# Logic, Computation & Set Theory, IIA/IIB

Dr P T Johnstone

Michaelmas 2001

# Contents



These notes are based upon the lecture course given by Dr Johnstone in Michaelmas 2001, though Dr Johnstone has no other connection with them. Mistakes are doubtless the fault of the typist. The reader is of course reminded that these notes are no substitute for attending lectures.

Comments and corrections to `steve@megson.org`.

– Steve Megson, August 2002

# Chapter 1

# Partially-Ordered Sets

**Definition 1.1** A PARTIAL ORDER on a set $A$ is a binary relation $\leq$ on $A$ such that

| | | |
|---|---|---|
| (1) | $\leq$ is reflexive | $(\forall x \in A)(x \leq x)$ |
| (2) | $\leq$ is transitive | $(\forall x, y, z \in A)((x \leq y)$ and $(y \leq z)$ implies $(x \leq z))$ |
| (3) | $\leq$ is antisymmetric | $(\forall x, y \in A)((x \leq y)$ and $(y \leq x)$ implies $(x = y))$ |

A POSET is a set equipped with a partial order.

**Example 1.2**

(a) The usual ordering $\leq$ on $\mathbb{N}$ or $\mathbb{R}$ is a partial order, and in fact a TOTAL ORDER, so for any two elements $x, y$ we have either $x \leq y$ or $y \leq x$. Note that $x \leq y$ holds in $\mathbb{N}$ iff $\exists z \in \mathbb{N}$ s.t. $x + z = y$ and holds in $\mathbb{R}$ iff $\exists z \in \mathbb{R}$ s.t. $x + z^2 = y$.

(b) The relation $|$ on $\mathbb{N}$ defined by $x|y$ iff $x$ divides $y$ (ie $\exists z \in \mathbb{N}$ s.t. $xz = y$) is a partial order.

(c) Let $\mathcal{P}A$ denote the power set of $A$, ie the set of all subsets of $A$. Then the relation $\subseteq$ on $\mathcal{P}A$ is a partial order.

(d) If $B$ is a subset of a poset $(A, \leq_A)$ then the restriction of $\leq_A$ to $B$ is a partial order on $B$. So, for example, the set of subspaces of a vector space $V$ is partially-ordered by $\subseteq$.

(e) Given a set $\Sigma$ of abstract symbols ('letters') we can form the set $\Sigma^*$ of all finite strings of symbols in $\Sigma$ ('words').
We define the SUBWORD ORDERING on $\Sigma^*$ by $u \leq v$ if $\exists$ words $w, x$ such that $v = wux$ ($u$ occurs somewhere in $v$)
We define the PREFIX ORDERING by $u \leq v$ if $\exists x$ such that $v = ux$ ($u$ is the start of $v$).

(f) A PARTIAL FUNCTION from $A$ to $B$ is a function defined on a subset of $A$ and taking values in $B$. We write '$f : A \rightharpoonup B$' for '$f$ is a partial function from $A$ to $B$'. (eg $x \mapsto \log x$ and $x \mapsto \frac{1}{x}$ are partial functions $\mathbb{R} \rightharpoonup \mathbb{R}$)
We write $[A \rightharpoonup B]$ for the set of all partial functions from $A$ to $B$. On this set we have a partial order $\leq$ defined by $f \leq g$ iff $g$ extends $f$, ie $\operatorname{dom} g \supseteq \operatorname{dom} f$ and $g(x) = f(x)$ wherever $f(x)$ is defined.

We can represent finite posets pictorially as follows:
Represent the elements of $A$ by vertices and instances $x \leq y$ of the order relation by lines $\begin{smallmatrix} & & y \\ & / & \\ x & & \end{smallmatrix}$ such that $y$ is higher up the page than $x$. However, if we have $\begin{smallmatrix} & z & \\ & \diagdown & \\ x & & y \\ & \diagup & \end{smallmatrix}$ then we don't need to draw the line $\begin{smallmatrix} z \\ | \\ x \end{smallmatrix}$

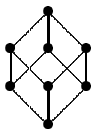
We say that $y$ COVERS $x$ in a poset (and write $x \lhd y$) if $x < y$ (ie $x \leq y$ and $x \neq y$) but if $z$ satisfies $x \leq z \leq y$ then $z = x$ or $z = y$.

**Lemma 1.3** *In a finite poset, $x \leq y$ holds iff either $x = y$ or there is a finite chain $x \lhd z_1 \lhd z_2 \cdots \lhd z_n \lhd y$.*

*Proof:* $\Leftarrow$ : obvious
$\Rightarrow$ : If $x \leq y$, then either $x = y$ or $x < y$ If $x < y$, then either $x \lhd y$ or $\exists z$ with $x < z < y$
If $x < z < y$ then either $x \lhd z \lhd y$ or $\exists z'$ with $x < z < z' < y$ or $\exists z''$ with $x < z'' < z < y$
Since the poset is finite, this process must terminate. □

**Definition 1.4** The HASSE DIAGRAM of a finite poset $(A, \leq)$ has the elements of $A$ as vertices and instances of $\lhd$ as edges drawn so that if $x \lhd y$ then the vertex representing $y$ is higher up the page than that representing $x$.

**Example 1.5** (a) Consider $\mathcal{P}A$ where $A = \{a, b, c\}$



(b) Consider the poset $\mathrm{Sub}(V)$ of the non-cyclic group $V$ of order 4.

**Definition 1.6** Let $(P, \leq)$ be a poset, $S$ a subset of $P$

(i) By a GREATEST ELEMENT of $S$, we mean an element $x \in S$ such that $x \geq y$ for all $y \in S$;

(ii) By a MAXIMAL ELEMENT of $S$, we mean an element $x \in S$ such that if $x \leq y$ for any $y \in S$, then $x = y$;

(iii) By a UPPER BOUND of $S$, we mean an element $x \in P$ such that $x \geq y$ for all $y \in S$;

(iv) By a LEAST UPPER BOUND (or SUPREMUM or JOIN) for $S$ we mean an upper bound $x$ for $S$ such that if $y$ is any upper bound for $s$, then $x \leq y$.

Dually, we define the GREATEST LOWER BOUND/INFIMUM/MEET for $S$.

*Remarks:* A greatest element of $S$, if it exists, is a maximal element. Conversely, if $S$ is totally ordered then any maximal element of $S$ is a greatest element.
A greatest element of $S$, if it exists, is a least upper bound for $S$. Conversely, a least upper bound for $S$ is a greatest element of $S$ iff it belongs to $S$.

**Example 1.7** If $S \subseteq \mathcal{P}A$ for some $A$, then the least upper bound of $S$ in $\mathcal{P}A$ exists and is the set $\bigcup S = \{x \in A \mid (\exists B \in S)\,(x \in B)\}$, and similarly the greatest lower bound of $S$ is $\bigcap S = \{x \in A \mid (\forall B \in S)\,(x \in B)\}$.

In general, we write $\bigvee S$ and $\bigwedge S$ for the sup and inf of $S$, if they exist.

*Warning:* $\bigvee S$ depends on the poset $P$ of which $S$ is considered to be a subset, as well as on $S$.
eg. If $P$ is the poset of all subspaces of a vector space $V$, and $S = \{U, W\}$ then $\bigvee S = U + W$, but if $S$ is considered as a subset of $\mathcal{P}V$, then $\bigvee S = U \cup W$. If necessary, write $\bigvee_P S$ for the sup of $S$ in $P$.

**Lemma 1.8** *Let $(P, \leq)$ be a poset. Then every subset of $P$ has a sup $\Leftrightarrow$ every subset of $P$ has an inf.*

*Proof:* The two implications are duals of each other, we will prove $\Rightarrow$.

Let $S$ be an arbitrary subset of $P$. Consider the set $L \subseteq P$ of all lower bounds of $S$. By assumption, $\bigvee L$ exists in $P$. For all $x \in S$, $y \in L$ we have $y \leq x$ so every element of $S$ is an upper bound for $L$, so every $x \in S$ satisfies $\bigvee L \leq x$. So $\bigvee L$ is a lower bound for $S$, ie $\bigvee L \in L$. Hence $\bigvee L$ is the greatest element of $L$, ie. $\bigvee L = \bigwedge S$. □

**Definition 1.9**

(i) We say that a poset is COMPLETE if every subset of $P$ has a sup;

(ii) By a CHAIN in a poset $(P \leq)$ we mean a non-empty subset of $P$ which is totally-ordered by $\leq$;

(iii) We say $(P \leq)$ is CHAIN-COMPLETE if every chain has a sup in $P$.

*Warning:* If $P$ is a complete poset and $Q \subseteq P$, $Q$ *may* be complete without being closed under sups in $P$. If $\bigvee_P S \in Q$ for some $S \subseteq Q$ then it is necessarily a least upper bound for $S$ in $Q$, but $\bigvee_Q S$ can exist without being equal to $\bigvee_P S$.
eg. If $P = \mathcal{P}G$ for a group $G$ and $Q$ the set of subgroups of $G$, then both $P$ and $Q$ are complete, but $\bigvee_Q S \neq \bigvee_P S$ in general.

**Example 1.10**

(i) For any set $A$, $\mathcal{P}A$ is complete;

(ii) $(\mathbb{R}, \leq)$ is not complete, but $\mathbb{R} \cup \{\pm\infty\}$ is complete.

*Note* that any complete poset $P$ has a least element $0 = \bigvee \emptyset$ and a greatest element $1 = \bigvee P$.

(iii) $(\mathbb{N}, \text{divisibility})$ is complete. It has a least element of 1. For a finite set $S \subseteq \mathbb{N}$, $\bigvee S$ is the lcm of the members of $S$. For an infinite $S \subseteq \mathbb{N}$, 0 is the only upper bound for $S$, since any non-zero natural number has finitely many factors, and so $0 = \bigvee S$. (We take $0 \in \mathbb{N}$)

(iv) The set $[A \rightharpoonup B]$ is not complete in general. If $f$ and $g$ are both defined at some $x \in A$ and satisfy $f(x) \neq g(x)$, then $\{f, g\}$ has no upper bound in $[A \rightharpoonup B]$. but $[A \rightharpoonup B]$ is chain-complete. If $S \subseteq [A \rightharpoonup B]$ is a chain then any two members of $S$ must agree on the intersection of their domains. So if $A' = \bigcup \{\text{dom}\, f \mid f \in S\}$ then there's a unique function $g$ with domain $A'$ such that if $x \in A'$ then $g(x) = f(x)$ for some $f \in S$ such that $x \in \text{dom}\, f$. Then $g$ is the least upper bound for $S$ in $[A \rightharpoonup B]$.

(v) The set $\Sigma^*$ of finite strings of members of $\Sigma$ is not chain complete (in either of the two orderings considered). We can find an infinite chain $\{w_0, w_1, \dots\}$ such that $w_i \leq w_j$ for all $i \leq j$ having no upper bound in $\Sigma^*$.

Given posets $P$ and $Q$ we say a function $f : P \to Q$ is ORDER PRESERVING if $x \leq y$ in $P$ implies $f(x) \leq f(y)$ in $Q$.

**Theorem 1.11** *(Tarski-Knaster Theorem)*
*If $P$ is a complete poset, then any order-preserving map $f : P \to P$ has a fixed point.*

*Proof:* Consider the set $S = \{x \in P \mid x \leq f(x)\}$ of 'pre-fixed points' of $f$. Let $y = \bigvee S$. If $x \in S$, then we have $x \leq f(x) \leq f(y)$ (order-preserving). So $f(y)$ is an upper bound for $S$. Hence $y = \bigvee S \leq f(y)$, ie. $y \in S$. Since $f$ is order-preserving we also have $f(y) \leq f(f(y))$, so $f(y) \in S$ and hence $f(y) \leq \bigvee S = y$. Hence $f(y) = y$, and $y$ is a fixed point. □

**Corollary 1.12** *(Cantor-Bernstein Theorem)*
*If $A$ and $B$ are sets and $\exists$ injections $f : A \to B$ and $g : B \to A$, then $\exists$ a bijection $h : A \to B$.*

*Proof:* Suppose we can decompose $A = A_1 \sqcup A_2$ and $B = B_1 \sqcup B_2$ in such a way that $f(A_1) = B_1$ and $g(B_2) = A_2$. Then we can define $h : A \to B$ by $h(x) = \begin{cases} f(x) & \text{if } x \in A_1 \\ g^{-1}(x) & \text{if } x \in A_2 \end{cases}$ and $h$ will be bijective.

To find such partitions of $A$ and $B$, we use the Tarski-Knaster theorem. Consider the composite mapping

$$\mathcal{P}A \xrightarrow{f} \mathcal{P}B \xrightarrow{B\setminus(\cdot)} \mathcal{P}B \xrightarrow{g} \mathcal{P}A \xrightarrow{A\setminus(\cdot)} \mathcal{P}A$$

This is order-preserving since the 1st and 3rd maps are order-preserving and the other two are order-reversing. $\mathcal{P}A$ is complete, so there exists and $A_1 \subseteq A$ such that $f(A_1) = A_1$. If we now define $B_1 = f(A_1)$, $B_2 = B \setminus B_1$ and $A_2 = g(B_2)$ we have $A_1 = A \setminus A_2$ as required. □

*Remark:* The analogue of Theorem 1.11 for chain-complete posets is false. Given any set $P$, give it the DISCRETE PARTIAL ORDER ($x \leq y \Leftrightarrow x = y$). This is chain-complete since the only chains are singletons and any mapping $P \to P$ is order-preserving. So if $P$ has $\geq 2$ elements, we can find an order-preserving $f : P \to P$ with no fixed points.

For an arbitrary poset $(P, \leq)$, we say $f : P \to P$ is INFLATIONARY if $x \leq f(x)$ holds for all $x \in P$.

**Theorem 1.13** *(Bourbaki-Witt Theorem)*
*Let $(P, \leq)$ 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 $x \leq y = f(y)$. Moreover, if $f$ is order-preserving then there is a least such $y$.*

*Non-proof:* First consider $x$: if $x = f(x)$ then stop
If not, consider $f(x)$: if $f(x) = f(f(x))$ then stop
If not, consider $f(f(f(x)))\ldots$
If none of the $f^n(x)$ are fixed, consider $\bigvee\{f^n(x) \mid n \geq 0\}$: if this is a fixed point then stop
If not, consider $f\left(\bigvee\{f^n(x) \mid n \geq 0\}\right)\ldots$
How long can we go on doing this?

*Proof:* We define a subset $C$ of $P$ to be CLOSED if

(i) whenever $y \in C$, we have $f(y) \in C$

(ii) whenever $S \subseteq C$ is a chain, we have $\bigvee S \in C$

We note that any intersection of closed sets is closed, so for any $B \subseteq P$ there is a smallest closed subset $C$ with $B \subseteq C$, namely the intersection of all such $C$. In particular, write $C(x)$ for the smallest closed set with $x \in C$. Suppose we can prove that $C(x)$ is a chain, then $\bigvee C(x) \in C(x)$ and $f\left(\bigvee C(x)\right) \in C(x)$, so $f\left(\bigvee C(x)\right) \leq \bigvee C(x)$. Since $f$ is inlationary this implies that $\bigvee C(x)$ is a fixed point, and $x \leq \bigvee C(x)$ since $x \in C(x)$.

To prove that $C(x)$ is a chain:

*Step 1:* Observe that every $y \in C(x)$ satisfies $y \geq x$, since the set $\{y \in P \mid y \geq x\}$ is closed and has $x$ as a member, so it contains $C(x)$. We'll say that an element $y \in C(x)$ is NORMAL if $(\forall z \in C(x))\,(z < y \Rightarrow f(z) \leq y)$

*Step 2:* If $y$ is normal, then $\forall z \in C(x)$ we have either $z \leq y$ or $f(y) \leq z$ for if we define $T_y = \{z \in C(x) \mid z \leq y \text{ or } f(y) \leq z\}$, then $x \in T_y$ since $x \leq y$ by Step 1. If $z \in T_y$ then

| | | | | | |
|---|---|---|---|---|---|
| either | $z < y$ | or | $z = y$ | or | $f(y) \leq z$ |
| so either | $f(z) \leq y$ | or | $f(z) = f(y)$ | or | $f(y) \leq f(z)$ |
| so | $f(z) \in T_y$ | | | | |

If $S \subseteq T_y$ is a chain, then $\forall z \in S$, we have $z \leq y$ or $f(y) \leq z$ so

| | | | |
|---|---|---|---|
| either | $z \leq y\ \forall z \in S$ | or | $f(y) \leq z$ for some $z \in S$ |
| either | $\bigvee S \leq y$ | or | $f(y) \leq \bigvee S$ |
| so | $\bigvee S \in T_y$ | | |

Hence $T_y$ is a closed set containing $x$, so $T_y = C(x)$.

*Step 3:* Every element of $C(x)$ is normal, for if $N = \{y \in C(x) \mid y \text{ is normal}\}$, then $x \in N$ since the hypothesis $z < x$ is never satisfied for $z \in C(x)$. If $y \in N$ and $z < f(y)$, then $z \leq y$ by Step 2, so

$$\begin{array}{lrcr} & z < y & \text{or} & z = y \\ \text{so} & f(z) \leq y & \text{or} & f(z) = f(y) \\ \text{so} & f(z) \leq y & & \end{array}$$

So $f(y) \in N$.

If $S \subseteq N$ is a chain and $z < \bigvee S$, then we can't have $f(y) = z \; \forall y \in S$ since this would imply $y \leq z \; \forall y \in S$ and hence $\bigvee S \leq z$. So for some $y \in S$ we have $z \leq y$, and in fact we must have $z < y$ for some $y$ since otherwise we would have $z = \bigvee S$. Hence by normality we have $f(z) \leq y \leq \bigvee S$, so $\bigvee S \in N$. Hence $N$ is a closed set containing $x$, so $N = C(x)$.

Now, if $y, z$ are any two members of $C(x)$, then $y$ is normal by Step 3,

$$\begin{array}{llcll} \text{so} & z \leq y & \text{or} & f(y) \leq z & \text{by Step 2} \\ \text{so} & z \leq y & \text{or} y \leq z & \text{since } f \text{ is inflationary} & \end{array}$$

So we've proved that $C(x)$ is a chain.

Finally, suppose $f$ is also order-preserving. Then, for any fixed point $z$, the set $\{w \in P \mid w \leq z\}$ is closed. So if $z$ is fixed and $x \leq z$, we have $C(x) \leq \{w \in P \mid w \leq z\}$. So $\bigvee C(x) \leq z$. Hence $\bigvee C(x)$ is the least fixed point $\geq x$. □

**Corollary 1.14** *If $P$ is a chain-complete poset and $f : P \to P$ is order-preserving, then for any $x \in P$ with $x \leq f(x)$ there is a least $y \in P$ with $x \leq y = f(y)$. In particular, if $P$ has a least element then $f$ has a least fixed point.*

*Proof:* Let $P_1 = \{x \in P \mid x \leq f(x)\}$. Then $x \in P_1 \Rightarrow f(x) \in P_1$ since $f$ is order-preserving, and if $S \subseteq P_1$ and $\bigvee S$ exists in $P$, then $\bigvee S \in P_1$ (cf proof of 1.11). So $P_1$ is chain-complete and $f|_{P_1}$ is an inflationary map $P_1 \to P_1$. Apply Theorem 1.13 to $f_{P_1}$ to obtain the result.

For the last assertion, if $f$ is order-preserving then a least element is necessarily pre-fixed. □

We've seen that if $P$ is a chain-complete poset and has a least element, and $F : P \to P$ is order-preserving, then $f$ has a least fixed point. One application of this result concerns recursive definitions, eg. that of the factorial function

$$f : \mathbb{N} \to \mathbb{N} \;\; f(n) = \begin{cases} 1 & (\text{if } n = 0) \\ n.f(n-1) & (\text{if } n > 0) \end{cases}$$

To avoid the circularity of this definition, we consider it as defining a mapping $\Phi : [\mathbb{N} \to \mathbb{N}] \longrightarrow [\mathbb{N} \to \mathbb{N}]$

$$\Phi(f)(n) = \begin{cases} 1 & \text{if } n = 0 \\ n.f(n-1) & \text{if } n > 0 \text{ and } f(n-1) \text{ is defined} \end{cases}$$

Now $[\mathbb{N} \to \mathbb{N}]$ is chain-complete in the extension ordering, and has a least element (the everywhere undefined function) and it's easy to check that $f \leq g \Rightarrow \Phi(f) \leq \Phi(g)$. So $\Phi$ has a least fixed point, but in fact any fixed point of this $\Phi$ must be a total function, ie. a maximal element of $[\mathbb{N} \to \mathbb{N}]$, and so the fixed point is unique.

An application of the original (non-order-preserving) version of 1.13:

**Theorem 1.15** *(Zorn's Lemma)*
*Let $(P, \leq)$ be a chain-complete poset. Then for every element $x \in P$ there exists a maximal element $y \in P$ with $x \leq y$.*

To prove this we need to assume the AXIOM OF CHOICE: if $\{A_i \mid i \in I\}$ is a set-indexed family of sets and $A_i \neq \emptyset$ for each $i \in I$, then there exists a function $f : I \to \bigcup\{A_i \mid i \in I\}$ such that $f(i) \in A_i \ \forall i$ (such an $f$ is called a CHOICE FUNCTION for $\{A_i \mid i \in I\}$).

We apply it to the family $\{A_x \mid x \in P\}$ where

$$A_x = \begin{cases} \{y \in P \mid y > x\} & \text{if } x \text{ is not maximal} \\ \{x\} & \text{if } x \text{ is maximal} \end{cases}$$

By definition, we have $A_x \neq \emptyset \ \forall x \in P$, so the axiom yields a function $f : P \to P$ satisfying $f(x) \in A_x \ \forall x \in P$. Clearly, $x \leq f(x) \ \forall x \in P$, but the fixed points of $x$ are exactly the maximal elements of $P$, so the result follows from 1.13. □

Applications of Zorn's Lemma occur in many parts of maths, eg.

**Corollary 1.16** *Every linearly independent set in a vector space $V$ (over an arbitrary field) can be extended to a basis.*

*Proof:* Let $P$ be the set of all linearly independent subsets of $V$, ordered by inclusion. If $S \subseteq P$ is a chain, we'll show $B = \bigcup S = \bigcup\{A | A \in S\}$ is LI, so that it is a least upper bound for $S$ in $P$.

If $B$ were linearly dependent, we'd have a linear relation $\sum_{i=1}^{n} \lambda_i x_i = 0$, where the $x_i$ are distinct members of $B$ and not all $\lambda_i$ are zero. For each $i$, $\exists A_i \in S$ s.t. $x_i \in A_i$, but the $A_i$ are totally ordered by inclusion, so there's a largest one $A_{i_0}$ say, and then we have $x_i \in A_{i_0} \ \forall i, \ i \leq i \leq n$. So $A_{i_0}$ is LD, giving a contradiction.

So by 1.15 every member of $P$ is contained in a maximal member. We need to show that a maximal member $M$ of $P$ is a spanning set. Suppose not, then $V \setminus \langle M \rangle$ is non-empty; let $y \in V \setminus \langle M \rangle$. Then $M \cup \{y\}$ is LI, since a non-trivial linear relation would express $y$ as a linear combination of elements of $M$. So $M$ isn't maximal. □

We can also deduce the Axiom of Choice from Zorn's Lemma as follows:
Given $(A_i \mid i \in I)$, consider the set of all partial functions $f : I \to \bigcup\{A_i \mid i \in I\}$ such that $f(i) \in A_i$ whenever $f(i)$ is defined. This is easily seen to be chain-complete and it's non-empty since the empty function belongs to it. So it has a maximal element, $f$, say. If $f$ isn't total, choose any $i \in I \setminus \operatorname{dom} f$, and any $x \in A_i$, and define $g$ by

$$g(j) = \begin{cases} f(j) & \text{if } f(j) \text{ is defined} \\ x & \text{if } j = i \\ \text{undefined} & \text{otherwise} \end{cases}$$

Then $g$ is a partial choice function and $f < g$, giving a contradiction. □

**Definition 1.17** A LATTICE is a poset $L$ in which every finite subset has a sup and an inf. Equivalently, $L$ has a least element 0 and a greatest element 1, and for any doubleton set $\{x, y\} \subseteq L$ we have a JOIN $\bigvee\{x, y\} = x \vee y$ and a MEET $\bigwedge\{x, y\} = x \wedge y$.
(We can then construct $\bigvee\{x_1, \ldots, x_n\}$ as $(\cdots((x_1 \vee x_2) \vee x_3) \vee \cdots)$ and similarly for meets)

*Note* that the binary operations $\vee$ and $\wedge$ are both commutative and associative, and have 0 and 1 as their respective identities. They also satisfy the IDEMPOTENCY LAWS $x \vee x = x = x \wedge x$.

We call a lattice DISTRIBUTIVE if the distributive law $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ holds.

**Lemma 1.18** *For a lattice $L$, the distributive law $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ (1) holds for all $x, y, z \in L$ iff $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ (2) holds for all $x, y, z \in L$.*

*Proof:* We show (1) $\Rightarrow$ (2). Applying (1) to the RHS of (2) we get

$$\begin{aligned}
(x \vee y) \wedge (x \vee z) &= ((x \vee y) \wedge x) \vee ((x \vee y) \wedge z) \\
&= x \vee ((x \vee y) \wedge z) && \text{since } x \vee y \geq x \\
&= x \vee ((x \wedge z) \vee (y \wedge z)) && \text{using (1) again} \\
&= (x \vee (x \wedge z)) \vee (y \wedge z) && \text{since } \vee \text{ is associative} \\
&= x \vee (y \wedge z) && \text{since } x \geq x \wedge z
\end{aligned}$$

□

In a lattice, we say an element $y$ is a COMPLEMENT for $x$ if $x \vee y = 1$ and $x \wedge y = 0$.

**Lemma 1.19** *In a distributive lattice, each element has at most one complement.*

*Proof:* Suppose that $y, z$ are both complements of $x$. Then

$$\begin{aligned}
y = (y \wedge 1) &= y \wedge (x \vee z) \\
&= (y \wedge x) \vee (y \wedge z) \\
&= 0 \vee (y \wedge z) \\
&= (y \wedge z)
\end{aligned}$$

Similarly, $z = y \wedge z$, so $y = z$. □

We call a lattice $L$ a BOOLEAN ALGEBRA if it's distributive and every element of $L$ has a complement. We write $\neg x$ for the (unique) complement of $x \in L$.

**Example 1.20**

(i) For any set $A$, $\mathcal{P}A$ is a Boolean algebra. The distributive law $B \cap (C \cup D) = (B \cap C) \cup (B \cap D)$ holds since both sides have the same elements, and every $B \subseteq A$ has a complement $A \setminus B$.

(ii) Any totally ordered set with 0 and 1 is a lattice with $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$. Moreover, it's distributive since if $y \leq z$ then both $(x \wedge (y \vee z))$ and $(x \wedge y) \vee (x \wedge z)$ reduce to $(x \vee z)$. But a totally ordered set with $\geq 3$ elements is not a Boolean algebra.

(iii) The lattice of subgroups of the 4-elt non-cyclic group $\{e, x, y, z\}$ is not distributive since each of $\{e, x\}, \{e, y\}, \{e, z\}$ has 2 distinct complements.

If $L$ and $M$ are lattices, a LATTICE HOMOMORPHISM $f : L \to M$ is a function satisfying

$$\begin{aligned}
f(0_L) &= 0_M \\
f(1_L) &= 1_M \\
f(x \vee y) &= f(x) \vee f(y) && \forall x, y \in L \\
f(x \wedge y) &= f(x) \wedge f(y) && \forall x, y \in L
\end{aligned}$$

Such an $f$ is necessarily order-preserving since $x \leq y \Leftrightarrow x \wedge y = x$.

**Lemma 1.21** *Let $L$ be a distributive lattice and $a, b$ two elements of $L$ with $a \not\leq b$. Then there is a lattice homomorphism $f : L \to 2 = \{0, 1\}$ such that $f(a) = 1$ and $f(b) = 0$.*

*Proof:* Consider the set $P$ of all pairs $(A, B)$ of subsets of $L$ satisfying

- $a \in A$ and $A$ is closed under $\wedge$;
- $b \in B$ and $B$ is closed under *vee*;
- If $x \in A$ and $y \in B$, then $x \not\leq y$.

We order $P$ by pointwise inclusion ( $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2$ , $B_1 \subseteq B_2$ ). If $\{(A_i, B_i) \mid i \in I\}$ is a chain in $P$, then $\left(\bigcup_{i \in I} A_i, \bigcup_{i \in I} B_i\right) \in P$ since eg. if $x \in \bigcup A_i$ and $y \in \bigcup B_i$ then for some $i$, $x \in A_i$ and for some $j$, $y \in B_j$. Since we have either $A_i \subseteq A_j$ or $B_j \subseteq B_i$, $\exists k$ s.t. $x \in A_k$, $y \in B_k$. Hence $x \not\leq y$. Since $(\{a\}, \{b\}) \in P$, we conclude by Zorn's Lemma that $P$ has a maximal element $(A_0, B_0)$ say.

We need to show that $A_0 \cup B_0 = L$. Suppose that $c \in L \setminus (A_0 \cup B_0)$. There are three cases:

(1) Suppose $\forall x \in A_0, y \in B_0$ we have $x \wedge c \not\leq y$. Then we can define $A_1 = A_0 \cup \{c\} \cup \{c \wedge c \,|\, x \in A_0\}$ and $B_1 = B_0$ and then $(A_0, B_0) < (A_1, B_1) \in P$, which can't happen.

(2) Suppose $\forall x \in A_0, y \in B_0$ we have $x \not\leq c \vee y$. Then we can enlarge $B_0$ leaving $A_0$ unchanged, which again can't happen.

(3) Suppose both (1) and (2) fail, then $\exists\, x_1 \in A_0, y_1 \in B_0$ with $x_1 \wedge c \leq y_1$ and $x_2 \in A_0, y_2 \in B_0$ with $x_2 \leq c \vee y_2$. Putting $x = x_1 \wedge x_2$, $y = y_1 \vee y_2$ we get $x \in A_0, y \in B_0$ with $x \wedge c \leq y$, $x \leq c \vee y$. Now $x = x \wedge (c \vee y) = (x \wedge c) \vee (x \wedge y) \leq y \vee y = y$ contradicting $(A_0, B_0) \in P$.

So we now have a total function $f : L \to 2$ defined by $f(x) = \begin{cases} 1 & \text{if } x \in A_0 \\ 0 & \text{if } x \in B_0 \end{cases}$

Since $a \leq 1$ and $a \in A_0$, we can't have $1 \in B_0$ and so $1 \in A_0$, ie $f(1) = 1$. Similarly $f(0) = 0$.

If $x, y \in A_0$ then $x \wedge y \in A_0$ so $f(x \wedge y) = 1 = f(x) \wedge f(y)$. If either $x \in B_0$ or $y \in B_0$ then $x \wedge y \in B_0$ and so $f(x \wedge y) = 0 = f(x) \wedge f(y)$. Similarly, $f$ preserves $\vee$, so it's a lattice homomorphism. □

**Theorem 1.22** *(Birkhoff-Stone Representation Theorem)*

*(1) A lattice $L$ is distributive iff it is isomorphic to a sub-lattice of $\mathcal{P}A$ for some $A$ (ie. a subset closed under finite unions and intersections).*

*(2) $L$ is a Boolean algebra iff it is isomorphic to a subset of $\mathcal{P}A$ closed under finite unions and complements.*

*Proof:* $\Leftarrow$ holds since any sublattice of $\mathcal{P}A$ is distributive.

$\Rightarrow$ Given $L$, let $A$ be the set of all lattice homomorphisms $L \to 2$ and define $\Phi : L \to \mathcal{P}A$ by $\Phi(x) = \{f \in A \,|\, f(x) = 1\}$. We claim that $\Phi$ is a lattice homomorphism.

$$\begin{aligned}
&\Phi(1) = A \quad \text{since } f(1) = 1 \forall f \in A \\
&\Phi(0) = \emptyset \quad \text{since no } f \in A \text{ satisfies } f(0) = 1 \\
&f(x \wedge y) = 1 \Leftrightarrow f(x) = 1 \text{ and } f(y) = 1, \text{ so } \Phi(x \wedge y) = \Phi(x) \cap \Phi(y) \\
&f(x \vee y) = 1 \Leftrightarrow f(x) = 1 \text{ or } f(y) = 1, \text{ so } \Phi(x \vee y) = \Phi(x) \cup \Phi(y)
\end{aligned}$$

Lemma 1.21 says that if $a \not\leq b$ then $\Phi(a) \not\subseteq \Phi(b)$, hence $\Phi$ is injective. So $\Phi$ is a lattice homomorphism from $L$ to the sublattice $\{\Phi(x) \,|\, x \in L\}$ of $\mathcal{P}A$.

If $L$ is Boolean, then $\Phi$ also preserves complements, so its image is a sub-Boolean algebra of $\mathcal{P}A$. □

# Chapter 2

# The Propositional Calculus

Lecture 6

The idea behind the propositional calculus is that we assume that we are given a set $P$ of 'primitive' propositions $p, q, r, \ldots$ about which we assume only that they are capable of being assigned truth-values 1 (true) or 0 (false), and we then study how these can be combined with propositions such as

| | |
|---|---|
| $(p \wedge q)$ | (p and q) |
| $(p \vee q)$ | (p or q) |
| $\neg p$ | (not p) |
| $(p \Rightarrow q)$ | (p implies q) |

The compound propositions are assigned values by 'truth tables' which tell us the value of a compound in terms of the primitive propositions involved in it.

| $p$ | $q$ | $p \wedge q$ | $p \vee q$ | $p \Rightarrow q$ | $\neg p$ | $\neg p \vee q$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 |

Note that $p \Rightarrow q$ has the same truth-table as $\neg p \vee q$, so we could define $\Rightarrow$ in terms of $\vee$ and $\neg$.

We also need two 'nullary connectives' which produce a proposition from no primitives: $\top$ (true) and $\bot$ (false), which have preassigned truth values.

We can define everything we need from $\Rightarrow$ and $\bot$:

| Observe that | | has the same truth-table as | |
|---|---|---|---|
| | $\top$ | | $(\bot \Rightarrow \bot)$ |
| | $\neg p$ | | $(p \Rightarrow \bot)$ |
| | $(p \vee q)$ | | $((p \Rightarrow \bot) \Rightarrow q)$ |
| | $(p \wedge q)$ | | $((p \Rightarrow (q \Rightarrow \bot)) \Rightarrow \bot)$ [1] |

**Definition 2.1** Let $P$ be a set of primitive propositions. The set $\mathcal{L}(P)$ of PROPOSITIONAL FORMULAE (or COMPOUND PROPOSITIONS) over $P$ is defined recursively by

(1) $p \in P \;\Rightarrow\; p \in \mathcal{L}(P)$

(2) $\bot \in \mathcal{L}(P)$

(3) $s, t \in \mathcal{L}(P) \;\Rightarrow\; (s \Rightarrow t) \in \mathcal{L}(P)$

Formally, this means that $\mathcal{L}(P)$ is the smallest subset of $\Sigma^*$ (where $\Sigma = P \cup \{\bot, \Rightarrow, (, )\}$) with the three closure properties (i) - (iii), i.e. the intersection of all subsets with these properties.

By a VALUATION of $P$ we mean a function $v : P \to \{0, 1\}$. We extend this to the unique function $\bar{v} : \mathcal{L}(P) \Rightarrow \{0, 1\}$ satisfying

$$\bar{v}(t) = \begin{cases} v(t) & \text{if } t \text{ is primitive} \\ 0 & \text{if } t = \bot \\ \bar{v}(s) \Rightarrow \bar{v}(u) & \text{if } t = (s \Rightarrow u) \end{cases}$$

[1] Determining quite why this is clear is left as an exercise for the reader

We say a compound proposition $t$ is a TAUTOLOGY if $\bar{v}(t) = 1$ for all valuations $v$ of the primitive propositions which occur in $t$.

More generally, if $S \subseteq \mathcal{L}(P)$ is a set of HYPOTHESES or PREMISES, we say $t$ is a SEMANTIC CONSEQUENCE of $S$, or that $S$ SEMANTICALLY ENTAILS $t$ if

$$\bar{v}(t) = 1 \text{ for all } v \text{ such that } \bar{v}(s) = 1 \forall s \in S.$$

We write '$S \models t$' for '$S$ semanically entails $t$'.
We abbreviate '$\emptyset \models t$' to '$\models t$' for '$t$ is a tautology'.

**Example 2.2**

(a) The formula $t = ((p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r)))$ is a tautology:

| $p$ | $q$ | $r$ | $p \Rightarrow q$ | $p \Rightarrow r$ | $(p \Rightarrow q) \Rightarrow (p \Rightarrow r)$ | $q \Rightarrow r$ | $p \Rightarrow (q \Rightarrow r)$ | $t$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

(b) We show $\{s, (s \Rightarrow t)\} \models t$ for any formulae $s, t$.

| $s$ | $t$ | $s \Rightarrow t$ |
|---|---|---|
| 0 | 0 | 1 |
| 0 | 1 | 1 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

Hence the only valuation with $\bar{v}(s) = \bar{v}(s \Rightarrow t) = 1$ has $\bar{v}(t) = 1$.

We want a 'logical calculus' which will enable us to derive all valid semantic entailments from a finite list of such enailments. The one we take (the HILBERT-STYLE axiomatization of the propositional calculus) has three AXIOMS:

| | |
|---|---|
| $(p \Rightarrow (q \Rightarrow p))$ | Axiom K |
| $((p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r)))$ | Axiom S |
| $(((p \Rightarrow \bot) \Rightarrow \bot) \Rightarrow p)$ | Axiom T (tertium non datur) |

(strictly speaking, any substitution instance of one of the three, i.e. any formula obtained by substituting arbitrary formulae $s, t, u$ for the primitive propositions $p, q, r$ is considered to be an axiom)

and we have one RULE OF INFERENCE, the rule that from $s$ and $(s \Rightarrow t)$ we may infer $t$.

**Definition 2.3** By a DEDUCTION (from a set $S$ of hypotheses) in the propositional calculus we mean a finite sequence $t_1, t_2, \ldots, t_n$ of formulae such that for each $i \leq n$ we have either

(1) $t_i$ is obtainable from (K), (S) or (T) by substitution;

(2) $t_i \in S$;

(3) there exist $j, k < i$ such that $t_k$ is the formula $(t_j \Rightarrow t_i)$.

$t_n$ is called the CONCLUSION of a deduction. If there exists a deduction of $t_n$ from $S$ then we say that $S$ SYNTACTICALLY ENTAILS $t_n$, and write $S \vdash t_n$. If $S$ is empty, we write $\vdash t_n$ and say that $t_n$ is a THEOREM of the propositional calculus. A deduction from the empty set of hypotheses is called a PROOF.

**Example 2.4** Lecture 7

(a) We show $\vdash (p \Rightarrow p)$

| | |
|---|---|
| $(p \Rightarrow (p \Rightarrow p))$ | $K[p \,/\, q]$ |
| $(p \Rightarrow ((p \Rightarrow p) \Rightarrow p))$ | $K[(p \Rightarrow p) \,/\, q]$ |
| $((p \Rightarrow ((p \Rightarrow p) \Rightarrow p)) \Rightarrow ((p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p)))$ | $S[(p \Rightarrow p), p \,/\, q, r]$ |
| $((p \Rightarrow (p \Rightarrow p)) \Rightarrow (p \Rightarrow p))$ | MP(2,3) |
| $(p \Rightarrow p)$ | MP(1,4) |

(b) We show $\{(p \Rightarrow q), (q \Rightarrow r)\} \vdash (p \Rightarrow r)$

| | |
|---|---|
| $((p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r)))$ | $S$ |
| $(q \Rightarrow r)$ | (hypothesis) |
| $((q \Rightarrow r) \Rightarrow (p \Rightarrow (q \Rightarrow r)))$ | $K[(q \Rightarrow r), p \,/\, p, q]$ |
| $(p \Rightarrow (q \Rightarrow r))$ | MP(2,3) |
| $((p \Rightarrow q) \Rightarrow (p \Rightarrow r))$ | MP(4,1) |
| $(p \Rightarrow q)$ | (hypothesis) |
| $(p \Rightarrow r)$ | MP(5,6) |

Note that if we're trying to deduce $(s \Rightarrow t)$ from a set of hypotheses $S$, it suffices to find some $u$ for which both $(s \Rightarrow u)$ and $(u \Rightarrow t)$ are deducible from $S$ since then we can 'patch together' the two deductions using the one above.

**Theorem 2.5** *(Soundness Theorem)*
*Let $S \subseteq \mathcal{L}(P)$ and $t \in \mathcal{L}(P)$. If $S \vdash t$, then $S \models t$. In particular, every theorem is a tautology.*

*Proof:* Let $(t_1, t_2, \ldots, t_n = t)$ be a deduction of $t$ from $S$, and let $v : P \to \{0,1\}$ be a valuation such that $\bar{v}(s) = 1 \; \forall s \in S$ (a MODEL for $S$). We show that $\bar{v}(t_i) = 1 \; \forall i$ by induction on $i$.

If $t_i$ is an axiom, then $\bar{v}(t_i) = 1$ since the axioms are tautologies.

If $t_i \in S$ then $\bar{v}(t_i) = 1$ by assumption.

If there exist $j, k < i$ such that $t_k = (t_j \Rightarrow t_i)$, then by induction hypothesis $\bar{v}(t_j) = 1$ and $\bar{v}(t_k) = 1$. Then $\bar{v}(t_j) = 1$ and $\bar{v}(t_i) = 0$ would imply $\bar{v}(t_k) = 0$, so $\bar{v}(t_i) = 1$. □

**Theorem 2.6** *(Deduction Theorem)*
*Let $S \subseteq \mathcal{L}(P)$, $s, t \in \mathcal{L}(P)$. Then $S \vdash (s \Rightarrow t)$ if and only if $S \cup \{s\} \vdash t$.*

*Proof:* If we have a deduction of $(s \Rightarrow t)$ from $S$, we can add the two lines

| | |
|---|---|
| $s$ | (hypothesis) |
| $t$ | (MP) |

to get a deduction of $t$ from $S \cup \{s\}$.

Conversely, suppose $(t_1, t_2, \ldots, t_n = t)$ is a deduction of $t$ from $S \cup \{s\}$. We replace each line in the deduction by a sequence of lines in a deduction from $S$, as follows:

If $t_i$ is an axiom or $t_i \in S$, we write down

| | |
|---|---|
| $t_i$ | (axiom or hypothesis) |
| $(t_i \Rightarrow (s \Rightarrow t_i))$ | (K) |
| $(s \Rightarrow t_i)$ | (MP) |

If $t_i = s$, we write down the deduction of $(s \Rightarrow s)$ from Example 2.4 (a).

If $t_k = (t_j \Rightarrow t_i)$ for some $j, k < i$, then we already have $(s \Rightarrow t_j)$ and $(s \Rightarrow t_k)$ in our deduction. We now add

| | |
|---|---|
| $((s \Rightarrow (t_j \Rightarrow t_i)) \Rightarrow ((s \Rightarrow t_j) \Rightarrow (s \Rightarrow t_i)))$ | $S[s, t_j, t_i \ / \ p, q, r]$ |
| $((s \Rightarrow t_j) \Rightarrow (s \Rightarrow t_i))$ | (MP) |
| $(s \Rightarrow t_i)$ | (MP) |

□

For example, if we want to show

$$\{(p \Rightarrow q), (q \Rightarrow r)\} \vdash (p \Rightarrow r),$$

it's sufficient to show

$$\{p, (p \Rightarrow q), (q \Rightarrow r)\} \vdash r,$$

but we can do this by writing the three hypotheses and applying MP twice.

Using the Deduction Theorem, we can reduce the proof that $S \models t$ implies $S \vdash t$ to the particular case $t = \bot$ for if $S \models t$, then $S \cup \{\neg t\}$ has no models, so $S \cup \{\neg t\} \models \bot$.

If we could deduce $S \cup \{\neg t\} \models \bot$, then by the Deduction Theorem we'd have $S \vdash \neg(\neg t)$. To this deduction, we could append the lines

| | |
|---|---|
| $(\neg(\neg t) \Rightarrow t)$ | (T) |
| $t$ | (MP) |

and hence show $S \vdash t$.

**Lemma 2.7** *(Adequacy Theorem) If $S \models \bot$, then $S \vdash \bot$.*

*Proof:* We prove the contra-positive: if $S \nvdash \bot$ (i.e. $S$ is CONSISTENT) then $S \not\models \bot$ (i.e. $S$ has a model).

Consider the set $C$ of all consistent subsets of $\mathcal{L}(P)$ ordered by inclusion. We claim that $C$ is chain-complete. If $\{S_i | i \in I\}$ is a chain in $C$, then $\bigcup_{i \in I} S_i$ is consistent, since a deduction from it uses only finitely many members of $\bigcup_{i \in I} S_i$, all of which must belong to some particular $S_i$. Hence by Zorn's Lemma there is a maximal consistent set $\bar{S}$ with $S \subseteq \bar{S}$. What can we say about such a set?

(a) $\bar{S}$ is DEDUCTIVELY CLOSED, so if $\bar{S} \vdash t$ then $t \in \bar{S}$. If $t \notin \bar{S}$ then $\bar{S} \cup \{t\} \vdash \bot$, but if we also have $\bar{S} \vdash t$ this would yield $\bar{S} \vdash \bot$.

(b) For any $t$, either $t \in \bar{S}$ or $\neg t \in \bar{S}$. If $t \notin \bar{S}$ then $\bar{S} \cup \{t\} \vdash \bot$, so $\bar{S} \vdash \neg t$ by the Deduction Theorem, so $\neg t \in \bar{S}$ by (i).

Now consider the mapping $f : \mathcal{L}(P) \to \{0, 1\}$ defined by

$$f(t) = \begin{cases} 1 & \text{if } t \in \bar{S} \\ 0 & \text{if } t \notin \bar{S} \end{cases}$$

Clearly $f(\bot) = 0$ since $\bar{S}$ is consistent. We claim that, for any $s$ and $t$, we have $f(s \Rightarrow t) = f(s) \Rightarrow_2 f(t)$, so that $f$ is the unique extension of its restriction $v : P \to 2$ to the set of primitive propositions. Then $\bar{v}(s) = 1 \ \forall s \in S$, since $S \subseteq \bar{S}$, so $v$ is a model of $S$.

Case 1: Suppose $t \in \bar{S}$. Then $(s \Rightarrow t) \in \bar{S}$ since we have a deduction

$$\begin{array}{l} t \\ (t \Rightarrow (s \Rightarrow t)) \\ (s \Rightarrow t) \end{array}$$

So we have $f(s \Rightarrow t) = 1 = (f(s) \Rightarrow_2 f(t))$.

Case 2: Suppose $s \notin \bar{S}$. Then $\neg s \in \bar{S}$, and we have a deduction of $(s \Rightarrow t)$ from $\{\neg s\}$ (Q4 on sheet 2).
So $f(s \Rightarrow t) = 1 = (f(s) \Rightarrow_2 f(t))$.

Case 3: Suppose $s \in \bar{S}$, $t \notin \bar{S}$. Then $(s \Rightarrow t) \notin \bar{S}$ since $\{s, (s \Rightarrow t)\} \vdash t$.
So $f(s \Rightarrow t) = 0 = (f(s) \Rightarrow_2 f(t))$.

Hence $f$ is a homomorphism with respect to $\Rightarrow$ and $\bot$, and hence it is the unique extension $\bar{v}$ of the valuation $v = f|_P$. So $\bar{v}(s) = 1$ for all $s \in S$ since $S \subseteq \bar{S}$. □

**Remarks**

(a) For any valuation $v : P \to 2$, the set $\{t \in \mathcal{P} | \bar{v}(t) = 1\}$ is maximal consistent. It's deductively closed by the proof of Theorem 2.5, and hence consistent since it doesn't contain $\bot$. Also, for any $t$, either $t$ or $\neg t$ belongs to the set, so it's maximal consistent. Hence valuations of $P$ correspond bijectively to maximal consistent subsets of $\mathcal{L}(P)$.

(b) If $P$ is countable (which implies that $\mathcal{L}(P)$ is countable) then we can prove Lemma 2.7 without using Zorn's Lemma. Simply enumerate the members of $\mathcal{L}(P)$ as $t_1, t_2, t_3, \ldots$ and then, for each $i$, add $t_i$ to the set we've got so far if we can consistently do so, otherwise add $\neg t_i$. In this way we can enlarge any given consistent set to a maximal one.

(c) Lemma 2.7 is closely related to Lemma 1.21, which said that a distributive lattice has plenty of homomorphisms to 2. $\mathcal{L}(P)$ isn't a Boolean algebra, but it has a relation $\leq$ defined by $s \leq t$ iff $\{s\} \vdash t$ which is reflexive and transitive. If we form the quotient $\mathcal{L}(P)/\equiv$, where $s \equiv t$ means ($s \leq t$ and $t \leq s$) then we get a Boolean algebra, and any homomorphism $\mathcal{L}(P) \to 2$ must factor through to the quotient map $\mathcal{L}(P) \to \mathcal{L}(P)/\equiv$.

**Theorem 2.8** *(Completeness Theorem)*
*For any $S \subseteq \mathcal{L}(P)$ and any $t \in \mathcal{L}(P)$, we have $s \models t$ iff $s \vdash t$.*

*Proof:* ($\Rightarrow$) is Theorem 2.5.
($\Leftarrow$) Suppose $s \models t$. Then $S \cup \{\neg t\} \models \bot$, so $S \cup \{\neg t\} \vdash \bot$ by Lemma 2.7. Hence $S \vdash \neg\neg t$ by Theorem 2.6, and $\{\neg\neg t\} \vdash t$ using axiom $(T)$, so $S \vdash t$ □

**Corollary 2.9** *(Compactness Theorem)*
*If $S \models t$, then there exists a finite $S' \subseteq S$ such that $S' \models t$.*

*Proof:* This is obvious for $\vdash$, since a deduction of $t$ from $S$ uses only finitely many members of $S$. □

In particular, Corollary 2.9 asserts that if every finite subset of $S$ has a model, then $S$ has a model.

**Example 2.10**
We show that any partial ordering on a set can be extended to a total ordering. Given a poset $(A, \leq)$ we consider a set $P = \{p_{x,y} | x, y \in A\}$ of primitive propositions. (Think of $p_{x,y}$ as the assertion that $x \leq y$ in the desired total ordering). Take $S$ to consist of

$$\{p_{x,y} | x \leq y \text{ in } A\} \cup \{(p_{x,y} \Rightarrow (p_{y,z} \Rightarrow p_{x,z})) | x, y, z \in A\} \cup$$
$$\cup \{(p_{x,y} \Rightarrow (p_{y,x} \Rightarrow \bot)) | x, y \in A, x \neq y\} \cup \{((p_{x,y} \Rightarrow \bot) \Rightarrow p_{y,x}) | x, y \in A\}$$

Then a model of $S$ 'is' a total ordering of $A$ which extends $\leq$. Let $S' \subseteq S$ be a finite subset. Then the set $P'$ of primitive propositions involved in members of $S'$ is also finite, and so is the set $A' = \{x \in A | \exists y \in A$ such that $p_{x,y} \in P'$ or $p_{y,x} \in P'\}$.

A total ordering of $A'$ which extends $\leq |_{A'}$ is certainly a model of $S'$. So we've reduced to proving the result for a finite poset $A'$, but we can do this by considering each pair $(x, y) \in A' \times A'$ in turn and deciding whether to add it or $(y, x)$ to the relation we have so far.

**Example 2.11**
A GRAPH consists of a set $V$ of vertices and a binary relation $E \subseteq V \times V$ which is irreflexive and symmetric. An N-COLOURING of a graph $(V, E)$ is a function $V \xrightarrow{f} \{1, 2, \ldots, n\}$ such that $f(x) \neq f(y)$ whenever $(x, y) \in E$.

**Theorem 2.12** *A graph is $n$-colourable iff all its finite subgraphs are $n$-colourable*

*Proof:* Given $(V, E)$ we write down a propositional theory $S$ over a set $P = \{p_{x,i} | x \in V\ ,\ i \in \{1, 2, \ldots, n\}\}$ of primitive propositions.

$$S = \{(p_{x,i} \Rightarrow (p_{x,j} \Rightarrow \bot)) \,|x \in V\ ,\ i \neq j\} \cup \{\bigvee_{i=1}^{n} p_{x,i} | x \in V\} \cup$$
$$\cup \{(p_{x,i} \Rightarrow (p_{y,i} \Rightarrow \bot)) \,|(x, y) \in E\ ,\ 1 \leq i \leq n\}$$

Then models of $S$ are $n$-orderings of $(V, E)$. But, for any finite $S' \subseteq S$, we can find a finite $V' \subseteq V$ such that an $n$-colouring of $(V', E \cap V' \times V')$ yields a model of $S'$. □

**Corollary 2.13** *(Decidability Theorem)*
*Given a finite set $P$ of primitive propositions and a finite $S \subseteq \mathcal{L}(P)$, there's an algorithm which determines whether or not $S \vdash t$ for any given $t \in \mathcal{L}(P)$.*

*Proof:* This is obvious for $\models$ (truth tables). □

# Chapter 3

# First Order Predicate Calculus

Lecture 9

We want a first-order language adequate for talking about mathematical structures. What features should the language have?

Examples of structures include things like groups and rings, where we have an underlying set $A$ equipped with operations $A^n \to A$ (e.g. a group has $M : A^2 \to A$ , $i : A^1 \to A$ and $e : A^0 \to A$ corresponding to multiplication, inverse and the identity element). We also want to study structures such as posets, where the primitive structure is given by RELATIONS, i.e. subsets of $A^n$ for some $n$ (e.g. $\leq \subseteq A^2$).

Corresponding to these our language will have sets $\Omega$ and $\Pi$ of OPERATION SYMBOLS and PREDICATE SYMBOLS, each equipped with a mapping to $\mathbb{N}$ specifying the ARITY of each operation/predicate. We assume $\Omega$ and $\Pi$ are disjoint, and write $\alpha : \Omega \to \mathbb{N}$ and $\alpha : \Pi \to \mathbb{N}$ for the arity mappings. An operation symbol of arity 0 is called a CONSTANT, and a predicate symbol of arity 0 is called a PRIMITIVE PROPOSITION.

**Definition 3.1** The FIRST ORDER LANGUAGE $\mathcal{L}(\Omega, \Pi)$ over the signature $(\Omega, \Pi)$ is defined as follows: we set $\Sigma = \Omega \cup \Pi \cup \{v, ', =, \bot, \Rightarrow, \forall, (, ), , \}$ and consider the following subsets of $\Sigma^*$

- The VARIABLES of $\mathcal{L}(\Omega, \Pi)$ are the smallest set $V$ such that $v \in V$ and if $x \in V$ then $x' \in V$. Variables are $v, v', v'', v''', \ldots$ but in practise we denote variables by letters $x, y, z$ or $x_1, x_2, x_3, \ldots$
- The TERMS of $\mathcal{L}(\Omega, \Pi)$ are the smallest set $T$ such that
  (1) $V \subseteq T$
  (2) If $\omega \in \Omega$, $\alpha(\omega) = n$ and $t_1, t_2, \ldots, t_n \in T$ then $\omega t_1 t_2 \ldots t_n \in T$.

  Note that brackets are not needed for this notation: any term has an unambiguous interpretation as a variable or an operation symbol applied to a string of terms. For example, if $M$ is a binary operation symbol, then $MxMyz$ and $MMxyz$ are distinct terms. To say that they are equal is to say that the binary operation denoted by $M$ is associative.

  Note that any constant symbol is a term.
- The ATOMIC FORMULAE of $\mathcal{L}(\Omega, \Pi)$ are of 2 kinds:
  (1) If $\phi \in \Pi$, $\alpha(\phi) = n$ and $t_1, t_2, \ldots, t_n$ are terms, then $\phi t_1 t_2 \ldots t_n$ is an atomic formula.
  (2) If $s$ and $t$ are terms, then $(s = t)$ is an atomic formula.

Finally, the language $\mathcal{L}(\Omega, \Pi)$ is the smallest subset of $\Sigma^*$ such that

(a) Every atomic formula is in $\mathcal{L}$

(b) $\bot \in \mathcal{L}$, and if $p, q \in \mathcal{L}$ then $(p \Rightarrow q) \in \mathcal{L}$

(c) If $p \in \mathcal{L}$ and $x$ is a variable which occurs in $p$ as a free variable, then $(\forall x)\, p \in \mathcal{L}$

FREE and BOUND variables are defined as follows: every variable occurring in a term or atomic formula is free, any variable occurring free in $p$ of in $q$ is free in $(p \Rightarrow q)$, and any variable other than $x$ which occurs free in $p$ is also free in $(\forall x)\, p$, but every free occurence of $x$ in $p$ becomes bound in $(\forall x)\, p$.

The formula $((x = y) \Rightarrow (\forall x)\,(x = x))$ contains $x$ as both a bound and a free variable.

We'll see that the formulae $(\forall x)\, p$ and $(\forall y)\, p[y/x]$ will have the same interpretation in any $(\Omega, \Pi)$ structure where $y$ is a variable not occurring in $p$ and $p[y/x]$ means '$p$ with $y$ substituted for all free occurrences of $x$'. So we can always assume that the names of bound variables don't clash with those of free variables.

| We write | | as shorthand for | |
|---|---|---|---|
| | $\top$ | | $(\bot \Rightarrow \bot)$ |
| | $\neg p$ | | $(p \Rightarrow \bot)$ |
| | $(p \vee q)$ | | $((p \Rightarrow \bot) = q)$ |
| | $(p \wedge q)$ | | $((p \Rightarrow (q \Rightarrow \bot)) \Rightarrow \bot)$ |
| | $(\exists x)$ p | | $\neg\,(\forall x)\,\neg p$ |

By an $(\Omega, \Pi)$-STRUCTURE we mean a set $A$ equipped with distinguished mappings $\omega_A : A^{\alpha(\omega)} \to A$ for each $\omega \in \Omega$ and $\phi_A : A^{\alpha(\phi)} \to 2$ for each $\phi \in \Pi$ (or equivalently distinguished subsets $[\phi]_A \subseteq A^{\alpha(\phi)}$ for each $\phi \in \Pi$).

By a CONTEXT we mean a finite list $(x_1, x_2, \ldots, x_n) = \vec{x}$ of distinct variables. A TERM-IN-CONTEXT is an expression $\vec{x}.t$ where$t$ is a term and $\vec{x}$ is a context including all the variables which occur in $t$, and a FORMULA-IN-CONTEXT is an expression $\vec{x}.p$ where $p$ is a formula and $\vec{x}$ includes all the free variables of $p$.

**Definition 3.2** Given an $(\Omega, \Pi)$-structure $A$, we interpret terms and formulae in context as follows:

If $\vec{x} = (x_1, x_2, \ldots, x_n)$ then $(\vec{x}.t)_A$ is a mapping $A^n \to A$ given by $(\vec{x}.x_i)_A$ is the mapping

$$(a_1, a_2, \ldots, a_n) \mapsto a_i$$

$(\vec{x}.\omega t_1 \ldots t_n)_A$ is the composite

$$A^n \xrightarrow{\left((\vec{x}.t_1)_A, \ldots, (\vec{x}.t_m)_A\right)} A^m \xrightarrow{\omega_A} A$$

We interpret formulae-in-context $(\vec{x}.p)$ as mappings $(\vec{x}.p)_A : A^n \to 2$ (or equivalently as subsets $[\vec{x}.p]_A \subseteq A^n$) by the following rules: Lecture 10

- If $p$ is $\phi t_1 t_2 \ldots t_m$ for some $\phi \in \Pi$ (with $\alpha(\phi) = m$), then $(\vec{x}.p)_A$ is

$$A^n \xrightarrow{\left((\vec{x}.t_1)_A, \ldots, (\vec{x}.t_m)_A\right)} A^M \xrightarrow{\phi_A} 2$$

- If $p$ is $(s = t)$, then $(\vec{x}.p)_A$ is

$$A^n \xrightarrow{\left((\vec{x}.s)_A, (\vec{x}.t)_A\right)} A^2 \xrightarrow{f} 2$$

  where $f(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases}$

- $(\vec{x}.\bot)_A$ is the constant function with value 0.

- If $p$ is $(q \Rightarrow r)$, then $(\vec{x}.p)_A$ is

$$A^n \xrightarrow{\left((\vec{x}.q)_A, (\vec{x}.r)_A\right)} 2^2 \xrightarrow{\Rightarrow_2} 2$$

  where $\Rightarrow_2$ is the mapping given by the usual truth-table for implication.

- If $p$ is $(\forall x_{n+1})\, q$ then we have an interpretation

$$f : A^n \times A \xrightarrow{(\vec{x}, x_{n+1} . q)_A} 2$$

We can regard this as a mapping $A^n \to 2^A$, where $2^A$ is the set of all functions $A \to 2$, sending $(a_1, \ldots, a_n)$ to the mapping

$$a_{n+1} \mapsto f(a_1, \ldots, a_n, a_{n+1})$$

and we define $(\vec{x}.p)$ to be the composite

$$A^n \longrightarrow 2^A \xrightarrow{g} 2 \quad \text{where } g(h) = \begin{cases} 1 & \text{if } h \text{ is the constant function with value 1} \\ 0 & \text{otherwise} \end{cases}$$

**Lemma 3.3** *(Substitution Lemma)*
*Let $A$ be a $(\Omega, \Pi)$-structure, $\vec{x}$ a context and $\left\{ \begin{smallmatrix} t \\ p \end{smallmatrix} \right\}$ a $\left\{ \begin{smallmatrix} term \\ formula \end{smallmatrix} \right\}$ whose (free) variables are in the string $\vec{x}$. Let $\vec{s} = (s_1, s_2, \ldots, s_n)$ be a string of terms of the same length as $\vec{x}$, and $\vec{y}$ a context conatining all the variables in the $s_i$. Then $\vec{x}.\left\{ \begin{smallmatrix} t \\ p \end{smallmatrix} \right\}[\vec{s}/\vec{x}]$ is the composite*

$$A^n \xrightarrow{\left((\vec{y}.s_1)_A, \ldots, (\vec{y}.s_n)_A\right)} A^n \begin{cases} \xrightarrow{(\vec{x}.t)_A} A & \textit{(term)} \\ \xrightarrow{(\vec{x}.p)_A} 2 & \textit{(formula)} \end{cases}$$

NB: Some (or all) of the terms $s_i$ may be the corresponding $x_i$. If all $s_i$ are $x_i$ then $\left\{ \begin{smallmatrix} t \\ p \end{smallmatrix} \right\}$ is unchanged, but the context can be reordered or enlarged. The special case of the Lemma where all $s_i$ are $x_i$ is called the Weakening Lemma.

*Proof:* By induction on the structure of $\left\{ \begin{smallmatrix} t \\ p \end{smallmatrix} \right\}$. Clearly true if $t$ is $x_i$ for some $i$, and all other cases follow by associativity of composition. □

Lemma 3.3 ensures that if we know the interpretation of a term or formula in its CANONICAL CONTEXT (i.e. the list of all (free) variables in the term or formula in the order of their first occurence) then we know its interpretation in every context.

A formula with no free variables is called a CLOSED FORMULA or SENTENCE. It will be interpreted, in a given structure $A$, as a map $A^0 \to 2$, i.e. as a truth-value.

By a THEORY over the signature $(\Omega, \Pi)$ we mean a set $T$ of sentences in $\mathcal{L}(\Omega, \Pi)$, called the (non-logical) AXIOMS of the theory. We say a structure $A$ is a model of $T$ if $(p)_A = 1$ for every $p \in T$.

**Example 3.4** The **theory of groups** has $\Omega = \{m, i, e\}$ with $\alpha(m) = 1, \alpha(i) = 1, \alpha(e) = 0$ and $\Pi = \emptyset$, and its axioms are

| | |
|---|---|
| $(\forall x)\,(\forall y)\,(\forall z)\,(mxmyx = mmxyz)$ | (associativity) |
| $(\forall x)\,(mex = x)$ | (identity) |
| $(\forall x)\,(mixx = e)$ | (inverse) |

Note that any structure satisfying these axioms also satisfies $(\forall x)\,(mxe = x)$ and $(\forall x)\,(mxix = e)$

**Example 3.5** The **theory of graphs** has $\Omega = \emptyset$, $\Pi = \{c\}$ with $\alpha(c) = 2$ (c being the relation that 2 vertices are connected) and axioms

| | |
|---|---|
| $(\forall x)\,\neg cxx$ | (irreflexive) |
| $(\forall x)\,(\forall y)\,(cxy \Rightarrow cyx)$ | (symmetric) |

We say a sentence $p$ is a SEMANTIC CONSEQUENCE of a theory $T$, and write $T \models p$, if every model of $T$ also satisfies $p$.

If we wish to talk about semantic entailment for formulae with free variables, we assume the variables all lie in some context $\vec{x}$, and we write $T \models_{\vec{x}} p$ to mean that, given any structure $A$ and any $n$-tuple $(a_1, \ldots, a_n)$ of elements of $A$ which make $(\vec{x}.q)_A\,(a_1, \ldots, a_n) = 1$ for all $q \in T$, then we also have $(\vec{x}.p)_A\,(a_1, \ldots, a_n) = 1$.

We now set up a deduction system for the predicate calculus: it operates with formulae-in-context rather than formulae. A with the propositional calculus, we have a list of (logical) axioms which may be assumed in any deduction, and rules of inference for deducing new formulae from old.

The axioms are:

| | |
|---|---|
| $\vec{x}.\,(p \Rightarrow (q \Rightarrow p))$ | Here $\vec{x}$ may be any suitable |
| $\vec{x}.\,((p \Rightarrow (q \Rightarrow r)) \Rightarrow ((p \Rightarrow q) \Rightarrow (p \Rightarrow r)))$ | context, and $p, q, r$ are any |
| $\vec{x}.\,(\neg\neg p \Rightarrow p)$ | formulae on $\mathcal{L}(\Omega, \Pi)$. |
| $\vec{x}.\,((\forall y)\,p \Rightarrow p[t/y])$ | $\vec{x}$ any context including the free variables of $p$ other than $y$, and the variables of $t$. |
| $\vec{x}.\,((\forall y)\,(p \Rightarrow q) \Rightarrow (p \Rightarrow (\forall y)\,q))$ | $y$ not free in $p$. |
| $(\forall x)\,(x = x)$ | For some variable $x$. |
| $\vec{z}.\,(\forall x)\,(\forall y)\,((x = y) \Rightarrow (p \Rightarrow p[y/x]))$ | $p$ any formula whose free variables are in $\vec{z}, x, y$. |

Our rules of inference are:

(1) MODUS PONENS: From $\vec{x}.p$ and $\vec{x}.(p \Rightarrow q)$ we may infer $\vec{x}.q$

(2) WEAKENING: From $\vec{x}.p$ we may infer $\vec{y}.p$ for any $\vec{y}$ which contains all the variables in $\vec{x}$

(3) GENERALIZATION: From $\vec{x}, y.p$ we may infer $\vec{x}.\,(\forall y)\,p$, provided that $y$ doesn't appear free in any hypothesis used in deducing $\vec{x}, y.p$

Lecture 11

**Theorem 3.6** *(Soundness Theorem)*
*Let $S$ be a set of formulae of $\mathcal{L}(\Omega, \Pi)$, $p$ a particular formula and $\vec{x}$ a context including the free variables of $p$ and of the members of $S$. If $S \vdash_{\vec{x}} p$, then $S \models_{\vec{x}} p$.*

*Proof:* Just like the proof of Theorem 2.5: the axioms of the predicate calculus are all tautologies, and the rules of inference derive valid conclusions if their hypotheses are valid. □

**Theorem 3.7** *(Deduction Theorem)*
*Let $S$ be a set of formulae, $p$ and $q$ two formulae and $\vec{x}$ a suitable context. Then*

$$S \vdash_{\vec{x}} (p \Rightarrow q) \quad \textit{iff} \quad S \cup \{p\} \vdash_{\vec{x}} q$$

*Proof:* This can mostly be copied from the proof of Theorem 2.6. For $\Leftarrow$, suppose we have a deduction $\vec{z}_1.q_1, \vec{z}_2.q_2, \ldots, \vec{z}_n.q_n = \vec{x}.q$ of $q$ from $S \cup \{p\}$ with the context $\vec{x}$. Suppose that $\vec{z}_i.q_i$ is obtained by the rule of generalization, i.e. $q_i$ is the formula $(\forall y)q_j$ for some $j < i$ and $\vec{v}_j = \vec{z}_i, y$. Then by induction we have a deduction of $\vec{z}_i, y.(p \Rightarrow q_j)$ from $S$, and hence we can obtain $\vec{z}_i.(\forall y)(p \Rightarrow q_j)$. If $y$ isn't free in $p$, we can deduce $\vec{z}_i.(p \Rightarrow (\forall y)q_j)$ which is what we want. □

If $y$ is free in $p$, we couldn't have used $p$ as a hypothesis in the deduction of $q_j$, so we have a deduction of $\vec{z}_i, y.q_j$ from $S$ and we can further deduce $\vec{z}_i.(\forall y)q_j$ whence we obtain $\vec{z}_i'.(p \Rightarrow q_i)$ for a context $\vec{z}_i'$ containing $\vec{z}_i$ plus any other variables free in $p$.

**Theorem 3.8** *(Completeness Theorem)*

$$S \vdash_{\vec{x}} p \quad \textit{iff} \quad S \models_{\vec{x}} p$$

*Proof:* ($\Rightarrow$) is just Theorem 3.6.

($\Leftarrow$) First reduce to the case when the context $\vec{x}$ is empty (so that all formulae in $S \cup \{p\}$ are sentences) by adding a string $\vec{c} = (c_1, c_2, \ldots, c_n)$ of new constants to the language and substituting them for the variables $x_1, \ldots, x_n$. We get $S' \models p'$ where $S'$ and $p'$ denote the formlae obtained by making these substitutions. If we can show that $s' \vdash p'$ then we can convert to a deduction $S \vdash_{\vec{x}} p$.

Next we use Theorem 3.7 to reduce to the case $p = \bot$. We now show the contrapositive, that if $S$ is a set of sentences such that $S \nvdash \bot$, then $S$ has a model.

*Step 1:* We can enlarge $S$ to a maximal consistent set of sentences, $S_1$, then as in the propositional case $S_1$ is deductively closed and for every sentence $p$ we have either $p \in S$ or $\neg p \in S$.

*Step 2:* Suppose $S_1 \vdash (\exists x)q$, where $q$ is a formula with one free variable $x$. We may add a new constant $c_q$ to $\mathcal{L}$, and add the sentence $q[c_q/x]$ to $S_1$ and the result will still be consistent, for if $S_1 \cup \{q[c_q/x]\} \vdash \bot$, then $S_1 \vdash \neg q[c_q/x]$. Since no member of $S_1$ mentions $c_q$, we may convert this into a deduction of $x.\neg q$ from $S_1$, whence we get $S_1 \vdash (\forall x)\neg q$ by generalization. But $S_1 \vdash \neg(\forall x)\neg q$ by assumption. So $S_1 \vdash \bot$, which gives a contradiction. Hence we can enlarge $S_1$ to a theory $S_2$ (in an enlarged language $\mathcal{L}_2$) such that whenever $S_1 \vdash (\exists x)q$, there exists a constant $c_q$ in $\mathcal{L}_2$ such that $S_2 \vdash q[c_q/x]$.

We now define theories $S_n$ in languages $\mathcal{L}_n$ for all $n \geq 3$.

- For odd $n$, $\mathcal{L}_n = \mathcal{L}_{n-1}$ and $S_n$ is the maximal consistent consistent extension of $S_{n-1}$.
- For even $n$, $\mathcal{L}_n$ is obtained by adding new constants to $\mathcal{L}_{n-1}$ and $S_n$ is obtained by adding sentences of the form $q[c_q/x]$ whenever $S_{n-1} \vdash (\exists x)q$.

Let $\mathcal{L}_\infty = \bigcup_{n=1}^{\infty} \mathcal{L}_n$ and $S_\infty = \bigcup_{n=1}^{\infty} S_n$. Then $S_\infty$ has the properties that all sentences are either provable or refutable and all existential sentences have witnesses.

Let $C$ be the set of all closed terms (ie terms with no variables) in $\mathcal{L}_\infty$. Define $A = C/\equiv$ where $s \equiv t$ iff $S_\infty \vdash (s = t)$. (Need to check that this is an equivalence relation on closed terms). We make $A$ into a structure for $(\Omega_\infty, \Pi_\infty)$: if $\omega \in \Omega_\infty, \alpha(\omega) = n$ and $[t_1], [t_2], \ldots, [t_n] \in A$, then $\omega_A([t_1], [t_2], \ldots, [t_n]) = [\omega t_1 t_2 \ldots t_n]$. Similarly if $\phi \in \Pi_\infty, \alpha(\phi) = n$, then

$$([t_1], [t_2], \ldots, [t_n]) \in [\phi]_A \quad \Leftrightarrow \quad S_\infty \vdash \phi t_1 t_2 \ldots t_n$$

Now we show by induction that if $p$ is any formula of $\mathcal{L}_\infty$, then

$$([[t_1], [t_2], \ldots, [t_n]) \in [\vec{x}.p]_A \quad \Leftrightarrow \quad S_\infty \vdash p\,[\vec{t}/\vec{x}]$$

In particular if $p$ is a sentence and $S_\infty \vdash p$, then $[p]_A$ is the whole of $A^0 = 1$, i.e. $p$ is valid in $A$. So $A$ is a model of $S_\infty$, and hence a model of $S$. □

**Corollary 3.9** *(Compactness Theorem) Let $S$ be a set of sentences. If every finite subset of $S$ has a model then $S$ has a model.*

*Proof:* If not, then $S \models \bot$, so $S \vdash \bot$, so there exists a finite $S' \subseteq S$ such that $S' \vdash \bot$. But this gives $S' \models \bot$, a contradiction. □

**Corollary 3.10** *(Löwenstein-Skolem Theorem)* Lecture 12
*If a first-order theory $T$ has an infinite model, then it has arbitrarily large models. Explicitly, for any set $I$ there exists a model $A$ together with an injection $I \to A$.*

*Proof:* To the language of $T$, we add a set $\{c_i : i \in I\}$ of new constants, and we let $T^* = T \cup \{\neg(c_i = c_j) : i, j \in I, i \neq j\}$. Then a $T^*$-model is explicitly a $T$-model equipped with an injection from $A$. By compactness, we know $T^*$ has a model provided every finite subset of it does, and the given infinite $T$-model provides a model for any finite subset of $T^*$ □

There's also a 'downward' L-S theorem: if $T$ has an infinite model and its language $\mathcal{L}(\Omega, \Pi)$ is countable (i.e. $\Omega \cup \Pi$ is countable) then $T$ has a countably infinite model. To prove this, we adjoin $\{c : n \in \mathbb{N}\}$ and axioms $\neg(c_n = c_m)$ for all $m \neq n$. This theory is consistent, and we use the technique of the proof of 3.8 to construct a model of it. This model is constructed from the closed terms of a countable language $\mathcal{L}_\infty$, so it's countable.

The L-S theorems say that if $A$ is an infinite stucture, then there can't be a first-order theory whose only model (up to isomorphism) is $A$.

G. Peano (1890) wrote down a system of 'postulates' which characterize $\mathbb{N}$ up to isomorphism:

| | |
|---|---|
| 0 is a natural number | $\Omega = \{0, s\}$ with $\alpha(0) = 0, \alpha(s) = 1$ |
| If $n$ is a natural number, so is its successor $sn$ | |
| 0 is not a successor | $(\forall n)\neg(sn = 0)$ |
| Distinct numbers have distinct successors | $(\forall m, n)((sm = sn) \Rightarrow (m = n))$ |
| If $N'$ is any subset of $\mathbb{N}$ such that $0 \in N'$ and whenever $n \in \mathbb{N}$ we have $sn \in \mathbb{N}$, then $n' = \mathbb{N}$ | Could take a scheme of axioms $((p[0/n] \wedge (\forall n)(p \Rightarrow p[sn/n])) \Rightarrow (\forall n)p)$ for any $p$ with one free variable $n$. But this doesn't ensure that the induction postulate holds for *all* subsets of $\mathbb{N}$. |

Unlike the propositional calculus, there is no decidibility theorem for the predicate calculus. In the next chapter we'll construct a theory $T$ such that there is no algorithm to decide whether or not $T \vdash p$ for a given $p$.

# Chapter 4

# Recursive Functions

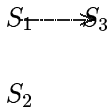
We want a precise mathematical formulation of the notion of an 'algorithmically-computable function'. We'll deal for simplicity with functions $\mathbb{N}^k \to \mathbb{N}$ for some $k \geq 1$. In order to do this, we need a mathematical model of a computer. Turing (1937) proposed the first such model. We'll use the notion of a register machine, due to M. Minsky.

**Definition 4.1** A REGISTER MACHINE has a sequence $R_1, R_2, \ldots$ of REGISTERS, each of which can store an arbitrary natural number. A PROGRAM for a register machine consists of a finite sequence of STATES $S_1, S_2, \ldots, S_k$ together with, for each $i > 0$, an instruction to be obeyed in state $i$, either

- add 1 to $R_j$ and move to state $S_k$
- if possible, subtract 1 from $R_j$ and move to state $S_k$, otherwise move to state $S_l$

$S_0$ is the halt state, and by convention $S_1$ is the starting state.

We represent programs by flow diagrams, e.g.




This computes a function of one variable, namely $f(n) = 2^n$.

Note that in general the function computed by a program may be a partial function: we say that $f : \mathbb{N}^k \to \mathbb{N}$ is COMPUTABLE if there exists a register machine program P such that P starts in state 1 with $r_1, r_2, \ldots, r_k$ in $R_1, R_2, \ldots, R_k$ and all other registers empty, then P will reach state 0 with $f(r_1, r_2, \ldots, r_k)$ in $R_1$ if $f(r_1, r_2, \ldots, r_k)$ is defined and will never reach state 0 if $f(r_1, r_2, \ldots, r_k)$ is undefined.

We can alternatively represent P by a finite sequence of instructions, e.g.

$$(1, -, 2, 3), (3, +, 1), (1, +, 4), (3, -, 5, 0), \ldots$$

There are only countably many programs, so there are only countably many computable functions — not every function $\mathbb{N} \to \mathbb{N}$ is computable.

**Theorem 4.2**

*(1) For any $i \leq k$, the projection function $\mathbb{N}^k \xrightarrow{\pi_i} \mathbb{N}$ defined by $\pi_i(n_1, n_1, \ldots, n_k) = n_i$ is computable.*

*(2) The constant function $\mathbb{N} \to \mathbb{N}$ with value 0, and the successor function $\mathbb{N} \xrightarrow{s} \mathbb{N}$ are computable.*

*(3) If $f : \mathbb{N}^k \rightharpoonup \mathbb{N}$ and $g_1, \ldots, g_k : \mathbb{N}^l \rightharpoonup \mathbb{N}$ are all computable, then so is the composite*

$$\mathbb{N}^l \xrightarrow{(g_1, g_2, \ldots, g_k)} \mathbb{N}^k \xrightarrow{f} \mathbb{N}$$

*(4) If $f : \mathbb{N}^k \to \mathbb{N}$ and $g : \mathbb{N}^{k+2} \to \mathbb{N}$ are computable, then so is the function $h : \mathbb{N}^{k+1} \to \mathbb{N}$ defined by*

$$h(n_1, \ldots, n_k, 0) = f(n_1, \ldots, n_k)$$
$$h(n_1, \ldots, n_k, sm) = g\,(n_1, \ldots, n_k, m, h(n_1, \ldots, n_k, m))$$

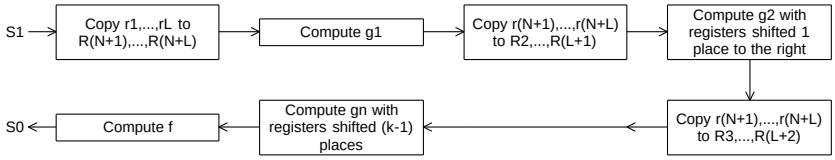
*(We say $h$ is defined by the* Primitive Recursion *from $f$ and $g$).*

*(5) If $f : \mathbb{N}^{k+1} \to \mathbb{N}$ is computable, then so is the function $g : \mathbb{N}^k \to \mathbb{N}$ defined by*
$g(n_1, \ldots, n_k) = m$ *if* $f(n_1, \ldots, n_k, m) = 0$ *but* $f(n_1, \ldots, n_k, j) > 0 \forall j < m$, *undefined otherwise*

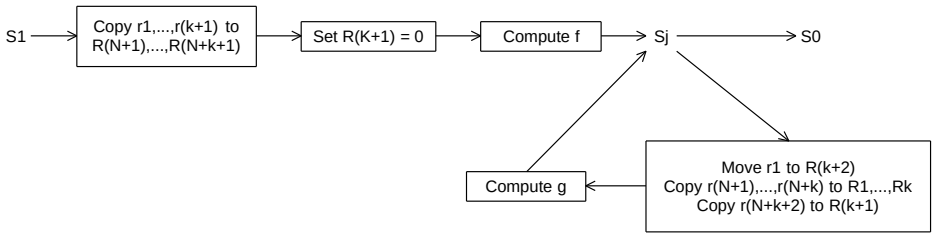
*(We say $g$ is defined by* Minimalization *from $f$).*

*Proof:* (1) and (2) are obvious. For (3) – (5) we draw schematic flow diagrams.

(3)




(4)




(5)

**Definition 4.3** (Due to A. Church) The set of Recursive Functions is the smallest set of partial functions $\mathbb{N}^k \rightharpoonup \mathbb{N}$ (for any $k > 0$) with the above closure properties (1) – (5) of Theorem 4.2. The set of Primitive Recursive Functions is the smallest set with the closure properties (1) – (4) of 4.2. Note that primitive recursive functions are total functions (however, not all total functions are primitive recursive). Theorem 4.2 implies that all recursive functions are computable.

**Example 4.4** (Some primitive recursive functions)

$$\begin{array}{lll} \mathrm{add}(n, 0) = n & \mathrm{times}(n, 0) = 0 & \exp(n, 0) = s0 \\ \mathrm{add}(n, sm) = s(\mathrm{add}(n, m)) & \mathrm{times}(n, sm) = \mathrm{add}(\mathrm{times}(n, m), n) & \exp(s, sm) = \mathrm{times}(\exp(n, m), n) \end{array}$$

**Theorem 4.5** *All computable functions are recursive.*

*Proof:* Suppose we have a program P for computing a function $f : \mathbb{N}^k \rightharpoonup \mathbb{N}$. Consider the function $g : \mathbb{N}^{k+1} \rightharpoonup \mathbb{N}$ defined by

$$g(n_1, \ldots, n_k, t) = 2^q.3^{r_1}.5^{r_2} \ldots {p_l}^{r_l}$$

where $q$ is the state reached by P after $t$ steps with inputs $(n_1, \ldots, n_k)$, $r_i$ is the number in the $i$th register, $p_j$ is the $j$th odd prime and $l$ is the number of the remotest register used by P. If P terminates in $< t$ steps then $g$ remains constant thereafter.

We claim that $g$ is primitive recursive. Clearly $g(n_1, \ldots, n_k, 0) = 2.3^{n_1} \ldots {p_k}^{n_k}$, which is a primitive recursive function of $n_1, \ldots, n_k$. We need to show that $g(n_1, \ldots, n_k, st)$ is a primitive recursive function of $g(n_1, \ldots, n_k, t)$. For this we need to show that the function

$$n \mapsto (n)_p = \text{ largest power of } p \text{ dividing } n$$

is primitive recursive for any prime $p$. Once we know that $(\cdot)_p$ is primitive recursive, it's clear that $g(n_1, \ldots, n_k, t) \mapsto g(n_1, \ldots, n_k, sn)$ is primitive. Now consider the function

$$h(n_1, \ldots, n_k) = \text{ least } t \text{ such that } (g(n_1, \ldots, n_k, t))_2 = 0 \text{ if this exists}$$

Then $f(n_1, \ldots, n_k) = (g(n_1, \ldots, n_k, h(n_1, \ldots, n_k)))_3$ □

*Remark:* Theorem 4.5 shows that any recursive function can be obtained from a primitive recursive function with a single application of minimalization.

Note that since the function $(n_1, n_2, \ldots, n_k) \mapsto 2^{n_1}3^{n_2} \ldots {p_{k-1}}^{n_k}$ is recursive and has recursive inverse functions $m \mapsto (m)_{p_k}$, we can reduce a recursive function $f$ of $k$ variables to a recursive function $g$ of 1 variable such that $f(n_1, n_2, \ldots, n_k) = g\left(2_{n_1}3^{n_2} \ldots {p_{k-1}}^{n_k}\right)$. Similarly, we can code the instructions in a register machine program by natural numbers; e.g. code $(j, +, k)$ by $2^j3^k$ and $(j, -, k, l)$ by $2^j3^k5^l$. Lecture 14

If the instructions in a program P are coded by the natural numbers $i_1, i_2, \ldots i_m$ then we can code P itself by the number $2^{i_1}3^{i_2} \ldots {p_{m-1}}^{i_m}$. Note that the function $\mathbb{N} \xrightarrow{f} \mathbb{N}$ defined by

$$f(n) = \begin{cases} 1 & \text{if } n \text{ is a code for some program} \\ 0 & \text{if not} \end{cases}$$

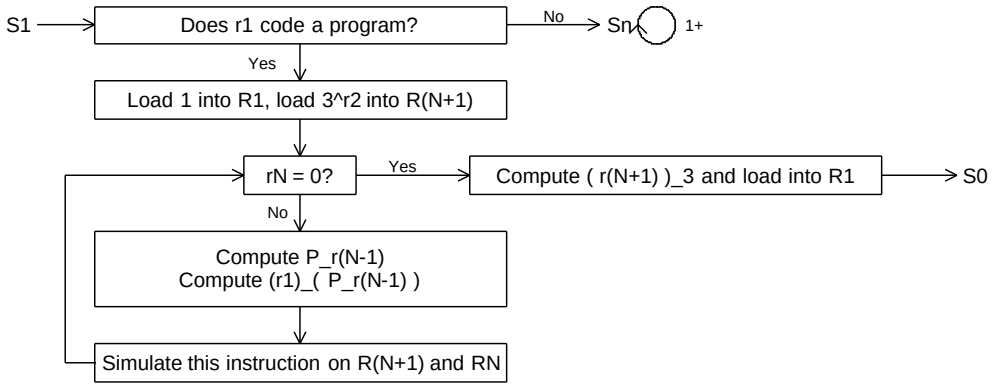
is itself recursive.

We write $P_n$ for the program coded by $n$, and $f_n : \mathbb{N} \mapsto \mathbb{N}$ for the unary (partial) function computed by $P_n$. (If $n$ isn't a program code, we interpret $f_n$ as the everywhere undefined function).

Note that, for any given recursive function $g$, the set $\{n \in \mathbb{N} | f_n = g\}$ is infinite.

**Theorem 4.6** *The function $u : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{N}$ defined by $u(n, m) = f_n(m)$ if $f_n(m)$ is defined, $u(n, m)$ undefined otherwise, is recursive.*

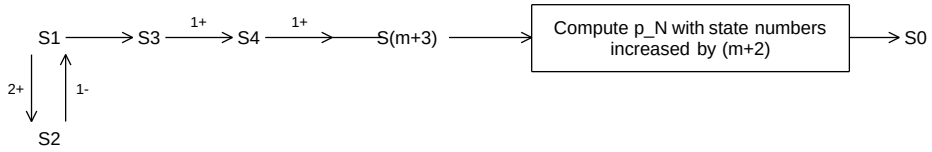
*Proof:* We describe a program for computing $u$:



A progam for computing $u$ is called a UNIVERSAL register machine program.

**Theorem 4.7** *(Parameterization Theorem, or s.m.n. theorem)*
*Suppose $g : \mathbb{N} \times \mathbb{N} \rightharpoonup \mathbb{N}$ is a recursive function of two variables. Then there exists a recursive function $h : \mathbb{N} \rightharpoonup \mathbb{N}$ such that $g(m,n) = f_{h(m)}(n)$ for all $m, n \in \mathbb{N}$.*

*Proof:* Let $P_N$ be a program for computing $g$. Then the following program computes the unary function $g(m,-)$:




We can compute the code number of this program recursively from $m$ (and $N$). This defines the function $h$.

**Definition 4.8** Let $S \subseteq \mathbb{N}^k$ for some $k$.

(1) We say $S$ is RECURSIVE (or decidable) if the function $\chi_S : \mathbb{N}^k \to \mathbb{N}$ defined by

$$\chi_S(n_1, \ldots, n_k) = \begin{cases} 1 & \text{if } (n_1, \ldots, n_k) \in S \\ 0 & \text{otherwise} \end{cases}$$

is recursive. (Eg the set $\{n | n \text{ codes a program}\}$ is recursive).

(2) We say $S$ is SEMI-RECURSIVE (or semi-decidable or recursively enumerable) if the partial function $\varphi_S$ defined by

$$\varphi_S(n_1, \ldots, n_k) = \begin{cases} 0 & \text{if } (n_1, \ldots, n_k) \in S \\ \text{undefined} & \text{otherwise} \end{cases}$$

is recursive.

**Example 4.9** The set $H = \{(m,n) | f_m(n) \text{ is defined}\}$ is semi-recursive, since we can compute $\varphi_H$ by first computing $u(m,n)$, then throwing away the answer and returning 0. $H$ is called the HALTING SET.

However, $H$ is not recursive. If it were, then we could define a recursive function $g$ of one variable as follows:
Compute $\chi_H(n,n)$, if this is 0 then output 0, otherwise compute $u(n,n) + 1$.
But, if this function is recursive, it must equal $f_n$ for some $n$ and $g(n) \neq f_n(n)$ for any $n$, which gives a contradiction.

We say that the 'halting problem' for recursive functions is undecidable. (Note that some people use $H$ for the set $\{n | f_n(n) \text{ is defined}\}$).

**Lemma 4.10**

*(1) For a subset $S$ of $\mathbb{N}$, the following are equivalent:*

*(a) $S$ is semi-recursive,*

*(b) There is a recursive function $f$ such that $f(n)$ is defined $\Leftrightarrow$ $n \in S$,*

*(c) There is a recursive function $f$ such that $n \in S \Leftrightarrow n$ is a value of $f$.*

*(2) A set $S \subseteq \mathbb{N}^k$ is recursive iff both $S$ and $\mathbb{N}^k \setminus S$ are semi-recursive.*

*Proof:*
(1) (a)⇒(b) is easy since $\varphi_S$ has domain $S$
(b)⇒(a) Given a program for computing $f$, we can modify it to compute $\varphi_S$.
(b)⇒(c) Given a program for computing $f$, we can modify it to compute the function $\psi_S$ defined by

$$\psi_S(n) = n \text{ if } n \in S, \text{ undefined otherwise.}$$

(c)⇒(b) Given a program for computing $f$, we can construct a program which, given input $n$, runs 'interlaced' computations of $f(0), f(1), f(2), \ldots$ in such a way that each terminating computation will eventually be completed. Each time a computation terminates, compare its output with $n$: if they are equal then stop, otherwise continue.

(2) ⇒ Given a program for computing $\chi_S$, we can modify it to compute $\varphi_S$ or $\varphi_{\mathbb{N}\setminus S}$. Lecture 15
⇐ Given programs for computing $\varphi_S$ and $\varphi_{\mathbb{N}^k\setminus S}$, we run them in parallel and output $\{\begin{smallmatrix}1\\0\end{smallmatrix}\}$ if $\{\begin{smallmatrix}\varphi_S\\ \varphi_{\mathbb{N}^k\setminus S}\end{smallmatrix}\}$ terminates.

□

Note that $\mathbb{N}^2 \setminus H = \{(m,n) | f_m(n) \text{ is undefined}\}$ is not semi-recursive since $H$ is semi-recursive but not recursive. Any finite subset of $\mathbb{N}$ is recursive, as are 'most' sets that one encounters in everyday mathematics.

In general, to prove that a subset $E$ is not recursive, we show that 'the halting problem can be reduced to $E$', i.e. that a program for computing $\chi_E$ could be modified to compute $\chi_H$.

**Theorem 4.11** *(Rice's Theorem)*
*Let $S$ be any set of (unary) recursive functions, which is neither the set of all recursive functions nor the empty set. Then $E = \{n | f_n \in S\}$ is not recursive.*

*Proof:* Let $g$ be a recursive function belonging to whichever of $S$ and its complement does not contain the empty function. Now consider the function $h(n,m) = g(m)$ if $u(n,n)$ and $g(m)$ are defined, undefined otherwise. $h$ is recursive since we can compute it by first computing $u(n,n)$ and then computing $g(m)$ if the first computation terminates. So by Theorem 4.7 we can write $h(n,m) = f_{k(n)}(m)$ for some recursive function $k : \mathbb{N} \to \mathbb{N}$. If $(n,n) \in H$, then $f_{k(n)} = g$. If $(n,n) \notin H$, then $f_{k(n)}$ is everywhere undefined.

So the composite $\mathbb{N} \xrightarrow{k} \mathbb{N} \xrightarrow{\chi_E} \mathbb{N}$ is the characteristic function of either $H_1 = \{n | (n,n) \in H\}$ or its complement. But $H_1$ is not recursive, so $E$ isn't recursive □

*Remark:* The proof of 4.11 shows that whichever of $E$ and $\mathbb{N}\setminus E$ contains the codes for the empty function is not semi-recursive.

**Theorem 4.12** *Let $h$ be a total recursive function $\mathbb{N} \to \mathbb{N}$. Then there exists $n \in \mathbb{N}$ such that $f_n = f_{h(n)}$.*

*Proof:* Consider the function $g(n,m) = u\,(h(u(n,n)), m)$. This is clearly recursive since $u$ and $h$ are recursive. So by 4.7 we can write $g(n,m) = f_{f_r(n)}(m)$ for some $r$. Now let $n = f_r(r)$. Then for any $m$ we have

$$\begin{aligned} f_n(m) &= g(r,m) \\ &= u(h(u(r,r)), m) \\ &= u(h(n), m) \\ &= f_{h(n)}(m) \end{aligned}$$

Observe that by the proof of 4.7 we know that $f_r$ is a total function, so in particular $f_r(r)$ is defined. □

Suppose $\mathcal{L} = \mathcal{L}(\Omega, \Pi)$ is a countable first-order language, i.e. the sets $\Omega$ and $\Pi$ are countable, hence so is the set $\Sigma = \Omega \cup \Pi \cup \{v, ', =, \bot, \Rightarrow, \forall, (, ), , \}$. Assume that we can choose an injection $\Sigma \xrightarrow{i} \mathbb{N}$ so that the image of $i$ is a recursive set and the arity functions $\Sigma \to \mathbb{N}$ and $\Pi \to \mathbb{N}$ become partial recursive functions when $\Omega$ and $\Pi$ are identified with subsets of $\mathbb{N}$.

Then we can code members of $\Sigma^*$ by natural numbers: the string $\sigma_1\sigma_2\ldots\sigma_k$ is coded by $2^{i(\sigma_1)}3^{i(\sigma_2)}\ldots {p_{k-1}}^{i(\sigma_k)}$. By the recursive nature of the rules for forming terms and formulae of $\mathcal{L}$, we see that the sets of codes for terms and for formulae are recursive.

Similarly, we can code contexts (finite strings of variables) by natural numbers, and the function which sends a formula to its canonical context (i.e. the list of its free variables in order of their first occurences) is coded by a recursive function $\mathbb{N} \to \mathbb{N}$.

**Definition 4.13** We say a first order theory $T$ is RECURSIVELY PRESENTED is its language has an enumeration as above, and the set of codes for the axioms of $T$ is recursive (If $T$ has finitely many axioms then this is certainly true).

Note that, for example, the first-order Peano arithmetic is recursively presented: the code for the formula

$$(\forall y_1, \ldots, y_n)\,(p[0/x] \Rightarrow ((\forall x)\,(p \Rightarrow p[sx/x]) \Rightarrow (\forall x)\,p))$$

can be calculated recursively from those for $p$ and the context $(x, y_1, \ldots, y_n)$.

The axioms of first-order logic are also coded by a recursive set, and the rules of inference are also 'recursively-encoded'. So, if $T$ is a recursively-presented theory, the set of deductions from $T$ is also coded by a recursive set, so the set $\{p | \top \vdash p\}$ is at least semi-recursive.

Construct a recursively presented theory $T$ such that $\{\lceil p \rceil \,|\top \vdash p\}$ is non-recursive (where $\lceil p \rceil$ denotes the number coding a formula $p$). We take $\Omega$ to consist of one binary operation (denoted by $(x, y) \mapsto xy$) and three constants $r, s, t$ (and $\Pi = \emptyset$). The first axiom says that the binary operation is associative: Lecture 16

$$(\forall x)(\forall y)(\forall z)\,((xy)z = x(yz))$$

All other axioms are equations between closed terms.

Suppose we are given a program P for computing the recursive function $u$ defined by $u(n, m) = f_n(m)$ and suppose P uses the registers $R_1, R_2, \ldots, R_k$. By a CONFIGURATION TERM we mean a closed term of the form

$$ts^{q+1}tr^{n_1}tr^{n_2}t \ldots r_{n_k}t$$

We think of this as corresponding to the configuration of the register machine when the program is in state $q$ and $R_1, R_2, \ldots, R_k$ hold the numbers $n_1, \ldots, n_k$. If the $q$th instruction in P is $(j, +, q')$, we write down all equations of the form

$$ts^{q+1}tr^{n_1}t \ldots r^{n_{j-1}}t = ts^{q'+1}tr^{n_1}t \ldots r^{n_{j-1}}tr$$

If the $q$th instruction is $(j, -, q', q'')$ then we write down the equations

$$ts^{q+1}tr^{n_1}t \ldots r^{n_{j-1}}tr = ts^{q'+1}tr^{n_1}t \ldots r^{n_{j-1}}t$$
$$ts^{q+1}tr^{n_1}t \ldots r^{n_{j-1}}tt = ts^{q''+1}tr^{n_1}t \ldots r^{n_{j-1}}tt$$

In the theory we've defined so far, two configuration terms $c_1, c_2$ are provably equal if the program can go from configuration $c_1$ to configuration $c_2$ in some number of steps. Since the 'forwards' behaviour of the register machine is deterministic, we see that $c_1$ and $c_2$ are provably equal iff there exists a configuration $c_3$ reachable from both $c_1$ and $c_2$.

Now we add the equations $tstr = tst$ and $tstt = tst$ which make all eventually terminating configurations provably equal in $T$ but have no effect on provable equality between non-terminating configurations.

Hence $u(n, m)$ is defined iff the equation $ts^2tr^ntr^mt^{k-1} = tst$ is provable in $T$. So if the set of codes deducible from $T$ were recursive, we could recursively determine whether or not $(n, m) \in H$.

**This is the end of the material which is examinable in IIA. The remainder of the course is examinable in IIB only.**

# Chapter 5

# Zermelo-Fraenkel Set Theory

Lecture 17

We aim to study the 'universe of sets' as a model of a first-order thoery in a language $\mathcal{L}$ with one primitive predicate $\in$ (apart from equality). We write $x \in y$ rather than $\in xy$. Later we'll add further operation and predicate symbols to $\mathcal{L}$, having proved that they're definable.

What should the axioms say? A first attempt (due to G. Frege, 1890):

We want sets to satifsy the AXIOM OF EXTENSIONALITY (1)

$$(\forall x)(\forall y)\left((\forall z)\left((z \in x) \Leftrightarrow (z \in y)\right) \Rightarrow (x = y)\right)$$

and the axiom-scheme of COMPREHENSION (2)

$$(\forall z_1, \ldots, z_n)(\exists y)(\forall x)\left((x \in y) \Leftrightarrow p\right)$$

where $p$ is a formula having $x, z_1, \ldots, z_n$ as free variables.

B.Russell observed that this is inconsistent: take $p$ to be the formula $\neg(x \in x)$ and we get a $y$ such that $(\forall x)\left((x \in y) \Leftrightarrow \neg(x \in x)\right)$. [1]Then we have $(y \in y) \Leftrightarrow \neg(y \in y)$ by substitution which yields $\perp$.

Possible responses:

(1) *Type Theory* (Russel): work with a many-sorted theory in which there are entities of different types, and the membership relation can only hold between entities of different types.

(2) *'New Foundations'* (W. Quine): We restrict comprehension to formulae $p$ which are *stratified*, ie we can assign 'levels' to the variables in $p$ so that $(x = y)$ appears only if $x$ and $y$ have the same level and $(x \in y)$ appears only if $y$ has level one higher than $x$

(3) *Zermelo(-Fraenkel) set theory* (E.Zermelo, 1904): replace comprehension by the axiom-scheme of SEPARATION

$$(\forall w_1, \ldots, w_n)(\forall x)(\exists y)(\forall z)\left((z \in y) \Leftrightarrow ((z \in n) \wedge p)\right)$$

where $p$ is any formula with free variables $z, w_1, \ldots, w_n$.

We adopt extensionality and separation as fundamental. Note that given extensionality, the $y$ whose existence is asserted by separation is uniquely determined by $w_1, \ldots, w_n, x$ and the formula $p$. We may, if we wish, introduce an $(n+1)$-ary function symbol for this $y$: we write it as $\{z \in x | p\}$ (the variable $z$ has the status of a bound variable here).

We need more axioms, e.g. to ensure that the universe (i.e. a model of our thoery) is non-empty. The first axiom in this group is the EMPTY-SET axiom: (3)

$$(\exists y)(\forall x)\neg(x \in y)$$

(we add the constant $\emptyset$ to $\mathcal{L}$ to denote this set).

The next is the PAIR-SET axiom (4)

[1]This is the set of all sets which do not have themselves as an element

$$(\forall x)(\forall y)(\exists z)(\forall w)\,((w \in z) \Leftrightarrow ((w = x) \vee (w = y)))$$

(we denote this $z$ by $\{x, y\}$ and we denote $\{x, x\}$ by $\{x\}$).

Then we add the UNION axiom (5)

$$(\forall x)(\exists y)(\forall z)\,((z \in y) \Leftrightarrow (\exists w)\,((z \in w) \wedge (w \in x)))$$

(we denote this $y$ by $\bigcup x$, and we denote $\bigcup\{x, y\}$ by $(x \cup y)$).

Note that if $x$ has a member (say $y$) then we can construct $\bigcap y$ as

$$\{z \in y \mid (\forall w)\,((w \in x) \Rightarrow (z \in w))\}$$

and $\bigcap \emptyset$ doesn't exist as a set. Similarly, we can construct $(x \cap y)$ and $(x \setminus y)$.

Next, the POWER-SET axiom [2] (6)

$$(\forall x)(\exists y)(\forall z)\,((z \in y) \Leftrightarrow (\forall w)\,((w \in z) \Rightarrow (w \in x)))$$

(we denote this $y$ by $\mathcal{P}x$).

Note that if we define $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$ and

$$\text{first}(t) = \begin{cases} \bigcup\bigcap t & \text{if } t \neq \emptyset \\ \emptyset & \text{if } t = \emptyset \end{cases}$$

$$\text{second}(t) = \begin{cases} \bigcup(\bigcup t \setminus \bigcap t) & \text{if } \bigcup t \setminus \bigcap t \neq \emptyset \\ \text{first}(t) & \text{otherwise} \end{cases}$$

Then $\text{first}(\langle x, y\rangle) = x$ and $\text{second}(\langle x, y\rangle) = y$ so $\langle x, y\rangle = \langle z, w\rangle \Leftrightarrow (x = z) \wedge (y = w)$ so we call $\langle x, y\rangle$ the (Kuratowskian) ORDERED PAIR of $x$ and $y$.

We write '$t$ is an ordered pair' for $t = \langle \text{first}(t), \text{second}(t)\rangle$ and we define

$$x \times y = \left\{ t \in \mathcal{P}\mathcal{P}\bigcup\{x, y\} \mid (t \text{ is an ordered pair}) \wedge (\text{first}(t) = x) \wedge (\text{second}(t) = y) \right\}$$

We define '$x$ is a FUNCTION' to mean Lecture 18

$$(\forall t)\,(t \in x) \Rightarrow ((t \text{ is an ordered pair}) \wedge (\forall y, z, w)\,((\langle y, w\rangle \in x \Rightarrow (\langle y, w\rangle \in x) \Rightarrow (z = w))))$$

and we write '$x : y \to z$' for

$$((x \text{ is a function}) \wedge (\forall w)\,((w \in y) \Leftrightarrow (\exists v)\,(\langle w, v\rangle \in x)) \wedge (\forall v)\,((\exists w)\,(\langle w, v\rangle \in x) \Rightarrow v \in z))$$

So we may write $z^y$ for the set $\{x \in \mathcal{P}(y \times z) \mid x : y \to z\}$.

Note that the axioms (1) – (6) also imply that any model $V$ of set theory must be infinite. e.g we can define the sequence of sets $V_n$ recursively by $V_0 = \emptyset$, $V_{n+1} = \mathcal{P}V_n$ and these are all distinct. But it's possible that $V$ is simply the 'union' of the $V_n$'s and hence that every set is finite.

We can also define (VON NEUMANN) NATURAL NUMBERS by

$$0 = \emptyset\ ,\ n + 1 = n \cup \{n\}$$

We define a set $x$ to be a SUCCESSOR SET if

$$(\emptyset \in x) \wedge (\forall y)\,((y \in x) \Rightarrow ((y \cup \{y\}) \in x))$$

And we take the axiom of INFINITY to be the assertion (7)

[2] Note that $(\forall w)\,((w \in z) \Rightarrow (w \in x))$ is equivalent to $(z \subseteq x)$

$$(\exists x)\,(x \text{ is a successor set})$$

If this holds then there's a unique smallest successor set, namely the intersection of all successor sets, which we denote by $\omega$.

$$\omega = \{y \in x \mid (\forall z)\,((z \text{ is a successor set}) \Rightarrow (y \in z))\}$$

Note that the axiom makes the empty-set axiom redundant.

To state the next axiom (Replacement) we need to introduce the notion of a CLASS. Informally a class is the collection of sets specified by some first-order formula $p$ having one free variable $x$. We denote it by $\{x \mid p\}$. Formally a class is an equivalence of first-order formulae $p$ with one free variable $x$ where $p$ is equivalent to $q$ if $(\forall x)\,(p \Leftrightarrow q)$ is provable.

Similarly we can define classes of pairs, classes of triples and so on using formulae with $2, 3, \ldots$ free variables.

We say a class of pairs $F$ is a FUNCTION CLASS if its representing formula $f$ has the property

$$(\forall x, y, z)\,(f \Rightarrow (f[z/y] \Rightarrow (y = z)))$$
$$\text{ie. } (\forall x, y, z)\,((\langle x, y\rangle \in F) \Rightarrow ((\langle x, z\rangle \in F) \Rightarrow (y = z)))$$

The AXIOM-SCHEME OF REPLACEMENT says that if we have a function-class $F$ and a set $x$ then the class $\{F(u) \mid u \in x\}$ is a set. This translates to the axiom-scheme (8)

$$(\forall z_1, \ldots, z_n)\,(\forall u, v, w)\,((f \Rightarrow (f[w/v] \Rightarrow (w = v))) \Rightarrow (\forall x)\,(\exists y)\,(\forall v)\,((v \in y) \Leftrightarrow (\exists u)\,((u \in x) \wedge f)))$$

where $F$ is any formula with free variables $u, v, z_1, \ldots, z_n$. This implies that the class $\{V_n \mid n \in \omega\}$ is a set and hence that $V_\omega = \bigcup\{V_n \mid n \in \omega\}$ is a set.

We can now continue by defining
$V_{\omega+1} = \mathcal{P}V_\omega$, $V_{\omega+2} = \mathcal{P}V_{\omega+1}, \ldots$, $V_{\omega+\omega} = \bigcup\{V_{\omega+n} \mid n \in \omega\}$, $V_{\omega+\omega+1} = \mathcal{P}V_{\omega+\omega}, \ldots$

It would be useful to know that every set belongs to some member of this sequence. To ensure this, we adopt the axiom of FOUNDATION (or Regularity), due to von Neumann, which says (9)

$$(\forall x)\,(\neg\,(x = \emptyset) \Rightarrow (\exists y)\,((y \in x) \wedge (x \cap y = \emptyset)))$$

**Theorem 5.1** *In the presence of the other axioms, ZF (Foundation) is equivalent to the scheme of $\in$-induction:*

$$(\forall z_1, \ldots z_n)\,((\forall x)\,((\forall y)\,((y \in x) \Rightarrow p[y/x]) \Rightarrow p) \Rightarrow (\forall x)\,p)$$

*where $p$ is a formula with free variables $x, z_1, \ldots, z_n$.*

*Proof:* $\Leftarrow$: Apply $\in$-induction to the formula ($x$ is a regular set) where this means

$$(\forall y)\,((x \in y) \Rightarrow (\exists z)\,((z \in y) \wedge (y \cap z = \emptyset)))$$

Clearly the assertion $(\forall x)\,(x \text{ is regular})$ is equivalent to Foundation. Suppose

$$(\forall w)\,((w \in x) \Rightarrow (w \text{ is regular})) \quad \text{and} \quad (x \in y)$$

Then either $x \cap y = \emptyset$ in which case we take $z = x$ or $(\exists w)\,(u \in x \cap y)$ in which case $w$ is regular since $w \in x$ and hence $(\exists z)\,((z \in y) \wedge (y \cap z = \emptyset))$ since $w \in y$. Hence $x$ is regular.

Before proving the second half of Theorem 5.1 we need: Lecture 19

**Lemma 5.2** $(\forall x)\,(\exists y)\,((x \subseteq y) \wedge (\forall z, w)\,(((z \in w) \wedge (w \in y)) \Rightarrow (z \in y)))$ *is provable in ZF.*
*A set $y$ with the property $((z \in w \in y) \Rightarrow (z \in y))$ is called* TRANSITIVE. *Any intersection of transitive sets is transitive, so there's a smallest one containing a given $x$ which we denote $TC(x)$ and call the* TRANSITIVE CLOSURE *of $x$.*

*Proof:* Informally, we want $\mathrm{TC}(x)$ to be $\bigcup \{x, \cup x, \cup \cup x, \dots\}$ so we have to ensure that the set above exists. To show this, we construct a function class with domain $w$ and range equal to the class above, namely

$$\{\langle y, z\rangle \mid (\exists f, n)\,(f \text{ is a function}) \wedge (n \in w) \wedge (\langle y, z\rangle \in f) \wedge (\forall u)\,((\exists v)\,(\langle u, v\rangle \in f) \Leftrightarrow (u \in n)) \wedge (\langle 0, x\rangle \in f) \wedge (\forall w, v)\,\big(((\langle u, v\rangle \in f) \wedge (u^+ \in n)) \Rightarrow (\langle u^+, \cup v\rangle \in f)\big)\}$$

where $w^+$ is shorthand for $w \cup \{w\}$. We can check this is a function class and its domain is the whole of $w$. Then it's easy to check that $x \subseteq \mathrm{TC}(x)$ and $\mathrm{TC}(x)$ is transitive. □

*Proof of 5.1 continued:* Let $p$ be a formula (with one free variable $x$, for simplicity) satisfying

$$(\forall x)\,((\forall y)\,((y \in x) \Rightarrow p[y/x]) \Rightarrow p)$$

Suppose for a contradiction we have $\neg\,(\forall x)\,p$, or equivalently $(\exists x)\,\neg p$. Let $x$ be a set for which $p$ fails. Consider the set $u = \{y \in \mathrm{TC}(x) \mid p[y/x]\}$. Then $x \in u$ so $u \neq \emptyset$ and hence there is a $y$ such that $(y \in w)$ and $(y \cap u = \emptyset)$. Now any $z$ with $z \in y$ belongs to $\mathrm{TC}(\{x\})$ but not to $u$, so $p[z/x]$ holds. Hence by induction hypothesis, $p[y/x]$ holds, which contradicts $y \in u$. So $(\forall x)p$ holds. □

**Definition 5.3** Let $R$ be a RELATION-CLASS (ie a class defined by a formula $r$ with two free variables). We say $R$ is WELL-FOUNDED if every non-empty set has an $R$-minimal member, ie

$$(\forall u)\,(\neg\,(u = \emptyset) \Rightarrow (\exists x)\,((\langle y, x\rangle \in R) \Rightarrow \neg\,(y \in u)))$$

We say $R$ is LOCAL if the $R$-predecessors of any set form a set, ie

$$(\forall x)\,(\exists u)\,(\forall y)\,((y \in u) \Leftrightarrow (\langle y, x\rangle \in R))$$

If $R$ is local, then for any set $x$ we can form a set $\mathrm{RC}(x)$ which contains $x$ and is '$R$-closed', ie

$$(\forall y, z)\,((((\langle y, z\rangle \in R) \wedge (z \in \mathrm{RC}(x))) \Rightarrow (y \in \mathrm{RC}(x)))$$

If $R$ is both local and well-founded, we can prove the scheme of $R$-INDUCTION:

$$(\forall x)\,((\forall y)\,((\langle y, x\rangle \in R) \Rightarrow p[y/x]) \Rightarrow p) \Rightarrow (\forall x)\,p$$

**Lemma 5.4** *If $R$ is a well-founded local relation-class, then there exists a well-founded local relation-class $\bar{R}$ such that $R \subseteq \bar{R}$ and $\bar{R}$ is transitive, ie*

$$((((\langle x, y\rangle \in R) \wedge (\langle y, z\rangle \in R)) \Rightarrow (\langle x, z\rangle \in R))$$

*Proof:* We define $\bar{R}$ by

$$\langle x, y\rangle \in \bar{R} \Longleftrightarrow (x \in \mathrm{RC}\,(\{y\}) \setminus \{y\})$$

It's immediate that $\bar{R}$ is local and $R \subseteq \bar{R}$. Also if $x \in \mathrm{RC}\,(\{x\})$ then $\mathrm{RC}\,(\{x\}) \subseteq \mathrm{RC}\,(\{y\})$, from which it follows that $\bar{R}$ is transitive.

To show $\bar{R}$ is well-founded suppose $u$ is a non-empty set and consider

$$\{x \in \mathrm{RC}\,(u) \mid (\exists y)\,((y \in u) \wedge (y \in \mathrm{RC}\,(\{x\})))\}$$

$\bar{u}$ is non-empty, since $u \subseteq \bar{u}$, so let $x$ be an $R$-minimal member of $\bar{u}$. If there exists $y \in \mathrm{RC}\,(\{x\}) \setminus \{x\}$ such that $y \in u$ then some $R$-predecessor of $x$ is also in $\bar{u}$, giving a contradiction.

So $\mathrm{RC}\,(\{x\}) \cap u = \{x\}$, hence $x \in u$ and $x$ is $R$-minimal in $u$.

□

**Theorem 5.5** *(R-recursion Theorem)*
*Let $R$ be a well-founded local relation-class, and $G$ a function-class of two variables (i.e. $G$ is a class of triples $(x,y,z)$ such that if $(x,y,z) \in G$ and $(x,y,z') \in G$ then $z = z'$) which is defined on the whole of $V \times V$. Then there is a unique function-class $F$, defined on $V$, such that* $(\forall x)\,(F(x) = G(x, \{F(y) \mid \langle y,x\rangle \in R\})$

*Proof: Uniqueness:* If $F$ and $F'$ are both solutions to the recursion equation, then we can prove $(\forall x)\,(F(x) = F'(x))$ by $R$-induction.

*Existence:* Define '$F$ is an attempt' to mean

$$(F \text{ is a function}) \wedge ((\exists x)\,(x \text{ is } R\text{-closed})) \wedge (\forall y)\,((\exists z)\,(\langle y,z\rangle \in F) \Leftrightarrow (y \in x)) \wedge$$
$$(\forall y)\,((y \in x) \Rightarrow F(y) = G\,(y, \{F(z) \mid \langle z,y\rangle \in R\}))$$

By $R$-induction, we can show that any two attempts agree on the intersection of their domains. So

$$\bigcup \{F \mid F \text{ is an attempt}\} = \{\langle x,y\rangle \mid (\exists F)\,((F \text{ is an attempt}) \wedge (\langle x,y\rangle \in F))\}$$

is a function-class, and it satisfies the recursion equation for all $x$ in its domain.

Suppose its domain is not the whole of $V$, i.e. we have an $x$ not in the domain of any attempt. Consider

$$\{y \in RC(\{x\}) \mid y \text{ is not in the domain of any attempt}\}$$

By 5.4, this has an $\bar{R}$-minimal member $y$, say. Now every member of $RC(\{y\}) \setminus \{y\}$ is the domain of some attempt, so by Replacement there is a set consisting of all pairs $\langle z,w\rangle$ where $z \in RC(\{y\}) \setminus \{y\}$ and $\langle z,w\rangle$ belongs to some attempt, and this is itself an attempt. Now define

$$F' = F \cup \{\langle y, G\,(y, \{F(z) \mid \langle z,y\rangle \in R\})\rangle\}$$

then $F'$ is an attempt and $y \in \mathrm{dom}F'$, giving a contradiction. □

Lecture 20

We can modify the Recursion Theorem to construct functions by recursion over some class $A$, given a relation-class $R$ which is 'well-founded relative to $A$', i.e.

$$(\forall x)\,(((x \subseteq A) \wedge \neg\,(x = \emptyset)) \Rightarrow (\exists y)\,(y \in x) \wedge (\forall z) \neg\,((z \in x) \wedge (\langle z,y\rangle \in R)))$$

and 'local relative to $A$' in a similar sense. In particular we can relativize the Recursion Theorem to a set $a$. In this case the proof can be simplified since the union of all attempts to construct a function satisfying the recursion equation is itself an attempt.

We say a relation-class $R$ is EXTENSIONAL (relative to a class $A$) if it satisfies

$$(\forall x, y \in A)\,((\forall z)\,((z \in A) \Rightarrow ((\langle z,x\rangle \in R) \Leftrightarrow (\langle z,y\rangle \in R)))) \Rightarrow (x = y))$$

**Corollary 5.6** *(Mostowski's Collapsing Theorem) Let $a$ be a set and $r \subseteq a \times a$ a binary relation which is well-founded and extensional. Then there is a unique pair $(b,f)$ such that $b$ is a transitive set and $f : (a,r) \to (b, \in \cap\, b \times b)$ is an isomorphism of sets-with-a-binary-relation.*

*Proof:* We define $f$ by $r$-recursion over $a$:

$$f(x) = \{f(y) \mid (y \in a) \wedge (\langle y,x\rangle \in r)\}$$

and we define $b = \{f(x) \mid x \in a\}$ (which we know is a set, by Replacement). $f$ is surjective by definition, and $\langle y,x\rangle \in r$ implies $f(y) \in f(x)$. To show the converse of this implication it's enough to show that $f$ is injective: we prove this by $r$-recursion over $a$.

We consider

$$t = \{x \in a \mid (\forall y)\,(((y \in a) \wedge (f(x) = f(y))) \Rightarrow (y = x))\}$$

We need to show that if every $r$-predecessor of $x$ belongs to $t$, then $x \in t$. Suppose $f(x) = f(y)$ for some $y$, then for $z$ such that $\langle z, y\rangle \in r$, we have $f(z) = f(u)$ for some $u$ with $\langle u, x\rangle \in r$, and hence $z = u$ since $u \in t$. So $\langle z, x\rangle \in r$. Similarly $\langle z, x\rangle \in r$ implies $\langle z, y\rangle \in r$. So by extensionality of $r$, $x = y$. Hence $x \in t$. Hence $a \setminus t = \emptyset$, so $f$ is injective.

*Uniqueness:* Suppose $f' : a \to b'$ is another solution. Then the composite $g = f' \circ f^{-1} : b \to b'$ is an isomorphism of transitive sets. Now use $\in$-induction on $b$ to prove $(\forall x)\,((x \in b) \Rightarrow (g(x) = x))$. So $b = b'$ and $f = f'$. $\square$

# Chapter 6

# Ordinals and Well-Orderings

A WELL-ORDERING is a well-founded relation which is a total ordering. Note that a well-founded relation must be irreflexive (since $\langle x, x\rangle \in r$ implies $\{x\}$ has no $r$-minimal member) and anti-symmetric (similar argument with pair-sets) and if $r$ satisfies the TRICHOTOMY LAW

$$(\forall x, y)\,((\langle x, y\rangle \in r) \vee (\langle y, x\rangle \in r) \vee (x = y))$$

then it's transitive, since if $\langle x, y\rangle \in r$ and $\langle y, z\rangle \in r$ but $\langle x, z\rangle \notin r$ then $\{x, y, z\}$ has no $r$-minimal member. Hence a well-founded relation is a well-ordering iff it's trichotomous.

A well-ordering is automatically existentional, so by Corollary 5.6 we deduce that for any well-ordered set $(a, <)$ there's a unique transitive set $\alpha$, trichotomously ordered by $\in$, such that $(a, <) \cong (\alpha\ \in)$.

**Definition 6.1** An ORDINAL is a transitive set $\alpha$ satisfying

$$(\forall \beta, \delta)\,(((\beta \in \alpha) \wedge (\gamma \in \alpha)) \Rightarrow ((\beta \in \gamma) \vee (\gamma \in \beta) \vee (\beta = \gamma)))$$

So Mostowski's Theorem says that every isomorphism class of well-ordered sets contains a unique ordinal. We write $O_n$ for the class of all ordinals.

**Lemma 6.2** *If $\alpha$ is an ordinal, then so is $\alpha^+ = \alpha \cup \{\alpha\}$.*

*Proof:* If $\gamma \in \beta \in \alpha^+$, then either $\beta \in \alpha$ or $\beta = \alpha$, so in either case $\gamma \in \alpha \subseteq \alpha^+$. So $\alpha^+$ is transitive.

| | | | |
|---|---|---|---|
| If $\gamma, \beta \in \alpha^+$, then either | $\gamma, \beta \in \alpha$ | | in which case $\gamma \in \beta$ or $\beta \in \gamma$ or $\beta = \gamma$ |
| or | $\beta \in \alpha$ and $\gamma = \alpha$ | | in which case $\beta \in \gamma$ |
| or | $\beta = \alpha$ and $\gamma \in \alpha$ | | in which case $\gamma \in \beta$ |
| or | $\beta = \alpha$ and $\gamma = \alpha$ | | in which case $\beta = \gamma$ |

So $\alpha^+$ is trichotomous. □

Since $\emptyset \in O_n$ we can conclude that $\omega \in O_n$.

**Lemma 6.3** *If $\alpha$ is an ordinal and $\beta \in \alpha$, then $\beta$ is an ordinal.*

*Proof:* First suppose that $\delta \in \gamma \in \beta$. Then $\delta, \gamma \in \alpha$ since $\alpha$ is transitive, and we must have one of $\delta \in \beta$, $\beta \in \delta$ or $\beta = \delta$ by trichotomy of $\alpha$. The second and third possibilities imply that that $\{\beta, \gamma, \delta\}$ has no $\in$-minimal member, so $\delta \in \beta$, i.e. $\beta$ is transitive. Similarly, if $\gamma, \delta$ are any 2 members of $\beta$, then $\gamma \in \alpha$ and $\delta \in \alpha$ so by trichotomy of $\alpha$ we have $(\gamma \in \delta) \vee (\delta \in \gamma) \vee (\gamma = \delta)$ □

**Lemma 6.4** *If $\alpha$ and $\beta$ are ordinals then either $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$*

*Proof:* Suppose $\alpha \not\subseteq \beta$. Then $\alpha \setminus \beta$ is non-empty, so it has a $\in$-least member $\gamma$, say. Now $\delta \in \gamma$ implies $\delta \in \alpha$ (by transitivity) and $\delta \in \beta$ by minimality of $\gamma$. And if $\delta \in \alpha \cap \beta$ then we have $\delta \in \gamma$, $\gamma \in \delta$ or $\gamma = \delta$ by trichotomy but $\gamma \in \delta$ and $\gamma = \delta$ both contradict $(\gamma \notin \beta) \wedge (\delta \in \beta)$, so $\delta \in \gamma$. So $\gamma = \alpha \cap \beta$, and in particular $\alpha \cap \beta \in \alpha$. Similarly, $\beta \not\subseteq \alpha$ would imply $\alpha \cap \beta \in \beta$. If both $\alpha \not\subseteq \beta$ and $\beta \not\subseteq \alpha$, then $\alpha \cap \beta \in \alpha \cap \beta$, contradicting foundation. □

Lecture 21

**Corollary 6.5** *For any two ordinals $\alpha, \beta$ we have one of $\alpha = \beta$, $\alpha \in \beta$ or $\beta \in \alpha$.*

*Proof:* If $\alpha \subseteq \beta$ but $\alpha \neq \beta$ then $\alpha = \alpha \cap \beta \subseteq \beta$ by the proof of 6.4. Similarly if $\beta \subseteq \alpha$ but $\beta \neq \alpha$, then $\beta \in \alpha$. □

Note that 6.3 and 6.5 together say that the class $O_n$ has all the properties of an ordinal, except that of being a set. Hence $O_n$ is not a set for if it were we'd have $O_n \in O_n$, contradicting foundation. (This is sometimes called the Burali-Forti paradox).

Note also that for ordinals $\alpha$ and $\beta$, $(\alpha \subseteq \beta) \Leftrightarrow ((\alpha = \beta) \text{ or } (\alpha \in \beta))$. ( $\Rightarrow$ from 6.5, $\Leftarrow$ by transitivity of $\beta$).

We sometimes write $\alpha < \beta$ for $\alpha \in \beta$ and $\alpha \leq \beta$ for $\alpha \subseteq \beta$. Note also that, is $a$ is any nonempty set of ordinals, then $\cap a$ is an ordinal and is the $\in$-least member of $a$.

**Lemma 6.6** *If $a$ is any set of ordinals, then $\cup a$ is an ordinal.*

*Proof:* $\cup a$ is transitive, since any union of transitive sets is transitive. By 6.3 the members of $\cup a$ are ordinals, so by 6.5 they satisfy the trichotomy law.

**Theorem 6.7** *Let $A$ be any class such that*

$$(\forall x)\,((x \in A) \Rightarrow (x^+ \in A))$$
$$\textit{and}\quad (\forall x)\,((\forall y)\,((y \in x) \Rightarrow (y \in A)) \Rightarrow (\cup x \in A))$$

*Then $O_n \subseteq A$.*

*Proof:* Suppose there exists an ordinal $\alpha$ not in $A$. Then there's a least such ordinal $\beta$, namely $\beta = \bigcap \{x \in \alpha^+ \mid x \notin A\}$. If $\beta$ has an $\in$-greatest member $\gamma$, then $\beta = \gamma^+$, since $(\forall \delta)\,((\delta < \beta) \Leftrightarrow (\delta \leq \alpha))$. But $\delta \in A$ by the construction of $\beta$, so $\beta \in A$, giving a contradiction.

If $\beta$ has no greatest member, then every $\gamma \in \beta$ is a member of some member of $\beta$, so $\beta \subseteq \cup\beta$. But $\cup\beta \subseteq \beta$ is equivalent to transitivity, so $\beta = \cup\beta$. By construction, every member of $\beta$ is in $A$, so this implies $\beta \in A$, giving a contradiction. □

We call $\beta$ a SUCCESSOR ORDINAL if $\beta = \gamma^+$ for some $\gamma$ and a LIMIT ORDINAL is $\beta = \cup\beta$.

6.7 implies that we can prove things about ordinals inductively, or define functions on $O_n$ recursively, by separately considering the cases of successor and limit ordinals. For example, the VON NEUMANN HIERARCHY of sets $V_\alpha$ is the function-class $O_n \to V$ defined by

$$V_{\alpha^+} = \mathcal{P}V_\alpha$$
$$V_\lambda = \bigcup\{V_{alpha} \mid \alpha < \lambda\} \text{ if } \lambda \text{ is a limit ordinal}$$

(Note that this includes $V_0 = \emptyset$). We could also equivalently say$V_\alpha = \{\mathcal{P}V_\beta \mid \beta \in \alpha\}$ for any $\alpha$.

Similarly we define the RANK of a set by $\in$-recursion over $V$:

$$\text{rank}(x) = \bigcup\{\text{rank}(y)^+ \mid y \in x\}$$

Note that $(\forall x)\,(\text{rank}(x) \in O_n)$ by $\in$-induction.

**Lemma 6.8** *For any set and ordinal $\alpha$, we have*

$$(x \in V_\alpha) \Leftrightarrow \textit{rank}(x) < \alpha$$
$$\textit{and}\quad (x \subseteq V_\alpha) \Leftrightarrow \textit{rank}(x) \leq \alpha$$

*Proof:* The second assertion is obtained from the first by substituting $\alpha^+$ for $\alpha$. We prove each direction of the first by induction:

$\Rightarrow$ : By induction over $\alpha$. Suppose $(\Rightarrow)$ holds for all ordinals $< \alpha$ and $x \in V_\alpha$. If $\alpha$ is a limit, then $x \in V_\beta$ for some $\beta < \alpha$, so $\text{rank}(x) < \beta < \alpha$. If $\alpha = \beta^+$ then $x \subseteq V_\beta$ so we have $\text{rank}(y) < \beta$ for all $y \in x$, and so $\text{rank}(y)^+ \leq \beta$ for all $y \in x$, and so $\text{rank}(x) \leq \beta \leq \alpha$.

$\Leftarrow$ : By induction over $x$. Suppose $(\Leftarrow)$ holds for all $y \in x$, and suppose $\text{rank}(x) < \alpha$. If $\alpha = \beta^+$, then we have $\text{rank}(x) \leq \beta$, so $\text{rank}(y) < \beta$ for all $y \in x$, so $x \in V_\beta$, so $x \in V_\alpha$. If $\alpha$ is a limit, then $\text{rank}(x) < \beta < \alpha$ for some $\beta$, so every $y \in x$ has $\text{rank}(y) < \beta$. So all $y \in x$ are members of $V_\beta$, so $x \in \mathcal{P}V_\beta = V_{\beta+} \subseteq V_\alpha$. $\square$

To define addition and multiplication of ordinals we may either think of them 'synthetically' as 'shadowing' operations on well-ordered sets, or define them recursively.

For addition, we may equip the disjoint union $\alpha \sqcup \beta = (\alpha \times \{0\}) \cup (\beta \times \{1\})$ with a well-ordering, namely

$$(\sigma, i) < (\delta, j) \;\Leftrightarrow\; (i < j) \text{ or } ((i = j) \text{ and } (\sigma < \delta))$$

(reverse lexicographic ordering) and we define $\alpha + \beta$ to be the order-type of this well-ordered set (i.e. the unique ordinal isomorphic to it).

Equivalently, we may define $\alpha + \beta$ by recursion on $\beta$:

$$\begin{aligned} \alpha + 0 &= \alpha \\ \alpha + (\gamma^+) &= (\alpha + \gamma)^+ \\ \alpha + \beta &= \bigcup\{\alpha + \gamma \mid \gamma < \beta\} \text{ if } \beta \text{ is a non-zero limit.} \end{aligned}$$

Similarly, we define $\alpha\beta$ synthetically as the order-type of $\alpha \times \beta$, with reverse lexicographical ordering, or equivalently

$$\begin{aligned} \alpha.0 &= 0 \\ \alpha\,(\gamma^+) &= \alpha\gamma + \alpha \\ \alpha\beta &= \bigcup\{\alpha\gamma \mid \gamma < \beta\} \text{ if } \beta \text{ is a limit.} \end{aligned}$$

Note that these operations are not commutative:

$$\omega + 1 = \omega^+ > \omega \qquad \text{but} \qquad 1 + \omega = \bigcup\{1 + n \mid n < \omega\} = \omega$$

$$\omega.2 = \omega + \omega > \omega \qquad \text{but} \qquad 2\omega = \bigcup\{2n \mid n < \omega\} = \omega$$

Lecture 22

**Lemma 6.9** *(i) If $\alpha \leq \beta$ then $\alpha^+\gamma \leq \beta^+\gamma$*

*(ii) If $\alpha \leq \beta$ then $\alpha\gamma \leq \beta\gamma$*

*(iii) If $\beta < \gamma$ then $\alpha + \beta < \alpha + \gamma$*

*(iv) If $\beta < \gamma$ and $\alpha \neq 0$, then $\alpha\beta < \alpha\gamma$*

*Proof:* For (i) and (ii) use induction on $\gamma$. Clearly true if $\gamma = 0$. If $\gamma = \delta^+$ then

$$\begin{aligned} (\alpha + \gamma) = (\alpha + \delta)^+ &\leq (\beta + \delta)^+ = \beta + \gamma \\ \alpha\gamma = \alpha\delta + \alpha &\leq \beta\delta + \beta = \beta\gamma \quad \text{(assuming (iii))} \end{aligned}$$

If $\gamma$ is a limit, then $\alpha + \delta = \bigcup\{\alpha + \delta \mid \delta < \gamma\} \leq \bigcup\{\beta + \delta \mid \delta < \gamma\} = \beta + \gamma$ and similarly for multiplication.

For (iii) and (iv) observe that if $\beta < \gamma$, then $\alpha \sqcup \beta$ is a proper initial segment of $\alpha \sqcup \gamma$ and similarly for $\alpha \times \beta$ and $\alpha \times \gamma$, provided $\alpha \neq 0$. $\square$

**Lemma 6.10** *(i)* $0 + \alpha = \alpha \ \forall \alpha$ *and* $0.\alpha = 0 \ \forall \alpha$

*(ii)* $1.\alpha = \alpha = \alpha.1 \ \forall \alpha$

*(iii)* $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma) \ \ \forall \alpha, \beta, \gamma$

*(iv)* $\alpha\,(\beta + \gamma) = \alpha\beta + \alpha\gamma \ \forall \alpha, \beta, \gamma$

*(v)* $\alpha\,(\beta.\gamma) = (\alpha\beta)\,\gamma \ \forall \alpha, \beta, \gamma$

*Proof:* (i) For $0 + \alpha = \alpha$ use induction. True for $\alpha = 0$. If true for $\beta$, then true for $\beta^+$. If true for all $\beta <$ some non-zero limit $\lambda$, then true for $\lambda$. For $0.\alpha = 0$ use synthetic argument: $\emptyset \times \alpha = \emptyset$.

(ii) $\alpha.1 = \alpha\,(0^+) = \alpha.0 + \alpha = 0 + \alpha = \alpha$. For $1.\alpha = \alpha$ use induction on $\alpha$.

(iii), (iv) and (v) use synthetic arguments: The obvious bijection

$$(\alpha \sqcup \beta) \sqcup \gamma \to \alpha \sqcup (\beta \sqcup \gamma)$$

is order-preserving, and similarly for

$$\alpha \times (\beta \sqcup \gamma) \to (\alpha \times \beta) \sqcup (\alpha \times \gamma)$$

and

$$(\alpha \times \beta) \times \gamma \to \alpha \times (\beta \times \gamma)$$

□

Can also define ordinal exponentation $\alpha^\beta$, and verify laws like $\alpha^{\beta+\gamma} = \alpha^\beta.\alpha^\gamma$, $\alpha^{\beta\gamma} = \left(\alpha^\beta\right)^\gamma$ etc.

*Problem:* Which sets are well-orderable? The first answer is Hartog's Lemma:

**Lemma 6.11** *For any set $a$, there is an ordinal $\gamma(a)$ for which there is no injection $\gamma(a) \to a$.*

*Proof:* There is a set, $g$, whose members are all well-orderings of subsets of $a$, since we can construct it from $\mathcal{P}(a \times a)$ by separation. There is a function-class which assigns to every well-ordered set its order-type. So we may define $\gamma(a)$ to be the set of ordinals $\alpha$ for which there exists $x \in g$ and an order-isomorphism from the subset well-ordered by $x$ to $\alpha$. $\gamma(a)$ is transitive, since any subset of a well-ordered set is well-ordered, so it's an ordinal, and it can't be injected into $a$. □

In 1908, E.Zermelo published a proof assuming the axiom of choice that every set can be ordered. In fact,

**Theorem 6.12** *The following assertions about a model of set theory are equivalent:*

*(i) The Axiom of Choice: if $(a_i \mid i \in I)$ is a family of non-empty sets, then there exists $f : I \to \bigcup\{a_i \mid i \in I\}$ such that $f(i) \in a_i \ \forall i$;*

*(ii) Zorn's Lemma;*

*(iii) Every set can be well-ordered;*

*(iv) Given any two sets $a$ and $b$, there exists either an injection $a \to b$ or an injection $b \to b$.*

*Proof:* (i) ⇒ (ii) was proved in Theorem 1.15.

(ii) ⇒ (iii) Given a set $a$, consider the set $g$ of all well-orderings of subsets of $a$, ordered by $(a', <) \leq (a'', <')$ iff $a'$ is an initial segment of $(a'', <')$ and $<= (<' \cap (a' \times a'))$. Then $(g, \leq)$ is chain-complete: if $\{(a_i, <_i) \mid i \in I\}$ is a chain of members of $g$, then $(\bigcup\{a_i \mid i \in I\}, \bigcup\{<_i \mid i \in I\})$ is well-ordered, and is an upper bound for the $(a_i, <_i)$ in $g$. Let $(a', <)$ be a maximal element of $g$. If $a' \neq a$, pick $x \in a \setminus a'$ and define $a'' = a \cup \{x\}$, $<' = < \cup \{\langle y, x\rangle \mid y \in a\}$, then $(a', <) \lneqq (a'', <')$ in $g$, giving a contradiction.

(iii) ⇒ (iv): If $a$ and $b$ are two sets, then by (iii) we have bijections $a \to \alpha$and $b \to \beta$ where $\alpha, \beta \in O_n$. By 6.4 we have either $\alpha \subseteq \beta$ or $\beta \subseteq \alpha$.

(iv) $\Rightarrow$ (iii): Follows from 6.11: if (iv) holds then there mst be an injection $a \to \gamma(a)$. But any subset of a well-ordered set is well-ordered, so we obtain a well-ordering of $a$.

(iii) $\Rightarrow$ (i): Choose some well-ordering $<$ of $\bigcup\{a_i \mid i \in I\}$. Then the function $f(i) =$ the $<$-least element of $a_i$ is a choice function. $\square$

*Remark:* There are direct proofs of (ii) $\Rightarrow$ (iv), (iii) $\Rightarrow$ (i) and (i) $\Rightarrow$ (iii). In fact there's a stronger result than (i) $\Leftrightarrow$ (iii): in any model of ZF set theory, a set $a$ has a well-ordering iff there exists a function $g : \mathcal{P}(a) \setminus \{\emptyset\} \to a$ such that $g(b) \in b$ for all non-empty $b \subseteq a$. (Such a $g$ is called a choice function for the set $a$).

Consider the following assertions: Lecture 23

(1) Every vector space has a basis;

(2) Every linearly independent set in a vector space can be extended to a basis;

(3) Every subspace of a vector space has a complement.

We proved Zorn's Lemma $\Rightarrow$ (2) in Chapter 1. (2) $\Rightarrow$ (1) and (2) $\Rightarrow$ (3) are easy. J.D.Halpern (1966) proved (3) $\Rightarrow$ Axiom of Choice. A.R.Blass (1983) proved (1) $\Rightarrow$ Axiom of Choice.

(4) Every non-trivial ring has a maximal ideal;

(5) Every proper ideal in a ring is contained in a maximal ideal.

Zorn's Lemma $\Rightarrow$ (5) $\Leftrightarrow$ (4) are easy. W.Hodges (1979) showed that ( (5) for unique factorization domains) $\Rightarrow$ Axiom of Choice.

(6) Every product of compact topological spaces is compact.

A.N.Tychonoff (1929) proved Axiom of Choice $\Rightarrow$ (6). J.L.Kelley (1950) proved (6) $\Rightarrow$ Axiom of Choice.

(4′) Every non-trivial commutative ring has a prime ideal;

(5′) Every proper ideal in a commutative ring is contained in a prime ideal.

These are both equivalent to the Boolean Prime Ideal Theorem, which is the statement of Lemma 1.21 for Boolean algebras (and also equivalent to 1.21 for arbitrary distributive lattices). The Boolean prime ideal theorem also implies the Compeleteness Theorem for propositional logic.

(6′) Every product of compact Hausdorff spaces is compact.

Łoś–Ryll-Nardzewski (1954) proved (6′) $\Leftrightarrow$ Boolean Prime Ideal Theorem.

J.D.Halpern (1969) constructed a model of set theory in which BPIT holds but the Axiom of Choice fails. However, BPIT implies that every set can be totally ordered (example in Chapter 2) which in turn implies that any family $(a_i \mid i \in I)$ of non-empty finite sets has a choice function. (To prove this, take a total ordering of $\bigcup\{a_i \mid i \in I\}$ and then choose the least element of each $a_i$).

Informally, a CARDINAL is an equivalence class of sets under the equivalence relation

$$a \equiv b \iff (\exists f)\,(f \text{ a bijection from } a \text{ to } b)$$

Apart from $\{\emptyset\}$, all such classes are proper classes, so we need to find a way of representing them by sets. If we assume AC, then every equivalence class contains an ordinal, so it contains a least ordinal, which we use to represent the class.

**Definition 6.13** We say an ordinal $\alpha$ is INITIAL is there is no bijection from $\alpha$ to any smaller ordinal.

All finite ordinals are initial. We can enumerate the infinite initial ordinals as a sequence $(\omega_\alpha \mid \alpha \in O_n)$ as follows:

$$\begin{aligned} \omega_0 &= \omega \\ \omega_{\alpha+} &= \gamma\,(\omega_\alpha) \\ \omega_\lambda &= \bigcup \{\omega_\beta \mid \beta < \lambda\} \qquad \text{if } \lambda \text{ is a non-zero limit ordinal.} \end{aligned}$$

Given cardinals $m$ and $n$, we write $m \leq n$ to mean that there is an injection from a set of cardinality $m$ to one of cardinality $n$. The Cantor-Bernstein Theorem (1.12) implies that $\leq$ is a partial order. We saw earlier that AC is equivalent to saying that $\leq$ is a total order.

If $A$ and $B$ are sets of cardinality $m$ and $n$, we write $m+n$ for the cardinality of the set $A \sqcup B$, $m.n$ for the cardiality of $A \times B$ and $m^n$ for the cardinality of the set $A^B$ of all functions $B \to A$.

When thinking of $\omega_\alpha$ as a cardinal, we normally denote it $\aleph_\alpha$ (so $\aleph_\alpha.\aleph_\beta$ denotes the cardinal product whereas $\omega_\alpha.\omega_\beta$ denotes the ordinal product).

**Lemma 6.14** *The identities*

$$\begin{aligned} m+n &= n+m \\ m.n &= n.m \\ m.(n+p) &= m.n + m.p \\ m^{n+p} &= m^n.m^p \\ m^{n.p} &= (m^n)^p \\ (m.n)^p &= m^p.n^p \end{aligned}$$

*all hold.*

*Proof:* For example, to prove $(m.n)^p = m^p.n^p$, let $A, B, C$, be sets of ordinalities $m, n, p$. We have a bijection $(A \times B)^C \to A^C \times B^C$ which sends a function $f : C \to A \times B$ to the pair $\langle \pi_1 \circ f\,,\ \pi_2 \circ f \rangle$ where $\pi_i$ are the projections. □

**Lemma 6.15** *For any ordinals $\alpha, \beta$, we have*

$$\aleph_\alpha + \aleph_\beta = \aleph_\alpha.\aleph_\beta = \aleph_{\max\{\alpha,\beta\}}$$

*Proof:* Assume $\alpha \leq \beta$. Then we have

$$\aleph_\beta \leq \aleph_\alpha + \aleph_\beta \leq \aleph_\alpha.\aleph_\beta \leq \aleph_\beta.\aleph_\beta$$

so it suffices to prove $\aleph_\beta.\aleph_\beta \leq \aleph_\beta$. To show this, we construct a well-ordering of $\omega_\beta \times \omega_\beta$ in which every proper initial segment has cardinality $< \aleph_\beta$. We define the ordering by

$$\begin{aligned} \langle \gamma, \delta \rangle < \langle \gamma', \delta' \rangle \iff & \text{either} \gamma \cup \delta < \gamma' \cup \delta' \\ & \text{or } \gamma = \gamma \cup \delta = \gamma' \cup \delta' \text{ and } \delta < \delta' \\ & \text{or } \delta = \delta' = \gamma \cup \delta = \gamma' \cup \delta' \text{ and } \gamma < \gamma' \end{aligned}$$

Easy to check that this is a well-ordering, and for any $\langle \gamma, \delta \rangle$, the initial segment it determines is contained in $(\gamma \cup \delta) \times (\gamma \cup \delta)$. Now the cardinality of $\gamma \cup \delta$ is either finite or equal to $\omega_\theta$ for some $\theta < \beta$, whence by induction on $\beta$ we have $\text{card}\,((\gamma \cup \delta) \times (\gamma \cup \delta)) < \aleph_\beta$. Hence the ordering has order-type $\leq \omega_\beta$, so we have an injection $\omega_\beta \times \omega_\beta \to \omega_\beta$. □

For cardinal exponentation we have, for example: Lecture 24

**Lemma 6.16** *If $\alpha \leq \beta$ then $\aleph_\alpha{}^{\aleph_\beta} = 2^{\aleph_\beta}$*

*Proof:*

$$2^{\aleph_\beta} \leq {\aleph_\alpha}^{\aleph_\beta} \leq \left(2^{\aleph_\alpha}\right)^{\aleph_\beta} = 2^{(\aleph_\alpha \aleph_\beta)} = 2^{\aleph_\beta}$$

so the result holds by Cantor-Bernstein. □

*Problem:* What can we say about the function $O_n \to O_n$ defined by $2^{\aleph_\beta} = \aleph_{f(\beta)}$ ? Clearly $f(\beta) > \beta$ and $\beta \leq \gamma \Longrightarrow f(\beta) \leq f(\gamma)$. There are a few other restrictions, e.g. $f(0) \neq \omega$, but otherwise almost anything consistent with these conditions is possible.

# Chapter 7

# Consistency and Independence

The original aim in axiomatizing set theory was to produce a theory $T$ which would be consistent and COMPLETE, i.e. for every sentence $p$ in the language of $T$ we'd have either $T \vdash p$ or $T \vdash \neg p$.

However, K.Gödel (1931) showed that this aim can't be achieved by any recursively presented theory. Clearly, Zermelo-Fraenkel is recursively presented, so we can code its formulae and proofs by natural numbers. But we have a 'copy of $\mathbb{N}$' inside any model of ZF; more explicitly we have a (recursive) interpretation of Peano Arithmetic (formulated in a language with $0, s, +, \times$ as primitives) in ZF. In Peano Arithmetic, we can write down a formula with two free variables $x, y$ which asserts that $x$ is the code for a proof in ZF of the formula coded by $y$. Hence in ZF we can write down a formula $g$ which says 'ZF can't prove $g$'. Then neither $g$ nor $\neg g$ can be proved in ZF unless ZF is inconsistent. We can also write down a formula $\mathrm{Con}_{\mathrm{ZF}}$ which asserts 'ZF can't prove $\perp$', and in fact $\mathrm{ZF} \vdash (g \Leftrightarrow \mathrm{Con}_{\mathrm{ZF}})$, so if ZF is consistent then $\mathrm{ZF} \nvdash \mathrm{Con}_{\mathrm{ZF}}$. We could define $\mathrm{ZF}^+$ to be the theory $\mathrm{ZF} \cup \{\mathrm{Con}_{\mathrm{ZF}}\}$, then $\mathrm{ZF}^+ \vdash \mathrm{Con}_{\mathrm{ZF}}$, but $\mathrm{ZF}^+ \nvdash \mathrm{Con}_{\mathrm{ZF}^+}$.

The most we can hope to do is to prove RELATIVE CONSISTENCY results of the form

$$\mathrm{ZF} \vdash (\mathrm{Con}_{T_1} \Rightarrow \mathrm{Con}_{T_2})$$

where $T_1$ and $T_2$ are theories related to ZF in some way. To prove this we need to construct an interpretation of $T_2$ in $T_1$, but we usually think of this as a way of constructing a model of $T_2$ from a model of $T_1$.

If $p$ is one of the axioms of ZF, an INDEPENDENCE PROOF for $p$ is a proof of

$$\mathrm{Con}_{\mathrm{ZF}} \Longrightarrow \mathrm{Con}_{(\mathrm{ZF} \setminus \{p\} \cup \{\neg p\})}$$

A RELATIVE CONSISTENCY PROOF for $p$ is a proof of

$$\mathrm{Con}_{\mathrm{ZF} \setminus \{p\}} \Longrightarrow \mathrm{Con}_{\mathrm{ZF}}$$

A standard interpretation is one which involves the passage from a model $(V, \in)$ of ZF to a structure $(M, \in|_{M \times M})$ where $M$ is a transitive class. For example, taking $M = V_\omega$ = class of hereditarily finite sets, we get a model for all of ZF except the axiom of infinity. $M = V_{\omega+\omega}$ is a model for all axioms except Replacement.

For relative consistency of Foundation, we can replace $V$ by its 'well-founded part'. For independence of Foundation we can't use this method. Instead, we replace $(V, \in)$ by $(V, \in')$ where $x \in' y$ is defined to mean $x \in \sigma(y)$ where $\sigma$ is the permutation which interchanges $\phi$ and $\{\phi\}$ and leaves everything else fixed. In this structure, $\phi \in' \phi$ and $\{\phi\}$ plays the role of the empty set.

Gödel (1938) proved the relative consistency of the Axiom of Choice using a standard model $L$ (the constructible universe) which is defined as $\bigcup \{L_\alpha | \alpha \in O_n\}$ with $L_{\alpha+1}$ consisting on those subsets of $L_\alpha$ which are definable by first-order formulae with parameters in $L_\alpha$. There's a definable well-ordering on $L$, and it satisfies all the axioms of ZF.

Fraenkel (1822) and Mostowski (1939) produced independence proofs for AC relative to weaker versions of set theory. Fraenkel used ATOMS with no members, Mostowski used sets satisfying $x = \{x\}$.

P.Cohen (1963) developed a technique of 'forcing' which enables one to freely adjoin an injection $x \to \mathcal{P}\omega$ for an arbitrary set $x$. Applying this to Mostowski's model, one gets the failure of AC for $\mathcal{P}\omega$. Cohen's method also enables one to construct models of ZF in which the continuum hypothesis fails, i.e. $2^{\aleph_0} > \aleph_1$.