Lecture Notes
Philosophical Logic
PHIL20060

Johannes Stern
Department of Philosophy
University of Bristol
johannes.stern@bristol.ac.uk

January 2025

Preface

These lecture notes have been specifically designed for the unit Philosophical Logic (PHIL20060) at the University of Bristol. They draw on a variety of different lecture notes, textbooks. In particular they draw on material presented in

- Forall x Calgary
- *Philosophical Logic* by John MacFarlane
- *Philosophical Logic* by John Burgess
- *Modal Logics for Philosophers* by James W. Garsons

However our presentation often differs substantially from these textbooks.

These notes are regularly updated. Up-to-date versions can be downloaded from:

johannesstern.github.io/teaching/LectureNotesPHIL20060

Contents

Part I

Introduction

Chapter 1

Mathematical Toolkit

In our metalanguage, that is, the language we use to talk about the different systems of logic and the language we use to give the various semantics for the differ systems of logic we discuss we appeal to basic concepts from set-theory.

A **set** S is a collection of objects. There are two ways to specify a set. The first one is to write $\{o_1, o_2, o_3\}$ if the set contains exactly the three objects o_1, o_2, and o_3. However, if a set contains many objects this is not always practicable. This leads us to the second way to specify. We write $\{o \mid \Phi(o)\}$ meaning that the set consists of all objects o that have the property Φ.[1] For example, using this way of forming sets we can form the set of consisting of all even numbers:

$$E := \{n \mid n \text{ is an even number}\}.$$

If S is a set and we write **o** $\in$ S, then o **is a member** of S. For example, if $n \in E$ this means that n is an even number. Sets are *extensional*, which means that for sets S, S', is S and S' have the same members, then $S = S'$, i.e., S and S' are the same set.

Sets can have many members, infinitely many in fact as illustrated by considerng set E defined above. But a set can also have no member. For example, the set $\{o \mid o \neq o\}$ has no member—it is empty. The **empty set** is denoted by $\emptyset$.

Subset, Intersection, and Union The **union** of two sets A and B consists of the members of both A and B an denoted by $A \cup B$:

$$A \cup B := \{o \mid o \in A \text{ or } o \in B\}.$$

The **intersection** of two sets A and B consists of those objects that are members of both A and B and denoted by $A \cap B$:

$$A \cap B := \{o \mid o \in A \text{ and } o \in B\}.$$

[1] In standard one cannot always form the set $\{o \mid \Phi(o)\}$, since otherwise paradoxes—in particular *Russell's paradox*—may arise. For purposes this problems do not concern, but you might think about what happens if one forms the set $\{o \mid o \notin o\}$

A is a **subset** of B, if all members of A are also members of B and denoted by $A \subseteq B$:

$$A \subseteq B : \text{iff if } o \in A, \text{ then } o \in B.$$

For two sets A and B either $A \subseteq B$ or $B \subseteq A$, or there are objects which are members of A but not in B *and* objects that are members if B bur not in B.

For example, let $U := \{n \mid n \text{ is an uneven number}\}$ and $\mathbb{N}$ the set of natural numbers. The we have $U \subseteq \mathbb{N}$ and $E \subseteq \mathbb{N}$, but neither $U \subseteq E$ nor $E \subseteq U$. Indeed, we have $U \cap E = \emptyset$ and $U \cup E = \mathbb{N}$ (we day U and E are complements of each other relative to $\mathbb{N}$.

Ordered Pairs, Tuples, Relations and Functions In a set the order of the object doesn't matter. $\{o_1, o_2\}$ and $\{o_2, o_1\}$ are the same set. Sometimes order is important however and this leads to ordered pairs.

In an **ordered pair** $\langle o, u \rangle$ order is important. It matters that o is in first position and u in second position. Confusingly, ordered pairs can be defined in set theory, but they are not the set $\{o, u\}$.[2]

We can also define triples of objects $\langle o_1, o_2, o_3 \rangle$, quadruples, and, more general ***n*-tuples** $\langle o_1, \ldots, o_n \rangle$ where n is any natural number.

A **binary relation** is a set that has only ordered pairs as members and an ***n*-ary relation** is a set that has only n-tuples as members. Let $S_1, S_2, \ldots, S_n$ be sets. Then the **Cartesian product** between S_1 and S_2 is the set of ordered pairs with its first element from S_1 and its second element from S_2. The Cartesian product between S_1 and S_2 is denoted by $S_1 \times S_2$ and defined as follows:

$$S_1 \times S_2 := \{\langle o_1, o_2 \rangle \mid o_1 \in S_1 \text{ and } o_2 \in S_2\}.$$

A binary relation R between S_1 and S_2 is subset of $S_1 \times S_2$. Similarly, an n-ary relation between $S_1, \ldots, S_n$ is a subset of $S_1 \times \ldots \times S_n$.

- $\leq := \{\langle n, m \rangle \mid \langle n, m \rangle \in \mathbb{N} \times \mathbb{N} \text{ and } n \leq m\}$ is a binary relation on $\mathbb{N}$.
- $\{\langle n, m \rangle \mid \langle n, m \rangle \in \mathbb{N} \times U \text{ and } m = 2n + 1\}$ is a binary relation on $\mathbb{N} \cup U$.
- $\{\langle n, m, k \rangle \mid \langle n, m, k \rangle \in \mathbb{N} \times \mathbb{N} \times \mathbb{N} \text{ and } n \leq m \text{ and } m \leq k\}$ is a ternary relation on $\mathbb{N}$.

A **function** f is a binary relation **such that** for $f \subseteq S_1 \times S_2$: if $\langle o, u \rangle \in f$ and $\langle o, u_1 \rangle \in f$, then $u = u_1$. This means that if $f \subseteq S_1 \times S_2$ an element $o \in S_1$ can be related to at most one element of S_2.

Let X and Y be sets. We write $f : X \longrightarrow Y$ iff $f \subseteq X \times Y$, f is a function and *for all* $x \in X$ there is a $y \in Y$ such that $\langle x, y \rangle \in f$.

- $f = \{\langle n, m \rangle \mid \langle n, m \rangle \in \mathbb{N} \times U \text{ and } m = 2n + 1\}$ is a function $f : \mathbb{N} \longrightarrow U$.

[2]The standard way of defining ordered pairs set-theoretically is due to Wiener and Kuratowski:

$$\langle o, u \rangle := \{\{o\}, \{o, u\}\}.$$

That is, the $\langle o, u \rangle$ is defined as the set which has as its element a set with only o as its member (the singleton of o), and the set that both o and u as members.

Countable vs Uncountable Finite sets (sets with finitely many members) are **countable sets**. Let S be a set with infinitely many members. Then S is a **countable set** iff there is function $f : S \longrightarrow \mathbb{N}$ such that for every natural number n there is an $o \in S$ such that $f(o) = n$ and if $f(o) = f(u)$, then $o = u$.[3] It is uncountable otherwise. For example, the sets E and U are countable (and of course $\mathbb{N}$ itself), but the set of real numbers in not countable, as there is no such function.

Powerset A set is different to its members, e.g.,

$$o \neq \{o\} \neq \{\{o\}\} \neq \{\{\{o\}\}\} \neq \ldots$$

and similarly

$$\{1,2\} \neq \underbrace{\{\emptyset, \{1\}, \{2\}, \{1,2\}\}}_{S}.$$

If we inspect the set S above closely we notice that S consists of all the subsets of $\{1,2\}$ ($\emptyset$ is a subset of every set), i.e.

$$S := \{X \,|\, X \subseteq \{1,2\}.$$

S is called the **powerset** of $\{1,2\}$. The powerset of a given set S is denoted by $\mathcal{P}(S)$:

$$\mathcal{P}(S) := \{X \,|\, X \subseteq S\}.$$

Notice that if S has n members, then $\mathcal{P}(S)$ has 2^n members.

[3]Such a function f is called injective: a function $f : X \longrightarrow Y$ is injective iff $f(x) \neq f(x_1)$ for all $x, x_1 \in X$

Chapter 2

Classical Logic

In this section we introduce the basics of classical propositional logic, i.e., truth-functional logic (TFL).

2.1 Syntax

The language of propositional logic has the following vocabulary.

- a countable set of propositional variables:

  $$\mathsf{At}_{\mathcal{L}} := \{p_0, p_1, p_2, \dots\}$$

  (we often use 'p','q', or'r' instead to avoid the use of indices)

- parenthesis: '(' and ')'

- the propositional constant '$\bot$'

- the logical connectives (operator) '$\longrightarrow$' (read: 'if…, then…'), '$\wedge$' (read: 'and'), and '$\vee$' (read: 'or').[1]

We use greek letters φ, ψ, χ, … as metavariables for $\mathcal{L}$ formulae (please read Chapter 4 of forall$x$-Bristol to learn more about metavariables and the use/mention distinction):

Definition 1 (Formula)**.** *Well-formed formulae (sentences) of $\mathcal{L}$ are defined by the following inductive definition:*

1. *Propositional variables are $\mathcal{L}$ formulae,*
2. *$\bot$ is an $\mathcal{L}$-formula,*
3. *φ and ψ are formulae, then $(\varphi \wedge \psi), (\varphi \vee \psi)$, and $(\varphi \longrightarrow \psi)$ are $\mathcal{L}$-formula,*[2]
4. *nothing else is a formula.*

[1] In classical logic it would suffice to only have the logical connective '$\longrightarrow$' in our language and define '$\wedge$' and '$\vee$' (cf. below). This, however, does not work for intuitionistic logic.

[2] To be fully precise we should write ' '(' $\varphi \frown$ '$\wedge$' $\frown \psi$ ')' ' instead of '$(\varphi \wedge \psi)$' and similarly for $(\varphi \vee \psi)$ and $(\varphi \longrightarrow \psi)$. See Chapter 6 of forall$x$-Calgary for discussion.

Bracket conventions Brackets guarantee that there is one unique way of reading formulas and that there is no ambiguity. However, some brackets are redundant and can be omitted. We assume the following conventions:

- Outer brackets can be omitted;
- $\wedge$ and $\vee$ bind stronger than $\longrightarrow$, that is, we may write $\varphi \wedge \psi \longrightarrow \chi$ instead of $(\varphi \wedge \psi) \longrightarrow \chi$.

But of course brackets remain of crucial importance, e.g., (as we shall see) $(\varphi \wedge \psi) \vee \chi$ is different too $\varphi \wedge (\psi \vee \chi)$.

Defined symbols We introduce further logical connectives, i.e. $\neg$ ('not', 'it is not the case that') and $\leftrightarrow$ ('if and only if') by definition:

$$(\neg) \qquad \neg\varphi \equiv_{\mathsf{def}} (\varphi \longrightarrow \bot)$$

$$(\leftrightarrow) \qquad (\varphi \leftrightarrow \psi) \equiv_{\mathsf{def}} ((\varphi \longrightarrow \psi) \wedge (\varphi \longrightarrow \psi))$$

Officially, $\neg\varphi$ and $(\varphi \leftrightarrow \psi)$ are mere notational abbreviation, that is, a short way of writing $(\bot \longrightarrow \varphi)$ and $((\varphi \longrightarrow \psi) \wedge (\varphi \longrightarrow \psi))$ respectively.

2.1.1 An aside: proof by induction

To prove that something holds for all formulas of the language, we often use so-called proofs by induction. Perhaps you have already encountered proofs by induction in maths. For example, using a proof by induction one can show that all multiples of 4 can be divided by 2, that is, one can shows that $\frac{4\times n}{2} \in \mathbb{N}$ for all $n \in \mathbb{N}$. To this effect, one first shows that the claim holds for $n = 0$, which follows from the fact that $0 \in \mathbb{N}$. Now we assume that we already showed the claim for the number m and then show that in this case the claim also holds for $m + 1$. That is, we assume that $\frac{4\times m}{2} \in \mathbb{N}$. This is called the **Induction Hypothesis** (IH). Now we wish to show that $\frac{4\times(m+1)}{2} \in \mathbb{N}$:

$$\frac{4 \times (m + 1)}{2} = \frac{(4 \times m) + 4}{2} = \frac{4 \times m}{2} + 2$$

Since by IH $\frac{4\times m}{2}$ is a natural number and 2 is also a natural number, we can infer that $\frac{4\times(m+1)}{2} \in \mathbb{N}$. On the basis of this we can infer that the claim holds for **all** natural numbers.

In more abstract terms, proof by induction proceeds as follows:

1. One shows that some property Φ holds for 0
2. Next one shows that if Φ holds for an arbitrary natural number n (the induction hypothesis), then Φ also holds for $n + 1$.
3. Then one can conclude that Φ holds for **all** natural numbers.

Frequently, an alternative form of induction is employed, so-called **strong induction**. If one uses strong induction then Step 2 above is altered as follows:

2’ As induction hypothesis one assumes that for all $m < n$ for an arbitrary natural number n the property Φ holds. Using this induction hypothesis one then shows that then Φ also holds for n.

If 1 and 2’ have been established, then one can again infer that Φ holds for all natural numbers.

How can we use a proof by induction to prove something about all formulas? For example, one can prove that all well-formed formulas have an even number of brackets. To this effect, we count the number of embeddings of logical connectives of a formula, that is, the so-called complexity of a formula. The complexity of a formula is defined by the following inductive definition:

- propositional variables and $\bot$ have complexity 0;
- if $\varphi \doteq (\psi \rightarrow \chi)$ or $\varphi \doteq (\psi \wedge \chi)$ or $\varphi \doteq (\psi \vee \chi)$, then the complexity of φ is the maximum of the complexities of ψ and χ plus 1.

One then shows using a proof by strong induction that formulas of any complexity have an even number of brackets, i.e., all formula have an even number of brackets:

- if φ has complexity 0, then φ is a propositional variable or constant and has 0 brackets.
- Now assume that φ has complexity $n > 0$. Then $\varphi \doteq (\psi \rightarrow \chi)$ or $\varphi \doteq (\psi \wedge \chi)$ or $\varphi \doteq (\psi \vee \chi)$. By Induction Hypothesis we can assume that ψ and χ have an even number of brackets, as they have a complexity $< n$. Therefore $(\psi \rightarrow \chi), (\psi \wedge \chi)$, and $(\psi \vee \chi)$ have an even number of brackets. We can infer that φ has an even number of brackets.
- We can conclude that all formulas have an even number of brackets.

2.2 Classical Semantics for Propositional Logic

The aim of the semantics is to say under which conditions (not whether) a sentence is true or false, and whether a sentence logically follows from a set of sentences. A key notion of the semantics is that of a (classical) valuation.

Definition 2 (Valuation). *A valuation is a function $v : At_{\mathcal{L}} \rightarrow \{\mathsf{T}, \mathsf{F}\}$ that assigns to every propositional variable either the value true (T) or the value false (F).*[3]

This means that for some propositional variable p_j we have $v(p_j) \in \{\mathsf{T}, \mathsf{F}\}$. Our next job is to specify the truth value of complex formulas relative to a valuation v. To this effect we define, relative to a valuation v, an interpretation function $\mathtt{I}_v$ that assigns to every $\mathcal{L}$-formula either the value T or the value F.

Definition 3 (Interpretation). *Let v be a valuation. The I_v is is a function that assigns to every formula φ either the value* T *or the value* F *such that:*

(i) if φ is a propositional variable, then $\mathtt{I}_v(\varphi) = v(\varphi)$;

[3]In the literature the truth values are often denoted by 1 and 0 instead of T and F.

(ii) if φ is $\bot$, then $\mathtt{I}_v(\varphi) = \mathsf{F}$*;*

(iii) if φ is a formula $(\psi \wedge \chi)$, then:

$$\mathtt{I}_v(\varphi) := \begin{cases} \mathsf{T}, & \text{if } \mathtt{I}_v(\psi) = \mathtt{I}_v(\chi) = \mathsf{T} \\ \mathsf{F}, & \text{otherwise}; \end{cases}$$

(iv) if φ is a formula $(\psi \vee \chi)$, then:

$$\mathtt{I}_v(\varphi) := \begin{cases} \mathsf{T}, & \text{if } \mathtt{I}_v(\psi) = \mathsf{T} \text{ or } \mathtt{I}_v(\chi) = \mathsf{T} \\ \mathsf{F}, & \text{otherwise}; \end{cases}$$

(v) if φ is a formula $(\psi \longrightarrow \chi)$, then:

$$\mathtt{I}_v(\varphi) := \begin{cases} \mathsf{T}, & \text{if } \mathtt{I}_v(\psi) = \mathsf{F} \text{ or } \mathtt{I}_v(\chi) = \mathsf{T} \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

If $I_v(\varphi) = \mathsf{T}$ we also write $v \Vdash \varphi$.

This might sound fairly abstract but clauses (iii) - (v) can be neatly presented in terms of truth-tables:

φ	ψ	$\varphi \wedge \psi$
T	T	T
T	F	F
F	T	F
F	F	F

φ	ψ	$\varphi \vee \psi$
T	T	T
T	F	T
F	T	T
F	F	F

φ	ψ	$\varphi \longrightarrow \psi$
T	T	T
T	F	F
F	T	T
F	F	T

Using truth tables we can determine whether a formula is true or false relative to different valuations. For example, we can ask which truth values a valuation needs to assign to the propositional variable p and q for the formula $p \vee q \longrightarrow \bot \vee q$. We can construct a truth table for the formula as follows:

1. create a vertical column for each propositional variable;
2. create a horizontal row for each possible combination of assigning truth values to the propositional variable (each possible valuation);
3. create a vertical column for each subformula of the target formula (including the target formula);
4. successively complete each horizontal row.

Valuation	p	q	$\bot$	$p \vee q$	$\bot \vee q$	$p \vee q \longrightarrow \bot \vee q$
v_1	T	T	F	T	T	T
v_2	T	F	F	T	F	F
v_3	F	T	F	T	T	T
v_4	F	F	F	F	F	T

For 2 propositional variables there are 4 different ways to assign truth values to the propositional variables, that is, four different valuations. More generally for n propositional variables there are 2^n different ways of assigning truth values to the propositional variables, that is, 2^n different possible valuations.

Looking at the truth table we can see that $p \vee q \longrightarrow \bot \vee q$ is true if q is true or both p and q are false. So $p \vee q \longrightarrow \bot \vee q$ relative to valuations v_1, v_3 and v_4, that is, $v_j \Vdash p \vee q \longrightarrow \bot \vee q$ for $j \in \{1, 3, 4\}$.

We have specified when an $\mathcal{L}$-formula is true relative to a valuation. The interesting aspect of the semantics is that once the truth values of the propositional variables occurring in a given formula have been specified, the truth value of the formula is fully determined. This is implicit in our Definition 3 and becomes apparent by looking at the truth table of a formula. It is also the reason why classical propositional logic is also called truth-functional logic.

Truth-functionality For a truth-functional logic we expect the following principle to hold for all formulas φ, ψ, and χ:

$$\text{if } \mathrm{I}_v(\psi) = \mathrm{I}_v(\chi), \text{ then } \mathrm{I}_v(\varphi(\psi)) = \mathrm{I}_v(\varphi(\chi)).$$

This means that the truth-value of a formula does not change if we replace (substitute) a subformula of φ, i.e. ψ, by a formula χ that has the same truth truth-value as ψ. In this course we will encounter several logics for which this does not hold. Such logics are not truth-functional and sometimes called intensional logics.

2.2.1 Logical consequence and other semantic notions

The central notion of logic is that of **logical consequence**. A formula φ is a logical consequence of a set of formulas Γ iff for all valuations such that all members of Γ are true, φ is also true. In other words φ is a logical consequence of Γ iff there is no valuation such that all members of Γ are true, but φ is false.

Definition 4 (Logical Consequence, Logical Equivalence)**.** *Let φ be a formula and Γ a set of formulas. Then φ is a logical consequence of Γ (in symbols: $\Gamma \vDash \varphi$) iff*

$$\textit{for all valuations } v\textit{: if } v \Vdash \gamma \textit{ for all } \gamma \in \Gamma\textit{, then } v \Vdash \varphi.$$[4]

If $\{\varphi\} \vDash \psi$ and $\{\psi\} \vDash \varphi$ we say that φ and ψ are **logically equivalent**.

We will sometimes be sloppy with notation and write $\varphi \vDash \psi$ or $\varphi, \psi \vDash \chi$ instead of the correct $\{\varphi\} \vDash \psi$ and $\{\varphi, \psi\} \vDash \chi$ respectively.

We say a formula φ **logically follows** from a set Γ iff φ is a logical consequence of Γ. Notice that a formula can logical follow from the emptyset, i.e., if $\Gamma = \emptyset$. In this case, we write $\vDash \varphi$ and call φ a **logical truth**. By definition a logical truth is true relative to every valuation (Exercise:

[4] As mentioned in the paragraph leading into Definition 4 the definiens is equivalent to: there is no valuations v such that $v \Vdash \gamma$ for all $\gamma \in \Gamma$ and not $v \Vdash \varphi$.

why?). In propositional logic a logical truth is often called a **tautology**. A logical truth is also said to be **valid**.[5] In contrast to a tautology that is true relative to every valuation, a **logical contradiction** is false relative to every valuation, that is, a formula φ is a logical contradiction iff $\varphi \vDash \bot$ (Exercise: why?). In particular this means that $\bot$ is a logical contradiction. A formula that is neither a tautology nor a logical contradiction is called **contingent**. If φ is a contingent formula then there is a valuation v_1 such that $v_1(\varphi) = \mathsf{T}$, but there is also a valuation v_2 such that $v_2(\varphi) = \mathsf{F}$.

Finally, a set of formulas Γ is **satisfiable** or **jointly consistent** iff there exists a valuation relative to which all members of Γ are true, i.e., if there exists a valuation v such that $v(\gamma) = \mathsf{T}$ for all $\gamma \in \Gamma$. Γ is called **inconsistent** or **insatisfiable** otherwise. Notice that if $\varphi \in \Gamma$ for some logical contradiction φ, then Γ is inconsistent.

Some logical consequences For all $\mathcal{L}$-formulas φ, ψ and χ:

1. $\varphi \vDash \varphi$
2. $\bot \vDash \varphi$
3. $\varphi \wedge \psi \vDash \varphi$
4. $\varphi \vDash \varphi \vee \chi$
5. $\varphi, \varphi \longrightarrow \psi \vDash \psi$
6. $\neg(\varphi \wedge \psi) \vDash \neg\varphi \vee \neg\psi$
7. $\neg\varphi \vee \neg\psi \vDash \neg(\varphi \wedge \psi)$
8. $\varphi \vee \psi \vDash \neg(\neg\varphi \wedge \neg\psi)$
9. $\neg(\neg\varphi \wedge \neg\psi) \vDash \varphi \vee \psi$
10. $\varphi \longrightarrow \psi \vDash \neg\varphi \vee \psi$
11. $\neg\varphi \vee \psi \vDash \varphi \longrightarrow \psi$
12. $\varphi \vee \psi, \neg\varphi \vDash \psi$

Exercise 5. *Verify claims 1-12 using truth tables. Can you read of some logical equivalences from these claims?*

[5] Strictly speaking validity is a property of arguments. An argument with premises $\varphi_1, \dots, \varphi_n$ and conclusion ψ is said to be valid iff ψ is a logical consequence of $\{\varphi_1, \dots, \varphi_n\}$. To say that a formula is valid is thus to say that the degenerate argument without premises is valid, that is, that the formula is always/unconditionally true.

Logical truths For all $\mathcal{L}$-formulas φ, ψ and χ:

(a) $\varphi \to \varphi$

(b) $\neg\bot$

(c) $\bot \leftrightarrow \varphi \wedge \neg\varphi$

(d) $\varphi \vee \neg\varphi$

(e) $\varphi \to (\psi \to \varphi)$

(f) $\varphi \to (\psi \to \varphi \wedge \psi)$

(g) $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$

(h) $(\varphi \to \chi) \to ((\psi \to \chi) \to (\varphi \vee \psi \to \chi))$

(i) $(\varphi \to \psi) \to (\neg\psi \to \neg\varphi)$

are logical truths (Exercise: verify using truth tables).

There is an immediate connection between logical truths and logical contradiction: if φ is a logical truths, then $\neg\varphi$ is a logical contradiction and vice versa (Exercise: explain why).

For classical logic we have an important theorem that connects logical consequence and logical truth. This theorem is called the **Deduction theorem**:

Theorem 6 (Deduction theorem)**.** *Let φ and ψ be $\mathcal{L}$-forumae and Γ a set of $\mathcal{L}$-formulae. Then*

$$\Gamma \cup \{\varphi\} \models \psi \textit{ iff } \Gamma \models \varphi \to \psi.$$

In the Deduction theorem we may again take $\Gamma = \emptyset$. Then the Deduction tells us that ψ is a logical consequence of φ iff $\varphi \to \psi$ is a logical truth.

Exercise 7. *Using the Deduction theorem can you:*

(i) add to the list of logical truths by exploiting the logical consequences displayed in 1-12;

(ii) add to list of logical consequences by exploiting the list of logical truths in (a)-(i)?

Exercise 8. *Verify the following claims either using truth tables, or by arguing informally for the claim:*

1. *if $\Gamma \subseteq \Gamma'$ and $\Gamma \models \varphi$, then $\Gamma' \models \varphi$;*
2. *$\Gamma \models \varphi$ and $\Gamma \models \psi$ iff $\Gamma \models \varphi \wedge \psi$;*
3. *if $\Gamma \models \varphi$, then $\Gamma \models \varphi \vee \psi$;*
4. *if $\Gamma \models \varphi$ and $\Gamma \models \varphi \to \psi$, then $\Gamma \models \psi$;*
5. *$\Gamma, \varphi \models \psi$ iff $\Gamma \models \varphi \to \psi$;*
6. *if $\Gamma \models \varphi \vee \psi$, and $\Gamma, \varphi \models \chi$, and $\Gamma, \psi \models \chi$, then $\Gamma \models \chi$;*
7. *if $\Gamma, \varphi \to \bot \models \bot$, then $\Gamma \models \varphi$.*

2.3 Proofs and Natural Deduction

If a formula is a logical consequence of a set Γ, we know that if all formulas in Γ are true, then φ is true like wise. So φ follows from the truth of Γ and we can check that using truth tables. This means that for classical propositional logic we can always *decide* whether $\Gamma \vDash \varphi$ or not. However, this is peculiar for classical propositional logic and is not the case for first-order logic, i.e., for first-order logic we cannot always decide whether $\Gamma \vDash \varphi$ or not. Moreover, truth tables are not always practicable and can quickly grow out of hand if formulas grow bigger and contain many propositional variables. We would like to have a different method for establishing that a formulas φ follows from a set of formulae Γ.

This leads to so-called deductive systems and the idea of **deriving** a formula φ from a set of formulae Γ. The are a variety of different deductive systems and different types of derivation but the basic idea is to give some basic rules, which tell us when we can infer a formula from some other formulas. E.g. on such rule is the rule of *modus ponens* or *conditional elimination* which says that if one has derived the formula φ and one has also derived the formula $\varphi \longrightarrow \psi$, then one can infer the formula ψ. Notice that if φ and $\varphi \longrightarrow \psi$ are true, then we are guaranteed that ψ is also true (cf. 5 above). Modus ponens is a **truth preserving** rule and that is a property all rules need to have because ultimately this guarantees that if we derive a formula φ from a set of formulae Γ, then $\Gamma \vDash \varphi$. This property is called **soundness** and we require all the deductive systems to be sound in that sense.

2.3.1 Natural Deduction

We now present a (Fitch-style) natural deduction system (cf. forall$x$-Calgary but note the different choice of logical connectives. A proof of φ from a set of formual $\Gamma = \{\gamma_1, \dots, \gamma_n\}$ will look as follows:

$$\begin{array}{c|ll}
1 & \gamma_1 & \\
\vdots & \vdots & \\
n & \gamma_n & \\
\cline{2-2}
\vdots & \vdots & \\
l & \psi & \text{Rule, j, k} \\
\vdots & \vdots & \\
m & \varphi & \text{Rule, i, j, l}
\end{array}$$

Each line of the proof is numbered. One starts the proof by listing the premises and then draws a horizontal line underneath. From then on one has to say for each line of the proof by which rule and using which previous lines one has derived the line, that is, unless one makes a further assumption and starts a **subproof**. Every subproof start with an additional assumption that needs to be eventually discharged using some rule, e.g., the rule of **conditional introduc-**

tion:

$$
\begin{array}{l|l l}
\vdots & \vdots & \\
m & \quad \begin{array}{|l} \varphi \\ \hline \vdots \\ \psi \end{array} & \\
\vdots & & \\
k & & \text{Rule j,n} \\
k+1 & \varphi \longrightarrow \psi & \longrightarrow\text{I}, m{-}k \\
\vdots & \vdots &
\end{array}
$$

In line we have introduced the additional assumption φ, which enabled us to infer ψ. We then used the rule of conditional introduction to **discharge** the assumption and to derive $\varphi \longrightarrow \psi$. Without discharging φ everything we derive will be conditional on φ (on the truth of φ). In addition to conditional introduction the rules *disjunction elimination* and *proof by contradiction* (also called *reductio ad absurdum*) are rules that makes use of subproofs and allow us to close specific subproof and discharge specific assumptions. Importantly, once a subproof has been closed, i.e., the assumption discharged, we cannot any line of the subproof at a later stage in the proof. Otherwise, we would reintroduce formulas that can only be proved if we rely on the additional, now discharged assumption.

The basic rules of our system of natural deduction

Conjunction Introduction

$$
\begin{array}{l|l l}
m & \varphi & \\
n & \psi & \\
\vdots & \vdots & \\
j & \varphi \land \psi & \land\text{I}, m, n
\end{array}
$$

Conjunction Elimination

$$
\begin{array}{l|l l}
m & \varphi_1 \land \varphi_2 & \\
\vdots & \vdots & \\
n & \varphi_i & \land\text{E}, m
\end{array}
$$

Disjunction Introduction

Disjunction Elimination

$$
\begin{array}{r|ll}
m & \varphi_i & \\
\vdots & \vdots & \\
j & \varphi_1 \lor \varphi_2 & \lor\mathrm{I},\ m
\end{array}
$$

$$
\begin{array}{r|ll}
m & \varphi \lor \psi & \\
\vdots & \vdots & \\
j & \quad \begin{array}{|l} \varphi \\ \hline \vdots \\ \chi \end{array} & \\
\vdots & & \\
n-1 & & \\
n & \quad \begin{array}{|l} \psi \\ \hline \vdots \\ \chi \end{array} & \\
\vdots & & \\
k & & \\
k+1 & \chi & \lor\mathrm{E},\ m,\ j\text{–}n-1,\ n\text{–}k
\end{array}
$$

Conditional Introduction

$$
\begin{array}{r|ll}
m & \quad \begin{array}{|l} \varphi \\ \hline \vdots \\ \psi \end{array} & \\
\vdots & & \\
n & & \\
n+1 & \varphi \longrightarrow \psi & \longrightarrow\mathrm{I},\ m\text{–}n
\end{array}
$$

Conditional Elimination

$$
\begin{array}{r|ll}
m & \varphi & \\
n & \varphi \longrightarrow \psi & \\
\vdots & \vdots & \\
j & \psi & \longrightarrow\mathrm{E},\ m,\ n
\end{array}
$$

Proof by Contradiction

$$
\begin{array}{r|ll}
m & \quad \begin{array}{|l} \varphi \longrightarrow \bot \\ \hline \vdots \\ \bot \end{array} & \\
\vdots & & \\
n & & \\
n+1 & \varphi & \mathrm{PbC},\ m\text{–}n
\end{array}
$$

Rules for defined connectives

- **Negation**

$$
\begin{array}{r|ll}
m & \varphi \longrightarrow \bot & \\
\vdots & \vdots & \\
j & \neg\varphi & \text{Def. } \neg,\ m
\end{array}
\qquad
\begin{array}{r|ll}
m & \neg\varphi & \\
\vdots & \vdots & \\
n & \varphi \longrightarrow \bot & \text{Def. } \neg,\ m
\end{array}
$$

- **Biconditional**

m	$(\varphi \to \psi) \wedge (\psi \to \varphi)$			m	$\varphi \leftrightarrow \psi$	
$\vdots$	$\vdots$			$\vdots$	$\vdots$	
j	$\varphi \leftrightarrow \psi$	Def. $\leftrightarrow$, m		n	$(\varphi \to \psi) \wedge (\psi \to \varphi)$	Def. $\leftrightarrow$, m

2.3.2 Proofs and Derivation

If we can **derive** φ **from** Γ **we write** $\Gamma \vdash \varphi$. So $\Gamma \vdash \varphi$ says that there is a proof of φ starting from the formulas in Γ, while $\Gamma \vDash \varphi$ says that φ is a logical consequence of Γ, that is , that φ is true, if all members of Γ are true. Ultimately one wants $\vdash$ and $\vDash$ to coincide, but, as mentioned, a minimal requirement is that the deductive system is sound.

For $\Gamma \vdash \varphi$ there is not one unique way of deriving φ from Γ but infinitely many different ones. We now give some sample proofs:

- $\varphi \vdash \varphi$

1	φ	
2	$\varphi \wedge \varphi$	$\wedge$I, 1, 1
3	φ	$\wedge$E, 2

- $\varphi \longrightarrow \psi \vdash \neg\varphi \vee \psi$

1	$\varphi \longrightarrow \psi$	
2	$\quad \varphi \vee (\varphi \longrightarrow \bot) \longrightarrow \bot$	
3	$\quad\quad \varphi \longrightarrow \bot$	
4	$\quad\quad \varphi \vee (\varphi \longrightarrow \bot)$	$\vee$I, 3
5	$\quad\quad \bot$	$\longrightarrow$E, 2, 4
6	$\quad \varphi$	PbC, 3–5
7	$\quad \varphi \vee (\varphi \longrightarrow \bot)$	$\vee$I, 6
8	$\quad \bot$	$\longrightarrow$E, 2, 7
9	$\varphi \vee (\varphi \longrightarrow \bot)$	PbC, 2–8
10	$\quad \varphi$	
11	$\quad \psi$	$\longrightarrow$E, 1, 10
12	$\quad \neg\varphi \vee \psi$	$\vee$I, 11
13	$\quad \varphi \longrightarrow \bot$	
14	$\quad \neg\varphi$	Def. $\neg$, 13
15	$\quad \neg\varphi \vee \psi$	$\vee$I, 14
16	$\neg\varphi \vee \psi$	$\vee$E, 9, 10–12, 13–16

- $\vdash \varphi \longrightarrow \varphi$

1	$\quad \varphi$	
2	$\quad \varphi \wedge \varphi$	$\wedge$I, 1, 1
3	$\quad \varphi$	$\wedge$E, 2
4	$\varphi \longrightarrow \varphi$	$\longrightarrow$I, 1–3

- $\varphi \longrightarrow (\psi \longrightarrow \varphi \wedge \psi)$

1	$\quad \varphi$	
2	$\quad\quad \psi$	
3	$\quad\quad \varphi \wedge \psi$	$\wedge$I, 1, 2
4	$\quad \psi \longrightarrow \varphi \wedge \psi$	$\longrightarrow$I, 2–3
5	$\varphi \longrightarrow (\psi \longrightarrow \varphi \wedge \psi)$	$\longrightarrow$I, 1–4

- $\varphi \rightarrow \psi, \neg\psi \vdash \neg\phi$

$$
\begin{array}{r|l l}
1 & \varphi \rightarrow \psi & \\
2 & \neg\psi & \\
\hline
3 & \quad \varphi & \\
4 & \quad \psi & \rightarrow\text{E}, 1, 3 \\
5 & \quad \psi \rightarrow \bot & \text{Def. } \neg, 2 \\
6 & \quad \bot & \rightarrow\text{E}, 4, 5 \\
7 & \varphi \rightarrow \bot & \rightarrow\text{I}, 3\text{–}6 \\
8 & \neg\varphi & \text{Def. } \neg, 7
\end{array}
$$

In addition to these basic rules there are so-called **derived rules**. Derived rules are not strictly necessary: everything one can prove using derived rules one can prove using only the basic rules. However, derived rules make our life easier and can be used to shorten many rules substantially. For example, two handy rules are the rules of reiteration and *modus tollens*:

$$
\begin{array}{r|l l}
m & \varphi & \\
\vdots & \vdots & \\
k & \varphi & \text{R}, m
\end{array}
\qquad\qquad
\begin{array}{r|l l}
m & \varphi \rightarrow \psi & \\
\vdots & \vdots & \\
k & \neg\psi & \\
\vdots & \vdots & \\
j & \neg\varphi & \text{MT}, m, k
\end{array}
$$

The first proof we have given above shows you why the rule of reiteration is a derived rule and how we can do without; the last proof does the same for modus tollens.

Exercise 9. *Give a proof of the logical truths (a)-(i) in the natural deduction system using only basic rules. In 1-12 replace the symbol '$\vDash$' by '$\vdash$' and establish the resulting claims by giving proofs in the natural deduction system.*

Definition 10 (Theorems)**.** *Let φ be a formula. We say φ is a* **theorem** *(of classical logic) iff $\vdash \varphi$.*

Definition 10 tells us that a formula is called a theorem iff it can be derived without the use of any premises.

2.4 Soundness and Completeness

2.4.1 Soundness

We already mentioned that we expect a deductive system to be **sound** with respect to our semantics, i.e., if φ is derivable from a set of formulas Γ (the premises), then φ is a logical consequence of Γ :

(Snd) $$\text{If } \Gamma \vdash \varphi, \text{ then } \Gamma \vDash \varphi.$$

To show this we will show that for every line i of a proof in our natural deduction system $\Gamma_i \vDash \varphi_i$ where φ is the formula occurring on this line of the proof and Γ_i is the set of open (undischarged) assumptions on which φ_i depends. Notice that premises are assumptions that do not get discharged in the proof. Thus, if we have $\Gamma \vdash \varphi$. then $\Gamma \subseteq \Gamma_i$ for every line i of the proof.

Recall that $\Gamma \vdash \varphi$ implies that the proof ends with φ and that the only undischarged assumptions of the proof are the member of Γ. If the proof is of, say, length n and we show $\Gamma_n \vDash \varphi_n$, then we have established our claim, that is, $\Gamma \vDash \varphi$. By an induction on the length of the proof we show that for every line i of a natural deduction proof $\Gamma_i \vDash \varphi_i$.

We start with Line 1 of the proof. Then φ_1 is either a premise or, if $\Gamma = \emptyset$, it is an undischarged assumption. In both cases we have $\varphi_1 \in \Gamma_1$ and thus, trivially, $\Gamma_1 \vDash \varphi_1$. (If φ_1 is a premise, then $\Gamma_1 = \Gamma$. Otherwise, i.e., if φ_1 is an undischarged assumption, then $\Gamma_1 = \{\varphi_1\}$.)

As induction hypothesis we now assume that $\Gamma_j \vDash \varphi_j$ for all Lines j with $j < i$ and wish to show that $\Gamma_i \vDash \varphi_i$. We know that φ_j must either be an undischarged assumption or obtained using one of the rules from previous lines of the proof. In the former case we know that $\varphi_i \in \Gamma_i$ and thus $\Gamma_i \vdash \varphi_i$. For the latter case we now need to look at all the different rules in turn and show that the rules are truth preserving.

- **Conjunction Introduction**: Then $\varphi_i \doteq (\varphi_j \wedge \varphi_k)$ with $j, k < i$ and, by Induction Hypothesis, we obtain $\Gamma_j \vDash \varphi_j$ and $\Gamma_j \vDash \varphi_j$. Moreover, we know that $\Gamma_j, \Gamma_k \subseteq \Gamma_i$.[6] Then we can infer that $\Gamma_i \vDash \varphi_j$ and $\Gamma_i \vDash \varphi_k$ which entails $\Gamma_i \vDash \varphi_j \wedge \varphi_k$.

j	φ_j	$\Gamma_j \vDash \varphi_j$ by IH
k	φ_k	$\Gamma_j \vDash \varphi_j$ by IH
$\vdots$	$\vdots$	
i	$\varphi_j \wedge \varphi_k (\doteq \varphi_i)$	$\Gamma_j \cup \Gamma_k \subseteq \Gamma_i \vDash \varphi_i$ using Exercise 8:2

- **Conjunction Elimination**: Exercise—use Exercise 8:2!
- **Disjunction Introduction**: Exercise—use Exercise 8:3!
- **Conditional Elimination**: Exercise—use Exercise 8:4!

[6] Otherwise, Conjunction Introduction would not have been applied correctly.

- **Disjunction Elimination**: Then there are $j < l < k < i$ such that $i = k + 1$ and $\varphi_k = \varphi_l = \varphi_i$ and $\varphi_j \doteq (\varphi_{j_1} \vee \varphi_{j_2})$ with $j_1 < k$ and $j_2 < l$. Moreover by induction hypothesis we have $\Gamma_j \subseteq \Gamma_i$, $\Gamma_i \cup \{\varphi_{j_2}\} = \Gamma_k$, and $\Gamma_i \cup \{\varphi_{j_1}\} = \Gamma_l$. Then by Induction Hypothesis we obtain $\Gamma_i \vDash \varphi_{j_1} \vee \varphi_{j_2}$, $\Gamma_i \cup \{\varphi_{j_1}\} \vDash \varphi_l$ and $\Gamma_i \cup \{\varphi_{j_2}\} \vDash \varphi_k$ which entails $\Gamma_i \vDash \varphi_i$.

j	$\varphi_{j_1} \vee \varphi_{j_2}\ (\doteq \varphi_j)$	$\Gamma_j \vDash \varphi_{j_1} \vee \varphi_{j_2}$ by IH
$\vdots$	$\vdots$	
j_1	$\quad \varphi_{j_1}$	$\Gamma \subseteq \Gamma_{j_1} \vDash \varphi_{j_1}$ by IH
$\vdots$	$\quad \vdots$	
l	$\quad \varphi_l\ (\doteq \varphi_i)$	$\Gamma_{j_1} = \Gamma_l = \Gamma_i \cup \{\varphi_{j_1}\} \vDash \varphi_i$ by IH
j_2	$\quad \varphi_{j_2}$	$\Gamma \subseteq \Gamma_{j_2} \vDash \varphi_{j_2}$ by IH
$\vdots$	$\quad \vdots$	
k	$\quad \varphi_k\ (\doteq \varphi_i)$	$\Gamma_{j_2} = \Gamma_k = \Gamma_i \cup \{\varphi_{j_2}\} \vDash \varphi_i$ by IH
$k + 1$	$\varphi_{k+1}\ (\doteq \varphi_i)$	$\Gamma_i \vDash \varphi_i$ using Exercise 8:6

- **Conditional Introduction**: Exercise—use Exercise 8:5!
- **Proof by Contradiction**: Exercise—use Exercise 8:7!

This completes the proof of the soundness theorem, as we have shown that $\Gamma_i \vDash \varphi_i$ for any line i of a given proof of our natural deduction system.

Theorem 11. *The natural deduction system is sound wrt the semantics for propositional logic, i.e., for all formulas φ and sets of formulas Γ:*

$$\text{if } \Gamma \vdash \varphi, \text{ then } \Gamma \vDash \varphi.$$

Exercise 12. *Complete the missing steps in the proof of the Soundness theorem.*

2.4.2 Completeness

If we are working with a motivated and accepted semantics, soundness is the minimal condition we wish a deductive system to satisfy. Ultimately, we want, as mentioned, our deductive system to match our semantics, that is, in addition to (Snd) we also want the converse direction:

(Cmpl) $$\text{If } \Gamma \vDash \varphi, \text{ then } \Gamma \vdash \varphi.$$

This is known as the **Completeness** of the deductive system. Typically, Completeness is more difficult to establish than Soundness. If you are interested, the proof of Completeness of a deductive system of propositional (as well as quantificational) logic will be discussed in next years logic course.

If a deductive systems is sound and complete relative to the semantics, it is known to be **adequate**:

(Adq) $$\Gamma \vdash \varphi \text{ iff } \Gamma \models \varphi.$$

The system of natural deduction we introduced is adequate with respect to the semantics of propositional logic we introduced.

As a consequence of the adequacy of our natural deduction system we also know that the Deduction theorem will not only hold for '$\models$' but also for '$\vdash$', that is, we have the following equivalence:

$$\Gamma \cup \{\varphi\} \vdash \psi \text{ iff } \Gamma \vdash \varphi \to \psi.$$

Exercise 13. *Can we prove the deduction theorem for '$\vdash$' without appealing to the deduction theorem for '$\models$' and the adequacy of classical propositional logic? Reason informally.*

2.5 Hilbert-style deductive systems

There are different natural deduction systems (see here for an overview and discussion). But there are also different derivation systems/proof methods altogether. A prominent such method are so-called Hilbert-style deductive systems (sometimes also called *axiomatic systems* or the *axiomatic method.* The idea of this proof method is that we start with a set of axioms, that is, principles that are taken as a given. For classical propositional logic a list of axioms are:[7]

Ax1 $\varphi \to (\psi \to \varphi)$;

Ax2 $(\varphi \to (\psi \to \chi)) \to ((\varphi \to \psi) \to (\varphi \to \chi))$;

Ax3 $\varphi_i \to \varphi_1 \vee \varphi_2$ for $i \in \{1, 2\}$;

Ax4 $(\varphi \to \chi) \to ((\psi \to \chi) \to (\varphi \vee \psi \to \chi))$;

Ax5 $\varphi_1 \wedge \varphi_2 \to \varphi_i$ for $i \in \{1, 2\}$;

Ax6 $\varphi \to (\psi \to \varphi \wedge \psi)$;

Ax7 $\bot \to \varphi$;

Ax8 $\neg\varphi \vee \varphi$.

In addition to these axioms we have Modus Ponens (MP) as our **rule of proof**:

$$\frac{\varphi, \varphi \to \psi}{\psi}$$

A **proof** of a formula φ from a set of premises Γ in a Hilbert-style deductive system is a sequence of formulas that ends with the formula φ such that every formula in the sequence is either (i) a member of Γ, (ii) an axiom, or (iii) obtained from previous formulas in the sequence by modus ponens (MP).

[7] There a many different possible axiomatizations of classical propositional logic. Of course in all these axiomatizations one can prove exactly the same formulas.

Example 14. $\vdash p \wedge q \to p \vee q$

1. $p \to p \vee q$ *Ax3*
2. $(p \to p \vee q) \to (p \wedge q \to (p \to p \vee q))$ *Ax1*
3. $p \wedge q \to (p \to p \vee q)$ *1,2, MP*
4. $(p \wedge q \to (p \to p \vee q)) \to ((p \wedge q \to p) \to (p \wedge q \to p \vee q))$ *Ax2*
5. $(p \wedge q \to p) \to (p \wedge q \to p \vee q)$ *3,4, MP*
6. $p \wedge q \to p$ *Ax5*
7. $p \wedge q \to p \vee q$ *5,6, MP*

Rules of proof vs rules in natural deduction From the perspective of classical logic a Hilbert-style deductive system can in some sense be seen as a particular system of natural deduction: axioms are just rules without assumptions. They can be introduced at any point in a natural deduction proof and the rule of MP is basically the rule of conditional elimination. This way of viewing Hilbert-style deductive systems may just about work for classical propositional logic, but it is not quite correct.

The key point is that there is a crucial difference between the rule MP in a Hilbert-style proof system and the rule of conditional elimination in a natural deduction system. The former is a rule of proof in the sense that we can only infer ψ from φ and $\varphi \to \psi$ if φ and $\varphi \to \psi$ have already been derived, i.e., they occur at some previous point in the proof (sequence of formulas). In contrast, conditional elimination may be applied to assumptions, that is, formulas that are merely assumed and have not been established. This shows in that conditional elimination can be used within subproofs.

Due to the deduction theorem for classical logic conditional elimination and the rule MP turn out to be equivalent, but that is only *because* the deduction theorem holds and this cannot be assumed at the outset. The bottom line is that rules of natural deduction can typically be applied to assumptions, whereas rules of proof cannot or at least this cannot always be assumed. From a semantic perspective rules of proof can only be applied to logical truths and consequences of Γ.

2.6 Uniform substitution, propositional variables, propositional constants and the Formality of Logic

Propositional variables do not have a fixed meaning, that is, they do not have a fixed truth value. This sets them apart from propositional constants that do have a fixed mean and truth value: the truth value of '$\bot$' is always false. It is a hallmark of logic that a logic does not discriminate between different propositional variables, that is, the theorems of logic are closed under the rule of *Uniform Substitution*: let $\varphi(p)$ and ψ be a formulas and p be a propositional variable occurring $\varphi(p)$. Then

$$\text{if } \vdash \varphi(p), \text{ then } \vdash \varphi(\psi/p)$$

$\varphi(\psi/p)$ is the formula resulting from $\varphi(p)$ by replacing *every* occurrence of p in $\varphi(p)$ by ψ.[8]

To understand the strength of Uniform substitution it is important to appreciate that ψ is an arbitrary formula: if $\vdash \varphi(p)$, we can take any formula and substitute it fro p and the resulting formula remains derivable. Uniform substitution is a salient feature of the so-called *formality* of logic, i.e., that logic is about the form of sentences and arguments and not about their content: logic is topic neutral. Logic does not force a particular understanding of the propositional variable p on us. This will be the case for all logic we discuss in this unit, be they classical or non-classical.

In contrast, Uniform substitution does not hold for propositional constants, as propositional constants have fixed meaning. For example, in classical logic we have $\vdash \bot \longrightarrow q$ but we do not have, say, $\vdash p \longrightarrow q$, that is, we cannot uniformly replace '$\bot$' in a formula and preserve theoremhood.

[8]There is a more general formulation of Uniform substitution:

$$\text{if } \Gamma(p) \vdash \varphi(p), \text{ then } \Gamma(\psi/p) \vdash \varphi(\psi/p).$$

If a formula $\varphi(p)$ can be derived from Γ, then $\varphi(\psi/p)$ can be derived from $\Gamma(\psi/p)$ where $\Gamma(\psi/p)$ is the set of formulae resulting from Γ if all occurrences of p in members of Γ are replaced by ψ.

Part II

Modal Logics

Chapter 3

Modal Logic: Natural Deduction

- different modal notions
- different modal systems
- modal square of oppositions

3.1 Syntax

The language of modal logic an $\mathcal{L}_\Box$ extends the language of classical propositional logic an $\mathcal{L}$ by the logical operator (connective) '$\Box$'. The well-formed formulas of the language of modal logic are defined as follows:

1. if φ is a propositional variable, then φ is an $\mathcal{L}_\Box$-formula,
2. if φ is $\bot$, then φ is an $\mathcal{L}_\Box$-formula,
3. if φ is a formula, then $\Box\varphi$ is an $\mathcal{L}_\Box$-formula,
4. φ and ψ are formulas, then $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $(\varphi \longrightarrow \psi)$ are $\mathcal{L}_\Box$-formula,
5. nothing else is a formula.

In addition to the new logical operator '$\Box$' we introduce a "possibilty"-operator $\Diamond$ by definition:

$$(\Diamond) \qquad \Diamond\varphi \equiv_{\mathsf{def}} \neg\Box\neg\varphi$$

Strictly speaking $\Diamond\varphi$ is thus just an abbreviation of the formula

$$\Box(\varphi \longrightarrow \bot) \longrightarrow \bot.$$

3.2 Natural Deduction for Modal Logic

To provide a natural deduction system for Modal Logic we need a new type of subproof, that is, a **□-subproof**, which start with the □-symbol instead of a formula:

$$
\begin{array}{l|l}
\vdots & \vdots \\
m & \quad \begin{array}{|l} \Box \\ \hline \vdots \end{array}
\end{array}
$$

So far subproofs were used to introduce additional assumptions. □-subproofs in contrast are very different, they, in general, don't allow the use of *any* lines outside the subproof. There are two exceptions to this rule:

(i) it is a premise marked with ⋆ (a **Global Premise**), or

(ii) we can introduce a formula φ into the □-subproof, if there is a previous line of the form $\Box\varphi$ that is outside the □-subproof (more details below).

The idea is that in a □-subproof we should think of every formula to be prefixed by a □. A normal or *local* premise is assumed truth for the sake of the argument, In contrast if φ is a global premise, we assume not only φ to be true, but also $\Box\varphi$, $\Box\Box\varphi$, $\Box\Box\Box\varphi$, ...[1]

On the basis of these remarks, the following Introduction and Elimination rules for the modal operator hopefully make sense:

□-Introduction rule

$$
\begin{array}{l|lr}
m & \quad \begin{array}{|l} \Box \\ \hline \vdots \\ \varphi \end{array} & \\
\vdots & & \\
n & & \\
n+1 & \Box\varphi & \Box\text{I}, m\text{–}n
\end{array}
$$

No line above Line m may be used in the □-subproof unless they are marked by ⋆ or a rule explicitly permits its .

□-Elimination rule

$$
\begin{array}{l|l}
m & \Box\varphi \\
n & \quad \begin{array}{|lr} \Box & \\ \hline \vdots & \\ \varphi & \Box\text{E}, m \\ \vdots & \end{array} \\
\vdots & \\
k & \\
\vdots & \\
\vdots & \vdots
\end{array}
$$

Line k is not embedded in a □-subproof starting in between Lines n and k and the □-subproof starting Line n is not embedded in a further □-subproof starting in between Lines m and n.

[1]Preempting possible world semantics for modal logic a global premise is true at all possible worlds while a local premise is only required to be true in the present/relevant world.

Why do we have to be so careful with introducing formulas into □-subproofs? If we could cite (use) any odd previous line in a □-subproof we could show that if a formula is true then it is necessary:

1	φ	
2	$\quad$ □	
3	$\quad \varphi$	R, 1
4	$\Box\varphi$	□I, 2–3

1	$\quad \varphi$	
2	$\quad\quad$ □	
3	$\quad\quad \varphi$	R, 1
4	$\quad \Box\varphi$	□I, 2–3
5	$\varphi \longrightarrow \Box\varphi$	⟶I, 1–4

Two proofs with incorrect applications of the rule □In.

Of course, Line 3 in both proofs is not licensed by the □I rule, as we use: no formula that appears before the start of the □-subproof can be cited unless it is marked by ⋆ or it is licensed by the rule □E.

The whole point for introducing a modal operator was to distinguish between "truth" or "being the case" and being modally true. Without the restriction on citing formulas in □-subproofs the enterprise would thus fail, as the distinction truth and modal truth would collapse.

Let's turn to the □-elimination rule □E and its restriction. One might ask two questions:

1. Why are we only allowed to eliminate the □ in □-subproofs and not anywhere in the proof?
2. Why are we only allowed to eliminate the □ in *immediate* □-subproofs, that is, why can't □E reach across multiple □-subproofs?

Ad 1: If we were allowed to eliminate □ at any point in a proof we could derive $\Box\varphi \longrightarrow \varphi$:

1	$\quad \Box\varphi$	
2	$\quad \varphi$	□E, 1
3	$\Box\varphi \longrightarrow \varphi$	⟶I, 1–2

Incorrect application of the rule □E.

While $\Box\varphi \longrightarrow \varphi$ seems plausible if we understand □ as *it is necessary that*, or *it is know that* we wish to construct a modal logic that can be tailored to fit with many different modalities, in particular, deontic modalities for which the □ is read as *must* or *should* as well as the doxastic modalities such as *agent a believes that.* For such a reading of the $\Box$-connective $\Box\varphi \longrightarrow \varphi$ seem very implausible and so does an unrestricted $\Box$-elimination rule.

Ad 2: This would again deliver too much, e.g., we could derive $\Box\varphi \longrightarrow \Box\Box\varphi$, which is not a principle we want to accept for all readings/understandings of $\Box$. Famously, Tim Williamson in *Knowledge and its Limits* argues this principle, known as the KK-principle when applied to knowledge, should not be assumed if $\Box$ is understood in term of *it is known that.*

1	$\quad \mid \Box\varphi$	
2	$\quad \mid \quad \mid \Box$	
3	$\quad \mid \quad \mid \quad \mid \Box$	
4	$\quad \mid \quad \mid \quad \mid \varphi$	$\Box$E, 1
5	$\quad \mid \quad \mid \Box\varphi$	$\Box$I, 3–4
6	$\quad \mid \Box\Box\varphi$	$\Box$I, 2–5
7	$\Box\varphi \longrightarrow \Box\Box\varphi$	$\longrightarrow$I, 1–6

Incorrect application of the rule $\Box$E.

Of course, the KK-principle was endorsed by Jaakko Hintikka in *Knowledge and Belief* but whether we should endorse that principle or not should not be decided by logic but on philosophical grounds. The "proof" also show why it is wrong to conclude Line 4 by $\Box$E and Line 1: each $\Box$-subproof can be understood as prefixing every formula in the subproof by a $\Box$. This means that in order use $\Box$E in the second subproof we should assume/know that $\Box\Box\varphi$ is true—but we don't have that information.

Rules for $\Diamond$

As for the defined logical connectives of classical logic we also have two rules for the defined modal operator $\Diamond$: $\Diamond$-**Introduction rule** $\Diamond$-**Elimination rule**

n	$\neg\Box\neg\varphi$	
m	$\Diamond\varphi$	Def. $\Diamond$, n

m	$\Diamond\varphi$	
n	$\neg\Box\neg\varphi$	Def. $\Diamond$, m

3.2.1 The minimal normal modal system K

Extending the natural deduction system for classical propositional logic by the rules $\Box$I and $\Box$E leads to the basic/minimal **normal modal logic** K.[2]

[2] A normal modal logic is a modal logic which extends the natural deduction system for classical logic by the rule $\Box$I and $\Box$E and, possibly, further rules.

Definition 15 (Modal System K)**.** *Let φ be a formula of $\mathcal{L}_\Box$ and $\mathcal{G}, \Gamma$ sets of formulas. Then we write $\mathcal{G}, \Gamma \vdash_\mathsf{K} \varphi$ iff φ can be derived from the set of global premises $\mathcal{G}$ together with the set of local premises Γ in the natural deduction calculus consisting of the all the basic rules for classical propositional logic together with the rule $\Box$I and $\Box$E.*

Shortly we will discuss modal systems extending K but for now we will collect some theorems of K.

Theorems of K

Let φ, ψ, and χ be formulas $\mathcal{L}_\Box$. Then

(a) $\vdash_\mathsf{K} \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$

1	$\mid\ \mid\ \underline{\Box(\varphi \to \psi)}$	
2	$\mid\ \mid\ \mid\ \underline{\Box\varphi}$	
3	$\mid\ \mid\ \mid\ \mid\ \underline{\Box}$	
4	$\mid\ \mid\ \mid\ \mid\ \varphi \to \psi$	$\Box$E, 1
5	$\mid\ \mid\ \mid\ \mid\ \varphi$	$\Box$E, 2
6	$\mid\ \mid\ \mid\ \mid\ \psi$	$\to$E, 4, 5
7	$\mid\ \mid\ \mid\ \Box\psi$	$\Box$I, 3–7
8	$\mid\ \mid\ \Box\varphi \to \Box\psi$	$\to$I, 2–7
9	$\mid\ \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$	$\to$I, 1–9

(b) $\vdash_K \Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$.

1	$\Box\varphi$	
2	$\quad\neg\Diamond\neg\varphi \to \bot$	
3	$\quad\neg\neg\Diamond\neg\varphi$	Def. $\neg$, 2
4	$\quad\Diamond\neg\varphi$	Exercise!, 3
5	$\quad\neg\Box\neg\neg\varphi$	Def. $\Diamond$, 4
6	$\quad\quad\Box$	
7	$\quad\quad\varphi$	$\Box$E, 1
8	$\quad\quad\neg\neg\varphi$	Exercise!, 7
9	$\quad\Box\neg\neg\varphi$	$\Box$I, 6–8
10	$\quad\bot$	Exercise!, 5, 9
11	$\neg\Diamond\neg\varphi$	PbC, 2–10
12	$\Box\varphi \to \neg\Diamond\neg\varphi$	$\to$I, 1–11
13	$\neg\Diamond\neg\varphi$	
14	$\quad\Box\varphi \to \bot$	
15	$\quad\quad\neg\Box\neg\neg\varphi \to \bot$	
16	$\quad\quad\neg\neg\Box\neg\neg\varphi$	Def. $\neg$, 15
17	$\quad\quad\Box\neg\neg\varphi$	Exercise!, 16
18	$\quad\quad\Box\varphi$	Exercise!, 17
19	$\quad\quad\bot$	$\to$E, 14, 18
20	$\quad\neg\Box\neg\neg\varphi$	PbC, 15–19
21	$\quad\Diamond\neg\varphi$	Def. $\Diamond$, 20
22	$\quad\bot$	Exercise, 13, 21
23	$\Box\varphi$	PbC, 14–22
24	$\neg\Diamond\neg\varphi \to \Box\varphi$	$\to$I, 13–23
25	$\Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$	Exercise, 12, 24

(c) $\vdash_K \neg\Box\varphi \leftrightarrow \Diamond\neg\varphi$

(d) $\vdash_K \Box\neg\varphi \leftrightarrow \neg\Diamond\varphi$

(e) $\vdash_{\mathsf{K}} \Box(\varphi \wedge \psi) \leftrightarrow \Box\varphi \wedge \Box\psi$

(f) $\vdash_{\mathsf{K}} \Box\varphi \vee \Box\psi \rightarrow \Box(\varphi \vee \psi)$

(g) $\vdash_{\mathsf{K}} \Diamond(\varphi \wedge \psi) \rightarrow \Diamond\varphi \wedge \Diamond\psi$

(h) $\vdash_{\mathsf{K}} \Diamond(\varphi \vee \psi) \leftrightarrow \Diamond\varphi \vee \Diamond\psi$

Exercise 16. *Prove the theorems above for which no proof has been given.*

Derived rules for K

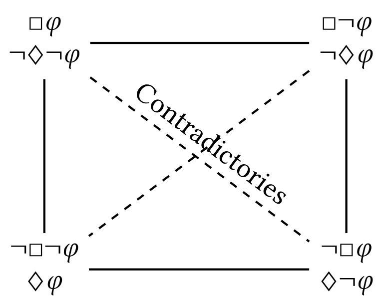
As should be evident from inspecting the proof of (b) above things can get complicated working only with the basic rules of K. To make our life easier we introduced a couple of derived rules. From now on we simply write

$$\frac{\varphi}{\psi}$$

instead of

$$\begin{array}{r|ll} \vdots & \vdots & \\ m & \varphi & \\ \vdots & \vdots & \\ n & \psi & \quad \text{Rule, } m \\ \vdots & \vdots & \end{array}$$

Derived Rules for $\Diamond$ The modal square of oppositions is a neat way of displaying the relation between $\Box$ and $\Diamond$:



All these equivalences can proved in the modal logic K. However, sometimes this is cumbersome and it will make our live easier to have them as derived rules:

$$\frac{\neg\Box\varphi}{\Diamond\neg\varphi}\ (\Diamond\neg 1) \qquad \frac{\Box\varphi}{\neg\Diamond\neg\varphi}\ (\neg\Diamond\neg 1) \qquad \frac{\Box\neg\varphi}{\neg\Diamond\varphi}\ (\neg\Diamond 1)$$

$$\frac{\Diamond\neg\varphi}{\neg\Box\varphi}\ (\Diamond\neg 2) \qquad \frac{\neg\Diamond\neg\varphi}{\Box\varphi}\ (\neg\Diamond\neg 2) \qquad \frac{\neg\Diamond\varphi}{\Box\neg\varphi}\ (\neg\Diamond 2)$$

Recall that we don't need a rule for moving from $\Diamond\varphi$ to $\neg\Box\neg\varphi$, as this follows by the Definition of $\Diamond$. We introduce a further rule that enables us to perform logical transformation within the scope of the $\Diamond$, in fact we give the rule in two different versions:

$$
\begin{array}{l|ll}
1 & \underline{\varphi \rightarrow \psi} & \star \\
\vdots & \vdots & \\
m & \Diamond\varphi & \\
\vdots & \vdots & \\
n & \Diamond\psi & \Diamond\text{-D}\star,\ 1,\ m
\end{array}
$$

$$
\begin{array}{l|l|l|ll}
k & \Diamond\varphi & & & \\
l & & \underline{\Box} & & \\
l+1 & & & \underline{\varphi} & \\
\vdots & & & \vdots & \\
m & & & \psi & \\
m+1 & \Diamond\psi & & & \Diamond\text{-D},\ k,\ l\text{–}m
\end{array}
$$

The subproof starting in Line l is not embedded in a $\Box$-subproof starting in between Lines l and k.

The rule $\Diamond$-D$\star$ makes us of the notion of a global premise it tells us that

$$\{\varphi \longrightarrow \psi\}, \{\Diamond\varphi\} \vdash_{\mathsf{K}} \Diamond\psi.$$

Here, the set $\{\varphi \longrightarrow \psi\}$ is the set of global premises, whilst the other set is the set of local premises. Notice that there are no application conditions for the rule $\Diamond$-D$\star$. At least in principles $\Diamond\varphi$ could be embedded in multiple $\Box$-subproofs and we could still use the rule ($\Diamond$-D$\star$ does not allow us to close a subproof however). $\Diamond$-D is a peculiar rule as it allows us to simultaneously close two subproofs.

We now show that $\Diamond$-D$\star$ is derivable rule, that is, we assume $\varphi \longrightarrow \psi$ is a global premise and assume that we obtain that we have obtained $\Diamond\varphi$ in Line m. We make now assumption on whether Line m is embedded in any $\Box$-subproofs or not.

1	$\varphi \to \psi$	$\star$
$\vdots$	$\vdots$	
m	$\Diamond \varphi$	
$m+1$	$\mid\ \Diamond \psi \to \bot$	
$m+2$	$\mid\ \Box\neg\psi$	$\neg\Diamond 2,\ m+1$
$m+3$	$\mid\ \mid\ \Box$	
$m+4$	$\mid\ \mid\ \varphi \to \psi$	R, 1
$m+5$	$\mid\ \mid\ \neg\psi$	$\Box$E, $m+2$
$m+6$	$\mid\ \mid\ \neg\varphi$	MT, $m+4$, $m+5$
$m+7$	$\mid\ \Box\neg\varphi$	$\Box$I, $m+3$–$m+6$
$m+8$	$\mid\ \Diamond \psi \to \bot$	$\neg\Diamond 1,\ m+7$
$m+9$	$\mid\ \bot$	$\to$E, m, $m+8$
$m+10$	$\Diamond \psi$	PbC, $m+1$–$m+9$

In this derivation MT is the derived rule Modus Tollens.

3.2.2 Normal Modal Logics beyond K

To capture important properties of alethic modalities such as logical and metaphysical necessity or doxastic modalities such as belief or knowledge we need to add further rules to the natural deduction system for K. We have already discussed that for alethic modalities the principles $\Box\varphi \to \varphi$ seems very plausible and the same hold if we understand $\Box$ as ‘it is known that’. To this effect we now extend the modal logic K by introducing further rules. We are led to number of different normal modal logics.

The normal modal logic KD

The characteristic principle of the modal logic KD is the principles

$$\Box\varphi \to \Diamond\varphi \tag{D}$$

This principle plays a crucial for deontic modalities (hence D). If you read the $\Box$ as *must do* and $\Diamond$ as *can do* (that is, *must not do not*), the D seems to convey the important fact that if you must do something, then you can do it. Differently if you obliged to do something, then it is (should be) permissible to do it.

As a rule of natural deduction the D-principle turns into:

$$\frac{\Box\varphi}{\Diamond\varphi}\ \text{(D-R)}$$

Importantly, one can prove in KD that theorems are possible (i.e., something is possible) and, equivalently, that $\bot$ is not necessary. Surprisingly, this is not possible in the modal logic K. K did still allow for a reading of $\Box$ as *it is false that* (it did also allow reading $\Box$ as *it is true that*.

In KD one can prove:

- $\vdash_{\mathsf{KD}} \Diamond\varphi$, if φ is a theorem of KD.
- $\vdash_{\mathsf{KD}} \neg\Box\bot$

We prove the latter claim the former is left as an exercise:

$$
\begin{array}{r|l|l}
1 & \quad \Box\bot & \\
\hline
2 & \quad \Box(\bot \to \bot) \to \bot & \text{D-R, 1} \\
3 & \qquad \Box & \\
\hline
4 & \qquad \bot \to \bot & \text{CL} \\
5 & \quad \Box(\bot \to \bot) & \Box\text{I, 3–4} \\
6 & \quad \bot & \to\text{E, 2, 5} \\
7 & \Box\bot \to \bot & \to\text{I, 1–6}
\end{array}
$$

Line 4 is a tautology of classical logic and the proof for deriving line 4 is left as an exercise.

The normal modal logic KT

The modal principles

$$(\mathsf{T}) \qquad \Box\varphi \to \varphi$$

is the characteristic principles of the modal logic KT. (T) asserts the factivity of the modal notion under consideration, that is, that it implies truth. It is thus constitutive for modal notion such as the different forms of alethic modalities ('necessity') and epistemic modalities ('it is known that'). Discuss: are there other factive notion, i.e., notions for which we should adopt the T principle.

As a rule of natural deduction the T-principle turns into:

$$\frac{\Box\varphi}{\varphi}\ (\text{T-R})$$

This rule yields the dual derived rule:

$$\frac{\varphi}{\Diamond\varphi}\ (\text{T-DR})$$

The modal logic KT thus extends the logic K by the rule T-R. The logic KD is a sublogic of KT, that is, the rule D-R is a derived rule of the logic KT. (Exercise!)

Some noteworthy theorems of KT are:

- $\vdash_{\mathsf{KT}} \varphi \longrightarrow \Diamond\varphi$
- $\vdash_{\mathsf{KT}} \Box\varphi \longrightarrow \Diamond\varphi$

The normal modal logic S4

The modal logic S4 is one candidate logic for knowledge. It is characterized by two constitutive principles: the modal principle T and the principle

$$\text{(4)} \qquad \Box\varphi \longrightarrow \Box\Box\varphi$$

This means that we obtain the modal logic S4 by adding the rule

$$\frac{\Box\varphi}{\Box\Box\varphi}\ \text{(4-R)}$$

to the modal logic KT. Again we obtain a derived dual rules:

$$\frac{\Diamond\Diamond\varphi}{\Diamond\varphi}\ \text{(4-DR)}$$

S4 is sometimes also called KT4, which would be the more systematic name.

Let's collect some theorems of S4:

- $\vdash_{\mathsf{S4}} \Box\varphi \leftrightarrow \Box\Box\varphi$
- $\vdash_{\mathsf{S4}} \Diamond\Diamond\varphi \leftrightarrow \Diamond\varphi$

The normal modal logic S5

Lastly, we introduce the modal logic S5 which extends S4 by the rule

$$\frac{\Diamond\varphi}{\Box\Diamond\varphi}\ \text{(E-R)}$$

Unsurprisingly, we get the derived, dual rule:

$$\frac{\Diamond\Box\varphi}{\Box\varphi}\ \text{(E-DR)}$$

In addition to the characteristic principles of S4 the modal logic S5 has thus the following characteristic principle:

$$\text{(E)} \qquad \Diamond\varphi \longrightarrow \Box\Diamond\varphi.$$

E tells us that if something is possible, then it is possible out of necessity. In S5 blocks of $\Box$'s and $\Diamond$'s can always be reduced to one of the modalities displayed in the modal square of opposition.

Let's collect some theorems:

- $\vdash_{\mathsf{S5}} \Diamond\varphi \leftrightarrow \Box\Diamond\varphi$
- $\vdash_{\mathsf{S5}} \Diamond\Box\varphi \leftrightarrow \Box\varphi$
- $\vdash_{\mathsf{S5}} \Diamond\Box\varphi \longrightarrow \Box\Diamond\varphi$
- $\vdash_{\mathsf{S5}} \varphi \longrightarrow \Box\Diamond\varphi$

Exercise 17. *Prove the above theorems of* S5*.*

Reduction of modalities In $\mathsf{S5}$ we have the noteworthy fact the chains of modal operators always reduce to one single operator. Let $\mathsf{O}_n \in \{\Box, \Diamond\}$ for all natural numbers j (that O_j is either the modal operator $\Box$ or the $\Diamond$-operator). Then for all formulas φ and all natural numbers n:

(i) $\vdash_{\mathsf{S5}} \mathsf{O}_1 \ldots \mathsf{O}_n \Box\varphi \leftrightarrow \Box\varphi$

(ii) $\vdash_{\mathsf{S5}} \mathsf{O}_1 \ldots \mathsf{O}_n \Diamond\varphi \leftrightarrow \Diamond\varphi$

Exercise 18. *Prove these two claim. You will have to do that using a proof by an induction on the number of operators* O. *If* $n = 1$ *there is only one operator* O. *There are two possible cases either* O *is* $\Box$ *or* O *is* $\Diamond$. *This means that for case (i) you need to establish:*

- $\vdash_{\mathsf{S5}} \Diamond\Box\varphi \leftrightarrow \Box\varphi$
- $\vdash_{\mathsf{S5}} \Box\Box\varphi \leftrightarrow \Box\varphi$

Once you have established that you have established the base case. You then assume you have established the claim for $n = m$ *and you wish to show that it also holds for* $n = m + 1$. *Applied to case (i) this means that you may assume:*

- $\vdash_{\mathsf{S5}} \mathsf{O}_1 \ldots \mathsf{O}_m \Box\varphi \leftrightarrow \Box\varphi$

and you need to show that the

- $\vdash_{\mathsf{S5}} \mathsf{O}_1 \ldots \mathsf{O}_m, \mathsf{O}_{m+1} \Box\varphi \leftrightarrow \Box\varphi$

Once you have established this claim you may infer that (i) holds for all n (and similarly for (ii)).

3.3 Rules of Natural Deduction vs Characteristic Principles

In the literature on modal logic the different systems are typically identified via their specific characteristic principles and not particular rules of natural deduction. However, it is not too difficult to derive the characteristic principle of a particular modal logic from the associated rule of natural deduction, or to show that given a particular characteristic principle the associated rule turns out to be a derivable rule. Let's look at the case of KT as an example:

- it is immediate to show that the characteristic principle (T) is provable in KT, that it, in the minimal modal logic K together with the rule T-R:

1	$\quad\Box\varphi$	
2	$\quad\varphi$	T-R, 1
3	$\Box\varphi \to \varphi$	$\to$I, 1–2

- if, in contrast, the principle (T) is a *global premise* (i.e., an axiom), then T-R becomes a derivable rule:

$$
\begin{array}{l|ll}
1 & \Box\varphi \longrightarrow \varphi & \star \\
\cline{2-2}
\vdots & \vdots & \\
k & \Box\varphi & \\
k+1 & \varphi & \longrightarrow\text{E}, 1, k
\end{array}
$$

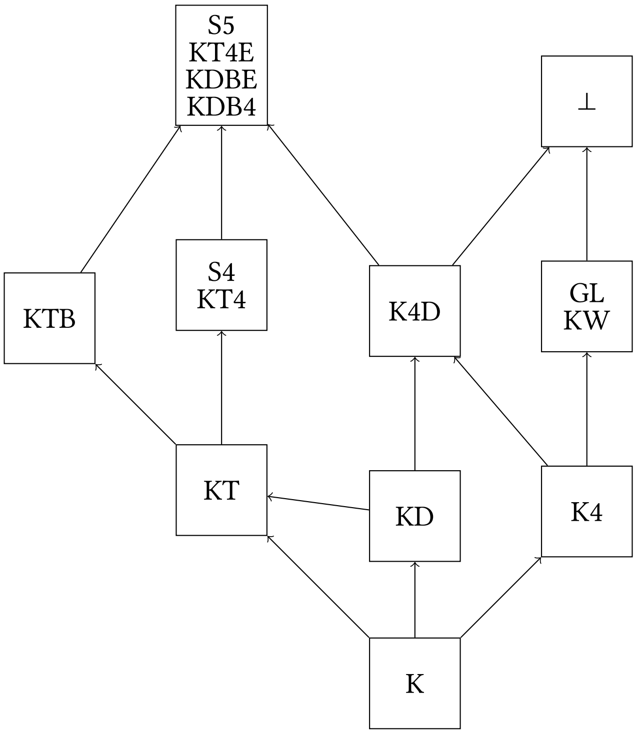
  Notice that only if (T) is a global premise can we infer Line $k + 1$ from Line k, if the latter is embedded in a $\Box$-subproof. However, T-R can be used within $\Box$-subproof (not across!!) and so we need to assumed that (T) is global premise.

Of course, the above is not particular to (T) and T-R but holds more generally: a formula $\varphi \longrightarrow \psi$ is a global premise/axiom iff $\dfrac{\varphi}{\psi}$ is a derivable rule.

Exercise 19. *Show that* $\dfrac{\varphi}{\psi}$ *is a derivable rule in a normal modal logic S extending* K *iff* $\varphi \vdash_S \psi$ *(φ is a local premise).*

The above exercise establishes a local version of the deduction theorem for $\vdash_{\mathsf{K}}$:

$$\mathcal{G}, \Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \psi \text{ iff } \mathcal{G}, \Gamma \vdash_{\mathsf{K}} \varphi \longrightarrow \psi.$$



Some modal systems and their relations—$\perp$ denotes an inconsistent system.

Modal rules

$\dfrac{\Box\varphi}{\Diamond\varphi}$ (D-R) $\qquad$ $\dfrac{\Box\varphi}{\varphi}$ (T-R) $\qquad$ $\dfrac{\Box\varphi}{\Box\Box\varphi}$ (4-R)

$\dfrac{\varphi}{\Box\Diamond\varphi}$ (B-R) $\qquad$ $\dfrac{\Diamond\varphi}{\Box\Diamond\varphi}$ (E-R) $\qquad$ $\dfrac{\Box(\Box\varphi \longrightarrow \varphi)}{\Box\varphi}$ (W-R)

Chapter 4

Possible World Semantics

In the previous section we introduced the new operator '$\Box$' and the defined, dual operator '$\Diamond$' together with an informal reading of the former as '*it is necessary that*' and the latter as '*it is possible that*'. The next step is to give a rigorous semantics for these modal operators. To this effect we cannot simply introduce a new truth table as the '$\Box$'-operator is not truth-functional: for any good semantics for modal logic it is **not** the case that for all propositional variables p and q (from now we will use capital V to denote a valuation, as we will us v as a variable for worlds)

$$\text{if } V(p) = V(q), \text{ then } \mathtt{I}_V(\Box p) = \mathtt{I}_V(\Box q).$$

If truth-tables are not the way forward, how should we interpret the modal operator?[1]

Taking a cue from Leibniz logicians introduced the notion of a *possible world* to their semantics and proposed understanding a formula $\Box\varphi$ as '*true in all (accessible) possible worlds*' and $\Diamond\varphi$ as '*true in some (accessible) possible world*'.[2] In the paraphrases we tacitly introduced the notion of an accessible world and thus the idea that not every possible world is accessible from a given other world, i.e., not every world can "see" every possible world. This means that if we wish to figure out whether $\Box\varphi$ is true at a given world, e.g., the actual world not *every* possible world will be relevant but only those that are accessible, i.e., those that can be seen by the given world. Different notions of "accessibility" is what will allow us to give sound and complete semantics for the different systems of modal logic we discussed in Section 3.2. Let's start making all this precise.

4.1 The basic Framework

4.1.1 Modal Frames

The first step is to make precise our talk of worlds and of worlds being accessible from other worlds. To this effect we introduce the notion of a **modal frame**.

[1]This question had logician genuinely worried until possible world semantics was developed.
[2]'Some' should be understood as 'there is at least one'.

Definition 20 (Modal Frame). *A tuple $F = \langle W, R \rangle$ is called a modal frame iff $W \neq \emptyset$ is a nonempty set of worlds and $R \subseteq W \times W$ is a binary relation between worlds. R is called the accessibility relation.*[3]

A modal frame F tells us what the possible worlds are (via the set W) and which worlds are accessible from a given world w, namely, those worlds v for which $\langle w, v \rangle \in R$ (we often write wRv): if $\langle w, v \rangle \in R$ then v is accessible from w. We will continue to use small letters u, v, w, z as world variables (if required we will also use indices, i.e., w_1, w_2, u_1 etc).

> **Example 21** (Examples of modal frame). *A frame $F = \langle W, R \rangle$ is given by specifying the set of worlds W and the accessibility relation R. We present five example below:*

(i) $W = \{w, v\}$ *and* $R = \{\langle w, v \rangle, \langle v, w \rangle\}$

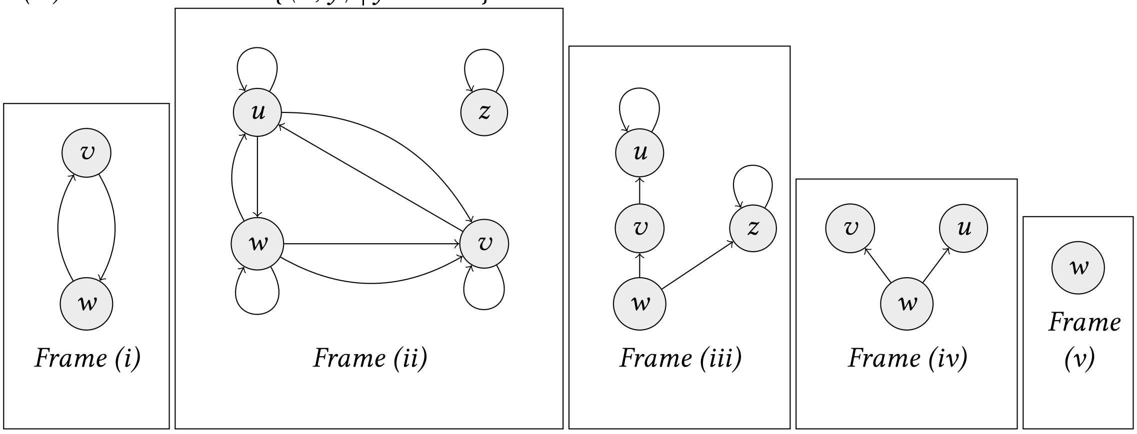
(ii) $W = \{w, v, u, z\}$ *and* $R = \{\langle w, v \rangle, \langle v, u \rangle, \langle w, u \rangle, \langle v, w \rangle, \langle u, v \rangle, \langle u, w \rangle, \langle w, w \rangle, \langle v, v \rangle, \langle u, u \rangle, \langle z, z \rangle\}$

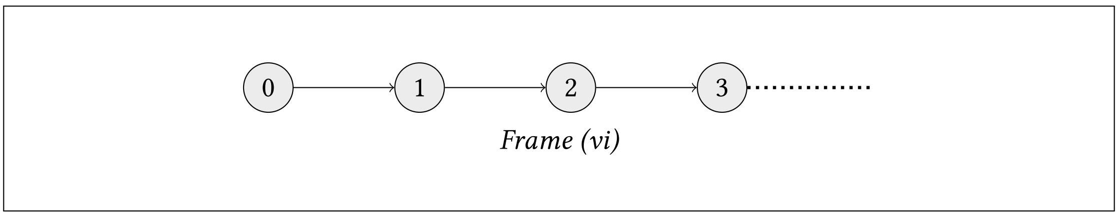
(iii) $W = \{w, v, u, z\}$ *and* $R = \{\langle w, v \rangle, \langle v, u \rangle, \langle u, u \rangle, \langle w, z \rangle, \langle z, z \rangle\}$

(iv) $W = \{w, v, u\}$ *and* $R = \{\langle w, v \rangle, \langle w, u \rangle\}$

(v) $W = \{w\}$ *and* $R = \emptyset$

(vi) $W = \mathbb{N}$ *and* $R = \{\langle x, y \rangle \mid y = x + 1\}$





[3]Recall that $W \times W = \{\langle w, v \rangle \mid w \in W \text{ and } v \in W\}$.

4.1.2 Valuations and Interpretations

Modal frame will be crucial in interpreting the modal operator, that is, in assigning a semantic value to formula of the form $\Box\varphi$ ($\Diamond\varphi$). But as in truth-functional logic before we can specify whether and under which conditions a formula of the form $\Box\varphi$ is true we need to give an account of the truth/falsehood of the propositional variables of the language. This job falls again to so-called *valuation functions* but in contrast to the case of truth-functional logic (clssical propositional logic) a valuation function is no longer a function from the set of propositional variables to the set of truth values, but assigns a truth value relative to a possible world to every propositional variable.

Definition 22 (Modal Valuation). *Let $F = \langle W, R\rangle$ be a* **modal frame**. *Then a modal valuation V on F is a function $V : \mathsf{At} \times W \longrightarrow \{\mathsf{T}, \mathsf{F}\}$.*

The definition guarantees that a valuation assigns a truth value to every propositional variable p and world w, that is, $V(p, w) \in \{\mathsf{T}, \mathsf{F}\}$. Starting from a valuation on a modal frame F we can, as in classical propositional logic, define an interpretation function that assigns truth values to all formulas of the modal language.

Definition 23 (Modal Interpretation/Model). *Let F be a modal frame and V a modal valuation on F. Then $\mathtt{I}_{(F,V)}$ is a function that assigns to every formula φ either the value T or the value F relative to a world w:*

(i) $\mathtt{I}_{(F,V)}(\varphi, w) = V(\varphi, w)$, *if φ is a propositional variable;*

(ii) $\mathtt{I}_{(F,V)}(\varphi, w) = \mathsf{F}$, *if φ is '$\bot$';*

(iii) if φ is a formula $\psi \wedge \chi$, then

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, w) = \mathsf{T} \textit{ and } \mathtt{I}_{(F,V)}(\chi, w) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

(iv) if φ is a formula $\psi \vee \chi$, then

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, w) = \mathsf{T} \textit{ or } \mathtt{I}_{(F,V)}(\chi, w) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

(v) if φ is a formula $\psi \longrightarrow \chi$, then

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, w) = \mathsf{F} \textit{ or } \mathtt{I}_{(F,V)}(\chi, w) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

(vi) if φ is a formula $\Box\psi$, then

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, v) = \mathsf{T} \textit{ for all } v \in W \textit{ s.t. } wRv; \\ \mathsf{F}, & \textit{otherwise.} \end{cases}$$

The pair (F, V) is sometimes called a **modal model** *(or possible world model). We write $((F, V)), w \Vdash \varphi$ iff $\mathtt{I}_{F,V}(\varphi, w) = \mathsf{T}$, and say that φ is true at w in the model (F, V). If for $(F, V), w \Vdash \varphi$ all $w \in W$, we say that φ is true in the model (F, V). If $(F, V), w \Vdash \varphi$ is true for all valuations V on F, i.e., for all models on the frame F, we write $F, w \Vdash \varphi$ and say that φ is true in the frame F at world w.*

Let's start unpacking the definition. First, Items (i-v) should not be surprising. They are simply the clause/truth-conditions we already know from truth-functional logic with the exception that we have introduced a world parameter to the valuation and the interpretation function. Otherwise nothing changes. (vi) tells us that a formula of the form $\Box\varphi$ is true in a model (F, V) relative to some world w iff φ is true at all worlds v which are accessible from w—which can be seen from w.

What about a formula of the form $\Diamond\varphi$? Under which condition will that formula be true in a model (M, V) at a world $w \in W$? Let's work out it's truth-condition by unpacking the definition of '$\Diamond$':

$$
\begin{aligned}
(F, V), w \Vdash \Diamond\varphi \text{ iff } & (F, V), w \Vdash \Box(\varphi \longrightarrow \bot) \longrightarrow \bot; \\
\text{iff } & \mathtt{I}_{(F,V)}(\Box(\varphi \longrightarrow \bot), w) = \mathsf{F} \text{ or } \mathtt{I}_{F,V}(\bot) = \mathsf{T}; \\
\text{iff } & \mathtt{I}_{(F,V)}(\Box(\varphi \longrightarrow \bot), w) = \mathsf{F}; \\
\text{iff } & \text{it is not the case that } \mathtt{I}_{(F,V)}(\varphi \longrightarrow \bot, v) = \mathsf{T} \text{ for all } v \in W \text{ s. t. } wRv; \\
\text{iff } & \mathtt{I}_{(F,V)}(\varphi \longrightarrow \bot, v) \neq \mathsf{T} \text{ for some } v \in W \text{ s. t. } wRv. \\
\text{iff } & \mathtt{I}_{(F,V)}(\varphi, v) = \mathsf{T} \text{ for some } v \in W \text{ s. t. } wRv.
\end{aligned}
$$

Example 24 (Truth in a Modal Model)**.** *Let F be Frame (iv) of Example 21 and let $V(p, w) = V(p, v) = V(p, u) = \mathsf{T}$. Then $(F, V), w \Vdash \Box p \longrightarrow p$ and indeed $(F, V) \Vdash \Box p \longrightarrow p$ (that is, we also have $(F, V), u \Vdash \Box p \longrightarrow p$ and $(F, V), v \Vdash \Box p \longrightarrow p$; why?). However, let V' be a valuation such that $V'(p, w) = \mathsf{F}$ and $V'(p, u) = V'(p, v) = \mathsf{T}$. Then $(F, V'), w \nVdash \Box p \longrightarrow p$ (i.e., $\Box p \longrightarrow p$ not true at w) and thus $(F, V') \nVdash \Box p \longrightarrow p$, but $(F, V), u \Vdash \Box p \longrightarrow p$ and $(F, V), v \Vdash \Box p \longrightarrow p$.*
If F is Frame (ii) of Example 21. Then $F \Vdash \Box p \longrightarrow p$, that is, $\Box p \longrightarrow p$ is true in F at every world w relative to every valuation V.

Exercise 25. *On which frames of Example 21 is the principle* D *true? Are there frames for which there are no models such that* D *is true at all worlds, i.e., are there worlds for which* D *has to be false at at least one world? Why or why not?*

4.2 Logical Consequence

The next step is to define an appropriate notion of logical consequence relative to the modal semantics, that is, to introduce the notion of logical consequence to possible world semantics. Recall that in truth-functional logic a sentence φ is a consequence of a set of sentences Γ iff for every valuation such that all members of Γ are true, φ is also true. In our modal semantics, however, there more parameter that need to be accounted for, that is, a formula is not only true relative to a valuation but relative to a valuation, a worlds, and a modal structure.

We postpone the discussion of *logical consequence* and first define when a formula follows from a set of formulas (premises) relative to a modal frame F. In Section we discussed two different types of premises: **global** and **local** premises. A global premise can be assumed to be true in every world of the modal frame while a local premise can only be assumed in the particular world we are currently in. Now

Definition 26 (φ follows from $(\mathcal{G},\Gamma)$ in F)**.** *Let F be a frame and the tuple $(\mathcal{G},\Gamma)$ consist of a set of global premises $\mathcal{G}$ and a set of local premises Γ. Then a formula φ follows from $(\mathcal{G},\Gamma)$ relative to F iff for all valuation V on F and all $w \in W$:*

$$\textbf{if } \underbrace{(F,V),v \Vdash \psi \textit{ for all } \psi \in \mathcal{G} \textit{ and } v \in W}_{\textit{all global premises are true in } (F,V)}, \textit{ and } \underbrace{(F,V),w \Vdash \gamma \textit{ for all } \gamma \in \Gamma}_{\textit{all local premises are true in } (F,V) \textit{ at } w}, \textbf{ then } (F,V),w \Vdash \varphi.$$

We write $(\mathcal{G},\Gamma) \vDash_F \varphi$ to denote that φ follows from $(\mathcal{G},\Gamma)$ relative to the frame F.

Example 27. *Consider the following claims:*

(a) $(\emptyset, \{\Box p\}) \vDash_F \Diamond p$

(b) $(\emptyset, \{\Box p\}) \vDash_F p$

(c) $(\emptyset, \{\Box p\}) \vDash_F \Box\Box p$

(d) $(\emptyset, \{p\}) \vDash_F \Box\Diamond p$

(e) $(\emptyset, \{\Diamond p\}) \vDash_F \Box\Diamond p$

(f) $(\{\Box p \to p\}, \emptyset) \vDash_F \Box p$

(g) $(\emptyset, \{\Box(p \to q)\}) \vDash \Box p \to \Box q$

Then Claim (a) only holds for Frames (i-iii) of Example 21, but not of Frames (iv) and (v). Claim (b) holds for Frame (ii) but not for all other frames of Example 21. Claim (c) holds for Frames (i), (ii), (iv), and (v), but not for Frame (iii). Claim (d) holds for Frames (i) and (ii), but not on Frames (iii-v). Claim (e) holds only for Frame (ii), but not for all other frames. Claim (f) is true for Frames (iv) and (v), but not for Frames (i-iii). Finally, claim (g) is true for all frames. (Exercise: Check this, and produce counterexample for the frames for which (a-f) do not hold. E.g. for (a) find a valuation on Frame (iv) such that there is a world at which $\Box p$ is true but $\Diamond p$ is false.)

4.2.1 Logical consequence for the modal logic K

Let's return to the notion of logical consequence and in particular logical consequence for the modal logic K. Recall that in truth-functional logic if a formula φ was a logical consequence of Γ then it was true whenever all members of Γ were true independently of any specific/particular assumptions. In contrast when we said that a formula followed from $(\mathcal{G},\Gamma)$ relative to F this still depended on the particular choice of F. We obtain a notion of logical consequence for our modal semantics by quantifying over all modal frames:

Definition 28 (Logical Consequence). *Let the tuple $(\mathcal{G}, \Gamma)$ consist of a set of global premises $\mathcal{G}$ and a set of local premises Γ and let φ be a formula. Then φ is a logical consequence of $(\mathcal{G}, \Gamma)$ iff*

$$(\mathcal{G}, \Gamma) \vDash_F \varphi$$

for all frames F. We write $(\mathcal{G}, \Gamma) \vDash_{\mathsf{K}} \varphi$ to denote that φ is a logical consequence of $(\mathcal{G}, \Gamma)$.

Exercise 29 (Validity). *A formula φ is said to be* K*-valid ($\vDash_{\mathsf{K}} \varphi$) iff $(\emptyset, \emptyset) \vDash_{\mathsf{K}} \varphi$. Show that*

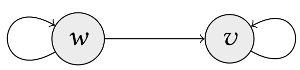
1. $\vDash_{\mathsf{K}} \Box\varphi \leftrightarrow \neg\Diamond\neg\varphi$
2. $\vDash_{\mathsf{K}} \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$
3. $\vDash_{\mathsf{K}} \neg\Box\varphi \leftrightarrow \Diamond\neg\varphi$
4. $\vDash_{\mathsf{K}} \Box\neg\varphi \leftrightarrow \neg\Diamond\varphi$
5. $\vDash_{\mathsf{K}} \Box(\varphi \wedge \psi) \leftrightarrow \Box\varphi \wedge \Box\psi$
6. $\vDash_{\mathsf{K}} \Box\varphi \vee \Box\psi \to \Box(\varphi \vee \psi)$
7. $\vDash_{\mathsf{K}} \Diamond(\varphi \wedge \psi) \to \Diamond\varphi \wedge \Diamond\psi$
8. $\vDash_{\mathsf{K}} \Diamond(\varphi \vee \psi) \leftrightarrow \Diamond\varphi \vee \Diamond\psi$

Example 30. *Consider the following claims:*

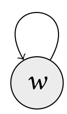
(i) $(\emptyset, \{\Box(\neg\varphi \to \psi)\}) \vDash_{\mathsf{K}} \Diamond\varphi \vee \Box\psi$

(ii) $(\emptyset, \{\Box\varphi\}) \vDash_{\mathsf{K}} \varphi$

Item (i) it true, i.e., $(\emptyset, \{\Box(\neg\varphi \to \psi)\}) \vDash_F \Diamond\varphi \vee \Box\psi$ for all frames (check it!). However (ii) is false, there are frame such that $(\emptyset, \{\Box\varphi\}) \nvDash_F \varphi$, e.g., Frames (i), (iii),(iv),(v) and (vi) of Example 21. $(\emptyset, \{\Box\varphi\}) \vDash_F \varphi$ is true, however, on Frame (ii) of Example 21 as well as the two frames below:



Reflexive frame 1



Reflexive frame 2

This suggest that while $(\emptyset, \{\Box\varphi\}) \vDash_F \varphi$ is not true relative to all frames F and thus φ is not a K*-logical consequences of $\Box\varphi$, it φ follows from $\Box\varphi$ on all reflexive frames, that is, frames in which every world sees itself.*

In the definition the subscript K to $\vDash_{\mathsf{K}}$ is suggestive and indeed the consequence relation $\vDash_{\mathsf{K}}$ coincides with the derivability relation $\vdash_{\mathsf{K}}$:

Proposition 31. *Let φ be a formula and $(\mathcal{G}, \Gamma)$ a pair of sets of sentences. Then*

$$(\mathcal{G}, \Gamma) \vdash_{\mathsf{K}} \varphi \textit{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{K}} \varphi.$$

This shows that we have developed an adequate semantics for the modal logic K. However, ultimately we also desire to have an adequate semantics for the other modal logics we have discussed. Fortunately, as we discuss in the next section, possible worlds semantics is equipped with the tools to accommodate the different modal logics.

4.3 Semantics for Modal Logics beyond K

The consequence relation $\vDash_{\mathsf{K}}$ is not an adequate consequence relation for the modal logics extending K. In particular, while the direction

$$\text{if } (\mathcal{G}, \Gamma) \vDash_{\mathsf{K}} \varphi, \text{ then } (\mathcal{G}, \Gamma) \vdash_S \varphi$$

holds for modal logics S extending K, **the converse does not hold**: there are frames F and valuations V such that we can find world w such that $(F, V), v \Vdash \psi$ for all $\psi \in \mathcal{G}$ and $v \in W$, and $(F, V), w \Vdash \gamma$ for all $\gamma \in \Gamma$, but $(F, V), w \Vdash \neg\psi$, **and** $(\mathcal{G}, \Gamma) \vdash_S \varphi$. The consequence relation $\vDash_{\mathsf{K}}$ is too permissive for modal logics extending K.

This suggests that logical consequence for modal logics extending K should not be defined relative to all frames but only relative to frames that meet a specific property. Example 30 provides us with a hint how this might go: we saw that the principle T was true on a frame in which the (every) world sees itself, that is, a reflexive frame. Indeed, T is true on every reflexive frames, i.e., it's true independently to the particular valuation V that is chosen. One can even show that if F is a frame, then

$$F \Vdash \Box\varphi \longrightarrow \varphi \text{ iff } F \text{ is reflexive}$$

$\Leftarrow$ Assume F is reflexive and assume that there is a valuation V and $w \in W$ such that $(F, V), w \Vdash \Box\varphi$. The latter can be unraveled to: for all $v \in W$: if wRv, then $(F, V), v \Vdash \varphi$. Thus in particular:

$$\text{if } wRw, \text{ then } (F, V), w \Vdash \varphi$$

By reflexivity we have wRw (for all $w \in W$) and thus by modus ponens $(F, V), w \Vdash \varphi$. This means that for all V on F and all $w \in W$:

$$(F, V), w \Vdash \Box\varphi \longrightarrow \varphi.$$

$\Rightarrow$ Assume $F \Vdash \Box\varphi \longrightarrow \varphi$ and, for reductio, we also assume that F is not reflexive. We now construct a valuation V for which there is a world w such that $(F, V), w \nVdash \Box p \longrightarrow p$. Let w be a non-reflexive world (i.e., not wRw) and then let V be a valuation such that $V(p, v) = \mathsf{T}$ for all $v \in W$ such that wRv, but $V(p, v) = \mathsf{F}$ otherwise. In particular $V(p, w) = \mathsf{F}$. Then $(F, V), w \Vdash \Box p$ but $(F, V), w \nVdash p$, which means that

$$F \nVdash \Box p \longrightarrow p.$$

We have arrived at a contradiction and may thus conclude that F is reflexive.

From this observation we can conclude that the logical consequence relation for KT needs to be defined relative to reflexive frames only and not arbitrary frames.

Definition 32 (KT-logical consequence)**.** *Let the tuple $(\mathcal{G}, \Gamma)$ consist of a set of global premises $\mathcal{G}$ and a set of local premises Γ and let φ be a formula. Then φ is a* KT*-logical consequence of $(\mathcal{G}, \Gamma)$ iff*

$$(\mathcal{G}, \Gamma) \vDash_F \varphi$$

for all **reflexive** *frames F. We write $(\mathcal{G}, \Gamma) \vDash_{\mathsf{KT}} \varphi$ to denote that φ is a* KT*-logical consequence of $(\mathcal{G}, \Gamma)$.*

With this definition in place we obtain the following adequacy result. Let $\mathcal{G}$ be a set of global premises, Γ a set of local premises and φ a formula. Then

$$(\mathcal{G}, \Gamma) \vdash_{\mathsf{KT}} \varphi \text{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{KT}} \varphi.$$

It turns out that the characteristic principles of different modal logics correspond to specific properties of the accessibility relation:

Table 4.1: **Characteristic modal principles and properties of the accessibility relation**

D	$\Box\varphi \longrightarrow \Diamond\varphi$	for all $w \in W$ there exists a $v \in W$ s.t. wRv	serial
T	$\Box\varphi \longrightarrow \varphi$	for all $w \in W$: wRw	reflexive
4	$\Box\varphi \longrightarrow \Box\Box\varphi$	for all $w, v, u \in W$: if wRu and uRv, then wRv	transitive
B	$\varphi \longrightarrow \Box\Diamond\varphi$	for all $w, v \in W$: if wRv, then vRw	symmetry
E	$\Diamond\varphi \longrightarrow \Box\Diamond\varphi$	for all $w, v, u \in W$: if wRu and wRv, then uRv	euclidian

Example 33 (4 and transitivity)**.** *Let F be a frame. Then*

$$F \vDash \Box\varphi \longrightarrow \Box\Box\varphi \textit{ iff } F \textit{ is transitive}$$

To show the claim. We first prove the left-to-right direction and then the converse direction.

To prove the former direction we assume that $F \vDash \Box p \longrightarrow \Box\Box p$ and assume for reductio (proof by contradiction) that F is not transitive, i.e., that there are worlds $w, v, u \in W$ such that wRv and vRu but not wRu. We then construct a valuation V such that $(F, V), w \Vdash \Box p$ but $(F, V), w \nVdash \Box\Box p$. This contradicts our initial assumption and allows us conclude that F is transitive. We define the valuation as follows:

$$V(p, z) := \begin{cases} \mathsf{T}, & \text{if } wRz; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

By definition we have $(F, V), w \Vdash \Box p$. However, since by definition $V(p, u) = \mathsf{F}$, we have $(F, V), v \nVdash \Box p$ and thus $(F, V), w \nVdash \Box\Box p$.

For the converse direction we assume that F is transitive and show $F \vDash \Box\varphi \longrightarrow \Box\Box\varphi$. The latter is equivalent to the following claim:

$$\textit{for all } V \textit{ for all } w \in W\textit{: if } (F, V), w \Vdash \Box\varphi\textit{, then } \underbrace{(F, V), w \Vdash \Box\Box\varphi}_{\dagger}$$

Now by definition † is equivalent to the following claim: for all $u \in W$: if wRu, then for all $v \in W$: if uRv, then $(F, V), v \Vdash \varphi$. With this in mind we prove our claim using the structure of an (informal) natural deduction proof (that is not how you have to do it, but hopefully makes the underlying reasoning clearer.

1	*F is transitive*	
2	$(F, V), w \Vdash \Box\varphi$	
3	wRu	
4	uRv	
5	$wRu \,\&\, uRv$	*Conjunction introduction,* 3, 4
6	wRv	*quantifier and conditional elimination,* 1, 5
7	*if* wRv *then* $(F, V), v \Vdash \varphi$	*Def. and quantifier elimination,* 2
8	$(F, V), v \Vdash \varphi$	*conditional elimination,* 6, 7
9	*if* uRv*, then* $(F, V), v \Vdash \varphi$	*conditional introduction,* 4–8
10	$\underbrace{\textit{for all } v \in W\textit{: if } uRv\textit{, then } (F, V), v \vDash \varphi}_{(F,V),u \Vdash \Box\varphi}$	*Univ. Q. Intro.,* 9
11	*if* wRu*, then* $(F, V), u \Vdash \Box\varphi$	*Cond. Intro,* 3–10
12	*for all* $u \in W$*: if* wRu*, then* $(F, V), u \Vdash \Box\varphi$	*Univ. Q. Intro.,* 11
13	$(F, V), w \Vdash \Box\Box\varphi$	*Def.,* 12
14	$(F, V), w \Vdash \Box\varphi \longrightarrow \Box\Box\varphi$	*Cond. Intro, Def.,* 2–13
15	$F \vDash \Box\varphi \longrightarrow \Box\Box\varphi$	*Univ. Q. Intro., Def.,* 14

Exercise 34. *Prove the remaining relations between the characteristic modal principles and particular properties suggested by Table 4.1. To show, e.g., that* D *is true on serial frames we pick an arbitrary serial frame F and choose an arbitrary valuation V and world $w \in W$ such that $(F, V), w \Vdash \Box\varphi$. Using the seriality of F we then need show that $(F, V), w \Vdash \Diamond\varphi$ (cf. $\Leftarrow$ above). For the converse direction, we assume that F is not serial, i.e., that there is a world w such that there is no $v \in W$ such that wRv. We then specify the minimal conditions for a valuation V such that $(F, V), w \Vdash \Box p$ and then attempt to show $(F, V), w \nVdash \Diamond p$ (cf. $\Rightarrow$ above). To this effect V can be extended as appropriate as long as the minimal conditions are still met (not necessary in the case of* D*).*

We now define the consequence relations for the modal logics $\mathsf{KD}, \mathsf{S4}$, and $\mathsf{S5}$. Let $\mathcal{G}$ be a set of global premises, Γ a set of local premises and φ a formula. Then

$$(\mathcal{G}, \Gamma) \vDash_{\mathsf{KD}} \varphi \text{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{F}} \varphi \text{ for all serial frames } F.$$
$$(\mathcal{G}, \Gamma) \vDash_{\mathsf{S4}} \varphi \text{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{F}} \varphi \text{ for all reflexive and transitive frames } F.$$
$$(\mathcal{G}, \Gamma) \vDash_{\mathsf{S5}} \varphi \text{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{F}} \varphi \text{ for all reflexive, transitive, and euclidian frames } F.$$

Before moving on to giving adequacy results for the logics KD, $\mathsf{S4}$, and $\mathsf{S5}$ let us quickly look at the properties of the accessibility relation associated with $\mathsf{S5}$. In the problem sheet on natural deduction for modal logic we saw that $\mathsf{S5} = \mathsf{KT4E} = \mathsf{KT4B}$. This also shows in the following fact:

Fact 35. *Let $R \subseteq W \times W$. Then*

R is reflexive, transitive, and euclidean iff R is reflexive, symmetric, and transitive.

A reflexive, symmetric and transitive relation is also called an *equivalence* relation (the name should make sense if you read 'R' as 'is equivalent to' and check the properties of R against that reading). That is, the $\mathsf{S5}$-consequence relation can also be characterise as follows:

$$(\mathcal{G}, \Gamma) \vDash_{\mathsf{S5}} \varphi \text{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{F}} \varphi \text{ for all equivalence frames } F.$$

We now collect the adequacy results of the different modal systems:

Proposition 36 (Adequacy of KD, $\mathsf{S4}$, and $\mathsf{S5}$)**.** *Let $\mathcal{G}$ be a set of global premises, Γ a set of local premises and φ a formula. Then*

$$(\mathcal{G}, \Gamma) \vdash_{\mathsf{KD}} \varphi \textit{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{KD}} \varphi;$$
$$(\mathcal{G}, \Gamma) \vdash_{\mathsf{KT}} \varphi \textit{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{KT}} \varphi;$$
$$(\mathcal{G}, \Gamma) \vdash_{\mathsf{S4}} \varphi \textit{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{S4}} \varphi;$$
$$(\mathcal{G}, \Gamma) \vdash_{\mathsf{S5}} \varphi \textit{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{S5}} \varphi.$$

4.3.1 S5 and Universal Frames

In some philosophico-logico works one can find the truth-conditions for the modal operator '$\Box$' stated without appealing to the accessibility relation. There we are presented with a non-empty set of worlds W, a valuation V and the following truth-conditions for a formula of the form $\Box\varphi$ at a world $w \in W$:

$$\mathtt{I}_{(W,V)}(\Box\varphi, w) := \begin{cases} \mathsf{T}, & \text{if } \mathtt{I}_{(W,V)}(\varphi, v) = \mathsf{T}, \text{ for all } v \in W; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

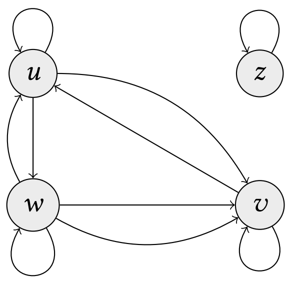
When theorists give this semantics they are typically working with the modal logic $\mathsf{S5}$ and are assuming the following definition of the $\mathsf{S5}$-consequence relation.

$$(\mathcal{G}, \Gamma) \vDash_{\mathsf{S5}} \varphi \text{ iff } (\mathcal{G}, \Gamma) \vDash_{F} \varphi \text{ for all universal frames } F$$

where $\mathcal{G}$ be a set of global premises, Γ a set of local premises and φ a formula.

What is a universal frame and how does that fit with our previous definition of the S5-consequence relation? A frame $F = (W, R)$ is universal iff wRv for all $w, v \in W$, i.e., iff every world in W is accessible from every other world in W. In universal frame the accessibility relation is redundant, as saying that a formula is true in all accessible worlds in W is just saying that the formula is true in all worlds in W. This answers the question why we can omit the accessibility, if we are working with universal frames. It does not however explain, why it is acceptable to define the S5-consequence relation by appeal to universal frames only.

Let's recall what an equivalence frames look like:

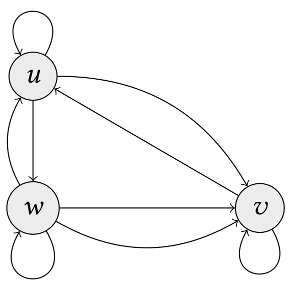


Example of an equivalence frame

As we can see by looking at the equivalence frame above the worlds w, v and u all see each other (and itself), but they don't see z, which sees itself. If, in the above frame, we were to focus only on the worlds w, v and u we could also drop the accessibility relation: let $W' = \{w, v, u\} \subseteq W$. Then for any $w' \in W'$ and valuation V on F:

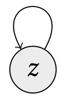
$$\begin{aligned}(F, V), w' \Vdash \Box\varphi &\text{ iff } (F, V), v \Vdash \Box\varphi \text{ if } w'Rv \text{ for all } v \in W\\ &\text{ iff } (F, V), v \Vdash \varphi \text{ if } w'Rv \text{ for all } v \in W'\\ &\text{ iff } (F, V), v' \Vdash \varphi \text{ for all } v \in W'\end{aligned}$$

This shows that the following subframe F' of F is a universal frame:



The universal frame F'

Similarly the subframe F'' of F with $W'' = \{z\} \subseteq W$ is also a universal frame:



The universal frame F''

This illustrates that if an equivalence frame has worlds that are not accessible to each other it can be seen to consist of multiple universal frames: Frame F above consists of the subframes F' and F''. However, if all worlds of an equivalence frame are connected then the equivalence frame is a universal frame. By the same token we know that every universal frame is an equivalence frame but, as we have seen, this does not hold the other way around (F is an equivalence frame but not a universal frame).

However, one can prove the following proposition which will allow us to show/argue that the two different definitions of the S5 consequence relation coincide:

Proposition 37. *Let F be an equivalence frame, V a valuation, and $w \in W$. Then there exists a universal frame $F' = (W', R')$ with $w \in W' \subseteq W$ and for all $u, v \in W$: if $uR'v$, then uRv, and there exists a valuation V' such that $V(p, u) = V'(p, u)$ for all $u \in W'$ and $p \in \mathsf{At}$ such that for all formulas φ:*

$$(F, V), w \Vdash \varphi \textit{ iff } (F', V'), w \Vdash \varphi.$$

As mentioned, this proposition can be used to show the following proposition:

Proposition 38.

$$(\mathcal{G}, \Gamma) \vDash_F \varphi \textit{ for all equivalence frames } F \textit{ iff } (\mathcal{G}, \Gamma) \vDash_F \varphi \textit{ for all universal frames } F,$$

Proof. Since every universal frame is an equivalence frame, we only need to prove the right to left direction. Now, suppose there are sets $\mathcal{G} = \{\psi_1, \dots, \psi_n\}$ and $\Gamma = \{\gamma_1, \dots, \gamma_m\}$ and a formula φ such that $(\mathcal{G}, \Gamma) \vDash_F \varphi$ for all universal frames F but that there exists an equivalence frame F' and valuation V with world w such that $F', V \Vdash \psi_j$ for $1 \leq j \leq n$ and $(F', V), w \Vdash \gamma_j$ for j with $1 \leq j \leq m$ but $(F', V), w \nVdash \varphi$. Then by Proposition 37 there must be a universal frame F'' and valuation V' such that $F'', V' \Vdash \psi_j$ for $1 \leq j \leq n$ and $(F'', V'), w \Vdash \gamma_j$ for j with $1 \leq j \leq m$ but $(F'', V'), w \nVdash \varphi$, which contradicts our assumption. □

This tells us that the two different way of defining the consequence relation of S5 do in fact coincide. So if we are only interested in the semantics for the modal logic S5, then it is acceptable to drop the accessibility relation from the truth-conditions for formulas of the form $\Box\varphi$ (and thus also for formulas of the form $\Diamond\varphi$).

Chapter 5

The Deduction Theorem for Modal Logic

Classical propositional logic enjoys the deduction theorem, i.e., in classical logic a formula ψ is a logical consequence of a formula φ iff $\varphi \longrightarrow \psi$ is a logical truth. In modal logic the situation is more problematic, as we now have two different types of premise sets: global premises and local premises. Accordingly, the deduction theorem for modal logic consists of two distinct deduction theorems: the *local* and the *global* deduction theorem.

The local deduction theorem is basically the deduction theorem we know from classical propositional logic.

Proposition 39 (Local Deduction Theorem)**.** *Let $\mathcal{G}$ be a set of global premises, Γ a set of local premises and φ, ψ formulas. Let L be one of the previous systems of modal logics discussed. Then*

($\vdash$) $$(\mathcal{G}, \Gamma \cup \{\varphi\}) \vdash_{\mathsf{L}} \psi \textit{ iff } (\mathcal{G}, \Gamma) \vdash_{\mathsf{L}} \varphi \longrightarrow \psi;$$

($\vDash$) $$(\mathcal{G}, \Gamma \cup \{\varphi\}) \vDash_{\mathsf{L}} \psi \textit{ iff } (\mathcal{G}, \Gamma) \vDash_{\mathsf{L}} \varphi \longrightarrow \psi;$$

The local deduction theorem for modal logic tells us that a formula ψ is derivable from/is a logical consequence of the premises $(\mathcal{G}, \Gamma)$ together with the additional local premise φ iff the conditional $\varphi \longrightarrow \psi$ is derivable from/is a logical consequence of $(\mathcal{G}, \Gamma)$.

The global deduction theorem for modal logic is of very different character, as it connects global premises and local premises: it is a statement about how to express global premises in terms of local premises and vice versa.

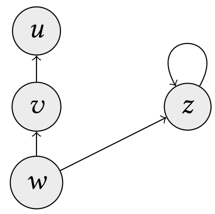
Proposition 40 (Global Deduction Theorem)**.** *Let $\mathcal{G}$ be a set of global premises, Γ a set of local premises and φ, ψ formulas. Let L be one of the previous systems of modal logics discussed. Then*

($\vdash$) $$(\mathcal{G} \cup \{\varphi\}, \Gamma) \vdash_{\mathsf{L}} \psi \textit{ iff } (\mathcal{G}, \Gamma \cup \{\Box^n \varphi \,|\, 0 \leq n\}) \vdash_{\mathsf{L}} \psi;$$

($\vDash$) $$(\mathcal{G} \cup \{\varphi\}, \Gamma) \vDash_{\mathsf{L}} \psi \textit{ iff } (\mathcal{G}, \Gamma \cup \{\Box^n \varphi \,|\, 0 \leq n\}) \vDash_{\mathsf{L}} \psi;$$

where $\Box^0 \varphi := \varphi, \Box^1 \varphi := \varphi, \dots, \Box^{n+1} \varphi := \Box\Box^n \varphi, \dots$, that is, for every natural numbers n the formula $\Box^n \psi \in \{\Box^n \varphi \,|\, 0 \leq n\}$.

How are we to make sense of the global deduction theorem? Recall that relative to a given model (F, V) a global premise φ is assumed to be true at all worlds, this tells us that φ is a local premise, i.e., it can be assumed to be true locally, that is, at the world under consideration. But if φ is true at every world, then it is true in all worlds accessible from a given world and thus $\Box\varphi$ must be true at every possible world and thus can be assumed to be true locally. We can repeat the reasoning to conclude that $\Box\Box\varphi$ is true at every world and thus a local premise. Now if at a world w we assume all members of the set $\Gamma \cup \{\Box^n\varphi \mid 0 \leq n\}$ are true. This means that φ will be true at any world that may become relevant for evaluating whether a formula is true at w or not. In other words, in this case φ is true in every world v such that there is a **path** from w to v. There is a path from a world w_1 to a world w_2, if starting from w_1 by following the accessibility relation we eventually arrive in w_2.



For example, in the frame above there is a path from w to all other worlds, save itself, despite the fact that u is not accessible from w (only v and z are). Yet, u is accessible from w in two consecutive steps and whether a formula is true or false at u will have impact on truth and falsity of formulas in w, e.g., if 'p' is true at u we know that $\Diamond\Box p$ is true at w. Yet, whether or not 'p' is true at u has no impact on the formulas that are true or false at the world z as there is no path between z and u: z and u are not connected. Now, if all members of the set $\{\Box^n\varphi \mid 0 \leq n\}$ are true at a world w, this is tantamount to saying that φ is true at all worlds that relevant for evaluating whether some formula ψ is true or false, that is, it amounts to assuming ψ as a global premise.

Chapter 6

Possible World Semantics and the Actual World★

In possible worlds semantics the key notion is that of a formula being true at a world. However, we might be particularly interested what is true in our world, that is, what is true in the actual world. In the semantics we introduced in the previous chapter we have no way of distinguishing between the actual world and any other world of the set of worlds W. This motivates adding a distinguished element w^*, the actual world, to possible world frames. On this view a modal frame is a tuple $F = \langle W, R, w^* \rangle$ with $w^* \in W$. The addition of the actual world does not affect the definition of a modal interpretation, that is, under which conditions a formula is true at a word in a model (F, V)—we write (F, w^*, V) to denote the model and to highlight that the frame has a distinguished element for the actual world. With the actual world in the picture we may reevaluate how to define the notion of logical consequence.

To make our life easier we focus on when a formula φ is a logical consequence of a set of local premises Γ, i.e., we set aside the role of global premises. So far, relative to a logic L we defined $\Gamma \vDash_{\mathsf{L}} \varphi$ as follows: $\Gamma \vDash_{\mathsf{L}} \varphi$ iff for all L-frames F, all valuations V on F and all worlds $w \in W$:

$$\text{if } (F, w^*, V), w \Vdash \gamma \text{ for all } \gamma \in \Gamma, \text{ then } (F, w^*, V), w \Vdash \varphi.$$

But now that we can single out the actual world amongst all the possible world, wouldn't it make sense to define logical consequence relative to the actual world, i.e., shouldn't we define logical consequence as *real world consequence*?

Definition 41 (Real world consequence)**.** *Let Γ be a set of local premises and φ a formula. Then $\Gamma \vDash_{@} \varphi$ iff for all universal frames F and all valuations V on F*

$$\textit{if } (F, w^*, V), w^* \Vdash \gamma \textit{ for all } \gamma \in \Gamma, \textit{ then } (F, w^*, V), w^* \Vdash \varphi.$$

In the definition above we focused on universal frames, but of course the definition can be generalized to other frames. The appeal of real world consequence is that it arguably is a better approximation of the notion of consequence we employ in ordinary reasoning. The downside however is that the neat systems of modal logic we discussed are not sound with respect to real world consequence. As we have seen, these systems of modal logic are sound and complete, that is, adequate with respect to the general notion of logical consequence as defined in the previous sections.

A way to recover the notion of real world consequence (and validity) whilst working with the more general notion of logical consequence is to introduce a further operator to our language, an actuality operator '@'. To this effect we expand the definition of a well-formed formula by the following clause:

- if φ is a formula, the $@\varphi$ is a formula.

$@\varphi$ may be read as '*it is actually the case that* φ' or '*actually,* φ'. Semantically the formula will be interpreted as 'φ *is true in the actual world*': let F be a universal frame, V a valuation and $w \in W$. Then

$$\mathtt{I}_{(F,w^*,V)}(@\varphi, w) := \begin{cases} \mathsf{T}, & \text{if } (\mathtt{I}_{(F,w^*,V)}(\varphi, w^*) = \mathsf{T} \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

A formula $@\varphi$ is true at a world w iff φ is true at the actual world w^* and false otherwise.

With the @-operator in our language we can move back and forth between real-world consequence and the general notion of logical consequence.

Proposition 42. *Let* Γ *be a set of premises and* φ *a formula. Then*

$$\Gamma \vDash_{S5} \varphi \textit{ iff } \vDash_{@} \Box(\bigwedge \Gamma \rightarrow \varphi);$$
$$\Gamma \vDash_{@} \varphi \textit{ iff } \vDash_{S5} @(\bigwedge \Gamma \rightarrow \varphi).$$

$\bigwedge \Gamma$ *is short for the formula* $\gamma_1 \wedge \ldots \wedge \gamma_n$ *for* $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$.[1]

To complete the picture we expand the modal system S5 by natural deduction rules for @. First, we introduce the notion of an @-subproof:

$$\begin{array}{l|l} \vdots & \vdots \\ m & \quad \begin{array}{|l} @ \\ \hline \vdots \end{array} \\ \vdots & \\ \vdots & \vdots \end{array}$$

An @-subproof works very much like a $\Box$-subproof in that we may only use/cite lines outside the @-subproof, if they are marked by $\star$, or if it is explicitly permitted by a rule:

[1] Without loss of generality we may always assume Γ to be a finite set.

@-Introduction rule

$$
\begin{array}{l|l|l}
m & \quad\begin{array}{|l} @ \\ \hline \vdots \\ \varphi \end{array} & \\
\vdots & & \\
n & & \\
n+1 & @\varphi & @\mathrm{I},\ m\text{–}n
\end{array}
$$

No line above Line m may be used in the @-subproof unless they are marked by ⋆, are within an open @-subproof or a rule explicitly permits it.

@-Elimination rule

$$
\begin{array}{l|l|l}
m & @\varphi & \\
n & \quad\begin{array}{|l} @ \\ \hline \vdots \\ \varphi \\ \vdots \end{array} & \\
\vdots & & \\
k & & @\mathrm{E},\ m \\
\vdots & & \\
\vdots & \vdots &
\end{array}
$$

Line n may be emedded in a □-subproof starting between Lines m and n.

Notice that there is no condition on the @-elimination rule, that is, we can introduce φ even to @-subproofs that are further embedded. Besides @introduction/elimination we need two further rules to complete our S5-natural deduction system with @:

$$
\frac{@\varphi}{\neg @\neg\varphi}\ (@1) \qquad\qquad \frac{\Box\varphi}{@\varphi}\ (@2)
$$

We list a couple of theorems of the logic:

(i) $\vdash_{\mathsf{S5@}} @(\varphi \rightarrow \psi) \leftrightarrow (@\varphi \rightarrow @\psi)$

(ii) $\vdash_{\mathsf{S5@}} @(\varphi \vee \psi) \leftrightarrow (@\varphi \vee @\psi)$

(iii) $\vdash_{\mathsf{S5@}} @\neg\varphi \leftrightarrow \neg @\varphi$

(iv) $\vdash_{\mathsf{S5@}} @\varphi \leftrightarrow \Box @\varphi$

(v) $\vdash_{S5@} @\varphi \leftrightarrow @@\varphi$

1	$\mid \; \underline{@\varphi}$	
2	$\mid \; \mid \; \underline{@}$	
3	$\mid \; \mid \; \mid \; \underline{@}$	
4	$\mid \; \mid \; \mid \; \varphi$	@E, 1
5	$\mid \; \mid \; @\varphi$	@I, 3–4
6	$\mid \; @@\varphi$	@I, 2–5
7	$@\varphi \rightarrow @@\varphi$	→I, 1–6
8	$\mid \; \underline{@@\varphi}$	
9	$\mid \; \mid \; \underline{@\varphi \rightarrow \bot}$	
10	$\mid \; \mid \; \neg @\varphi$	Def, 9
11	$\mid \; \mid \; @\neg\varphi$	→E, (iii) above, 10
12	$\mid \; \mid \; \mid \; \underline{@}$	
13	$\mid \; \mid \; \mid \; \mid \; \underline{@}$	
14	$\mid \; \mid \; \mid \; \mid \; \neg\varphi$	@E, 11
15	$\mid \; \mid \; \mid \; @\neg\varphi$	@I, 13–14
16	$\mid \; \mid \; \mid \; \neg @\varphi$	→E, (iii) above, 15
17	$\mid \; \mid \; @\neg @\varphi$	@I, 12–16
18	$\mid \; \mid \; \neg @@\varphi$	→E, (iii) above, 17
19	$\mid \; \mid \; \bot$	→E, 8, 18
20	$\mid \; @\varphi$	PbC, 9–19
21	$@@\varphi \rightarrow @\varphi$	→I, 8–21
22	$@@\varphi \leftrightarrow @\varphi$	∧I, 7, 21

(vi) $\vdash_{S5@} @(\varphi \leftrightarrow @\varphi)$

Of all these theorems (iv) is arguably the most surprising and, perhaps, also somewhat controversial. It says that something is actually true iff it is necessarily actually true. But couldn't the actual world have been a different one, couldn't things have been different than they are? In some sense this certainly right, i.e., things could have been different than they are and in this case the actual world would have been a different one. However, the word 'actually' is a so-called *indexical* like ‘now’, ‘here’, or ‘I’ and ‘she’. These words once uttered in a context refer rigidly to their referent, that is, they refer to the same object in all possible worlds.[2] If this idea is accepted (iv) is true, as ‘actually’ refers to one and the same world, i.e., the actual world in every possible world. No matter which world we are at ‘actually’ always takes us back to the actual world: if something is actually the case than this is necessary so (of course this does **not** mean that from $@\varphi$ we can infer $\Box\varphi$). Some interesting discussion on this theme may be found here.

[2]See the SEP entry on Indexicals for further details or, alternatively, any proficient introduction to the philosophy of language such Bill Lycan’s: Philosophy of Language.

Part III

Non-classical Logics

Chapter 7

Intuitionistic Logic

Modal logic extends classical logic by further modal/intensional operators. This is motivated by the observation that there are a lot of things we cannot say in classical logic, as classical logic can only handle truth-functional contexts. In contrast, some theorists think that classical logic does (also) say too much, that is, that it makes too strong assumptions. One key point of contention is the so-called *Law of Excluded Middle* (LEM) which says that every proposition is either true or false, that is, the truth value of every proposition is decided and there are no truth values beyond *true* and *false*.

Intuitionism or Intuitionistic Logic is one particular instalment of this view, which has its origins in a particular view on the foundations of mathematics. Oversimplifying somewhat the issue is whether mathematics is about the discovery of structures that are "out there" for everyone to discover like particular species of trees (of course at least prima facie in contrast to trees mathematical structures seem to be of abstract rather than of concrete nature) or, alternatively, whether mathematics is about constructing the structures via our mental activity. While on the former view LEM seems well motivated: either something is out there or it isn't (either it is true or false), on the latter view the situation is different. On such a constructivist view LEM, i.e., $\varphi \vee \neg\varphi$ would intuitively amount to the claim that either we have established φ or we have established $\neg\varphi$. But even a classical logician would arguably refrain from making such a claim.

Such a constructivist understanding of LEM, however, also suggests that in intuitionistic logic the logical connectives are not truth-functional as in classical logic: How are we to understand the logical connectives in intuitionistic logic?

7.1 Classical vs Intuitionistic Logical Connectives

In classical logic some logical connectives can be defined in terms of other connectives. We already saw that the negation can be defined by appeal to the logical constant '$\bot$' and the conditional '$\longrightarrow$'. Indeed, one can even define conjunction and disjunction using '$\bot$' and '$\longrightarrow$':

$$\varphi \wedge \psi \equiv_{def} \underbrace{(\varphi \longrightarrow (\psi \longrightarrow \bot)) \longrightarrow \bot}_{\neg(\varphi \longrightarrow \neg\psi)}$$

$$\varphi \vee \psi \equiv_{def} \underbrace{(\varphi \rightarrow \bot) \rightarrow \psi}_{\neg\varphi \rightarrow \psi}$$

As a matter of fact in classical logic all truth-functions can be defined/represented using only the logical constant '$\bot$' and the conditional '$\rightarrow$'. Alternatively, we could use '$\neg$' together with '$\wedge$' (or '$\neg$' together with '$\vee$') for this purpose.

This means that in classical logic we could get by with a language that only has the logical connective '$\rightarrow$' and the '$\bot$' constant (or only the pair of connectives '$\neg$' and '$\wedge$', or, alternatively, '$\neg$' and '$\vee$').

Exercise 43. *Check these claims using truth-tables, that is*

(i) show that the definitions of $\varphi \wedge \psi$ and $\varphi \vee \psi$ in terms of '$\rightarrow$' and '$\bot$' yield the right truth-functions (that is the correct truth tables for $\wedge$ and $\vee$;

(ii) show that '$\bot$', '$\rightarrow$', '$\vee$' can be defined in terms of '$\wedge$' and '$\neg$';

(iii) show that '$\bot$', '$\rightarrow$', '$\wedge$' can be defined in terms of '$\vee$' and '$\neg$'.

In contrast, in intuitionistic logic we may define negation as we did in Chapter 2 but all the other connections no longer hold: neither conjunction nor disjunction can be defined in terms of the conditional and vice versa; neither disjunction nor conjunction can be defined in terms of the other and negation. This shows that, in some sense, e.g., disjunction in classical logic is very different to disjunction in classical logic.

A heuristic for understanding the connectives intuitionistically is to replace "truth" by "proof": in classical logic a disjunction is true iff one of the disjuncts is true. In intuitionistic logic we ask under which conditions there is a proof of a disjunction. The answer is: if there is a proof of one of the disjuncts. This seems innocent enough but let's now focus on an instance of LEM, $\neg p \vee p$. This is classically true, if either p is true, or p is false and thus $\neg p$ true. But now we ask under which conditions there is a proof of '$p \vee \neg p$'? From the intuitionistic perspective this is the case if either there is a proof of 'p' or if 'p' can be refuted. But that does not even hold for classical logic: it is not the case that for every formula φ we can either prove or refute it (in particular, without additional premises no propositional variable can be proved or refuted in classical logic)!

The idea of moving from *truth-* to *proof-conditions* leads to the so-called BHK-interpretation of intuitionistic logic (named after Brouwer, Heyting and Kolgomorov all of which are important figures in the foundations of intuitionistic logic and, more general, constructive mathematics:

- No proof of '$\bot$' exists;
- there is a proof of $\varphi \vee \psi$ iff there is a proof of φ or a proof of ψ;
- there is a proof of $\varphi \wedge \psi$ iff there is a proof of φ and a proof of ψ;
- there is a proof of $\varphi \rightarrow \psi$ iff every proof of φ can be transformed into a proof of ψ;

For this to make sense it is important to understand 'proof' in an intuitive sense and not as 'proof in intuitionstic logic'. The BHK-interpretation is meant as an heuristic that helps us grasp the meaning of the connectives in intuitionistic logic. If we understanding 'proof' as 'proof in intuitionstic logic' this heuristic would break down. This said, the BHK-interpretation is very much in line with introduction and elimination rules for $\vee, \wedge$, and $\rightarrow$ we encountered in natural deduction.

How, are we understand negation in intuitionistic logic. Our formal language will not change so $\neg\varphi$, again, is short for the formula $\varphi \rightarrow \bot$. Given the BHK-interpretation there is a proof of $\neg\varphi$ iff every proof of φ can be transformed into a proof of $\bot$. According to the BHK-interpretation (and hopefully in alignment with our intuitive understanding) that's the case if we can prove that there is no proof of φ, i.e., if we can refute φ. Given this understanding it is rather immediate why the interdefinability of the logical connective $\wedge$ and $\vee$ breaks down in intuitionistic logic (recall that in classical logic we may define $\varphi \wedge \psi$ by the formula $\neg(\neg\varphi \vee \neg\psi)$). Now, according to the BHK-interpretation there is a proof of $\varphi \wedge \psi$ iff there is a proof of φ and there is a proof of ψ. In contrast, $\neg(\neg\varphi \vee \neg\psi)$ amounts to the claim that there is a refutation of the fact that either φ is refutable or ψ is refutable. That is decisively not the same thing than saying that there is a proof for both φ, but also for ψ.

Summing up, the connectives in classical logic and intutionistic logic need to be thought of differently and, as a consequence, the intuitions we used about truth and falsity that we successfully used in reasoning about the connective in classical logic need to be treated with some care.

7.2 Natural Deduction for Intuitionistic Logic

We just discussed that the logical connective in intuitionistic logic are different beasts in comparison to classical logic. We now need to ask what effect this has for the rules of natural deduction we proposed in Chapter 2. We know that at least one rule needs to be changed or dropped for otherwise we could derive LEM, which we found intuitionistically unacceptable.

However, as mentioned, the heuristics for understanding the intuitionistic connectives sketched in the previous section align nicely with the elimination and introduction rules for '$\wedge$', '$\vee$', and '$\rightarrow$'. As an example, we consider the case of the conditional: the elimination rule states that from φ and $\varphi \rightarrow \psi$ we can infer ψ. According to our heuristics this amounts to the claim that if there is a proof of φ and there is way of transforming a proof of φ into a proof of ψ, then there must be a proof of ψ. This seems OK. The introduction rule for $\rightarrow$ is equally plausible under the proposed heuristics: if assuming we are presented with a proof of φ, we can derive ψ, i.e., we can give a proof of ψ, then there is a proof of $\varphi \rightarrow \psi$.

If the introduction and elimination rules are intuitionistically acceptable only one rule remains: proof by contradiction. Recall that the rule of proof by contradiction allows us to infer φ if the assumption $\varphi \rightarrow \bot$ allows us to derive a contradiction, i.e., if the assumption that φ is false leads to a contradiction, then φ must be true.

$$
\begin{array}{r|l|ll}
m & & \varphi \longrightarrow \bot & \\ \cline{3-3}
\vdots & & \vdots & \\
n & & \bot & \\
n+1 & \multicolumn{2}{l}{\varphi} & \text{PbC}, m\text{–}n
\end{array}
$$

Clearly, this rule is suspicious from the perspective of intuitionistic logic: Line $n+1$ seems only justified, if every formula φ is either true or false, which is an assumption we are not licensed to make in intuitionistic logic. Similarly, if from the assumption that there is a proof of $\varphi \longrightarrow \bot$, we infer the there is a proof of $\bot$. Then, arguably, we are licensed to infer that there is no proof of $\varphi \longrightarrow \bot$—again that's not the same thing as saying that there is a proof of φ. In sum, in intuitionistic logic the rule PbC needs to be dropped.

We obtain a natural deduction system for intuitionistic logic by adopting all basic rules of the natural deduction system for classical logic save the rule PbC, and by adding the rule of explosion—in classical logic we can, using PbC, show that $\bot$E is a derived rule (exercise)—which we will denote by $\bot$E:

$$
\begin{array}{r|ll}
m & \bot & \\
\vdots & \vdots & \\
n & \varphi & \bot\text{E}, m
\end{array}
$$

At least prima facie this rule seems intutionistically acceptable for we have explicitly ruled out that there a proof of $\bot$. So, surely, if there is a proof of $\bot$, then anything follows. .

Basic Rules for Intuitionistic Logic (IL)

Conjunction Introduction

$$
\begin{array}{r|ll}
m & \varphi & \\
n & \psi & \\
\vdots & \vdots & \\
j & \varphi \land \psi & \land\text{I}, m, n
\end{array}
$$

Conjunction Elimination

$$
\begin{array}{r|ll}
m & \varphi_1 \land \varphi_2 & \\
\vdots & \vdots & \\
n & \varphi_i & \land\text{E}, m
\end{array}
$$

$i \in \{1, 2\}$

Disjunction Introduction

Disjunction Elimination

$$
\begin{array}{l|ll}
m & \varphi_i & i \in \{1, 2\} \\
\vdots & \vdots & \\
j & \varphi_1 \vee \varphi_2 & \vee\mathrm{I},\ m
\end{array}
$$

$$
\begin{array}{l|l|ll}
m & \multicolumn{2}{l}{\varphi \vee \psi} & \\
\vdots & \multicolumn{2}{l}{\vdots} & \\
j & & \varphi & \\
\vdots & & \vdots & \\
n-1 & & \chi & \\
n & & \psi & \\
\vdots & & \vdots & \\
k & & \chi & \\
k+1 & \multicolumn{2}{l}{\chi} & \vee\mathrm{E},\ m,\ j\text{–}n-1,\ n\text{–}k
\end{array}
$$

Conditional Introduction

$$
\begin{array}{l|l|ll}
m & & \varphi & \\
\vdots & & \vdots & \\
n & & \psi & \\
n+1 & \multicolumn{2}{l}{\varphi \longrightarrow \psi} & \longrightarrow\mathrm{I},\ m\text{–}n
\end{array}
$$

Conditional Elimination

$$
\begin{array}{l|ll}
m & \varphi & \\
n & \varphi \longrightarrow \psi & \\
\vdots & \vdots & \\
j & \psi & \longrightarrow\mathrm{E},\ m,\ n
\end{array}
$$

Explosion (⊥-Elimination)

$$
\begin{array}{l|ll}
m & \bot & \\
\vdots & \vdots & \\
n & \varphi & \bot\mathrm{E},\ m
\end{array}
$$

Notice since PbC is not a rule of our natural deduction system, many derived rule of classical logic will not be derived rules of intuitionistic logic.

Theorems of IL

We collect some interesting theorems of/facts about intuitionistic logic to get a better sense of the contrast between classical and intuitionistic logic:

(i) $\neg(\varphi \wedge \neg\varphi)$

(ii) $\varphi \longrightarrow \neg\neg\varphi$

(iii) $\varphi, \neg\varphi \vdash_{\mathrm{IL}} \psi$

(iv) $\varphi \to \psi \vdash_{\mathsf{IL}} \neg\psi \to \neg\varphi$

(v) $\neg\neg\neg\varphi \to \neg\varphi$

(vi) $\neg\neg(\varphi \vee \neg\varphi)$

1	$\varphi \vee \neg\varphi \to \bot$	
2	φ	
3	$\varphi \vee \neg\varphi$	$\vee$I, 2
4	$\bot$	$\to$E, 1, 3
5	$\varphi \to \bot$	$\to$I, 2–4
6	$\varphi \vee \neg\varphi$	$\vee$I, 5
7	$\bot$	$\to$E, 1, 6
8	$\underbrace{(\varphi \vee \neg\varphi \to \bot) \to \bot}_{\neg\neg(\varphi\vee\neg\varphi)}$	$\to$I, 1–7

(vii) $\neg(\varphi \vee \neg\varphi) \to \neg\varphi$

(viii) $\neg(\varphi \vee \neg\varphi) \to \neg\neg\varphi$

(ix) $\neg\neg\varphi \wedge \neg\neg(\varphi \to \psi) \to \neg\neg\psi$

1	$\neg\neg\varphi \wedge \neg\neg(\varphi \to \psi)$	
2	$\neg\neg\varphi$	$\wedge$E, 1
3	$\neg\neg(\varphi \to \psi)$	$\wedge$E, 1
4	$\psi \to \bot$	
5	$\varphi \to \psi$	
6	φ	
7	ψ	$\to$E, 5, 6
8	$\bot$	$\to$E, 4, 7
9	$\varphi \to \bot$	$\to$I, 6–8
10	$\bot$	$\to$E, 2, 9
11	$(\varphi \to \psi) \to \bot$	$\to$I, 5–10
12	$\bot$	$\to$E, 3, 11
13	$(\psi \to \bot) \to \bot$	$\to$I, 4–12
14	$\neg\neg\varphi \wedge \neg\neg(\varphi \to \psi) \to \neg\neg\psi$	$\to$I, 1–13

Exercise 44. *Give the missing proofs for the theorems of IL.*

Let us comment on some of the items above. (i) is the so-called principle of non-contradiction (PNC). IT says that there cannot be any contradiction. Classically it is equivalent to LEM, but intuitionistically LEM and PNC. While PNC is a theorem of IL adding LEM to IL leads us back to classical logic: intuitionistically saying that there are no contradictions is very different from asserting LEM. (ii) is so-called double negation introduction (DNI). Recall that classically φ and $\neg\neg\varphi$ are equivalent for every formula φ. However, as we shall discuss below in intuitionistic logic we only have DNI and not the converse direction, that is double negation elimination (DNE). (iii) is a form of explosion. This form of explosion holds intuitionistically but as we discuss in the next chapter it does not hold for some of the many-valued logics we discuss in the next chapter. (vi) highlights again that φ and $\neg\neg\varphi$ are not equivalent in IL: in the scope of a double negation LEM holds. Interestingly, this fact together with (ix) ultimately allows us to establish the following claim:

$$\text{if } \vdash_{\mathrm{CL}} \varphi, \text{ then } \vdash_{\mathrm{IL}} \neg\neg\varphi.$$

That is, if φ is a theorem of classical logic, then $\neg\neg\varphi$ is a theorem of intuitionistic logic. We refer to Chapter 6 in *Burgess: Philosophical Logic* for a proof of this claim.

Importantly, as discussed above neither LEM nor DNE are theorems of IL. If they were we could use LEM and DNE as derivable rules, i.e., could introduce them at any point in our natural deduction proof. But in this case PbC would turn out to be a derivable rule again:

- From LEM to PbC

m	$\quad\mid\ \varphi \to \bot$	
$\vdots$	$\quad\mid\ \vdots$	
n	$\quad\mid\ \bot$	
$n+1$	$(\varphi \to \bot) \to \bot$	$\to$I, m–n
$n+2$	$\varphi \vee (\varphi \to \bot)$	LEM
$n+3$	$\quad\mid\ \varphi \to \bot$	
$n+4$	$\quad\mid\ \bot$	$\to$E, $n+1$, $n+3$
$n+5$	$\quad\mid\ \varphi$	$\bot$E, $n+4$
$n+6$	$\quad\mid\ \varphi$	
$n+7$	$\quad\mid\ \varphi$	R, $n+6$
$n+8$	φ	$\vee$E, $n+2$, $n+3$–$n+5$, $n+6$–$n+7$

- From DNE to PbC

m	$\mid \; \underline{\varphi \to \bot}$	
$\vdots$	$\mid \; \vdots$	
n	$\mid \; \bot$	
$n+1$	$(\varphi \to \bot) \to \bot$	$\to$I, m–n
$n+2$	$((\varphi \to \bot) \to \bot) \to \varphi$	DNE
$n+3$	φ	$\to$E, $n+1$, $n+2$

As the rule PbC is what sets intuitionistic logic and classical logic apart the foregoing shows that either LEM or DNE collapses intutionistic logic into classical logic. A consequence of the observation that both LEM and DNE lead us back to classical logic is that the two are equivalent over classical logic in the following sense:

(*) $\neg\neg(\varphi \vee \neg\varphi) \to \varphi \vee \neg\varphi \vdash_{\mathrm{IL}} \varphi \vee \neg\varphi$

(**) $\varphi \vee \neg\varphi \vdash_{\mathrm{IL}} \neg\neg\varphi \to \varphi$

(*) is straightforward due to Item (vi). For (**) we reason as follows:

1	$\mid \; \underline{\varphi \vee \neg\varphi}$	
2	$\mid \; \mid \; \underline{\varphi}$	
3	$\mid \; \mid \; \mid \; \underline{\neg\neg\varphi}$	
m	$\mid \; \mid \; \mid \; \varphi$	R, 2
$m+1$	$\mid \; \mid \; \neg\neg\varphi \to \varphi$	$\to$I, 3–m
$m+2$	$\mid \; \mid \; \underline{\neg\varphi}$	
$m+3$	$\mid \; \mid \; \mid \; \underline{\neg\neg\varphi}$	
$\vdots$	$\mid \; \mid \; \mid \; \vdots$	
n	$\mid \; \mid \; \mid \; \varphi$	
$n+1$	$\mid \; \mid \; \neg\neg\varphi \to \varphi$	$\to$I, $m+3$–n
$n+2$	$\mid \; \neg\neg\varphi \to \varphi$	$\vee$E, 1, 2–$m+1$, $m+2$–$n+1$
$n+3$	$\varphi \vee \neg\varphi \to (\neg\neg\varphi \to \varphi)$	$\to$I, 1–$n+2$

Exercise 45. *Give the proof of* (*) *and supply the missing steps in the proof of* (**).

7.3 Semantics for Intuitionistic Logic

So far we understood the enterprise of semantics as the quest of specifying conditions under which the sentences of our formal languages are true (false). Reflecting on our discussion of Section 7.1 this understanding seems to sit ill with our understanding of the connectives in intuitionistic logic, which seem to be built on "proof-conditions" rather than "truth-conditions". However, there is a way of how to provide a truth-conditional semantics for IL. The key insight is that, as we mentioned in passing in Part II, 'provability' can be understood as a particular form of $\Box$-modality. It turns out that using this insight one can give a sound and complete possible world semantics for IL.

Definition 46 (IL-Frames)**.** $\mathsf{F} = \langle W, \leq, \rangle$ *is an* IL*-frame iff* $W \neq \emptyset$ *and* $\leq \subseteq W \times W$ *ia reflexive and transitive relation on* W*.*

In other words a IL-frame is just a modal $\mathsf{S4}$-frame. However, valuations on IL-frames will need to satisfy a particular constraint that is sometimes called *persistency*: if a propositional variable is true at a world, then it is true at all worlds that are accessible from that world.

Definition 47 (IL-valuation)**.** *A function* $V : \mathsf{At} \times W \longrightarrow \{\mathsf{T}, \mathsf{F}\}$ *is called an* IL*-valuation on an* IL*- frame iff for all* $w, v \in W$ *with* $w \leq v$ *and all* $p \in \mathsf{At}$*:*

$$\textit{if } V(p, w) = \mathsf{T}, \textit{ then } V(p, v) = \mathsf{T}.$$

Definition 48 (IL-Interpretation/Model)**.** *Let* F *be an* IL*-frame and* V *an* IL*-valuation on* F*. Then the interpretation* $\mathtt{I}_{(F,V)}$ *is a function that assigns to every formula* φ *either the value* T *or the value* F *relative to a world* w*:*

(i) $\mathtt{I}_{(F,V)}(\varphi, w) = V(\varphi, w)$*, if* φ *is a propositional variable;*

(ii) $\mathtt{I}_{(F,V)}(\varphi, w) = \mathsf{F}$*, if* φ *is '*$\perp$*';*

(iii) if φ *is a formula* $\psi \wedge \chi$*, then*

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, w) = \mathsf{T} \textit{ and } \mathtt{I}_{(F,V)}(\chi, w) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

(iv) if φ *is a formula* $\psi \vee \chi$*, then*

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, w) = \mathsf{T} \textit{ or } \mathtt{I}_{(F,V)}(\chi, w) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

(v) if φ *is a formula* $\psi \longrightarrow \chi$*, then*

$$\mathtt{I}_{(F,V)}(\varphi, w) := \begin{cases} \mathsf{T}, & \textit{if } \mathtt{I}_{(F,V)}(\psi, v) = \mathsf{F} \textit{ or } \mathtt{I}_{(F,V)}(\chi, v) = \mathsf{T} \textit{ for all } v \textit{ s.t. } w \leq v; \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

The pair (F, V) is sometimes called a **IL-model**. *We write $((F, V)), w \Vdash \varphi$ iff $\mathtt{I}_{F,V}(\varphi, w) = \mathtt{T}$, and say that φ is true at w in the model (F, V). If for $(F, V), w \Vdash \varphi$ all $w \in W$, we say that φ is true in the model (F, V). If $(F, V), w \Vdash \varphi$ is true for all valuations V on F, i.e., for all models on the frame F, we write $F, w \Vdash \varphi$ and say that φ is true in the frame F at world w.*

It is worth pausing and reflect about what this means for the truth conditions for negation, that is, formulas of the form $\neg\varphi$. The latter is of course shorthand for $\varphi \longrightarrow \bot$ and we obtain:

$$\begin{aligned}(F, V), w \Vdash \neg\varphi \text{ iff } & (F, V), w \Vdash \varphi \longrightarrow \bot \\ & \text{iff for all v s.t. } w \leq v\text{: } (F, V), v \nVdash \varphi \text{ or } (F, V), v \Vdash \bot \\ & \text{iff for all v s.t. } w \leq v\text{: } (F, V), v \nVdash \varphi\end{aligned}$$

This tell us that $\neg\varphi$ is true at a world iff φ is not true in any world that is accessible from w—at no point will φ turn out true. From the terms of modal logic the intuitionistic formulas $\varphi \longrightarrow \psi$ and $\neg\varphi$ should thus be understood as $\Box(\varphi \longrightarrow \psi)$ and $\Box\neg\varphi$ respectively with the connectives understood classically in the modal formulas.

Example 49. *To illustrate the semantics let's investigate the conditions under which $\varphi \vee \neg\varphi$ and, respectively, $\neg\neg(\varphi \vee \neg\varphi)$ is true at a world w in an* IL*-model (F, V):*

- $\varphi \vee \neg\varphi$:

  $$\begin{aligned}(F, V), w \Vdash \varphi \vee \neg\varphi \textit{ iff } & (F, V), w \Vdash \varphi \textit{ or } (F, V), w \Vdash \neg\varphi \\ \textit{iff } & (F, V), w \Vdash \varphi \textit{ or for all } v \textit{ s.t. } w \leq v\textit{:}(F, V), v \nVdash \varphi\end{aligned}$$

  This certainly does not hold for every world in an IL*-model: there are worlds w such that $(F, V), w \nVdash \varphi$ but there exists $v \geq w$ such that $(F, V), v \Vdash \varphi$.*

- $\neg\neg(\varphi \vee \neg\varphi)$:

  $$\begin{aligned}(F, V), w \Vdash \neg\neg(\varphi \vee \neg\varphi) & \textit{ iff for all } v \textit{ s.t. } w \leq v\textit{: } (F, V), v \nVdash \neg(\varphi \vee \neg\varphi) \\ & \textit{ iff for all } v \textit{ s.t. } w \leq v \textit{ there exists } u \textit{ s.t. } v \leq u\textit{: } (F, V), u \Vdash \varphi \vee \neg\varphi \\ & \textit{ iff f. a. } v \textit{ s.t. } w \leq v \textit{ there ex. } u \textit{ s.t. } v \leq u\textit{: } (F, V), u \Vdash \varphi \textit{ or } (F, V), u \Vdash \neg\varphi \\ & \textit{ iff f. a. } v \textit{ s.t. } w \leq v \textit{ there ex. } u \textit{ s.t. } v \leq u\textit{: } (F, V), u \Vdash \varphi \textit{ or f. a. } z \geq u\textit{: } (F, V), z \nVdash \varphi\end{aligned}$$

  So $\neg\neg(\varphi \vee \neg\varphi)$ is true at w if either there is world accessible from w in which φ will be true or φ will not be true in any world accessible from w. This is trivially the case on our semantics, i.e., $\neg\neg\varphi$ is a logical truth of IL.

As a consequence of the semantics the persistency property property that we imposed on propositional variables generalizes to all formulas of the language:

Proposition 50 (Persistency). *Let (F, V) be an* IL*-model. Then for all $w, v \in W$ and all formulas φ:*

$$\textit{if } w \leq v \textit{ and } (F, V), w \Vdash \varphi, \textit{ then } (F, V), v \Vdash \varphi.$$

That is, if φ is true at a world w, then φ will be true at all worlds accessible from w.

Proof. The proof is a proof by induction on the complexity φ (roughly the number of steps needed to construct a formula out of atomic formulas), that is, one first shows the claim for atomic formulas. Then one shows the claim for formulas of the form $\psi \wedge \chi$, $\psi \vee \chi$, and $\psi \longrightarrow \chi$ where on is allowed that the claim has already been established for the formulas ψ and χ. □

Logical consequence and Adequacy

Logical consequence for IL is defined as one would expect:

Definition 51 (IL-logical consequence). *Let* Γ *be a set of formulas and* φ *a formula. We say that* φ *is an* IL*-logical consequence of* Γ *iff for every* IL*-model* (F, V) *and every world* $w \in W$*:*

$$(F, V), w \Vdash_{\mathrm{IL}} \gamma, \textit{ for all } \gamma \in \Gamma, \textit{ then } (F, V), w \Vdash_{\mathrm{IL}} \varphi.$$

We write $\Gamma \vDash_{\mathrm{IL}} \varphi$ *to convey that* φ *is an* IL*-logical consequence of* Γ.

If F is an IL-frame we write $\Gamma \vDash_F \varphi$ to denote that φ follows from Γ on F, i.e., for every IL-valuation V and a every world w:

$$(F, V), w \Vdash_{\mathrm{IL}} \gamma, \text{ for all } \gamma \in \Gamma, \text{ then } (F, V), w \Vdash_{\mathrm{IL}} \varphi.$$ [1]

We can now state the soundness and completeness of IL with respect to possible world semantics for intuitionistic logic.

Proposition 52 (Adequacy of IL). *Let* Γ *be a set of formulas and* φ *a formula. Then*

$$\Gamma \vdash_{\mathrm{IL}} \varphi \textit{ iff } \Gamma \vDash_{\mathrm{IL}} \varphi$$

Of course, we may choose $\Gamma = \emptyset$. Then the adequacy result tells us that a formula is a theorem of IL iff it is true in every IL-model, that is, if the formula is IL-valid.

Using the adequacy result of IL we can highlight a distinctive feature of IL, a feature that sets it apart from classical logic, but also some many-valued logics: the **disjunction property**. The disjunction property says that a disjunction $\varphi \vee \psi$ is a theorem of the logic iff one of the disjunction is a theorem of the logic:

(DP) $$\text{if } \vdash_{\mathrm{IL}} \varphi \vee \psi, \text{ then } \vdash_{\mathrm{IL}} \varphi \text{ or } \vdash_{\mathrm{IL}} \psi$$

[1] This means that:

$$\Gamma \vDash_{\mathrm{IL}} \varphi \text{ iff } \Gamma \vDash_F \varphi \text{ for every IL-frame } F.$$

Proof of DP. Suppose $\vdash_{\text{IL}} \varphi \vee \psi$, and $\nvdash_{\text{IL}} \varphi$ and $\nvdash_{\text{IL}} \psi$, then we can find two IL-models (F', V') and (F'', V'') and worlds $w' \in W'$ and $w'' \in W''$ such that: $(F', V'), w' \nVdash \varphi$ and $(F'', V''), w'' \nVdash \psi$. In particular, we can choose F' such that $w' \leq' v$ for all $v \in W'$ and if $v \leq' w$, then $v = w'$. Similarly, F'' can be chosen such that $w'' \leq'' v$ for all $v \in W''$ and if $v \leq'' w''$, then $v = w''$. We now construct an new IL-frame $F = \langle W, \leq \rangle$ out F' and F'' as follows:

$$\begin{aligned} W &:= W' \cup W'' \cup \{w^\star\} \\ \leq &:= \leq' \cup \leq'' \cup \{\langle w^\star, v\rangle \mid v \in W\} \end{aligned}$$

F is reflexive and transitive and thus an IL-frame. Let V be a valuation such that for all $p \in \mathsf{At}$ and $w \in W$:

$$V(p, w) := \begin{cases} V'(p, w), & \text{if } w \in W'; \\ V''(p, w), & \text{if } w \in W''; \\ V(p, w), & \text{otherwise.} \end{cases}$$

Then $(F, V), w^\star \nVdash \varphi \vee \psi$ and thus there is an IL-frame such that $F \nVdash \varphi \vee \psi$ contradicting our initial assumption. (Assume otherwise and assume $(F, V), w^\star \Vdash \varphi$. Then by persistency $(F, V), w' \Vdash \varphi$. Contradiction. Similarly if we assume that $(F, V), w^\star \Vdash \psi$.)

We can conclude that (DP) holds. □

From the perspective of intuitionistic logic DP seems compelling and aligns nicely with the understanding of the connectives introduced in Section 7.1. In contrast, DP does not hold for classical logic as can be readily observed by noting that $\vdash_{\text{CL}} \varphi \vee \neg\varphi$ for every formula φ, but clearly it is not the case that $\vdash_{\text{CL}} \varphi$ or $\vdash_{\text{CL}} \neg\varphi$ for every formula φ. For example, let φ be any propositional variable.

Let's conclude this section by pointing out that IL enjoys the **deduction theorem**:

$\vdash_{\text{IL}}$	$\Gamma \cup \{\varphi\} \vdash_{\text{IL}} \psi$ iff $\Gamma \vdash_{\text{IL}} \varphi \longrightarrow \psi$
$\vDash_{\text{IL}}$	$\Gamma \cup \{\varphi\} \vDash_{\text{IL}} \psi$ iff $\Gamma \vDash_{\text{IL}} \varphi \longrightarrow \psi$

7.3.1 Kripke semantics: From Intuitionistic Logic to Classical Logic

We now ask whether we can single out IL-frames that on which the intuitive connective collapse into their classical counterparts. Put differently, on which IL-frames does classical logic hold. It turns out that these are precisely the euclidean IL-frames, that is, equivalent frames.

Lemma 53. *Let F be an IL-frame. Then*

$$F \Vdash \varphi \vee \neg\varphi \text{ iff } F \text{ is euclidean.}$$

Proof. For the left-to-right direction we assume that $F \Vdash p \vee \neg p$ and for reductio assume that F is not euclidean. Then there are $w, u, v \in W$ such that $w \leq u$ and $w \leq v$ and $u \nleq v$. We now define an IL-valuation as follows: $V(p, w) = V(p, v) = \mathsf{F}$ and $V(p, u) = \mathsf{T}$. Then $(F, V), w \nVdash p$ and $(F, V), w \nVdash \neg p$ (since $w \leq u$ and $V(p, u) = \mathsf{T}$) and we can infer that $(F, V), w \nVdash p \vee \neg p$.

Moreover, since V respects persistency it is clearly an IL-valuation and we have arrived at a contradiction. We conclude that F must be euclidean.

For the converse direction we assume that F is euclidean and show that $F \Vdash \varphi \vee \neg\varphi$. That is, we need to show that for every valuation V and word w: $(F, V), w \Vdash p$ or $(F, V), w \Vdash \neg p$. To this effect we assume $(F, V), w \nVdash p$ and show that $(F, V), w \Vdash \neg p$.[2] For reduction we now assume that $(F, V), w \nVdash \neg p$, i.e., there is a world z such that $w \leq z$ and $(F, V), z \Vdash p$. Since we have $w \leq z$ and $w \leq w$ (IL-frames are reflexive) we infer $z \leq w$ since F is euclidean. But then, contrary to assumption, V is not an IL-valuation as it does not respect persistency because we have $z \leq w$, $V(p, z) = \mathsf{T}$ but $V(p, w) = \mathsf{F}$. So we have arrived at a contradiction and may conclude that $F \Vdash \varphi \vee \neg\varphi$. □

Since adding LEM to the intutionistic logic has the consequence of collapsing IL into classical logic, we can infer that over euclidean IL frames the IL-connectives collapse into their classical counterparts. Interestingly this fact leads to the observation that there are logics in between intuitionistic logic and classical logic. We illustrate this fact by the following lemma:

Lemma 54. *Let F be an IL-frame. Then*

$$F \Vdash \neg\varphi \vee \neg\neg\varphi \textit{ iff } F \textit{ is convergent.}$$

A frame is convergent iff for all $w, v, u \in W$ such that wRu and wRv there is a $z \in W$ such that uRz and vRz.

Proof. See problem sheet. □

The lemma shows (i) that $\neg\varphi \vee \neg\neg\varphi$ is not provable in intutionistic logic as the principle will not be true on all IL-frames (only convergent ones). (ii) Yet, it also shows that intuitionistic logic will not collapse into classical logic if we add the rule

$$\begin{array}{r|ll} \vdots & \vdots & \\ k & \neg\varphi \vee \neg\neg\varphi & \quad \text{WEM} \end{array}$$

to intutionistic logic since not all convergent frames are also euclidean. That is, by adding WEM to IL we get a logic in between intutionistic logic and classical logic. This logic is sometimes called KC.

Summing up possible world semantics is a powerful, adequate semantics for intuitionistic logic which can fruitfully employed to obtain result about intutionistiv logic viz. the proof of DP. It is another question whether the semantics provides us with an understanding of intuitionistic logic that is acceptable to an intuitionist, as it ultimately turns the intuitionists "proof-conditions" and constructive understanding of the connective into modal truth conditions that rely on the the truth conditions of classical connective in possible world semantics. We do no discuss the prospect of an intuitionistically acceptable semantic. Some remarks to this effect can be found, e.g., in Chapter 6 of *Burgess: Philosophical Logic.* Some alternative semantics for IL are discussed in Chapter 11 of *Goble: The Blackwell Guide to Philosophical Logic*

[2] We use the fact that classically $\varphi \vee \psi$ is equivalent to $\neg\varphi \longrightarrow \psi$.

Chapter 8

Many-valued logic

In intuitionistic logic the idea that the logical connectives are truth-functional, that is, that they can be analysed using truth tables was rejected. Ultimately, that led the intuitionist to reject LEM, DNE, as well as, proof by contradiction. Many-valued logic embraces a truth-functional understanding of logical connectives in terms of truth tables, but it introduces further truth-values such as *neither true nor false* and *both true and false* into the picture.[1]

Why should we allow for further truth values than true or false? In the philosophical literature the most prominent motivations for going beyond truth and falsity are:

- presupposition failure
- vagueness
- semantic paradox

Presupposition failure Consider the following sentences:

- The king of France is bald.
- Every American king lives in New York.

Are they true or false? Typically, if asked to answer this question, many competent speakers will answer "can't say" or "there is no king of France/American king", or "neither". The idea is that to make sense of the sentence we need to *presuppose* the esistence of the king of France or American kings. If this *presupposition* is not met, the sentence is meaningless and not truth apt. It does not receive a (classical) truth value.[2]

To give a more mathematical/computational spin to the idea of presupposition failure. Think of the assignment of truth values as an algorithm to which you input sentence of a language. Presupposition failures arise if the algorithm does not produce an output for a given input, i.e., if it does not compute…

[1]In principle, there is no limit to the number of different truth-values one can assumed. For example, there are even logics that assume infinitely many and uncountably many truth-values. If you are interested see *Introduction to Non-Cassical Logic* by Graham Priest.

[2]See the SEP entry for more on presuppositions and presupposition failure.

Vagueness You all have this one jumper you can't quite tell whether it is green or blue. Perhaps it neither or perhaps it's both. Well according to classical logic the jumper is either blue or it is not blue, and similarly for green. But also a blue jumper is not a green jumper and accordingly classical logic will force us to say whether the sentence 'this is a blue jumper' is true or false. We do not have the option of answering neither or both. If we were to allow for further truth values than $\mathsf{T}$and F, this would be possible.[3]

Semantic paradox Let L be the following sentence:

L is not true

Now we ask the question whether L is true. Assume it is true. Then what L says is the case, that is, L is not true: L is true and L is not true. By PbC we can infer L is not true. But then it is not the case that L is not true and by double negation elimination L is true. Again we have arrived at a contradiction. It seems that we must conclude that either L is both true and false, or that it is neither true or false.

If further truth-values are introduced one has to decide how to evaluate complex formulas: for example, if a propositional variable p were to receive the truth value *neither true nor false*, what is truth value of '$p \wedge q$' depending on the different truth values that can be assigned to q? Answering this question is the main task of the semantics for many-values logic. However, before we start answering semantic questions let's make our syntax precise.

8.1 Syntax

We will now change our language. In contrast to the previous chapters **we now take '$\neg$' to be a primitive symbol and '$\rightarrow$' as a defined symbol**, that is, our language $\mathcal{L}$ now consists of:[4]

- countably many propositional variables: $p_0, p_1, p_2, \ldots$
- propositional constants: $\bot$
- logical connectives: $\neg, \wedge, \vee$
- auxiliary symbols: (,)

We continue to refer to the set consisting of all propositional variables by At. The well-formed formulas of the language are defined by an inductive definition as expected:

1. if φ is a propositional variable, then φ is an $\mathcal{L}$ formula,
2. if φ is a propositional constant, then φ is an $\mathcal{L}$-formula,
3. if φ is a formula, the $\neg\varphi$ is a formula,

[3] You may again consult the SEP entry on vagueness for more on the phenomena of vagueness.

[4] Strictly speaking this is not necessary, but it is compelling to think of the conditional as defined symbol, as the conditional, as we shall see, no longer has the salient properties we typically would expect conditionals to have.

4. φ and ψ are formulas, then $(\varphi \wedge \psi)$ and $(\varphi \vee \psi)$ are $\mathcal{L}$-formula,

5. nothing else is a formula.

If in the present chapter we write $\varphi \longrightarrow \psi$ we take that to be shorthand for the formulas $\neg\varphi \vee \psi$. '$\varphi \leftrightarrow \psi$' is shorthand for '$(\varphi \longrightarrow \psi) \wedge (\psi \longrightarrow \varphi)$'.

8.2 Semantics for many-valued logics

Semantics for many-valued logic is a generalization of the semantics for classical propositional logic (TFL): formulas can have truth values other than T or F. We give semantics for some well-known three valued and four valued logics

8.2.1 Weak and Strong three-valued semantics

In three-valued semantics we add one further truth value to T and F, that is, the value nb. For the time being nb can be understood as either *neither true nor false* or *both true and false.* Of course, giving these two different reading one might expect that depending on which understanding of nb one opts for a different logic should result, but surprisingly these two different understandingd of nb will only become relevant once we define the consequence relation of the three-valued logics.

However, as in the semantics for classical logic, the starting point of the three-valued semantics is the notion of valuation:

Definition 55 (Three-valued valuation)**.** *A three valued valuation is a function* $V : \mathsf{At} \longrightarrow \{\mathsf{T}, \mathsf{nb}, \mathsf{F}\}$ *that assign to every propositional variable an element (truth-value) of the set* $\{\mathsf{T}, \mathsf{nb}, \mathsf{F}\}$.

The next step is to specify how complex formulas get assigned their truth value, that is, to specify conditions when formulas $\neg\varphi$, $\varphi \wedge \psi$, and $\varphi \vee \psi$ receive value T, nb, and F respectively. There are two different rationale/heuristics that one can adopt to this effect and these lead to two different semantics and logics.

Semantics for strong three-valued logics

On the first, the rationale is that we should assign a classical truth value to a formula, if we can decide which truth value a formula would receive in classical logic. For example, if the value of φ is T, then we know that $\varphi \vee \psi$ would receive value T in classical logic—independently of the truth value of ψ. Similarly, if the truth value φ is false, then we know that $\varphi \wedge \psi$ has to be false, again independently of the truth value of ψ—and in strong three-valued logics independently of whether ψ receives a classical truth value (Tor F) or the new truth value nb. Following this rationale we are led to the following truth tables:

φ	$\neg\varphi$
T	F
nb	nb
F	T

φ	ψ	$\varphi \wedge \psi$
T	T	T
T	nb	nb
T	F	F
nb	T	nb
nb	nb	nb
nb	F	F
F	T	F
F	nb	F
F	F	F

φ	ψ	$\varphi \vee \psi$
T	T	T
T	nb	T
T	F	T
nb	T	T
nb	nb	nb
nb	F	nb
F	T	T
F	nb	nb
F	F	F

Exercise 56. *Determine the truth tables of $\varphi \longrightarrow \psi$ and $\varphi \leftrightarrow \psi$.*

With these truth tables in mind we define the interpretation function for the strong three-valued logics, that is the function that assigns a truth value to all formulas of the language relative to a valuation.

Definition 57 (Strong three-valued interpretation)**.** *Let V be a three-valued valuation. Then an (strong three-valued)interpretation I_V is a function that assigns to every formula of the language an element (truth value) of the set $\{\mathsf{T}, \mathsf{nb}, \mathsf{F}\}$ relative to V such that:*

(i) if φ is a propositional variable, then $\mathrm{I}_V(\varphi) = V(\varphi)$;

(ii) if φ is $\bot$, then $\mathrm{I}_V(\varphi) = \mathsf{F}$;

(iii) if φ is a formula $\neg\psi$, then

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{F} \\ \mathsf{F}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{T} \\ \mathsf{nb}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{nb} \end{cases}$$

(iv) if φ is a formula $(\psi \wedge \chi)$, then:

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathrm{I}_V(\chi) = \mathsf{T} \\ \mathsf{F}, & \textit{if } \mathrm{I}_v(\psi) = \mathsf{F} \textit{ or } \mathrm{I}_V(\chi) = \mathsf{F} \\ \mathsf{nb}, & \textit{otherwise;} \end{cases}$$

(v) if φ is a formula $(\psi \vee \chi)$, then:

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{T} \textit{ or } \mathrm{I}_V(\chi) = \mathsf{T} \\ \mathsf{F}, & \textit{if } \mathrm{I}_V(\psi) = \mathrm{I}_V(\chi) = \mathsf{F} \\ \mathsf{nb}, & \textit{otherwise.} \end{cases}$$

So far everything we said about strong three valued semantics works for both understanding nb as *neither true nor false* and understanding nb as *both true and false*. But now comes the tricky part of defining logical consequence and it is at this point that the difference between the two different understandings matters.

The general idea underlying the definition of logical consequence is that of *truth-preservation*: a formula φ is a logical consequence of a set of formulas Γ iff in all models in which all members of Γ are true, φ is also true.[5] When is a formula φ true in a model based on a valuation function V ($V \Vdash \varphi$)? So far the answer to this question was straightforward: if the interpretation function associated with the model assigns the value T to φ. If we understand the truth value nb as *neither true nor false*, then this still seems to be the correct answer, i.e., $V \Vdash \varphi$ is $\mathtt{I}_V(\varphi) = \mathsf{T}$ where $\mathtt{I}_V$ is a strong three-valued interpretation as specified in Definition 57. However, if we adopt the understanding of nb as *both true and false*, then a formula φ should also be true in a model based on V if $\mathtt{I}_V(\varphi) = \mathsf{nb}$. After all if φ is both true and false relative to V, then it is, in particular, true in the model, i.e., $V \Vdash \varphi$ (of course φ is also false in the model).

These two different accounts of truth in a model give rise to two different three-valued logics: *Strong Kleene Logic* ($\mathsf{K3}$) and *Logic of Paradox* (LP). The former associated with understanding nb as *neither true nor false*, whereas the latter is is based on understanding nb as *both true and false*.

Definition 58 (Truth in a model). *Let V be a three-valued valuation and $\mathtt{I}_V$ the strong three-valued interpretation based on V. Then*

$$V \Vdash_{\mathsf{K3}} \varphi \equiv_{\mathsf{def}} \mathtt{I}_V(\varphi) = \mathsf{T};$$
$$V \Vdash_{\mathsf{LP}} \varphi \equiv_{\mathsf{def}} \mathtt{I}_V(\varphi) \in \{\mathsf{T}, \mathsf{nb}\}.^{6}$$

With the definition of truth in a model for $\mathsf{K3}$ and LP at hand we can move toward introducing the notion of logical consequence for these logics. Indeed we will define three different notion of logical consequence: logical consequence for $\mathsf{K3}$, logical consequence for LP, and, finally, logical consequence for so-called symmetric strong Kleene logic $\mathsf{KS3}$.[7] $\mathsf{KS3}$ can be understood as the logic that does not differentiate between the two different ways of understanding of nb: either understanding is acceptable.

Definition 59 (Logical consequence of $\mathsf{K3}$ and LP). *Let Γ be a set of formulas and φ a formula. Then*

$$\Gamma \vDash_{\mathsf{K3}} \varphi \textit{ iff for all } V\textit{: if } V \Vdash_{\mathsf{K3}} \gamma \textit{ for all } \gamma \in \Gamma\textit{, then } V \Vdash_{\mathsf{K3}} \varphi;$$
$$\Gamma \vDash_{\mathsf{LP}} \varphi \textit{ iff for all } V\textit{: if } V \Vdash_{\mathsf{LP}} \gamma \textit{ for all } \gamma \in \Gamma\textit{, then } V \Vdash_{\mathsf{LP}} \varphi;$$
$$\Gamma \vDash_{\mathsf{KS3}} \varphi \textit{ iff } \Gamma \vDash_{\mathsf{K3}} \varphi \textit{ and } \Gamma \vDash_{\mathsf{LP}} \varphi.$$

[5] Of course, in possible world semantics this was further relativized to worlds, but the gist remains the same.

[6] The truth values for which a formula is defined to be *true in a model* are also called the designated (truth) values of the logic. Accordingly, the designated value of $\mathsf{K3}$ is just T (as for CL), whereas the designated valued of LP are T and nb.

[7] While the names $\mathsf{K3}$ and LP are fairly established in the literature. There is no agreed upon name for $\mathsf{KS3}$. So you might find this logic discussed using a different name.

Reflecting on the definition of $\models_{\mathsf{KS3}}$ we see that it really gives us a consequence relation that is agnostic between the two different readings of nb: a formula φ is a KS3-consequence of a set of formulas Γ if φ is a logical consequences of Γ if nb is understood as *neither true nor false* and if nb is understood as *both true and false.*

On the basis of the definition $\Gamma \models_{\mathsf{KS3}} \varphi$ we see immediately that if a formula φ is a logical consequence of Γ for the logic KS3, then is is a logical consequence in the logics K3 and LP:

$$\text{if } \Gamma \models_{\mathsf{KS3}} \varphi, \text{ then } \Gamma \models_{\mathsf{K3}} \varphi;$$
$$\text{if } \Gamma \models_{\mathsf{KS3}} \varphi, \text{ then } \Gamma \models_{\mathsf{LP}} \varphi.$$

Let's then start the investigation into these logics by looking at their common core, i.e., the consequence relation for KS3. The first thing to note is that KS3 will not have schematic logical truths like LEM in classical logic, i.e., in classical logic $\varphi \vee \neg\varphi$ is logical truth for every formulas φ. In KS3 only formulas that contain the propositional constant $\bot$, can be logical truths: If $\bot$ does not occur in a formula φ, then φ is not a logical truth of KS3. This will also hold for K3 but not for LP. For this reason in studying the logic KS3 we need to study the inferences of the logic.

Example 60 (Inferences of KS3)**.** *Let collect some interesting inferences for* KS3*:*

1. $\models_{\mathsf{KS3}} \neg\bot$
2. $\varphi \models_{\mathsf{KS3}} \varphi$
3. $\bot \models_{\mathsf{KS3}} \varphi$
4. $\models_{\mathsf{KS3}} \bot \longrightarrow \varphi$
5. $\neg\neg\varphi \models_{\mathsf{KS3}} \varphi$
6. $\varphi \models_{\mathsf{KS3}} \neg\neg\varphi$
7. $\varphi \wedge \psi \models_{\mathsf{KS3}} \varphi$
8. $\varphi \models_{\mathsf{KS3}} \varphi \vee \chi$
9. $\neg(\varphi \wedge \psi) \models_{\mathsf{KS3}} \neg\varphi \vee \neg\psi$
10. $\neg\varphi \vee \neg\psi \models_{\mathsf{KS3}} \neg(\varphi \wedge \psi)$
11. $\varphi \vee \psi \models_{\mathsf{KS3}} \neg(\neg\varphi \wedge \neg\psi)$
12. $\neg(\neg\varphi \wedge \neg\psi) \models_{\mathsf{KS3}} \varphi \vee \psi$
13. $\varphi \wedge \neg\varphi \models_{\mathsf{KS3}} \psi \vee \neg\psi$

However, if is also important to see what kind of inferences we do not get in KS3*, i.e., there are φ and ψ such that:*

13. $\varphi, \neg\varphi \not\models_{\mathsf{KS3}} \bot$

14. $\not\models_{\mathsf{KS3}} \varphi \vee \neg\varphi$

15. $\varphi, \varphi \longrightarrow \psi \not\models_{\mathsf{KS3}} \psi$

Item 15 tells us the conditional elimination/modus ponens is not available in KS3, which also informs us that we **loose the following direction of the deduction theorem** (Exercise: why?):

$$(\dagger) \qquad \text{if } \Gamma \models \varphi \longrightarrow \psi, \text{ then } \Gamma \cup \{\varphi\} \models \psi.$$

Interestingly, as we will show in more detail in Section 8.3, K3 is the logic we obtain from KS3 if the rule of conditional elimination is added to KS3, that is, if the (†) direction of the deduction theorem is re-established.

Example 61. *Some facts about* K3

- $\varphi, \neg\varphi \models_{\mathsf{K3}} \bot$
- $\not\models_{\mathsf{K3}} \varphi \vee \neg\varphi$[8]
- $\varphi, \varphi \longrightarrow \psi \models_{\mathsf{K3}} \psi$
- *if* $\Gamma \models_{\mathsf{K3}} \varphi \longrightarrow \psi$*, then* $\Gamma \cup \{\varphi\} \models_{\mathsf{K3}} \psi$.
- (‡) *does not hold for* $\models_{\mathsf{K3}}$.

The converse direction of the deduction theorem, i.e.,

$$(\ddagger) \qquad \text{if } \Gamma \cup \{\varphi\} \models \psi, \text{ then } \Gamma \models \varphi \longrightarrow \psi$$

also fails for KS3. This direction corresponds to the rule of conditional introduction (Exercise: why?). If conditional introduction is added to KS3 we obtain the logic LP, and if both conditional elimination and conditional introduction are added to KS3, then we are back to classical logic.

Example 62. *Some facts about* LP

- $\varphi, \neg\varphi \not\models_{\mathsf{LP}} \bot$
- $\models_{\mathsf{LP}} \varphi \vee \neg\varphi$
- $\models_{\mathsf{LP}} \varphi$*, if* $\models_{\mathrm{CL}} \varphi$.
- $\varphi, \varphi \longrightarrow \psi \not\models_{\mathsf{LP}} \psi$
- *if* $\Gamma \cup \{\varphi\} \models_{\mathsf{LP}} \psi$*, then* $\Gamma \models_{\mathsf{LP}} \varphi \longrightarrow \psi$.
- (†) *does not hold for* $\models_{\mathsf{LP}}$.

[8]To be precise, this means that there is a formula φ such that $\models_{\mathsf{K3}} \varphi \vee \neg\varphi$, that is, $\models_{\mathsf{K3}} \varphi \vee \neg\varphi$ does not hold for all formula φ. This comment applies whenever we write $\not\models$.

Before moving on to the semantics for weak three-valued logics, it's worth commenting on Item 13, which is the characteristic inference of KS3 and distinguishes the logic from the four-valued logic FDE, which we will discuss shortly. In nutshell, 13 says the following: for a given valuation V, if $V \Vdash \varphi \wedge \neg\varphi$, then the third truth value must be understood as *both true and false.*

Semantics for weak three-valued logics

The second rationale for assigning truth values to complex formulas can by summed up by the following slogan: non-classical truth values are infectious. Once a formula φ has been infected by non-classically, non-classicality is passed on to all formulas φ is a subformula of, i.e., φ is occurs in said formula. For example, if the value of φ is nb, then φ infects $\varphi \vee \psi$ and the formula also receives the value nb—independently of the truth value of ψ. We are led to the following truth tables:

φ	$\neg\varphi$
T	F
nb	nb
F	T

φ	ψ	$\varphi \wedge \psi$
T	T	T
T	nb	nb
T	F	F
nb	T	nb
nb	nb	nb
nb	F	nb
F	T	F
F	nb	nb
F	F	F

φ	ψ	$\varphi \vee \psi$
T	T	T
T	nb	nb
T	F	T
nb	T	nb
nb	nb	nb
nb	F	nb
F	T	T
F	nb	nb
F	F	F

Exercise 63. *Determine the truth tables of $\varphi \longrightarrow \psi$ and $\varphi \leftrightarrow \psi$.*

In principle, we could again champion two understandings of nb and then define two different consequence relations as we did for strong three-valued logics. In practice, weak three-valued logics are almost exclusively discussed for the understanding of nb as *neither true nor false* and we'll confine ourselves to this reading.

Definition 64 (Weak three-valued interpretation)**.** *Let V be a valuation. Then an (weak three-valued) interpretation I_V is a function that assigns to every formula of the language an element (truth value) of the set* $\{\mathsf{T}, \mathsf{nb}, \mathsf{F}\}$ *relative to V such that:*

(i) if φ is a propositional variable, then $\mathtt{I}_V(\varphi) = V(\varphi)$;

(ii) if φ is $\bot$, then $\mathtt{I}_V(\varphi) = \mathsf{F}$;

(iii) if φ is a formula $\neg\psi$, then

$$\mathtt{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \text{if } \mathtt{I}_V(\psi) = \mathsf{F} \\ \mathsf{nb}, & \text{if } \mathtt{I}_V(\psi) = \mathsf{nb} \\ \mathsf{F}, & \text{if } \mathtt{I}_V(\psi) = \mathsf{T} \end{cases}$$

(iv) if φ is a formula $(\psi \wedge \chi)$, then:

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathrm{I}_V(\chi) = \mathsf{T} \\ \mathsf{nb}, & \textit{if } \mathrm{I}_v(\psi) = \mathsf{nb} \textit{ or } \mathrm{I}_V(\chi) = \mathsf{nb} \\ \mathsf{F}, & \textit{otherwise}; \end{cases}$$

(v) if φ is a formula $(\psi \vee \chi)$, then:

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{T} \textit{ or } \mathrm{I}_V(\chi) = \mathsf{T}, \textit{ and } \mathrm{I}_V(\psi) \neq \mathsf{nb} \neq \mathrm{I}_V(\chi); \\ \mathsf{nb}, & \textit{if } \mathrm{I}_v(\psi) = \mathsf{nb} \textit{ or } \mathrm{I}_V(\chi) = \mathsf{nb} \\ \mathsf{F}, & \textit{otherwise.} \end{cases}$$

We write $V \Vdash_{\mathsf{B3}} \varphi$ iff $\mathrm{I}_V(\varphi) = \mathsf{T}$.

In the definition $V \Vdash_{\mathsf{B3}} \varphi$ is, as per usual, short for φ is true in the model of *Bochvar's (internal) three-valued logic* based on V. Arguably, B3 is better known as *weak Kleene logic* (in contrast to *strong Kleene logic* (K3).

One can establish the following relation between B3 and K3.

Proposition 65. *Let V be a three-valued valuation and φ a formula. Then*

$$\textit{if } V \Vdash_{\mathsf{B3}} \varphi \textit{ then } V \Vdash_{\mathsf{K3}} \varphi.$$

That is, B3-truth implies K3-truth (but of course not the other way around. Exercise: give a counterexample to the converse direction).

Definition 66 (Logical consequencee for B3)**.** *Let Γ be a set of formulas and φ a formula. Then*

$$\Gamma \vDash_{\mathsf{B3}} \varphi \textit{ iff for all } V\textit{: if } V \Vdash_{\mathsf{B3}} \gamma \textit{ for all } \gamma \in \Gamma, \textit{ then } V \Vdash_{\mathsf{B3}} \varphi.$$

We list some examples of (non-) relations of logical consequence distinctive for B3:

(i) $\varphi, \neg\varphi \vDash_{\mathsf{B3}} \psi$

(ii) $\varphi, \varphi \longrightarrow \psi \vDash_{\mathsf{B3}} \psi$

(iii) $\varphi \vee \psi \vDash_{\mathsf{B3}} (\varphi \wedge \psi) \vee (\varphi \vee \neg\psi) \vee (\neg\varphi \vee \psi)$

(iv) $(\varphi \wedge \psi) \vee (\varphi \vee \neg\psi) \vee (\neg\varphi \vee \psi) \vDash_{\mathsf{B3}}$

(v) $\neg(\varphi \wedge \psi) \vDash_{\mathsf{B3}} (\neg\varphi \wedge \psi) \vee (\neg\varphi \vee \neg\psi) \vee (\varphi \vee \neg\psi)$

(vi) $(\neg\varphi \wedge \psi) \vee (\neg\varphi \vee \neg\psi) \vee (\varphi \vee \neg\psi) \vDash_{\mathsf{B3}} \neg(\varphi \wedge \psi)$

(vii) $\varphi \nvDash_{\mathsf{B3}} \varphi \vee \psi$

(viii) $\neg\varphi \nvDash_{\mathsf{B3}} \neg(\varphi \wedge \psi$

(ix) $\not\models_{\mathsf{B3}} \varphi$, if φ does not contain $\bot$.

(x) $\varphi, \varphi \longrightarrow \psi \models_{\mathsf{B3}} \psi$

(xi) if $\Gamma \Vdash_{\mathsf{B3}} \varphi \longrightarrow \psi$, then $\Gamma \cup \{\varphi\} \Vdash_{\mathsf{B3}} \psi$.

While Items (i)-(ii) and (ix)-(xi) are shared with K3, the remaining items highlight the peculiarities of B3, e.g., that the truth of a disjunction implies not only that one of the disjuncts are true, but also that both disjuncts have classical truth values (otherwise the disjunction would receive the value nb). The positive claims (iii)-(vi) also highlight that while B3-truth in a model implies K3-truth in a model, it is not the case that B3-logical consequence implies K3-logical consequence: none of the claims (iii)-(vi) hold for the K3 consequence relation. However, as for K3 we obtain one direction of the deduction theorem ((xi)), but don't obtain the converse direction.

Exercise 67. *Show that Items (i)-(xi) above are true.*

8.2.2 Four-valued semantics

So far we have discussed logics with three-valued semantics. As discussed, depending on whether the third truth-value is understood as *neither true nor false* or *both true and false* we obtain different logics. But if there are two different ways of understanding nb, why not introduce separate truth value for the two competing reading of nb: why not work in a four-valued semantics in which we have a value n that is to be understood as *neither true nor false* and a value b, which we take to stand for *both true and false*? Carrying out this idea will lead us to the logic of FDE (*First Degree Entailment*).

Definition 68 (Four-valued valuation)**.** *A four valued valuation is a function* $V : \mathsf{At} \longrightarrow \{\mathsf{T}, \mathsf{b}, \mathsf{n}, \mathsf{F}\}$ *that assigns to every propositional variable an element (truth-value) of the set* $\{\mathsf{T}, \mathsf{b}, \mathsf{n}, \mathsf{F}\}$.

As before the next step is to specify how complex formulas get assigned their truth value, that is, to specify conditions when formulas $\neg\varphi$, $\varphi \wedge \psi$, and $\varphi \vee \psi$ receive value T, b, n, and F respectively. At this point we commit to the rationale adopted for the strong three-valued interpretation, i.e., that we should assign a classical truth value to a formula, if we can decide which truth value a formula would receive in classical logic. We obtain the following truth tables:[9]

φ	$\neg\varphi$
T	F
b	b
n	n
F	T

$\vee$	T	b	n	F
T	T	T	T	T
b	T	b	?	b
n	T	?	n	n
F	T	b	n	F

$\wedge$	T	b	n	F
T	T	b	n	F
b	b	b	?	F
n	n	?	n	F
F	F	F	F	F

[9]To make our truth tables more compact we represent the truth tables for $\varphi \wedge \psi$ in a slightly different way. The first column on the left provides the possible values of φ, which the first upper row provides the truth values of ψ. The remaining truth values of the table give the relevant truth value of the conjunction relative to the truth values of its conjuncts (and similarly for disjunction).

It remains of course to determine which truth value ought to take the position of the question mark in the respective truth table: what is the truth value of a conjunction for which one conjunct has value b and the other has value n; and, similarly, what is the truth value of a disjunction for which one disjunct has value b and the other has value n.

For conjunction we reason as follows: if one conjunct has value b, this means that it is both true and false. Since one conjunct is false, the conjunction is false. It remains to be checked whether the conjunction is only false or whether is also true, that is, whether the conjunction should have value b. Yet, one conjunct of the conjunction has value n, which tells us that the conjunction cannot be true. A conjunction for which one conjunct has value b and the other has value n, will thus have value F.

For disjunction we reason in a parallel fashion: if one disjunct has value b, this means that it is both true and false. Since one disjunct is true, the disjunction is true. It remains to be checked whether the disjunction is only true or whether it is also false, that is, whether the disjunction should have value b. Yet, one disjunct of the disjunction has value n, which tells us that the disjunction cannot be false. A disjunction for which one disjunct has value b and the other has value n, will thus have value T.

We arrive at the following truth tables for conjunction and disjunction:

$\vee$	T	b	n	F
T	T	T	T	T
b	T	b	T	b
n	T	T	n	n
F	T	b	n	F

$\wedge$	T	b	n	F
T	T	b	n	F
b	b	b	F	F
n	n	F	n	F
F	F	F	F	F

This leads to the following definition of a four-valued interpretation function:

Definition 69 (Four-valued interpretation)**.** *Let V be a four-valued valuation. Then an (four-valued)interpretation I_V is a function that assigns to every formula of the language an element (truth value) of the set $\{\mathsf{T},\mathsf{b},\mathsf{n},\mathsf{F}\}$ relative to V such that:*

(i) if φ is a propositional variable, then $\mathrm{I}_V(\varphi) = V(\varphi)$;

(ii) if φ is $\bot$, then $\mathrm{I}_V(\varphi) = \mathsf{F}$;

(iii) if φ is a formula $\neg\psi$, then

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{F} \\ \mathsf{b}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{b} \\ \mathsf{n}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{n} \\ \mathsf{F}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{T} \end{cases}$$

(iv) if φ is a formula $(\psi \wedge \chi)$, then:

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathrm{I}_V(\chi) = \mathsf{T} \\ \mathsf{b}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{b} \textit{ and } \mathrm{I}_V(\chi) \in \{\mathsf{T},\mathsf{b}\}, \textit{ or } \mathrm{I}_V(\psi) = \mathsf{T} \textit{ and } \mathrm{I}_V(\chi) = \mathsf{b} \\ \mathsf{n}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{n} \textit{ and } \mathrm{I}_V(\chi) \in \{\mathsf{T},\mathsf{n}\}, \textit{ or } \mathrm{I}_V(\psi) = \mathsf{T} \textit{ and } \mathrm{I}_V(\chi) = \mathsf{n} \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

(v) *if φ is a formula $(\psi \vee \chi)$, then:*

$$\mathrm{I}_V(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{T} \textit{ or } \mathrm{I}_V(\chi) = \mathsf{T}, \textit{ or } \mathrm{I}_V(\psi) \neq \mathrm{I}_V(\chi) \in \{\mathsf{n},\mathsf{b}\} \textit{ and } \mathrm{I}_V(\psi) \in \{\mathsf{n},\mathsf{b}\} \\ \mathsf{b}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{b} \textit{ and } \mathrm{I}_V(\chi) \in \{\mathsf{F},\mathsf{b}\}, \textit{ or } \mathrm{I}_V(\psi) = \mathsf{F} \textit{ and } \mathrm{I}_V(\chi) = \mathsf{b} \\ \mathsf{n}, & \textit{if } \mathrm{I}_V(\psi) = \mathsf{n} \textit{ and } \mathrm{I}_V(\chi) \in \{\mathsf{F},\mathsf{n}\}, \textit{ or } \mathrm{I}_V(\psi) = \mathsf{F} \textit{ and } \mathrm{I}_V(\chi) = \mathsf{n} \\ \mathsf{F}, & \textit{otherwise;} \end{cases}$$

We write $V \Vdash_{\mathsf{FDE}}$ iff $\mathrm{I}_V(\varphi) \in \{\mathsf{T},\mathsf{b}\}$.

According to the definition a formula is said to be true in the FDE-model iff the formula has either value Tor value b(and thus similar to the definition of truth in LP-model except that in the latter semantic we were only concerned with three truth-values.

Definition 70 (Logical consequence for FDE)**.** *Let Γ be a set of formulas and φ a formula. Then*

$$\Gamma \vDash_{\mathsf{FDE}} \varphi \textit{ iff for all } V\textit{: if } V \Vdash_{\mathsf{FDE}} \gamma \textit{ for all } \gamma \in \Gamma, \textit{ then } V \Vdash_{\mathsf{FDE}} \varphi.$$

The logic FDE is very similar to the logics KS3. Indeed, let Γ is a set of formula and φ a formula. Then

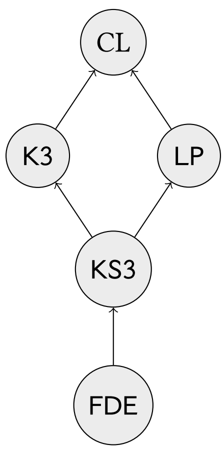
$$\text{if } \Gamma \vDash_{\mathsf{FDE}} \varphi, \text{ then } \Gamma \vDash_{\mathsf{K3}} \varphi.$$

The difference between FDE and KS3 is that the inference

$$\varphi \wedge \neg\varphi \vDash \psi \vee \neg\psi$$

holds for $\vDash_{\mathsf{KS3}}$ but does not hold for all choices of φ and ψ for FDE. For example, let φ be the propositional variable 'p' and ψ the propositional variable 'q'. Choose a valuation V such that $V(p) = \mathsf{b}$ and $V(q) = \mathsf{n}$. Then $V \Vdash_{\mathsf{FDE}} p \wedge \neg p$, but $V \nVdash_{\mathsf{FDE}} q \vee \neg q$ and we obtain $p \wedge \neg p \nVdash_{\mathsf{FDE}} q \vee \neg q$. This valuation is not available in three-valued logics and that is why the inference holds for $\vDash_{\mathsf{KS3}}$.

From FDE **to** CL We can sum up our previous discussion by the following picture:



To put the picture/diagram in words, we get the following relations between the different (strong) non-classical logics:

- if $\Gamma \vDash_{\mathsf{FDE}} \varphi$, then $\Gamma \vDash_{\mathsf{KS3}} \varphi$;
- if $\Gamma \vDash_{\mathsf{KS3}} \varphi$, then $\Gamma \vDash_{\mathsf{K3}} \varphi$;
- if $\Gamma \vDash_{\mathsf{KS3}} \varphi$, then $\Gamma \vDash_{\mathsf{LP}} \varphi$;
- if $\Gamma \vDash_{\mathsf{K3}} \varphi$, then $\Gamma \vDash_{\mathrm{CL}} \varphi$;
- if $\Gamma \vDash_{\mathsf{LP}} \varphi$, then $\Gamma \vDash_{\mathrm{CL}} \varphi$;
- K3 + LP = CL.

8.3 Natural Deduction

In this section we introduce natural deduction systems for the logics FDE, KS3, K3, and LP. We do this by first introducing the natural deduction system for FDE and then show how to extend this system to obtain the natural deduction system for the stronger logics.

For the connective $\wedge$ and $\vee$ the elimination and introduction rules remain unchanged. However, since the de morgan rules are sound for the strong three- and four-valued logics we need to also give introduction and elimination rules for negated conjunctions and negated disjunctions. This is not required in classical logic, as the respective rules/de morgan properties could be established using the rule of proof by contradiction, which is not acceptable for the many-valued logics save LP. We also need to give introduction and elimination rules for negation: these will be double negation introduction (DNI) and elimination (DNE). Instead of PcB we will also adopt suitable introduction and elimination rules for $\bot$.

8.3.1 Basic rules for FDE

Double Negation Introduction

$$\begin{array}{r|ll} m & \varphi & \\ n & \neg\neg\varphi & \text{DNI}, m \end{array}$$

Double Negation Elimination

$$\begin{array}{r|ll} m & \neg\neg\varphi & \\ n & \varphi & \text{DNE}, m \end{array}$$

Conjunction Introduction

$$\begin{array}{r|ll} m & \varphi & \\ n & \psi & \\ j & \varphi \wedge \psi & \wedge\text{I}, m, n \end{array}$$

Conjunction Elimination

$$\begin{array}{r|ll} m & \varphi_1 \wedge \varphi_2 & \\ n & \varphi_i & \wedge\text{E,1, for } i \in \{1, 2\} \end{array}$$

$\neg$ Conjunction Introduction

$\neg$ Conjunction Elimination

$$
\begin{array}{l|ll}
m & \neg\varphi_i & i \in \{1,2\} \\
n & \neg(\varphi_1 \wedge \varphi_2) & \neg\wedge\mathrm{I},\ m
\end{array}
$$

$$
\begin{array}{l|l}
j & \neg(\varphi \wedge \psi) \\
k & \quad \begin{array}{|l} \underline{\neg\varphi} \\ \vdots \\ \chi \end{array} \\
\vdots & \\
m & \\
m+1 & \quad \begin{array}{|l} \underline{\neg\psi} \\ \vdots \\ \chi \end{array} \\
\vdots & \\
n & \\
n+1 & \chi \qquad\qquad \neg\wedge\mathrm{E},\ j,\ k\text{–}m,\ m+1\text{–}n
\end{array}
$$

Disjunction Introduction

$$
\begin{array}{l|ll}
m & \varphi_i & i \in \{1,2\} \\
n & \varphi_1 \vee \varphi_2 & \vee\mathrm{I},\ m
\end{array}
$$

Disjunction Elimination

$$
\begin{array}{l|l}
j & \varphi \vee \psi \\
k & \quad \begin{array}{|l} \underline{\varphi} \\ \vdots \\ \chi \end{array} \\
\vdots & \\
m & \\
m+1 & \quad \begin{array}{|l} \underline{\psi} \\ \vdots \\ \chi \end{array} \\
\vdots & \\
n & \\
n+1 & \chi \qquad\qquad \vee\mathrm{E},\ j,\ k\text{–}m,\ m+1\text{–}n
\end{array}
$$

¬ Disjunction Introduction

$$
\begin{array}{l|ll}
m & \neg\varphi & \\
n & \neg\psi & \\
j & \neg(\varphi \vee \psi) & \neg\wedge\mathrm{I},\ m,\ n
\end{array}
$$

¬ Disjunction Elimination

$$
\begin{array}{l|ll}
m & \neg(\varphi_1 \vee \varphi_2 & \\
n & \neg\varphi_i & \neg\vee\mathrm{E},\ 1, i \in \{1,2\}
\end{array}
$$

⊥-Introduction

$$
\begin{array}{l|ll}
m & \neg\bot & \bot\mathrm{I}
\end{array}
$$

⊥-Elimination

$$
\begin{array}{l|ll}
m & \bot & \\
n & \varphi & \bot\mathrm{E},\ m
\end{array}
$$

The DNI/DNE rules should be self-explanatory and well motivated given the truth table for negation discussed in the previous section. The same should hold ¬∧I and ¬∧E, as well as, ¬∨I and $\neg\vee$E. The rule $\bot$I serves as counterpart to LEM in the non-classical setting. LEM of course is not a sound rule for FDE, however, we may always introduce $\neg\bot$, as even in these non-classical logics we are guaranteed that $\neg\bot$ will always have value T. $\bot$E is a rule of explosion. It allows as infer any formula φ from the $\bot$-constant. Notice that in the present setting this is not the same as inferring an arbitrary formula ψ from $\varphi \wedge \neg\varphi$. The latter inference is most decidedly not a sound inference of FDE.

8.3.2 Further rules for related systems

Excluded Middle

m	$\neg\bot$	
n	$\varphi \vee \neg\varphi$	EM, m

Explosion

k	φ	
m	$\neg\varphi$	
n	$\bot$	EXP, k, m

GG

k	φ
m	$\neg\varphi$
n	$\psi \vee \neg\psi$

Definition 71 (KS3). *The logic* KS3 *is obtained from* FDE *by adding the rule* GG *to the basic rules of* FDE*.*

Definition 72 (K3). *The logic* K3 *is obtained from* FDE *by adding the rule* EXP *to the rules of* FDE*.*

Definition 73 (LP). *The logic* LP *is obtained from* FDE *by adding the rule* EM *to the rules of* FDE*.*

Proposition 74. *We obtain the following relations between the different logics:*

(i) *if* $\Gamma \vdash_{\mathsf{FDE}} \varphi$*, then* $\Gamma \vdash_{\mathsf{KS3}} \varphi$*;*

(ii) *if* $\Gamma \vdash_{\mathsf{KS3}} \varphi$*, then* $\Gamma \vdash_{\mathsf{K3}} \varphi$ *and* $\Gamma \vdash_{\mathsf{LP}} \varphi$*;*

(iii) *if* $\Gamma \vdash_{\mathsf{K3}} \varphi$*, then* $\Gamma \vdash_{\mathsf{CL}} \varphi$*;*

(iv) *if* $\Gamma \vdash_{\mathsf{LP}} \varphi$*, then* $\Gamma \vdash_{\mathsf{CL}} \varphi$*;*

(v) CL = K3 + EM = LP + EXP

Proof. We take up each item individually:
immediate.

(ii) We need to show that GG is a derivable rule in both K3 and LP

- K3

$$\begin{array}{l|l l}
k & \varphi & \\
m & \neg\varphi & \\
m+1 & \bot & \mathrm{EXP},\ k,\ m \\
m+2 & \psi \lor \neg\psi & \bot\mathrm{E},\ m+1
\end{array}$$

- LP

$$\begin{array}{l|l l}
k & \varphi & \\
m & \neg\varphi & \\
m+1 & \neg\bot & \bot\mathrm{I} \\
m+2 & \psi \lor \neg\psi & \mathrm{EM},\ m+1
\end{array}$$

(iii) Immediate

(iv) Immediate

(v) We show that in FDE + EM conditional introduction is a derivable rule and that in FDE + EXP conditional elimination is a derivable rule:

- FDE + EM. We assume that there is an FDE-derivation $\varphi \vdash_{\mathsf{FDE}} \psi$.

$$\begin{array}{l|l l}
m & \varphi \lor \neg\varphi & \bot\mathrm{I} + \mathrm{EM} \\
m+1 & \quad \begin{array}{|l} \varphi \\ \hline \end{array} & \\
\vdots & \quad \begin{array}{|l} \vdots \end{array} & \\
n & \quad \begin{array}{|l} \psi \end{array} & \\
n+1 & \quad \begin{array}{|l} \neg\varphi \lor \psi \end{array} & \lor\mathrm{I},\ n \\
n+2 & \quad \begin{array}{|l} \neg\varphi \\ \hline \end{array} & \\
n+3 & \quad \begin{array}{|l} \neg\varphi \lor \psi \end{array} & \lor\mathrm{I},\ n+2 \\
n+4 & \underbrace{\neg\varphi \lor \psi}_{\varphi \to \psi} & \lor\mathrm{E},\ m,\ m+1\text{–}n+1,\ n+2,\ n+3
\end{array}$$

- FDE + EXP

k	$\neg\varphi \vee \psi$ ($\equiv_{\mathsf{def}}$ $\varphi \longrightarrow \psi$)	
m	φ	
$m+1$	$\quad \mid \neg\varphi$	
$m+2$	$\quad \mid \bot$	EXP, m, $m+1$
$m+3$	$\quad \mid \psi$	$\bot$E, $m+2$
$m+4$	$\quad \mid \psi$	
$m+5$	$\quad \mid \psi$	R, $m+4$
$m+6$	ψ	$\vee$E, k, $m+1$–$m+3$, $m+4$–$m+5$

It remains to be shown that PbC is a derivable rule in FDE + EM + EXP:

l	$\quad \mid \varphi \longrightarrow \bot$	
$\vdots$	$\quad \mid \vdots$	
k	$\quad \mid \bot$	
$k+1$	$(\varphi \longrightarrow \bot) \longrightarrow \bot$	
m	$\neg(\varphi \longrightarrow \bot) \vee \bot$	Def, $k+1$
$m+1$	$\quad \mid \neg(\neg\varphi \vee \bot)$	
$m+2$	$\quad \mid \neg\neg\varphi$	$\neg\vee$E, $m+1$
$m+3$	$\quad \mid \varphi$	DNE, $m+2$
$m+4$	$\quad \mid \bot$	
$m+5$	$\quad \mid \varphi$	$\bot$E, $m+4$
$m+6$	φ	$\vee$E, m, $m+1$–$m+3$, $m+4$–$m+5$

□

Exercise 75. *Establish the inferences displayed in Items 1-12 of Example 60 by proving them in the natural deduction system for* FDE.

Proposition 76 (Adequacy Many-valued Logic)**.** *Let* Γ *be a set of formula and* φ *a formula. Then*

$$\Gamma \vDash_{\mathsf{FDE}} \varphi \textit{ iff } \Gamma \vdash_{\mathsf{FDE}} \varphi;$$
$$\Gamma \vDash_{\mathsf{KS3}} \varphi \textit{ iff } \Gamma \vdash_{\mathsf{KS3}} \varphi;$$

$$\Gamma \vDash_{\mathsf{K3}} \varphi \textit{ iff } \Gamma \vdash_{\mathsf{K3}} \varphi;$$
$$\Gamma \vDash_{\mathsf{LP}} \varphi \textit{ iff } \Gamma \vdash_{\mathsf{LP}} \varphi.$$

Part IV

First-order Logic

Chapter 9

First-order Logic

In propositional logic propositional variables are the basic building block of sentences/formulas. Intuitively they stand for declarative sentences of English. However, we frequently wish to analyse the subsentential structure of sentences, as opposed to the sentential structure that tells us how to compose new sentences out of existing sentences. This leads to the introduction of individual constants (names), predicate constants, individual variables, and quantifiers to the language.

9.1 Syntax

The language of first-order logic (predicate logic) consists of the following vocabulary:

- a countable set of individual variables:

  $$\mathsf{Var} := \{x_0, x_1, x_2, x_3, \ldots\}$$

  (for ease of readablility we use x, y, and z—possibly with indices).

- a countable set of individual constants (names):

  $$\mathsf{Const} := \{c_0, c_1, c_2, c_3, \ldots\}$$

  (in practice we use small caps letters $a, b, \ldots, r$ for names).

- a countable set of predicate constants of arbitrary finite *arity*, i.e., n, m, k, and l below are natural numbers:

  $$\mathsf{Pred} := \{P_0^n, P_1^m, P_2^k, P_3^l, \ldots\}$$

  (in practice we often use capital letters $P, Q, R \ldots$ for predicate constants).

- the propositional constant '$\bot$';

- the logical connectives '$\wedge$','$\vee$,' and '$\rightarrow$';

- the universal quantifier '$\forall$';

- auxiliary symbols (brackets: '(','))'

The universal quantifier '$\forall$' is read as '*everything*' or '*for all*'.

The next step is to define what the well-formed formulas of the language are. In propositional logic we used the 'formula' and 'sentence' somewhat interchangeably. In first-order logic we have to be more careful: a sentence will be a formula with no free variable. To make the definition of a sentence precise we need to say what the free variables of a formula are. Indeed this will be defined in tandem with the notion of a well-formed formula of first-order logic.

First, however we introduce the notion of a term:

Definition 77 (Term)**.** *A* **term** *of first-order logic is either an individual variable of an individual constant. Terms are denoted by $s, t, u, \ldots$ (possible with indices). Let t be a term then the* **set of free variables** *of t ($\mathsf{FV}(t)$) is defined as follows:*

$$\mathsf{FV}(t) := \begin{cases} \{t\}, & \textit{if } t \in \mathsf{Var} \\ \emptyset, & \textit{otherwise.} \end{cases}$$

Definition 78 (Formula,Sentence)**.** *Well-formed formulas of the language of first-order logic are defined in tandem with the set of free variables of a given formula:*

1. *If P^k is a predicate constant with arity k and $t_1, \ldots, t_k$ are first-order terms, then $P^k t_1, \ldots, t_k$ is a formula and $\mathsf{FV}(P^k t_1, \ldots, t_k) = \mathsf{FV}(t_1) \cup \ldots \cup \mathsf{FV}(t_1)$;*

2. *$\perp$ is a formula and $\mathsf{FV}(\perp) = \emptyset$;*

3. *if φ and ψ are formulas, then $\varphi \wedge \psi$, $\varphi \vee \psi$, and $\varphi \longrightarrow \psi$ are formulas, and*
$$\mathsf{FV}(\varphi \wedge \psi) = \mathsf{FV}(\varphi \vee \psi) = \mathsf{FV}(\varphi \longrightarrow \psi) = \mathsf{FV}(\varphi) \cup \mathsf{FV}(\psi);$$

4. *if φ is a formula and $x \in \mathsf{Var}$, then $\forall x \varphi$ is a formula and $\mathsf{FV}(\forall x \varphi) = \mathsf{FV}(\varphi) - \{x\}$.*[1]

5. *Nothing else is a formula.*

A formula φ is called a **sentence** *iff it has no free variables, i.e., iff $\mathsf{FV}(\varphi) = \emptyset$.*

Defined Symbols On top of the primitive symbols of the language we also have defined symbols that are introduced via the following definitions:

$$\neg\varphi \equiv_{\mathsf{def}} \varphi \longrightarrow \perp$$
$$\varphi \leftrightarrow \psi \equiv_{\mathsf{def}} (\varphi \longrightarrow \psi) \wedge (\psi \longrightarrow \varphi)$$
$$\exists x \varphi \equiv_{\mathsf{def}} \neg\forall x \neg\varphi.$$

Whilst '$\forall$' is read as '*everything*' or '*for all*' the existential quantifier '$\exists$' will be read as '*there is*', '*there exist*' or '*something*'.

---

[1]Let S and S_1 be sets. Then
$$S - S_1 := \{o \in S \mid o \notin S_1\}.$$

9.2 Semantics

In the semantics for propositional logic it was sufficient to assign truth values to the propositional variables via a valuation function, and then give rules how to assign truth values to all formulas of the language relative to a given valuation. In the semantics for first-order language we do not assign truth values to simple sentences using a valuation. Rather we assign a **denotation** to the individual constants and predicate constants of the language, which in turn will allow us to determine whether certain simple sentences are true. The denotation of an individual constant will be an object of the **domain** whereas the denotation of an n-place predicate constant will be an n-ary relation on the domain. A domain together with a denotation function constitute a **model** of first-order logic.[2]

Definition 79 (Model)**.** *A model M of first-order logic is a pair (D, J) such that $D \neq \emptyset$ is a non-empty domain (set) and J is a denotation function such that:*

(i) $J(c) \in D$*, for all individual constants (names) c;*

(ii) $J(P^k) \subseteq \underbrace{D \times \ldots \times D}_{k\text{-times}}$*, for every predicate P of arbitrary arity k.*

When is a sentence true in a model (D, J)? Consider the following sentences together with a proposed formalization:

1. Mary is happy—Pm.
2. Mary likes Catrin—Qmc.
3. Mary is moving from Bristol to London—$Rmbl$.

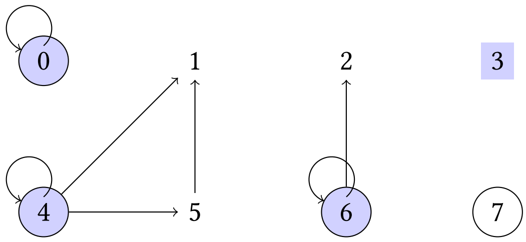
Then, intuitively, we would like to say that 1 is true iff $J(m) \in J(P)$, i.e., iff the Mary is amongst the things of the domain that are happy; 2 is true iff $\langle J(m), J(c)\rangle \in J(Q)$, i.e., iff the liking relation holds between Mary and Catrin; and analogously 3 is true iff $\langle J(m), J(b), J(c)\rangle \in J(R)$ (notice the latter requires the domain to consist of people and places/cities).

Example 80. *A model (D, J) for a first-order language with three predicate constants (B^1, C^1, P^2) and two individual constant (a, b)*

- $D = \{0, 1, 2, 3, 4, 5, 6, 7\}$
- $J(a) = 4$; $J(b) = 5$
- $J(B) = \{0, 3, 4, 6\}$ *(is shaded blue in the diagram)*
- $J(C) = \{0, 4, 6, 7\}$ *(has a circle around it in the diagram)*

[2]'Denotation' functions are frequently referred to as 'interpretations' in the literature. Since I used the word 'Interpretation' differently in the previous sections, that is, as a function that assigns truth values to all formulas of the language I prefer to speak of denotation functions.

- $J(P) = \{\langle 0,0\rangle, \langle 4,4\rangle, \langle 6,6\rangle, \langle 4,1\rangle, \langle 4,5\rangle, \langle 5,1\rangle, \langle 6,2\rangle\}$ *(there is an arrow from __ to __ in the diagram)*



Exercise 81. *Intuitively, are Paa, Pab, Pba, Ba, Bb, Ca, and Cb true in the above model?*

The denotation function provides us with an interpretation/denotation of the individual constants and the predicate constants, but does not tell us how to handle variables. Variables can be seen as holes of a sentence that need to be filled before we can say whether the sentence is true or false. For example, before we can say whether the somewhat incomplete sentence

$$\_\__x \text{ is a great philosophy unit.}$$

is true or false, we need to specify how to fill the x-gap in the sentence, that is, we need to say whether the gap is supposed to be filled by the *Philosophical Logic* unit or, more plausibly, by the *Philosophy of Language* unit (or some other unit altogether). Similarly, we can only say whether a formula, say Px, is true once we have specified which objects of the domain the variables occurring in the formula stand for, e.g., which object of the domain the variable x in the formula Px stands for. That is precisely the role of so-called variable assignments. Variable assignments are functions that assign to every variable of the language an object of the domain of the model.

Definition 82 (Assignment Function, x-variant)**.** *Let (D, J) be a model. An assignment function $\beta : \mathsf{Var} \longrightarrow D$ assigns an object of the domain D of the model to every variable of the language. If β is an assignment function, then $\beta(x : d)$ is an assignment function such that for all $y \in \mathsf{Var}$:*

$$\beta(x : d)(y) := \begin{cases} d, & \text{if } x \doteq y; \\ \beta(y), & \text{otherwise.} \end{cases}$$

$\beta(x : d)$ is called an x-variant of β.

An x-variant of an assignment function β thus assigns exactly the same objects to the variables of the language as β except, possibly, for the variable x.

Definition 83 (Term denotation)**.** *Let t be a term, J a denotation function and β an assignment function. Then the denotation of the term t is defined as follows:*

$$t^{J,\beta} := \begin{cases} J(t), & \text{if } t \in \mathsf{Const} \\ \beta(t), & \text{if } t \in \mathsf{Var}. \end{cases}$$

We can now assign truth values to formulas of the language in a model relative to a variable assignment.

Definition 84 (Truth). *Let (D, J) be a first-order model and β an assignment function. Then $\mathrm{I}^{\beta}_{(D,J)}$ is a function that assigns truth values to all formulas of the language relative the (D, J) and β such that:*

1. *If φ is a formula $P^k t_1, \dots, t_k$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } \langle t_1^{J,\beta}, \dots, t_k^{J,\beta} \rangle \in J(P^k); \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

2. *If φ is $\perp$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \mathsf{F}.$$

3. *If φ is a formula $\psi \wedge \chi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } I^{\beta}_{(D,J)}(\psi) = I^{\beta}_{(D,J)}(\chi) = \mathsf{T}; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

4. *If φ is a formula $\psi \vee \chi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } I^{\beta}_{(D,J)}(\psi) = \mathsf{T} \text{ or } I^{\beta}_{(D,J)}(\chi) = \mathsf{T}; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

5. *If φ is a formula $\psi \longrightarrow \chi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } I^{\beta}_{(D,J)}(\psi) = \mathsf{F} \text{ or } I^{\beta}_{(D,J)}(\chi) = \mathsf{T}; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

6. *If φ is a formula $\forall x \psi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } I^{\beta(x:d)}_{(D,J)}(\psi) = \mathsf{T} \text{ for all } d \in D; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

We write $(D, J) \Vdash \varphi[\beta]$ iff $\mathrm{I}^{\beta}_{(D,J)}(\varphi) = \mathsf{T}$. A formula φ is true in the model (D, J) iff $(D, J) \Vdash \varphi[\beta]$ for every variable assignment β. In this case we simply write $(D, J) \Vdash \varphi$.

Interestingly, the definition has the consequence that the following two claims are equivalent over a model M if φ is a sentence, i.e., if φ has no free variables (exercise: why?):

(I) There exist a variable assignment β such that $M \Vdash \varphi[\beta]$.

(II) $M \Vdash \varphi[\beta]$ for all variable assignments β.

The equivalence of (I) and (II) highlights that it is irrelevant which objects are assigned to particular variables in the case of sentences, that is, a particular choice of assignment function is only relevant if we are considering formulas that have free variables.

A further consequence of the definition is the following clause for the existential quantifier:

$$\mathrm{I}^{\beta}_{D,J}(\exists x \psi) := \begin{cases} \mathsf{T}, & \text{if } \mathrm{I}^{\beta(x:d)}_{(D,J)}(\psi) = \mathsf{T} \text{ for some } d \in D; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

Exercise 85. *Let M be the model given in Example 80. Specify one assignment function such that following formulas are true, and one assignment function such that the formulas are false:*

1. $Bx \wedge Pxy \longrightarrow Cy$
2. $Bx \wedge \neg Cx \longrightarrow Pyx$
3. $Bx \longrightarrow \forall y(Pxy \longrightarrow By)$
4. $\forall x(Pxy \longrightarrow Bx \wedge Cx) \longrightarrow \neg By \wedge \neg Cy$
5. $Bx \longrightarrow \neg Cx$

For example, 5 is true for the assignment β with $\beta(x) = 3$, but false for the assignment β' such that $\beta'(x) = 4$.

Definition 86 (Logical consequence). *Let Γ be a set of formulas and φ a formula. Then $\Gamma \vDash \varphi$ iff for all first order models M and assignment functions β:*

$$\text{if } M \Vdash \gamma[\beta] \text{ for all } \gamma \in \Gamma, \text{ then } M \Vdash \varphi[\beta].$$[3]

Some logical consequence of FOL:

1. $\forall x Px \vDash Pc$
2. $\vDash \forall x \forall Qxy \longrightarrow \forall y Qcy$
3. $\forall x Px \vDash \forall y Py$
4. $\forall x Px \vDash \neg \exists x \neg Px$
5. $\forall x Qxx \vDash Qcc$
6. $Qcc \vDash \exists x Qxc$
7. $\vDash \forall x(Px \longrightarrow Rx) \longrightarrow (\forall x Px \longrightarrow \forall x Rx)$

[3] If Γ and φ are sentences, quantification over variable assignment can be omitted.

8. $\vDash \forall x(Px \wedge Wx) \leftrightarrow \forall xPx \wedge \forall xWx$

9. $\exists x(Px \wedge Rx) \vDash \exists xPx \wedge \exists xRx$

10. $\vDash \exists x(Px \vee Rx) \leftrightarrow \exists yPy \vee \exists yRy$

Proposition 87 (Deduction Theorem)**.** *Let* Γ *be a set of and* φ, ψ *FOL-formulas. Then*

$$\Gamma \cup \{\varphi\} \vDash \psi \textit{ iff } \Gamma \vDash \varphi \longrightarrow \psi.$$

9.3 Natural Deduction for FOL

9.3.1 Preliminaries: Substitutions

In order to spell out the quantifier introduction and elimination rules for first-order logic we need to talk about substitutions of a term for another term of in a formula. For example, in 1 the 'Pc' is obtained from '$\forall xPx$' by removing the quantifier '$\forall x$' and by substituting every occurrence of the variable 'x' in the formula 'Px' by the constant 'c'. To say that this type of inference holds not only for the specific formulas occurring in 1, but for any type of formula we need to say more precisely of what it means to substitute a term for another term in a formula. Indeed we distinguish between two different types of substitutions:

Let φ be a formula and s, t terms. Then

- $\varphi(s/t)$ denotes the formula in which **every** free occurrence of t is replaced by s; This can be defined inductively as follows:
  - $t_1(s/t) := \begin{cases} s, & \text{if } t_1 \textbf{ is the term } t; \\ t_1, & \text{otherwise.} \end{cases}$
  - $\varphi(s/t) := \begin{cases} Pt_1(s/t), \dots, t_n(s/t), & \text{if } \varphi \doteq Pt_1, \dots, t_n; \\ \bot, & \text{if } \varphi \doteq \bot; \\ \psi_1(s/t) \wedge \psi_2(s/t), & \text{if } \varphi \doteq \psi_1 \wedge \psi_2; \\ \psi_1(s/t) \vee \psi_2(s/t), & \text{if } \varphi \doteq \psi_1 \vee \psi_2; \\ \psi_1(s/t) \longrightarrow \psi_2(s/t), & \text{if } \varphi \doteq \psi_1 \longrightarrow \psi_2; \\ \forall v[\psi(s/t)], & \text{if } \varphi \doteq \forall v\psi. \end{cases}$
- $\varphi[s/t]$ denotes a formula in which **at least one** free occurrence of t—but not necessary all occurrences—are replaced by s.[4]

Going back to our examples of logical consequences in FOL, we can see the two types of substitutions at play. For example, 5 can be written as

$$\forall xQxx \vDash Qxx(c/x),$$

[4] The second type of substitution is, strictly speaking, not an operation on formulas because it does not yield a unique output for any given formula. For example, $Pss[t/s]$ can be either the formula Pts, the formula Pst, or the formula Ptt. This problem will not concern us here.

while the other type of substitution is at play in 6:

$$Qcc \vDash \exists x \underbrace{Qxc}_{Qcc[x/c]} .$$

These two examples are instances of the rules of ∀-elimination and ∃-introduction respectively. With respective to the rule of ∃-introduction a further complication arises: suppose we are presented with the formula $\forall x(Px \wedge Rc)$ and want to use the rule of ∃-introduction. Then we need to be careful to choose a variable different to x for otherwise we could infer $\exists x \forall x(Px \wedge Qx)$, which is equivalent to $\forall x(Px \wedge Qx)$, which we should not be able to infer from $\forall x(Px \wedge Rc)$. To avoid this problem we need to make sure that the variable we substitute for c will not be bound by a quantifier.

Definition 88 (Free for)**.** *Let φ be a sentence and c an individual constant. Then a variable x* **is free for** *c in φ iff c does not occur in the scope of a quantifier binding x in φ.*

Example 89. *Let us look at the following examples:*

(i) $\forall y(Py \wedge Rc)$

(ii) $\forall yPy \wedge Rc$

(iii) $\exists yQyc \wedge Pc$

(iv) $\forall z(\exists yQyc_1 \wedge Rc_2)$

In (i) c is not free for y, but free for all other variables. In (ii) c is free for all variables (including y). In (iii) c is not free for y, but free for all other variables. In (iv) c_1 is not free for both y and z (but all other variables) while c_2 is not free for z but free for all other variables (including y).

9.3.2 Rules of Natural Deduction

The natural deduction system for FOL extends the natural deduction system for classical propositional logic by introduction and elimination rules for the universal quantifier '∀' (recall that officially the existential quantifier '∃' is a defined symbol). However, it is important to notice that **derivations only consist of sentences of FOL**, that is, no formula with free variables is allowed to occur in a FOL-derivation. The natural deduction system of *FOL* consists of the rules of classical propositional logic together with:

∀-**Introduction rule**

m	$\varphi(c)$	
$m+1$	$\forall x\varphi(x/c)$	∀I, m

provided x is free for c in φ and c does not occur in any premise or undischarged assumption.

∀-**Elimination rule**

m	$\forall x\varphi$	
$m+1$	$\varphi(c/x)$	∀E, m

The condition on c in the ∀-introduction rule guarantee that no special assumption are made with respect to c, i.e., that the use of c is generic (in the sense of let n be a number…). So the ∀-introduction rule tells us that if we have derived $\varphi(c)$ for a constant c without introducing any specific assumption with respect to c, then we may conclude that $\varphi(x)$ must be true of every object of the domain. The ∀-elimination rule is almost self-explanatory: if something holds for all objects of the domain then it must specifically hold of the object denoted by c.

Even though the existential quantifier is only a defined symbol we will include basic rules for ∃ to our natural deduction system. However, strictly speaking, these rules are derived rules that can be obtained from the introduction and elimination rules for ∀ together with the definition of the existential quantifier.

∃-Introduction rule

$$
\begin{array}{l|ll}
m & \varphi(c) & \\
m+1 & \exists x \varphi[x/c] & \exists\mathrm{I},\ m
\end{array}
$$

provided x is free for c in φ and where $\varphi[x/c]$ denotes a formula that results from φ by substituting one or more occurrences of c by x.

∃-Elimination rule

$$
\begin{array}{l|lll}
k & \exists x \varphi & & \\
m & & \begin{array}{|l} \varphi(c/x) \\ \hline \end{array} & \\
n & & \begin{array}{|l} \chi \end{array} & \\
n+1 & \chi & & \exists\mathrm{E},\ k,\ m\text{–}n
\end{array}
$$

c is new to the derivation and does not occur in χ.

The ∃-elimination rule may be confusing at first sight. It is arguably best understood in analogy with the rule of disjunction elimination: disjunction elimination tells us that if χ follows from either disjunct then χ follows from the disjunction alone. Now if we assume a formula $\exists x \varphi$ is true, then we know that $\phi(x)$ must be true for at least on object of the domain. In some sense $\exists x \varphi$ is a disjunction with as many disjuncts as there are members of the domain: $\phi(x)$ holds of the first object of the domain or it holds of the second object of the domain or …
The spirit of ∃-elimination is to say that if a formula χ follows for every of these aforementioned disjuncts, then we may infer χ from the formula $\exists x \chi$. In practice, this means that we pick a constant c about which we have no prior information (it is new to the derivation). Then, if we can derive χ (where c does not occur in χ) from $\varphi(c)$, we may discharge the assumption $\varphi(c)$ and conclude that χ can be derived from $\exists x \varphi$.

We now show that both ∃-introduction and ∃-elimination are derived rules. We start with the former and assume that in a derivation we obtain a sentence $\varphi(c)$.

$$
\begin{array}{lll}
m & \varphi(c) & \\
m+1 & \quad \mid \neg\exists x\varphi[x/c] & \\
m+2 & \quad \mid \neg\neg\forall x\neg\varphi[x/c] & \text{definition of } \exists,\ m+1 \\
m+3 & \quad \mid \forall x\neg\varphi[x/c] & \text{TFL, } m+2 \\
m+4 & \quad \mid \neg\underbrace{(\varphi[x/c])(c/x)}_{\varphi(c)} & \forall\text{E, } m+3 \\
m+5 & \quad \mid \bot & \rightarrow\text{E, } m,\ m+4 \\
m+6 & \exists x\varphi[x/c] & \text{PbC, } m+1\text{–}m+5
\end{array}
$$

Throughout we have assumed that x is free for c in φ.

Let's turn to the $\exists$-elimination: we assume that in a derivation we obtain a sentence $\exists x\varphi$ and that together with $\varphi(c/x)$ for some c on can derive χ.

$$
\begin{array}{lll}
m & \exists x\varphi & \\
m+1 & \neg\forall x\neg\varphi & \text{definition, } m \\
m+2 & \quad \mid \neg\chi & \\
m+3 & \quad \mid \quad \mid \varphi(c/x) & \\
m+4 & \quad \mid \quad \mid \chi & \text{assumption, } m,\ m+3 \\
m+5 & \quad \mid \quad \mid \bot & \rightarrow\text{E, } m+2,\ m+4 \\
m+6 & \quad \mid \neg\varphi(c/x) & \rightarrow\text{I, } m+3\text{–}m+5 \\
m+7 & \quad \mid \forall x\,\underbrace{\neg\varphi(c/x)(x/c)}_{\neg\varphi} & \forall\text{I, } m+6 \\
m+8 & \quad \mid \bot & \rightarrow\text{E, } m+1,\ m+7 \\
m+9 & \chi & \text{PbC, } m+2\text{–}m+8
\end{array}
$$

In Section 9.2 we required the domain of a model to be non-empty. Similarly, using the natural deduction system for FOL we can now show that there must be at least one object:

$$
\begin{array}{lll}
1 & \quad \mid Pc & \\
2 & \quad \mid Pc & \text{R, } 1 \\
3 & Pc \rightarrow Pc & \rightarrow\text{I, } 1\text{–}2 \\
4 & \exists x(Px \rightarrow Px) & \exists\text{I, } 3
\end{array}
$$

This shows that first-order logic is not fully ontologically neutral; it postulates the existence of at least one object.

Exercise 90. *Establish the following claims using the natural deduction system of FOL:*

1. $\forall x Px \vdash Pc$
2. $\vdash \forall x \forall y Qxy \longrightarrow \forall y Qcy$
3. $\forall x Px \vdash \forall y Py$
4. $\forall x Px \vdash \neg\exists x Px$
5. $\forall x Qxx \vdash Qcc$
6. $Qcc \vdash \exists x Qxc$
7. $\vdash \forall x(Px \longrightarrow Rx) \longrightarrow (\forall x Px \longrightarrow \forall x Rx)$
8. $\vdash \forall x(Px \wedge Wx) \leftrightarrow \forall x Px \wedge \forall x Wx$
9. $\exists x(Px \wedge Rx) \vdash \exists x Px \wedge \exists x Rx$
10. $\vdash \exists x(Px \vee Rx) \leftrightarrow \exists y Py \vee \exists y Ry$

Chapter 10

Adding Identity

So far, our language does not have a designated symbol that allows us to express identity, i.e., we have no designated way to say that two terms t_1 and t_2 denote the same object of the domain: $t_1 = t_2$ is not formula of our language. Of course, a two-place predicate constant P *can* be interpreted as identity in some model (D, J), i.e., $J(P) = \{\langle d, d\rangle \mid d \in D\}$, but what we are ultimately after is an specific symbol (predicate) that gets interpreted as identity in *all* models. Our motivation is that inferences such as

1	There is at most one cat that is black.
2	There are at least two cats
3	There is at least one cat that is not black.

seem to be logical inferences and not inferences that depend on the way we understand 'cat' and 'black', i.e., the inference should be sound independent of the particular interpretation of the predicates 'cat' and 'black'.[1]

In this chapter we extend the language of FOL by the identity symbol and extend the semantics and the system of natural deduction accordingly. The language of FOL with identity extends the language of FOL by the symbol '='. We need to add one clause to Definition 78, that is, the definition which specifies which expressions are the well-formed formulas of the language:

- if s and t are first-order terms, then $s = t$ is a formula and $\mathsf{FV}(s = t) = \mathsf{FV}(s) \cup \mathsf{FV}(t)$.

As a convention and for reasons of readability we shall write $\forall x(s = t)$ instead of the correct $\forall x s = t$ and similarly for the existential quantifier (formulas such as $\forall x\, x = s$ or $\exists x x = s$ become difficult to read without paranthesis...)

As for the semantics of FOL everything remains the same except that we have to add one further clause to Definition 84, which tells us when a formula of the form $s = t$ receives value T (F):

[1]The example is from p. 27 of MacFarlane: Philosophical Logic

- If φ is a formula $s = t$, then

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } s^{J,\beta} = t^{J,\beta} \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

Strictly speaking we should of course have distinguished between the identity symbol of the language of FOL with identity and the identity relation we employ in stating the truth conditions of formulas of the form ‘s=t’. The truth conditions tell us that a formula of the form ‘s=t’ is true iff ‘s’ and ‘t’ denote the same object.

With the semantics in place one can check that the following formulas are logical truths of FOL with identity for all terms s and t:[2]

(i) $s = s$

(ii) $s = t \longrightarrow t = s$

(iii) $s = t_1 \wedge t_1 = t_2 \longrightarrow s = t_2$

(iv) $s = t \wedge \varphi(s/x) \longrightarrow \varphi(t/x)$

(i)-(iii) state that identity is reflexive, symmetric and transitive. (iv) is the so-called substitution principle: it tells us that if ‘s’ and ‘t’ denote the same object, then we can substitute ‘s’ and ‘t’ in any context, i.e., formula without altering the truth value. This principle is uncontroversial for extensional languages such as the language of FOL with identity, but somewhat controversial for intensional languages such as the language of first-order modal logic (with identity).[3] (Exercise: why?)

(i) and (iv) also motivate the rules of natural deduction that are required for handling identity:

=-Introduction rule

$$\begin{array}{r|ll} m & c = c & = \mathrm{I} \end{array}$$

.

=-Elimination rule

$$\begin{array}{r|ll} k & c_1 = c_2 & \\ m & \varphi & \\ n & \varphi[c_1/c_2] & = \mathrm{E},\, k,\, m \end{array}$$

$\varphi[c_1/c_2]$ denotes a formula that results from φ by substituting one or more occurrences of c_2 by c_1.

Identity is important as it allows us to formalize claims such as:

[2]Strictly speking, these formulas are logical truths only if they are sentences, if they have no free variables. However, even if the have free variables they will be true in all models under every variable assignment.

[3]One obtains the language of first-order modal logic by extending the vocabulary of the language of FOL with identity by the modal operator ‘□’

- there is at most one thing such that φ:

$$\forall x(\varphi(x) \longrightarrow \forall y(\varphi(y) \longrightarrow x = y)).$$

- there is a unique thing such φ:

$$\exists x(\varphi(x) \wedge \forall y(\varphi(y) \longrightarrow x = y)).$$

- there are at least two things such that φ:

$$\exists x \exists y(\varphi(x) \wedge \varphi(y) \wedge \neg x = y).^4$$

- there are at least n things such that φ

$$\exists x_1 \dots \exists x_n(\varphi(x_1) \wedge \dots \varphi(x_n) \wedge \neg x_1 = x_2 \wedge \dots \wedge \neg x_1 = x_n \wedge \dots \wedge \neg x_{n-1} = x_n).$$

- there are at most n things such that φ—exercise
- there are exactly n things such that φ—exercise

Such claims, which play an important role in mathematical, scientific, and ordinary discourse, could not be expressed without the identity symbol in the language. In particular we can now establish the inference we used to motivate adding a designated identity symbol to the language in our natural deduction system. To this effect let 'C' be the predicate 'is a cat' and 'B the predicate is 'is black'. We want to show:

$$\exists x(Cx \wedge Bx \wedge \forall y(Cy \wedge By \longrightarrow x = y)), \exists x \exists y(Cx \wedge Cy \wedge \neg x = y) \vdash \exists x(Cx \wedge \neg Bx).$$

[4]Strictly speaking in order to respect the formation rules for FOL-formulas we should write

$$\exists x \exists y((\varphi(x) \wedge \varphi(y)) \wedge \neg x = y)$$

or

$$\exists x \exists y(\varphi(x) \wedge (\varphi(y) \wedge \neg x = y)).$$

However, since both formulas are logically equivalent and, more generally, conjunction is associative, we omit the brackets in this case and similar cases below.

1	$\vert\ \exists x(Cx \wedge Bx \wedge \forall y(Cy \wedge By \longrightarrow x = y))$	
2	$\vert\ \exists x \exists y(Cx \wedge Cy \wedge \neg x = y)$	
3	$\vert\ \vert\ \neg\exists x(Cx \wedge \neg Bx)$	
4	$\vert\ \vert\ \neg\neg\forall x \neg(Cx \wedge \neg Bx)$	def. ∃, 3
5	$\vert\ \vert\ \forall x \neg(Cx \wedge \neg Bx)$	DNE, 4
6	$\vert\ \vert\ \forall x(Cx \longrightarrow Bx)$	from 5: Exercise
7	$\vert\ \vert\ \vert\ Ca \wedge Ba \wedge \forall y(Cy \wedge By \longrightarrow a = y)$	
8	$\vert\ \vert\ \vert\ \forall y(Cy \wedge By \longrightarrow a = y)$	∧E, 7
9	$\vert\ \vert\ \vert\ \vert\ \exists y(Cb \wedge Cy \wedge \neg b = y)$	
10	$\vert\ \vert\ \vert\ \vert\ \vert\ Cb \wedge Cc \wedge \neg b = c$	
11	$\vert\ \vert\ \vert\ \vert\ \vert\ Cb \longrightarrow Bb$	∀E, 6
12	$\vert\ \vert\ \vert\ \vert\ \vert\ Cc \longrightarrow Bc$	∀E, 6
⋮	⋮	
13	$\vert\ \vert\ \vert\ \vert\ \vert\ Cb \wedge Bb$	
14	$\vert\ \vert\ \vert\ \vert\ \vert\ Cc \wedge Bc$	
15	$\vert\ \vert\ \vert\ \vert\ \vert\ Cb \wedge Bb \longrightarrow a = b$	∀E, 8
16	$\vert\ \vert\ \vert\ \vert\ \vert\ a = b$	⟶E, 13, 15
⋮	⋮	
17	$\vert\ \vert\ \vert\ \vert\ \vert\ a = c$	
18	$\vert\ \vert\ \vert\ \vert\ \vert\ b = c$	=E, 16, 17
19	$\vert\ \vert\ \vert\ \vert\ \vert\ \neg b = c$	∧E, 10
20	$\vert\ \vert\ \vert\ \vert\ \vert\ \bot$	⟶E, 18, 19
21	$\vert\ \vert\ \vert\ \vert\ \bot$	∃E, 10–20
22	$\vert\ \vert\ \vert\ \bot$	∃E, 9–21
23	$\vert\ \vert\ \bot$	∃E, 7–22
24	$\vert\ \exists x(Cx \wedge \neg Bx)$	PbC, 3–23

10.1 Adequacy, Compactness, and Expressive Limitations

The natural deduction for first-order logic with identity ($\text{FOL}_=$) is adequate with respect to our semantics.

Theorem 91 ($\text{FOL}_=$-Adequacy). *Let Σ be a set of $\text{FOL}_=$-sentences and φ a $\text{FOL}_=$-sentence. Then*

$$\Sigma \vdash \varphi \textit{ iff } \Sigma \models \varphi$$

The Adequacy theorem for $\text{FOL}_=$ has an interesting consequence which is know as Compactness theorem:

Theorem 92 (Compactness). *Let Σ be a set of $\text{FOL}_=$-sentences. Then*

$$\Sigma \textit{ is satisfiable iff every finite } \Sigma' \subseteq \Sigma \textit{ is satisfiable.}$$

Recall that a set Σ is satisfiable iff there is a model (D, J) such that

$$(D, J) \Vdash \psi$$

for all $\psi \in \Sigma$. With this in mind we turn to the proof of the right-to-left direction of the theorem (the converse direction is trivial).

Proof of Compactness Theorem. We assume that every finite $\Sigma' \subseteq \Sigma$ is satisfiable and show by a proof by contraction that Σ is satisfiable. To this effect, we assume that Σ is not satisfiable. Then we have $\Sigma \models \bot$ (Why? Exercise!) and by the adequacy of $\text{FOL}_=$ we obtain $\Sigma \models \bot$, that is, there is a derivation of $\bot$ from Σ. However, natural deduction proofs are of finite length and so there must be a finite $\Sigma' \subseteq \Sigma$ such that $\Sigma' \vdash \bot$. Applying the adequacy result again we obtain $\Sigma' \models \bot$ which implies that Σ' is not satisfiable contra our assumption. We can conclude that Σ is satisfiable. $\square$

One important moral of the Compactness theorem is that if an infinite set of sentences is inconsistent, then there must be a finite number of sentences in the set which are jointly inconsistent!

The Compactness theorem also points us towards expressive imitations of first-order logic. In the previous section we saw that in $\text{FOL}_=$ we can require there to be to be exactly n-many things where n is some natural number, that is, for each natural number n we can find a sentence φ_n such that for every model (D, J) of first-order logic

$$(D, J) \Vdash \varphi_n \text{ iff } D \text{ has exactly } n \text{ elements.}$$

For example, for $n = 2$ we have

$$\varphi_n := \exists x \exists y(\neg x = y \wedge \forall z(x = z \vee y = z)).$$

A natural idea would then be to try and find a sentence or, perhaps a set of sentence that requires there to be finitely many things, but not of a specific finite size. As we shall see, we cannot find such a sentence (set of sentences) in $\text{FOL}_=$.

Theorem 93. *There is no set of* $\mathrm{FOL}_=$*-sentences* Σ *such that*

$$(D, J) \Vdash \psi, \text{ for all } \psi \in \Sigma \text{ iff } D \text{ is a finite set}$$

for every model (D, J).

Proof. To show this claim we assume that there is such a set Σ expand the language of $\mathrm{FOL}_=$ by constants c_j for every natural number j, that is by infinitely many new constants. Now we define

$$\Sigma' := \Sigma \cup \{\neg c_j = c_k \mid j < k \text{ for } j, k \in \mathbb{N}\}.$$

Next we show that every finite subset $\Gamma \subseteq \Sigma'$ is satisfiable. If Γ is a finite subset of Σ' then there will only be a finite number of negated identity statements. Let $\Gamma^= := \{\neg c_j = c_k \mid j < k \text{ for } j, k \in \mathbb{N}\} \cap \Gamma$, $D^\Gamma := \{n \in \mathbb{N} \mid \neg c_n = c_k \in \Gamma \,\&\, k \in \mathbb{N}\}$ and define $J_\Gamma(c_j) = j$. Clearly (D_Γ, J_Γ) is a model of $\Gamma^=$. But D_Γ is also finite and so by assumption $(D_\Gamma, J_\Gamma) \vDash \Sigma$. Hence $(D_\Gamma, J_\Gamma) \vDash \Gamma$ since $\Gamma \subseteq \Sigma \cup \Gamma^=$. Then, since every finite subset of Σ' is satisfiable we conclude by the Compactness theorem that Σ' is satisfiable, i.e., has a model. But any model of Σ' needs to have an infinite domain, as for every non identical natural numbers j and k we know that $\neg c_j = c_k \in \Sigma'$. This implies that the domain must have at least as many elements as there are natural numbers, i.e., the domain of any model of Σ' needs to be infinite. Let (D, J) a model of Σ'. Then (D, J) is also a model of Σ. So we have found a model of Σ with infinite domain, and we can conclude that there is no set of sentences Σ that is true precisely on models with finite domains. □

The theorem shows that while in $\mathrm{FOL}_=$ we can define/express quantifiers

- there are n many things

$\mathrm{FOL}_=$ does not have the expressive power to define/express quantifiers such as

- there are finitely many things
- there are infinitely many things

Amongst other reasons this has led logicians to propose to extend first-order logic with so-called *generalized quantifiers* (cf. MacFarlane, Chapter 2). Yet, such extensions have the consequence that either

- there will not be an adequate proof system or
- that we no longer have a decision algorithm that tells us whether something is a proof or not.

Thus introducing generalized quantifies to our logic comes at an important cost. Indeed, one can show that first-order logic is (in some sense) the strongest logic for which one can obtain an adequacy result without altering the notion proof.

Chapter 11

Definite Descriptions

In the previous chapter there were two types of terms—so-called singular terms—that can stand for objects of the domain: individual constants and variables. In natural language, however, there seem to be other types of terms, that is, types of expression we use to denote objects which are not names (and not variables):

- the king of France
- the square root of two
- the square root of -1
- Catrin's jumper
- Richard's flat

Such expressions are called definite description (in contrast to indefinite description such as *a blue jumper* etc.). The idea is that such expressions provide us with a description that enables us to uniquely single out the relevant object. In natural language definite descriptions assume the same syntactic position as names and it thus seems appealing to introduce a new category of singular terms. However, while definite description play an important role in language there are good arguments to think that we should, perhaps not think of them as singular terms.

Bivalence Consider the sentences

K1 The king of France is bald.

K2 The king of France is not bald.

Are K1 and K2 true or false? In classical logic if K1 is true, then K2 must be false and vice versa. Since there is no king of France, then it seems that K1 must be false (if it has a classical truth value). But doesn't the same reasoning also suggest that it is also not the case that the king of France is not bald, i.e., that K2 is also false. Yet, this contradicts the law of excluded middle—are we forced to accept that there must be a third truth values like"undefined"? Recall that in Chapter 8 we used this type of argument to motivate a third truth value. Can we resist this argument?

Existence In classical logic every individual constant denotes, that is, we can prove for every constant c of the language that $\exists x(x = c)$ (Exercise: how?). But we use definite descriptions to deny the existence of an object satisfying the description, e.g.,

E1 The golden mountain does not exist.

E2 The present king of France does not exist.

So perhaps definite descriptions are not singular terms?

Logicality Consider the following argument: *The logic teacher is wearing a blue jumper. Hence, there is a logic teacher wearing a blue jumper.* On the face of it one would think that the conclusion of the argument is a logical consequence of the premise. But if we handle definite descriptions like ordinary terms this will not be the case: let W be the predicate '*is wearing a blue jumper*', L the predicate '*is a logic teacher*' and t the term 'the logic teacher'. Then the premise can be formalized as 'Wt' and the conclusion as '$\exists x(Lx \wedge Wx)$'. $\exists x(Lx \wedge Wx)$ is not a logical consequence of Wt (Exercise: give a counterexample to this effect).

11.1 Definite Description as Incomplete Symbols

From the perspective of the language of FOL with identity, a description is simply a formula with a free variable, e.g., the Formula Fx (we may read 'F' as '*is a king of France*'. We then add a new operator ι that allows us to form and represent the definite description *the F* as the expression ιxFx. Let G be a predicate (read '*is bald*'). Then we represent '*the F is G*'—the king of France is bald—by '$G\iota xFx$'. For reasons that should become clear shortly we call '$G\iota xFx$' a *pseudo-formula* and 'ιxFx' a *pseudo-term*. By the same token we obtain a general rule for constructing definite descriptions as follows:

- if φ is a formula and x a variable, then $\iota x\varphi$ is a pseudo-term.

Thus the pseudo-term $\iota x\varphi$ may be read as *the x such that φ*. Such expressions are called pseudo-terms as they occupy term positions, i.e., syntactic positions that are occupied by individual constants. However, at least according to the theory proposed by Bertrand Russell definite descriptions are not singular terms and a pseudo-formula such as $G\iota xFx$ is not an actual formula of FOL, but an abbreviations of a more complex formula/claim.

Famously, Bertrand Russell argued that by saying '*the F is G*' we are making three distinct claims. We are saying that:

(a) there exists at least one F

$$\exists xFx$$

(b) there exists at most one F

$$\forall x(Fx \rightarrow \forall y(Fy \rightarrow x = y))$$

(c) everything that is F is also G

$$\forall x(Fx \rightarrow Gx)$$

Jointly these three claims are logically equivalent to the following formula:

$$(D_{FG}) \qquad \exists x(Fx \wedge \forall y(Fy \longrightarrow x = y) \wedge Gx)$$

(Exercise: show this claim in the natural deduction system of FOL with identity.)

Accordingly, Russell held that $G\iota xFx$ is only abbreviation of the formula (D_{FG}) above and even though ιxFx seems to occupy a term position in $G\iota xFx$, it is not actually a term of FOL as it does not occur in (D_{FG}). The definite description ιxFx is contextually eliminated (the formula Gx being the context). It is for that reason that $G\iota xFx$ is a called pseudo-formula and ιxFx a pseudo-term.

So to sum up a definite description $\iota x\varphi$ is not a term of the language FOL and it only superficially occupies term position in, e.g., a formula ψ. Indeed, if we write $\psi(\iota x\varphi)$ this denotes the formula

$$\exists x(\varphi(x) \wedge \forall y(\varphi(y) \longrightarrow x = y) \wedge \psi(x))$$

Ultimately, this means that the *logical form* of $G\iota xFx$ is not subject-predicate; $G\iota xFx$ is not an atomic formula like Gc where c is an individual constant. Rather the logical form of $G\iota xFx$ is given by the complex formula (D_{FG}).

Russell's theory and bivalence Now that we know how to handle definite description let's look back at Sentences K1 and K2 and the question of bivalence. Given our understanding of the predicates F and G, K1 may be formalized as $G\iota xFx$ and K2 as $\neg G\iota xFx$. Applying Russell's account of definite descriptions K1 need to be understood in terms of the formula (D_{FG}). Since there is no king of France, i.e., the denotation of F will be empty, (D_{FG}) will be false as expected. What about $\neg G\iota xFx$? If we apply Russell's theory we obtain the following formula

$$(\dagger) \qquad \exists x(Fx \wedge \forall y(Fy \longrightarrow x = y) \wedge \neg Gx)$$

which, indeed, is also false. But notice that (†) is not the negation of (D_{FG})—nor is it equivalent to the negation of (D_{FG}). (†) says that the unique object that is the king of France is not bald, i.e., it says that the king of France is non-bald. On this reading $\neg G\iota xFx$ should indeed be false, whilst the actual negation of (D_{FG}), i.e.,

$$\neg\exists x(Fx \wedge \forall y(Fy \longrightarrow x = y) \wedge \neg Gx)$$

will be true as required by bivalence. This points to a structural ambiguity of $\neg G\iota xFx$ which comes down to the scope of the definite description: according to the first reading—the *wide scope* reading (the DD scopes over the entire formula)—-we have $[\neg Gx](\iota xFx/x)$ and on the second—the *narrow scope* reading (the DD does not scope over the negation symbol)— we have $\neg[Gx(\iota xFx/x)]$. $[\neg Gx](\iota xFx/x)$ says that the king of France is non-bald whereas the second, $\neg[Gx(\iota xFx/x)]$, says that it is not the case that the king of France is bald. $\neg G\iota xFx$ comes out false on the first but true on the second reading, and it is the second reading that is the negation of (D_{FG}). Russell's analysis thus respects bivalence but also gives a nice account of why K2 may be taken to be false.

Russell's theory and negative existentials In FOL with identity claiming that an individual constant c does not denote leads to a contradiction, as in FOL with identity we can prove $\exists x(x = c)$. So is it contradictory to say that the king of France does not exist, i.e., is $\neg\exists y(y = \iota x Fx)$ a logical contradiction? Recall that $\neg\exists y(y = \iota x Fx)$ is a pseudo-formula and not a formula of the language of FOL with identity. To find out whether $\neg\exists y(y = \iota x Fx)$ is logical contradiction we need to apply Russell's theory and eliminate the definite description. As in the discussion of bivalence we need to be careful with respect to which context, i.e., which formula we are eliminating the definite description. Since we surely do not want to say that the king of France has the property of nonexistence we should work with the narrow scope reading $\neg[\exists y(y = \iota x Fx)]$. Applying Russell's theory we obtain:

$$\neg\exists x(Fx \wedge \forall z(Fz \longrightarrow x = z) \wedge \exists y(x = y))$$

Since the denotation of F will be empty the non-existence claim comes out true, which again suggests that we may want to distinguish definite descriptions and individual constants.[1]

Russell's theory and logicality Let 'L' be the predicate '*is a logic teacher*' and 'W' the predicate '*is wearing a blue jumper*. Then the sentence '*the logic teacher is wearing a blue jumper*' can be formalized by the pseudo-formula $W\iota x Lx$ and the sentence '*there is a logic teacher wearing a blue jumper*' can be formalized as '$\exists y(Ly \wedge Wy)$'. Using Russell's theory of definite description $W\iota x Lx$ can be eliminated in favour of the formula $\exists x(Lx \wedge \forall y(Ly \longrightarrow x = y) \wedge Wx)$. It is an easy exercise to show that assuming the latter formula one can give a proof of $\exists y(Ly \wedge Wy)$ (Exercise!), which appealing to the soundness of the system of natural deduction allows one to conclude that

$$\exists x(Lx \wedge \forall y(Ly \longrightarrow x = y) \wedge Wx) \models \exists y(Ly \wedge Wy).$$

11.2 Outlook and Discussion

Are definite descriptions singular terms? We just gave a couple of considerations to the contrary but this by no means settles the debate. One can push back against the arguments we have just discussed in many ways (and somewhat independently of whether one takes definite descriptions to be singular terms or not). According to Russell's account a sentence like K1 *entails* or *implies* that there is a unique king of France. However, Strawson forcefully argued that Russell got it wrong and that K1 does not entail the existence of a unique king of France but for speakers to understand K1 we must *presuppose* that there is a unique king of France, that is, in absence of a king of France we fail to understand K1. According to Strawson K1 presupposes the existences of a unique king of France rather than entail it. Strawson's analysis motivates an analysis in terms of a three-valued logic such as K3. For further more linguistically oriented discussion we refer to Chapter 2 in Lycan: Philosophy of Language, Chapter 2.3 of Abbott: Reference, or also the entry on Description in the *Stanford Encyclopaedia of Philosophy*.[2]

[1] On the wide scope reading $\neg\exists y(y = \iota x Fx)$ as $[\neg\exists y(x = y)](\iota x Fx/y)$ the claim will be false. This is as one would expect it: there is no object one ascribing a property to (even if it is the property of non-existence).

[2] See also Chapter §2 (in particular §2.2) in MacFarlane: Philosophical Logic for further useful discussion, and an explanation of why Russell's theory amounts to understanding 'the' as quantifier/determiner.

Similarly, our discussion of negative existentials is somewhat wanting: isn't it an odd feature of classical first-order logic that one can prove that every constant denotes, that is, that if we have a name in the language then we can *prove* that there exists a referent to that name? Maybe this is an acceptable presupposition if we are working in mathematics (?), but surely this is problematic in connection to natural languages. We refer to the next chapter (Chapter 12) for discussion.

However, no matter one's view on definite descriptions one important take home message of this chapter is that scope is important when it comes to definite description. We saw that definite descriptions give rise to narrow and wide scope readings with respect to negation, but this phenomenon is not limited to negation. Consider the sentence:

NP The number of planets is necessarily greater than 5.

which we may formalize taking 'N' to be the predicate ' is a number of planets' as

$$\Box(\iota x N x > 5).[3]$$

Now potentially there are two different ways we may understand NP:

NP1 The number which is the number of planets is necessarily greater than 5.

NP2 It is necessary that there are more than 5 planets.

(Exercise: discuss whether you think that both readings are actually available or not). Now, NP1 which is the so-called *de re*-reading of NP is true (8 is necessarily greater than 5), but NP2 the so-calle *de dicto*-reading of NP is false (our solar system could have only had 4 planets). These two different readings of NP correspond to whether the definite description $\iota x N x$ takes scope over the modal operator $\Box$ or not. The *de re*-reading corresponds to the definite description taking wide scope, i.e.,

$$\exists x(Nx \wedge \forall y(Ny \longrightarrow x = y) \wedge \Box y > 5)$$

while the *de dicto*-reading corresponds to the narrow scope reading:

$$\Box\exists x(Nx \wedge \forall y(Ny \longrightarrow x = y) \wedge y > 5).$$

[3]Strictly speaking we should write $\Box B\iota N x\, 5$ with B being the predicate 'greater than'.

Chapter 12

Free Logic

As we have already seen in discussing definite description in classical first-order logic with identity we can prove for every constant c that c denotes an object of the domain. This seems highly counterintuitive, at least if we think of these constants as names of natural languages such as Pegasus, Santa Clause, or Sherlock Holmes. The argument is shockingly simple:

1	$c = c$	=1
2	$\exists x(x = c)$	$\exists$I, 1

Surely, if we are involved in ontological/metaphysical debates, we need to be careful before concluding that Pegasus exists, or, at least, that Pegasus exists in the same way that James Ladyman exists. This should not simply follow from the fact that there is some name in our language, and logic alone. Famously, this was one of the reasons (there are other reasons) which motivated Bertrand Russell to claim that basically all proper names in natural language are definite descriptions in disguise.[1] If, say, Pegasus is not a name but a definite description then the logical form of the sentence 'Pegasus is Pegasus' is not an identity statement as in Line 1 of the above proof, but a complex quantified sentence that would come out as false.

However, while Russell's view gives a nice systematic solution to the problem raised by so-called non-denoting terms, the view that proper names are definite description in disguise is highly contentious and nowadays rarely accepted. An alternative is to alter the logic of identity and quantification to the effect that the logic no longer presupposes that every individual constant (closed term) denotes an object of the domain of quantification. There a two ways how this can be implemented in the realms of classical logic. The first is in some sense Meinongean in spirit and holds that not all objects "exist", which, in this context, we understand as saying that the domain of quantification does not exhaust the set of all objects: if an object is in the domain of quantification, it exists. Otherwise the object is a nonexistent object. This option may be philosophical hard to swallow: what does it mean for an object not to exist? Yet, the resulting logic so-called **Positive Free Logic (PFL)** is very elegant and easy to use. The second way simply allows for some constants not to denote. This is so-called **Negative Free Logic (NFL)**.

[1] See Chapter 3 in Lycan: Philosophy of Language for a discussion of Russell's view.

The difference between PFL and NFL can be summed as follows. In PFL the rule (=I) holds for all constants whether they denote an existent of non-existent object and atomic formulas of the form $Pc_1, \ldots, c_n$ can be true even if some constant c_j $(1 \leq j \leq n)$ does not denote an "existing" object. In contrast in NFL atomic formulas will only be true if all constants that occur in the formula denote. This means that the (=I)-rule needs to be replaced in NFL by

(i) $\forall x(x = x)$

This shows that in both PFL and NFL the rule of ∀-elimination has to be revised (recall that the only quantifier rule we used to show the ∃-introduction rule was a derivable rule was the rule of ∀-elimination).

12.1 Natural Deduction and Semantics for PFL

12.1.1 Natural Deduction for PFL

We just mentioned that for both PFL and NFL the rule of ∀-elimination needs to be altered. Indeed for PFL we will only need to change the quantifier introduction and elimination rules. In free logic before we need to make sure that the individual constant under consideration exists before we can eliminate the universal quantifier. This leads to the **Free ∀-Elimination** rule:

$$
\begin{array}{l|ll}
m & \forall x \varphi & \\
n & \exists y(c = y) & \\
n+1 & \varphi(c/x) & \text{F-}\forall\text{E}, m, n
\end{array}
$$

We also need to alter the ∀-Introduction rule. The rule ∀I allows us to conclude that $\forall x \varphi(x/c)$, if we have derived φ and c does not occur in a premise or an undischarged assumption. However, this rule does not guarantee the c denotes an object of the domain of quantification. The **Free ∀-Introduction** rule requires us to operate under assumption that c exists:

$$
\begin{array}{l|l|ll}
m & & \exists x(x = c) & \\
\cline{3-3}
\vdots & & \vdots & \\
n & & \varphi(c) & \\
n+1 & \multicolumn{2}{l}{\forall x \varphi(x/c)} & \text{F-}\forall\text{I}, m\text{–}n
\end{array}
$$

where c does not occur in an undischarged assumption or premise.

One obtains the natural deduction system for PFL from the natural deduction system for FOL with identity by replacing the rules ∀E and ∀l by the rules F-∀E and F-∀I respectively. No further changes are required.

One consequence of the new rule F-∀E and F-∀I is that ∃I and ∃E are no longer derivable rules (nor are they sound rules). Instead we obtain the new derived rules of **Free ∃-Introduction/Elimination**:

Free ∃-Introduction rule

$$\begin{array}{l|ll} m & \varphi(c) & \\ n & \exists y(y = c) & \\ n+1 & \exists x\varphi[x/c] & \text{F-}\exists\text{I}, m, n \end{array}$$

provided x is free for c in φ and where $\varphi[x/c]$ denotes a formula that results from φ by substituting one or more occurrences of c by x.

Free ∃-Elimination rule

$$\begin{array}{l|l|ll} k & \exists x\varphi & & \\ m & & \exists y(y = c) \wedge \varphi(c/x) & \\ \cline{3-3} n & & \chi & \\ n+1 & \chi & & \text{F-}\exists\text{E}, k, m\text{–}n \end{array}$$

c is new to the derivation and does not occur in χ.

Exercise 94. *Show that F-∃I and F-∃E are derivable rules of PFL.*

Using F-∃I we can no longer infer $\exists x(x = c)$ from the premise $c = c$. $\exists x(x = c)$ is precisely the information we need to apply F-∃I. Trivial existence proofs are no longer possible in PFL as it was intended.

We give two theorems of PFL:

- $\vdash_{\mathsf{PFL}} \forall x(x = x)$

$$\begin{array}{l|l|ll} 1 & & \exists x(x = c) & \\ \cline{3-3} 2 & & c = c & =1 \\ 3 & \forall x(x = x) & & \text{F-}\forall\text{I}, 1\text{–}2 \end{array}$$

- $\vdash_{\mathsf{PFL}} \forall x\exists y(x = y)$

$$\begin{array}{l|l|ll} 1 & & \exists y(y = c) & \\ \cline{3-3} 2 & & \exists y(y = c) & \text{R}, 1 \\ 3 & \forall x\exists y(x = y) & & \text{F-}\forall\text{I}, 1\text{–}2 \end{array}$$

12.1.2 Semantics for PFL

Turning to the semantics the main novelty will be that there are now objects that are not in the domain of quantification.

Definition 95 (PFL-model)**.** *A PFL-model M is a tuple (U, D, J) such that $D \subseteq U$ where D is a (possibly empty) domain of quantification and U the universe of discourse of the model. J is a denotation function such that:*

(i) $J(c) \in U$*, for all individual constants (names)* c*;*

(ii) $J(P^k) \subseteq \underbrace{U \times \ldots \times U}_{k\text{-times}}$*, for every predicate* P *of arbitrary arity* k*.*

According to the definition individual constants denote objects in the universe of the model. We are not guaranteed that these objects are also members of the domain of quantification. Similarly, the interpretation of a predicate is defined on the universe and not the domain so a formula, e.g., Pc can be true independently of whether the denotation of c is in the domain of quantification. In contrast, variables will be assigned objects of the domain of quantification

Definition 96 (Assignment Function, x-variant)**.** *Let* (U, D, J) *be a PFL-model. An assignment function* $\beta : \mathsf{Var} \longrightarrow D$ *assigns an object of the domain* D *of the model to every variable of the language. If* β *is an assignment function, then* $\beta(x : d)$ *is an assignment function such that for all* $y \in \mathsf{Var}$*:*

$$\beta(x : d)(y) := \begin{cases} d, & \text{if } x \doteq y; \\ \beta(y), & \text{otherwise.} \end{cases}$$

$\beta(x : d)$ *is called an* x*-variant of* β*.*

With these definitions in place we can assign truth values to the formula of the language as in Definition 84 (together with the clause for = discussed in Chapter 10 with the sole proviso that instead of a model of FOL, we are working with a PFL-model. Everything else remains unchanged.

Exercise 97. *Show that* $\forall x Px \longrightarrow \exists x Px$ *is not a logical truth of PFL.*

12.2 Natural Deduction and Semantics for NFL

12.2.1 Natural Deduction for NFL

In NFL constants may simply fail to denote and in this case atomic formulas in which they occur will be false: if the constant c fails to denote an object of the domain of the model, then formulas like $c = c$ or Pc will be false. As a consequences, the identity rule =I needs to be replaced for NFL by the following rule:

$$\begin{array}{r|ll} m & \forall x(x = x) & \text{Q=I} \end{array}$$

The rule =E remains unchanged for if one of the constants c_1 and c_2 does not denote, then $c_1 = c_2$ will be false. This guarantees that the rule =E remains truth preserving. Consequently, if $c_1 = c_2$ is true, then both c_1 and c_2 must denote (exist). More generally, if $\varphi(c)$ is an atomic formula in which c occurs and $\varphi(c)$ is true, then c must denote (exist). We introduce a rule to this effect in NFL:

$$\begin{array}{r|ll} m & \varphi(c) & \\ m+1 & \exists x(x = c) & \text{CE, } m \end{array}$$

whenever φ is an atomic formula and c occurs in φ.

12.2.2 Semantics for NFL

A model M of NFL is just a tuple (D, J) where J is a partial function in the sense that J does not necessarily assign an object of D to every constant c.

Definition 98 (NFL-model)**.** *A NFL-model M is a tuple (D, J) such that D where D is a (possibly empty) domain of quantification and J is a denotation function such that:*

(i) $J(c) \in D$, if J is defined for the individual constant c;

(ii) $J(P^k) \subseteq \underbrace{D \times \ldots \times D}_{k\text{-times}}$, for every predicate P of arbitrary arity k.

We say that the denotation $t^{J,\beta}$ of a term t is defined in a model (D, J) relative to an assignment function β iff t is a variable or if t is a constant and $J(t) \in D$. With this in mind we can proceed to assign truth values to all formulas of the language relative to an NFL-model and an assignment function.

Definition 99 (Truth)**.** *Let (D, J) be an NFL-model and β an assignment function. Then $\mathrm{I}^{\beta}_{(D,J)}$ is a function that assigns truth values to all formulas of the language relative to (D, J) and β such that:*

1. *If φ is a formula $P^k t_1, \ldots, t_k$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } t_1^{J,\beta}, \ldots, t_k^{J,\beta} \text{ are defined and } \langle t_1^{J,\beta}, \ldots, t_k^{J,\beta} \rangle \in J(P^k); \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

2. *If φ is $s = t$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } t^{J,\beta}, s^{J,\beta} \text{ are defined and } s^{J,\beta} = t^{J,\beta}; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

3. *If φ is $\bot$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \mathsf{F}.$$

4. *If φ is a formula $\psi \wedge \chi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \text{if } I^{\beta}_{(D,J)}(\psi) = I^{\beta}_{(D,J)}(\chi) = \mathsf{T}; \\ \mathsf{F}, & \text{otherwise.} \end{cases}$$

5. *If φ is a formula $\psi \vee \chi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } I^{\beta}_{(D,J)}(\psi) = \mathsf{T} \textit{ or } I^{\beta}_{(D,J)}(\chi) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise.} \end{cases}$$

6. *If φ is a formula $\psi \longrightarrow \chi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } I^{\beta}_{(D,J)}(\psi) = \mathsf{F} \textit{ or } I^{\beta}_{(D,J)}(\chi) = \mathsf{T}; \\ \mathsf{F}, & \textit{otherwise.} \end{cases}$$

7. *If φ is a formula $\forall x \psi$, then*

$$I^{\beta}_{(D,J)}(\varphi) := \begin{cases} \mathsf{T}, & \textit{if } I^{\beta(x:d)}_{(D,J)}(\psi) = \mathsf{T} \textit{ for all } d \in D; \\ \mathsf{F}, & \textit{otherwise.} \end{cases}$$

We write $(D,J) \Vdash \varphi[\beta]$ iff $\mathrm{I}^{\beta}_{(D,J)}(\varphi) = \mathsf{T}$. A formula φ is true in the model (D,J) iff $(D,J) \Vdash \varphi[\beta]$ for every variable assignment β. In this case we simply write $(D,J) \Vdash \varphi$.

Exercise 100. *Show that $\forall xPx \longrightarrow \exists xPx$ is not a logical truth of NFL.*

12.3 Free Logics vs Universally Free Logics

We motivated the free logics PFL and NFL by arguing against the presupposition of classical first-order logic that every individual constant (term) of the language denotes an object of the domain of quantification. It does not seem sufficient to infer from the existence of a name in our language to the existence of an object that is denoted by said name. The names 'Pegasus', 'Sherlock Holmes', or 'Mary Poppins' all seem to be a case in point. However, the logics we just introduced go one step further in that they not only get rid of the presupposition that every constant (term) denotes, but of the presupposition that something exists: recall in classical first-order logic we can prove the existence of at least one object. This is reflected in the semantics of first-order logic by the requirement that the domain of the first-order model has to be non-empty, and by the fact that the following formulas are logical truths of FOL:

(a) $\forall x \varphi \longrightarrow \exists x \varphi$;

(b) $\exists x \varphi$, if $\varphi(c/x)$ is a tautology (derivable in propositional logic);

(c) $\exists x(x = x)$;

In both the semantics for PFL and NFL we allowed the domain of quantification to be empty and the above formulas are not logical truths (or theorems) of PFL and NFL. Logics that allow for the domain to be non-empty are called **Universally Free Logic**s and, as introduced, PFL and NFL are universally free logics.

But both PFL and NFL can be turned into logics that require the domain of quantification to be nonempty, that is, they can be turned into **Free Logic**s rather than Universally Free Logics. In their respective semantics this is of course simply achieved by adding the requirement that $D \neq \emptyset$ where D is the domain of quantification of the PFL-model (NFL-model). Turning to the systems of natural deduction for PFL and NFL this condition can be enforced by adding the following rule to both systems of natural deduction:

m	$\exists x(x = x)$	E!

To end this chapter we show that by adding E! to the respective natural deduction systems we can prove (a) and (b) above (and of course (c)). First, we prove that assuming E! one can prove in both PFL and NFL:

(†) $$\exists x \exists y(x = y)$$

1	$\exists x(x = x)$	E!
2	$\quad \exists y(y = c) \wedge c = c$	
3	$\quad \exists y(y = c)$	$\wedge$E, 2
4	$\quad \exists x \exists y(x = y)$	F-$\exists$I, 3, 3
5	$\exists x \exists y(x = y)$	F-$\exists$E, 1, 2–4

To establish (a) we now reason as follows:

1	$\forall x\varphi$	
2	$\neg\exists x\varphi$	
3	$\neg\neg\forall x\neg\varphi$	Def. $\exists$, 2
4	$\forall x\neg\varphi$	DNE, 3
5	$\exists x(x = c)$	
6	$\varphi(c/x)$	F-$\forall$E, 1, 5
7	$\neg\varphi(c/x)$	F-$\forall$E, 4, 5
8	$\bot$	$\longrightarrow$E, 6, 7
9	$\neg\exists x(x = c)$	$\bot$E, 8
10	$\forall y\neg\exists x(x = y)$	F-$\forall$I, 5–9
11	$\neg\neg\forall y\neg\exists x(x = y)$	DNI, 10
12	$\neg\exists y\exists x(x = y)$	Def. $\exists$, 11
13	$\exists y\exists x(x = y)$	†
14	$\bot$	$\longrightarrow$E, 12, 13
15	$\exists x\varphi$	PbC, 5–14
16	$\forall x\varphi \longrightarrow \exists x\varphi$	$\longrightarrow$I, 1–15

We now show that we can derive (b). Notice that if φ is a tautology, we can assume that there is a proof of φ using the rules of propositional logic alone. This means φ is derivable in PFL and NFL:

1	$\exists x(x = c)$	
2	$\varphi(c)$	Taut
3	$\forall x\varphi(x/c)$	F-$\forall$I
4	$\forall x\varphi \longrightarrow \exists x\varphi$	(a) above
5	$\exists x\varphi$	$\longrightarrow$E, 3, 4