Chapter 1: Posets and Zorn's Lemma

1.1 Definition

By a partial order on a set $A$, we mean a binary relation $\leq$ on $A$ which is:

- reflexive: $(\forall x \in A)(x \leq x)$
- transitive: $(\forall x, y, z \in A)(x \leq y \text{ and } y \leq z \Rightarrow x \leq z)$
- antisymmetry: $(\forall x, y \in A)(x \leq y \text{ and } y \leq x \Rightarrow x = y)$

We say that $\leq$ is a total order if it also satisfies

$(\forall x, y \in A)(x \leq y \text{ or } y \leq x)$

By a poset, we mean a partially ordered set. Similarly, a total is a totally ordered set.

1.2 Examples

a) The usual ordering on $\mathbb{R}$, $\leq$ defined by $x \leq y \iff x - y$ is a square, is a total ordering.

b) For any set $A$, the set $\mathcal{P}(A)$ of all subsets of $A$ is partially ordered by $\subseteq$.

c) For any group, $G$, the set $\text{Sub}(G)$ of all subgroups of $G$ is partially ordered by $\subseteq$, and similarly for a vector space $V$.

d) The set $\mathbb{N}_0$ of natural numbers (including 0) is partially ordered by $|$ where $m | n \iff \frac{n}{m} \in \mathbb{N}_0$. 
(Aside: If \( \leq \) is reflexive and transitive on a set \( A \), then the relation \( \sim \) defined by \( x \sim y \iff (x \leq y \text{ and } y \leq x) \) is an equivalence relation on \( A \), and \( \leq \) induces a partial ordering on the quotient set \( A/\sim \).

e) Suppose that given a set \( S \) of 'letters', a word in \( S \) is a finite (possibly empty) sequence of members of \( S \). The set \( S^* \) of all words over \( S \) has several partial orderings:
- the prefix ordering by \( r \leq w \iff w = ru \) for some \( u \in S^* \)
- the infix ordering by \( r \leq w \iff w = urx \) for some \( u, x \in S^* \)

f) Given sets \( A, B \), by a partial function \( f : A \rightarrow B \) we mean a function defined on a subset of \( A \) and taking values in \( B \).

We write \( [A \rightarrow B] \) for the set of all partial functions \( A \) to \( B \).

This set is partially ordered by \( f \leq g \iff g \text{ extends } f \), i.e. \( g(x) \) is defined and equals \( f(x) \) wherever \( f(x) \) is defined.

1.3 Definition
Let \( (A, \leq) \) be a poset. We say that \( b \) covers \( a \) in \( A \), and write \( a \downarrow \downarrow b \) if \( a \leq b \), \( a \neq b \) and \((\forall c)((a \leq c \text{ and } c \leq b) \Rightarrow (a = c \text{ or } c = b)\).

If \( A \) is finite, then \( a \leq b \) holds \( \iff \) either \( a = b \) or there exists a finite chain \( a \downarrow \downarrow c_1 \downarrow \downarrow c_2 \downarrow \cdots \downarrow \downarrow c_n \downarrow \downarrow b \).

The Hasse diagram of a finite poset represents the elements of \( P \) by points, with an upward line from \( x \) to \( y \) whenever \( x \downarrow \downarrow y \).
1.4 Examples

a) \( P\{x, y, z\} \) has 8 elements, and its Hasse diagram is

\[
\begin{array}{c}
\{x, y, z\} \\
\{x, y\} \quad \{x, z\} \quad \{y, z\} \\
\{x\} \quad \{y\} \quad \{z\} \\
\emptyset
\end{array}
\]

b) Let \( G \) be the non-cyclic group of order 4. The Hasse diagram of \( \text{Sub}(G) \) is

\[
\begin{array}{c}
G \\
H_1 \quad H_2 \quad H_3 \\
\{e\}
\end{array}
\]

1.5 Definition

Let \((P, \leq)\) be a poset, \( S \subseteq P \).

a) By a greatest element of \( S \), we mean an element \( s \in S \) such that \( t \leq s \) for all \( t \in S \).

b) By an upper bound for \( S \), we mean an element \( p \in P \) such that \( s \leq p \) for all \( s \in S \).

c) By a least upper bound for \( S \), we mean a least element of \[ \{ p \in P \mid p \text{ is an upper bound for } S \} \]
p is the greatest element of $S \Rightarrow (p)$ is a least upper bound for $S$ and $p \in S$.

We say that $(P, \leq)$ is complete if every $S \subseteq P$ has a least upper bound.

1.6 Examples

a) $PA$ is complete. If $\{B_i : i \in I\}$ is a subset of $PA$, then $\bigcup_{i \in I} B_i$ is a least upper bound for it.

b) $(R, \leq)$ is not complete, since it lacks greatest and least element.

But $R \cup \{\pm \infty\}$ is complete.

c) Sub $(G)$ is complete. Given a set $\{H_i : i \in I\}$ of subgroups of $G$, $\bigcup_{i \in I} H_i$ needn't be a subgroup, but there is a smallest subgroup $\langle \bigcup_{i \in I} H_i \rangle$ containing it.

1.7 Lemma

Suppose that $(P, \leq)$ is a complete poset. Then, so is $(P, \geq)$.

Proof

Given $S \subseteq P$, we have to show that $S$ has a greatest lower bound.

Let $T = \{p \in P : p$ is a lower bound for $S\}$, and let $t = \bigvee T$, the least upper bound of $T$. For any $s \in S$, we have $p \leq s \forall p \in T$, so $s$ is an upper bound for $T$. Hence $t \leq s$.

So $t$ is a lower bound for $S$, i.e. $t \in T$, and so $t$ is the greatest element of $T$. 
We denote the least upper bound of a set $S$, if it exists, by $\text{US}$, and call it the join of $S$, and the greatest lower bound $\text{MS}$, the meet of $S$.

1.8 Definition

a) By a chain in a poset $(P, \leq)$, we mean a non-empty subset $C \subseteq P$ which is totally ordered by the restriction of $\leq$.

b) We say $(P, \leq)$ is chain complete if every chain $C \subseteq P$ has a join in $P$.

1.9 Examples

a) Let $G$ be a group, and let $P$ be the set of abelian subgroups of $G$ ordered by inclusion. If $H, K \in P$ contain elements $h, k$ such that $hk \neq kh$, then $\{H, k\}$ has no upper bound in $P$, so $P$ is not complete.

But $P$ is chain complete: If $\{H_l : l \in I\}$ is a chain in $P$, then $\bigcup_{l \in I} H_l$ is a subgroup, since if $H, K \in \bigcup_{l \in I} H_l$, then $h \in H_l$ and $k \in H_l$ for some $i, j$, but then $hk \in H_i \cap H_j$, and hence $hk \in \bigcup_{l \in I} H_l$. Similarly, if $H \in \bigcup_{l \in I} H_l$, then $HK = KH$. So $\bigcup_{l \in I} H_l$ is an abelian subgroup, and hence a least upper bound for $\{H_l : l \in I\}$ in $P$.

b) Let $A, B$ be two sets, and consider the poset $[A \to B]$ of partial functions $A \to B$, ordered by extension. This is
not complete, since if \( f, g \) are both defined at \( x \in A \) and \( f(x) \neq g(x) \), then \( \{f, g\} \) has no upper bound.

But if \( \{f_i : i \in I\} \) is a chain in \( [A \rightarrow B] \) then we have \( f_i(x) = f_j(x) \) wherever \( f_i, f_j \) are both defined at \( x \), so there is a unique \( f \) with domain \( \bigcap_{i \in I} \text{dom } f_i \) satisfying

\[
    f(x) = f_i(x) \quad \forall x \in \text{dom } f_i, \quad \forall i.
\]

So \( f = \bigvee \{f_i : i \in I\} \).

**Recursive Definitions**

We define the factorial function as the function

\[
f : \mathbb{N} \rightarrow \mathbb{N} \text{ satisfying } f(n) = \begin{cases} 1 & \text{if } n = 0, \\ n f(n-1) & \text{if } n > 0. \end{cases}
\]

How do we know that this defines anything? Think instead of

a function \( \Phi : [\mathbb{N} \rightarrow \mathbb{N}] \rightarrow [\mathbb{N} \rightarrow \mathbb{N}] \) defined by

\[
    \Phi(f)(n) = \begin{cases} 1 & \text{if } n = 0, \\ f(n-1) & \text{if } n > 0 \text{ and } f(n-1) \text{ is defined undefined otherwise}. \end{cases}
\]

Then we define the factorial function to be the unique \( f \) such that \( \Phi(f) = f \), assuming that this exists.

Note that \( \Phi \) is order preserving: i.e. \( f \leq g \Rightarrow \Phi(f) \leq \Phi(g) \)

**1.10 Theorem (Kuratowski-Tarski)**

Let \( (P, \leq) \) be a complete poset and \( f : P \rightarrow P \) an order preserving map. Then \( f \) has a fixed point.

**Proof**

Consider the set \( S = \{ x \in P \mid x \leq f(x) \} \) of pre-fixed points.

Let \( \Phi y = VS \).
Thus, $\forall x \in S$, we have $x \leq y$. Hence $x \leq f(x) \leq f(y)$. So $f(y)$ is an upper bound for $S$. Hence $y \leq f(y)$, i.e., $y \in S$. Since $f$ is order-preserving, we also have $f(y) \leq f(f(y))$, so $f(f(y)) \in S$. Hence $f(y) \leq y$, and $f(y) = y$ by antisymmetry.

**Corollary (Cantor-Bernstein Theorem)**

Suppose given sets $A$, $B$, and injections $f : A \to B$, $g : B \to A$. Then there exists a bijection $A \to B$.

**Proof**

Consider the composite $PA \xrightarrow{f^{-1}} PB \xrightarrow{g^{-1}} PA \xrightarrow{PA \to PA} PA$.

This is order-preserving since $f^{-1}$ and $g^{-1}$ are order-preserving and $B \setminus A$, $A \setminus B$ are order reversing. So it has a fixed point $A_0$. Let $B_0 = f(A_0)$, $B_1 = B \setminus B_0$, $A_1 = g[B_1]$, then $A_0 = A_1$.

![Venn Diagram](image)

I.e.,

where $f$ maps $A_0$ bijectively to $B_0$, and $g$ maps $B_1$ bijectively to $A_1$. So we define $h : A \to B$ by $h(x) = f(x)$ if $x \in A_0$, and $h(x) = g^{-1}(x)$ if $x \in A_1$.

Given a poset $P$, we say $f : P \to P$ is inflationary if $x \leq f(x)$ for all $x \in P$. 

1:12 Theorem (Bourbaki - Witt) Let \( P \) be a chain complete poset and \( f: P \to P \) an inflationary map. Then for any \( x \in P \), there exists \( y \in P \) with \( x \leq y = f(y) \).

N. Bourbaki, Sur la théorème de Zorn, Arch Math 2(1950), 434-436
E. Witt, Beweisstudien zum Satz von M. Zorn, Math Nachr 43(1961), 434-436
E. Witt, Sobre el teorema de Zorn, Rev. Mat. Hispánica 82-85

Non-proof

Set \( x_0 = x \). If it isn't a fixed point, set \( x_1 = f(x_0) \).
If \( x_1 \) is not, set \( x_2 = f(x_1) \).
If \( x_0 < x_1 < x_2 < \ldots \), set \( x_\infty = \bigvee \{ x_n \mid n \in \mathbb{N} \} \)
If \( x_\infty \) isn't a fixed point, \( x_\infty + 1 = f(x_\infty) \).
\( x_\infty + 2 = f(x_\infty + 1) \).

For heaven's sake, this must stop!
1.12 Theorem (Bourbaki-Witt)

Let \( P \) be a chain-complete poset and \( f : P \to P \) an inflationary map. Then, for every \( x \in P \), there exists \( y \in P \) with \( x \leq y = f(x) \).

Proof:

Define a subset \( C \subseteq P \) to be closed if

a) For all \( y \in C \), \( f(y) \in C \) and

b) For all chains \( S \subseteq C \), \( \forall \in S \in C \)

Note that arbitrary intersections of closed sets are closed.

Hence \( C(x) = \bigwedge \{ D \subseteq P \mid D \text{ closed}, x \in D \} \) is the smallest closed set containing \( x \).

(Idea: \( C(x) \) is the set of all \( x \)'s that we constructed in the non-proof.)

Suppose we can show that \( C(x) \) is a chain. Then \( y = \bigwedge C(x) \) is a member of \( C(x) \) since \( C(x) \) is closed, and hence we also have \( f(y) \in C(x) \). So \( f(y) \leq y \); but we have \( y \leq f(y) \) since \( f \) is inflationary. So \( y = f(y) \) (and \( y \leq x \) since \( x \in C(x) \)).

To show that \( C(x) \) is totally ordered:

Step 1:

\( x \) is the least element of \( C(x) \), since \( \uparrow(x) = \{ y \in P \mid x \leq y \} \) is closed and \( x \in \uparrow(x) \), and \( C(x) \subseteq \uparrow(x) \).
Now call an element $y$ of $C(x)$ normal if
\[(\forall z \in C(x))(z < y \Rightarrow f(z) \leq y).\]

**Step 2**
If $y$ is normal, then for all $z \in C(x)$, we have either $z \leq y$ or $f(y) \geq z$.

To prove this, consider the set $T_y = \{ z \in C(x) | z \leq y \text{ or } f(y) \geq z \}$. We have $x \in T_y$ since $x \leq y$ by Step 1. Suppose $z \in T_y$: then either $z < y$, in which case $f(z) \leq y$ by normality. Or, $z = y$, in which case $f(z) = f(y)$. Or, $z > f(y)$, in which case $f(z) \leq f(y)$ since $f$ is inflationary. So in every case $f(z) \in T_y$. Now suppose $S \subseteq T_y$ is a chain; if all $z \in S$ satisfy $z \leq y$, then $\forall s \in S \Rightarrow f(s) \leq f(y)$, so $\forall s \in S \Rightarrow s \leq f(y)$, so $\forall s \in S \Rightarrow s \in T_y$. So $T_y$ is a closed set containing $x$, and hence $T_y = C(x)$.

**Step 3**
Every $y \in C(x)$ is normal.

Consider the set $N = \{ y \in C(x) | y \text{ is normal} \}$. Since the hypothesis $z \leq x$ is never satisfied, by Step 1.

Suppose $y \in N$, and $z < f(y)$. Then by Step 2, we have $z \leq y$. So either $z < y$, in which case $f(z) \leq y \leq f(y)$, by normality of $y$, or $z = y$, so $f(z) = f(y) \leq f(y)$. Let $S \subseteq N$ be a chain, and suppose $z < \forall s \in S$ (in $N$ want).
Then \( z \neq y \) for some \( y \in S \), so \( z \neq f(y) \), so \( z \leq y \) by Step 2, so \( z < y \). So \( f(z) \leq y \leq VS \) by normality of \( S \). So \( VS \in N \). 

Hence by Steps 2 and 3, \( \mathbb{C}(x) \) is a chain, so \( VC(x) \) is a fixed point of \( f \) above \( x \).

**1.13 Corollary**

Let \( P \) be a chain-complete poset, and \( f : P \rightarrow P \) an order-preserving map. Then, for any \( x \in P \), with \( x \leq f(x) \), there's a least \( y \geq x \) with \( y = f(y) \). In particular, if \( P \) has a least element \( 0 \), then \( f \) has a least fixed point.

**Proof**

Let \( P' = \{ x \in P \mid x \leq f(x) \} \). Then \( x \in P' \Rightarrow f(x) \in P' \) since \( f \) is order-preserving. And if \( S \subseteq P' \) has a join \( VS \) in \( P \), then for all \( x \in S \) we have \( x \leq VS \), and so \( x \leq VS \) since \( f(VS) \) is an upper bound for \( S \) and hence \( VS \leq f(VS) \).

So by 1.12 every \( x \in P' \) lies below a fixed point \( y = VC(x) \).

If \( z \) is any fixed point with \( z \geq x \), then \( V(z) = \{ w \in \mathbb{C}(x) \mid w \leq z \} \) is a closed set containing \( x \), so \( VC(x) \leq V(z) \), so \( VC(x) \leq z \).

Recall the recursive definition of \( n! \):

\[
\varphi(f)(n) = 1 \quad \text{if} \quad n = 0, \quad \varphi(f)(n) = n f(n-1) \quad \text{if} \quad n > 0, \quad f(n-1)
\]
\[ \Phi \text{ is an order preserving map } [\mathbb{N} \to \mathbb{N}] \to [\mathbb{N} \to \mathbb{N}] \]

\[ [\mathbb{N} \to \mathbb{N}] \text{ is chain complete and has a least element, so there is a least } f \text{ with } \Phi(f) = f. \text{ But, for this } f, \text{ we have } 0 \in \text{ dom } f \text{ and } (\forall n)(n \in \text{ dom } f \implies n+1 \in \text{ dom } f). \text{ So } f \text{ is a total function and hence a maximal element of } [\mathbb{N} \to \mathbb{N}], \text{ i.e. } f \preceq g \implies f = g. \]

Hence \( f \) is the unique fixed point of \( \Phi \).

For the next result, we need the Axiom of Choice which asserts the if \([A_i : i \in I]\) is any set of non-empty sets, then there exists a choice function \( f : I \to \bigcup_{i \in I} A_i \), i.e. a function such that \( f(i) \in A_i \), for all \( i \in I \).

1.4 Corollary (Zorn's Lemma)

Assume the Axiom of Choice. Let \( P \) be a chain complete poset.

Then, for any \( x \in P \), there exists a maximal element \( y \) of \( P \) with \( x \preceq y \).

Proof:

For \( x \in P \), let \( A_x = \{ y \in P : y \succ x \} \) if \( x \) is not maximal

\[ = \{ x \} \text{ if } x \text{ is maximal} \]

Then \( A_x \) is non-empty for all \( x \), so we have a choice function \( f : P \to P \) with \( f(x) \in A_x \) \( \forall x \). Thus \( f \) is necessarily inflationary, and its fixed points are exactly the maximal elements of \( P \).

So the result follows from 1.12.
1) Move Examples

a) We show that Zorn's Lemma implies the Axiom of Choice.

Given a family \( \{ A_i : i \in I \} \) of non-empty sets, let \( P \) be the set of partial choice functions, i.e., partial functions \( f : I \to \bigcup_{i \in I} A_i \) satisfying \( f(i) \in A_i \) whenever \( f(i) \) is defined.

Order \( P \) by \( f \preceq g \iff g \text{ extends } f \). Then \( P \) is chain-complete being closed under joins of chains in \( \bigcup_{i \in I} A_i \).

\( P \neq \emptyset \) since the everywhere undefined function is in \( P \), so it has a maximal element, \( f \). Say. Suppose \( f \) is not total. Pick \( i_0 \in I \setminus \text{dom } f \), and pick \( x_0 \in A_{i_0} \). Now define \( g(i) = f(i) \) if \( i \in \text{dom } f \), \( x_0 \) if \( i = i_0 \), undefined otherwise.

Then \( g \in P \) and \( f \preceq g \).

b) (Hamel's Theorem) Every vector space has a basis. Let \( V \) be a vector space (over some field \( F \)). Consider the poset \( P \) of linearly independent subsets \( S \subseteq V \), ordered by inclusion. \( P \) is chain complete.

If \( \{ S_i : i \in I \} \) is a chain of linearly independent sets, then \( \bigcup_{i \in I} S_i \) is linearly independent, since if we had a linear relation \( \sum_{i=1}^{n} A_i x_i = 0 \) with \( x_i \in \bigcup_{i \in I} S_i \) for all \( j \), and then \( j \in I \) such that \( x_j \in S_j \) for all \( j \), and hence we must have \( A_j = 0 \) for all \( j \).

Now, let \( S \) be a maximal element of \( P \).
Suppose that $S$ is not a basis. Then we can pick $v \in V \setminus \langle S \rangle$ and consider $T = S \cup \{v\}$. Then, $T$ is linearly independent and $S \subset T$.

(c) (Maximal Ideal Theorem) Let $R$ be a ring (with 1). Then, any proper ideal of $R$ is contained in a maximal (proper) ideal.

Let $P$ be the set of proper ideals of $R$, ordered by inclusion.

We must show that $P$ is chain complete. Let $\{I_j : j \in J\}$ be a chain of proper ideals. Then $\bigcup_{j \in J} I_j$ is an ideal, since it is an additive subgroup (cf 1.9(a)). But an ideal $I$ is proper

$\implies 1 \notin I$. So $1 \notin \bigcup_{j \in J} I_j$ and hence $\bigcup_{j \in J} I_j$ is a proper ideal.

1.16 Definition

(a) A lattice is a poset $L$ which has joins and meets for all finite subsets. In particular, $L$ has a least element $0 = \bigcap \emptyset$ and a greatest element $1 = \bigvee \emptyset$. Also, for any $\{x, y\} \subseteq L$, we have $x \vee y = \bigvee \{x, y\}$ and $x \wedge y = \bigwedge \{x, y\}$.

These suffice, since we can construct $\bigvee \{x_1, \ldots, x_n\}$ as $\left(\cdots (x_1 \vee x_2) \vee x_3) \vee \cdots \right) \vee x_n$.

(Exercise: check that $\vee$ and $\wedge$ are commutative and associative.)

A mapping $f : L \to M$ between lattices is called a lattice homomorphism if $f(0) = f(0)$, $f(1) = f(1)$, and $f(x \vee y) = f(x) \vee f(y)$, $f(x \wedge y) = f(x) \wedge f(y)$.

A lattice homomorphism is order-preserving since $x \leq y$.

$\iff x \vee y = y \iff x \wedge y = x$. 
The course is false: If \( G \) is a group, the inclusion \( \text{Sub}(G) \to PG \) does not preserve \( \lor \) or \( \land \).

b) We say that a lattice \( L \) is distributive if the identity
\[ x \land (y \lor z) = (x \land y) \lor (x \land z) \]
holds for all \( x, y, z \in L \).

c) We say that \( y \) is a complement for \( x \in L \) if \( x \land y = 0 \) and \( x \lor y = 1 \). By a Boolean Algebra, we mean a distributive lattice in which every element has a complement.

1.17 Examples

a) For any \( A, PA \) is a Boolean Algebra. It is distributive, since
\[ x \leq y \leq z \implies x \land (y \lor z) = x \land (y \lor z) = x \land z = x \]
and \( (x \land y) \lor (x \land z) = x \lor x = x \).

b) Let \( L \) be a totally ordered set with greatest and least elements. Then \( L \) is a lattice, with
\[ x \lor y = \max \{x, y\} \quad \text{and} \quad x \land y = \min \{x, y\}. \]

\( L \) is distributive: e.g., if \( x \leq y \leq z \), then
\[ x \land (y \lor z) = x \land z = x \]
and \( (x \land y) \lor (x \land z) = x \lor x = x \).

Or, if \( y \leq z \leq x \), then
\[ x \land (y \lor z) = x \land z = z \]
and \( (x \land y) \lor (x \land z) = y \lor z = z \).

(Exercise: Check the other cases.)
But if $L$ has more than 2 elements, it is not Boolean.

e) If $G$ is the non-cyclic group of order 4, then $\text{Sub}(G)$ is not distributive. 
\[
x \land (y \lor z) = x \land z = x
\]
but $(x \land y) \lor (x \land z) = \{e\} \lor \{e\} = \{e\} 

Note that here, $x$ has two distinct complements, $y$ and $z$.

1.18 Lemma

i) If $L$ is distributive, so is $L^{op}$

ii) In a distributive lattice, every element has at most one complement.

Proof:

i) We have to show $x \lor (y \land z) = (x \lor y) \land (x \lor z)$. But 
\[
(x \lor y) \land (x \lor z) = (x \lor y) \land x \lor (x \lor y) \land z
\]
\[
= x \lor (x \land z) \lor (y \lor z) = x \lor (y \land z)
\]

ii) Suppose $y, z$ are both complements for $x$. Then:
\[
y \land (x \lor z) = y \land 1 = y 
\]
but
\[
(y \lor x) \lor (y \land z) = 0 \lor (y \land z) = y \land z
\]
and similarly $z = y \land z$, so $y = z$. 
We write 2 for the two element lattice \([0, 1]\).

1.9 Lemma

Let \(a, b\) be elements of a distributive lattice \(L\) with \(a \neq b\). Then there exists a lattice homomorphism \(f : L \to 2\) with \(f(a) = 1\) and \(f(b) = 0\).

Proof

Let \(P\) be the set of ordered pairs \((A, B)\) of subsets of \(L\) satisfying:

i) \(x \in A\), \(x \leq y \Rightarrow y \in A\), and \(A\) is closed under finite meets.

ii) \(x \leq y, y \in B \Rightarrow x \in B\), and \(B\) is closed under finite joins.

iii) \(A \cap B = \emptyset\).

We partially order \(P\) by \((A_1, B_1) \leq (A_2, B_2) \iff A_1 \leq A_2, B_1 \leq B_2\).

\(P\) is chain complete: given a chain \(\{(A_i, B_i) \mid i \in I\}\),

\(\bigcup_i A_i\) satisfies the closure properties (i), and similarly

\(\bigcup_i B_i\) satisfies (ii).

Suppose \(x \in (\bigcup_i A_i) \cap (\bigcup_i B_i)\), then \(x \in A_i\) for some \(i\), and \(x \in B_j\) for some \(j\). But either \(A_i \leq A_j\) or \(B_i \leq B_j\), so either \(x \in A_j \cap B_i\) or \(x \in A_i \cap B_j\).

So \((\bigcup_i A_i, \bigcup_i B_i) \in P\), and is a least upper bound for the chain.

Now \((\top(A), \top(B)) \in P\), since \(a \neq b\), so we can find a maximal element \((A_0, B_0)\) of \(P\) with \(a \in A_0\), \(b \in B_0\).
Suppose $A_0 \cup B_0 \neq L$, let $c \in L \setminus (A_0 \cup B_0)$.

Let $A_1 = \{ y \in L \mid y \geq x \wedge c \text{ for some } x \in A_0 \}$

then $A_1$ satisfies (i) and $A_0 \subseteq A_1$, since $c = 1 \wedge c \in A_1$.

Hence $(A_1, B_0) \notin \mathcal{P}$ by maximality of $(A_0, B_0)$, so $A_1 \cap B_0 \neq \emptyset$, i.e. $\exists x \in A_0$ such that $x \wedge c \in B_0$.

Similarly, considering $B_1 = \{ x \in L \mid x \leq y \vee c \text{ for some } y \in B_0 \}$

we deduce that there exist $y \in B_0$ such that $y \vee c \in A_0$.

Now $x \wedge (y \vee c) \in A_0$ since it is a meet of two elements of $A_0$, but it equals $(x \wedge y) \vee (x \wedge c)$ which is a join of two elements of $B_0$.

So $A_0 \cup B_0 = L$, and the total function $f : L \to 2$ defined by

$f(x) = \begin{cases} 1 & x \in B_0 \\ 0 & x \in A_0 \end{cases}$

is a lattice homomorphism with $f(1) = 1$, $f(0) = 0$.

1.20 Theorem (Birkhoff–Stone)

Any distributive lattice is isomorphic to a sub-lattice of some power set $\mathcal{P}(A)$, i.e. to a family of subsets of $A$ closed under finite unions and intersections. In particular, any Boolean algebra is isomorphic to a sub-Boolean algebra of $\mathcal{P}(A)$, i.e. a subset of $\mathcal{P}(A)$ closed under finite unions.

Proof

Given a distributive lattice $L$, let $A$ be the set of all lattice homomorphisms $L \to 2$. 
Define $\Phi : L \to \text{PA}$ by $\Phi(x) = \{ f \in A | f(x) = 1 \}$

$\Phi$ is a lattice homomorphism: we have $f \in \Phi(x, y) \Leftrightarrow f(x, y) \Leftrightarrow f(x) = 1 \Leftrightarrow f \in \Phi(x) \land \Phi(y)$

and similarly $f \in \Phi(x \land y) \Leftrightarrow f(x \land y) = 1 \Leftrightarrow$ either $f(x) = 1$ or $f(y) = 1 \Leftrightarrow f \in \Phi(x) \lor \Phi(y)$.

Also, $\Phi$ is injective since if $a \neq b$ then $1 \cdot b$ yields $f \in A$ belonging to just one of $\Phi(a)$ and $\Phi(b)$.

So $\Phi$ is an isomorphism from $L$ to the sub-lattice

$\{ \Phi(x) | x \in L \}$ of $\text{PA}$.

In particular, if $L$ is Boolean, then this sub-lattice is a sub-

Boolean-algebra of $\text{PA}$.

Chapter II: The Propositional Calculus

2.1 Definition

A primitive proposition is an abstract symbol $p$ with incapable of

being assigned a truth value 0 (false) or 1 (true).

Given a set $P$ of primitive propositions, a valuation of $P$

is a function $V : P \to 2$. Given such a set $P$, the

compound propositions (or propositional formulae) over $P$ are the

members of the set $L(P)$ defined recursively by:

i) If $p \in P$ then $p \in L(P)$

ii) If $s, t \in L(P)$, then $(s \Rightarrow t) \in L(P)$

iii) $\perp \in L(P)$
Given a valuation $\mathcal{V} : P \rightarrow 2$, we extend it to a function $\mathcal{V} : \mathcal{L}(P) \rightarrow 2$ by

i) $\mathcal{V}(p) = \mathcal{V}(p)$ if $p \in P$

ii) $\mathcal{V}(\neg s) = 0$ if and only if $\mathcal{V}(s) = 1$ and $\mathcal{V}(\neg s) = 0$

iii) $\mathcal{V}(\top) = 0$

The clause for $\mathcal{V}(s \rightarrow t)$ can be presented as a truth table:

$\begin{array}{ccc}
  s & t & s \rightarrow t \\
  0 & 0 & 1 \\
  0 & 1 & 0 \\
  1 & 0 & 1 \\
  1 & 1 & 1 \\
\end{array}$

We can use the same idea to evaluate the truth or falsity of more complicated propositions.

We define $\neg p$ to be $(p \Rightarrow \bot)$ since this has truth table:

$\begin{array}{ccc}
  p & \neg p \\
  0 & 1 \\
  1 & 0 \\
\end{array}$

We now define $\top = \top \bot$, and also $\mathcal{V}(p \Rightarrow q) = \mathcal{V}(p \lor q)$ to be $(\neg p \Rightarrow q)$, $(p \lor q) = \mathcal{V}(p \Rightarrow q)$.

$\begin{array}{cccccc}
  p & q & \neg p & p \lor q & q & \neg p \Rightarrow q \\
  0 & 0 & 1 & 1 & 0 & 0 \\
  0 & 1 & 1 & 1 & 1 & 1 \\
  1 & 0 & 0 & 1 & 0 & 1 \\
  1 & 1 & 0 & 1 & 0 & 0 \\
\end{array}$

$(p \Rightarrow q) \Rightarrow (p \Rightarrow q)$ is defined to be $((p \Rightarrow q) \lor (q \Rightarrow p))$.
2.2 Lemma. (Functional Completeness of Propositional Calculus)

For any $n > 0$, each function $2^n \to 2$ occurs as the truth table of some $s \in \mathcal{L}(\{p_1, p_2, \ldots, p_n\})$

Proof. (By induction on $n$)

For $n = 0$, there are two functions $2^0 = \{\star\} \to 2$, corresponding to the two elements 0 and 1, and these are the truth tables of $\bot$ and $T = (\bot \Rightarrow \bot)$ respectively.

Assume true for $n$, and let $f : 2^{n+1} \to 2$. Define $f_0, f_1 : 2^n \to 2$ by $f : (x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n, \star)$. We can find $s_0, s_1, s \in \mathcal{L}(\{p_1, \ldots, p_n\})$ whose truth tables are $f_0, f_1$. Now consider $s = (s \land \neg p_{n+1}) \lor (s \land p_{n+1})$.

It is immediately true that this has $f$ as its truth table. \(\Box\)

When do two formulae have the same truth table?

Note that this happens for $s, t$ if and only if $(s \Rightarrow t)$ has truth table the constant function with value 1.

2.3 Definition

Let $p$ be a set of primitive propositions, let $s \in \mathcal{L}(p)$ and let $t \in \mathcal{L}(p)$. We say that $s$ semantically entails $t$ and write $s \models t$, if any valuation $v : p \to 2$ such that $v(s) = 1 \forall s \in S$ also satisfies $v(t) = 1$.

In the particular case with $S = \emptyset$, we simply write $t \models t$ and call it a tautology.
2.4 Examples

a) $F = (p \rightarrow (q \rightarrow p))$ since this formula has truth table

<table>
<thead>
<tr>
<th>$p$</th>
<th>$q$</th>
<th>$(q \rightarrow p)$</th>
<th>$p \rightarrow (q \rightarrow p)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

b) $F = (((p \rightarrow (q \rightarrow r)) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))) = S$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$q$</th>
<th>$(p \rightarrow q)$</th>
<th>$(q \rightarrow r)$</th>
<th>$(p \rightarrow r)$</th>
<th>$(p \rightarrow q) \rightarrow (p \rightarrow r)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

c) $\{p, (p \rightarrow q)\} \vdash q$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$q$</th>
<th>$(p \rightarrow q)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

2.5 Definition

The axioms of the propositional calculus are all formulas of the three forms

- (S) $S \rightarrow (t \rightarrow S)$
- ((S ⊃ (t ⊃ u))) ⊃ ((S ⊃ t) ⊃ (S ⊃ u))
- ((S ⊃ t) ⊃ (t ⊃ u)) ⊃ S

or ((S ⊃ t) ⊃ (t ⊃ u)) ⊃ S

We have one rule of inference: from S, and (S ⊃ t), we may infer (deduce) t.

A deduction from a set S of hypotheses is a finite list $(S_1, S_2, \ldots, S_n)$ of propositional formulas, such that for each i, we have one of:

1. $S_i \in S$
2. $S_i$ is an axiom
3. There exist $s, k < i$ such that $S_k = (S_i \rightarrow S_i)$
We say that $S$ syntactically entails $t$, and write $S \vdash t$ if there exists a deduction from $S$ whose last line is $t$.
We call a deduction from $\emptyset$ a proof, and write $\vdash t$, and say that $t$ is a theorem if it is deducible from $\emptyset$.

2.6 Examples

a) The following is a proof of $(s \Rightarrow s)$:
1. $(s \Rightarrow (s \Rightarrow s))$ (K)
2. $(s \Rightarrow ((s \Rightarrow s) \Rightarrow s))$ (K) with $s$, $(s \Rightarrow s)$
3. $((s \Rightarrow (s \Rightarrow s) \Rightarrow s) \Rightarrow ((s \Rightarrow (s \Rightarrow s)) \Rightarrow (s \Rightarrow s)))$ (axiom S) with $s$, $(s \Rightarrow s)$
4. $((s \Rightarrow (s \Rightarrow s)) \Rightarrow (s \Rightarrow s))$ (MP from lines 2 and 3)
5. $(s \Rightarrow s)$ (MP from lines 1 and 4)

b) The following is a deduction of $(s \Rightarrow u)$ from $\{ (s \Rightarrow t), (t \Rightarrow u) \}$
1. $((s \Rightarrow (t \Rightarrow u)) \Rightarrow ((s \Rightarrow t) \Rightarrow (s \Rightarrow u)))$ (axiom S)
2. $(s \Rightarrow u)$ (hypothesis)
3. $((t \Rightarrow u) \Rightarrow (s \Rightarrow (t \Rightarrow u)))$ (axiom K)
4. $(s \Rightarrow (t \Rightarrow u))$ (MP from 2, 3)
5. $((s \Rightarrow t) \Rightarrow (s \Rightarrow u))$ (MP from 1, 4)
6. $(s \Rightarrow t)$ (hypothesis)
7. $(s \Rightarrow u)$ (MP from 5, 6)
2.7 Lemma (Soundness Theorem)

If \( S \vdash t \), then \( S \models t \). In particular, every Theorem is a tautology.

Proof

Let \( (s_1, s_2, \ldots, s_n = t) \) be a deduction of \( t \) from \( S \).

We show \( S \models s_i \) for all \( i \), by induction on \( i \).

Clearly, \( s_i \in S \implies S \models s_i \).

If \( s_i \) is an axiom, then \( 0 \models s_i \) and \( 0 \models S \).

If we have \( j, k < i \) such that \( s_k = (s_j \models s_i) \), then by induction any model \( \mathcal{V} \) for \( S \) (a valuation making all of \( S \) true) satisfies \( \mathcal{V}(s_j) = 1 \) and \( \mathcal{V}(s_k) = 1 \), so by 2.4 (c) we have \( \mathcal{V}(s_i) = 1 \).

2.8 Theorem (Deduction Theorem)

Let \( S \models \forall (P), s, t \in \forall (P) \). Then

\( S \vdash (s \models t) \) if and only if \( S \cup \{s\} \vdash t \)

Proof

(\( \Rightarrow \)) Given a deduction \( s_1, \ldots, s_n = (s \models t) \) from \( S \), add the true lines \( s_{n+1} = S \) (hypothesis), then \( S_{n+2} = t \) (MP).

(\( \Leftarrow \)) Suppose given a deduction \( (t_1, t_2, \ldots, t_n = t) \) of \( t \) from \( S \cup \{s\} \) we show by induction on \( i \) that \( S \vdash (s \models t_i) \). 

If \( t_i \) is an axiom or a member of \( S \), we write down (MP)

\( t_i \) (axiom or hypothesis), \( t_i \models (s \models t_i) \) (axiom \( k \)), \( s \models t \).
If $t_i \in S$, we write down the 5 line proof of $(S \Rightarrow S)$ in 2.6 a).

If $\exists s, k < i$ with $t_k = (t_i \Rightarrow t_i)$, we write down deduction of $(S \Rightarrow t_i)$ and of $(S \Rightarrow (t_i \Rightarrow t_i))$, then write down $(S \Rightarrow (t_i \Rightarrow t_i)) \Rightarrow ((S \Rightarrow t_i) \Rightarrow (S \Rightarrow t_i))$ (axioms) and apply MP twice.

$S \vdash t$: syntactic entailment, a deduction exists

$S \models t$: semantic entailment, if $S$ true means $t$ is true

Model: a valuation $V: \mathbb{L}(P) \rightarrow \mathbb{Z}$ on which $V(s) = 1$ for $s \in S$. 
S \leftarrow (9) V \\
\text{method to } S_0 \text{ in } \text{JbM}
2.9 Theorem (Completeness Theorem)

Given $S \models s(t)$, $t \in \mathcal{L}(P)$, we have $S \vdash t \iff S \models t$

Proof:

$(\Rightarrow)$ in Lemma 27.

$(\Leftarrow)$ Using 2.8, we may reduce this problem to the particular case $t = \top$:

If $S \models t$, then $S \cup \{\neg t\} \not\models \bot$ and if $S \cup \{\neg t\} \not\models \bot$, then by 2.8 we have $S \vdash \neg \neg t$ and we can convert a deduction of $\neg \neg t$ from $S$ into a deduction of $t$ by adding the axiom $(\neg \neg t \Rightarrow t)$ and applying modus ponens.

In the case $t = \bot$, we prove the contrapositive implication.

If $S$ is consistent (i.e. $S \not\models \bot$) then $S$ has a model (i.e. $S \not\models \bot$). Consider the set $Q$ of all consistent subsets of $\mathcal{L}(P)$, ordered by inclusion. $Q$ is chain-complete: if $\{T_i : i \in I\}$ is a chain, then $\bigcup_{i \in I} T_i$ is consistent since a deduction of $\bot$ from it would involve only finitely many members of $\bigcup_{i \in I} T_i$, and we could find $i \in I$ such that all of them belong to $T_i$. Hence, by Zorn's Lemma there exists a maximal consistent set $\overline{S}$ containing $S$.

$\overline{S}$ is deductively closed i.e. $\overline{S} \vdash t$ implies $t \in \overline{S}$ since if $\overline{S} \vdash t$, then anything deducible from $\overline{S} \cup \{t\}$ is deducible from $\overline{S}$, and hence by maximality $\overline{S} \cup \{t\} = \overline{S}$. 
Also, for every $t \in \lambda(P)$ we have either $t \in \bar{S}$ or $7t \in \bar{S}$ since if $t \notin \bar{S}$, then $\bar{S} \cup \{t\} \vdash \bot$ and by 2.8 $\bar{S} \vdash \bot$, so $7t \in \bar{S}$.

Now we define $v : P \to 2$ by $v(p) = \emptyset$ if and only if $p \in \bar{S}$.

We claim that the canonical extension $v \cdot \lambda(P) \to 2$ of $r$ satisfies $v(t) = 1$ if and only if $t \in \bar{S}$. (\textit{\#})

We prove this by induction on the structure of $t$:

- true by definition of $t \in P$, and true for $\bot$ since $\bar{S}$ is consistent. Suppose (\textit{\#}) holds for both $s$ and $t$, and consider $(s \Rightarrow t)$.

Case 1: $t \in \bar{S}$, so $v(t) = 1$. Then $v(s \Rightarrow t) = 1$, but $\{t\} \vdash (s \Rightarrow t)$, so since $\bar{S}$ is deductively closed, we have $(s \Rightarrow t) \in \bar{S}$.

Case 2: $s \notin \bar{S}$, so $v(s) = 0$, and again we have $v(s \Rightarrow t) = 1$. But $\forall s \in \bar{S}$, and we have $\{\bot\} = (s \Rightarrow t)$ (\textit{\#}4, \textit{\#}5 \textit{\#}2) and so $(s \Rightarrow t) \in \bar{S}$.

Case 3: $s \in \bar{S}$, $t \notin \bar{S}$, then $v(s) = 1, v(t) = 0$, so $v(s \Rightarrow t) = 0$.

But $(s \Rightarrow t) \notin \bar{S}$, since $\{s, (s \Rightarrow t)\} \vdash \bot$.

Hence $r$ is a model for $\bar{S}$, and in particular a model for $S$.

2.10 Remarks

a) Given a valuation $v : P \to 2$, the set $\{s \in \lambda(P) \mid v(s) = 1\}$ is deductively closed, consistent, and for every $s$ it contains either $s$ or $\neg s$. So it is a maximal consistent set.
Hence, the maximal consistent subsets of $\mathbb{L}(P)$ are exactly the sets of formulae true under some valuation.

6) If $P$ is countable, then there is a proof of 2.9 which does not use Zorn's Lemma. $P$ countable $\Rightarrow \mathbb{L}(P)$ countable, so we can enumerate the members of $\mathbb{L}(P)$, as $(t_0, t_1, t_2, \ldots)$

Given a consistent set $S$, we can enlarge it to a maximal consistent set as follows:

- go through the $t_t$ one by one, and at the $n$th stage if $t_n$ can be consistently added to our current set $S_n$, we do so, and otherwise, we add $\neg t_n$. $S_{n+1} = S_n \cup \{ t_n \}$ or $S_{n+1} = S_n \cup \{ \neg t_n \}$

(This works since if $S_n \cup \{ t_n \} \vdash \bot$, then $S_n \vdash \neg t_n$. Then $S = \bigcup_{n \in \mathbb{N}} S_n$. $S$ is consistent since it is the union of a chain of consistent sets, and for every $t \in \mathbb{L}(P)$, we have either $t \in S$ or $\neg t \in S$. So $S$ is maximal consistent.

7) Given $S \subseteq \mathbb{L}(P)$, we define a relation $\leq_S$ on $\mathbb{L}(P)$ by $t \leq_S u$ if $S \vdash t \supset u$. This is reflexive and transitive, so if we define $\equiv_S$ by $t \equiv_S u$ if $t \leq_S u$ and $u \leq_S t$, then $\equiv_S$ is an equivalence relation, and $\leq_S$ induces a partial order on the quotient set $\mathbb{B}(S) = \mathbb{L}(P) / \equiv_S$ (the Lindenbaum algebra of a set $S$). We claim that $\mathbb{B}(S)$ is a Boolean algebra: the meet of $[S]$ and $[t]$ is $[S \land t]$, and their join is $[S \lor t]$. The complement of $[S]$ is $[\neg S]$. 

Theorem 7.5
$S$ is consistent $\iff [T] \neq [L]$, i.e. $B(S)$ has $\geq 2$ elements.

So by 1.19 this implies that there exists a homomorphism $f : B(S) \to 2$. Then the composite $\lambda(P) \to B(S) \xrightarrow{f} 2$ coincides with $\bar{v}$, where $v$ is its restriction to $P \subseteq \lambda(P)$.

Again, models for $S$ correspond bijectively to homomorphisms $B(S) \to 2$.

2.11 Corollary (Decidability Theorem)
Let $S \equiv \lambda(P)$ be finite. Then there is an algorithm which determines, for any $t \in \lambda(P)$, whether or not $S \vdash t$.

Proof:
This is obvious for $\Gamma$ (truth tables).
2.12 Corollary (Compactness Theorem)
Let $\mathcal{S} \subseteq \mathcal{L}(P)$. If every finite subset of $\mathcal{S}$ has a model, then $\mathcal{S}$ has a model.

Proof.
$\mathcal{S}$ has a model if and only if $\mathcal{S} \not\models \bot$
But we know that if $\mathcal{S} \models \bot$ then there is a finite $\mathcal{S}' \subseteq \mathcal{S}$
with $\mathcal{S}' \models \bot$, the same holds for $\bot$.

2.13 Remark.
Let $\mathcal{V}$ be the set of all valuations of $\mathcal{P}$ (= all functions $P \to \mathbb{2}$).
If we define $\mathcal{U}_t$, for any $t \in \mathcal{L}(P)$, to be $\{v \in \mathcal{V} \mid v(t) = 1\}$
then the $\mathcal{U}_t$ form a base for a topology on $\mathcal{V}$, since
$\mathcal{U}_s \cap \mathcal{U}_t = \mathcal{U}(s \wedge t)$. The Compactness Theorem is exactly
the assertion that this space is compact, since
$\{\mathcal{U}_t \mid t \in \mathcal{S}\}$ covers $\mathcal{V} \iff \{\mathcal{U}_t \mid t \in \mathcal{S}\} \models \bot$.
(In fact, $\mathcal{V}$ is homeomorphic to the product of Props
of the discrete space $\{0, 1\}$, so the Compactness Theorem
follows from Tychonoff's Theorem).

2.14 Examples

a) We use Compactness to show that any partial ordering
on a set can be extended to a total ordering.
Given a poset $(\mathcal{A}, \preceq)$, take $\mathcal{P}$ to be the set of all
propositions $p_{xy}, x \neq y$ in $\mathcal{A}$. 

New Axiom:
Axiom 1

We add the new axiom:

New Axiom 2

We add the new axiom:...
Take $S$ to consist of all propositions $p_{xy} (x < y \in A)$

\[(p_{xy} \Leftrightarrow \neg p_{yx}) \quad x \neq y \in A.\]

\[(p_{xy} \Rightarrow (p_{xz} \Rightarrow p_{x,z})) \quad (x, y, z \text{ any } 3 \text{ elements of } A).\]

Then $v$ is a model for $S$ if and only if

\begin{align*}
&\{(x, y) \mid \text{either } x = y \text{ or } v(p_{xy}) = 1\}
&\text{is a total ordering extending } \leq.
\end{align*}

So such an ordering exists providing that every finite subset $S' \subseteq S$ has a model. Given such an $S'$, we can find a finite $A' \subseteq A$ such that all the $p_{xy}$'s occurring in members of $S'$ have $[x, y] \subseteq A'$. So we can construct a model for $S'$ by constructing a total ordering of $A'$ which extends $(\leq \cap A' \times A')$, and this can be done without Zorn's Lemma, by making a finite number of choices.

\[G = (E \cup V)\]

b) Let $G$ be a connected graph. By an $n$-colouring of $G$, we mean a function $V \rightarrow \{1, 2, \ldots, n\}$ such that if $\{x, y\}$ is an edge, then $c(x) \neq c(y)$.

We claim that a graph is $n$-colourable if and only if all of its finite subgraphs are $n$-colourable.

To prove this, let $P = \{\sigma_{xi} \mid x \in V, 1 \leq i \leq n\}$ and let $S = \{([\sigma_{xi} \Rightarrow (\sigma_{xi} \Rightarrow \bot)] \cup \{x \in V, i \neq i\})

\begin{align*}
&\cup \{V_i \leq p_{xi} \mid x \in V\} \cup \{(\sigma_{xi} \Rightarrow (\sigma_{xi} \Rightarrow \bot))
&\mid \exists x, y \in V, 1 \leq i \leq n\}.
\end{align*}
Then a model for $S'$ is an $n$-colouring of $(V, E)$.

If $S' \subseteq S$ is finite, then we can find a finite $V' \subseteq V$ such that the $\pi x$ occurring in members of $S'$ all satisfy $x \in V'$, so an $n$-colouring of $(V', E' = E \cap \mathcal{P}(V'))$ will yield a model for $S'$.

Chapter 3: The Predicate Calculus

3.1 Definition

By a signature, we mean a pair $\Sigma = (\Sigma, \Pi)$, where $\Sigma$ is a set of operation symbols, each equipped with an arity $\alpha(\omega) \in \mathbb{N}$ and $\Pi$ is a set of relation symbols $\pi$ with a similar arity.

Given a signature $\Sigma$, a $\Sigma$-structure is a set $A$ equipped with operation $\omega_A : A^{\alpha(\omega)} \to A$, for each $\omega \in \Sigma$, and relation $\pi_A \subseteq A^{\alpha(\pi)}$ for each $\pi \in \Pi$.

We could (and many logicians do) dispense with operations symbols, by replacing them with each $n$-ary operation $\omega$ by an $(n+1)$-ary relation $\omega^*$, where $(x_1, \ldots, x_n+1) \in [\omega^*]$ if and only if $x_{n+1} = \omega(x_1, \ldots, x_n)$.

3.2 Definition

Given a signature $\Sigma = (\Sigma, \Pi)$, we define the set of terms over $\Sigma$ as follows:

a) We have a supply of variables $x, y, z, \ldots$ (or $x, x', x''$, etc.) which are terms.
b) If $t_1, \ldots, t_n$ are terms, then $W$ an $n$-ary operation on $K$, then $w(t_1, t_2, \ldots, t_n)$ is a term.

8. If $A$ is a $\Sigma$-structure, we recursively define the interpretation $t_A$ in $A$ of each term $t$ involving variables in the set $\{x_1, \ldots, x_n\}$ as a function $A^n \to A$:

- If $t = x_i$ for some $i$, then $t_A(a_1, \ldots, a_n) = a_i$.
- If $t = w(u_1, u_2, \ldots, u_m)$ where $\alpha(w) = m$, then $t_A$ is the composite
  $$A^n \xrightarrow{(u_1)_A, \ldots, (u_m)_A} A^m \xrightarrow{\alpha A} A$$
  (interpreted by juxtaposition)

  E.g. if we have a binary operation $\cdot$, then $\alpha(xy)z$ is interpreted in a structure $A$ as the mapping $(a, b, c) \mapsto a(bc)$ and $\alpha(xz)y$ is interpreted by the mapping $(a, b, c) \mapsto (ab)c$. 
3.3 Definition

Let $Σ = (\mathcal{L}, IT)$ be a first order signature. The set $η(Σ)$ of (first-order) formulae over $Σ$ is defined as follows:

a) atomic formulae which are of two kinds: if $t \in IT$, $α(t) = n$ and $t_1, \ldots, t_n$ are terms then $t_1(t_2, \ldots, t_n)$ is an atomic formula; and if $s, t$ are two terms, then $(s = t)$ is an atomic formula.

b) $\top$ is a formula, and if $\varphi, \psi$ are formulae, so is $(\varphi \Rightarrow \psi)$.
(We introduce $\top, (\varphi \land \psi), (\varphi \lor \psi), (\varphi \Rightarrow \psi)$ as in Chapter 2.)

c) If $\varphi$ is a formula, and $x$ is a free variable of $\varphi$, then $(\forall x)\varphi$ is a formula. (We introduce $(\exists x)\varphi$ as shorthand for $\forall x(\exists x)\varphi$.)

The (finite) set $FV(\varphi)$ of free variables in $\varphi$ is defined by:

- for an atomic formula $\varphi$, $FV(\varphi)$ is the set of all variables which occur in $\varphi$.
- $FV(\top) = \emptyset$, $FV(\varphi \Rightarrow \psi) = FV(\varphi) \cup FV(\psi)$
- $FV(\forall x\varphi) = FV(\varphi) \setminus \{x\}$

Note that a variable may appear both free and bound in the same formula: e.g. $(\forall x(y) \land (\exists x)z(x, y))$. We say that $\varphi, \psi$ are $α$-equivalent if $\psi$ is obtained from $\varphi$.
by renaming its bound variables.

So \((\pi_1(\alpha, y)) \wedge (\forall w) \pi_2(z, w)\) is \(\alpha\)-equivalent to \((*)\).

Given \(\Sigma\)-structure \(A\), we interpret any formula \(\varphi\) with \(FV(\varphi) \subseteq \{x_1, x_2, \ldots, x_n\}\) as a set \([\varphi]_A \subseteq A^n\), or equivalently a function \(\varphi_A : A^n \to 2\), as follows:

If \(\varphi \equiv \top\), \(\varphi_A\) is the composite
\[
A^n \xrightarrow{(\text{and})} A^2 \xrightarrow{\pi_A} 2
\]
where \(\pi_A\) is the characteristic function of \([\top]_A \subseteq A^n\).

If \(\varphi \equiv (s = t)\), \(\varphi_A\) is the composite
\[
A^n \xrightarrow{(s=t)} A^2 \xrightarrow{5} 2,
\]
where \(5(x, y) = 1\) iff \(x = y\).

If \(\varphi \equiv \bot\), \(\varphi_A\) is the constant function with value 0.

If \(\varphi \equiv (\forall \theta) \psi\), \(\varphi_A\) is the composite
\[
A^n \xrightarrow{(\forall \theta)} A^2 \xrightarrow{\Theta_\theta} 2,
\]
where \(\Theta_\theta(x, y) = 1\) whenever \(x = 1, y = x\).

If \(\varphi \equiv \forall \theta \psi\), then
\[
[\varphi]_A = \{ \langle x_1, \ldots, x_n \rangle \in A^n \mid \forall x_{n+1} \in A, \langle x_1, \ldots, x_n, x_{n+1} \rangle \in [\psi]_A \}.
\]

By a context for \(\varphi\), we mean a sequence \((x_1, \ldots, x_n)\) of distinct variables including all the free variables of \(\varphi\).

By the canonical context for \(\varphi\), we mean a list of its free variables in the order of their first (free) occurrence in \(\varphi\).

Unless otherwise stated, we interpret formulae in their canonical context.
3.4 Definition

We say that a formula \( \varphi \) is satisfied in a structure \( A \), and write \( A \models \varphi \), if \( \varphi_A : A^\varphi \rightarrow 2 \) is the constant function with value 1 (equivalently, \( \varphi_A = A^\varphi \)). Note that if \( x \in \text{FV}(\varphi) \), then \( A \models \varphi \iff A \models (\forall x)\varphi \).

By the universal closure of \( \varphi \), we mean the formula
\[
(\forall x_1)(\forall x_2) \cdots (\forall x_n) \varphi,
\]
where \( (x_1, x_2, \ldots, x_n) \) is the canonical context for \( \varphi \). We abbreviate this to \( (\forall x_1, x_2, \ldots, x_n) \varphi \).

A sentence or closed formula is one with no free variables.

By a first-order theory over \( \Sigma \), we mean a set \( T \) of sentences in \( \text{L}(\Sigma) \). We say that \( A \) is a model for \( T \) (and write \( A \models T \)) if \( A \models \varphi \) for all \( \varphi \in T \).

3.5 Examples

a) The theory of groups is written over a signature \((\Sigma, IT)\) with
\[
\Sigma = \{ M, i, e \}, \quad \text{with} \quad \alpha(M) = 2, \ \alpha(i) = 1, \ \alpha(e) = 0, \ \text{and} \quad IT = \emptyset.
\]
The axioms of \( T \) are:
\[
(\forall x, y, z)(MxMyz = Mx(Myz))
\]
\[
(\forall x)(Mxe = x)
\]
Then a \( T \)-model is exactly a
\[
(\forall x)(Mxi = x)
\]
group.

b) The theory of fields has \( \Sigma = \{ +, \times, 0, 1, - \} \) and \( IT = \emptyset \).
We write down the axioms of commutative ring, with 1 as universal closure of equations.
Add the axioms \((10 \neq 1) \Rightarrow \perp\) and
\((\forall x)(x = 0) \lor (\exists y)(x \cdot y = 1)\)

c) For the theory of posets, we take \(Q = \emptyset\), \(IT = \{1, \ldots, 3\}\) (write \((s \leq t)\) for \(s \leq (s, t)\)) and axioms,
\((\forall x)(x < x)\)
\((\forall x, y, z)(x < y \Rightarrow (y \leq z) \Rightarrow (x < z))\)
\((\forall x, y)(x \leq y \Rightarrow (y \leq x) \Rightarrow (x = y))\)

For sets, we add the axiom \((\forall x, y)(x \leq y) \lor (y \leq x)\)

d) The theory of (combinatorial) projective planes has \(Q = \emptyset\),
\(IT = \{\Pi, \lambda, \Theta\}\) with \(\alpha(\Pi) = \alpha(\lambda) = 1\), \(\alpha(\Theta) = 2\),
and axioms \((\forall x)(\Pi(x) \lor \lambda(x))\)
\((\forall x) \top (\Pi(x) \land \lambda(x))\)
\((\forall x, y)(x \leq y) \Rightarrow (\Pi(y) \lor \lambda(y))\)
\((\forall x, y)(\Pi(x) \land \Pi(y) \land (x = y)) \Rightarrow (\exists z)(\Pi(\Pi(z) \land \Pi(y) \land (z = x))\lor...
\quad (\forall w)((\lambda(x) \land \lambda(y) \land (y = w)) \Rightarrow (z = w)))))

and the dual of this.
Substitution

If \( \Phi \) is a formula, \( x \) a variable, and \( t \) a term, we write \( \Phi [t/x] \) for the formula obtained from \( \Phi \) by replacing each free occurrence of \( x \) in \( \Phi \) by a copy of \( t \), provided that none of the variables of \( t \) occur bound in \( \Phi \), otherwise, we must first rename the bound variables of \( \Phi \) so that they don't clash with variables in \( t \).

Similarly, we write \( \Phi [t_1/t_2/x_1, x_2] \), note that this is not the same as \( \Phi [t_1/x_1, t_2/x_2] \).

3.6 Definition

Suppose that \( T \) is a set of sentences and \( \Phi \) a sentence. We say that \( T \) semantically entails \( \Phi \), and write \( T \models \Phi \), if every model for \( T \) satisfies \( \Phi \).

If \( T \) and \( \Phi \) are free variables, then we say \( T \models \Phi \) if \( T' \models \Phi' \), where \( \Phi' \) is obtained from \( \Phi \) by adding new constants (= nullary operation symbols) \( c_i \) to our signature for each free variable \( x_i \) in \( \Phi \), and setting \( \Phi' = \Phi [c_1, c_2, \ldots, c_{n-1}, x_{n-1}, c_n, \ldots] \). \( T' \) is obtained from \( T \) by substituting the \( c_i \) and then for the \( x_i \) in the members of \( T \), and then quantifying universally over any remaining free variables.

3.7 Definition

The axioms of the predicate calculus are all formulas of the following 7 forms:

\[
(\Phi \Rightarrow (\Psi \Rightarrow \Phi)) \quad (k)
\]

\[
(\Phi \Rightarrow (\Psi \Rightarrow \chi)) \Rightarrow ((\Phi \Rightarrow \Psi) \Rightarrow (\Phi \Rightarrow \chi)) \quad (S)
\]
(77 \phi \Rightarrow \psi) \quad (T)

((\forall x) \phi \Rightarrow \psi[\forall x]) \quad (\forall e \in FV(\phi)) \quad (I)

((\forall x) (\psi \Rightarrow \psi)) \Rightarrow ((\forall e (\forall x) \psi)) \quad (\forall e \in FV(\psi)) \quad (\forall e)

(\forall x) (x = x) \quad (R)

((\forall x, \psi) (x = y) \Rightarrow ((\psi \Rightarrow (\forall x) \psi[\forall y] x))) \quad (Sub)

Our rules of inference are:

1. From \phi and (\psi \Rightarrow \psi), we may infer \psi, provided either \phi has
   a free variable or \psi doesn't. Note that if \phi has a free
   variable, then \phi \models \psi, and \phi \models (\psi \Rightarrow \psi), but \phi \not\models \psi
   (MA)

2. From \psi, we may infer (\forall x) \phi, provided that \psi does not
   occur free in any hypothesis used in the deduction of \phi.
   (Generalisation)

By a deduction in the predicate calculus from a set of
hypotheses T, we mean a finite sequence \psi_1, \psi_2, \ldots, \psi_n
such that for each i we have either \psi_i an axiom
or \psi_i obtained from earlier formulas by the rules
of inference.

We say that T syntactically entails \phi if there is a
deduction from T whose last member is \phi, and write T \vdash \phi.

3.8 Lemma (Soundness Theorem)

If T \vdash \phi then T \models \phi.

Proof: (by induction on the length of a deduction of \phi from T).
(essentially the same as 2.7)

3.9 Proposition (Deduction Theorem)

Suppose either that \phi has a free variable, \phi doesn't.
Then
\[ T \vdash (\forall x \exists y) \iff T \cup \{ \exists x \forall y \} \vdash \neg \forall x \exists y \]

Proof:
\[ \exists \]  Exactly like 2.8
\[ \iff \]  Mostly like 2.8, but we have to consider the case when
\[ \forall x (\forall x \exists y) \]  was obtained by an application of (Gen).

Then, by our inductive hypothesis, we have a deduction of
\[ (\forall x \exists y) \]  from \( T \), and we can add
\[ (\forall x \exists y) \]  (Gen)
\[ ((\forall x \exists y) \vdash \exists y \vdash (\exists y (\forall x \exists y))) \]  (IP)
\[ (\forall x \exists y \vdash (\exists y (\forall x \exists y))) \]  (MP)
provided \( \exists x \notin \text{FV}(\forall) \).

But if \( x \) is free in \( \forall \), then \( \forall \) wasn't used in the deduction of
\[ x, \]  so we actually have \( T \vdash (\exists x (\forall x \exists y)) \), from which we can
obtain \( T \vdash (\forall x \exists y) \) by (Le) plus (MP).

3.10 Theorem (Completeness Theorem)
\[ T \vdash y \iff T \vdash \neg y \]

Proof:
\[ \exists \]  is 3.8
\[ \iff \]  We use 3.9 plus axiom (T) to reduce to the case \( y = 1 \)
Then, reduce to the case when \( T \) is a set of sentences: we need to show that if \( T \) is consistent (i.e., \( T \vdash \bot \)) then \( T \) has a model.

Basic idea: suppose that \( T \) is maximal consistent (i.e., for
any sentence \( y \) we have \( y \in T \) or \( \neg y \in T \)) and
\( \forall x \exists y \) witnesses, i.e., whenever \( \text{FV}(\forall) = \{x\} \) and
\( T \vdash (\exists x) \forall \) there is a closed term \( t \) (one with no variables)
such that $T\vdash \Gamma[\epsilon/x]$.  

Then $T$ has a term model, i.e. if $C$ is the set of closed terms, and $\equiv$ is the equivalence relation on $C$ given by $s \equiv t$ iff $T \vdash (s = t)$.  Then the set $A = C/\equiv$ can be given the structure of a model of $T$.  

4
We use two constructions:

a) enlarging $T$ to a maximal consistent set of sentences.
   (i.e. for all sentences $\phi$ we have $\phi \in T$ or $\neg \phi \in T$)

b) adding witnesses. We need to show that if $T'$ is consistent and $T' \vdash (\exists x)\phi$ where $\text{FV}(\phi) = \{x\}$, then $T'$ is consistent where $T'$ is obtained from $T$ by adding a new constant $c_0$ to $T$ and adding a new axiom $\phi[c_0/x].$

Suppose that $T'$ is not consistent; i.e. $T' \vdash L$. Then $T' \vdash \forall x \phi[\phi/x]$ by 3.9. Since $T$ does not mention $c_0$, we can convert this into a deduction of $\forall x \phi[\phi/x]$ from $T$, and hence $T \vdash (\forall x)\phi[\phi/x]$ by (Gen). But $T \vdash (\forall x)\phi[\phi/x]$ by assumption, so $T \vdash L$.

Now, given $T$, we enlarge $T$ by adding constants $c_0$ for each finite $\phi$ such that $T \vdash (\exists x)\phi$, and enlarge $T$ by adding the axioms $\phi[\phi/x]$ for each such $\phi$.

This is still consistent, since a deduction of $L$ from it could involve only finitely many of the new axioms.

Now, starting from $T = T_0$, we construct an infinite sequence $(T_n)$ of theories, such that for even $n$, $T_{n+1}$ is a completion of $T_n$, and for odd $n$, $T_{n+1}$ is obtained by adding witnesses to $T_n$. 
Then $T_0 = \bigcup_{n=0}^{\infty} T_n$ is consistent, complete, and has witnesses.

Now let $C$ be the set of closed terms over the signature of $T_0$, and define $\equiv$ on $C$ by $s \equiv t$ iff $T_0 \vdash (s = t)$. We can make the quotient $M = \Gamma/\Sigma$ into a $\Sigma$-structure by setting $\omega^M([t_1], [t_2], \ldots, [t_n]) = [\omega t_1 t_2 \ldots t_n]$ and $([t_1], [t_2], \ldots, [t_n]) \in [\mathcal{I}]_m$ iff $T_0 \vdash \rho (t_1, \ldots, t_n)$.

Now, an induction over the structure of $\phi$ shows that for any $\phi$ with $FV(\phi) = \{x_1, \ldots, x_n\}$, we have

$([t_1], [t_2], \ldots, [t_n]) \in [\phi]_m$ iff $T_0 \vdash \phi[t_1, t_2, \ldots, t_n/x_1, \ldots, x_n]$

Thus, for a sentence $\phi$, we have $M \models \phi$ iff $T_0 \vdash \phi$.

In particular, since $T_0 \subseteq T_0$, $M$ is a model of $T_0 = T$.

**Remark**

Note that if our original signature $\Sigma_0$ is countable, then so is $\Sigma_{\infty}$ and therefore the model $M$ is countable.

**3.11 Compactness (Compactness Theorem)**

If $T$ is a set of sentences such that every finite subset of $T$ has a model, then $T$ has a model.

**Proof**

Obvious for "is consistent".
3.12 Cookley (Upward Löwenheim–Skolem Theorem)

Suppose that a first order theory $T$ has an infinite model. Then it has models of arbitrarily large cardinality.

Proof

Given a set $I$, extend the signature of $T$ by adding new constants $\{c_i : i \in I\}$, and extend $T$ to $T'$ by adding new axioms $\forall x (c_i = c_j)$ for all $i \neq j$. Given a finite subset $T'' \subseteq T'$, we can construct a model for $T''$ by taking our infinite $T$-model $M$ and assigning distinct values in $M$ to all the constants which occur in members of $T''$, since there are only finitely many of these.

So by 3.11 $T'$ has model $M'$; but this is just a $T$-model equipped with an injection $I \rightarrow M'$, sending $i$ to $(c_i)$. \(\square\)

3.13 Cookley (Downward Löwenheim–Skolem Theorem)

If a theory $T$ over a countable signature has an infinite model, then it has a countably infinite model.

Proof

Apply the argument of 3.12 with a countably infinite model of constants. This produces a consistent theory $T'$ over a countable signature. By the remark after 3.10, $T'$ has a countable model, which must be countably infinite.
A theory \( T \) is called categorical if it has just one model, up to isomorphism.

We can have categorical first order theories with finite models, but not with infinite ones. There are categorical axiomatizations for infinite structures (e.g., Peano's Postulates for \( \mathbb{N} \), 1899):

a) \( 0 \) is a natural number

b) If \( a \) is a natural number, then \( S(a) \) is a natural number

c) \((\forall n) \rightarrow (0 = S(n))\)

d) \((\forall n, m) \rightarrow (S(m) = S(n) \Rightarrow (m = n))\)

e) If \( p(n) \) is a property of natural numbers such that \( p(0) \) holds and \((\forall n) (p(n) \Rightarrow p(s(n)))\), then \((\forall n) p(n)\).

* a) and b) give \( \mathbb{Z} = (\mathbb{N}, \mathbb{N}) \), with \( \mathbb{N} = \{0, S\} \), \( \mathbb{N} = \emptyset \), where \( 0 \) is nullary and \( S \) is unary.

Note that e) involves a quantification over subsets of the intended structure \( \mathbb{N} \). We could replace e) by a scheme of axioms of the form

\[(\forall y_1, \ldots, y_n) ((\forall x)(y \equiv (\forall x)((y \equiv S(x) \land (x \equiv y))) \Rightarrow (\forall y)(y))\]

with one for each formula \( \varphi \) with \( \text{FV}(\varphi) = \{x, y_1, \ldots, y_n\} \)

but this yields e) only for countably many subsets of \( \mathbb{N} \).
3.14 Definition

First-order Peano Arithmetic is the theory over the signature with
\[ \sigma = \{0, s, a, m^3, T\} \] and axioms

\[(\forall x) \neg (sx = 0)\]

\[(\forall x, y) ((sx = sy) \Rightarrow (x = y))\]

\[(\forall x) (axo = x)\]

\[(\forall x, y) (axsy = saxy)\]

\[(\forall x) (mx0 = 0)\]

\[(\forall x, y) (mxsy = axsyx)\]

\[(\forall y, \ldots, y_n) ((\forall [0/x] \land (\forall x)(\forall \Rightarrow 4 [sx/3c]) \Rightarrow (\forall x)4)\]

for each formula \( \varphi \) with \( FV(\varphi) = \{x, y, \ldots, y_n\} \)
3.14 Definition: First Order Peano Arithmetic

This is the theory over the signature with $S = \{ 0, s, +, \cdot \}$ and $T = \emptyset$, and axioms

$(\forall x) \neg (s x = 0)$
$(\forall x, y) ((s x = s y) \Rightarrow (x = y))$
$(\forall x) (x + 0 = x)$
$(\forall x, y) (x \cdot s y = s(x \cdot y))$
$(\forall x) (m \cdot x \cdot 0 = 0)$
$(\forall x, y) (m \cdot x \cdot s y = s(m \cdot x \cdot y))$
$(\forall y_1, \ldots, y_n) ((\forall x \exists \exists \varphi) \land (\forall x) (\varphi \Rightarrow \varphi[s x / x, y_i])) \Rightarrow (\forall x) \varphi$

for each formula $\varphi$ with $\text{FV}(\varphi) = \{ x, y_1, \ldots, y_n \}$

3.15 Definition

a) A first order theory $T$ is called complete if for every sentence $\varphi$ in the language of $T$, we have either $T \models \varphi$ or $T \not\models \varphi$ (but not both).

b) We say that two $\Sigma$-structures $M, N$ are elementarily equivalent if, for all sentences $\varphi$ in $L(\Sigma)$, $M \models \varphi$ if and only if $N \models \varphi$.

3.16 Proposition

A first order theory is complete if and only if all its models are elementarily equivalent.

Proof

Suppose $T$ is complete. Then, if $M, N$ are $T$-models
$M = \emptyset$ implies $T \vdash \forall$, so $T + \forall$, so $N = \emptyset$.

Conversely, suppose that all $T$-models are elementarily equivalent. Then, if $T \vdash \forall$, $T + [\forall]$ is consistent, so has a model, so all $T$-models satisfy $\forall$, so $T + \forall$.

We say a theory $T$ is $K$-categorical, for some cardinal $K$, if any two $T$-models of cardinality $K$ are isomorphic.

For example: the theory of dense total orders without top and bottom elements is (complete and) countably categorical, since any countable model is isomorphic to $\mathbb{Q}$. However, it is not (cardinality $\mathbb{R}$)-categorical, since $\mathbb{R}$ and $\{ x \in \mathbb{R} \mid (x > 0) \lor (x \in \mathbb{Q}) \}$ are both models, and the second contains countable intervals but the first does not.

Another example: the theory of algebraically closed fields of characteristic 0 is complete, but not countably categorical:

- (the algebraic closure of $\mathbb{Q}$) and $\overline{\mathbb{Q}(\pi)}$ are non-isomorphic countable models.

By the L"{o}wenheim–Skolem argument, if a first order theory has finite models of arbitrarily large cardinality, then it must have infinite models.

So there is no first order theory whose models are finite groups.

We can axiomatise infinite groups: take the theory of groups...
and add \( \forall x, \ldots, x_n \exists y \left( \bigwedge (x_i = y) \right) \) for each \( n > 1 \). But there is no finite set of axioms whose models are exactly infinite groups. If there were, say \( A_1, \ldots, A_n \), then all axioms for groups \( 3 \cup \bigwedge i=1^n A_i \) would be a first order axiomatisation of finite groups.

Chapter 4: Zermelo-Fraenkel Set Theory

We want to construct a first-order theory ZF whose intended models are "universes of sets".

Start with a signature containing a binary predicate symbol \( \in \) (and nothing else apart from \( = \)).

The fundamental axiom is **Extensionality** which says:

\[
(\forall x, y) \left( \forall z \left( (z \in x) \iff (z \in y) \right) \right) \Rightarrow (x = y)
\]

We also need the ability to "collect" elements into sets:

Frege (1893) suggested the **comprehension scheme**

\[
(\exists y) (\forall x) \left( (x \in y) \iff (y \in x) \right)
\]

where \( \forall \) is any formula with \( \text{FV}(\forall) = \{ x \} \).

**Russell's Paradox**: consider the formula \( \exists y (\forall x) (x \in y) \iff (x \in x) \)

\[
(\exists y) (\forall x) \left( (x = y) \iff (x \in x) \right)
\]

Is \( y \in y \)?

If \( y \in y \) then \( \forall y \left[ y/x \right] \) i.e. \( 7(y \in y) \)

If \( 7(y \in y) \) then \( 7(\forall y \left[ y/x \right]) \), i.e. \( (y \in y) \)

To avoid this, there are three main possibilities:

1. Russell advocated type theory, which is a many-sorted theory.
a) which the formula $x = y$ is only meaningful if $x$, $y$ have the same sort, and $x \in y$ if $\text{sort}(x) = \text{sort}(y) + 1$ (and so on). But this isn't set theory, since it abandons extensionality.

b) W. Quine 'New Foundations' (NF):

We work with sets, but restrict the Comprehension Scheme to stratifiable formulas $\phi$, i.e. to formulas which could be interpreted in type theory. This seems uncomfortably strong in comparison with

c) Zermelo (1904) proposed replacing Comprehension by the Separation Scheme:

$$(\forall z_1, \ldots, z_k)(\forall y)(\exists x)(\forall z)(x \in y \iff (x \in y_1 \land \phi))$$

where $\phi$ is any formula with $\text{FV}(\phi) = \{x, z_1, \ldots, z_k\}$.
1.1 Definition

Zermelo set theory is the first-order theory with the following axioms:

i) \((\forall x, y)(\forall z) ((z \in x) \iff (z \in y)) \Rightarrow (x = y)\) (Ext)

ii) \((\forall z_1, \ldots, z_n)(\forall y_1)(\exists y_2)(\forall x)((x \in y_2) \iff ((x \in y_1) \land F(y)))\) (Sep)

where \( F \) is any formula with \( \text{FV}(F) = \{x, z_1, \ldots, z_n\}\)

[Note that this axiom implicitly defines an \((n+1)\)-ary operation on our universe \( V \), sending \((z_1, \ldots, z_n, y_1)\) to the unique \( y_2 \) whose existence it asserts. We write \( \{x \in y_1 \mid F\} \) for a term denoting this \( y_2 \); note that \( x \) appears in this term as a bound variable.]

iii) \((\exists x)(\forall y) \neg (y \in x)\) (EmptySet)

[We introduce the constant symbol \( \emptyset \) as a name for this \( x \).]

iv) \((\forall x, y)(\exists z)(\forall w)((w \in z) \iff ((w = x) \lor (w = y)))\) (Pair)

[We write \( \{x, y\} \) for the set \( z \) whose existence is asserted by this, and abbreviate \( \{x, x\} \) for \( \{x\} \).]

v) \((\forall x)(\exists y)(\forall z)(((z \in y) \iff \exists w((z \in w) \land (w \in x)))\) (Union)

[We write \( U x \) for this \( y \), and abbreviate \( U\{x, y\} \) to \( x \cup y \). Note that we don't need an extra axiom to define \( \cap x \) if \( x \neq \emptyset \), and we don't want to define \( \cap \emptyset \). We can also form]

\[ x \setminus y = \{x \in y \} \quad \text{and} \quad x \cap y = \{z \in x \mid z \in y\} \]
vi) \((\forall x)(\exists y)(\forall z)(z \leq y) \iff (\forall w)(w \in z \Rightarrow (w \in x))\)

[We write \(\forall x\) for the \(y\) whose existence is asserted by this, and abbreviate \((\forall w)(w \in z \Rightarrow (w \in x))\) to \((z \leq x)\)]

We can now define the ordered pair \(\langle x, y \rangle\) to be \(\{\{x\}, \{x, y\}\}\) (Kuratowski–Wiener ordered pair). We define

\[
\text{First}(t) = U \cap t \text{ if } t \neq \emptyset \text{ and } \text{First}(\emptyset) = \emptyset, \text{ and} \\
\text{Second}(t) = U(U \setminus U \cap t) \text{ if } U \setminus U \cap t \neq \emptyset \text{ and} \\
\text{Second}(t) = \text{First}(t) \text{ otherwise.} \text{ Then we have} \\
\text{First}(\langle x, y \rangle) = x, \text{ Second}(\langle x, y \rangle) = y, \text{ for all } x, y, \text{ so that} \\
\langle x, y \rangle = \langle z, w \rangle \iff x = z, y = w. \\
\]

We write \(x\) is an ordered pair for \(x = \langle \text{First}(x), \text{Second}(x) \rangle\).

We can form the product set \(x \times y\) as

\[
\{z \in P P U [x, y] | (z \text{ is an ordered pair}) \land (\text{First}(z) \in x) \land (\text{Second}(z) \in y)\}.
\]

We write \(x\) is a function for

\[
(\forall y)((y \in x) \Rightarrow (y \text{ is an ordered pair}) \land (\forall u, v_1, v_2)(u, v_1) \in x \land (u, v_2) \in x \Rightarrow (u = v_1 = v_2)).
\]

We write \(x : y \rightarrow z\) for \((x\text{ is a function}) \land (\forall w)((w \in x) \Rightarrow (\text{First}(w) \in y) \land (\text{Second}(w) \in z))\)

\[
\land (\forall u)(u \in y \Rightarrow (\exists v)(u, v) \in x)).
\]

and we can form the set \(Z^y\) of all functions \(y\) to \(Z\) as

\[
\{x \in P(y \times z) | x : y \rightarrow z\}.
\]

vii) \((\exists x)((\emptyset \in x) \land (\forall y)((y \in x) \Rightarrow (y \cup \{y\} \in x)))

[We abbreviate \(y \cup \{y\}\) by \(y^+\), the successor of \(y\).]
A set with the closure properties specified in (Inf) is called a successor set. By (Sep), if a successor set exists, there is a smallest one, namely
\[ \{ z \in x \mid (y \text{ is a successor set}) \equiv (z \in y) \} \]
where \( x \) is any given successor set. We write \( w \) for the smallest successor set.

Note that \( \varphi^{++...+} = \varnothing \) is a set with \( n \) elements.

We could alternatively have used the sequence \( \varnothing, \varnothing\varnothing, \varnothing\varnothing\varnothing, \ldots \)
or the sequence \( \varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \ldots \) to construct an infinite sequence of sets.

Note that we have a formula which is satisfied by precisely the ordered pairs \( \langle \varnothing, \varnothing \rangle, \langle \varnothing^+, \varnothing\varnothing \rangle, \langle \varnothing^{++}, \varnothing\varnothing\varnothing \rangle, \ldots \)
This formula defines a 'function' (not yet a set), whose 'domain' is a set, but whose 'range' is not yet a set.

A. Fraenkel (1923) introduced a new axiom scheme of Replacement which solves this problem (and also makes the pair set axiom redundant)
Sets and Classes

By a class, we mean an equivalence class of formulas \( \phi \) with one free variable \( x \), where we consider \( \phi \) and \( \psi \) to be equivalent if \((\forall x)(\phi \iff \psi)\) is deducible from the axioms of our set theory. (Equivalently, \( \phi \) and \( \psi \) have the same interpretation in any model.)

We say that \( \phi \) and \( \psi \) are extensionally equivalent if this holds.

We write \( \{x \mid \phi \} \) for the class corresponding to \( \phi \).

We will also denote classes by letters such as \( M \), and write \((t \in M)\) to mean that \( \phi[x/t] \) is deducible, where \( \phi \) is a formula defining \( M \).

We say the class \( \{x \mid \phi \} \) is a set if \((\exists y)(\forall x)(x \in y \iff \phi)\) is deducible. If we can deduce \((\exists y)(\forall x)(x \in y \iff \phi)\),

then we call \( \{x \mid \phi \} \) a proper class.

E.g. the universe \( V \) is a class, corresponding to the formula \((x = x)\) and it is a proper class, as is the Russell class \( \{x \mid \neg(x \in x)\} \).

Similarly, a class of pairs is an extensional-equivalence class of formulas \( \phi \) with two free variables \( x, y \) denoted \( \{<x, y> \mid \phi\} \). We say that \( t \in \{<x, y> \mid \phi\} \) iff \( t \) is an ordered pair \(<u, v>\), and \( \phi[u/v, x, y] \) is deducible. Similarly for classes of \( n \)-tuples \((n \geq 3)\).
A function class is a class of pairs \( F \) for which
\[
(\forall x, y, y') (\langle x, y \rangle \in F \land \langle x, y' \rangle \in F) \Rightarrow (y = y')
\]
The domain of the function class \( F \) is the class \( \{ x \mid (\exists y) (\langle x, y \rangle \in F) \} \).

4.2 Definition (Zermelo–Fraenkel set theory - ZF).

ZF is obtained from Zermelo set theory by adding the axiom scheme of Replacement and the axiom of Foundation, where:

Replacement says:

\[
(\forall z, \ldots, z_n) \left( \exists x \left( \psi(z, \ldots, z_n, z, x) \right) \right) \Rightarrow \left( \exists y \left( \psi(z, \ldots, z_n, x, y) \right) \right)
\]

for any \( \psi \) with free variables \( z, \ldots, z_n \) satisfying

\[
(\forall z, \ldots, z_n, u, v, v') \left( (u \land \psi[v'/v]) \Rightarrow (v = v') \right)
\]

Foundation says:

\[
(\forall x) (\forall (x = \emptyset) \Rightarrow (\exists y) (\forall (y \in x) \land (x \cap y = \emptyset)))
\]

Using Replacement, we can extend the sequence of sets \( \emptyset, \emptyset \emptyset, \emptyset \emptyset \emptyset, \ldots \) (usually denoted \( V_0, V_1, V_2, \ldots \)) by forming the set \( \{ V_n \mid n \in \omega \} \), and then forming \( V_\omega = \bigcup \{ V_n \mid n \in \omega \} \).

Then we can form \( V_{\omega+1} = \mathcal{P}(V_\omega) \), \( V_{\omega+2} = \mathcal{P}(V_{\omega+1}) \), and so on.

And \( V_{\omega+\omega} = \bigcup \{ V_n \mid n \in V_{\omega+1} \} \).

Informally, foundation says that every set is a member of some \( V_n \) (Proved later).
4.3 Definition

By 'x is a transitive set' we mean

\((\forall y, z)((y \in x) \land (z \in y)) \Rightarrow (z \in x)\)

Equivalently, \(x \in P x\) or \(U x \subseteq x\).

Given any set \(x\), there is a smallest transitive set \(TC(x)\) with

\(x \subseteq TC(x)\), namely \(U \{x, Ux, UUx, \ldots\}\). This is transitive

since \(y \in TC(x)\) implies \(y \subseteq U^nx\) for some \(n\), then any member

of \(y\) is in \(U^{n+1}x\).

Conversely, if \(x \subseteq Z\) and \(Z\) is transitive, then \(Ux \subseteq UZ \subseteq Z\),

\(U^2 x \subseteq U^2 Z \subseteq UZ \subseteq Z\) and so on, i.e. \(TC(x) \subseteq Z\).

Note also that \(TC(\{x\})\) is the smallest transitive set \(Z\) with \(x \in Z\).

Exercise: show that \(Vx\) is a transitive set (power set and union of
transitive sets are transitive).

4.4 Proposition

In the presence of the other axioms of ZF, the axiom of Foundation
is equivalent to the axiom-schema of \(\in\)-Induction:

\((\forall x, z)(\forall y)((y \in x) \Rightarrow (y \in z) \Rightarrow \forall y(y \in x) \Rightarrow y \in z)\)

where \(y\) is any formula with \(\text{FV}(y) = \{x, z, \ldots, z\}\).

Proof

\((\Leftarrow)\) Define 'x is a regular set' to mean

\((\exists y)((x \subseteq y) \Rightarrow (\exists z)((z \subseteq y) \land (z \cap y = \emptyset)))\)

Then (Foundation) \(\Rightarrow (\forall x)(x \text{ is a regular set})\). So it suffices
to prove that (\(\forall y)((y \subseteq x) \Rightarrow (y \text{ regular}) \Rightarrow x \text{ regular})\)
But if all members of $x$ are regular and $x \subseteq \mathbb{Z}$, then

either $(x \cap \mathbb{Z} = \emptyset)$, in which case $x$ is a member of $\mathbb{Z}$
disjoint from $\mathbb{Z}$, or $(\exists y)((y \in x) \land (y \in \mathbb{Z}))$, in which case
$y$ is regular since $y \in x$, and so $\mathbb{Z}$ has a member disjoint from $x$.

$(\Rightarrow)$ For the converse, we do the case $n = 0$, i.e., no parameter.

Suppose $(\forall x)(\forall y)((y \in x) \Rightarrow y \notin [\gamma/x]) \Rightarrow \mathcal{Q})$

but $(\exists x) \mathcal{Q}$. Now, for this $x$, $\mathcal{U} = \{y \in TC([x \cup z]) \mid \mathcal{Q} \} \neq \emptyset$
is non-empty since it contains $x$.

So by (Foundation) $(\exists u)(\mathbb{Z} \cup u \land (\mathbb{Z} \cup u) = \mathbb{Z})$

Then $\mathcal{Q}$ holds at all members of $\mathbb{Z}$, since they are members of
$TC([x \cup z]) \cup u$, but $\mathcal{Q} \setminus [z/x]$
4.5 Definition

a) Let \( R \) be a relation class (i.e., an extensional equivalence class of formulas with two free variables). We say that \( R \) is well-founded if it satisfies

\[(\forall x)(\neg (x = \emptyset) \Rightarrow (\exists y)((y \in x) \land (\forall z)((z, y) \in R) \Rightarrow z \in x))]\]

[So the Axiom of foundation says that \( \in \) is well-founded]

b) More generally, let \( M \) be a class. We say that \( R \) is well-founded relative to \( M \) if it satisfies

\[(\forall x)((x \in M) \land \neg (x = \emptyset)) \Rightarrow (\exists y)((y \in x) \land (\forall z)((z, y) \in R) \Rightarrow z \in x))]\]

c) We say \( R \) is local relative to \( M \) if the \( R \)-predecessors in \( M \) of any element of \( M \) form a set, i.e.

\[(\forall x \in M)(\exists y)(\exists z)((z, y) \in R) \Rightarrow ((z \in M) \land (z, x) \in R))]\]

where \((\forall x \in M) \Psi \) means \((\forall x)((x \in M) \Rightarrow \Psi)\) and

\[(\exists x \in M) \Psi \) means \((\exists x)((x \in M) \land \Psi)\)

If \( R \) is local relative to \( M \), then for any subset \( x \) of \( M \) we can form a set \( RC(x) \), which is the subset of \( y \) satisfying

\[(x \in y) \land ((\forall u, v)((\forall e)(v \in y) \land (u, v) \in R) \land (u \in M)) \Rightarrow (u, x)]]\)

Specifically, \( RCM(x) = U \{x, F(x), F(F(x)), \ldots\}\)

where \( F(x) \) denotes the set of \( R \)-predecessors of members of \( x \) which belong to \( M \).

Hence we can prove a proposition.
4.6 Proposition

If \( R \) is well-founded and local relative to \( M \), then we have a principle of \( R \)-induction over \( M \):

\[
(\forall x \in M)(\forall y \in M)(y < y, x) \in R \Rightarrow [\psi[y/x]] \Rightarrow [\psi] \Rightarrow (\forall x \in M)\psi
\]

Proof (Just like 4.4)

Suppose that \( \psi \) satisfies the inductive condition, and suppose we are given \( x \in M \) for which \( \psi \) holds. Form

\[
\{ y \in R^M([x]) \mid \psi[y/x] \} \text{. This is non-empty and } x \in M \text{, so has an } R \text{-minimal member } y, \text{ say.}
\]

Then \( \psi[y/x] \) holds for all \( z \in M \) with \( z < y, x \), but \( \psi[y/x] \) fails.

4.7 Lemma

Suppose that \( R \) is well-founded and local relative to a class \( M \).

Then there is a transitive relation \( \bar{R} \) such that \( R \subseteq \bar{R} \), and \( \bar{R} \) is well-founded and local relative to \( M \).

Proof

We define \( \bar{R} \) by \( (y, x) \in \bar{R} \iff y \in (R^M([x]) \setminus \{x\}) \)

Clearly \( \bar{R} \supseteq R \), from the construction of \( R^M([x]) \).

\( \bar{R} \) is transitive, since \( (y, x) \in \bar{R} \) implies \( R^M([y]) \subseteq R^M([x]) \) and hence from \( (z, y) \in \bar{R} \) we can deduce \( (z, x) \in \bar{R} \).

It is local, since the \( \bar{R} \)-predecessors of \( x \) are the members of the set \( R^M([x]) \setminus \{x\} \).
For well-foundedness, suppose given $\emptyset \neq x \in M$ such that
$x$ has no minimal member. Form the set
\[ y = \{ z \in R_{M}(x) \mid (\exists u \in M)((u, z) \in R) \land (u \in x) \} \]
Then $y \neq \emptyset$ since $x \in y$, and $y$ has no $R$-minimal member.

4.8 Theorem (R-Recursion Theorem)
Suppose that $R$ is well-founded and local relative to a class $M$.
and suppose that given a function class $G$ of two variables which is
defined on $M \times V$ (i.e., $G(x, y)$ is defined whenever $x \in M$).
Then there is a unique function class $F$ of one variable defined
on $M$, and satisfying
\[ (\forall x \in M) \ F(x) = G(\langle x, \{ F(y) \mid (y \in M) \land (y, x) \in R \rangle \rangle) \]
(*)

Proof
For uniqueness, suppose that $F_1$, $F_2$ both satisfy (*). Then we
may prove that $(\forall x \in M)(F_1(x) = F_2(x))$ by $R$-induction over $M$.

For existence, we define the notion of an attempt as follows:
(f is an attempt) means $(f$ is a function) $\land (\text{dom} \ f \subseteq M) \land$
\[ (\text{dom} \ f = R_{M}(\text{dom} \ f)) \land (\forall x)((x \in \text{dom} \ f) \Rightarrow (f(x) = G(\langle x, \{ f(y) \mid (y \in M) \land (y, x) \in R \rangle \rangle)).) \]
Note that if $F_1$, $F_2$ are attempts, then $\text{dom} \ F_1 \cap \text{dom} \ F_2$ is an
$R$-closed subset of $M$, and we can prove
$(\forall x \in \text{dom} \ F_1 \cap \text{dom} \ F_2)(F_1(x) = F_2(x))$, by $R$-induction over this set.
Hence if we define $F$ by
\[ \langle x, y \rangle \in F \iff (\exists f)(f \text{ is an attempt}) \land (f(x) = y) \]
then $F$ is a function class, with $\text{dom } F = M$, and $F$ satisfies (*) for all $x \in \text{dom } F$. So we need to show $\text{dom } F = M$ i.e. $(\forall x \in M)(\exists f)(f \text{ is an attempt}) \land (x \in \text{dom } F))$

Suppose not: given $x \in M$ not in the domain of any attempt, consider $\{y \in RCM(M) \mid x \in \text{dom } y\}$ is not in the domain of any attempt.

This set is non-empty since it contains $x$, so it has an $F$-minimal member $y$, say.

Now $(\forall z \in RCM(M) \setminus \{y\})(\exists ! w)(\exists f)(f \text{ is an attempt}) \land (z, w \in F))$

So the set of all such pairs $<z, w>$ is a function $F_0$ with domain $RCM(M) \setminus \{y\}$, and $F_0$ is an attempt.

Now define $F_1 = F_0 \cup \{<y, z \in RCM(M) \setminus \{y\} | \langle z, x \rangle \in F_0(I))\}$

Then $F_1$ is an attempt with domain $RCM(M) \setminus \{y\}$, contradicting the assumption that $y$ is not in the domain of any attempt. $\times$
4.9 Remark

Given a set $a$ and a well-founded relation $r \subseteq a \times a$, we can prove the existence of functions defined by recursion over $r$ by a simpler argument: in this case, the union of all attempts is a set (by Replacement), and hence in itself an attempt. If this attempt ($f$, say), is not total, then an $r$-minimal member of $a \setminus \text{dom } f$ yields a contradiction.

4.10 Example

The relation $s \subseteq \omega \times \omega$ defined by $\langle x, y \rangle \in s \iff y = x^+$ is well-founded, as is the strict order-relation on $\omega$ (which we may take to be $e \cap (\omega \times \omega)$).

These yield the two versions of mathematical induction:

$$
(\forall x)(x \in \omega) \land (\phi(x)) \land (\forall y \in \omega)((\forall x)(x \leq y \iff (\exists x')(y = x')) \iff (x = \omega))
$$

$$
(\forall x)(x \in \omega) \land (\forall y \in \omega)((\forall x')(\exists x) (y = x') \iff (\exists x)(x \in x')) \iff (x = \omega)
$$

4.11 Definition

A binary relation-class $R$ is said to be extensional on a class $M$ if $(\forall x, y \in M)(\forall z \in M)((\exists x') \in R \iff (\exists y') \in R) \iff (x = y)$.

4.12 Theorem (Mostowski's Theorem)

Let $a$ be a set and $r \subseteq a \times a$ an extensional well-founded relation. Then there exists a unique pair $\langle b, f \rangle$ such that $b$ is a transitive set and $f: \langle a, r \rangle \to \langle b, e \cap b \times b \rangle$ is an isomorphism of sets-with-binary-relation.
Proofs

Uniqueness: suppose given \(<b, f>, <b', f'>\) both satisfying the condition. Then \(y = f' \circ f^{-1}: b \to b'\) is an isomorphism \(<b, e> \to <b', e>\) and we can prove that \((b \circ e \circ b')(g(x) = x)\) by \(e\)-induction over \(b\). Hence \(b = b'\) and \(f = f'\).

Existence: we define \(f\) by \(r\)-recursion over \(a\):
\[ f(x) = \begin{cases} f(a) & \text{if } (y = a) \land (y, x) \in r \end{cases} \]
and we define \(b = \{ f(x) \mid x \in a \}\) (which is a set by replacement).

Clearly \(f\) is injective and \(<x, y> \in r\) implies \(f(x) \in f(y)\).

For the converse of the latter, it suffices to show that \(f\) is injective.

Since we know that if \(f(x) \in f(y)\) then \(f(x) = f(z)\) for some \(z\) with \(<z, y> \in r\), we show that \(f\) is injective by \(r\)-induction:

Suppose \((\forall y \in a)(\langle y, x > \in r) \Rightarrow (\forall z \in a)(f(x) = f(z) \Rightarrow z = y))\)

and suppose \(f(x) = f(z)\).

Then \(\langle y, x > \in r\) \Rightarrow (\exists u)(\langle u, z > \in r) \land (f(y) = f(w))\)

where \((\forall y)(\langle y, x > \in r) \Rightarrow (\langle y, z > \in r)\)) by the induction hypothesis.

Similarly if \(<u, z> \in r\) then we deduce \(f(w) = f(y)\) for some \(y\) with \(<y, x > \in r\), whence \(u = y\), and so \(<u, x > \in r\).

Hence \(x\) and \(z\) have the same \(r\)-predecessor, so by extensionality \(x = z\).
4.13 Definition

We say that a binary relation \( R \) is trichotomous on a class \( M \) if
\[
\forall x, y \in M \left( (x, y) \in R \lor (y, x) \in R \lor (x = y) \right)
\]
Note that if \( R \) is well-founded, then these three possibilities are mutually exclusive, since if two of them hold then \( \{ x, y \} \) has no \( R \)-minimal member.

Also, if \( R \) is well-founded and trichotomous, then it is transitive, since if \( (x, y) \in R \) and \( (y, z) \in R \) but not \( (x, z) \in R \) then \( \{ x, y, z \} \) has no \( R \)-minimal member.

So if \( R \) is well-founded, and trichotomous, then it is a total ordering of \( M \).

Also, a well-founded trichotomous relation is extensional, since if \( x \) and \( y \) have the same \( R \)-predecessors, then we cannot have \( (x, y) \in R \) or \( (y, x) \in R \).

4.14 Corollary

Given a set \( a \) and a well-founded trichotomous relation \( R \) on \( a \), there is a unique pair \( \langle b, F \rangle \) such that \( b \) is transitive and totally ordered by \( E \), and \( F : \langle a, R \rangle \rightarrow \langle b, E \rangle \) is an isomorphism of ordered sets.

Note: \( R \) is well-founded and trichotomous on \( a \\
\Leftrightarrow \) every non-empty \( b \subseteq a \) has an \( R \)-least upper bound.
We use the term well-ordering for a well-founded trichotomy relation.

Chapter 5: Ordinals

5.1 Definition

An ordinal is a transitive set \( x \) which is well-ordered by \( \in \).

Clearly \( \emptyset \) is an ordinal, and so are \( \emptyset^+ \), \( \emptyset^{++} \), and \( \omega \). To prove this, we need some lemmas.

5.2 Lemma

If \( x \) is an ordinal, then \( x^+ = x \cup \{x\} \).

Proof

Transitivity: if \( x \in y \in x^+ \) then either \( y \in x \) or \( y = x \).

So in either case \( x \in x \) and hence \( x \in x^+ \).

Trichotomy: if \( x, y \in x^+ \), then we have one of

\( (x \in x) \land (y = y) \) in which case we have \( (x \in y) \lor (y \in x) \lor (x = y) \),

\( (x \in x) \land (y = y) \) in which case \( x \in y \)

\( (x = x) \land (y \in x) \) \quad y \in x

\( (x = x) \land (y \in x) \) \quad x = y
5.2 Lemma
If α is an ordinal, xo is α⁺ = α ∪ {α}

5.3 Lemma
Every member of an ordinal is an ordinal.
Proof
Let α be an ordinal, x ∈ α. If z ∈ y ∈ x, then y ∈ x and
z ∈ α by transitivity of α. So we have one of (z ∈ x), (z ∈ x)
or (x ∈ z). But if either (z ∈ x) or (x ∈ z), then {x, y, z}
has no ∈-minimal member. So (z ∈ x), hence x is transitive.
If y, z are both members of x, then y, z ∈ α by transitivity, so
we have (y ∈ z) v (y = z) v (z ∈ y).

5.4 Lemma
If α, β are ordinals, then either (α ≤ β) or (β ≤ α)
Proof
Suppose α ≠ β. Then α \ β ≠ ∅, so it has an ∈-least member r,
may. We can show that r = α ∩ β:
- If r ∈ r, then r ∈ r and r ∈ α \ β, so r ∈ α ∩ β
- Conversely, if r ∈ α ∩ β, then r and r are both members of α, so
we have one of (r ∈ r), (r = r), (r ∈ r)
But either (r = r) or (r ∈ r) would imply r ∈ β, since r ∈ β.
So by extensionality we have r = α ∩ β, and in particular
α ∩ β = α.
Similarly, if $\beta \neq \alpha$, then $\alpha \cap \beta \in \beta$. So if neither inclusion holds, $(\alpha \cap \beta) \in (\alpha \cap \beta)$, contradicting Foundation $\Box$

5.5 Corollary

i) For ordinals $\alpha$ and $\beta$, $(\alpha \leq \beta)$ is equivalent to $(\alpha \leq \beta) \lor (\alpha = \beta)$

ii) For any two ordinals $\alpha$ and $\beta$, we have $(\alpha \leq \beta) \lor (\alpha = \beta) \lor (\beta \leq \alpha)$

Proof

i) $(\alpha \leq \beta)$ implies $\alpha \leq \beta$ since $\beta$ is transitive, so $(\subseteq)$ is obvious.

But if $(\alpha \leq \beta)$ and $(\alpha \neq \beta)$, then $\alpha = (\alpha \cap \beta) \in \beta$ by the proof of 5.4.

ii) This is immediate from 5.4 and 5.5).

Writing $\text{On}$ for the class of ordinals, we have shown that

$\alpha \leq \beta \in \text{On}$ implies $\alpha \in \text{On}$ and that $\alpha, \beta \in \text{On}$ implies

$(\alpha \leq \beta) \lor (\alpha = \beta) \lor (\beta \leq \alpha)$

5.6 Corollary (Buchi-Forti Paradox)

$\text{On}$ is a proper class.

Proof

If $\text{On}$ were a set, we would have $\text{On} \in \text{On}$ $\star$

5.7 Lemma

If $\alpha$ is a set, then $\text{U}(\alpha)$ is a subset of $\text{On}$.

Proof

$\text{U}(\alpha)$ is transitive, since it is a union of transitive sets.
member of \( Ux \) satisfying trichotomy by 5.5(ii)

5.8 Theorem

Let \( M \) be any class satisfying
\[
(\forall x)(x \in M \Rightarrow (x^+ \in M)) \quad \text{and} \quad (\forall x)(x \subseteq M \Rightarrow (Ux \in M))
\]

Then \( \text{On} \subseteq M \).

Proof

Suppose not. Then there is an \( \varepsilon \)-least \( \alpha \in \text{On} \setminus M \). Suppose that \( \alpha \) has an \( \varepsilon \)-greatest member \( \beta \) say. Then \( \beta \in M \), and we have
\[
(\forall \gamma)(\gamma \in (\varepsilon \alpha) \Rightarrow ((\varepsilon \beta) \vDash (\gamma = \beta))) \quad \text{i.e.} \quad \alpha = \beta^+
\]

If \( \alpha \) has no greatest member, then \( (\forall \beta \in \alpha)(\exists x \in \alpha)(\beta \in x) \), so \( \alpha = Ux \). But \( \varepsilon \)-minimality of \( \alpha \) means that \( \alpha \subseteq M \).

5.9 Proposition

\[
(\exists \alpha)((\alpha \in \text{On}) \Rightarrow (\alpha \text{ is transitive}) \land (\forall \beta \in \alpha)(\beta \text{ is transitive}))
\]

Proof

(\( \Rightarrow \)) is immediate from 5.3.

(\( \Leftarrow \)) Let \( \text{On}' \) denote the class of transitive sets whose members are all transitive. If \( \text{On}' \neq \text{On} \), let \( \alpha \) be an \( \varepsilon \)-minimal member of \( \text{On}' \setminus \text{On} \). Then all members of \( \alpha \) are in \( \text{On} \), so they satisfy trichotomy by 5.5(ii). \( \alpha \) is transitive by definition, so \( \alpha \in \text{On} \).
Let $\alpha$ be an ordinal. We say that $\alpha$ is a successor if it has an $\varepsilon$-greatest member, or equivalently, if $\alpha = \beta^+$ for some $\beta$. 

If not, we say that $\alpha$ is a limit; this is equivalent to saying $\alpha = \cup \beta$.

(Note that $0$ is a limit; if we want to exclude it, we will refer to non-zero limits.)

Example: $\omega = \cup \omega$ is a non-zero limit ordinal.

We tend to denote (non-zero) limit ordinals by $\lambda$.

5.10 Definition

a) We define a function-class $\text{rk}$ (rank) by $\varepsilon$-recursion over $V$:

$$\text{rk}(x) = \bigcup \{ \text{rk}(y) + 1 | y \in x \}.$$ 

Clearly, $\text{rk}(x) \in \omega$ for all $x$.

b) We define a function-class $x \mapsto V_\alpha$ by $\varepsilon$-recursion over $\omega$:

$$V_\alpha = \bigcup \{ P_{V_\beta} | \beta \in \alpha \}.$$ 

Equivalently:

$V_0 = \emptyset$, $V_{\alpha+} = P_{V_\alpha}$, $V_\lambda = \bigcup \{ V_\beta | \beta < \lambda \}$ for $\lambda$ a non-zero limit.

5.11 Theorem

For $x \in V$ and $\alpha \in \omega$, we have

$$x \in V_\alpha \iff \text{rk}(x) < \alpha \quad (\text{i.e. } \text{rk}(x) \in \alpha)$$

and $x \in V_\alpha \iff \text{rk}(x) \leq \alpha \quad (\text{i.e. } \text{rk}(x) \subseteq \alpha)$

Proof:

The second assertion is the special case $\alpha = \beta^+$ of the first, so we need only prove the first.
Suppose true for all $\beta < \alpha$, and suppose $x \in V\alpha$. Then any
$y \in x$ belongs to $V\beta$ for some $\beta < \alpha$. Hence $(\forall y \in x)(\exists \beta < \alpha)(rk(y) < \beta)$. So

If $\alpha = \beta^+$ is a successor, we have

$(\forall y \in x)(rk(y) < \beta)$

so $(\forall y \in x)(rk(y)^+ \leq \beta)$, so $rk(x) \leq \beta < \alpha$.

If $\alpha$ is a limit, we have $(\forall y \in x)(\exists \beta < \alpha)(rk(y) < \beta)$

so $(\forall y \in x)(\exists \beta < \alpha)(rk(x) = \beta)$ and then

$(\forall y \in x)(rk(y) < \beta)$. So $(\forall y \in x)(rk(y)^+ \leq \beta)$

so $rk(x) \leq \beta^+ < \alpha$. 
Proposition

i) \( x \in V_\alpha \iff \alpha < \text{rk}(x) < \alpha \)

ii) \( x \in V_\alpha \iff \text{rk}(x) \leq \alpha \)

Proof

i) Follows from i) by substituting \( \alpha^+ \) for \( \alpha \).

ii) By \( \varepsilon \)-induction on \( \alpha \):

Assume \( \forall x \forall \beta < \alpha \forall y \in V_\beta \forall z \in V_\beta \implies \text{rk}(x) < \beta \)

and assume that \( x \in V_\alpha \). If \( \alpha = \beta^+ \) is a successor, then \( x \in V_\beta \) i.e. \( \forall y \in x \forall y \in V_\beta \). So by the induction hypothesis \( \forall y \in x \forall y \in V_\beta \). Hence

\[
\text{rk}(x) = \bigcup \{ \text{rk}(y)^+ \mid y \in x \} < \beta^+ = \alpha
\]

If \( \alpha \) is a limit, then since \( V_\alpha = \bigcup \{ V_\beta \mid \beta < \alpha \} \) we have \( x \in V_\beta \) for some \( \beta < \alpha \). So by the induction hypothesis \( \text{rk}(x) < \beta \).

(\( \Rightarrow \)) By \( \varepsilon \)-induction on \( x \).

Suppose \( \forall y \in x \forall \beta < \alpha \forall y \in V_\beta \implies \text{rk}(x) < \alpha \).

Therefore if \( \alpha = \beta^+ \), then \( \text{rk}(y) < \beta \) for all \( y \in x \), so by the induction hypothesis, \( \forall y \in x \forall y \in V_\beta \) i.e. \( x \in \mathcal{P}V_\beta = V_\alpha \).

If \( \alpha \) is a limit, then there exists \( \beta \) with \( \text{rk}(x) < \beta < \alpha \). So \( \text{rk}(y) < \beta \) for all \( y \in x \), so by the induction hypothesis, \( y \in V_\beta \) for all \( y \in x \) i.e. \( x \in \mathcal{P}V_\beta = V_{\beta^+} \subseteq V_\alpha \).
The set-theoretic picture of the universe:

If we are given a model \( V \) of 

\[ \text{ZF} \setminus \text{[Foundation]} \]

we can still define \( \text{On} \) as the class of transitive sets well-ordered by \( \in \),

and we can still define the function class \( x \mapsto V_x \)

In this context, foundation is equivalent to the assertion

\[(\forall x)(\exists x \in \text{On})(x \in V_x)\]

Indeed, in this model, \( V[\{V_x | x \in \text{On}\}] \) is the class of regular sets, as defined in 4.4, i.e. \( \{x \mid (\forall y)(x \in y) \implies (\exists z \in y) y \setminus z = \emptyset \} \)

\textbf{Arithmetic of Ordinals}

We can define \( \alpha + \beta \), \( \alpha \cdot \beta \), either recursively or synthetically.

\textbf{5.12 Definition}

\textbf{a)} The recursive definition say

\[ \alpha + 0 = \alpha \quad , \quad \alpha + \beta^+ = (\alpha + \beta)^+ \]

\[ \alpha + \lambda = V[\{\alpha + \beta \mid \beta < \lambda\} \text{ if } \lambda \text{ is a non-zero limit.} \]

\[ \alpha \cdot 0 = 0 \quad \alpha \cdot \beta^+ = (\alpha \cdot \beta) + \alpha \]

\[ \alpha \cdot \lambda = V[\{\alpha \cdot \beta \mid \beta < \lambda\} \text{ if } \lambda \text{ is a limit.} \]

\textbf{b)} The synthetic definitions say

\[ \alpha + \beta \text{ is the order-type of } \alpha \sqcup \beta = (\alpha \times \{0\}) \cup (\beta \times \{1\}) \]

ordered so that every \( (\beta, 0) \) with \( \beta \in \alpha \) comes before every \( (\beta', 1) \) with \( \beta' \in \beta \) (i.e. with reverse-lexicographic ordering)
\( \alpha \cdot \beta \) is the order-type of the product \( \alpha \times \beta \), with reverse lexicographic ordering (we must check that this is a well ordering).

**5.13 Lemma**

The two definitions in 5.12 coincide.

**Proof**

We need to show that the synthetic definition satisfies all clauses of the recursive definition. 

\( \alpha +_s 0 = \otp(\alpha \times [0] \cup \emptyset) = \alpha \)

\( \alpha +_s (\gamma^+) = \otp(\alpha \times [0] \cup \{r \times \{l, r\} \}) \)

\[ = \otp(\alpha \times [0] \cup \{r \times \{1\}\}^+) = (\alpha + r)^+ \]

by the induction hypothesis.

\( \alpha +_s \beta = \otp(\alpha \times [0] \cup \beta \times \{1\}) \)

\[ = \otp(U[\alpha \times [0] \cup \beta \times \{1\} \mid \beta < \alpha]) \]

and each \( \alpha \times [0] \cup \beta \times \{1\} \) occurs as an initial segment of \( \alpha \times [0] \cup \beta \times \{1\} \), so this is \( U[\alpha + \beta \mid \beta < \alpha] \)

Similarly, (for \( \cdot \)'s synthetic multiplication)

\( \alpha \cdot_0 0 = \otp(\emptyset) = 0 \)

\( \alpha \cdot_+ (\gamma^+) = \otp(\alpha \times \gamma \cup \alpha \times \{r\}) = \otp(\alpha \times r) + \otp(\alpha \times \{r\}) \)

\[ = \alpha \cdot r + \alpha \]

\( \alpha \cdot \beta = \otp(U[\alpha \times \beta \mid \beta < \alpha]) \) and each \( \alpha \times \beta \) is an initial segment of \( \alpha \times \beta \), so this is \( U[\alpha \cdot \beta \mid \beta < \alpha] \)
5.15 Lemma

i) If $\beta < r$ then $\alpha + \beta < \alpha + r$

ii) If $\beta \leq r$, then $\alpha + \beta \leq \alpha + r$

iii) If $\alpha \neq 0$ and $\beta < r$, then $\alpha \cdot \beta < \alpha \cdot r$

iv) If $\alpha \leq \beta$ then $\alpha \cdot r \leq \beta \cdot r$

Proof

We prove i) and iii) synthetically. If $\beta < r$, then $\alpha \cup \beta$ is a proper initial segment of $\alpha \cup r$, so $\text{otp} (\alpha \cup \beta) < \text{otp} (\alpha \cup r)$.

Similarly, if $\alpha \neq 0$, then $\alpha \times \beta$ is a proper initial segment of $\alpha \times r$, so again $\text{otp} (\alpha \times \beta) < \text{otp} (\alpha \times r)$.

i) and iv) are proved by induction on $r$:

$\alpha + 0 = \alpha \leq \beta = \beta + 0$

If $\alpha + r \leq \beta + r$ for all $r < \lambda$, then

$\alpha + \lambda = U \left\{ \alpha + r | r < \lambda \right\} \leq U \left\{ \beta + r | r < \lambda \right\} = \beta + \lambda$.

The proof of iv) is similar.

Note that $1 + \omega = U \left\{ n+1 | n < \omega \right\} = \omega = 0 + \omega$ and $2 \cdot \omega = U \left\{ 2 \cdot n | n < \omega \right\} = \omega = 1 \cdot \omega$.

Note also that $\square + \beta$ and $\square \cdot \beta$ are not continuous at limits i.e. $\omega + 1 \neq U \left\{ n+1 | n < \omega \right\}$ and $\omega \cdot 2 = \omega + \omega \neq U \left\{ n \cdot 2 | n < \omega \right\}$.
5.15 Lemma

a) \(0 + \alpha = \alpha\) for all \(\alpha\)

b) \(\alpha \cdot 0 = 0\) for all \(\alpha\)

c) \(\alpha \cdot 1 = \alpha = 1 \cdot \alpha\) for all \(\alpha\)

d) \(\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma\) for all \(\alpha, \beta, \gamma\)

e) \(\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma\) for all \(\alpha, \beta, \gamma\)

f) \(\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma\) for all \(\alpha, \beta, \gamma\)

Proof (sketch)

All this can be proved synthetically, by establishing an order-isomorphism between two well-ordered sets. For example, e) says that \(\alpha \cdot (\beta + \gamma)\) is order isomorphic to \((\alpha \cdot \beta) + (\alpha \cdot \gamma)\) but the left-hand side has elements of the form \(\langle \alpha', \langle \beta', 0 \rangle \rangle\) or \(\langle \alpha', \langle \gamma', 1 \rangle \rangle\) where \(\alpha' < \alpha, \beta' < \beta, \text{ and } \gamma' < \gamma\). The right-hand side has elements of the form \(\langle \langle \alpha', \beta' \rangle, 0 \rangle, \langle \langle \alpha', \gamma' \rangle \rangle\).

The obvious bijection, \(\langle x, \langle y, z \rangle \rangle \mapsto \langle \langle x, y \rangle, z \rangle\) is an order isomorphism. With the exception of the first equality in c), which follows from a), plus \(\alpha \cdot 0^+ = \alpha \cdot 0 + \alpha\), we can also prove by induction, over \(\alpha\) in cases a) - c), and over \(\gamma\) in cases (d-f).

E.g. the inductive proof of f):

\[\alpha \cdot (\beta \cdot \gamma) = \alpha \cdot 0 = 0 = (\alpha \cdot \beta) \cdot 0\]

\[\alpha \cdot (\beta \cdot \gamma^+) = \alpha \cdot (\beta \cdot \gamma + \beta) = \alpha \cdot (\beta \gamma) + \alpha \cdot \beta \quad \text{by e)}\]

\[= (\alpha \cdot \beta) \cdot \gamma + \alpha \cdot \beta \quad \text{(induction hypothesis)} = (\alpha \cdot \beta) \cdot \gamma^+\]
If \( \lambda \) is a limit, \((\alpha \cdot \beta) \cdot \lambda = U\{ (\alpha \cdot \beta) \cdot r \mid r < \lambda \}\)

\[ = U\{ \alpha \cdot (\beta \cdot r) \mid r < \lambda \} \] by our induction hypothesis.

But \( \beta \cdot \lambda = U\{ \beta \cdot r \mid r < \lambda \} \), and unless \( \beta = 0 \) (in which case both sides are zero), the sequence \( \beta \cdot r, r < \lambda \) is strictly increasing.

So \( \beta \cdot \lambda \) is a limit and hence \( \alpha \cdot (\beta \cdot r) = U\{ \alpha \cdot s \mid s < \beta \cdot r \}\).

But \( s < \beta \cdot \lambda \) implies that \( s < \beta \cdot r \) for some \( r < \lambda \) and hence \( \alpha \cdot s < \alpha \cdot (\beta \cdot r) \) (again, unless \( \alpha = 0 \)).

So \( U\{ \alpha \cdot s \mid s < \beta \cdot r \} = U\{ \alpha \cdot (\beta \cdot r) \mid r < \lambda \} \).

Note that \((\alpha + \beta) \cdot \lambda \neq \alpha \cdot r + \beta \cdot r\) in general.

E.g. \((1+1) \cdot \omega = 2 \cdot \omega = \omega \) but \(1 \cdot \omega + 1 \cdot \omega = \omega + \omega > \omega\).

5.16 Lemma (Division Algorithm for \( \omega \)).

Let \( \alpha, \beta \in \omega \) with \( \beta \neq 0 \). Then there exist unique \( r, s \in \omega \) with \( \alpha = \beta \cdot r + s \) and \( s < \beta \).

Proof.

Since \( \beta \neq 1 \), we have \( \beta \cdot r \geq 1 \cdot r = r \), so there exists \( r' \) such that \( \beta \cdot r' > \alpha \). Consider the least such \( r' \): it cannot be a limit, since \( \beta \cdot r = U\{ \beta \cdot r \mid r < \lambda \} \), so \( r' = s^+ \), where \( \beta \cdot r \leq \alpha < \beta \cdot r^+ = \beta \cdot r + \beta \).

Now, there is a least \( s' \) for which \( \beta \cdot r + s' > \alpha \), and this \( s' \) must be a successor \( s^+ \) (moreover, \( s', s \beta, s \), so \( s < \beta \)).

Then \( \beta \cdot r + s \leq \alpha < \beta \cdot r + s^+ = (\beta \cdot r + s)^+ \), so \( \alpha = \beta \cdot r + s \).
Conversely, if $\alpha = \beta \cdot r + s$ with $s < \beta$, then $\beta \cdot r \leq \alpha < \beta \cdot (r^+)$. 

$r$ is uniquely determined, and $\beta \cdot r + s \leq \alpha < \beta \cdot r + s^+$, so $s$ is uniquely determined.

**Ordinal Exponentiation**

What should $\alpha^\beta$ be? Our first guess is that it should be the order-type of the set $F(\beta, \alpha)$ for all functions $\beta \to \alpha$.

This set can be totally ordered by lexicographic ordering (i.e., $f < g$ if $f(r) < g(r)$ for the least $r$ such that $f(r) \neq g(r)$), but this is not a well ordering if $\beta$ is infinite and $\alpha > 2$.

Since if we define $f_i : \omega \to 2$ by $f_i(0) = 1$ if $i = j$, 0 otherwise, then the sequence $(f_i)$ is strictly decreasing, no $f_i : i \in \omega$ has a least member.

To get around this, we use reverse lexicographic ordering and cut down to the set $F_0(\beta, \alpha) \subseteq F(\beta, \alpha)$ of functions of finite support, i.e., those $f$ such that $f(r) = 0$ for all but finitely many $r$.

**5.17 Lemma**

$F_0(\beta, \alpha)$ is well ordered by reverse lexicographic ordering.

**Proof (by induction on $\beta$).**

Let $S$ be a non-empty subset of $F_0(\beta, \alpha)$. Pick $F \in S$; if $F$ is identically zero then it is the least member of $S$. 

3
Otherwise, there is a greatest $\beta' < \beta$ such that $f(\beta') \neq 0$.

Now let $S_1 = \{ g \in S \mid g(\tau) = 0 \text{ for all } \tau > \beta' \}$

Then every member of $S$ precedes every member of $S \setminus S_1$, so we need to find the least element of $S_1$. Let $\alpha'$ be the least element of $\{ g(\beta') \mid g \in S \}$ and let $S_2 = \{ g \in S, g(\beta') = \alpha' \}$

Again, every member of $S_2$ precedes every member of $S \setminus S_2$ (so we need to find the least element of $S_2$).

But $S_2$ is order-isomorphic to $\{ g \beta' \mid g \in S_2 \}$ which is a non-empty subset of $F_0(\beta', \alpha)$ and hence has a least element by our induction hypothesis.

We define $\alpha^\beta$ to be the order-type of $(F_0(\beta', \alpha), \preceq)$.

5.18 Lemma

Ordinal exponentiation satisfies the recursive definition

$\alpha^0 = 1$, $\alpha^{\beta+} = \alpha^\beta \cdot \alpha$

$\alpha^\lambda = \bigcup \{ \alpha^\beta \mid \beta < \lambda \}$ if $\lambda$ is a non-zero limit.

Proof

$F_0(\varnothing, \alpha) = \{ \varnothing \}$ which is the ordinal $1$.

We have a bijection $F_0(\beta^+, \alpha) \to F_0(\beta, \alpha) \times \alpha$ given by $f \mapsto \langle f|_\beta, f(\beta) \rangle$ which is an order isomorphism when both sides are ordered by reverse lexicographic order.
If $\kappa$ is a limit, $F_\kappa(\tau, \alpha)$ is the union of the subsets $G(\beta, \alpha) \mid \beta < \kappa$
where $G(\beta, \alpha) = \{ f \in F_\beta(\tau, \alpha) \mid f(r) = 0 \ \text{for all } r \geq \beta \}$

But $G(\beta, \alpha)$ is order-isomorphic to $F_\beta(\beta, \alpha)$ and the
$G(\beta, \alpha)$ are initial segments of $F_\beta(\tau, \alpha)$.

So the order-isomorphisms $G(\beta, \alpha) \cong F_\beta(\beta, \alpha) \cong \alpha^\beta$
fit together to yield an order-isomorphism $F_\kappa(\tau, \alpha) \cong \bigcup \{ \alpha^\beta \mid \beta < \kappa \}$. 

5.19 Lemma

a) \( \alpha^{(\beta + r)} = \alpha^\beta \cdot \alpha^r \) for all \( \alpha, \beta, r \in \omega_1 \)

b) \( \alpha^{(\beta \cdot r)} = (\alpha^\beta)^r \)

Proof

We prove a) synthetically: there is a bijection

\( F : (\beta + r, \alpha) \to F_0 (\beta, \alpha) \times F_0 (r, \alpha) \)

sending \( F \to <F_0, F, > \) where \( F_0 (\beta') = F(<\beta', 0>) \)

and \( F_0 (r') = F(<r', 1>) \),

which is easily seen to be an order isomorphism when both sides are

ordered by reverse lexicographic ordering. We prove b) by induction on \( r \):

\( \alpha^{(\beta \cdot 0)} = \alpha^0 = 1 = (\alpha^\beta)^0 \)

\( \alpha^{(\beta \cdot r + 1)} = \alpha^{(\beta \cdot r + \beta)} = \alpha^{(\beta \cdot r)} \cdot \alpha^\beta \text{ by a) } \)

\( = (\alpha^\beta)^r \cdot \alpha^\beta \text{ by the induction hypothesis. } \)

\( = (\alpha^\beta)^{r+1} \)

If \( r = \lambda \) is a limit, we deal first with the cases \( \alpha = 0 \) or 1, or

\( \beta = 0 \). Otherwise, \( \alpha^\alpha \) and \( (\alpha^\beta)^\alpha \) are strictly increasing

functions, since we have \( \alpha \geq 2 \) and \( \alpha^\beta \geq 2 \).

So \( \alpha^{(\beta \cdot \lambda)} = \cup \{ \alpha^{(\beta \cdot r)} \mid r < \lambda \} \) since \( \beta \cdot \lambda \) is non-terminating,

\( = \cup \{ \alpha^{(\beta \cdot r)} \mid r < \lambda \} \) since \( \{ \beta \cdot r \mid r < \lambda \} \) is confluent in \( \beta \cdot \lambda \),

\( = U \{ (\alpha^\beta)^r \mid r < \lambda \} \) by the induction hypothesis,

\( = (\alpha^\beta)^\lambda \)
We don't have \((\alpha \cdot \beta)^\tau = \alpha^\tau \cdot \beta^\tau\) in general.

e.g. \((\omega \cdot 2)^2 = \omega \cdot (2 \cdot \omega) \cdot 2 = \omega \cdot \omega \cdot 2 = \omega^2 \cdot 2\)

but \(\omega^2 \cdot 2^2 = \omega^2 \cdot 4 > \omega^2 \cdot 2\)

But it can also fail even when \(\alpha \cdot \beta = \beta \cdot \alpha\) if \(\tau\) is infinite:

e.g. \((2 \cdot 2)^\omega = 4^\omega = \bigcup \{ 4^n \mid n < \omega \} = \omega\)

but \(2^\omega \cdot 2^\omega = \omega \cdot \omega = \omega^2 > \omega\)

Chapter 6: Choice and Well-Ordering

We've seen that \(\omega < \omega + 1 < \omega + 2 < \ldots < \omega + \omega = \omega \cdot 2\)

\(\omega < \omega \cdot 2 < \omega \cdot 3 < \ldots < \omega \cdot \omega = \omega^2\)

\(\omega < \omega^2 < \omega^3 < \ldots < \omega^\omega\)

\(\omega < \omega^\omega < \omega^{\omega^\omega} < \ldots < \varepsilon_0 = \text{the least } \alpha \text{ such that } \alpha = \omega^\alpha\)

\(\varepsilon_0 < \varepsilon_1 < \ldots < \varepsilon_{\omega} < \ldots < \varepsilon_\omega = \text{the least } \alpha \text{ such that } \alpha = \varepsilon_\alpha\)

But these are all countable! Do there exist any uncountable ordinals?

Yes: Consider the set of all countable ordinals. More generally,

6.1 Lemma (Hartogs' Lemma)

For any set \(a\), there exists an ordinal \(\gamma(a)\) for which there is no injection \(\gamma(a) \to a\).

Proof:

Consider the set \(b = P(\alpha \times a)\) of all well-orderings of subsets of \(a\). From Mostowski's Theorem, we have a function class \(F\) assigning to each well-ordered set its order type.
Logic and Set Theory

So \( L \{ \{ x \mid x \in \{ y \mid y \in \{ \emptyset, \{ y \} \} \} \} \} \) is a set by Replacement.

Clearly, this set is a transitive subset of \( \text{On} \), and hence an ordinal \( \gamma(a) \).

But if we had an injection \( f : \gamma(a) \to a \), then \( \{ f(\beta) \mid \beta < \gamma(a) \} \) would be a well ordered subset of \( a \) with order type \( \gamma(a) \).

6.2 Corollary

We get a 'new' proof of the Bourbaki-Witt theorem 1.12:

Suppose, given a chain complete poset \( P \), an inflationary map \( \hat{f} : P \to P \), and an element \( x \in P \), we wish to find a fixed point \( \geq x \).

We define a function class \( G : \text{On} \to P \) by recursion:

\[
G(0) = x
\]

\[
G(\alpha +) = \hat{f}(G(\alpha))
\]

\[
G(\alpha) = \bigvee \{ G(\beta) \mid \beta < \alpha \} \text{ for } \alpha \text{ a non-zero limit.}
\]

(This works since \( \alpha \leq \beta \Rightarrow G(\alpha) \leq G(\beta) \), and so \( \{ G(\alpha) \mid \alpha < \beta \} \) is a chain.

The restriction of \( G \) to \( \gamma(P) \) isn't injective, i.e., \( \exists x < \beta < \gamma(P) \)

with \( G(x) = G(\beta) \).

But now \( \alpha \leq \beta \), so \( G(\alpha +) = \hat{f}(G(\alpha)) \leq G(\beta) = G(\alpha) \) and hence

\( G(\alpha) \) is a fixed point of \( \hat{f} \), lying above \( x = G(0) \). \( \square \)

This proof requires the 'full strength' of ZF set theory, whereas the proof in chapter I works in Zermelo set theory (i.e., without Replacement).

Question: Does every set actually have a well-ordering?

Answer: (Zermelo, 1904) Yes, iff we assume the axiom of choice.
6.3 Definition

Given a set $a$, we write $P^+a$ for $\mathcal{P}(a \setminus \emptyset)$

By a choice function for $a$, we mean a function $f: P^+a \to a$

such that $f(b) \in b$ for all $b \in P^+a$.

Note that, if $\{a_i; i \in I\}$ is a set of non-empty sets, then it is a subset

of $P^+U\{a_i; i \in I\}$, so a choice function for $U\{a_i; i \in I\}$

will yield a choice function $I \to U\{a_i; i \in I\}$ as considered in Chapter I.

Thus, we can formulate the axiom of choice $\forall \alpha \exists \beta (\alpha \subseteq \beta)$ (in a choice form).

6.4 Theorem (Zermelo's Theorem)

In a model of ZF set theory, the sets which have choice functions are

exactly the well-orderable sets.

Proof

If $<$ is a well-ordering of $a$, then we get a choice function

g: $P^+a \to a$ by setting $g(b) = \text{the } \leq \text{-least member of } b$.

Conversely, suppose that we are given $g: P^+a \to a$ (assume $a \neq \emptyset$).

We define a function class $F: \alpha \to a$ by recursion:

$F(\alpha) = \begin{cases} g(\alpha) \setminus \{F(\beta); \beta < \alpha \} & \text{if} \exists F(\beta); \beta < \alpha \end{cases}$

By $b.1$, $F(\alpha)$ is not injective. So there must be some $\alpha < r(\alpha)$

such that $F(\alpha)$ is defined by the second clause, i.e. $\exists F(\beta); \beta < \alpha \subseteq \alpha$

Let $\alpha$ be the least ordinal for which this happens: then $F|\alpha$ is injective.

Hence this is a bijection $\alpha \to a$, so $\{\langle F(\alpha), F(\beta) \rangle; \beta < \alpha \}$

is a well-ordering of $a$. 
6.5 Remarks

a) The proof we gave for 6.4 is not Zermelo’s original. His proof doesn’t use replacement (and is similar to the proof of Bourbaki-Witt in 1.12).

b) We can prove the Well-Ordering Theorem directly from Zorn’s Lemma:

Given a set \( a \), consider the set \( P = \{ \langle b, \prec \rangle \mid (b = a) \wedge (\text{ordering of } b) \} \).

We order \( P \) by setting \( \langle b_1, \prec_1 \rangle \leq \langle b_2, \prec_2 \rangle \iff b_1 \subseteq b_2 \), \( b_1 \) is a \( \prec_2 \)-initial segment of \( b_2 \), and \( \prec_1 = \prec_2 \cap (b_1 \times b_1) \).

Then \( P \) is chain-complete: given a chain \( \{ \langle b_i, \prec_i \rangle \mid i \in I \} \), the pair \( \langle U \{ b_i \mid i \in I \}, U \{ \prec_i \mid i \in I \} \rangle \) is a member of \( P \) and a least upper bound for the chain.

\( P \neq \emptyset \), as \( \langle \phi, \phi \rangle \in P \), so it has a maximal element \( \langle b_0, \prec_0 \rangle \) say. If \( b_0 \neq a \), pick \( x \in a \setminus b_0 \) and set \( b_1 = b_0 \cup \{ x \}, \prec_1 = \prec_0 \cup \{ \langle y, x \rangle \mid y \in b_0 \} \).

Then \( \langle b_0, \prec_0 \rangle < \langle b_1, \prec_1 \rangle \).

So \( \prec_0 \) is a well-ordering of \( a \).

c) Results proved using Zorn’s Lemma can also be proved using the Well-Ordering Theorem.

e.g. Hamel’s Theorem: given a vector space \( V \), well order (the underlying set of) \( V \) as \( \{ \langle \beta, \alpha \rangle \mid \beta < \alpha \} \). Now define a sequence of subsets \( S_\beta \subseteq V \), \( \beta \leq \alpha^+ \), by \( S_0 = \emptyset \), and if \( x \beta \in \langle S_\beta \rangle \), set \( S_{\beta^+} = S_\beta \), otherwise \( S_{\beta^+} = S_\beta \cup \{ x \beta \} \).
If $\alpha$ is a limit ordinal, $S_\alpha = \bigcup \{ S_\beta \mid \beta < \alpha \}$

By induction, $S_\beta$ is linearly independent for all $\beta$, and since $x_\beta \in <S_\beta>$ for all $\beta$, we have $<S_\alpha>$ = $V$ i.e. $S_\alpha$ is a basis.

Note also that we can prove the uncountable case of the completeness theorem for propositional logic in the same style in which we did the countable case (cf. 2.18(b)), by well-ordering the set $L(P)$ of a compound propositions.

6.6 Definition

We say that an ordinal is initial if there is no bijection $\alpha \to \beta$ with $\beta < \alpha$. Clearly, every $n \in \mathbb{N}$ is initial, as is $\omega$ itself, and $\omega_1 = \tau(\omega)$, the set of all countable ordinals.

More generally, we define a function class $\alpha \to \omega_\alpha$, from $\omega_\alpha$ to $\omega_\alpha$, by recursion:

$\omega_0 = \omega$, $\omega_{\alpha + 1} = \tau(\omega_\alpha)$

$\omega_\alpha = \bigcup \{ \omega_\xi \mid \xi < \alpha \}$ if $\alpha$ is a non-zero limit ordinal.

6.7 Lemma

The ordinals of the form $\omega_\alpha$ are exactly the infinite initial ordinals.

Proof

We prove that every $\omega_\alpha$ is initial by induction on $\alpha$:

This is clear for $\alpha = 0$.

For a successor $\alpha^+$, $\omega_{\alpha^+} = \tau(\omega_\alpha)$ does not inject into any...
any $\beta < \omega \omega^*$, since every such $\beta$ injects into $\omega \omega$.

For a limit $\alpha$, $\omega \alpha$ is the limit of a strictly increasing sequence.

So if we had an injection $\omega \alpha \to \beta$ for some $\beta < \omega \alpha$,
we would have an injection $\omega \beta \to \omega \alpha$ for some $\alpha < \beta$,
and hence an injection $\omega \omega^* \to \omega \alpha$.

Now let $\beta$ be an infinite initial ordinal. We have $\omega \alpha \geq \alpha$ for all $\alpha$,
so there is a least $\alpha$ with $\omega \alpha > \beta$. This $\alpha$ cannot be a non-zero limit,
and cannot be zero since $\beta$ is infinite.

So $\alpha$ is a successor $\omega^+$, and then we have $\omega \omega \leq \beta < \omega \omega^*$.

So we have injection $\omega \omega \to \beta$, $\beta \to \omega \omega$, hence by

Cantor-Bernstein (1.11) there is a bijection $\omega \omega \to \beta$. Since $\beta$ is
initial, this implies that $\beta = \omega \omega$

Informally, a cardinal is an equivalence class of sets under the
relation of equipollence (i.e. that of being in bijective correspondence).

But every equivalence class except $\emptyset$ is a proper class, so
we need to find a way of representing this by sets.

If we assume the Axiom of Choice, then every equivalence class
contains an ordinal, and hence contains a unique initial ordinal,
which we can take as a canonical representative of the class
(i.e. we define $\text{card}(x)$ to be the unique initial ordinal in
bijection with $x$.)
Without AC, we can't find a unique representative for each class, but we can find a representative subset of it, as follows:

Define the essential rank of a set \( x \) as

\[
\bigwedge \{ \beta \leq \text{rank}(x) \mid (\exists y)((x, y) = \beta) \land (\exists \text{ a bijection } y \rightarrow x)\}
\]

and then define \( \text{card } x = \{ y \in V_{\text{essential rank }x} \mid (\exists \text{ a bijection } x \rightarrow y)\} \)

From now on, all we assume about the function class \( \text{card } \) is

\[(x, y) \left( \text{card}(x) = \text{card}(y) \right) \iff (\exists \text{ a bijection } x \rightarrow y)\]

We need a new notation for \( \text{card } \): following Cantor, we denote it by \( \aleph_x \) (aleph-alpha).

Given cardinals \( m, n \), we write \( m \leq n \) to mean that there exists an injection \( x \rightarrow y \) where \( \text{card}(x) = m \), \( \text{card}(y) = n \).

Cantor-Bernstein implies that this is a partial ordering of the class of cardinals.

The assertion that it is a total ordering is equivalent to the Axiom of Choice.

We define binary operations \( +, \cdot \) and \( \cap \) on the class of cardinals as follows:

\[ m + n = \text{card}(x \uplus y) \text{ where } \text{card}(x) = m, \text{card}(y) = n, \]

\[ m \cdot n = \text{card}(x \times y) \text{ similarly, and} \]

\[ m^\circ = \text{card}(x^y), \text{ where } x^y \text{ denotes the set of all functions } y \rightarrow x \]
6.8 Lemma

Let \( m, m', n, n', p \) be cardinals.

a) If \( m' \leq m \) and \( n' \leq n \), then \( m' + n' \leq m + n \) and \( m' \cdot n' \leq m \cdot n \).

b) If \( m' \leq m \), then \( m^n \leq m' \cdot n \); and if \( 0 \neq n' \leq n \), then \( m' \leq m^n \).

c) \( m + (n + p) = (m + n) + p \) and \( m \cdot (n \cdot p) = (m \cdot n) \cdot p \).

d) \( m + n = n + m \) and \( m \cdot n = n \cdot m \).

e) \( m \cdot (n + p) = m \cdot n + m \cdot p \).

f) \( m^{(n + p)} = m^n \cdot m^p \), \( m^{n \cdot p} = (m^n)^p \), \((m \cdot n)^p = m^p \cdot n^p \).

Proof (Selected Items)

Let \( a, b, c \) be sets with cardinalities \( m, n, p \) respectively.

b) Given an injection \( f: a' \to a \), the map \( g \circ f \circ g \) is an injection \( a'^b \to a^b \).

Given an injection \( h: b' \to b \) with \( b' \neq \emptyset \), pick \( x_0 \in b' \) and define \( k: b \to b' \) by \( k(y) = \begin{cases} x_0 & \text{if } h(x) = y \\ y & \text{if } h(x) \in \text{im}(h) \end{cases} \).

Then \( k \) is injective, and the map \( g \circ f \circ g \circ k \) is an injection \( a'^b \to a^b \).

f) We want a bijection \( a^{(b \times c)} \to a^{b \times a^c} \); take the mapping \( f \mapsto (f \circ i, f \circ j) \), where \( i: b \to b \upharpoonright c \) is the map \( y \mapsto \langle y, 0 \rangle \), and \( j: c \to b \upharpoonright c \), \( z \mapsto \langle z, 1 \rangle \).

Similarly, we want a bijection \( a^{b \times a^c} \to (a^b)^c \); take the mapping \( f \mapsto f^\wedge \), where \( f^\wedge(z)(y) = f(\langle y, z \rangle) \).

We also want a bijection \( (a \times b)^c \to a^c \times b^c \); take the mapping \( f \mapsto \langle p \circ f, q \circ f \rangle \), where \( p(\langle x, y \rangle) = x \) and \( q(\langle x, y \rangle) = y \).
Proposition

For any ordinal \( \alpha \), \( S_\alpha \cdot S_\alpha = S_\alpha \)

Proof (by induction on \( \alpha \))

We construct a well-ordering of \( \omega \cdot \omega \), which has order-type \( \omega \cdot \omega \)
We order pairs \( \langle r, s \rangle \in \omega \cdot \omega \) by setting

\[ \langle r, s \rangle < \langle r', s' \rangle \]

\[ \Leftrightarrow \text{either } \max \{r, s\} < \max \{r', s'\} \]

or \( \max \{r, s\} = \max \{r', s'\} \text{ and } r < r' \)

or \( r = \max \{r, s\} = \max \{r', s'\} = r' \text{ and } s < s' \)

To show that this is a well-ordering, let \( S \) be a non-empty subset of \( \omega \cdot \omega \).

Let \( \beta = \min \{\max \{r, s\} \mid \langle r, s \rangle \in S\} \) and let

\[ S_\beta = \{\langle r, s \rangle \in S \mid \max \{r, s\} = \beta\} \]

Let \( r_0 = \min \{r \mid \langle r, s \rangle \in S_\beta\} \); if \( r_0 < \beta \) then

\[ \langle r_0, \beta \rangle \text{ is the least element of } S_\beta \text{, and hence of } S. \]

If \( r_0 = \beta \), set \( s_0 = \min \{s \mid \langle \beta, s \rangle \in S_\beta\} \); then

\[ \langle \beta, s_0 \rangle \text{ is the least element of } S_\beta \text{, and hence of } S. \]

Now any proper initial segment \( \langle r, s \rangle \) of \( \omega \cdot \omega \) is contained in \( \beta \cdot \beta \), where \( \beta = (\max \{r, s\})^+ \). Then either \( \beta < \omega \), in which case \( \text{card}(\beta \cdot \beta) < S_{r_0} \leq S_\alpha \)
or \( \text{card}(\beta) = S_{\alpha'} \), for some \( \alpha' < \alpha \), in which case
card(\(\beta \times \beta\)) = \(\aleph_\alpha\) < \(\aleph_\alpha\) by the induction hypothesis.

So every proper initial segment of \(\text{WA} \times \text{WA}\) has order type < \(\text{WA}\) hence the order type of \(\text{WA} \times \text{WA}\) is \(\leq \text{WA}\). But \(\text{WA}\) injects into \(\text{WA} \times \text{WA}\) (say by \(\beta \mapsto (\beta, \alpha)\)) so the order type cannot be < \(\text{WA}\). Hence card \((\text{WA} \times \text{WA})\) = \(\aleph_\alpha\)

6.8 (g)

If \(m \geq 2\) and \(n \geq 2\), then \(m + n \leq m \cdot n\)

Proof

Let \(x_0, x_1\) be distinct elements of \(\alpha\), and \(y_0, y_1\) distinct elements of \(\beta\). Define \(f: \alpha \cup \beta \rightarrow \alpha \times \beta\) by

\[
f(\langle x_0, y_0 \rangle) = \langle x, y_0 \rangle \text{ for all } x \in \alpha
\]

\[
f(\langle y_0, y_1 \rangle) = \langle x_0, y \rangle \text{ if } y \in \beta \setminus \{y_0\}
\]

\[
f(\langle y_1, y_1 \rangle) = \langle x_1, y_1 \rangle \text{ if } y = y_0.
\]

It is easy to check that \(f\) is injective.

6.10 Corollary

For all \(\alpha\) and \(\beta\), we have

\(\aleph_\alpha + \aleph_\beta = \aleph_\alpha \cdot \aleph_\beta = \aleph_{\text{max} \{\alpha, \beta\}}\)

Proof

Assume \(\alpha \leq \beta\). Then \(\aleph_\beta = 0 + \aleph_\beta \leq \aleph_\alpha + \aleph_\beta \leq \aleph_\alpha \cdot \aleph_\beta \leq \aleph_\beta \cdot \aleph_\beta = \aleph_\beta\) by 6.9

By Cantor-Bernstein, all these must be equal.
Hence, if the Axiom of Choice holds, then addition and multiplication of infinite cardinals is commutative. Conversely, if $m \cdot m = m$ for all infinite $m$, then the Axiom of Choice holds. To prove this, we need:

6.11 Lemma

If $m + n = m \cdot n$, then either $n \leq m$, or there exists a surjection from a set of cardinality $n$ to one of cardinality $m$.

Proof

Let $a, b$ be sets of cardinality $m, n$ respectively. Consider the composite $b \xrightarrow{j} a \cup b \xrightarrow{p} a$ where $f$ is our given bijection, $j$ is the inclusion and $p$ is the projection. If this is surjective, we are done. Otherwise, pick $x \in a \setminus \text{im}(pFj)$. Now the restriction of $F^{-1}$ to $\{x\} \times b$ takes values in $a \times \{a\} = a \cup b$, so it defines an injection $b \to a$.

6.12 Proposition

Suppose $m \cdot m = m$ for all $m \geq M_0$. Then the Axiom of Choice holds.

Proof

We show that any set $a$ can be well-ordered. If $a$ injects into $\omega$, this is easy; if not then $\text{card}(a) > \omega$, so card $(a \cup \text{im}(pFj)) \geq \omega$. Hence, writing $n$ for card $b$, we have $(m+n) \cdot (m+n) = m+n$.

But $m \cdot n = (m+0) \cdot (0+n) \leq m+n$, so $m \cdot n \leq M+n$.

We also have $m+n \leq m \cdot n$ by 6.8(9).

So by Cantor-Bernstein we have $m+n = m \cdot n$, but $n \neq m$. 

4
Since \( R(a) \) doesn't inject into \( a \).

So we have a bijection \( g : R(a) \to a \). Hence we can well-order \( a \) by setting

\[
\forall x < y \iff (\text{the least } \beta \text{ with } g(\beta) = x) < (\text{the least } \beta \text{ with } g(\beta) = y).
\]
6.13 Lemma
For any cardinal \( m \), we have \( m < 2^m \).

Proof
For any set \( a \), the map \( x \mapsto \{ x \} \) is an injection \( a \to \mathcal{P}a \).
But there is no injection \( a \to \mathcal{P}a \), by Cantor's Diagonal Argument.

6.14 Lemma
If \( \beta \leq \alpha^+ \), then \( 2^{\aleph_\beta} = 2^{\aleph_\alpha} \) (assuming the Axiom of Choice).

Proof
We have \( 2^{\aleph_\beta} \leq 2^{\aleph_\alpha} \leq 2^{\aleph_\alpha^+} \leq \left( 2^{\aleph_\alpha^+} \right)^{\aleph_\alpha} = 2^{\aleph_\alpha \cdot \aleph_\alpha} = 2^{\aleph_\alpha} \)

\[
\text{since } 2 \leq \aleph_\beta \text{ since } \beta \leq \alpha^+ \text{ since } 2^{\aleph_\alpha} > \aleph_\alpha
\]
So by Cantor-Bernein all the \( < \)'s are \( \leq \)’s.

Hence, by assuming the Axiom of Choice, we are principally interested in the function class \( F : \mathbb{C} \to \mathbb{C} \) defined by \( k_F(x) = 2^{\aleph_\alpha} \).

Cantor conjectured (the Continuum Hypothesis) that \( F(\alpha) = 1 \).

The Generalized Continuum Hypothesis asserts that \( F(\alpha) = \alpha^+ \) for all \( \alpha \), or (without assuming the Axiom of Choice) that \( m < 2^m \)
(i.e. \( m \) is covered by \( 2^m \)) for all infinite \( m \). In fact, the latter statement implies the Axiom of Choice.

Gödel (1930) showed that if \( ZF \) is consistent, so is \( ZF + GCH + AC \).

Cohen (1964) showed that if \( ZF \) is consistent, then so is \( ZF + AC + \neg CH \).
Given a family of cardinals, \( \{ m_i : i \in I \} \), we define \( \sum_{i \in I} m_i \) to be the cardinality of \( \prod_{i \in I} a_i = U \{ a_i \times \{ i \} : i \in I \} \), where each \( a_i \) has cardinality \( m_i \), and we define \( \prod_{i \in I} m_i \) to be the cardinality of \( \prod_{i \in I} a_i = \{ f : I \to U \{ a_i : i \in I \} \mid f(i) \in a_i \text{ for all } i \in I \} \). Note that it requires the Axiom of Choice to show that operations on cardinals are well-defined.

6.15 Lemma (König's Lemma)

If \( m_i < n_i \) for all \( i \in I \), then \( \sum_{i \in I} m_i < \prod_{i \in I} n_i \).

Proof

\( \leq \) is proved very similarly to 6.8(a). To show \( \geq \), suppose \( a_i, b_i \) are sets of cardinalities \( m_i, n_i \) respectively, and consider a map \( \prod_{i \in I} a_i \to \prod_{i \in I} b_i \). For each \( i \), consider the composite \( a_i \xrightarrow{\phi_i} \prod_{i \in I} a_i \xrightarrow{\pi_i} \prod_{i \in I} b_i \). This can't be surjective since card \( a_i \) < card \( b_i \), so we can find \( y_i \in b_i \) not in its image. But then the map \( i \mapsto y_i \) is an element of \( \prod_{i \in I} b_i \) not in the image of \( \phi_i \) for any \( i \), and hence not in the image of \( \phi \).

6.16 Corollary

Let \( A \) be a limit ordinal with \( \text{cf} (A) = \omega \), i.e. such that \( A = U \{ \alpha_i : i \in \omega \} \) for some increasing sequence of ordinals \( \alpha_i < A \). Then \( 2^{\aleph_0} \neq \aleph_A \).
Proof

Let \( \alpha_0 = \omega \), and, for \( i > 1 \), set \( \alpha_i = \omega \backslash \omega_{\alpha_{i-1}} \). Then \( \text{card}(\omega_\alpha) = \kappa_\alpha \) for each \( \alpha \). Set \( m_\alpha = \kappa_\alpha \) and \( n_\alpha = \kappa_\alpha \) in König's Lemma; \( m_\alpha < n_\alpha \) for all \( \alpha \); \( \sum_{\alpha \in \omega} m_\alpha < \prod_{\alpha \in \omega} n_\alpha \).

But \( \sum_{\alpha \in \omega} m_\alpha = \kappa_\omega \) since there is a bijection from \( \omega_\alpha = \bigcup_{\alpha \in \omega} \alpha \) to \( \prod_{\alpha \in \omega} \omega_\alpha \) and \( \omega_\alpha = \kappa_\alpha \).

However, \( \left( 2^{\kappa_\omega} \right)^{\aleph_0} = 2^{\left( \kappa_\omega \cdot \aleph_0 \right)} = 2^{\aleph_0} \), \( \omega \not= \aleph_0 \).

Chapter 7: Problems of Consistency and Independence

Recall: since ZF is a first order theory in a countable language, it must (if consistent) have countable models (note that the function enumerating the elements of the model cannot be definable by a function class). We might still hope that ZF is complete (i.e. for every sentence \( \phi \) we have either \( \text{ZF} + \phi \) or \( \text{ZF} \vdash \neg \phi \)), or at least that we could complete it by adding a finite set of extra axioms or axiom schemes.

There are easy independence proofs of Foundation, Infinity, Power Set and Union axioms relative to the rest of ZF (see question 3 and 9 on Example Sheet 3).

We can also prove consistency of Foundation relative to \( \text{ZF} \backslash \{ \text{Foundation} \} \) by cutting down to the class of regular sets (cf. Theorem 4.4).
Gödel showed that \( \text{GCH} \) and \( \text{AC} \) are consistent relative to \( \text{ZF} \), by "slowing down" the von Neumann hierarchy: he defined the constructible universe \( L \) by setting

\[
L_0 = \emptyset, \quad L_{\alpha+1} = \text{Def}(L_\alpha) = \{ x \in L_\alpha | x = \{ y \in L_\alpha | \phi \} \text{ where } \phi \text{ is a formula with parameters in } L_\alpha \} \\
L_\lambda = \bigcup \{ L_\alpha | \alpha < \lambda \} \text{ for } \lambda \text{ a limit ordinal}.
\]

\[
L = \bigcup \{ L_\alpha | \alpha \in \text{Ord} \}
\]

\( L \) does satisfy all the axioms of \( \text{ZF} \) (the Power Set axiom is the difficult one), but we can show by induction that \( L_\alpha \) has a definable well-ordering for each \( \alpha \), and hence that \( L \) can be well ordered.

\( L_\omega \) satisfies \( \text{GCH} \).
Let $\Psi(x) = \neg \text{Thm}_T(x)$

Let $c : \mathbb{N} \to \mathbb{N}$ be the coding function, i.e. $c(n) = n^7$
and $mb_{x} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be the function such that

$\text{mb}_{x}(\Psi^{n}, \overline{t^{7}}) = \Psi^{\text{mb}_{x}(\Psi^{n}, \overline{t^{7}})}$

Let $\Psi(y) = \Psi(\text{mb}_{x}(y, c(y)))$

Let $m = \Psi(x)^7$

Now, $\Psi(m) = \Psi(\text{mb}_{x}(m, c(m))) = \Psi(\text{mb}_{x}(\Psi^{c(m)}^{7}, m^{7}))$

$= \Psi(\overline{m^{7}}^{7}) = \neg \text{Thm}_T(\Psi(m))$

Clearly $T \vdash \Psi(m)$ unless $T$ is inconsistent.

If $T$ is $\omega$-consistent (a slight strengthening of consistency), then
$T \vdash \neg \Psi(m)$ either. We can formalise the assertion that $T$ is consistent, as the formula $\text{Con}_T = \neg \text{Thm}_T(\overline{\bot^{7}})$

We claim that $T \vdash (\Psi(m) \iff \text{Con}_T)$, and hence $T \not\vdash \text{Con}_T$ unless $T$ is inconsistent.

We have $\vdash (\bot \Rightarrow \Psi(m)) \Rightarrow \text{PA} \vdash \text{Thm}_T(\neg (\bot \Rightarrow \Psi(m))^{7})$

$\Rightarrow \text{PA} \vdash (\text{Thm}_T(\bot^{7}) \Rightarrow \text{Thm}_T(\neg \Psi(m)^{7}))$

$\Rightarrow \text{PA} \vdash (\Psi(m) \Rightarrow \text{Con}_T)$.

Conversely, $\text{PA} \vdash (\text{Thm}_T(\neg \Psi(m)^{7}) \Rightarrow \text{Thm}_T(\neg \text{Thm}_T(\neg \Psi(m)^{7})))$

i.e. $\text{PA} \vdash (\text{Thm}_T(\neg \Psi(m)^{7}) \Rightarrow \text{Thm}_T(\neg \neg \Psi(m)^{7}))$

and hence formalising a propositional deduction we get
PA \vdash (\text{Thm}_T (\neg \Psi(m))) \Rightarrow \text{Thm}_T (\neg \Gamma)

i.e. \ PA \vdash (\text{Cons} \Rightarrow \Psi(m))
P. Cohen developed a technique called 'forcing' which can be used to adjoin a 'generic' object to a model of set theory. He showed that if we adjoin a generic function \( f : a \to Pw \), then

a) \( f \) is injective

b) 'Cardinals are preserved', i.e. if there is no injection \( b \to c \) in the original model then there is not one in the extended model.

So if \( a = \beta Pw \), then the set \( f(Pw) \) is a subset of \( Pw \) whose cardinality is strictly between those of \( w \) and \( Pw \).

This shows that if ZFC is consistent, then \( \alpha \) is ZFC + \( \exists \alpha \text{CH} \).

A. Fraenkel and A. Mostowski (1920s) proved the independence of AC from a theory ZFA ('ZF with atoms'), in which Extension fails.

E. Specker modified this by replacing atoms with 'antiminima', i.e. sets \( x \) satisfying \( x = \{ x \} \) (so that Extension holds but Foundation fails) - see question 12, sheet 3.

However, the well-founded part of any such model is the original model Cohen observed that given a Fraenkel-Mostowski-Specker model, with a set \( a \) having no choice function, if we adjoin a generic function \( a \to Pw \), we get a subset of \( Pw \) with no choice function.

Now by cutting down to the well-founded part of the model, we get a model of ZFU \( \{\neg AC \} \).
Gödel's Incompleteness Theorem

Informally, no 'sufficiently complicated' theory can be both
complete and consistent. Suppose, given a theory $T$ (e.g. $\text{ZF}$) in
a countable language, we can encode all formulae $\varphi$ in our language
by natural numbers $\num{\varphi}$, and also finite sequences of formulae
by natural numbers.

Write $\text{ded}_T(M,n)$ for the assertion that $M$ encodes a deduction of
the formula coded by $n$, from the axioms of $T$.

If $T$ is an extension of $\text{ZF}$, we also have a model of
$\text{PA}$ (first order Peano arithmetic) inside any model of $T$.

Hence we have an interpretation of $\text{PA}$ in $T$,
i.e. a translation from formulae of $\text{PA}$ to formulae of $T$ which
maps the axioms of $\text{PA}$ to deducible sentences in $T$.

We write $\text{Thm}_T(x,y)$ for the formula $(\exists x) \text{ded}_T(x,y)$.

Then we have $T \vdash \varphi$ implies $\text{PA} \vdash \text{Thm}_T(\num{\varphi})$ and we also
have $\text{PA} \vdash (\text{Thm}_T(\num{\varphi}) \Rightarrow \text{Thm}_T(\num{\text{Thm}_T(\num{\varphi})}))$.

We can also formalize Modulo Poreno; we have a function
$\text{imp} : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ such that (definable in Peano Arithmetic) such that
$\text{imp}(\num{\varphi}, \num{\psi}) = \num{(\varphi \Rightarrow \psi)}$
and then show $\text{PA} \vdash (\forall x,y) (\text{Thm}_T(\text{imp}(x,y)) \Rightarrow \text{Thm}_T(x \Rightarrow \text{Thm}_T(x)))$. 