \beta < \alpha\}) \quad \text{holds.}$$

Note. In imitating the last proof, (3) gets replaced by

"(3) for each α in W, with $\alpha > 1$, property $P_\alpha(\{O_\beta : \beta < \alpha\})$ holds."

Later in the proof, N gets replaced by α_0 in W, so that $\mathrm{dom}(h_0) = \{\beta \in W : \beta < \alpha_0\}$.

Example. The plane has a subset that intersects every straight line in exactly two points. [This is *not* obvious. Try to prove it directly!]

Proof. The set $\mathcal{S}$ of all straight lines has exactly $\mathfrak{c}$ elements, since each line is determined by its slope and y-intercept. We index $\mathcal{S}$ by the ordinal number $\alpha_\mathfrak{c}$ so that $\mathcal{S} = \{L_\beta : \beta < \alpha_\mathfrak{c}\}$ and $\mathrm{card}(\{\beta : \beta < \alpha\}) < \mathfrak{c}$ for $\alpha < \alpha_\mathfrak{c}$. Let A_1 be any 2-element subset of L_1. Given 2-element subsets A_β of L_β for $\beta < \alpha < \alpha_\mathfrak{c}$, we note that the line L_α has infinitely many points that are not in the set $\cup\{L_\beta : \beta < \alpha\}$. [Details: Each L_β can intersect L_α in at most one point, since two straight lines with two common points must be identical. Consequently, the intersection of L_α with $\cup\{L_\beta : \beta < \alpha\}$ has at most $\mathrm{card}(\{\beta : \beta < \alpha\})$ points while L_α has $\mathfrak{c}$ points.] Hence there exists a 2-element subset A_α of L_α that is disjoint from $\cup\{L_\beta : \beta < \alpha\}$. The set $A = \cup\{A_\beta : \beta < \alpha_\mathfrak{c}\}$ is the desired set. ♣

Note. I wrote this proof in typical mathematical style, but I implicitly applied the Transfinite Inductive Construction, since I had to have the transfinite sequence $\{A_\beta : \beta < \alpha_\mathfrak{c}\}$ before I could define A.

Here's the formal set-up. My allowable objects are 2-element subsets of the plane. Property $P_\alpha(\{A_\beta : \beta < \alpha\})$ is

$$\text{"each} \quad A_\beta \subseteq L_\beta \quad \text{and} \quad A_\beta \quad \text{is disjoint from} \quad \cup\{L_\gamma : \gamma < \beta\}."$$

Remark. The example remains valid if we replace $\mathcal{S}$ by all circles, or all ellipses, etc. Also 2 can be replaced by any finite number or by $\aleph_0$.

For measure theory fans. Using Fubini's theorem, one can show that if the set A above is Lebesgue measurable in $\mathbf{R}^2$, then A must have measure 0.

Can A be forced to have measure 0? Yes. Here $\mathcal{S}$ is all straight lines. Let C be the Cantor set in $[0, 1]$ with measure 0. Let $C + \mathbf{N} = \cup\{C + n : n \in \mathbf{N}\}$, which also has measure 0. Let U be the union of all circles in $\mathbf{R}^2$ centered at 0 whose radii belong to the set $C + \mathbf{N}$. Then the set A can be selected inside U because every straight line intersects U in a set with $\mathfrak{c}$ elements. Also, U has measure 0 in $\mathbf{R}^2$, as can be seen by integrating its characteristic function using polar coordinates. So A also will have measure 0 in $\mathbf{R}^2$.

Can A be selected to be a Borel set? I don't know, and (at least until recently) this is an open question.

Conclusion

The approach in these notes has been completely pragmatic and non-philosophical, avoiding questions about "truth" or the wisdom of using the Axiom of Choice and its equivalents. An excellent article concerning the history of these axioms and their current status is by Solomon Feferman, *Does mathematics need new axioms?*, American Mathematical Monthly **106** (1999), 99-111. Another place to start learning about the subtleties involved is: Chapter 13 of Stan Wagon's book *The Banach-Tarski Paradox*, Cambridge University Press, 1985.

Appendix

The following proof is due to Paul Chernoff (1992).

Tychonoff's Theorem. Let $X = \prod_{i\in I} X_i$, where each X_i is a compact topological space. Then X is compact in the product topology.

Proof. It suffices to show that every net $\langle f_\alpha\rangle_{\alpha\in A}$ in X has a cluster point in X.

For each nonempty set $J \subseteq I$ and g in $\prod_{i\in J} X_i$, we say g is a *partial cluster point* of $\langle f_\alpha\rangle_{\alpha\in A}$ if g is a cluster point of the net $\langle f_\alpha|_J\rangle_{\alpha\in A}$ of restrictions to J.

Let $\mathcal{F}$ be the set of all partial cluster points of $\langle f_\alpha\rangle_{\alpha\in A}$. $\mathcal{F}$ is nonempty because if $J = \{i_0\}$ then $\langle f_\alpha|_J\rangle_{\alpha\in A}$ is a net in X_{i_0}, so it has a cluster point since X_{i_0} is compact.

We partially order $\mathcal{F}$ by extension:

$$g_1 \le g_2 \quad \text{if} \quad \operatorname{dom}(g_1) \subseteq \operatorname{dom}(g_2) \quad \text{and} \quad g_1(i) = g_2(i) \quad \text{for} \quad i \in \operatorname{dom}(g_1).$$

Consider a chain $\mathcal{C}$ in $\mathcal{F}$. By the proposition on page 3, $g_0 = \cup_{g\in\mathcal{C}} g$ is a function with domain $J = \cup_{g\in\mathcal{C}} \operatorname{dom}(g)$.

Claim. g_0 is in $\mathcal{F}$, i.e., g_0 is a partial cluster point of $\langle f_\alpha\rangle_{\alpha\in A}$.

Prf. We need to show that g_0 is a cluster point of the net $\langle f_\alpha|_J\rangle_{\alpha\in A}$. Consider a basic neighborhood of g_0 in $\prod_{i\in J} X_i$ having the form

$$W = \{h \in \prod_{i\in J} X_i : h(i) \in U_i \quad \text{for} \quad i \in F\},$$

where F is a finite subset of J and U_i is open in X_i for each $i \in F$. Since $\mathcal{C}$ is a chain, there exists g in $\mathcal{C}$ so that $F \subseteq \operatorname{dom}(g)$. Consider $\alpha \in A$. Since g is a cluster point of $\langle f_\alpha|_{\operatorname{dom}(g)}\rangle_{\alpha\in A}$, there exists β in A so that $\beta \succeq \alpha$ and $f_\beta(i) \in U_i$ for $i \in F$. Therefore $f_\beta|_J$ belongs to W. Since α in A is arbitrary and the basic neighborhood W of g_0 is arbitrary, this shows that g_0 is a cluster point of $\langle f_\alpha|_J\rangle_{\alpha\in A}$. Hence g_0 is in $\mathcal{F}$, which verifies the claim.

Since g_0 is clearly an upper bound for $\mathcal{C}$, we see that every chain in $\mathcal{F}$ has an upper bound in $\mathcal{F}$. So by Zorn's Lemma, $\mathcal{F}$ has a maximal element g^*. If $\operatorname{dom}(g^*) = I$, then $\langle f_\alpha\rangle_{\alpha\in A}$ has a cluster point in X and we're done.

So assume that $\operatorname{dom}(g^*) = J^* \neq I$. Select k in $I \setminus J^*$. Since g^* is in $\mathcal{F}$, a subnet $\langle f_{\alpha_\beta}|_{J^*}\rangle_{\beta\in B}$ converges to g^*. Since X_k is compact, the net $\langle f_{\alpha_\beta}(k)|_{J^*}\rangle_{\beta\in B}$ in X_k must have a cluster point p. Define h on $J^* \cup \{k\}$ so that $h = g^*$ on J^* and $h(k) = p$. Then h is a partial cluster point of $\langle f_\alpha\rangle_{\alpha\in A}$. Since h is bigger than g^* in the order of $\mathcal{F}$, and since h is in $\mathcal{F}$, this contradicts the maximality of g^*. ♣

INDEX