Modal Logic for Artificial Intelligence

Jan Broersen & Tristan van der Vlugt

May 17, 2020

Abstract & Acknowledgement

This reader is written to be used as course notes for the course "Modal Logic for AI" for students of Artificial Intelligence at Utrecht University. The first part of the notes are largely based on earlier notes by Rosja Mastop.

This is still very much a work in progress, as can be witnessed by the missing third part. The bibliography is not finalised either.

CONTENTS

INTRODUCTION

In this reader you will find an introduction to modal logic. We assume the reader has had a first encounter with propositional and predicate logic and is able to understand elementary mathematical texts. For those students who lack this understanding or who want to refresh their memory, it is highly advised to work through Part A.

MOTIVATION

Formal logic can be roughly described as the study of the rules of argumentation. Whenever we try to reason about a subject, we wish for our conclusions to be a reasonable consequence of the assumptions we make. Formal logic gives us a way to study this reasonability of conclusions and to quantify the validity of an argument. However, not every argument is alike. There are many different ways to argue whether a certain argument is valid, and many depend on the context. For this purpose there have been developed many types of logic.

Possibly one of the simplest kinds of logic is **classical propositional logic**. Here, a proposition is any statement that is **truth functional**, or in other words, any statement can be determined to be true or false. The word **classical** refers to the fact that this is one of the oldest kinds of logic, having been studied extensively by the ancient Greek logicians. It lies at the basis of much of Western analytical philosophy and is considered by many to be a very natural way of reasoning. An example of an argument in classical propositional logic is the following:

The bus is not crowded.
If it is raining, then the bus is crowded.

Therefore, it is not raining.

We have two propositions, namely *"it is raining"* and *"the bus is crowded"*, which have been related to each other by an implication (*if ..., then ...*). The argument shows that if we know the truth value of *"the bus is crowded"* to be false, then we can conclude that the truth value of *"it is raining"* has to be false as well. This conclusion is based on two acceptable rules from classical propositional logic, namely the rules of noncontradiction (a statement and its negation cannot be simultaneously true), and the rule of modus ponens (an implication together with its antecedent allows us to conclude the consequent).

Not every kind of reasoning that is made in natural language can be translated accurately with propositional logic. For example, the following argument seems to follow not from the truth value of its statements, but from a relation between properties it describes:

Tuna are fish.
If something is a fish, then it has fins.

Therefore, tuna have fins.

If we try to translate this argument using propositional logic, we run into a problem: the second statement does not state anything about tuna specifically, therefore from a propositional viewpoint there is nothing relating the first statement to the second statement. Instead, the second statement tells us something about a relation between properties. We therefore have to use a different logic to study this type of argument. In this case, **classical predicate logic**, also known as **first-order logic**, appears to be adequate. A **predicate** is some statement that could be made about subjects. In the above example we have two predicates *"is a fish"* and *"has fins"*, and a single subject *"tuna"*. The first statement states that our subject *"tuna"* has the property *"is a fish"*, and our second statement relates the two predicate by an implication. Formally this looks something like this:

$$\begin{array}{l} \text{isFish}(\text{tuna}) \\ \forall x(\text{isFish}(x) \to \text{hasFins}(x)) \\ \hline \text{hasFins}(\text{tuna}) \end{array}$$

The conclusion follows from two acceptable rules, namely the rule of universal instantiation (when a predicate holds for all subjects, it holds for any specific subject) and once again the rule of modus ponens.

Naturally there are many other arguments that can be made that cannot be adequately studied using propositional logic. Consider the following examples of arguments:

Possibly it is raining.
Necessarily, when it rains, the bus will be crowded.

Therefore, possibly the bus is crowded.

I am obligated to pay rent for my house.
When I pay rent for my house, I am allowed to live in my house.

Therefore, I ought to be allowed to live in my house.

At some point in the future, I will finish a marathon.
Before I finish a marathon, I finish half a marathon.

Therefore, if I have not finished half a marathon yet, I will eventually do so.

It is more likely that I lose the lottery than that I win it.
It is more likely that I survive until 2030 than that I don't.

Therefore, there is a chance that I both lose the lottery and survive until 2030.

John knows that Mary knows whether it rains.
John knows that it rains.

Therefore, Mary knows that it rains.

All of these arguments are usually studied using different logics than first-order logic. The first argument is studied with alethic logic, the second with deontic logic, the third with temporal logic, the fourth with probabilistic logic and the last with epistemic logic. This sounds like a whole bunch of completely different systems, but in a way all of them are related by a common property: each of these logics uses *modality* to express its statements.

A *modality* in an argument gives a means to express differing degrees of possibility and necessity and can be used to reason about hypothetical situations. In some interpretations this possibility is objective, such as in alethic logic or temporal logic, while in others it depends on the subject, such as in epistemic logic. In a broad sense, modal logic is the study of such modalities and the rules that naturally come with their interpretations. These rules differ between the different interpretation, yet there is also a lot of similarity. Modal logic is the field that studies the different systems that use modality in their reasoning, either by studying the system individually, or studying the similarities between systems.

In this reader, we will spend the first half working on a good general frameworks for what are called *normal modal logics*. In the second half, we focus more specifically on systems for epistemic logic.

APPLICATION TO ARTIFICIAL INTELLIGENCE

A natural question you might ask at this moment is why knowledge of modal logic is useful in the field of artificial intelligence. We will discuss two subjects where modal logic plays a significant role.

Multi-Agent Systems
A Multi-Agent System is a system of multiple interacting entities, called agents, that have to cooperate to achieve a goal. One could, for example, imagine several agents working together to solve a complex task, where each agent individually has to decide what to do based on only a limited amount of information. It could very well be that no single agent has a complete overview of the whole situation, but through collaboration the agents might become able to solve the complex task.

It is necessary for such agents to be able to make decisions based on their own state and on how an action will change their state and the states of other agents. The logic behind such decision making can be described and studied accurately with modal logic.

As an example, consider the following puzzle:

> Three agents, Anne, Bill and Cate, each are assigned a positive integer, one of which is the sum of the other two integers. None of the agents knows their own integer, but each agent knows the integer that is assigned to the other two agents. Finally all agents are aware of everything that has just been said.
>
> Anne, Bill and Cate successively announce to each other, truthfully, that they do not know their own integer; first Anne announces this, then Bill announces this and finally Cate announces this. After these announcements Anne announces that she knows her own number: 42.

We see that although the agents had incomplete information at the start, Anne managed to know her number from reasoning with information *about* information. She never directly got told what her assigned number was, but she could still figure this out using logic.

This type of problem will be studied in the later chapter about epistemic logic.

Natural Language Processing
Another aspect of artificial intelligence in which modal logic is useful, is in natural language processing. This becomes quite apparent already from observing the examples of arguments with modalities that were given earlier.

For example, an intelligent entity needs to have a good concept of time. Preferably, we want such an entity to be able to properly understand and answer the following question:

> *"On Thursdays I go to the gym before I watch my weekly drama series.*
> *I do not go to the gym until after my work has ended.*
> *Do I have to be afraid to hear spoilers for my weekly drama series at work?"*

In order to understand the meaning of these sentences and answer them correctly, we need to be able to logically reason with the concept of time. Temporal logic is a form of modal logic that is suitable for this, and thus temporal logic can play a very important role in modelling the temporal reasoning of an intelligent entity.

MATHEMATICAL AND LOGICAL BACKGROUND

In this section we will very briefly give an overview of the logical and mathematical language that will be used in this reader. In particular there will be a summary of classical propositional logic (CPL), soundness and completeness and the basis of set theory and mathematics. Clearly this is not meant as a complete overview, and it is recommended to take a look at an introductory resource on classical logic, set theory and / or foundations of mathematics if these concepts are unclear to you.

LOGIC

In a very loose description, logic is the study of reasoning and argumentation. In a more precise way, logic studies when it is valid to draw a conclusion from some given premises. Formal logic is the specific area of logic that does not study the *content* of arguments, but the *form* of the argument.

There are two main ways of studying the forms of arguments. One is the **semantic** approach, where we use **models** to study arguments, and the other is the **syntactic** approach, where we use **proofs** to study arguments. In this section we will develop both of these approaches for **classical propositional logic** (CPL). This should be familiar to you, but it will serve as a useful comparison to the development of the same concepts for modal logic in later chapters.

Before we can start reasoning, we need something to reason with: we need a language to talk in. A **logical language** could be seen as a set of rules that tells us which things are sentences that we could use, and which things are not. A logical language $\mathcal{L}$ usually consists of:

- a set of **constants**,
- a set of **atomic variables** At, also called **propositional variables** or **atoms**, and
- a set of **logical connectives**.

For CPL we have the language $\mathcal{L}_{\text{CPL}}$, which has:

- two constants $\{\bot, \top\}$,
- (countably) infinitely many atomic variables $\mathsf{At} = \{p, q, r, p_0, p_1, \dots, q_0, \dots\}$,
- five connectives $\{\vee, \wedge, \rightarrow, \leftrightarrow, \neg\}$, and
- two auxiliary symbols: the brackets (and), used to avoid ambiguity.

A **formula** in a logical language is a combination of these symbols that follows some grammatical rules. In the case of CPL a string of symbols of the language $\mathcal{L}_{\text{CPL}}$ is a formula if and only if it can be build (in finitely many steps) from the following rules:

- If $c \in \mathcal{L}_{\text{CPL}}$ is a constant, then c is a formula (i.e. $\bot$ and $\top$ are formulas).
- If $p \in \mathsf{At}$ is an atomic variable, then p is a formula.
- If φ is a formula, then $(\neg\varphi)$ is a formula,
- If φ and ψ are formulas, then $(\varphi \vee \psi)$, $(\varphi \wedge \psi)$, $(\varphi \rightarrow \psi)$ and $(\varphi \leftrightarrow \psi)$ are formulas.

Anything that can not be built by following these rules is not a formula of CPL. We will denote the set of all propositional formulas as Fml_{CPL}.

Brackets are important to prevent ambiguous sentences to be formed. Usually excessive brackets are left out whenever there is no ambiguity. In this, it is standard that $\neg$ binds stronger than $\vee$

and $\wedge$, which in turn bind stronger than $\rightarrow$ and $\leftrightarrow$. This means that, for example, $\neg p \vee q \rightarrow r \wedge s$ should be understood as $(((\neg p) \vee q) \rightarrow (r \wedge s))$

Now that we have a language to work with, we will investigate the **semantics** of these symbols, or in other words, their meaning. In particular we will decide for each formula whether it is true or not in a given context. The context in which we study the truth of a formula is called a **model**, often denoted with the symbols $\mathcal{M}$ and $\mathcal{N}$. The structure of a model can be widely different for different kinds of logic. For example, the models for CPL are different in structure than the ones we will develop for modal logic. Before we talk about what models for CPL are, let us introduce some terminology that is generally used for most logics.

- A **model** $\mathcal{M}$ gives us a method to determine for any formula φ if it is **true in** $\mathcal{M}$, which we denote by $\mathcal{M} \vDash \varphi$, or if it is **false in** $\mathcal{M}$, denoted by $\mathcal{M} \nvDash \varphi$.
- If φ is a formula that is true in *any* possible model, then we write this as $\vDash \varphi$ and call φ a **tautology** or a **validity**.
- Given a set of formulas $\Psi = \{\psi_1, \psi_2, \dots\}$, we say a formula φ is a **consequence of** Ψ, denoted as $\Psi \vDash \varphi$, if any model that makes every formula of Ψ true, also makes φ true.
- A formula φ is **satisfiable** if there is some model $\mathcal{M}$ such that φ is true in $\mathcal{M}$. A formula that is not satisfiable (i.e. there are no models that make it true) is called a **contradiction**. A **contingent** formula is a formula that is satisfiable, but not a tautology, or in other words, a formula is contingent if there are models that make it true as well as models that make it false.
- Two formulas φ and ψ are **logically equivalent** if in any model $\mathcal{M}$ either both φ and ψ are true, or both φ and ψ are false.

We will now investigate the models of CPL. A propositional model $\mathcal{M}$ is a function that *interprets* each of the atomic variables as either **true** or **false**. In other words, $\mathcal{M}$ consists of a function $V : \mathsf{At} \rightarrow \{\mathbf{T}, \mathbf{F}\}$. This function is called a **valuation**. We say that $\mathcal{M} \vDash p$ for an atomic variable p if and only if $V(p) = \mathbf{T}$. Equivalently we could say that $\mathcal{M} \nvDash p$ for an atomic variable p if and only if $V(p) = \mathbf{F}$.

The truth of a formula φ is defined recursively by looking at how φ is built up from the grammar of our language*:

- It always holds that $\mathcal{M} \vDash \top$ and $\mathcal{M} \nvDash \bot$.
- $\mathcal{M} \vDash \neg\varphi$ iff $\mathcal{M} \nvDash \varphi$.
- $\mathcal{M} \vDash \varphi \wedge \psi$ iff $\mathcal{M} \vDash \varphi$ and $\mathcal{M} \vDash \psi$.
- $\mathcal{M} \vDash \varphi \vee \psi$ iff $\mathcal{M} \vDash \varphi$ or $\mathcal{M} \vDash \psi$.
- $\mathcal{M} \vDash \varphi \rightarrow \psi$ iff $\mathcal{M} \nvDash \varphi$ or $\mathcal{M} \vDash \psi$.
- $\mathcal{M} \vDash \varphi \leftrightarrow \psi$ iff ($\mathcal{M} \vDash \varphi$ and $\mathcal{M} \vDash \psi$) or ($\mathcal{M} \nvDash \varphi$ and $\mathcal{M} \nvDash \psi$).

This might look like a difficult definition, but it is in fact the same thing as building a truth table. A row in a truth table for a formula φ is nothing more than one of the possible valuations for the atomic variables in φ. In the columns for the connectives, we give a valuation to the connective using the exact same rules as are dictated by the recursive definition above.

For example if we have a connective $\wedge$ that connects φ with ψ, then we give a row in the column of $\wedge$ in the truth table the value $\mathbf{T}$ if and only if both of the columns for φ and ψ have the value $\mathbf{T}$ in that row. A formula φ is a tautology if the most prominent column of φ only has values

---

[*] This kind of recursion is called **recursion** over the **complexity** (or **structure**) of the formula. A related notion is **induction** over the complexity of a formula, which we will see in multiple examples in this reader.

T, which means that any possible valuation function (and thus any model $\mathcal{M}$) makes φ true. This corresponds with $\vDash \varphi$.

Now that we have given a semantical meaning to formulas of our language, the other thing we could study is the laws that allow us to reason about arguments. This is where we study the *structure* of arguments. If we have an argument with some premises and a conclusion, and this conclusion is reached from the premises using a very specific set of syntactical manipulations, then we call such an argument a **proof**. The syntactical manipulations that are allowed make up what is called a **proof system**. As with the semantics, we will first introduce some terminology and then look at the specific case for CPL.

- A proof system consists of **axiom schemes** and **derivation rules**.
- A formula φ is considered an **axiom** if its **syntactical structure** (i.e. the way the formula is build up recursively) is the same as the syntactical structure of one of the axiom schemes. For example, if $(A \wedge B) \to A$ is an axiom scheme, then any formula that has the form $(\varphi \wedge \psi) \to \varphi$ for any formulas φ and ψ is considered an axiom. This holds even if φ and ψ have additional structure, e.g. $((p \vee q) \wedge \neg p) \to (p \vee q)$ is indeed an instance of the aforementioned axiom scheme.
- An derivation rule consists of the schemes of premises from which the scheme of a conclusion follows. A formula φ follows from formulas $\psi_1, \dots, \psi_n$ if the **syntactical structure** of φ and of the ψ_i's is the same as the syntactical structures of the conclusion and the premises respectively of one of the derivation rules. For example, if an derivation rules says that from the premises $A \to B$ and $B \to A$ one can conclude that $A \leftrightarrow B$, then for any formulas φ and ψ one can conclude $\varphi \leftrightarrow \psi$ from the premises $\varphi \to \psi$ and $\psi \to \varphi$.
- A **proof** of a formula φ is a (finite) sequence of formulas $\psi_0, \psi_1, \dots, \psi_n$ such that $\psi_n = \varphi$ and each ψ_i is either an axiom or follows from (some of) the formulas ψ_j with $j < i$ using an derivation rule. In other words, in writing a proof of φ, you can start with axioms and rewrite them using the derivation rules, until the result is the formula φ.
- If there exists a proof of φ, then we call φ a **theorem** and we write $\vdash \varphi$.
- Given a set of formulas $\Psi = \{\psi_1, \psi_2, \dots\}$, we say φ is **provable from** Ψ if there is a proof that starts with finitely many formulas in Ψ and uses only axioms and derivation rules to conclude φ. This is denoted as $\Psi \vdash \varphi$.
- A set of formulas Ψ is **consistent** if there is no formula φ such that $\Psi \vdash \varphi$ and $\Psi \vdash \neg\varphi$. If Ψ is not consistent, it is called **inconsistent**. If Ψ is inconsistent, then $\Psi \vdash \varphi$ for any formula φ.

As an example for a proof system for CPL, consider the following **Hilbert system**:

Axiom schemes:

$$A \to (B \to A)$$
$$(A \to (B \to C)) \to ((A \to B) \to (A \to C))$$
$$A \wedge B \to A \qquad A \wedge B \to B$$
$$A \to A \vee B \qquad B \to A \vee B$$
$$A \to (B \to (A \wedge B))$$
$$(A \to C) \to ((B \to C) \to (A \vee B \to C))$$
$$\neg\neg A \to A$$

Derivation rules:

Modus Ponens (MP): $\Gamma \vdash \varphi$ and $\Gamma \vdash \varphi \to \psi$ implies $\Gamma \vdash \psi$

A proof in this system is then a finite sequence of formulas, starting with some instances of the axioms, and all other formulas being the result of the modus ponens rule. For example, here is a proof of $\vdash \varphi \vee \psi \to \psi \vee \varphi$.

1.	$\varphi \to \psi \vee \varphi$	Axiom
2.	$\psi \to \psi \vee \varphi$	Axiom
3.	$(\varphi \to \psi \vee \varphi) \to ((\psi \to \psi \vee \varphi) \to (\varphi \vee \psi \to \psi \vee \varphi))$	Axiom
4.	$(\psi \to \psi \vee \varphi) \to (\varphi \vee \psi \to \psi \vee \varphi)$	MP $1, 3$
5.	$\varphi \vee \psi \to \psi \vee \varphi$	MP $2, 4$

Here line 3 uses the second to last axiom of the system defined above.

The two notions of *semantical* and *syntactical* reasoning look very different, but if we want these two systems to be of any use, we naturally want them to correspond to the same thing. In particular, it is very important that we can prove only true formulas, since otherwise our proof system does not really make sense from the viewpoint of our semantic interpretation. This property, that a formula φ is a *tautology* whenever it is a *theorem*, is called **soundness**. Soundness is a basic requirement of any logic that has both a semantical and a syntactical interpretation. In short, soundness is exactly the statement that $\vdash \varphi \Rightarrow \models \varphi$.

If we want to show that the proof system from above is sound, we have to show that each of the syntactical manipulations can draw a conclusion only if this conclusion is valid with respect to the semantics. This boils down to showing that each axiom is a tautology, and that each derivation rule holds up semantically.

For example, the rule of Modus Ponens is valid semantically, since if $\mathcal{M} \models \varphi \to \psi$, then by the definition of the truth of a formula, we see $\mathcal{M} \not\models \varphi$ or $\mathcal{M} \models \psi$. As we also assume $\mathcal{M} \models \varphi$, it can not be the case that $\mathcal{M} \not\models \varphi$, and thus $\mathcal{M} \models \psi$ must be true.

As an example of an axiom being a validity, consider $A \to (B \to A)$. If $\mathcal{M}$ is any model for CPL, then $\mathcal{M} \models A \to (B \to A)$ using the definition of the truth of a formula, since if $\mathcal{M} \models A$, then also $\mathcal{M} \models B \to A$, and thus $\mathcal{M} \models A \to (B \to A)$, while if $\mathcal{M} \not\models A$, then we again see that $\mathcal{M} \models A \to (B \to A)$.

Another common requirement for a logic, is that any true statement can be proved. If a logic has this property, then in a certain sense our proof system is *complete* in that it can prove anything that is true, hence we call this property **completeness**. In short, completeness is exactly the statement that $\models \varphi \Rightarrow \vdash \varphi$.

Even if a proof system is sound with respect to some semantical system, this does not need to imply that it is complete. For example, if we take the Hilbert system from before, but leave out the final axiom ($\neg\neg A \to A$), the result is a proof system that is still sound with respect to truth tables: any formula that can be proved in the Hilbert system will be valid when evaluated using a truth table. However, the resulting logic is no longer complete with respect to truth tables. There are formulas that are valid using the method of truth tables that can no longer be proved in the Hilbert system. The logic that we just described is called **intuitionistic propositional logic**. Although it is not complete with respect to truth tables, there are other semantical systems for which intuitionistic propositional logic is complete.[*]

---

[*] One of such semantical systems is closely related to modal logic, in particular to Kripke semantics with partially ordered frames (what these terms mean is introduced in Chapters 1 and 2).

SETS

Most of mathematics can be written in the language of sets. We will not go into a lot of detail as to what a set precisely is, but an intuitive idea is that a set is a collection of objects, called the **elements** of the set. A set can be written as a list of its elements by putting the elements between accolades: $A = \{b, c, d\}$ means that A is a set with elements b, c and d and no other elements. If a is an element of set A, we write this as $a \in A$, and if a is not an element of A, then we write $a \notin A$. If all elements of A are also elements of B, we call A a **subset** of B, denoted $A \subseteq B$. To describe a subset $A \subseteq B$ that contains exactly those elements of B that satisfy a property P, we use the set builder notation $A := \{x \in B \mid P(x)\}$. When the set B is understood from context, we also just write $A := \{x \mid P(x)\}$.

The following are examples of sets:

- $\varnothing$ is a set that contains no elements. It is called the empty set, and is unique in that it contains no elements.
- $\{\varnothing\}$ is a set that contains only the empty set. Notably, this set is not empty itself.
- $\{1, 2, 3, 4, 5\}$ is a set that contains the numbers 1 through 5.
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$ is the set of **integers**, $\mathbb{N} = \{0, 1, 2, \dots\}$ is the set of nonnegative integers, also know as **natural numbers**, and $\mathbb{R}$ is the set of **real numbers**, also known as the number line. We have $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{R}$.

If two sets contain the same elements, then they are the same sets. This is the main difference between a set and a sequence, list or array. In particular, this means the order of notation or multiplicity of elements does not matter: $\{1, 2, 3\}$ and $\{3, 1, 3, 2, 2, 2\}$ contain the same elements and hence are the same set. The next is a small list of how we can manipulate sets to build new sets:

- A **singleton** $\{A\}$ is a set containing a single element, in this case the element A.
- If $A_1, \dots, A_n$ is a list of elements, then an ***n*-tuple** is a sequence $\langle A_1, \dots, A_n\rangle$. The difference between $\{A_1, \dots, A_n\}$ and $\langle A_1, \dots, A_n\rangle$ is that the order and multiplicity of the elements does not matter in a set, but it *does matter* in an n-tuple. As an example we have $\{A_1, A_2\} = \{A_2, A_1\}$, whereas $\langle A_1, A_2\rangle \neq \langle A_2, A_1\rangle$ (unless $A_1 = A_2$ of course). We usually call a 2-tuple an **(ordered) pair** and a 3-tuple a **triple**.
- Given two sets A and B, define the **union** $A \cup B$ as the set $\{x \mid x \in A \vee x \in B\}$, the **intersection** $A \cap B$ as the set $\{x \mid x \in A \wedge x \in B\}$ and the **relative complement** $A \setminus B$ as the set $\{x \mid x \in A \wedge x \notin B\}$. For example, for the sets $\{1, 2\}$ and $\{2, 3\}$ we have

  the union: $\{1, 2\} \cup \{2, 3\} = \{1, 2, 3\}$,
  the intersection: $\{1, 2\} \cap \{2, 3\} = \{2\}$ and
  the complements: $\{1, 2\} \setminus \{2, 3\} = \{1\}$ and $\{2, 3\} \setminus \{1, 2\} = \{3\}$.
- For any set A there exists a set $\mathcal{P}(A) = \{X \mid X \subseteq A\}$ containing all the subsets of A. This set is called the **power set** of A. For example, the power set of $\{1, 2\}$ is:

  $$\mathcal{P}(\{1, 2\}) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}.$$
- For any sets A and B, the **cartesian product** $A \times B = \{\langle a, b\rangle \mid a \in A \wedge b \in B\}$ is the set of all ordered pairs with their first element coming from A and their second element coming from B. For example:

  $$\{1, 2\} \times \{2, 3\} = \{\langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 2\rangle, \langle 2, 3\rangle\}.$$

Sets happen to be very useful concepts to describe more complex mathematical structures with. For this reader we will particularly need the concepts of a relation between sets and of functions between sets.

- If A and B are sets, then a **relation** $R \subseteq A \times B$ is a set. We write the statement "$\langle a, b\rangle \in R$" also with infix notation as $a \mathrel{R} b$, or as a **relates to** or **reaches** b. If $A = B$, we also call R a **relation on** A. For example:

  > $<$ is a relation on $\mathbb{N}$ and we have $\langle 2, 5\rangle \in <$, usually written as $2 < 5$.
  >
  > The set of all pairs $\langle a, n\rangle \in \mathbb{Z} \times \mathbb{N}$ such that n is an natural number that divides integer a (without remainder) is a relation. The pairs $\langle 12, 3\rangle, \langle -16, 8\rangle$ and $\langle 1, 1\rangle$ are elements of this relation, but the pairs $\langle -9, 15\rangle$ and $\langle 17, 4\rangle$ are not.

- If $R \subseteq A \times B$ is a relation and $C \subseteq A$ and $D \subseteq B$, then the **restriction** of R to $C \times D$ is the subset of R defined by $R \upharpoonright C \times D = \{\langle a, b\rangle \in R \mid a \in C \wedge b \in D\}$. We see that $R \upharpoonright C \times D$ is a relation from C to D. For example:

  > The restriction of $<$ as a relation on $\mathbb{N}$ to $< \upharpoonright \{1, 2\} \times \{1, 2, 3\}$ is the relation $\{\langle 1, 2\rangle, \langle 1, 3\rangle, \langle 2, 3\rangle\}$.

- If A and B are sets, then a **function** or a **map** $f : A \to B$ is a relation from A to B that contains exactly one ordered pair $\langle a, b\rangle$ for each $a \in A$. If $\langle a, b\rangle \in f$, we also write this as $f(a) = b$ or as $f : a \mapsto b$ and we call b the **image** of a over f, and we say that f **maps** or **sends** a to b. The **domain** of f is the set $\mathrm{dom}(f) = \{a \in A \mid \langle a, b\rangle \in f\}$ and is equal to A. The **range** of f is the set $\mathrm{ran}(f) = \{b \in B \mid \langle a, b\rangle \in f\}$. For example:

  > The function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ contains the elements $\langle x, x^2\rangle$ for any $x \in \mathbb{R}$. For example, $\langle 2, 4\rangle \in f$ and $\langle -3, 9\rangle \in f$, usually written as $f(2) = 4$ or as $f : 2 \mapsto 4$. The range of f is the set of all $x \in \mathbb{R}$ such that $x \geq 0$.

- As a special case of a restriction, if $f : A \to B$ is a function and $C \subseteq A$, then the restriction of f to C is a function $f \upharpoonright C : C \to B$. For example:

  > If f is the function from the last example, then $f \upharpoonright \{-5, 1, 2, 3\}$ is the function $\{\langle -5, 25\rangle, \langle 1, 1\rangle, \langle 2, 4\rangle, \langle 3, 9\rangle\}$ and is a function with domain $\{-5, 1, 2, 3\}$ and range $\{1, 4, 9, 25\}$.

  Note that we need to be careful with restricting B to a subset $D \subseteq B$, as the result might no longer be a function: it could be that $f(c) = b$ for some $c \in C$ and $b \notin D$, which would mean that $\mathrm{dom}(f)$ is no longer all of C if we restrict the range of f to D.

- A function $f : A \to B$ is **injective** if every $a \in A$ is sent to a unique $b \in B$, that is, for any $a, a' \in A$ such that $a \neq a'$ we also have $f(a) \neq f(a')$. An injective function is also called an **injection** or **one-to-one**.

- A function $f : A \to B$ is **surjective** if every $b \in B$ is the image of some element $a \in A$, that is, for any $b \in B$ there is an $a \in A$ such that $f(a) = b$. A surjective function is also called a **surjection** or **onto**.

- A function $f : A \to B$ is **bijective** if it is both injective and surjective. A bijective function is also called a **bijection** or a **one-to-one correspondence**.

Finally we will give a very short introduction to the **cardinality** of a set. Each set has a cardinality, which intuitively describes the size of a set. We write the cardinality of a set A as $|A|$. If A and B are sets, then $|A| \leq |B|$ if and only if there exists an injective function

$f : A \to B$. In other words, A is smaller or equal in size to B if we can pick a unique element in B for each element in A. Two sets A and B have the same cardinality $|A| = |B|$ if there are two injective functions $f : A \to B$ and $g : B \to A$, or equivalently if there is a bijective function $h : A \to B$.

For example, the set $A = \{1, 2, 3\}$ and the set $B = \{3, 5, 7, 10\}$ have cardinalities $|A| = 3$ and $|B| = 4$. We see that $|A| \leq |B|$, since we have the function $f = \{\langle 1, 3\rangle, \langle 2, 5\rangle, \langle 3, 7\rangle\}$, also known as the function $f(x) = 2x + 1$. This function is an injection, since every element in A is mapped to a unique element in B: no two elements in A map to the same value.

Using injective functions to see which set has a larger cardinality seems unnecessary for the example above: we could have just counted the elements to get to the same conclusion. However, things get a lot more difficult with infinite sets.

A set X is **finite** if it has exactly n distinct elements, for some natural number n. If a set X is **infinite** and $|X| = |\mathbb{N}|$, then we call the set X **countably infinite**. A set is **countable** if it is countably infinite or finite. Note that $|X| \leq |\mathbb{N}|$ is true if and only if there is an injective function $f : X \to \mathbb{N}$. What this function f does, is give a number to each of the elements of X. This means that the elements of a countable set X can be **enumerated** in a numbered list.

As an example, consider the set of all integers $\mathbb{Z}$. You might think that there are more elements in $\mathbb{Z}$ than there are in $\mathbb{N}$, since $\mathbb{N}$ contains only the positive integers, while $\mathbb{Z}$ contains both the positive *and* the negative integers. However, we can enumerate $\mathbb{Z}$ with the following function $f : \mathbb{Z} \to \mathbb{N}$:

$$f(x) = \begin{cases} -2x & \text{if } x \text{ is negative} \\ 2x + 1 & \text{if } x \text{ is not negative} \end{cases}$$

You can check that this is an injective function, and thus that $|\mathbb{Z}| \leq |\mathbb{N}|$. We also have $|\mathbb{N}| \leq |\mathbb{Z}|$, since we can simply map each natural number to itself to get an injective function.

Not every infinite set is countable. For example the set $\mathbb{R}$ of all real numbers is **uncountable**. The proof of this fact is famous for being elegant and surprising to those who see it for the first time, and is considered by many mathematicians to be *beautiful.* It is called **Cantor's diagonal argument**, but because we will not use uncountable sets in this reader, we will not write the proof itself. It is highly recommended to look up the proof yourself if you are interested.

PART I

SEMANTICS OF MODAL LOGIC

1 SEMANTICS OF MODAL LOGIC

As we saw in the introduction, modal logic is a broad family of logics that allow us to reason about modal qualifiers on propositional statements. There are many different interpretations of modal logic, related to many different modalities, each having their own semantic and syntactic systems to work with. We will mostly work with just one of these interpretations for the first half of this book, before we move to other modalities.

The kind of modality that we will use for our exposition is called the *alethic modality*. The word *alethic* comes from the Greek word ἀλήθεια (aletheia), meaning *truth*. An alethic modality is that of statements being *"necessarily true"*, *"possibly true"* or *"impossible"*. We use two symbols, the box operator $\Box$ and the diamond operator $\Diamond$, to express these modalities. If φ is some statement, then we interpret $\Box\varphi$ as φ being necessary, and $\Diamond\varphi$ as φ being possible.

Example 1.1
As an example, consider the following argument:

> Possibly, it is raining.
> Necessarily, when it rains the bus will be crowded.
> Therefore, possibly the bus will be crowded.

We can describe this argument using our modal operators $\Box$ and $\Diamond$ as:

$$\frac{\begin{array}{l}\Diamond raining \\ \Box(raining \to crowded\ bus)\end{array}}{\Diamond crowded\ bus}$$

◁

In order to study why this argument is valid, we will develop a formal semantics for modal logic called **Kripke semantics**. Before we do this, let's take a brief moment to remember what the semantics of propositional logic and first-order logic looked like.

In propositional logic, when we want to evaluate the truth of a formula, we first need a context to interpret the formula in. This context is given by assigning a truth value to each propositional variable. For example, if p is false, and q and r are true, then the formula $p \vee q \to r$ is true. This is what we consider to be a propositional model: we interpret each of the propositional variables as true or false to interpret the whole formula.

If we want to discover what the general truth of the formula is, we do this by considering all possible models. In propositional logic this relates to the concept of a truth table, in which we systematically go through all possible valuations for p, q and r to discover exactly when the formula above is true. In case of the formula $p \vee q \to r$, a truth table will show that it is true if and only if r is true or if both p and q are false.

In first-order logic, the context is given by interpreting a formula inside a domain of discourse. Here a model will consist of a set of elements, relations and functions, such that we can interpret each of the constants, relations and functions of our first-order language with those from the model. For example, if we have the set $\{John, Bob, Mary\}$ of three people, with *John* being *Bob*'s father, and *Bob* being *Mary*'s father, then we can interpret the first-order formula $\forall xyz(x < y \wedge y < z \to x < z)$ by letting the relation $x < y$ correspond to "y is a father of x". Given this interpretation, if we let $x = Mary$, $y = Bob$ and $z = John$, then $x < y \wedge y < z \to x < z$ is false, since *John* is not *Mary*'s father (but her grandfather).

Once again, to discover the general truth of a formula, we have to consider all possible models. Doing this, we discover that the formula $\forall xyz(x < y \wedge y < z \to x < z)$ holds for exactly those sets with a relation $<$ that is transitive.

The Kripke semantics we will study works in the same way: it provides a context in which we can study the truth of modal formulas.

1.1 LANGUAGE

To begin, we will need to define exactly what constitutes a modal formula. We do this with the following recursive definition.

Definition 1.2 — *Language of modal logic*
The **language of modal logic** $\mathcal{L}_{\mathrm{ML}}$ is an extension of the language of propositional logic with the symbols $\Box$ and $\Diamond$. A **modal formula** is defined recursively:

- The constants $\top$ and $\bot$ are modal formula.
- Every propositional variable is a modal formula.
- If φ and ψ are modal formulas, then $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \to \psi$, $\varphi \leftrightarrow \psi$ and $\neg\varphi$ are modal formulas.
- If φ is a modal formula, then $\Box\varphi$ and $\Diamond\varphi$ are modal formulas.
- Nothing else is a modal formula.

We denote the set of modal formulas with $\mathrm{Fml}_{\mathrm{ML}}$. ◁

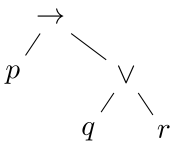
The first three clauses of the definition above are precisely the clauses of the language of propositional logic, so we see that any propositional formula is also a modal formula. The fourth clause gives us two new unary operators, $\Box$ and $\Diamond$, alongside the classical operators. As mentioned before, these operators are commonly called *box* and *diamond.*

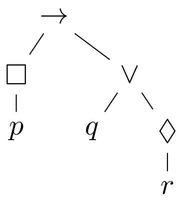
It should be noted that this definition uses more logical operators than is strictly necessary. Similarly to how the classical operator $\vee$ is superfluous, because $\varphi \vee \psi$ can be defined in terms of $\wedge$ and $\neg$ as $\neg\varphi \wedge \neg\psi$, we can leave out one of the operators $\Box$ or $\Diamond$ from the definition without losing any strength in the kind of sentences that we can express. We will come back to this fact in Lemma 1.14.

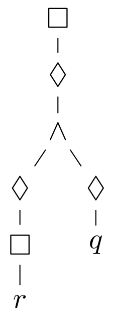
Example 1.3
The following are examples of modal formulas:

$$p \to q \vee r$$
$$\Box p \to q \vee \Diamond r$$
$$\Box\Diamond(\Diamond\Box r \wedge \Diamond q)$$

To show that these formulas are indeed well formed, we can build a semantic tree, exploring how the formula is build from the propositional variables:







On the other hand, the following are not considered modal formulas by the definition above:

$p\Box q$

$\forall x \Box P(x)$

$p\Box \to q$

In the first, it appears as if $\Box$ is used as a binary operator, while it should be a unary operator.

The second uses first-order quantifiers mixed with modal operators. In this reader we're working with propositional modal logic, and thus quantifiers are not part of our modal language. Note that it is possible to define a stronger kind of logic, that uses both quantifiers and modality. However, in this reader we will not treat this kind of modal logic, called predicate modal logic.

The third appears to apply a modal operator on a logical connective, as if it was a translation of *"p necessarily implies q"*. The correct way to translate this sentence is as $\Box(p \to q)$, as can be seen from the semantic tree of the latter formula. ◁

1.2 Kripke Models

Alethic modal logic essentially adds the concepts of *necessity* and *possibility* to the propositional language. This means that the formulas of modal logic are no longer truth-functional: the truth of $\Diamond\varphi$ does not depend solely on the truth of φ, but is dependent on another, broader, context. Intuitively we want to express with $\Diamond\varphi$ that out of all conceivable situations, there is some situation in which φ is true. Similarly, with $\Box\varphi$, we wish to express that φ is true in any conceivable situation.

To illustrate this, we look back at the argument of Example 1.1. If we are in a room without windows or access to a local weather report, we have no way of knowing whether it is raining or not. We therefore consider two kinds of worlds, one in which it rains and one it which it doesn't. The actual world could be of either of these two kinds, we just don't know which it is. Similarly, we have no way to know if the bus is crowded or not.

The first assumption, that it is possibly raining, states that we consider it possible that our actual world is of the kind where it rains. The second assumption on the other hand, states that in any world we can conceive of, if it rains, then the bus is crowded. Since we consider the actual world to be possibly of the kind where it rains, we also consider the actual world to be possibly of the kind where the bus is crowded, and therefore the conclusion is legitimate.

To formalise this idea, we will describe a kind of model that can be used to model modal logic. These models are called Kripke models, after the logician Saul Kripke. We start by defining the underlying structure of a Kripke model. These structures are called Kripke frames.

Definition 1.4 — *Kripke frame*
A **Kripke frame** is a pair $\mathcal{F} := \langle W, R \rangle$, where

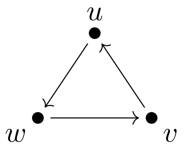
- W is a nonempty set, whose elements are called the **worlds** of $\mathcal{F}$, and
- $R \subseteq W \times W$ is a relation, called the **accessibility relation**. If $(w, v) \in R$, we also write this as $w \; R \; v$ and say that v is **accessible** from w, and that w **accesses** v, or that w can **reach** v, or that w **sees** v ◁

The worlds in a Kripke frame are representations of different kinds of worlds. The accessibility relation defines which kinds of worlds are conceivable from the perspective of a certain world in the frame. So if a world w accesses v, then in case the actual world were w, we consider it a possibility that v is the actual world.

Note that from the perspective inside the frame, we do not know which world is the actual world, similarly to how in the example above we do not now if it rains in the actual world or not.

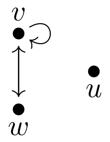
We can draw Kripke frames in a diagram by letting the worlds be points and drawing the relation as arrows between the points.

Example 1.5
Let $\mathcal{F}_1 = \langle W_1, R_1 \rangle$, where $W_1 = \{w, v, u\}$ and $R_1 = \{\langle w, v \rangle, \langle v, u \rangle, \langle u, w \rangle\}$. We can draw this as follows:



We see that $w \; R_1 \; v$, for example, because there is an arrow going from w to v, or because the ordered pair $\langle w, v \rangle$ is an element of R_1. ◁

Example 1.6
Let $\mathcal{F}_2 = \langle W_2, R_2 \rangle$, where $W_2 = \{w, v, u\}$ and $R_2 = \{\langle w, v \rangle, \langle v, w \rangle, \langle v, v \rangle\}$. We can draw this as follows:

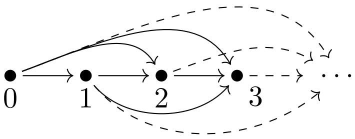


From this frame several interesting properties become apparent:

- Two worlds can reach each other, as we can see with w and v. We draw this with a double-headed arrow. Intuitively, this means that from the perspective of either of the worlds w or v, the other world is considered to be possibly the actual world.
- A world can reach itself, such as v in this frame, which is drawn by an arrow going from the world to itself. This means intuitively that if v were the actual world, then it is considered a possibility that v is the actual world.
- It is possible that a world accesses two different worlds, such as $v \; R \; w$ and $v \; R \; v$ both holding in this frame. Intuitively this means that from the perspective of v, there are multiple worlds that can conceivably be the actual world.

- It is possible for a world to not access any world, such as u in this frame. Intuitively this means that there is no world conceivable, if u were the actual world. We call such a world a **blind world**.
- The worlds in a frame do not need to be connected, but could consist of several separate clusters not accessible from each other by the accessibility relation. ◁

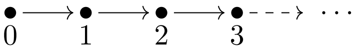
Example 1.7
Let $\mathcal{F}_3 = \langle W_3, R_3 \rangle$, where $W_3 = \mathbb{N} = \{0, 1, 2, 3, \dots\}$ and $R_3 = \{\langle a, b \rangle \mid a < b\}$ (we could also write $R_3 = <$, since $a\ R_3\ b$ if and only if $a < b$). We can draw this as follows:



Note that the worlds of this frame are numbers. This is not a problem, since Definition 1.4 did not specify what kind of objects worlds should be. Indeed, intuitively we so far have been talking about worlds as if they are some representation of a possible reality, but in general it does not really matter what kind of objects we take for our worlds, since the kind of object does not matter in the slightest in how we interpret the semantics of modal logic. We could take our set of worlds to be *anything*, from names such as w, v or u, to numbers $0, 1$ or 2, to actual things such as *cow*, *baseball* or *planet earth*.

The frame $\mathcal{F}_3$ has an infinite number of worlds, so it is impossible to draw it completely. Instead, we just drew the first four worlds and implied that it continues forever by using three dots.

We also see that there are too many relation arrows to draw all at once. However, this relation R_3 has a special property, called transitivity. This means that whenever $x\ R_3\ y$ and $y\ R_3\ z$, then also $x\ R_3\ z$. With transitive relations we usually only draw the most essential arrows, and state that the relation is supposed to be transitive. This results in a much cleaner diagram:



Note however, that it is *essential* to mention that this relation is supposed to be transitive, as the diagram that is drawn above could also depict the frame $\mathcal{F}_4 = \langle W_4, R_4 \rangle$ with $W_4 = \mathbb{N}$ and $R_4 = \{\langle a, a+1 \rangle \mid a \in \mathbb{N}\}$; that is, the relation where a number a only accesses its immediate successor $a + 1$. ◁

The idea of a Kripke model, is that in each world certain propositions are true and others are false. The most basic propositions are the propositional variables. Intuitively, we will pair each world of a frame with a set of propositional variables that are considered true in the world. The resulting structure is what is called a Kripke model.

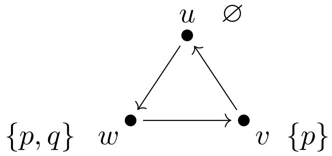
Definition 1.8 — *Kripke model*
A **Kripke model** is a triple $\mathcal{M} := \langle W, R, V \rangle$, where

- $\langle W, R \rangle$ is a Kripke frame, and
- $V : W \to \mathcal{P}(\mathsf{At})$ is a function, called the **valuation function**.

If $\mathcal{F} = \langle W, R \rangle$ is a frame, then we say $\mathcal{M}$ is **based** on $\mathcal{F}$ if the set of worlds and the accessibility relation of $\mathcal{M}$ are those from $\mathcal{F}$. ◁

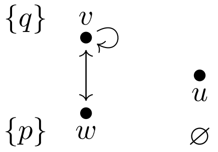
When drawing a diagram of a Kripke model, we usually write the valuation $V(w)$ for a world w next to the world's label.

Example 1.9
Let $\mathcal{F}_1$ be the frame from Example 1.5. We give a model $\mathcal{M}_1 = \langle W_1, R_1, V_1 \rangle$ based on $\mathcal{F}_1$ by letting $V(w) = \{p, q\}$, $V(v) = \{p\}$ and $V(u) = \varnothing$. We can draw this in a diagram as:



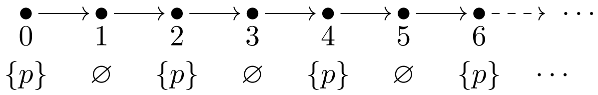
◁

Example 1.10
Let $\mathcal{F}_2$ be the frame from Example 1.6. We give a model $\mathcal{M}_2 = \langle W_2, R_2, V_2 \rangle$ based on $\mathcal{F}_2$ by letting $V(w) = \{p\}$, $V(v) = \{q\}$, $V(u) = \varnothing$:



◁

Example 1.11
Let $\mathcal{F}_3$ be the frame from Example 1.7. We give a model $\mathcal{M}_3 = \langle W_3, R_3, V_3 \rangle$ based on $\mathcal{F}_3$ by letting $V(a) = \{p\}$ if a is even and $V(a) = \varnothing$ if a is odd, giving the following transitive diagram:



◁

Remember that a propositional model is nothing more than a valuation function that maps the set of propositional variables to a truth value. Another way of looking at this, is to consider the set of all propositional variables that are true under a given valuation. In other words, by considering the set of true propositional variables, we get a propositional model.

Since the valuation function of a Kripke model maps each world to a set of propositional variables, the valuation function essentially pairs each world in the frame with a propositional model. This point of view is reasonable, since we consider the different worlds to be different alternatives to the actual world, where propositions might be true in one world and false in another, yet each world on its own is just a propositional model. The following definition will reflect this idea by defining recursively how to interpret modal formula from the perspective of a world in the model.

Definition 1.12 — *Truth in a world*
Let $\mathcal{M} = \langle W, R, V \rangle$ be a Kripke model and $w \in W$. We define φ to be true in w, denoted $\mathcal{M}, w \vDash \varphi$, by recursion on the structure of φ as follows:

$$\begin{array}{lll}
\mathcal{M}, w \vDash \top & \multicolumn{2}{l}{\text{(is always the case).}} \\
\mathcal{M}, w \nvDash \bot & \multicolumn{2}{l}{(\mathcal{M}, w \vDash \bot \text{ is never the case).}} \\
\mathcal{M}, w \vDash p & \text{iff} & p \in V(w) \text{ for atomic variables } p \in \mathsf{At}. \\
\mathcal{M}, w \vDash \varphi \wedge \psi & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi. \\
\mathcal{M}, w \vDash \varphi \vee \psi & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ or } \mathcal{M}, w \vDash \psi. \\
\mathcal{M}, w \vDash \neg\varphi & \text{iff} & \mathcal{M}, w \nvDash \varphi. \\
\mathcal{M}, w \vDash \varphi \to \psi & \text{iff} & \mathcal{M}, w \nvDash \varphi \text{ or } \mathcal{M}, w \vDash \psi. \\
\mathcal{M}, w \vDash \varphi \leftrightarrow \psi & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ if and only if } \mathcal{M}, w \vDash \psi. \\
\mathcal{M}, w \vDash \Box\varphi & \text{iff} & \text{for all } x \in W \text{ if } w\ R\ x \text{ then } \mathcal{M}, x \vDash \varphi. \\
\mathcal{M}, w \vDash \Diamond\varphi & \text{iff} & \text{there exists an } x \in W \text{ such that } w\ R\ x \text{ and } \mathcal{M}, x \vDash \varphi.
\end{array} \quad \triangleleft$$

Note that the clauses for the classical operators are evaluated with respect to the same world. For example, the truth of $\varphi \wedge \psi$ in a world w does only depend on the truth of φ and of ψ in the same world w. This means that the truth of a propositional formula in a world only depends on the valuations given in that world. Hence, the idea that each world represents a propositional model is justified under these semantics.

Furthermore, we see that the clause for $\Box$ gives us that a formula $\Box\varphi$ is true in a world w if and only if φ is true in all the worlds accessible from w. This coincides with the idea that it is necessary that φ is true, when φ is true in all the world that could conceivably be the actual world from the perspective of w. Similarly, the clause for $\Diamond$ gives us that $\Diamond\varphi$ holds exactly when there is at least one conceivable world in which φ holds.

Example 1.13
In the model, $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ from Examples 1.9, 1.10 and 1.11, we see that the following statements are true:

- $\mathcal{M}_1, w \vDash \Diamond p \wedge q$

  This is true if and only if $\mathcal{M}_1, w \vDash \Diamond p$ and $\mathcal{M}_1, w \vDash q$. The latter is true, since q is in the valuation $V_1(w)$ of w. For the former to be true, there must exist some world $x \in W_1$ such that $w\ R_1\ x$ and $\mathcal{M}_1, x \vDash p$. Since $w\ R_1\ v$ and $\mathcal{M}_1, v \vDash p$, we see that this is indeed true.

- $\mathcal{M}_2, v \vDash \Box(\neg p \to q)$

  This is true if and only if every world $x \in W_2$ such that $v\ R_2\ x$ is such that $\mathcal{M}_2, x \vDash \neg p \to q$. The only worlds accessible from v are w and v itself, therefore we need to check that $\mathcal{M}_2, w \vDash \neg p \to q$ and $\mathcal{M}_2, v \vDash \neg p \to q$. The first of these is true since $\mathcal{M}_2, w \nvDash p$, and thus the antecedent of the implication is false in w, while the second of these is true since $\mathcal{M}_2, v \vDash q$, and thus the consequent of the implication is true in v.

- $\mathcal{M}_2, u \vDash \Box(p \wedge \neg p)$

  This is true if and only if $\mathcal{M}_2, x \vDash p \wedge \neg p$ in all worlds x such that $u\ R_2\ x$. We could immediately see that no world can make $p \wedge \neg p$ true, since it is a classical contradiction. However, since u is a blind world, it it nevertheless vacuously true that for all worlds x reachable from u we have $\mathcal{M}_2, x \vDash p \wedge \neg p$, therefore the statement holds.

- $\mathcal{M}_3, 0 \vDash \Diamond p \wedge \Diamond\neg p$

  To see that $\mathcal{M}_3, 0 \vDash \Diamond\neg p$, we simply note that $0\ R_3\ 1$ and $\mathcal{M}_3, 1 \vDash \neg p$. To see that also

$\mathcal{M}_3, 0 \vDash \Diamond p$, we remind ourselves that $\mathcal{M}_3$ is transitive, and thus that $0 \; R_3 \; 2$ as well. Since $\mathcal{M}_3, 2 \vDash p$, the statement is true. ◁

Before we move on to general validity of modal formulas, we give an example of how to prove that two formulas are equivalent. The following lemma is an example of such a proof, and a special one at that: it shows that we can express the operator $\Box$ in terms of $\Diamond$ (and negation), and the operator $\Diamond$ in terms of $\Box$ (and negation). This proves that we can indeed leave one of the two operators out of our language, without losing any expressive power; we could describe the exact same things with just a $\Box$ operator, as we could with both a $\Box$ and a $\Diamond$ operator.

Lemma 1.14 — *Duality lemma*
For any model $\mathcal{M} = \langle W, R, V \rangle$ and any world $w \in W$:

$$\mathcal{M}, w \vDash \Diamond\varphi \quad \Leftrightarrow \quad \mathcal{M}, w \vDash \neg\Box\neg\varphi$$

and

$$\mathcal{M}, w \vDash \Box\varphi \quad \Leftrightarrow \quad \mathcal{M}, w \vDash \neg\Diamond\neg\varphi$$

◁

Proof. We will prove that $\mathcal{M}, w \vDash \Box\varphi$ if and only if $\mathcal{M}, w \vDash \neg\Diamond\neg\varphi$, and leave the other proof as an exercise. The proof follows completely from Definition 1.12 and elementary logic:

$$\begin{array}{rcll}
\mathcal{M}, w \vDash \Box\varphi & \Leftrightarrow & \forall v \in W (w \; R \; v \Rightarrow \mathcal{M}, v \vDash \varphi) & \\
& \Leftrightarrow & \forall v \in W (w \; R \; v \Rightarrow \mathcal{M}, v \nvDash \neg\varphi) & \text{(Definition 1.12, } \neg \text{ case)} \\
& \Leftrightarrow & \forall v \in W (w \not{R} \; v \text{ or } \mathcal{M}, v \nvDash \neg\varphi) & \\
& \Leftrightarrow & \forall v \in W \neg (w \; R \; v \text{ and } \mathcal{M}, v \vDash \neg\varphi) & \text{(De Morgan's laws)} \\
& \Leftrightarrow & \neg\exists v \in W (w \; R \; v \text{ and } \mathcal{M}, v \vDash \neg\varphi) & (\forall\neg \text{ equals } \neg\exists) \\
& \Leftrightarrow & \mathcal{M}, w \vDash \neg\Diamond\neg\varphi & \blacksquare
\end{array}$$

Note that we used the first-order quantifiers $\forall$ and $\exists$ in the above proof. This is because the above proof is not a proof *using* modal logic, but a proof *about* modal logic. We use first-order logic to talk about modal logic, its models and to prove theorems about modal logic. This is different from the proof theory we will treat in later chapters, where the focus is on proving statements made *within* modal logic.

1.3 VALIDITY

Both in propositional logic and in first-order logic, we are often interested in those formulas that are true in any given model. We call such formulas *tautologies*, and they are valid in any given context. In this section we will describe several notions of truth for a modal formula, each of them stronger than the previous notion.

Definition 1.15 — *Frame class*
A **frame class** is a class of Kripke frames. ◁

Definition 1.16 — *General validity*
Let $\mathcal{M} = \langle W, R, V \rangle$ be a model, $\mathcal{F} = \langle W, R \rangle$ a frame, $\mathcal{C}$ a frame class and φ a modal formula. Then we have the following generalisations of truth:

$$\begin{aligned} \mathcal{M} \vDash \varphi &\quad \text{iff} \quad \mathcal{M}, w \vDash \varphi \text{ for all worlds } w \in W. \\ \mathcal{F} \vDash \varphi &\quad \text{iff} \quad \mathcal{M} \vDash \varphi \text{ for all models } \mathcal{M} \text{ based on } \mathcal{F}. \\ \vDash_{\mathcal{C}} \varphi &\quad \text{iff} \quad \mathcal{F} \vDash \varphi \text{ for all frames } \mathcal{F} \text{ in } \mathcal{C}. \\ \vDash \varphi &\quad \text{iff} \quad \mathcal{F} \vDash \varphi \text{ for all Kripke frames } \mathcal{F}. \end{aligned}$$

We say φ is valid on the model $\mathcal{M}$ if $\mathcal{M} \vDash \varphi$, valid on the frame $\mathcal{F}$ if $\mathcal{F} \vDash \varphi$ and **generally valid** if $\vDash \varphi$. $\lhd$

We will talk more about validity on the level of frame classes in the next chapter. For now we will just give some examples of the other three levels of validity.

Example 1.17

Let $\mathcal{M}_3$ be the model from Example 1.11 and let $\mathcal{F}_3$ be the frame it is based on. Then we see that the following statements are true.

- $\mathcal{M}_3 \vDash \Diamond p$

  Since $\mathcal{M}_3, i \vDash p$ for any even number / world $i \in \mathbb{N}$, and since $j \; R_3 \; 2j$ because $j < 2j$ (see the definition of R_3 in Example 1.7), we see that every world accesses some world in which p is true.

- $\mathcal{F}_3 \nvDash \Diamond p$

  To show that this is false, consider the model $\mathcal{M}_3' = \langle W_3, R_3, V_3' \rangle$ based on $\mathcal{F}_3$, where we let the valuation of every world be empty, that is $V_3'(i) = \varnothing$ for every $i \in \mathbb{N}$. Then p is false in every world of the model $\mathcal{M}_3'$, and thus $\Diamond p$ is false in every world of the model too.

- $\mathcal{F}_3 \vDash \Diamond \top$

  By Definition 1.12 we see that $\top$ is true in any world under any valuation, so specifically it is true in any world of any model based on $\mathcal{F}_3$. Therefore, to see that $\mathcal{F}_3 \vDash \Diamond \top$, we have to show that any world of $\mathcal{F}_3$ reaches at least one other world of $\mathcal{F}_3$. This is clearly true.

- $\nvDash \Diamond \top$

  This is true, because we can find a model based on a different frame where $\Diamond \top$ does not hold in one of the worlds. For example, the model $\mathcal{M}_2$ of Example 1.10 has a blind world u. Since there is no world accessible from u, we see that $\mathcal{M}_2, u \nvDash \Diamond \top$, and thus $\Diamond \top$ is not a general validity. $\lhd$

A generally valid modal formula is the modal analogue of a tautology. It is a formula that is true under any given model, and in any given world. The following proposition gives an example of such a generally valid formula.

Proposition 1.18 — *Kripke axiom*

For any formulas φ and ψ, the formula $\Box(\varphi \to \psi) \to (\Box \varphi \to \Box \psi)$ is generally valid. This formula is called **Axiom K**, after Kripke. $\lhd$

> *Proof.* Let $\mathcal{M} = \langle W, R, V \rangle$ be a model, $w \in W$ a world and φ, ψ arbitrary formulas. We want to prove that $\mathcal{M}, w \vDash \Box(\varphi \to \psi) \to (\Box \varphi \to \Box \psi)$, so by semantics of $\to$ we have to show that $\mathcal{M}, w \vDash \Box(\varphi \to \psi)$ implies $\mathcal{M}, w \vDash \Box \varphi \to \Box \psi$. Hence let our first assumption be that $\mathcal{M}, w \vDash \Box(\varphi \to \psi)$. Because we now want to prove $\mathcal{M}, w \vDash \Box \varphi \to \Box \psi$, we can make the second assumption $\mathcal{M}, w \vDash \Box \varphi$ and try to prove $\mathcal{M}, w \vDash \Box \psi$.
>
> By the semantics of the $\Box$-operator and $\Box(\varphi \to \psi)$ being true in w, if $v \in W$ is any world accessible from w, then $\mathcal{M}, v \vDash \varphi \to \psi$. By the semantics of $\to$, we know this

says that if $\mathcal{M}, v \vDash \varphi$, then also $\mathcal{M}, v \vDash \psi$. Furthermore, we know that $\mathcal{M}, v \vDash \varphi$ is actually the case, because $\Box\varphi$ is also true in w by the second assumption. Therefore we can conclude that $\mathcal{M}, v \vDash \psi$.

As v was an arbitrary world such that $w\ R\ v$, we see by the semantics of $\Box$ that $\mathcal{M}, w \vDash \Box\psi$. This shows that $\mathcal{M}, w \vDash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$. Since w and $\mathcal{M}$ were arbitrary, we can conclude that $\vDash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$. ■

In the chapters about proof theory, we will discover that Axiom **K** plays an important role, as does the following proposition.

Proposition 1.19 — *Principle of necessitation*
If φ is generally valid, then $\Box\varphi$ is generally valid. This implication is called the *Principle of Necessitation.* ◁

Proof. If φ is generally valid, then in any world w of any model $\mathcal{M}$ it holds that $\mathcal{M}, w \vDash \varphi$. But this means that also for any world v such that $w\ R\ v$ it holds that $\mathcal{M}, v \vDash \varphi$, since φ is generally valid. By the semantics of $\Box$ we see that therefore $\mathcal{M}, w \vDash \Box\varphi$ and since w and $\mathcal{M}$ were arbitrary we see that $\Box\varphi$ is generally valid. ■

We will introduce one last notion of validity, namely that of conditional validity. It will be a necessary concept to discuss the soundness and completeness theorems of the proof systems we will develop later on.

Definition 1.20 — *Conditional validity*
Let Γ be a set of modal formulas. We write $\mathcal{M}, w \vDash \Gamma$ if $\mathcal{M}, w \vDash \psi$ for all $\psi \in \Gamma$.

A formula φ is **valid under condition** Γ if for every model $\mathcal{M}$ and every world w the following holds: if $\mathcal{M}, w \vDash \Gamma$, then also $\mathcal{M}, w \vDash \varphi$. We denote this with $\Gamma \vDash \varphi$. If $\Gamma = \{\psi_1, \ldots, \psi_n\}$, we usually write $\psi_1, \ldots, \psi_n \vDash \varphi$ instead of $\{\psi_1, \ldots, \psi_n\} \vDash \varphi$.

Let $\mathcal{C}$ be a frame class. We write $\Gamma \vDash_{\mathcal{C}} \varphi$ when for every model $\mathcal{M}$ **based on a frame** in $\mathcal{C}$ and world w of $\mathcal{M}$, if $\mathcal{M}, w \vDash \Gamma$, then $\mathcal{M}, w \vDash \varphi$. ◁

Example 1.21
In this example we will show that $\Box(p \lor q) \vDash \Box p \lor \Diamond q$. Suppose that $\mathcal{M}$ is a model and w is a world of $\mathcal{M}$. If $\mathcal{M}, w \vDash \Box(p \lor q)$, then in every world v that is reachable from w we have $\mathcal{M}, v \vDash p \lor q$. From Definition 1.12 we know that $\mathcal{M}, v \vDash p \lor q$ is true if and only if $\mathcal{M}, v \vDash p$ or $\mathcal{M}, v \vDash q$.

If there is some world v reachable from w such that $\mathcal{M}, v \vDash q$, we see that $\mathcal{M}, w \vDash \Diamond q$, and thus $\mathcal{M}, w \vDash \Box p \lor \Diamond q$. On the other hand, if such a world v does not exist, then for all worlds v reachable from w we have $\mathcal{M}, v \nvDash q$. Since $\mathcal{M}, v \vDash p$ or $\mathcal{M}, v \vDash q$ must be true, we see by elimination that $\mathcal{M}, v \vDash p$ must hold in all worlds v reachable from w, and thus $\mathcal{M}, w \vDash \Box p$. This also implies that $\mathcal{M}, w \vDash \Box p \lor \Diamond q$.

Therefore, for any model $\mathcal{M}$ and world w of $\mathcal{M}$ we see that $\mathcal{M}, w \vDash \Box(p \lor q)$ implies $\mathcal{M}, w \vDash \Box p \lor \Diamond q$. ◁

1.4 OTHER INTERPRETATIONS OF MODALITY

So far, the interpretation we have worked with, is that $\Box\varphi$ means *"it is necessary that φ"* and $\Diamond\varphi$ means *"it is possible that φ"*. However, one of the strengths of modal logic is that it is a very good framework for other interpretations. In this section we discuss a few of them that are well studied and of importance to artificial intelligence.

Epistemic logic

The word *epistemic* comes from the Greek word ἐπιστήμη (episteme), meaning *knowledge* or *science.* The epistemic interpretation of the box modality $\Box\varphi$, which is usually written as $K\varphi$, is that *"it is known that φ".*

In language, the expressions of this modality do overlap with ways to express alethic modality. For example, although alethically speaking it necessarily either rains or does not rain at a given location and a given time, it is perfectly natural for someone not present at the location to say *"it could be raining or not raining right row"*, expressing an epistemic position of uncertainty. In the event that it is not actually raining, it is an alethic necessity that it does not rain, while it is an epistemic possibility that it does rains.

There is a distinction to be made between alethic and epistemic modality: although both modalities are signified by the same words in many (if not all) languages, the alethic modality refers to an objective truth, whereas the epistemic modality refers to a subjective truth, as it is perceived by the speaker.

Most of the later chapters of this reader will be concerned with epistemic logic.

Doxastic logic

The word *doxastic* comes from the Greek word δοξασία (doxasia), meaning *belief* or *opinion.* The doxastic interpretation of the box modality $\Box\varphi$, which is usually written as $B\varphi$, is that *"it is believed that φ".*

The study of doxastic modality is very closely connected to the study of epistemic modality, and both form the main subject in the field of epistemology. Philosophically the lines between knowledge and belief can be blurry, but it is usual to distinguish *believing φ* from *knowing φ* by knowledge only being possible of true statements. As such, belief can reasonably be interpreted in such a way that belief of falsehoods is possible, while knowledge of falsehoods is not.

Particularly interesting is the study of the dynamics of belief, that is, how belief can be revised based on new information becoming available. One way of studying these dynamics will be treated in one of the later chapters of this reader.

Deontic logic

The word *deontic* comes from the Greek word δέον (deon), meaning something along the lines of *the proper thing / way* or *the right thing / way.* The deontic modality expresses sentences such as *"φ is obligated"* ($\Box\varphi$), *"φ is allowed"* ($\Diamond\varphi$) and *"φ is prohibited"* ($\Box\neg\varphi$).

Next to the epistemic and alethic interpretations, this is yet another interpretation of the modality expressed by words like *can* and *may.* Apart from a purely philosophical or formal significance, the study of deontic logic has its practical use in the studies of law, morality, ethics and society. Often a characteristic of deontic logic is that obligated things are allowed, and as such any $\Box$ operator can be subverted to a $\Diamond$ operator in the deontic modality.

TEMPORAL LOGIC

As the name does suspect, temporal logic is concerned with modalities of time. Temporal logic uses two separate modalities, sometimes denoted as $[P]$ and $[F]$ and their duals $\langle P\rangle$ and $\langle F\rangle$, used to express statements such as:

"at some point in the ***past*** *it was the case that φ"*	$\langle P\rangle\varphi$,
"at some point in the ***future*** *it will be the case that"*	$\langle F\rangle\varphi$,
"it has always been the case that φ	$[P]\varphi$,
"it will always be the case that φ"	$[F]\varphi$.

Temporal logic has its use in a wide variety of fields, for example to give a formal representation of expressions related to time in formal linguistics, or in computing science as a way to describe the temporary states of a computation, useful in program verification.

PROVABILITY LOGIC

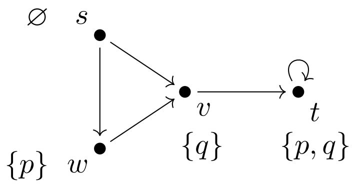
Provability logic is mainly of interest to pure logicians, but as a modal logic, it has some properties that might serve as interesting examples in some parts of this reader.

Provability logic concerns itself with the logic behind consistency proofs. In particular, given some logical system Λ, for example the system of *Peano Arithmetic*, we interpret the box operator $\Box\varphi$ as the statement that φ is provable in Λ. This gives us a means to express the provability of statements within Λ.

One characteristic of provability logic, is that the formula $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ holds true. This formula is a translation of Löb's theorem for Peano arithmetic, and it is commonly known as **Axiom GL**, after the logicians Kurt Gödel and Martin Hugo Löb.

1.5 EXERCISES

Exercise 1.1. Consider the model drawn below:



For each of the following formulas, give all the worlds in which the formula is true:

a) $\Diamond\Diamond p \wedge \Box\neg p$
b) $\Diamond\Box\bot$
c) $\Box\Diamond p \to \Box q$
d) $\Box\neg\Box(p \to q) \wedge \neg\Diamond q$
e) $\Box\Box\Box(\neg q \to \Diamond(\neg q \to \Diamond(\neg q \to \Diamond p)))$

Exercise 1.2. For each of the following formulas, give a model such that the formula is valid on the model.

a) $(\Diamond p \to \Diamond\Diamond q) \wedge (\Diamond q \to \Box p)$

b) $(\Box p \wedge \Diamond\Diamond\neg p) \vee (\Box\neg p \wedge \Diamond\Diamond p) \vee \Diamond(p \to \Box p)$
c) $\Box\Diamond\bot$
d) $\Box(\Box p \to \Diamond q) \to \Diamond(\Diamond p \to \Box q)$

Exercise 1.3. For each of the following formulas, give two different frames on which the formula is valid.
a) $\Box p \to p$
b) $\Box p \to \Diamond p$
c) $\Box\Diamond p \to \Diamond p$
d) $\Box(\Box p \to p) \to \Box p$
e) $\neg\Box\bot$

Exercise 1.4. For each of the following formulas, prove that it is generally valid or give a countermodel to show it is not.
a) $\Diamond(\varphi \wedge \psi) \leftrightarrow (\Diamond\varphi \wedge \Diamond\psi)$
b) $\Box\varphi \vee \Box\psi \to \Box(\varphi \vee \psi)$
c) $\Box(\varphi \vee \psi) \to \Box\varphi \vee \Box\psi$

Exercise 1.5. Show that if $\Diamond\varphi$ is generally valid, then φ is generally valid.

Exercise 1.6. Consider the formula $\varphi = \Box p \wedge \Diamond\Diamond\neg p$.
a) Give a model $\mathcal{M}$ and a world w such that $\mathcal{M}, w \vDash \varphi$.
b) Explain why φ can not be valid on any model.

Exercise 1.7. Suppose we defined $\Diamond\varphi$ as $\neg\Box\neg\varphi$ (as in the duality lemma). Show that the case for $\mathcal{M}, w \vDash \Diamond\varphi$ in Definition 1.12 follows from the definition for $\mathcal{M} \vDash \Box\varphi$.

Exercise 1.8. Consider the formula $\varphi = \Diamond\top \wedge (\Diamond p \vee \Diamond q \to \Box q)$.
a) Give a model $\mathcal{M}$ such that $\mathcal{M} \vDash \varphi$.
b) Explain why φ can not be valid on any frame.

Exercise 1.9. Let φ, ψ be modal formulas and $\mathcal{M}$ be a model based on frame $\mathcal{F} = \langle W, R\rangle$.
a) Suppose that $\mathcal{M}, w \vDash \varphi$ implies $\mathcal{M}, w \vDash \psi$ for every $w \in W$.
Can we conclude that then $\mathcal{M} \vDash \varphi$ implies $\mathcal{M} \vDash \psi$?
b) Suppose that $\mathcal{M} \vDash \varphi$ implies $\mathcal{M} \vDash \psi$.
Can we conclude that then $\mathcal{M}, w \vDash \varphi$ implies $\mathcal{M}, w \vDash \psi$ for every $w \in W$?
c) Suppose that $\mathcal{M} \vDash \varphi$ implies $\mathcal{M} \vDash \psi$ for all models $\mathcal{M}$ based on frame $\mathcal{F}$.
Can we conclude that then $\mathcal{F} \vDash \varphi$ implies $\mathcal{F} \vDash \psi$?

Exercise 1.10. For each of the following statements, give a proof if it is valid or give a counter model if it is not:
a) $\Box\varphi \ \vDash \ \Diamond\varphi$
b) $\neg\Diamond\varphi,\ \neg\Diamond\neg\varphi \vDash \Box\bot$
c) $\Box\varphi,\ \Diamond\psi \ \vDash \ \Box(\varphi \wedge \Diamond\psi)$
d) $\Box\Box\varphi \to \Box\psi \ \vDash \ \Diamond\Diamond\neg\varphi \vee \Diamond\psi$
e) $\varphi \to \Box\varphi,\ \Diamond\varphi \to \Diamond\Diamond\bot \ \vDash \ \neg(\varphi \wedge \Diamond\psi)$

Exercise 1.11. Can the formula $(\Box(\Box p \to p) \to \Box p) \wedge \neg\Box\bot$ be valid on any frame?

2 CHARACTERISABILITY

As we saw near the end of the last chapter, there are many modalities that give rise to a modal logic. All of the systems that were described in the last paragraph can be modelled using Kripke models. The exact interpretation of what a *world* is and what it means to be *accessible* depends on the interpretation of the modality. For example, in temporal logic a world describes a point in time, and the future and past accessibility relations between worlds describes which points in time occur after and before which other points. On the other hand, in epistemic or doxastic logic the accessible worlds are those situations that cannot be distinguished using the knowledge (in case of epistemic) or beliefs (in case of doxastic) one has in the current world.

However, not every possible Kripke model makes sense for every interpretation of the modal operators. For example, in temporal logic, we do not want a point in the future to access a point in the past with the future accessibility relation. In doxastic logic we would want it to be possible to have a false belief, and thus that the actual world does not access itself, whereas we do not want this to happen in epistemic logic, as we generally interpret knowledge to be truthful.

Therefore it makes sense to study which Kripke models are sensible for a given logic and how to describe such models. In this chapter we will give two methods to describe such collections of models. The first way is to describe properties that the accessibility relation and the set of worlds have to satisfy. The second way is to describe collections of models, is by taking the collection of all models in which a given modal formulas is valid. Such collections are thus described by a modal formula.

Some collections of models can only be described by a first-order description, while other collections can only be described by a modal formula. There are also collections that can be described using both methods. That is, there are collections of models with a first-order description that consist of exactly those models in which a certain modal formula is true, hence in such cases the two methods of describing a collection of models correspond to the same collection of models. The proofs of such correspondences are called **correspondence proofs** and will play a major part in the current and the next two chapters.

2.1 FIRST-ORDER DEFINABLE FRAME CLASSES

We start with the method of describing collections of models by describing their shape using a first-order formula. We have already seen one of these descriptions before, in Example 1.7: in some models the relation has the property of being transitive. This means that whenever a world w accesses a world v and v in turn accesses some world u, then w accesses the world u as well. This is a property that can be described using a first order formula:

$$\forall w \forall v \forall u (w \; R \; v \wedge v \; R \; u \rightarrow w \; R \; u).$$

We then say that a frame $\mathcal{F} = \langle W, R \rangle$ is transitive when it satisfies this formula if the quantifiers range over the set of worlds W. In other world, $\mathcal{F}$ is transitive if the above formula states a truth about any worlds $w, v, u \in W$ and the accessibility relation R. Logically we say that the pair $\langle W, R \rangle$ is a first-order model for the formula above.

The formula above **defines** a class of frames, namely the class of transitive frames.

Definition 2.1 — *First-order definable*
A frame class $\mathcal{C}$ is **first-order definable** if it is definable by a sentence in the first-order language with relation symbol R that quantifies over the set of worlds W. ◁

Needless to say, there are a lot of first-order formulas that can be used to describe a frame class. To give an overview of some of the most important properties, we give the following list of examples.

Example 2.2 — *First-order properties*
The following are examples of common first-order sentences that define frame classes:

$\langle W, R \rangle$ is ...

... **reflexive**	iff	$\forall w (w \; R \; w)$
... **irreflexive**	iff	$\forall w (w \not R \; w)$
... **coreflexive**	iff	$\forall w \forall v (w \; R \; v \to w = v)$
... **transitive**	iff	$\forall w \forall v \forall u (w \; R \; v \land v \; R \; u \to w \; R \; u)$
... **(right) Euclidean**	iff	$\forall w \forall v \forall u (w \; R \; v \land w \; R \; u \to v \; R \; u)$
... **left Euclidean**	iff	$\forall w \forall v \forall u (w \; R \; v \land u \; R \; v \to w \; R \; u)$
... **dense**	iff	$\forall w \forall v (w \; R \; v \to \exists u (w \; R \; u \land u \; R \; v))$
... **symmetric**	iff	$\forall w \forall v (w \; R \; v \to v \; R \; w)$
... **asymmetric**	iff	$\forall w \forall v (w \; R \; v \to v \not R \; w)$
... **antisymmetric**	iff	$\forall w \forall v (w \; R \; v \land v \; R \; w \to w = v)$
... **serial**	iff	$\forall w \exists v (w \; R \; v)$
... **inverse serial**	iff	$\forall w \exists v (v \; R \; w)$
... **universal**	iff	$\forall w \forall v (w \; R \; v)$
... **totally disconnected**	iff	$\forall w \forall v (w \not R \; v)$
... **connected**	iff	$\forall w \forall v (w \; R \; v \lor v \; R \; w)$
... **semi-connected**	iff	$\forall w \forall v (w \neq v \to w \; R \; v \lor v \; R \; w)$
... **piecewise connected**	iff	$\forall w \forall v \forall u ((u \; R \; w \land u \; R \; v) \to (w \; R \; v \lor v \; R \; w))$
... **deterministic**	iff	$\forall w \forall v \forall u (w \; R \; v \land w \; R \; u \to v = u)$
... **convergent**	iff	$\forall w \forall v \forall u (w \; R \; v \land w \; R \; u \to \exists z (v \; R \; z \land u \; R \; z))$
... **preordered**	iff	it is **reflexive** and **transitive**
... an **equivalence**	iff	it is **preordered** and **symmetric**
... **partially ordered**	iff	it is **preordered** and **antisymmetric**
... **totally ordered**	iff	it is **partially ordered** and **connected**

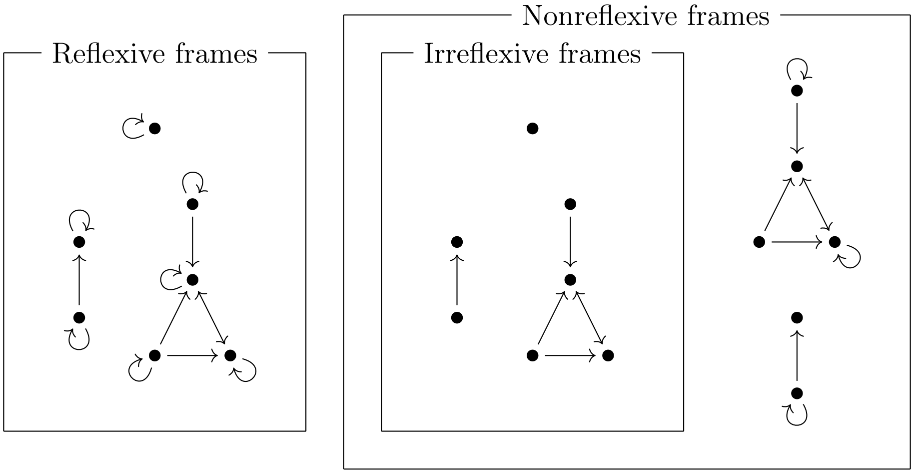
◁

Not all of these properties are equally interesting from the perspective of modal logic. The ones that will pop-up most frequently throughout this course are *reflexive*, *transitive*, *Euclidean*, *symmetric* and *serial* frames, as well as the frames defined by combinations of these properties. Most of the other properties are written here to serve as illustration for some (counter)examples of modally definable frames and to be used in exercises in this and in later chapters.

Before we move on, we will give a few diagrams to illustrate what the most important properties mentioned above actually look like.

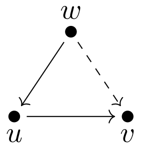
Reflexive and Irreflexive Frames

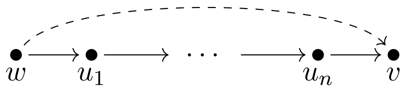
Reflexive frames are defined as frames in which every world can access itself. To contrast this, an irreflexive frame is a frame in which no world accesses itself. Not every frame has to be reflexive or irreflexive, as there could be frames that contain some worlds that access themselves, and some other worlds that do not. A frame that is not reflexive is called *nonreflexive*. Every irreflexive frame is nonreflexive, but not every nonreflexive frame has to be irreflexive.



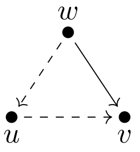
Transitive, Dense and Euclidean Frames

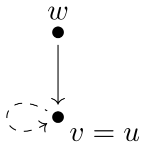
A transitive frame is a frame where a world w accesses any world v that could be reached by a finite path from w. In particular, if you have the relation $w\ R\ u\ R\ v$, then there is also a relation $w\ R\ v$. As a diagram, this could be seen as the dotted line being implied by solid lines:

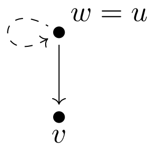




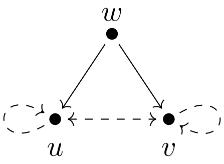
A dense frame is a frame where w can access an intermediate world u for every world v that is accessible from w, such that u also accesses v. Where transitivity says that any route of multiple steps can be taken in a single step, denseness says that any route taking a single step can be taken in multiple steps. It is however important to note that the intermediate world in a dense frame does not have to be a *different* world than w or v. As a diagram:







A (right) Euclidean frame is a frame where any two worlds accessible by w are also accessible by each other. We usually call right Euclidean frames just Euclidean frames, since left Euclidean frames are not very useful in modal logic (as we will discover in Exercise 4.4d). As with dense frames, we should not forget to look at the case where $u = v$: if a world is accessible from w in a Euclidean frame, then it is accessible from itself.



Symmetric, Asymmetric and Antisymmetric Frames

A symmetric frame is a frame where every accessibility relation can be inverted. For any arrow going in one direction, there is an arrow going in the opposite direction, hence any path can be traversed back in the opposite directon. A symmetric relation has the following implication, and is usually drawn with a double arrow:

An asymmetric frame is a frame where none of the accessibility relations has an inverse. In particular, this means that asymmetric frames are irreflexive, since if $w\ R\ v$ and $w = v$, then also $v\ R\ w$. As a diagram, this could be recognised by the lack of double arrows and reflexive arrows.

An antisymmetric frame is a frame where the only relations with an inverse are reflexive relations. In other words, if w and v can access each other, then $w = v$. It could be interpreted as a frame that is asymmetric only between any two *different* worlds w and v.

Serial, Totally Disconnected and Universal frames

A serial frame is a frame in which every world can reach another world. In other words, there are no blind worlds in the frame. Every reflexive frame is serial, as are the frames from Example 1.5 and Example 1.7.

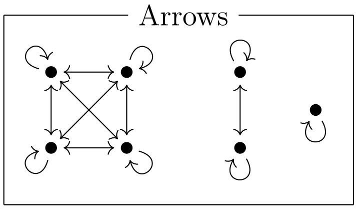
Totally disconnected frames are frames in which every world is blind. In such a frame we have $R = \varnothing$.

Universal frames are frames in which any two worlds can access each other. In these frames we have that $R = W \times W$. Universal frames are another example of serial frames.

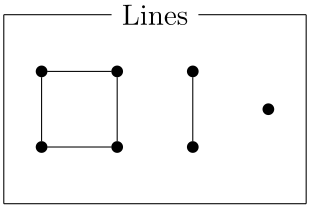
Equivalence Frames

Equivalence frames are frames that consist of clusters of worlds, such that each cluster in itself forms a universal frame. Another way to view equivalence frames is to create a partition of the set of worlds, and connect all the worlds inside the same part with each other. Equivalence frames satisfy a lot of properties, such as being reflexive, transitive, dense, Euclidean, symmetric, serial and piecewise connected.

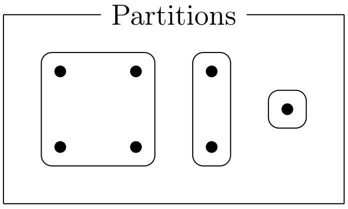
Since there are many arrows to be drawn in an equivalence frame, it is common to leave out the arrows that are not necessary, and simply mention that the frame is an equivalence frame. Another common practice is to draw lines instead of arrows, since each arrow is symmetric. A third way to draw equivalence frames, is by drawing the partitioning of the set of worlds as circles around worlds instead. Below are these three depictions drawn for the same equivalence frame:



Arrows



Lines



Partitions

2.2 Modal Definability

We are now ready to look at the second way to describe a frame class. Where first-order definitions of frame classes give a description of the shape of the frame, a modally definable frame class consists of those frames that make a certain set of formulas valid on the frame. As such, we could see the first-order definable frame classes as a definition using an outside view of what a model is supposed to look, whereas modally definable frame classes use an inside view to describe the frame, by describing what the model looks like from the perspective of the worlds in the model, as described by the validity of the formulas.

Definition 2.3 — *Modally definable*
A frame class $\mathcal{C}$ is **modally definable** if there is a set of modal formulas Φ such that $\mathcal{F} \in \mathcal{C}$ if and only if $\mathcal{F} \vDash \varphi$ for all $\varphi \in \Phi$. If Φ is such a set of modal formulas that defines $\mathcal{C}$, then we say Φ **characterises** the frame class $\mathcal{C}$. ◁

In a way, we could say that the a set of formulas Φ characterises a frame class $\mathcal{C}$ if the modal operators behave exactly as prescribed by the formulas in Φ, when we restrict our attention to only the frames in $\mathcal{C}$.

For example, if we interpret the modal operators $[F]$ and $\langle F\rangle$ as the "future" operators of temporal logic, then $\langle F\rangle\varphi$ means that φ holds at some moment in the future, while $[F]\varphi$ means that φ holds for any moment in the future. From intuition, we want $[F]\varphi \to \langle F\rangle\varphi$ to be valid under this interpretation: if something is true for any moment in the future, then certainly there exists some point in the future in which it is true. We say that $[F]\varphi \to \langle F\rangle\varphi$ is a characterisation of a property we expect of a temporal logic. Therefore the frames that are reasonable under the temporal interpretation are characterised by this formula.

Theorem 2.4 — *Frame correspondence*
Let $\mathcal{F}$ be a frame, then:

$\mathcal{F}$ is ...

... reflexive	iff	$\mathcal{F} \vDash \Box\varphi \to \varphi$	for all φ
... coreflexive	iff	$\mathcal{F} \vDash \Diamond\varphi \to \varphi$	for all φ
... transitive	iff	$\mathcal{F} \vDash \Box\varphi \to \Box\Box\varphi$	for all φ
... dense	iff	$\mathcal{F} \vDash \Diamond\varphi \to \Diamond\Diamond\varphi$	for all φ
... serial	iff	$\mathcal{F} \vDash \Box\varphi \to \Diamond\varphi$	for all φ
... deterministic	iff	$\mathcal{F} \vDash \Diamond\varphi \to \Box\varphi$	for all φ
... symmetric	iff	$\mathcal{F} \vDash \varphi \to \Box\Diamond\varphi$	for all φ
... Euclidean	iff	$\mathcal{F} \vDash \Diamond\varphi \to \Box\Diamond\varphi$	for all φ
... convergent	iff	$\mathcal{F} \vDash \Diamond\Box\varphi \to \Box\Diamond\varphi$	for all φ
... piecewise connected	iff	$\mathcal{F} \vDash \Box(\Box\varphi \to \psi) \vee \Box(\Box\psi \to \varphi)$	for all φ and ψ
... totally disconnected	iff	$\mathcal{F} \vDash \Box\varphi$	for all φ ◁

We will only prove the theorem for *reflexive* frames, *transitive* frames and *totally disconnected* frames, leaving the others as Exercise 2.1.

Proof of reflexive correspondence. We have to show that $\mathcal{F}$ is reflexive if and only if $\mathcal{F} \vDash \Box\varphi \to \varphi$ for all φ.

($\Rightarrow$) Let $\mathcal{F} = \langle W, R \rangle$ be a reflexive frame. By the definition of a reflexive frame, we have that for any world $w \in W$ it holds that $w\ R\ w$. Let $\mathcal{M}$ be any model based on $\mathcal{F}$ and φ be any formula. If $\mathcal{M}, w \vDash \Box\varphi$, then φ is true in all worlds v such that $w\ R\ v$. In particular, since $w\ R\ w$, we see that $\mathcal{M}, w \vDash \varphi$. So we see that $\mathcal{M}, w \vDash \Box\varphi$ implies $\mathcal{M}, w \vDash \varphi$. Therefore $\mathcal{M}, w \vDash \Box\varphi \to \varphi$.

($\Leftarrow$) Let $\mathcal{F} = \langle W, R \rangle$ not be reflexive, then not for all[*] the worlds $w \in W$ does it hold that $w\ R\ w$. Therefore, let v be such a world with $v \not R\ v$. We define a model $\mathcal{M}$ based on $\mathcal{F}$ with valuation V such that $V(x) = \{p\}$ if and only if $v\ R\ x$. In other words, p is true in every world reachable from v, and false in any world not reachable from v. Since we chose v such that $v \not R\ v$, we see that $\mathcal{M}, v \nvDash p$. Yet, since every world reachable from v makes p true, it also holds that $\mathcal{M}, v \vDash \Box p$. Therefore $\mathcal{M}$ is a counter model for $\Box\varphi \to \varphi$, since $\mathcal{M}, v \nvDash \Box p \to p$. ■

Proof of transitive correspondence. We have to show that $\mathcal{F}$ is transitive if and only if $\mathcal{F} \vDash \Box\varphi \to \Box\Box\varphi$ for all φ.

($\Rightarrow$) Let $\mathcal{F}$ be a transitive frame and let $\mathcal{M}$ be a model based on $\mathcal{F}$ with $w \in W$ a world and φ any formula. Suppose that $\mathcal{M}, w \vDash \Box\varphi$, then in every world v such that $w\ R\ v$ we have that $\mathcal{M}, v \vDash \varphi$. If there is a world u such that $v\ R\ u$, then by transitivity we see that $w\ R\ v$ and $v\ R\ u$ implies that $w\ R\ u$, and thus $\mathcal{M}, u \vDash \varphi$ as well. Therefore, for any world u reachable from v we have $\mathcal{M}, u \vDash \varphi$, which implies that $\mathcal{M}, v \vDash \Box\varphi$. Since this holds for every world v such that $w\ R\ v$, we see that $\mathcal{M}, w \vDash \Box\Box\varphi$. Therefore $\mathcal{M}, w \vDash \Box\varphi \to \Box\Box\varphi$.

($\Leftarrow$) Let $\mathcal{F}$ be a frame that is not transitive. Then $\forall w \forall v \forall u (w\ R\ v \wedge v\ R\ u \to w\ R\ u)$ is false. The negation of this sentence is $\exists w \exists v \exists u (w\ R\ v \wedge v\ R\ u \wedge \neg w\ R\ u)$, thus $\mathcal{F}$ contains world w, v, u such that $w\ R\ v$ and $v\ R\ u$, but also $w \not R\ u$.[†] We define a valuation V on $\mathcal{F}$ such that $V(x) = \{p\}$ if and only if $w\ R\ x$ and let $\mathcal{M}$ be the model based on $\mathcal{F}$ with valuation V. We then see that $\mathcal{M}, w \vDash \Box p$, since p is true in any world x reachable from w, and we see that $\mathcal{M}, u \nvDash p$, since $w \not R\ u$. But since $v\ R\ u$, we see that $\mathcal{M}, v \nvDash \Box p$, and thus that $\mathcal{M}, w \nvDash \Box\Box p$. Therefore $\mathcal{M}, w \nvDash \Box\varphi \to \Box\Box\varphi$, since $\mathcal{M}, w \nvDash \Box p \to \Box\Box p$. ■

Proof of totally disconnected correspondence. We have to show that $\mathcal{F}$ is totally disconnected if and only if $\mathcal{F} \vDash \Box\varphi$ for all φ.

($\Rightarrow$) Let $\mathcal{F} = \langle W, R \rangle$ be totally disconnected, and let $\mathcal{M}$ be a model based on $\mathcal{F}$ with $w \in W$ a world and φ any formula. Since $w \not R\ v$ holds for all worlds $v \in W$ (including $v = w$), we see that the set of worlds accessible from w is empty. Therefore it is vacuously true that φ is true in all the worlds accessible from w, because there are no such worlds. Hence $\mathcal{M}, w \vDash \Box\varphi$ and thus $\mathcal{F} \vDash \Box\varphi$.

($\Leftarrow$) Let $\mathcal{F} = \langle W, R \rangle$ not be totally disconnected, then $\neg \forall w \forall v (w \not R\ v)$ holds, and this is equivalent to $\exists w \exists v (w\ R\ v)$. Therefore let $w, v \in W$ such that $w\ R\ v$, and let

---

[*] N.B.: this is something different than that $w \not R w$ for all worlds $w \in W$!

[†] note that the worlds w, v and u need not necessarily be three *different* worlds

$\mathcal{M} = \{W, R, V\}$ be any model based on $\mathcal{F}$ such that $p \notin V(v)$. If we take $\varphi := p$ as a formula, then $\mathcal{M}, w \nvDash \Box p$, since $\mathcal{M}, v \nvDash p$ and $w\ R\ v$. Therefore $\mathcal{F} \nvDash \Box\varphi$. ■

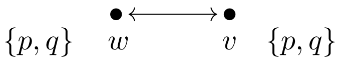
Note that all of the previous proofs follow the same structure. In each proof we have to show two directions, namely that a first-order frame property implies the validity of a modal formula and that a the validity of a modal formula implies a first-order frame property.

In the first direction, we take an *arbitrary* world in an *arbitrary* frame that has the frame property, and then show that in an *arbitrary* model the formula is valid in this world. It is important to note that we cannot make any assumptions about the frame, model and world apart from that they satisfy the first-order frame property. This is because we are essentially proving a universal statement: *every* frame with the first-order property makes the modal formula valid.

In the second direction it is easier to prove the contrapositive. That is, we take an *arbitrary* frame that does *not* satisfy the first-order property, and show that there *exists* a world and a model such that the modal formula is *not* valid in this world under this model. Here, we have once again a universal statement, namely that *every* frame not satisfying the first-order property also does not satisfy the modal formula. However, the statement that a model formula is not valid on a frame, means that we have to prove an existential statement: there *exists* some world in the frame and there *exists* some model based on the frame, such that in this model there is an instance of the formula that is false in the world. To summarise: for every frame that does not satisfy the first-order property we have to give a single counter model.

It is important to note that characterisation works on the level of frames, not models. The following example highlights this distinction.

Example 2.5
The formula $\Box\varphi \to \varphi$ is valid on any model based on a reflexive frame, as we have proved above. However, there are models of nonreflexive frames that make $\Box\varphi \to \varphi$ valid:



We will spend the next chapter to introduce a method for proving that $\Box\varphi \to \varphi$ is indeed valid on the model above for any formula φ, so we will come back to this in Example 3.7. For now, though, it is enough to note that there are other models based on this frame that do not make $\Box\varphi \to \varphi$ true, such as the following model:

We see that in this second model, $\Box p \to p$ does not hold in world w, since $w \nvDash p$, yet the only world reachable from w, namely v, makes p true, and thus is $w \vDash \Box p$ as well. ◁

Before we move on to the last section, we will present a way to combine several modally definable frame classes. This gives us a way to see that also the classes of preordered frames and equivalence frames are modally definable. The proof is left as Exercise 2.3.

Proposition 2.6 — *Combination of frame classes*
If $\mathcal{C}_1$ and $\mathcal{C}_2$ are frame classes that are modally definable, then $\mathcal{C}_1 \cap \mathcal{C}_2$ is modally definable. ◁

Theorem 2.7 — *Frame correspondence*
Let $\mathcal{F}$ be a frame, then:

$$\begin{array}{rlll} \mathcal{F} \text{ is ...} & & & \\ \text{... preordered} & \text{iff} & \mathcal{F} \vDash \Box\varphi \to \varphi \quad \text{and} & \\ & & \mathcal{F} \vDash \Box\varphi \to \Box\Box\varphi & \text{for all } \varphi \\ \text{... an equivalence frame} & \text{iff} & \mathcal{F} \vDash \Box\varphi \to \varphi, & \\ & & \mathcal{F} \vDash \Box\varphi \to \Box\Box\varphi \quad \text{and} & \\ & & \mathcal{F} \vDash \Diamond\Box\varphi \to \varphi & \text{for all } \varphi \end{array}$$

◁

2.3 Modally definable but not first-order definable

We now have two ways to define a frame class: with a first-order property or with a modal formula. We have also seen that several frame classes are described both by a first-order property and by a modal formula. The big question is of course if *every* frame class that is first-order definable is also modally definable. It turns out that this is not the case, as we will see in a later chapter. It also turns out that the reverse of this statement is false. In other words, not every modally definable frame class is first-order definable. We will give one example of such a frame class in the remainder of this section.

Definition 2.8 — *Conversely well-foundedness*
A frame $\langle W, R \rangle$ is **conversely well-founded** if for every nonempty subset $U \subseteq W$ there is a world $w \in U$ such that $w \not R\, v$ for any $v \in U$. ◁

One intuitive way to see conversely well-founded frames is that it is impossible to find an infinite path $w_0 \ R \ w_1 \ R \ w_2 \ R \ \cdots$ in the frame: for any world $w \in W$ there is a maximum finite distance that can be travelled by following the accessibility relation. An example of a conversely well-founded frame is the frame $\langle \mathbb{N}, > \rangle$, where the worlds are natural numbers and i reaches j if and only if $i > j$. Since there are only finitely many natural numbers smaller than any number, we see that this is indeed a conversely well-founded relation. In fact, this is even an example of a **transitive** conversely well-founded frame.

Proposition 2.9
Conversely well-founded frames are not first-order definable. Furthermore, transitive conversely well-founded frames are not first-order definable. ◁

The proof of this proposition lies beyond the scope of this course and requires some understanding of model theory. For those interested, it is a basic consequence of the compactness theorem for first-order logic.

Theorem 2.10 — *Frame correspondence*
$\mathcal{F}$ is transitive and conversely well-founded if and only if $\mathcal{F} \vDash \Box(\Box\varphi \to \varphi) \to \Box\varphi$. ◁

> *Proof.* ($\Rightarrow$) Let $\mathcal{M} = \langle W, R, V \rangle$ be a model based on a transitive conversely well-founded frame and let $w \in W$. We first assume that $\mathcal{M}, w \vDash \Box(\Box\varphi \to \varphi)$, then we wish to prove that $\mathcal{M}, w \vDash \Box\varphi$. If $\mathcal{M}, v \vDash \varphi$, for every v such that $w \ R \ v$, then there is nothing to prove, since $\mathcal{M}, w \vDash \Box\varphi$. So let's make as a second assumption that $\mathcal{M}, v \nvDash \varphi$ for some v such that $w \ R \ v$.
>
> We will show that we can find a subset $U \subseteq W$ such that there is no world $x \in U$ with $x \not R\, y$ for any $y \in U$, by using this second assumption. This implies that $\langle W, R \rangle$

is not conversely well-founded, which gives us a contradiction. Therefore the second assumption cannot be true, which means that $\mathcal{M}, v \vDash \varphi$ for all v such that $w\ R\ v$.

The subset $U := \{v \in W \mid w\ R\ v \land \mathcal{M}, v \nvDash \varphi\}$ is the subset we need. By the second assumption U is nonempty, so let $v \in U$. Then the first assumption implies that $\mathcal{M}, v \vDash \Box\varphi \to \varphi$. Since also $\mathcal{M}, v \nvDash \varphi$, we see that $\mathcal{M}, v \nvDash \Box\varphi$. This means there is some world v' such that $v\ R\ v'$ and $\mathcal{M}, v' \nvDash \varphi$. But R is transitive, so $w\ R\ v$ and $v\ R\ v'$ implies that $w\ R\ v'$. Therefore $v' \in U$, since it satisfies the definition of U. This means we can find an element $v' \in U$ such that $v\ R\ v'$ for any $v \in U$, meaning that U breaks the conversely well-foundedness of $\langle W, R\rangle$.

($\Leftarrow$) We leave this direction as Exercise 2.9. ■

This formula $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ from the theorem is called the Gödel-Löb formula, and it is a validity in provability logic.

As we saw in the theorem, the frame class of transitive conversely well-founded frames is modally definable. Interestingly enough the frame class of just the conversely well-founded frames is *not* modally definable (and by Proposition 2.9 it is not first-order definable either), so this gives us an example of a frame class that satisfies neither of the two definable notions. A proof of this fact goes beyond the scope of this course.

2.4 EXERCISES

Exercise 2.1. Complete the proof of Theorem 2.4 by giving correspondence proofs for the following frame classes

a) Dense frames
b) Euclidean frames
c) Serial frames
d) Deterministic frames
e) Symmetric frames
f) Convergent frames
g) Coreflexive frames
h) Piecewise connected frames

Exercise 2.2. Let $\mathcal{F} = \langle W, R\rangle$ be a frame. Prove that the following are equivalent:

(1) $\mathcal{F}$ is an equivalence frame.
(2) $\mathcal{F}$ is reflexive and Euclidean.
(3) For any world $w \in W$ if we take a frame $\mathcal{F}' = \langle W', R'\rangle$ with $W' = \{v \in W \mid w\ R\ v\} \cup \{w\}$ and R' is the restriction of R onto W', then $\mathcal{F}'$ is a universal frame.
(4) $\mathcal{F}$ is serial, transitive and symmetric.

Exercise 2.3. Finish the proof of Proposition 2.6. Start with taking two sets Φ_1 and Φ_2 characterising $\mathcal{C}_1$ and $\mathcal{C}_2$, then we claim that $\Phi_1 \cup \Phi_2$ characterises $\mathcal{C}_1 \cap \mathcal{C}_2$.

Exercise 2.4. Let the frame class defined by the formula $\Box^n\varphi \to \varphi$ be called **n-reflexive**. Give a first-order formula that describes this frame class and prove their correspondence.

Exercise 2.5. Let $\pi : \mathrm{Fml}_{\mathrm{ML}} \to \mathrm{Fml}_{\mathrm{ML}}$ be a function that translates modal formulas by replacing every occurrence of a $\Box$ by a $\Diamond$ and vice versa. For example, $\pi(\Box\Diamond q \to \Diamond\neg p) = \Diamond\Box q \to \Box\neg p$.

Let $\psi_1 = \triangle_1, \ldots, \triangle_n\varphi$ and $\psi_2 = \triangle_{n+1}, \ldots, \triangle_{n+m}\varphi$ for some $n, m \in \mathbb{N}$ and each $\triangle_i \in \{\Box, \Diamond\}$. Prove that a frame class $\mathcal{C}$ is defined by $\psi_1 \to \psi_2$ if and only if it is defined by $\pi(\psi_2) \to \pi(\psi_1)$.

Exercise 2.6. Let $\mathcal{F} = \langle W, R\rangle$ be a Euclidean frame with $W = \{w_0, w_1, w_2, \dots\} \cup \{v_0, v_1, \dots\}$ such that for every natural number $i \in \mathbb{N}$ we have $w_i \; R \; v_i$ and $v_i \; R \; w_{i+1}$.

a) Prove that $v_i \; R \; v_j$ for any natural numbers i and j.
(Hint: use induction on the difference between i and j)

b) Suppose that the frame described above was not Euclidean, but symmetric and transitive. Prove that it is universal.

Exercise 2.7. For a relation R, define the inverse R^{-1} such that $\langle w, v\rangle \in R$ if and only if $\langle v, w\rangle \in R^{-1}$.

a) Prove that R is symmetric if and only if $R = R^{-1}$

b) Prove that R is asymmetric if and only if $R \cap R^{-1} = \varnothing$

c) Prove that R is antisymmetric if and only if $R \cap R^{-1}$ is coreflexive.

∗ **Exercise 2.8.** Let $\mathcal{C}$ be a frame class defined by the following property:

$$\forall w \forall u \forall v (w \; R \; v \wedge w \; R \; u \wedge u \neq v \to v \; R \; u \vee u \; R \; v).$$

Prove that $\mathcal{C}$ is characterised by $\Diamond\varphi \wedge \Diamond\psi \to \Diamond(\varphi \wedge (\psi \vee \Diamond\psi)) \vee \Diamond(\psi \wedge (\varphi \vee \Diamond\varphi))$.

∗ **Exercise 2.9.** In this exercise we will finish the ($\Leftarrow$)-direction of the proof of Theorem 2.10. We will set up the framework of the proof and leave the details as the exercise. Let $\mathcal{F} = \langle W, R\rangle$ not be transitive conversely well-founded, then $\mathcal{F}$ is either transitive, but not conversely well-founded, or $\mathcal{F}$ is not transitive.

Suppose $\mathcal{F}$ is transitive, but not conversely well-founded, then there is a subset $U \subseteq W$ such that for all $v \in U$ there is a $v' \in U$ such that $v \; R \; v'$. Let $w \in U$ be any world and let $\mathcal{M} = \langle W, R, V\rangle$ be a model with $p \in V(x)$ if and only if $w \not R \, x$ or $x \notin U$.

a) Show that $\mathcal{M}, w \not\models \Box p$ and $\mathcal{M}, w \models \Box(\Box p \to p)$. For the latter, consider two cases for v with $w \; R \; v$: one where $v \in U$ and one where $v \notin U$.

On the other hand, suppose $\mathcal{F}$ is not transitive, then there are worlds $w, v, u \in W$ such that $w \; R \; v \; R \; u$ and $w \not R \, u$. Define a model $\mathcal{M} = \langle W, R, V\rangle$ such that $p \in V(x)$ if and only if $x \neq v$ and $x \neq u$.

b) Show that $\mathcal{M}, w \not\models \Box(\Box p \to p) \to \Box p$.

Exercise 2.10. Call $\forall w \exists v (w \; R \; v \wedge \forall u (v \; R \; u \to v = u))$ the **McKinsey condition**.

a) Show that if a frame $\mathcal{F}$ satisfies the McKinsey condition, then $\mathcal{F} \models \Box\Diamond\varphi \to \Diamond\Box\varphi$.

b) Show that there is a frame $\mathcal{F}$ such that $\mathcal{F} \models \Box\Diamond\varphi \to \Diamond\Box\varphi$ and $\mathcal{F}$ does **not** satisfy the McKinsey condition. Hence if we look at *all* Kripke frames, then $\Box\Diamond\varphi \to \Diamond\Box\varphi$ does *not* characterise the frames that satsfy the McKinsey condition. In fact the class of frames characterised by $\Box\Diamond\varphi \to \Diamond\Box\varphi$ is not first-order definable.

∗ **Exercise 2.11.** This is a continuation of the last exercise. We're going to show that if $\mathcal{F}$ is transitive and $\mathcal{F} \models \Box\Diamond\varphi \to \Diamond\Box\varphi$, then $\mathcal{F}$ satisfies the McKinsey condition. Therefore if we look just at the transitive frames, then $\Box\Diamond\varphi \to \Diamond\Box\varphi$ *does* characterises the frames that satisfy the McKinsey condition.

Let $\mathcal{F}$ not satisfy the McKinsey condition.

a) Show that there is a world w such that any world accessible from w can access at least two different worlds.

For this exercise, let's assume the set of worlds reachable from w is countable (the general case can be proved in a similar way using a stronger kind of induction, namely transfinite induction). We also assume w is not blind, since if w were blind, then trivially $\Box\Diamond\varphi \to \Diamond\Box\varphi$ would be false in w.

Let every world reachable from w be equal to v_i for some $i \in \mathbb{N}$ (possible equal to multiple v_i, if w only reaches finitely many worlds). Go through all the v_i's and create sets $X_i \subseteq W$ and $Y_i \subseteq W$ for every $i \in \mathbb{N}$ as follows:

1. Let $X_0 = Y_0 = \varnothing$.
2. if v_i can not access any world in X_i or in Y_i, take two distinct worlds x and y accessible from v_i and let $X_{i+1} = X_i \cup \{x\}$ and $Y_{i+1} = Y_i \cup \{y\}$.
3. if v_i can access a world $x \in X_i$ and a world $y \in Y_i$, let $X_{i+1} = X_i$ and $Y_{i+1} = Y_i$.
4. if $v_i \; R \; x$ for some $x \in X_i$ but there is no $y \in Y_i$ with $v_i \; R \; y$, then choose a y such that $y \notin X_i$ and $x \; R \; y$. Let $X_{i+1} = X_i$ and $Y_{i+1} = Y_i \cup \{y\}$.
5. if $v_i \; R \; y$ for some $y \in Y_i$ but there is no $x \in X_i$ with $v_i \; R \; x$, then choose an x such that $x \notin Y_i$ and $y \; R \; x$. Let $X_{i+1} = X_i \cup \{x\}$ and $Y_{i+1} = Y_i$.

Let $X = \bigcup_{i \in \mathbb{N}} X_i$ and $Y = \bigcup_{i \in \mathbb{N}} Y_i$.

b) The fourth and fifth step can only be well-defined if at stage i of the above construction, when v_i accesses some $x \in X_i$, then v_i accesses some $y \notin X_i$. Prove that this is the case. (Hint: if v_i can reach $x \in X_i$ and v_i does not reach any world outside X_i, then x does not reach any world outside X_i. Let $y \in X_i$ be such world with $x \; R \; y$ and consider the stages in which x and y were added to the set. Show that one of them can not have been added earlier than the other. Use induction to show that this gives a contradiction.)

c) Prove that every world accessible from w can access a world in X and a world in Y.

d) Prove that $X \cap Y = \varnothing$.

e) Give a model $\mathcal{M} = \langle W, R, V \rangle$ based on $\mathcal{F}$ such that $\mathcal{M}, w \nVdash \Box\Diamond\varphi \to \Diamond\Box\varphi$.

3 BISIMULATION

As we have stated in the last chapter, not every frame class is modally definable. The next chapter will be devoted to methods for proving that a frame class can not be modally defined. The idea behind such methods is that two different models can appear to be completely the same from the perspective of the worlds in the model. Such models will make the same formulas true and are equivalent to one another from a semantical viewpoint. This notion is made precise with the definition of a **bisimulation**: a relation that links to models together in such a way that they are modally indistinguishable from each other.

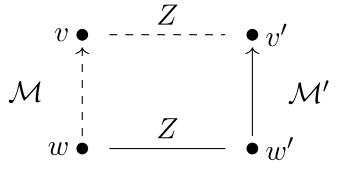
Definition 3.1 — *Bisimulation*
Let $\mathcal{M} = \langle W, R, V \rangle$ and $\mathcal{M}' = \langle W', R', V' \rangle$ be models. A **bisimulation** between $\mathcal{M}$ and $\mathcal{M}'$ is a relation $Z \subseteq W \times W'$ such that:

- (**atomic**) $V(w) = V'(w')$ for all $w \in W$ and $w' \in W'$ such that $w \; Z \; w'$,
- (**back**) $w' \; R' \; v'$ and $w \; Z \; w'$ implies that there is a $v \in W$ such that $w \; R \; v$ and $v \; Z \; v'$,
- (**forth**) $w \; R \; v$ and $w \; Z \; w'$ implies that there is a $v' \in W'$ such that $w' \; R' \; v'$ and $v \; Z \; v'$.

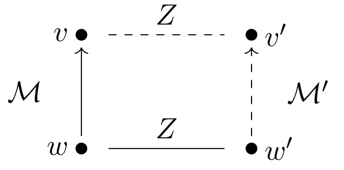
If for some $w \in W$ and $w' \in W'$ there is a bisimulation Z such that $w \; Z \; w'$, then we say w and w' are bisimilar, denoted as $\mathcal{M}, w \;\underline{\leftrightarrow}\; \mathcal{M}', w'$. $\lhd$

This may seem like a very abstract definition, so we will present a useful graphical representation of what a bisimulation is. Let w be a world in a model $\mathcal{M}$ and w' be a world in a model $\mathcal{M}'$. Then w is bisimilar to w' if the three bullet points in the definition are true. The **atomic** condition is easiest, and simply states that the valuation of w and w' must be the same. This makes sense: if we want w and w' to be indistinguishable, then surely we want them to have the same valuation.

The **back** and **forth** conditions, which are also sometimes called **zig** and **zag**, and can be represented by the following diagrams:



back



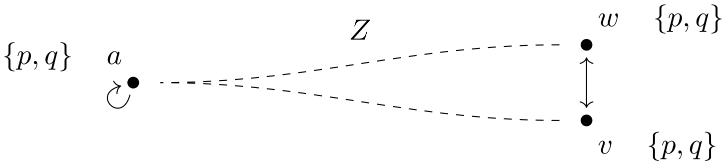
forth

The antecedent of the implication in the back and forth conditions are represented by the solid lines, while the consequent is represented by the dashed lines. For example, if the relations $w \; Z \; w'$ and $w' \; R' \; v'$, drawn with solid lines, in the back diagram exist, then there must be a world v such that the relations $w \; R \; v$ and $v \; Z \; v'$, drawn with dashed lines, also exist.

Note that we have not drawn the direction of the bisimulation. Technically speaking the bisimulation $Z \subseteq W \times W'$ has a direction as a relation. However, the back and forth clauses are symmetric, and thus by the following proposition the direction of the bisimulation does not really matter:

Proposition 3.2 — *Inverse bisimulation is bisimulation*
If $Z \subseteq W \times W'$ is a bisimulation between models $\mathcal{M}$ and $\mathcal{M}'$, then $Z^{-1} \subseteq W' \times W$ is a bisimulation between $\mathcal{M}'$ and $\mathcal{M}$. Hence $\mathcal{M}, w \,\underline{\leftrightarrow}\, \mathcal{M}', w'$ if and only if $\mathcal{M}', w' \,\underline{\leftrightarrow}\, \mathcal{M}, w$. ◁

Example 3.3
The following diagram gives a bisimulation Z between the models $\mathcal{M}$ with world a and $\mathcal{M}'$ with worlds w and v.

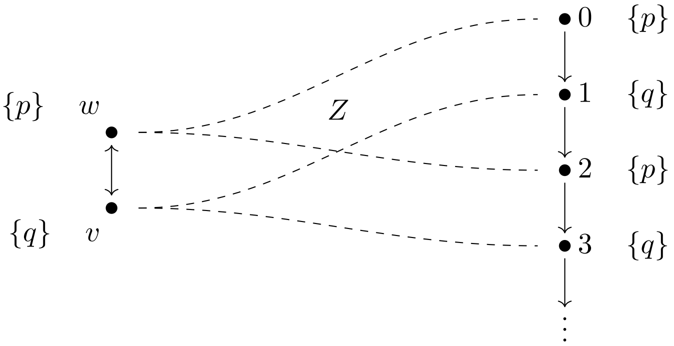


Clearly all worlds have the same valuation, thus the atomic condition is satisfied.

Next let us check the back condition. Since $a\ Z\ w$ and $w\ R'\ v$, we need to have a world b in $\mathcal{M}$ such that $a\ R\ b$ and $b\ Z\ v$. This world is the world a itself, since indeed $a\ R\ a$ and $a\ Z\ v$. In the same way the back condition works for $a\ Z\ v$ and $v\ R\ w$.

Finally we have to check the forth condition. Since $a\ R\ a$ and $a\ Z\ w$, we need to have a world u in $\mathcal{M}'$ such that $a\ Z\ u$ and $w\ R'\ u$. This world is of course v. We can show a similar thing for the relation $a\ Z\ v$. ◁

Example 3.4
The following diagram is another, slightly more interesting example of a bisimulation Z between the models $\mathcal{M}$ with worlds w, v and $\mathcal{M}'$ with the natural numbers $\mathbb{N} = \{0, 1, 2, \dots\}$ as worlds and the successor relation as accessibility relation (see the frame $\mathcal{F}_4$ from Example 1.7). The valuation of the worlds $n \in \mathbb{N}$ of $\mathcal{M}'$ is given by $V(n) = \{p\}$ if n is even and $V(n) = \{q\}$ if n is odd. We have $w\ Z\ n$ if and only if n is even, and $v\ Z\ n$ if and only if n is odd.



We will leave it as Exercise 3.2 to check that this is a bisimulation. ◁

The main point that makes bisimulation such a powerful tool is the following theorem. It states that whenever two worlds are bisimilar in a model, then they make the same formulas true.

Theorem 3.5 — *Bisimulation theorem*
If $\mathcal{M} = \langle W, R, V\rangle$ and $\mathcal{M}' = \langle W', R', V'\rangle$ are models, $w \in W$ and $w' \in W'$ are worlds and $\mathcal{M}, w \mathrel{\underline{\leftrightarrow}} \mathcal{M}', w'$, then the same formulas are true in w and w', or in other words, $\mathcal{M}, w \vDash \varphi$ if and only if $\mathcal{M}', w' \vDash \varphi$ ◁

> *Proof.* The proof is by induction on the structure of the formula φ. Assume that $\mathcal{M}, w \mathrel{\underline{\leftrightarrow}} \mathcal{M}', w'$ for some worlds w of $\mathcal{M} = \langle W, R, V\rangle$ and w' of $\mathcal{M}' = \langle W', R', V'\rangle$. Then there is a bisimulation $Z \subseteq W \times W'$ such that $w \; Z \; w'$.
>
> We will assume without loss of generality that every formula φ is built with just the operators $\wedge$, $\neg$ and $\Box$ and atomic variables. The logical constants $\bot$ and $\top$ and the other operators $\vee$, $\rightarrow$ and $\Diamond$ can be expressed in terms of the aforementioned operators*, and thus we can save ourselves some work by ignoring them.
>
> The base case is if φ is an atomic variable, say $\varphi = p$. In this case $\mathcal{M}, w \vDash p$ if and only if $p \in V(w)$. Since Z is a bisimulation and $w \; Z \; w'$, we have that $V(w) = V'(w')$, and thus $p \in V(w)$ if and only if $p \in V'(w')$. Therefore $\mathcal{M}, w \vDash p$ if and only if $\mathcal{M}', w' \vDash p$.
>
> For the inductive cases, we assume the inductive hypothesis that we have already shown that $\mathcal{M}, x \vDash \varphi$ if and only if $\mathcal{M}', x' \vDash \varphi$ and $\mathcal{M}, x \vDash \psi$ if and only if $\mathcal{M}', x' \vDash \psi$ for any worlds $x \in W$ and $x' \in W'$ such that $x \; Z \; x'$.
>
> First we will prove that the inductive hypothesis implies that $\mathcal{M}, w \vDash \varphi \wedge \psi$ if and only if $\mathcal{M}', w' \vDash \varphi \wedge \psi$. This is quite simple: $\mathcal{M}, w \vDash \varphi \wedge \psi$ if and only if $\mathcal{M}, w \vDash \varphi$ and $\mathcal{M}, w \vDash \psi$, which by the inductive hypothesis is true if and only if $\mathcal{M}', w' \vDash \varphi$ and $\mathcal{M}', w' \vDash \psi$, which in turn is true if and only if $\mathcal{M}', w' \vDash \varphi \wedge \psi$.
>
> Next we will prove that the inductive hypothesis implies that $\mathcal{M}, w \vDash \neg\varphi$ if and only if $\mathcal{M}', w' \vDash \neg\varphi$. Once again, $\mathcal{M}, w \vDash \neg\varphi$ if and only if $\mathcal{M}, w \nvDash \varphi$, which holds if and only if $\mathcal{M}', w' \nvDash \varphi$ by the inductive hypothesis, which in turn holds if and only if $\mathcal{M}', w' \vDash \neg\varphi$.
>
> Finally we have the least trivial case, that $\mathcal{M}, w \vDash \Box\varphi$ if and only if $\mathcal{M}', w' \vDash \Box\varphi$. Suppose that $\mathcal{M}, w \vDash \Box\varphi$, then for all worlds v such that $w \; R \; v$ we have $\mathcal{M}, v \vDash \varphi$. Suppose that $w' \; R' \; v'$ for some world $v' \in W'$. We have $w \; Z \; w'$ and $w' \; R' \; v'$, thus by the back condition there must be a world $v \in W$ such that $w \; R \; v$ and $v \; Z \; v'$. Since $\mathcal{M}, w \vDash \Box\varphi$, we have that $\mathcal{M}, v \vDash \varphi$ for this v, and by the induction hypothesis we then get $\mathcal{M}', v' \vDash \varphi$ as well. Because v' was arbitrary, we see that $\mathcal{M}', w' \vDash \Box\varphi$. Showing that $\mathcal{M}', w' \vDash \Box\varphi$ implies $\mathcal{M}, w \vDash \Box\varphi$ happens with the exact same argument, but this time using the forth condition instead of the back condition. ■

For those who see induction on the structure of a formula for the first time, it might be confusing that this proof implies that $\mathcal{M}, w \vDash \varphi$ if and only if $\mathcal{M}', w' \vDash \varphi$ for all formulas φ. The thing to note is that in evaluating a formula φ in some world w of model $\mathcal{M}$, we look at how the **subformulas** of φ are evaluated. We start by looking at the valuation of the atomic variables, and work our way up by interpreting the connectives accordingly. In other words, to find the validity of some complex formula φ, we look at the validity of the simpler components that make up the formula φ. The proof above gives us a recipe for how bisimulation preserves this validity at each step of building a complex formula from a simpler one.

---

*Namely: $\bot \equiv p \wedge \neg p$, $\top \equiv \neg(p \wedge \neg p)$, $p \vee q \equiv \neg(\neg p \wedge \neg q)$, $p \rightarrow q \equiv \neg(p \wedge \neg q)$ and $\Diamond p \equiv \neg\Box\neg p$.

Definition 3.6 — *Total bisimulation*
Let $\mathcal{M} = \langle W, R, V\rangle$ and $\mathcal{M}' = \langle W', R', V'\rangle$ be models. A bisimulation $Z \subseteq W \times W'$ is **total for** $\mathcal{M}'$ if the set $\{w' \in W' \mid \exists w \in W(w\ Z\ w')\}$ is all of W', in other words, every world in W' is reached by some $w \in W$ for bisimulation Z. Similarly Z is **total for** $\mathcal{M}$ if the set $\{w \in W \mid \exists w' \in W'(w\ Z\ w')\}$ is all of W. Equivalently Z is total for $\mathcal{M}$ if and only if Z^{-1} is total for $\mathcal{M}$. If Z is total for both $\mathcal{M}$ and $\mathcal{M}'$, we call Z a **total bisimulation**. ◁

Example 3.7
The bisimulations given in Examples 3.3 and 3.4 are both total.

Since the bisimulation in Example 3.3 is total, we see that the first model of Example 2.5 is bisimilar with a reflexive model. As a consequence of the bisimulation theorem, this means that the same formulas are true in the nonreflexive model of Example 2.5 as in the reflexive model. In particular, the formula $\Box\varphi \to \varphi$ is true in the nonreflexive model, since this formula characterises reflexive frames and hence must be valid on the bisimilar reflexive model. ◁

This example shows that we can extend the bisimulation theorem to the level of models, and indeed to the level of frames. We do this in the following two corollaries.

Corollary 3.8
If $\mathcal{M}$ and $\mathcal{M}'$ are models and there exists a bisimulation between $\mathcal{M}$ and $\mathcal{M}'$ total for $\mathcal{M}'$, then $\mathcal{M} \vDash \varphi \Rightarrow \mathcal{M}' \vDash \varphi$ for all formulas φ. ◁

> *Proof.* Let Z be a bisimulation between $\mathcal{M}$ and $\mathcal{M}'$ that is total for $\mathcal{M}'$. Suppose that $\mathcal{M} \vDash \varphi$, then for every world w of $\mathcal{M}$ we have $\mathcal{M}, w \vDash \varphi$. Let w' be any world of $\mathcal{M}'$, then by Z being total, there must be some world w in $\mathcal{M}$ such that $w\ Z\ w'$. Because $\mathcal{M}, w \vDash \varphi$, we see by the bisimulation theorem that then $\mathcal{M}', w' \vDash \varphi$ as well. Since w' was an arbitrary world of $\mathcal{M}'$, we see that $\mathcal{M}', w' \vDash \varphi$ for all worlds w' of $\mathcal{M}'$, and thus $\mathcal{M}' \vDash \varphi$. ■

Corollary 3.9
If $\mathcal{F}$ and $\mathcal{F}'$ are frames, and for any model $\mathcal{M}'$ based on $\mathcal{F}'$ there is a model $\mathcal{M}$ based on $\mathcal{F}$ such that there exists a bisimulation between $\mathcal{M}$ and $\mathcal{M}'$ total for $\mathcal{M}'$, then $\mathcal{F} \vDash \varphi \Rightarrow \mathcal{F}' \vDash \varphi$ for all formulas φ. ◁

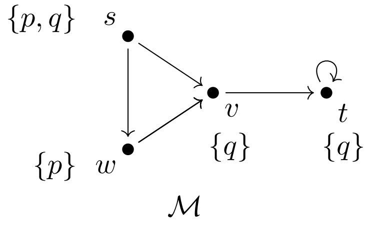
> *Proof.* The proof is left as an exercise. ■

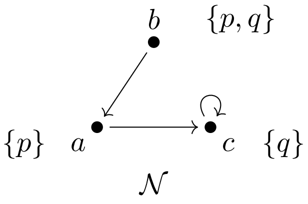
This last corollary gives us a way to show that two different frames make the same formulas true. If we have a frame class $\mathcal{C}$ containing $\mathcal{F}$, but not $\mathcal{F}'$, and $\mathcal{F}$ and $\mathcal{F}'$ are as in Corollary 3.9, then it follows that there can not be a modal formula that defines $\mathcal{C}$: if such a formula would hold on every frame in $\mathcal{C}$, then it would in $\mathcal{F}$, but using the corollary we then see that the frame $\mathcal{F}'$ also makes this formula true, while $\mathcal{F}'$ is not a frame in $\mathcal{C}$.

The next chapter will gives us some specific tools that build on this concept and can be used to show that frame classes are not modally definable.

3.1 EXERCISES

Exercise 3.1. Consider the following models:



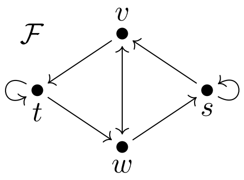


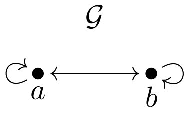
Determine for the following relations if they are a bisimulation, and if so, for which model they are total:

a) $Z_1 = \{\langle w, a\rangle, \langle v, c\rangle, \langle t, c\rangle, \langle s, b\rangle\}$
b) $Z_2 = \{\langle w, a\rangle, \langle t, c\rangle, \langle v, c\rangle\}$
c) $Z_3 = \{\langle a, w\rangle, \langle c, t\rangle, \langle c, v\rangle\}$
d) $Z_4 = \{\langle t, c\rangle, \langle c, v\rangle\}$
e) $Z_5 = \{\langle w, w\rangle, \langle s, s\rangle, \langle v, v\rangle, \langle t, v\rangle\}$
f) $Z_6 = \{\langle w, w\rangle, \langle s, s\rangle, \langle v, v\rangle, \langle t, v\rangle, \langle t, t\rangle\}$
g) $Z_7 = \{\langle a, v\rangle, \langle c, t\rangle\}$
h) $Z_8 = \varnothing$

Exercise 3.2. Check that the relation Z in Example 3.4 is a bisimulation.

Exercise 3.3. Consider the following two frames:





a) Give a valuation on $\mathcal{F}$ and $\mathcal{G}$ such that there is a total bisimulation.
b) Show that for any valuation on $\mathcal{G}$ there exists a valutation on $\mathcal{F}$ such that there is a total bisimulation between the two models.
c) Does there exist a nonempty bisimulation between any models based on $\mathcal{F}$ and $\mathcal{G}$ that is not total for one of the models?

Exercise 3.4. Prove Proposition 3.2.

Exercise 3.5. Give a direct proof of the $\Diamond$ case in the proof by induction of the bisimulation theorem (Theorem 3.5). That is, let Z be a bisimulation and assume the induction hypothesis that $\mathcal{M}, x \vDash \varphi$ if and only if $\mathcal{M}', x' \vDash \varphi$ for any x and x' such that $x\ Z\ x'$. Then prove that $\mathcal{M}, w \vDash \Diamond\varphi$ if and only if $\mathcal{M}', w' \vDash \Diamond\varphi$

Exercise 3.6. Prove Corollary 3.9.

Exercise 3.7. A frame is called **acyclic** if there is no sequence of worlds $w_1, w_2, \ldots, w_n$ such that $w_1\ R\ w_2\ R\ \cdots\ R\ w_n\ R\ w_1$ for any $n \in \mathbb{N}$. A **tree model** is a model based on an *asymmetric*, *acyclic* frame in which every world has a *at most one predecessor* ($\forall w \forall v \forall u (v\ R\ w \wedge u\ R\ w \to v = u)$).

a) Consider the model $\mathcal{M}$ from Exercise 3.1. Give a tree model $\mathcal{M}'$ with a world w' such that $\mathcal{M}, w \;\underline{\leftrightarrow}\; \mathcal{M}', w'$.
b) Consider the model $\mathcal{M}$ based on $\mathcal{F}$ from Exercise 3.2 with $V(w) = \varnothing$, $V(s) = \{p\}$, $V(v) = \{q\}$ and $V(t) = \{r\}$. Give a tree model $\mathcal{M}'$ with a world w' such that $\mathcal{M}, w \;\underline{\leftrightarrow}\; \mathcal{M}', w'$.

Exercise 3.8. Give examples of models as described below:

a) An infinite model that has a total bisimulation with a finite model.
b) An infinite model such that there exists no total bisimulation with any finite model.

Exercise 3.9. Prove that for any finite model, there is an infinite model such that there exists a total bisimulation between the two models.

$\ast$ **Exercise 3.10.** Prove or give a counter example to the following claims:

a) for any infinite frame, there is a model based on the frame that has a bisimulation with a finite model.
b) for any infinite frame, there is a model based on the frame that has a bisimulation with a finite model that is total for the infinite model.
c) for any infinite frame, there is a model based on the frame that has a bisimulation with a finite model that is total for the finite model.
d) for any infinite frame, there is a model based on the frame such that some world of the frame is bisimilar with a world in a finite model.

$\ast$ **Exercise 3.11.** Define the *depth* of a formula φ as the depth of the nesting of modal operators. To be precise, let the depth $\mathrm{dp}(\varphi)$ be defined recursively by

$$\begin{aligned} \mathrm{dp}(\varphi) &:= 0 \quad \text{if } \varphi \text{ contains no modal operators,} \\ \mathrm{dp}(\varphi \triangle \psi) &:= \max\{\mathrm{dp}(\varphi), \mathrm{dp}(\psi)\} \quad \text{if } \triangle \text{ is one of the operators } \wedge, \vee, \rightarrow \text{ or } \leftrightarrow, \\ \mathrm{dp}(\neg\varphi) &:= \mathrm{dp}(\varphi), \\ \mathrm{dp}(\Diamond\varphi) = \mathrm{dp}(\Box\varphi) &:= \mathrm{dp}(\varphi) + 1. \end{aligned}$$

For example $\Diamond\Diamond\Box(\Diamond p \wedge \Diamond\neg(\Box q \vee r))$ has depth 5.

If $w \in W$ is a world in a model $\mathcal{M} = \langle W, R, V \rangle$, let $\mathrm{acc}_n(w)$ be the set of worlds v accessible from w in less than or equal to n steps. To be precise, recursively define $\mathrm{acc}_0(w) = \{w\}$ and $\mathrm{acc}_{n+1}(w) = \mathrm{acc}_n(w) \cup \{v \in W \mid \exists u \in \mathrm{acc}_n(w) \text{ such that } u\, R\, v\}$.

a) Prove that for a formula φ and a world w of model $\mathcal{M}$, you only need to look at the valuation of $\mathrm{acc}_{\mathrm{dp}(\varphi)}(w)$ to determine whether $\mathcal{M}, w \vDash \varphi$.
b) Use the previous answer to prove that if Φ is a finite set of modal formulas and there is a model $\mathcal{M}$ with world w such that $\mathcal{M}, w \vDash \varphi$ for all $\varphi \in \Phi$, then there is a model $\mathcal{N}$ with world v such that $\mathcal{N}, v \vDash \varphi$ for all $\varphi \in \Phi$ and such that there is no infinite chain of worlds $v\, R\, v_1\, R\, v_2\, R\, v_3\, R \cdots$ in $\mathcal{N}$.
c) Show that the result of the previous question does not hold if Φ is an infinite set of modal formulas.

$\ast$ **Exercise 3.12.** The bisimulation theorem (Theorem 3.5) says that if two worlds are bisimilar, then they make the same formulas true. In this exercise we will investigate to what extend the reverse implication is true.

Call a frame $\langle W, R \rangle$ **finitely branching** when the set $\{v \mid w\, R\, v\}$ is finite for every $w \in W$.

a) Give an example of a frame that is not finitely branching.

Let $\mathcal{M} = \langle W, R, V \rangle$ and $\mathcal{M}' = \langle W', R', V' \rangle$ be models with worlds $w \in W$ and $w' \in W'$ such that $\mathcal{M}, w \vDash \varphi$ if and only if $\mathcal{M}', w' \vDash \varphi$ for any formula φ. The **Hennessy-Milner theorem** states that if $\mathcal{M}$ and $\mathcal{M}'$ are finitely branching, then $\mathcal{M}, w \,\underline{\leftrightarrow}\, \mathcal{M}', w'$.

b) Show that if the models are not finitely branching, then $\mathcal{M}, w \,\underline{\leftrightarrow}\, \mathcal{M}', w'$ does not have to be true.
(Hint: let $\mathcal{M}$ and $\mathcal{M}'$ be models such that both w and w' have infinitely many branches, but only w has a branch of infinite length, while w' has only branches of finite, but arbitrarily large length.)

4 NON-CHARACTERISABILITY

In this chapter we will look into three methods that can be used to prove that frames are not modally definable. The way these methods work, is that they present two frames that are related in a specific way that allows us to apply Corollary 3.9. If one of the two frames is part of a frame class, while the other is not, then Corollary 3.9 shows that there exists no modal formula that can distinguish between these two frames, and hence the frame class cannot be defined by any modal formula.

4.1 DISJOINT UNIONS

The first method is that of combining two frames in a disjoint union. Intuitively this could be seen as taking two copies of the frames and taking them side by side as a new frame. As none of the worlds of the two disjoint frames can reach any world in the other frame, the two copies do not influence which formulas are true on the frames. As such any formula that is true on both frames must still be true on the disjoint union.

To make this argument concrete, we introduce a formal way to describe disjoint unions.

Definition 4.1 — *Disjoint union of sets*
If X and Y are sets, then the **disjoint union** $X \sqcup Y$ is defined as the set $(X \times \{1\}) \cup (Y \times \{2\})$ which consists of elements (a, i) where $a \in X$ if $i = 1$ and $a \in Y$ if $i = 2$.

This definition can be generalised to any family of sets X_i for $i \in I$, then the disjoint union $\bigsqcup_{i \in I} X_i$ is the set $\bigcup_{i \in I}(X_i \times \{i\})$. ◁

Example 4.2
The disjoint union of $\{a, b, c\}$ and $\{a, c, d\}$ is the set $\{(a, 1), (b, 1), (c, 1), (a, 2), (c, 2), (d, 2)\}$.
The disjoint union of $\mathbb{N}$ copies of $\mathbb{N}$ is the set $\bigsqcup_{m \in \mathbb{N}} \mathbb{N} = \{(n, m) \mid n, m \in \mathbb{N}\}$. ◁

We can extend this definition to frames, where we take the disjoint union of their sets of worlds, and connect the worlds with a new accessibility relation in the same way they were originally connected.

Definition 4.3 — *Disjoint union of frames*
If $\mathcal{F}_1 = \langle W_1, R_1 \rangle$ and $\mathcal{F}_2 = \langle W_2, R_2 \rangle$ are frames, then $\mathcal{F}_1 \sqcup \mathcal{F}_2 = \mathcal{F} = \langle W, R \rangle$ is the frame with

$$W = W_1 \sqcup W_2$$
$$R = \{\langle (x, i), (y, j) \rangle \mid i = j \land x \; R_i \; y\}$$

So we see that $(x, i) \not R \; (y, j)$ whenever $i \neq j$, and $(x, i) \; R \; (y, i)$ whenever $x \; R_i \; y$. ◁

Lemma 4.4 — *Bisimulation of disjoint unions*
If $\mathcal{F}$ and $\mathcal{F}'$ are frames, then $\mathcal{F} \sqcup \mathcal{F}' \vDash \varphi$ if and only if $\mathcal{F} \vDash \varphi$ and $\mathcal{F}' \vDash \varphi$. ◁

Proof. ($\Leftarrow$) Let $\mathcal{M}$ be a model based on $\mathcal{F}$ with valuation V and $\mathcal{M}'$ a model based on $\mathcal{F}'$ with valuation V'. Then we define a model $\mathcal{M} \sqcup \mathcal{M}'$ on $\mathcal{F} \sqcup \mathcal{F}'$ with valuation V'' as follows: if (w, i) is a world in $\mathcal{F} \sqcup \mathcal{F}'$, then $V''((w, i)) = V(w)$ if $i = 1$ and $V''((w, i)) = V'(w)$ if $i = 2$. In other words, the worlds in the disjoint union that originally were worlds in $\mathcal{F}$ get the same valuation as their original has in $\mathcal{M}$, and similar for the worlds that originally were worlds from $\mathcal{M}'$.

We define a bisimulation Z between $\mathcal{M} \sqcup \mathcal{M}'$ and $\mathcal{M}$ that is total for $\mathcal{M}$ by letting $(w, 0)\ Z\ w$ for every world w of $\mathcal{M}$. We leave it as an exercise to check that this is indeed a bisimulation that is total for $\mathcal{M}$. Since our model $\mathcal{M}$ based on $\mathcal{F}$ was arbitrary, we satisfy the conditions of Corollary 3.9, thus $\mathcal{F} \sqcup \mathcal{F}' \vDash \varphi$ implies $\mathcal{F} \vDash \varphi$. Analogously we can show that $\mathcal{F} \sqcup \mathcal{F}' \vDash \varphi$ implies $\mathcal{F}' \vDash \varphi$.

($\Rightarrow$) Given any model $\mathcal{M}''$ based on $\mathcal{F} \sqcup \mathcal{F}'$ with valuation V'', we can define a model $\mathcal{M}$ on $\mathcal{F}$ with valuation V as follows: if w is a world in $\mathcal{F}$, then we let $V(w) = V''((w, 1))$. Similarly we define a model $\mathcal{M}'$ on $\mathcal{F}'$ with valuation V' such that $V'(w') = V''((w', 2))$ for every world w' of $\mathcal{F}'$.

Define bisimulations Z between $\mathcal{M}''$ and $\mathcal{M}$ and Z' between $\mathcal{M}''$ and $\mathcal{M}'$ given by $(w, 1)\ Z\ w$ for all worlds w of $\mathcal{F}$ and $(w', 2)\ Z'\ w'$ for all worlds w' of $\mathcal{M}'$. Again, we leave it an exercise to check that Z and Z' are bisimulations. Neither Z nor Z' is total for $\mathcal{M}''$, so we cannot immediately use Corollary 3.9. However, if both $\mathcal{F} \vDash \varphi$ and $\mathcal{F}' \vDash \varphi$, then $\mathcal{M}, w \vDash \varphi$ for every world w of $\mathcal{F}$ and $\mathcal{M}', w' \vDash \varphi$ for every world w' of $\mathcal{F}'$. Therefore, for any world w of $\mathcal{F}$ we have $\mathcal{M}, w \underline{\leftrightarrow} \mathcal{M}'', (w, 1)$ and thus by the bisimulation theorem $\mathcal{M}'', (w, 1) \vDash \varphi$. In the same way $\mathcal{M}'', (w', 2) \vDash \varphi$ for all worlds w' of $\mathcal{F}'$.

Together with the fact that every world of $\mathcal{M}''$ is of the form $(w, 1)$ for some world w of $\mathcal{F}$ or of the form $(w', 2)$ for some world w' of $\mathcal{F}'$, it follows that $\mathcal{M}'' \vDash \varphi$. ■

Corollary 4.5 — *Modal definability is closed under disjoint unions*
If $\mathcal{C}$ is a modally definable frame class and $\mathcal{F}_1, \mathcal{F}_2 \in \mathcal{C}$, then $\mathcal{F}_1 \sqcup \mathcal{F}_2 \in \mathcal{C}$. ◁

Proof. Let φ be a modal formula characterising $\mathcal{C}$, then $\mathcal{F}_1 \vDash \varphi$ and $\mathcal{F}_2 \vDash \varphi$, since $\mathcal{F}_1 \in \mathcal{C}$ and $\mathcal{F}_2 \in \mathcal{C}$. By Lemma 4.4 then $\mathcal{F}_1 \sqcup \mathcal{F}_2 \vDash \varphi$, and thus $\mathcal{F}_1 \sqcup \mathcal{F}_2 \in \mathcal{C}$. ■

We will now give an example to show how this technique can be used to prove that the class of universal frames is not modally definable.

Example 4.6
Let $\mathcal{C}_{\text{univ}}$ be the class of universal frames, and let $\mathcal{F}, \mathcal{F}'$ be two universal frames. Then $\mathcal{F} \sqcup \mathcal{F}'$ is not a universal frame, since we can take a world w of $\mathcal{F}$ and a world w' of $\mathcal{F}'$ and see that $(w, 1)$ does not reach $(w', 2)$ in $\mathcal{F} \sqcup \mathcal{F}'$.

Therefore $\mathcal{C}_{\text{univ}}$ is not closed under disjoint unions, and by Corollary 4.5 we see that $\mathcal{C}_{\text{univ}}$ is not modally definable. ◁

4.2 GENERATED SUBFRAMES

The second method we will discuss is that of a generated subframe. Intuitively, if we have a model $\mathcal{M}$ with a world w and a formula φ, then if we want to know whether $\mathcal{M}, w \vDash \varphi$, we only need to look at the worlds that are reachable from w, the worlds reachable from those worlds, and so on. Hence, if we take a smaller model that consists just of those worlds reachable from w, and from the worlds reachable from w, and so on, then in this model it is also true that $\mathcal{M}, w \vDash \varphi$. We call such a smaller model a generated subframe.

We need to give a more formal description of what it means for a world to be reachable from a world that is reachable from a world that is reachable from ... from w. The following definition gives us exactly this notion.

Definition 4.7 — *Transitive reflexive closure*
For a relation R, define the **transitive reflexive closure** R^* such that $x\ R^*\ y$ if and only if $x = y$ or there are $z_1, \ldots, z_n$ such that $x\ R\ z_1\ R \cdots R\ z_n\ R\ y$ (possibly with $n = 0$, which gives the case that $x\ R\ y$). ◁

An alternative definition is that R^* is the smallest relation such that $R \subseteq R^*$, and R^* is transitive and reflexive (see Exercise 4.11). We see that any worlds x and y that $x\ R^*\ y$ if y is reachable from x is a finite number of steps following the accessibility relation R. A generated subframe is the frame of exactly those worlds that reachable from w with the transitive reflexive closure of the accessibility relation.

Definition 4.8 — *Generated subframe*
If $\mathcal{F} = \langle W, R \rangle$ is a frame and $w \in W$ a world of $\mathcal{F}$, then we define the **subframe generated by** w as the frame $\mathcal{F}_w = \langle W_w, R_w \rangle$ with $W_w = \{v \in W \mid w\ R^*\ v\}$ and with $v\ R_w\ u$ if and only if $v, u \in W_w$ and $v\ R\ u$. ◁

Lemma 4.9 — *Bisimulation of generated subframes*
If $\mathcal{F} = \langle W, R \rangle$ is a frame and $w \in W$, then $\mathcal{F} \vDash \varphi$ implies $\mathcal{F}_w \vDash \varphi$. ◁

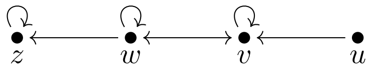
> *Proof.* To avoid confusion, we will denote a world x as x if we mean it as a world of $\mathcal{F}$, and as x' if we mean it as a world of $\mathcal{F}_w$. Of course this is just a notational convenience, since actually $x = x'$.
>
> Suppose that $\mathcal{M}$ is a model based on $\mathcal{F}$, then we define a model $\mathcal{M}_w$ based on $\mathcal{F}_w$ with the valuation V_w defined as $V_w(x') = V(x)$. That is, the valuation of the worlds in $\mathcal{M}_w$ is exactly the valuation of the same worlds in $\mathcal{M}$. We define a bisimulation Z between $\mathcal{M}$ and $\mathcal{M}_w$ by $x\ Z\ x'$ for all worlds x' in $\mathcal{M}_w$.
>
> To see that Z is a bisimulation, we first note that $V_w(x') = V(x)$, thus the atomic condition is satisfied.
>
> For the back condition, let $x\ Z\ x'$ and $x'\ R_w\ y'$ for some world y' in $\mathcal{M}_w$, then by how we defined R_w, we see that also $x\ R\ y$, and by how Z is defined we have $y\ Z\ y'$, so the back condition is fulfilled as well.
>
> Finally, for the forth condition, let $x\ Z\ x'$ and $x\ R\ y$ for some y. Since $x' \in W_w$, we see that $w\ R^*\ x$, by how we defined W_w. Therefore, either $x = w$ and thus $w\ R\ y$ and also $w\ R^*\ y$, or there are $z_1, \ldots, z_n$ such that $x\ R\ z_1\ R \cdots R\ z_n\ R\ x$. We can extend this second case with $x\ R\ y$ to see that also $w\ R^*\ y$. Therefore $y \in W_w$. This proves that $x'\ R_w\ y'$ and $y\ Z\ y'$, and therefore we see that the forth condition is satisfied as well.
>
> Because Z is total for $\mathcal{F}_w$, we see by Corollary 3.9 that $\mathcal{F} \vDash \varphi$ implies $\mathcal{F}_w \vDash \varphi$. ■

Corollary 4.10 — *Modal definability is closed under generated subframes*
If $\mathcal{C}$ is a modally definable frame class and $\mathcal{F} \in \mathcal{C}$, then for any world w of $\mathcal{F}$ also $\mathcal{F}_w \in \mathcal{C}$. ◁

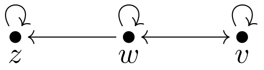
Proof. The proof is as in Corollary 4.5. ■

In the following example we use this technique to show that the class of nonreflexive frames is not modally definable.

Example 4.11
Let $\mathcal{C}_{\text{refl}}$ be the class of nonreflexive frames, and let $\mathcal{F}$ be the following nonreflexive frame:



The generated subframe $\mathcal{F}_v$ consists of all the worlds that are reachable from v in a finite number of steps following the accessibility relation. Since $v\ R\ w$, we see that w is a world of $\mathcal{F}_v$, and since $v\ R\ w\ R\ z$ we see that z is a world of $\mathcal{F}_v$ as well. However, there is no path going from v to u, therefore u is not a world of $\mathcal{F}_v$. We see that $\mathcal{F}_v$ looks like this:



Because all the relations between the worlds in $\mathcal{F}_v$ are inherited from $\mathcal{F}$, we see that each world in $\mathcal{F}$ can reach itself, and therefore $\mathcal{F}_v$ is reflexive, meaning $\mathcal{F}_v \notin \mathcal{C}_{\text{refl}}$.

We see that the class of nonreflexive frames is not closed under generated subframes. By Corollary 4.10 we conclude that the class of nonreflexive frames is not modally definable. ◁

4.3 P-MORPHISMS

The last method we will discuss is that of p-morphisms, and it is arguably the method that looks most like a total bisimulation. Intuitively we generalise the idea of a total bisimulation to the level of frames, so the valuation does no longer matter. Furthermore ask of the bisimulation that it becomes a surjective function, instead of a general relation. The idea is that we can define a valuation on the domain of the function for any given valuation of the range of the function. It follows that if a formula is true on the frame that is the domain, then it will be true on the frame that is the range as well.

Definition 4.12 — *P-morphism*
Let $\mathcal{F} = \langle W, R\rangle$ and $\mathcal{F}' = \langle W', R'\rangle$ be frames. A **p-morphism** (short for *pseudo epimorphism*) from $\mathcal{F}$ to $\mathcal{F}'$ is a function $f : W \to W'$ such that f satisfies the following conditions:

- (**surjection**) f is a surjection (i.e. for all $w' \in W'$ there is a $w \in W$ such that $f(w) = w'$),
- (**back**) if $f(w)\ R'\ v'$ for some $v' \in W'$, then there is $v \in W$ such that $w\ R\ v$ and $f(v) = v'$.
- (**forth**) if $w\ R\ v$, then $f(w)\ R'\ f(v)$,

We call $\mathcal{F}'$ the **p-morphic image** of $\mathcal{F}$ under f. Some authors use the description **bounded morphism** instead of p-morphism. ◁

As said, the atomic condition is removed, since p-morphisms are relations between frames, and not between models. Instead, there is the additional constraint that the p-morphism is a surjective function instead of a more general kind of relation. Because of this, the back and forth conditions look a little different, but in fact they state exactly the same thing. As with bisimulations, we can represent the back and forth conditions of a p-morphism with the following diagram:

$$\begin{tikzcd} v \arrow[r, dashed, "f"] & v' \\ w \arrow[u, dashed, "\mathcal{F}"] \arrow[r, "f"] & w' \arrow[u, "\mathcal{F}'"'] \end{tikzcd}$$

back

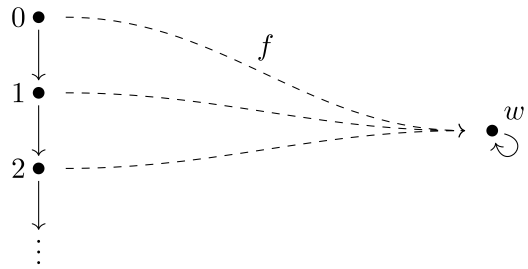
$$\begin{tikzcd} v \arrow[r, "f"] & v' \\ w \arrow[u, "\mathcal{F}"] \arrow[r, "f"] & w' \arrow[u, dashed, "\mathcal{F}'"'] \end{tikzcd}$$

forth

In this case f is a function, so we have drawn arrows to emphasise the direction of p-morphisms. In the back condition, we see that if f maps w to $f(w) = w'$, and $w'\ R'\ v'$, then there must be some world v in $\mathcal{F}$ such that $w\ R\ v$ and f maps v to $f(v) = v'$. Note that there needs to *exist* such world v, but it is not necessary for *every* world v with $f(v) = v'$ that also $w\ R\ v$. See Example 4.13 for an example where the latter is not true.

The forth condition has the assumption that $w\ R\ v$ and that $f(w) = w'$. In the definition of bisimulations, we needed that there *exists* some world v' in $\mathcal{F}'$ such that $v\ Z\ v'$. However, since f is a surjective function, we know that there exists a unique world v in $\mathcal{F}$ such that f sends v to $f(v) = v'$. Therefore the top arrow does not need to be dashed, since it is guaranteed to exist and is unique by f being a function. The only requirement is therefore that $w'\ R'\ v'$ holds.

Example 4.13
The following diagram is an example of a bisimulation between the (non-transitive) frame $\mathcal{F}_4$ from Example 1.7 and the reflexive frame with a single world.



To see that this f is a p-morphism, first note that f is indeed a surjective function: every world in the reflexive frame (there is only one) is in the range of f. Next, we look at the back condition. If we take any world $n \in \mathbb{N}$, then $f(n) = w$. There is just one world that is reachable from w, which is w itself, so we have $f(w)\ R\ w$. The back condition states that there must be some $m \in \mathbb{N}$ such that $n + 1 = m$ and such that $f(m) = w$. Clearly this m exists. Finally the forth condition states that if $n + 1 = m$ for some worlds $n, m \in \mathbb{N}$ (that is, if n reaches the world m), then also $f(n)\ R\ f(m)$. This is also clearly true, since $f(n) = f(m) = w$ and $w\ R\ w$. ◁

We will later look at a slightly less trivial p-morphism, but first we prove the lemma analogous to those for the generated subframe and disjoint union methods.

Lemma 4.14 — *Bisimulation of p-morphic images*
Let $\mathcal{F} = \langle W, R\rangle$ and $\mathcal{F}' = \langle W', R'\rangle$ be frames such that there is a p-morphism $f : W \to W'$, then $\mathcal{F} \vDash \varphi$ implies $\mathcal{F}' \vDash \varphi$. ◁

> *Proof.* Suppose that $\mathcal{F} \vDash \varphi$, and let $\mathcal{M}' = \langle W', R', V'\rangle$ be a model based on $\mathcal{F}'$. We define a valuation V on $\mathcal{F}$ to get a model $\mathcal{M}$ based on $\mathcal{F}$, by letting $V(w) = V'(f(w))$

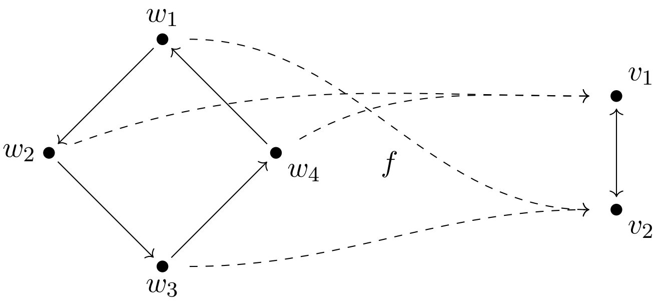
for each world $w \in W$. This means that the worlds in $\mathcal{M}$ get the same valuation as their image under f.

It is not difficult to see that f is a bisimulation: the atomic condition follows from how we defined V, and the back and forth conditions for the bisimulation follow directly from the back and forth conditions for the p-morphism f. Moreover, since f is a surjective function, we see that f is a total bisimulation. It follows from Corollary 3.9 that then $\mathcal{F} \vDash \varphi$ implies $\mathcal{F}' \vDash \varphi$. ■

Corollary 4.15 — *Modal definability is closed under p-morphic images*
If $\mathcal{C}$ is a modally definable frame class, $\mathcal{F} \in \mathcal{C}$, and $\mathcal{F}'$ is a p-morphic image of $\mathcal{F}$, then $\mathcal{F}' \in \mathcal{C}$. ◁

Proof. The proof is as in Corollary 4.5. ■

Example 4.16
Let $\mathcal{C}_{\text{antisym}}$ be the class of antisymmetric frames and let $\mathcal{F} = \langle W, R\rangle$ and $\mathcal{F}' = \langle W', R'\rangle$ be the following frames with a function $f : W \to W'$:



It can be checked that f is a p-morphism. $\mathcal{F}$ is clearly antisymmetric, and $\mathcal{F}'$ is clearly not. Therefore by Corollary 4.15 we conclude that the class $\mathcal{C}_{\text{antisym}}$ is not modally definable. ◁

4.4 GOLDBLATT-THOMASON THEOREM

It turns out that the three construction methods from this chapter are, together with a fourth method, exactly enough to make first-order definable frames modally definable. The fourth method of ultrafilter extensions requires some model theoretic tools if we wish to explain it properly and thus lies outside the scope of this course.

Theorem 4.17 — *Goldblatt-Thomason theorem*
A first-order definable frame class is modally definable if and only if it is closed under

- disjoint unions,
- generated subframes
- p-morphic images

and its complement is closed under ultrafilter extensions. ◁

The proof of this theorem does require quite some model theory, or alternatively it could be done using methods from universal algebra. Both methods lie outside the scope of this course, but for those who are interested, more details can be found in [1].

We can however prove the following weakening of the Goldblatt-Thomason theorem.

Corollary 4.18
If $\mathcal{C}$ is a frame class that is not closed under disjoint unions, or under generated subframes or under p-morphic images, then $\mathcal{C}$ is not modally definable. ◁

Proof. This is a combination of Corrolaries 4.5, 4.10 and 4.15. ■

4.5 EXERCISES

Exercise 4.1. Let $\mathcal{F}_1$ be the frame from Example 1.5 and $\mathcal{F}_2$ the frame from Example 1.6.

a) Give a formal description of the frame $\mathcal{F}_1 \sqcup \mathcal{F}_2$.
b) Give a formal description of the frame $(\mathcal{F}_2)_w$, that is, the subframe of $\mathcal{F}_2$ generated by w.
c) Is there a p-morphism between $\mathcal{F}_1$ and $(\mathcal{F}_2)_w$ from the last question? If so, give the p-morphism, if not, explain why.

Exercise 4.2. Prove for each of the following frames that it is closed under disjoint unions or give a counterexample to show it is not (see Example 2.2 for the definitions):

a) Irreflexive frames
b) Semi-connected frames
c) Partially ordered frames
d) Totally ordered frames
e) Deterministic frames

Exercise 4.3. Prove for each of the following frames that it is closed under generated subframes or give a counterexample to show it is not (see Example 2.2 for the definitions):

a) Euclidean frames
b) Universal frames
c) Antisymmetric frames
d) Connected frames
e) Serial frames
f) Inverse serial frames

Exercise 4.4. Prove for each of the following frames that it is closed under p-morphisms or give a counterexample to show it is not (see Example 2.2 for the definitions):

a) Connected frames
b) Asymmetric frames
c) Left Euclidean frames
d) Convergent frames
e) Partially ordered frames

Exercise 4.5. Determine whether the frames in the last three exercises are modally definable.

Exercise 4.6. Show that the following frame classes are not modally definable.

a) Frames that contain less than n worlds (for some natural number n)
b) Frames that contain more than n worlds
c) Frames that contain infinitely many worlds
d) Frames that contain an n-cycle (defined by $\exists w_1 \ldots \exists w_n(w_1 \; R \; w_2 \; R \cdots R \; w_n \; R \; w_1)$)
e) Acyclic frames (frames containing no cycle of any length)

Exercise 4.7. Let $\mathcal{F} = \langle W, R\rangle$ and $\mathcal{F}' = \langle W', R'\rangle$ be two frames.

A (graph) **homomorphism** between $\mathcal{F}$ and $\mathcal{F}'$ is a function $f : W \to W'$ such that $w \; R \; v$ implies $f(w) \; R' \; f(v)$ for any $w, v \in W$. An **epimorphism** is a surjective homomorphism.

An **isomorphism** between $\mathcal{F}$ and $\mathcal{F}'$ is a function $f : W \to W'$ such that f is a bijection, and for any $w, v \in W$ we have $w\ R\ v$ if and only if $f(w)\ R'\ f(v)$. Equivalently f is an isomorphism if it is a bijective homomorphism, with the inverse f^{-1} also being a homomorphism.

a) Give an example of a modally definable frame class that is not closed under epimorphisms.
b) Show that every isomorphism is a p-morphism.
c) Show that $\mathcal{F} \sqcup (\mathcal{F}' \sqcup \mathcal{F}'') \neq (\mathcal{F} \sqcup \mathcal{F}') \sqcup \mathcal{F}''$.
(Hint: consider what the worlds are.)
d) Show that $\mathcal{F} \sqcup (\mathcal{F}' \sqcup \mathcal{F}'')$ is isomorphic to $(\mathcal{F} \sqcup \mathcal{F}') \sqcup \mathcal{F}''$.

Exercise 4.8. If $\mathcal{C}$ is a frame class, let $\overline{\mathcal{C}}$ be the frame class containing all frames not in $\mathcal{C}$.

a) Let $\mathcal{C}$ and $\mathcal{D}$ be nonempty frame classes that are modally definable. Show that if $\mathcal{C} \cap \mathcal{D} = \varnothing$ then there is a frame $\mathcal{F} \notin \mathcal{C} \cup \mathcal{D}$.
b) Show that if both $\mathcal{C}$ and $\overline{\mathcal{C}}$ are modally definable, then $\mathcal{C} = \varnothing$ or $\overline{\mathcal{C}} = \varnothing$.

Exercise 4.9. Let $\mathcal{C}$ be a modally definable frame class with $\mathcal{F}_1 \sqcup \mathcal{F}_2 \in \mathcal{C}$ for some frames $\mathcal{F}_1$ and $\mathcal{F}_2$.

a) Show that $\mathcal{F}_1 \in \mathcal{C}$ and $\mathcal{F}_2 \in \mathcal{C}$.
b) Use this to prove that the frame class defined by $\exists w \exists v (w \not{R}\, v)$ is not modally definable.
c) Prove that the frame class defined by $\exists w \exists v (w \not{R}\, v)$ is not modally definable using a different technique.

$*$ **Exercise 4.10.** In Definition 4.1 we gave a definition for disjoint unions of an arbitrary (possibly infinite) number of sets.

a) Give a counterpart for Definition 4.3 to arbitrary disjoint unions.
b) Show that there are analogues for the proofs of Lemma 4.4 and Corollary 4.5 to arbitrary disjoint unions.
c) Show that the class of frames that contain finitely many worlds is not modally definable.

Exercise 4.11. Let R be any relation. Let R^* be the relation as defined in Definition 4.7. Let R_* be the smallest relation such that R_* is transitive, reflexive and $R \subseteq R_*$. Prove that $R^* = R_*$.

Part II

Proof Theory of Modal Logic

5 HILBERT SYSTEMS

This chapter and the next chapter will introduce two important proof systems for modal logics. First we will have look at one of the simplest systems, namely Hilbert calculus. Being very simple has one major drawback: it is not always easy to find a proof in Hilbert calculus. For this reason we will subsequently introduce a variant of analytic tableaux for modal logic. These systems are more complex, but also make finding proofs easier and more intuitive.

5.1 HILBERT SYSTEMS

The proof system we will introduce in this chapter is an extension of the Hilbert system for classical propositional logic that was mentioned in the introduction. As a reminder we will define what a Hilbert system is properly in the following definition.

Definition 5.1 — *Hilbert systems*
A Hilbert system S consists of a set of **axiom schemes** and a set of **derivation rules**.

A **proof** that a formula φ follows from a set of formulas Γ in a Hilbert system S is a finite sequence of formulas $\varphi_1, \varphi_2, \ldots, \varphi_n$ such that $\varphi_n = \varphi$ and each φ_i is either a formula in Γ, an instance of an axiom scheme or follows from one or more of the formulas $\varphi_1, \ldots, \varphi_{i-1}$ using a derivation rule.

If there exists a proof that a formula φ follows from a set of formulas Γ in Hilbert system S, we write this as $\Gamma \vdash_{\mathsf{S}} \varphi$ (pronounced as Γ proves φ, or Γ entails φ). If $\Gamma = \{\psi_1, \ldots, \psi_n\}$ we write $\psi_1, \ldots, \psi_n \vdash_{\mathsf{S}} \varphi$ and if Γ is empty, we simply write $\vdash_{\mathsf{S}} \varphi$. ◁

Definition 5.2 — *Hilbert system* M
The axiom schemes of M are:

(Axiom **1**) $A \to (B \to A)$
(Axiom **2**) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$
(Axiom **3**) $(\neg A \to \neg B) \to ((\neg A \to B) \to A)$
(Axiom **K**) $\Box(A \to B) \to (\Box A \to \Box B)$

The derivation rules of M are:

(Modus Ponens) From $\Gamma \vdash_{\mathsf{M}} \varphi$ and $\Gamma \vdash_{\mathsf{M}} \varphi \to \psi$ infer $\Gamma \vdash_{\mathsf{M}} \psi$
(Necessitation) From $\vdash_{\mathsf{M}} \varphi$ infer $\vdash_{\mathsf{M}} \Box\varphi$ ◁

The system M is very small: it has just four axioms and two derivation rules. Nevertheless it is expressive enough to be a proper proof system for the modal logic we introduced in the previous chapters.

Remark 5.3
Note that these axioms of M only use the symbols $\rightarrow$, $\neg$ and $\Box$. In particular, there are no rules for the operators $\wedge$, $\vee$, $\leftrightarrow$ or $\Diamond$, but it is important to mention that this does not restrict the expressivity of our modal language. This is because we can define the other operators by the following translation rules:

$$\begin{aligned} A \vee B &\equiv \neg A \rightarrow B \\ A \wedge B &\equiv \neg(A \rightarrow \neg B) \\ A \leftrightarrow B &\equiv \neg((A \rightarrow B) \rightarrow \neg(B \rightarrow A)) \\ \Diamond A &\equiv \neg\Box\neg B \end{aligned} \qquad \triangleleft$$

Before we move on to some examples of proofs, we will state the following theorem without proof. The theorem states that the simpler Hilbert system made up of just the axioms **1**, **2** and **3** and the modus ponens rule is a proof system for classical propositional logic

Theorem 5.4
The Hilbert system H with as axioms Axiom **1**, **2** and **3**, and as only derivation rule the rule of Modus Ponens is sound and complete with respect to the classical propositional semantics of truth tables. $\triangleleft$

What this means, is that if φ is a classical formula, then $\vDash_{\text{CPL}} \varphi$ by using truth tables if and only if $\vdash_{\mathsf{H}} \varphi$ using the system H. Since any of the axioms and rules of H are also axioms and rules of M, we see that any classical tautology can be proved in the system M as well.

The main difficulty with using Hilbert systems, is that finding a proof can be difficult. Many proofs using only the axioms of M will be long, difficult to read and can require quite a lot of creativity to discover.

Example 5.5
As an example, here is a proof of the classical tautology $\varphi \rightarrow \varphi$ using only the axioms of M. Although $\varphi \rightarrow \varphi$ being a tautology is entirely obvious from its semantical meaning (naturally anything implies itself), the way to reach this conclusion is not at all obvious at first glance:

1.	$\varphi \rightarrow (\varphi \rightarrow \varphi)$	Ax. **1**
2.	$\varphi \rightarrow ((\varphi \rightarrow \varphi) \rightarrow \varphi)$	Ax. **1**
3.	$(\varphi \rightarrow ((\varphi \rightarrow \varphi) \rightarrow \varphi)) \rightarrow ((\varphi \rightarrow (\varphi \rightarrow \varphi)) \rightarrow (\varphi \rightarrow \varphi))$	Ax. **2**
4.	$(\varphi \rightarrow (\varphi \rightarrow \varphi)) \rightarrow (\varphi \rightarrow \varphi)$	MP $2, 3$
5.	$\varphi \rightarrow \varphi$	MP $1, 4$

The proof is a sequence of formulas $\varphi_1, \ldots, \varphi_5$, which are given in the middle column. The left column simply denotes the step number of the formula. In the right column there is a clarification of which rule is being used to arrive at that step. We will go through each of the steps to explain how a proof in a Hilbert system works.

In the first step, we have an instance of Axiom **1**: $A \rightarrow (B \rightarrow A)$. The instance is given by the substitution of both $A = \varphi$ and $B = \varphi$. The second step is also an instance of Axiom **1**, but here we substitute $A = \varphi$ and $B = \varphi \rightarrow \varphi$. The third step is an instance of Axiom **2**: $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$. In this step we have the substitutions $A = \varphi$, $B = \varphi \rightarrow \varphi$ and $C = \varphi$.

The fourth step we apply the rule of modus ponens to steps 2 and 3. In particular, the modus ponens rule states that if we have proved an implication and its antecedent, then we can prove its consequent. Here we see that the formula from step 2 is the antecedent of the formula in step 3, and thus the result of applying modus ponens is that step 4 gives us the consequent of the implication in step 3.

A similar thing happens in the fifth step, where we have step 1 being the antecedent of step 4, and thus we can conclude the consequent of step 4, which is the formula from step 5. $\triangleleft$

Because our main interest is not in proving classical validities, but in proving validities of modal logic, we will work with a different Hilbert system than the system M for the remainder of this chapter.

Definition 5.6 — *Hilbert system* K
The axiom schemes of K consists of *all classical propositional tautologies* and the Axiom **K**: $\Box(A \to B) \to (\Box A \to \Box B)$.

The derivation rules of K are:

(Modus Ponens) From $\Gamma \vdash_{\mathsf{K}} \varphi$ and $\Gamma \vdash_{\mathsf{K}} \varphi \to \psi$ infer $\Gamma \vdash_{\mathsf{K}} \psi$
(Necessitation) From $\vdash_{\mathsf{K}} \varphi$ infer $\vdash_{\mathsf{K}} \Box\varphi$ $\triangleleft$

Note that this change does not make it possible to prove any new formulas, since we can convert a proof in system K to a proof in system M by simply inserting a proof in M for any of the classical tautologies in the proof. Such proofs exist by Theorem 5.4.

Example 5.7
A proof for $\varphi \to \varphi$ in system K is rather trivial:

1. $\varphi \to \varphi$ Taut.

Here the first step is a classical tautology, and thus we consider it an axiom of system K. Note that the formulas that are substituted in the classical tautology need not be classical formulas. For the same reason the above proof is correct, the following proof is correct as well:

1. $(\Box(\varphi \vee \psi) \vee (\psi \to \Diamond\varphi)) \wedge \neg\Box(\varphi \vee \psi) \to (\psi \to \Diamond\varphi)$ Taut.

This formula is an instance of the classical tautology $(A \vee B) \wedge \neg A \to B$ with substitutions $A = \Box(\varphi \vee \psi)$ and $B = \psi \to \Diamond\varphi$. $\triangleleft$

This is a good moment to take some extra note of the necessitation rule. Whereas the modus ponens rule allows for assumptions, in the form of the set of formulas Γ, the necessitation rule does not. In particular, the following rule is **invalid**:

(Necessitation with assumptions) From $\Gamma \vdash_{\mathsf{K}} \varphi$ infer $\Gamma \vdash_{\mathsf{K}} \Box\varphi$

We can see this in the following example.

Example 5.8 — *Counterexample to necessitation with assumptions*
It is evident that $\varphi \vdash_{\mathsf{K}} \varphi$ for any φ: if we assume that φ, then we can prove φ. If necessitation with assumptions was allowed, we would then be able to conclude that $\varphi \vdash_{\mathsf{K}} \Box\varphi$. This is clearly false. If φ is true, it does not have to be necessary. $\triangleleft$

Therefore, whenever we use the necessitation rule in a proof, we must make sure that the step it is applied to does not depend on any assumptions: it may only depends on axioms.

Before we will give another example of a proof, we will first prove a theorem that makes finding proofs a bit easier.

Theorem 5.9 — *Deduction theorem*
$\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \psi$ if and only if $\Gamma \vdash_{\mathsf{K}} \varphi \to \psi$. ◁

Proof. Suppose that $\Gamma \vdash_{\mathsf{K}} \varphi \to \psi$, then also $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \varphi \to \psi$. This is because a proof for the first entailment is also a proof for the second entailment: all the assumptions, axioms and derivation rules that are used in the first proof are also allowed to be assumptions, axioms and derivation rules of the second proof.

We can then use modus ponens, since $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \varphi \to \psi$ and $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \varphi$, to get $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \psi$.

On the other hand, suppose that $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \psi$. Then there is a proof of ψ with assumptions from $\Gamma \cup \{\varphi\}$. We will show that there is also a proof of $\varphi \to \psi$ with assumptions from just Γ. The proof is by induction on the length of the proof $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \psi$, and our induction hypothesis is that the deduction theorem holds for any proof of length less than the proof for $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \psi$.

- If $\psi \in \Gamma$ or ψ is an instance of an axiom, then $\Gamma \vdash_{\mathsf{K}} \psi$. Then we also have $\Gamma \vdash_{\mathsf{K}} \psi \to (\varphi \to \psi)$, since $\psi \to (\varphi \to \psi)$ is an instance of Axiom **1**. By application of modus ponens we get that $\Gamma \vdash_{\mathsf{K}} \varphi \to \psi$ as well.
- If $\psi \in \{\varphi\}$, then $\psi = \varphi$, so we have to show that $\Gamma \vdash_{\mathsf{K}} \varphi \to \varphi$. This is a classical tautology, or alternatively it was proved in Example 5.5.
- If ψ is the result of an application of Modus Ponens, then there are proofs $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \chi \to \psi$ and $\Gamma \cup \{\varphi\} \vdash_{\mathsf{K}} \chi$ which have a shorter length than the proof for ψ. The induction hypothesis states that the deduction theorem holds for all proofs of shorter length, and thus we can see that $\Gamma \vdash_{\mathsf{K}} \varphi \to (\chi \to \psi)$ and $\Gamma \vdash_{\mathsf{K}} \varphi \to \chi$. Let $\alpha_1, \ldots, \alpha_n$ be a proof of $\Gamma \vdash_{\mathsf{K}} \varphi \to (\chi \to \psi)$ and $\beta_1, \ldots, \beta_m$ be a proof of $\Gamma \vdash_{\mathsf{K}} \varphi \to \chi$. Then $\alpha_n = \varphi \to (\chi \to \psi)$ and $\beta_n = \varphi \to \chi$. The following is then a proof of $\Gamma \vdash_{\mathsf{K}} \varphi \to \psi$:

$$\begin{array}{rll}
1. & \alpha_1 & \\
\vdots & \vdots & \\
n. & \varphi \to (\chi \to \psi) \qquad (= \alpha_n) & \\
n{+}1. & \beta_1 & \\
\vdots & \vdots & \\
n+m. & \varphi \to \chi \qquad (= \beta_m) & \\
n+m{+}1. & (\varphi \to (\chi \to \psi)) \to ((\varphi \to \chi) \to (\varphi \to \psi)) & \text{Ax. } 2 \\
n+m{+}2. & (\varphi \to \chi) \to (\varphi \to \psi) & \text{MP } n,\ n+m{+}1 \\
n+m{+}3. & \varphi \to \psi & \text{MP } n{+}m,\ n+m{+}2
\end{array}$$

- If ψ is the result of an application of the Necessitation rule, then ψ is not based on any assumptions, and thus $\vdash_{\mathsf{K}} \psi$, and therefore $\Gamma \vdash_{\mathsf{K}} \psi$. We can then use the same reasoning as in the first case, above, where ψ is an assumption or an instance of an axiom. ■

In fact, the above proof works entirely for the system M as well.

Armed with the full arsenal of classical tautologies and the deduction theorem, we are now ready to give some examples of proofs for modal formulas in K.

Example 5.10

We will prove $\Box(\varphi \wedge \psi) \to \Box\varphi \wedge \Box\psi$. For this we use the following classical tautologies as axiom schemes:

(T. a) $A \wedge B \to A$
(T. a') $A \wedge B \to B$
(T. b) $(A \to B) \to ((A \to C) \to (A \to B \wedge C))$

We use the abbreviation $[\varphi_i]$ to refer to the formula of step i.

1.	$\varphi \wedge \psi \to \varphi$	T. a
2.	$\Box(\varphi \wedge \psi \to \varphi)$	Nec. 1.
3.	$\Box(\varphi \wedge \psi \to \varphi) \to (\Box(\varphi \wedge \psi) \to \Box\varphi)$	Ax. **K**
4.	$\Box(\varphi \wedge \psi) \to \Box\varphi$	MP $2, 3$
5.	$\varphi \wedge \psi \to \psi$	T. a'
6.	$\Box(\varphi \wedge \psi \to \psi)$	Nec 5.
7.	$\Box(\varphi \wedge \psi \to \psi) \to (\Box(\varphi \wedge \psi) \to \Box\psi)$	Ax. **K**
8.	$\Box(\varphi \wedge \psi) \to \Box\psi$	MP $6, 7$
9.	$[\varphi_4] \to ([\varphi_8] \to (\Box(\varphi \wedge \psi) \to \Box\varphi \wedge \Box\psi))$	T. b
10.	$[\varphi_8] \to (\Box(\varphi \wedge \psi) \to \Box\varphi \wedge \Box\psi)$	MP $4, 9$
11.	$\Box(\varphi \wedge \psi) \to \Box\varphi \wedge \Box\psi$	MP $8, 10$

In step 9, we have the substitution $A = \Box(\varphi \wedge \psi)$, $B = \Box\varphi$ and $C = \Box\psi$. We used the rule for Necessitation twice, in step 2 applied on step 1, and in step 6 applied on step 5. Since we can only apply Necessitation to formulas that do not depend on assumptions, we have to take care that steps 1 and 5 do not depend on assumptions. However, as there are no assumptions in the proof, this is not a problem here. ◁

Example 5.11

We will prove $\Box\Diamond\varphi, \Diamond\psi \vdash_{\mathsf{K}} \Diamond\Diamond\varphi$. For this we use the following classical tautologies as axiom schemes:

(T. a) $A \to (\neg A \to B)$
(T. b) $(A \to B) \to (\neg B \to \neg A)$

Furthermore, we make use of the translation rule mentioned in Remark 5.3 in step 12.

1.	$\Box\Diamond\varphi$	Assumption
2.	$\Diamond\psi$	Assumption
3.	$\Diamond\varphi \to (\neg\Diamond\varphi \to \neg\psi)$	T. a
4.	$\Box[\varphi_3]$	Nec. 3
5.	$\Box[\varphi_3] \to (\Box\Diamond\varphi \to \Box(\neg\Diamond\varphi \to \neg\psi))$	Ax. **K**
6.	$\Box\Diamond\varphi \to \Box(\neg\Diamond\varphi \to \neg\psi)$	MP $4, 5$
7.	$\Box(\neg\Diamond\varphi \to \neg\psi)$	MP $1, 6$
8.	$\Box(\neg\Diamond\varphi \to \neg\psi) \to (\Box\neg\Diamond\varphi \to \Box\neg\psi)$	Ax. **K**
9.	$\Box\neg\Diamond\varphi \to \Box\neg\psi$	MP $7, 8$
10.	$[\varphi_9] \to (\neg\Box\neg\psi \to \neg\Box\neg\Diamond\varphi)$	T. b
11.	$\neg\Box\neg\psi \to \neg\Box\neg\Diamond\varphi$	MP $9, 10$
12.	$\Diamond\psi \to \Diamond\Diamond\varphi$	Sub. $\neg\Box\neg A \equiv \Diamond A$
13.	$\Diamond\Diamond\varphi$	MP $2, 12$

Once again, we need to make sure our application of Necessitation in step 4 does not depend on the assumptions. This is the case, since it only depends on step 3, which is a classical tautology. $\triangleleft$

SOUNDNESS AND COMPLETENESS

Two of the most important properties of a proof system are **soundness** and **completeness**. Of these two the soundness theorem is most important, since it tells us that our proofs are valid. If we can prove $\Gamma \vdash_{\mathsf{K}} \varphi$, then we also want that in any model and world in which Γ is valid we have that φ is valid as well. Luckily the soundness theorem is also a lot easier to prove than the completeness theorem, therefore we will give the proof now.

Completeness means that whenever a formula φ is valid in every world, then φ is provable using our proof system. Completeness is unfortunately a lot more delicate than soundness, and there even exist sound modal logics that are not complete with respect to the semantics that we have used so far. Luckily, this is not the case for the Hilbert system K, which is complete with respect to Kripke semantics. The proof of the completeness theorem for the Hilbert system K will be subject of Chapter 7.

Theorem 5.12 — *Soundness theorem*
The Hilbert system K is sound with respect to Kripke semantics: if $\Gamma \vdash_{\mathsf{K}} \varphi$, then $\Gamma \vDash \varphi$. $\triangleleft$

> *Proof.* We prove soundness by induction on the length of a proof in the Hilbert system. Remember that a proof in a Hilbert system $\Gamma \vdash_{\mathsf{K}} \varphi$ is a finite sequence of formulas $\varphi_1, \varphi_2, \ldots, \varphi_n$. Also remember from Definition 1.20 that $\Gamma \vDash \varphi$ holds if every model $\mathcal{M}$ and every world w where $\mathcal{M}, w \vDash \Gamma$ is true, also $\mathcal{M}, w \vDash \varphi$.
>
> If $n = 1$, there are no derivation rules being used, so φ_1 is either a formula in Γ or an instance of an axiom scheme.
>
> If φ_1 is a formula in Γ, we see that $\Gamma \vDash \varphi$ holds, since for any Kripke model $\mathcal{M}$ with world w where $\mathcal{M}, w \vDash \psi$ for all $\psi \in \Gamma$ is true, we automatically have $\mathcal{M}, w \vDash \varphi$ being true as well. On the other hand, if φ_1 is an axiom scheme, we have to show that this axiom scheme is valid in any Kripke model. If φ_1 is a classical tautology, then it holds in every world of every model, so φ_1 is generally valid. Furthermore, we showed that Axiom **K** is generally valid in Proposition 1.18.
>
> We assume the induction hypothesis that for any Γ' and φ' we have that if $\Gamma' \vdash_{\mathsf{K}} \varphi'$ has a proof of length at most n, then $\Gamma' \vDash \varphi'$. Now suppose that $\Gamma \vdash_{\mathsf{K}} \varphi$ has proof $\varphi_1, \ldots, \varphi_{n+1}$ of length $n + 1$. Then for $i \leq n$ we see that $\varphi_1, \ldots, \varphi_i$ is a proof of $\Gamma \vdash_{\mathsf{K}} \varphi_i$ with length at most n, and thus by the induction hypothesis $\Gamma \vDash \varphi_i$ for every $i \leq n$. We have to show that $\Gamma \vDash \varphi_{n+1}$ as well.
>
> We have covered the cases where φ_{n+1} is a formula in Γ or an instance of an axiom scheme above, so what is left is the case where φ_{n+1} follows from previous formulas by one of the derivation rules.
>
> If this rule is Modus Ponens, then we have formulas φ_i and $\varphi_j = \varphi_i \to \varphi_{n+1}$ for some $i, j \leq n$, and thus if $\mathcal{M}, w \vDash \Gamma$, then also $\mathcal{M}, w \vDash \varphi_i$ and $\mathcal{M}, w \vDash \varphi_i \to \varphi_{n+1}$. Now $\mathcal{M}, w \vDash \varphi_i \to \varphi_{n+1}$ is defined to be true if and only if $\mathcal{M}, w \vDash \varphi_{n+1}$ or $\mathcal{M}, w \nvDash \varphi_i$. Since $\mathcal{M}, w \nvDash \varphi_i$ false, it must be true that $\mathcal{M}, w \vDash \varphi_{n+1}$. This proves that $\Gamma \vDash \varphi_{n+1}$.

> If this rule is Necessitation, then $\varphi_{n+1} = \Box\varphi_i$ for some i. By induction hypothesis and $\vdash_{\mathsf{K}} \varphi_i$ being true, it holds that $\vDash \varphi_i$. Then by the principle of necessitation (Proposition 1.19) we see that $\vDash \Box\varphi_i$ as well. ■

As an important corollary of the soundness theorem, we have now a way to show that a formula is not provable in the Hilbert system K. The corollary follows directly from taking the contrapositive statement of the soundness theorem.

Corollary 5.13 — *Counter model method*
If $\Gamma \nvDash \varphi$, then $\Gamma \nvdash_{\mathsf{K}} \varphi$. ◁

Example 5.14
We will show that $\Box\varphi, \Diamond\psi \nvdash_{\mathsf{K}} \Box\psi$.

For a counter model, we need a model $\mathcal{M} = \langle W, R, V\rangle$ and a world $w \in W$ such that $\mathcal{M}, w \nvDash \Box\psi$. Therefore there must be a world $v \in W$ such that $w\ R\ v$ and $\mathcal{M}, v \nvDash \psi$. However, we also want $\mathcal{M}, w \vDash \Box\varphi$ and $\mathcal{M}, w \vDash \Diamond\psi$ to be true, so we need another world $u \in W$ such that $w\ R\ u$ and $\mathcal{M}, u \vDash \psi$.

So, let $\mathcal{M}$ be the model $W = \{w, v, u\}$ with $R = \{(w, v), (w, u)\}$ and $V(w) = \varnothing$, $V(v) = \{p\}$ and $V(u) = \{p, q\}$. Then $\mathcal{M}, w \vDash \Box p$ and $\mathcal{M}, w \vDash \Diamond q$ while $\mathcal{M}, w \nvDash \Box q$.

By Corollary 5.13 it then follows that $\Box\varphi, \Diamond\psi \nvdash_{\mathsf{K}} \Box\psi$. ◁

5.2 NORMAL MODAL LOGICS

Nothing stops us from adding additional axioms to create new Hilbert systems. This is good, since different interpretations of the modality operators give the need for different axioms.

For example, if we consider $\Box\varphi$ with an epistemic interpretation of φ being known, then we usually consider knowledge to be correct, and thus $\Box\varphi \to \varphi$ seems to be a general validity in this interpretation. On the other hand, if we consider the deontic interpretation of this formula, then it says that φ will happen whenever φ ought to happen. This seems to be invalid: although one ought not to steal, it does not follow that stealing does not happen.

Definition 5.15 — *Modal axioms*
The following are common modal axiom schemes:

(Axiom **T**) $\Box A \to A$
(Axiom **B**) $A \to \Box\Diamond A$
(Axiom **D**) $\Box A \to \Diamond A$
(Axiom **4**) $\Box A \to \Box\Box A$
(Axiom **5**) $\Diamond A \to \Box\Diamond A$ ◁

Lemma 5.16
None of the axioms given in Definition 5.15 is provable in K. ◁

> *Proof.* We have to show that $\nvdash_{\mathsf{K}} \varphi$ for φ any of the axioms. The proof uses the method from Corollary 5.13 and is left as Exercise 5.2 ■

We can create new Hilbert systems by adding some of the modal axioms to the system K.

Definition 5.17 — *Modal logics*
The following are common modal logics:

Logic	Axiom schemes
K	**K**
KT	**K**, **T**
KD	**K**, **D**
K4	**K**, **4**
S4	**K**, **T**, **4**
S5	**K**, **T**, **5** or **K**, **D**, **B**, **4**
KD45	**K**, **D**, **4**, **5**

◁

All of these logics have a nice interpretation using Kripke semantics. In particular, each axiom defines a frame class, and the resulting logic will be sound with respect to this frame class. We mean by this that for any proof of $\Gamma \vdash_\Lambda \varphi$ in some system Λ that extends K with additional axioms, we have that $\Gamma \vDash_{\mathcal{C}_\Lambda} \varphi$ where $\mathcal{C}_\Lambda$ is the frame class defined by the additional axioms.

Such modal logics that are sound with respect to some frame class are called **normal modal logics**, and are defined by the following property:

Definition 5.18 — *Normal modal logic*
A modal logic Λ is called **normal** if:

- $\vdash_\Lambda \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$, and
- $\vdash_\Lambda \varphi$ implies $\vdash_\Lambda \Box\varphi$. ◁

Modal logics that are not normal are called **weak modal logics**. These will be the subject of a later chapter.

Example 5.19
The logic K is normal. Any logic that extends K with additional axioms is normal, so all the logics from Definition 5.17 are normal. ◁

To prove the soundness theorems for the abovementioned normal logics, we use the correspondence proofs that we have seen in Chapter 2.

Theorem 5.20 — *Soundness theorem*

- The Hilbert system KT is sound with respect to the class of reflexive frames.
- The Hilbert system KB is sound with respect to the class of symmetric frames.
- The Hilbert system KD is sound with respect to the class of serial frames.
- The Hilbert system K4 is sound with respect to the class of transitive frames.
- The Hilbert system K5 is sound with respect to the class of Euclidean frames. ◁

Proof. The proofs will go exactly as the proof of Theorem 5.12, except that we have to additionaly prove the soundness of the additional axiom. We will prove the soundness theorem for KT and KB and leave the other three cases as an exercise.

Let $\mathcal{C}_{\text{refl}}$ be the class of reflexive frames. The only case for KT that does not follow from the proof of Theorem 5.12, is that $\Gamma \vdash_{\mathsf{KT}} \varphi$ implies $\Gamma \vDash_{\mathcal{C}_{\text{refl}}} \varphi$ when $\varphi = \Box\psi \to \psi$ is an instance of axiom **T**. Let $\mathcal{M}$ be any model based on a reflexive frame, and w be any world of $\mathcal{M}$, and suppose that $\mathcal{M}, w \vDash \Box\psi$, then by $\mathcal{M}$ being reflexive we see that w reaches itself, and thus $\mathcal{M}, w \vDash \psi$. Therefore $\mathcal{M}, w \vDash \Box\psi \to \psi$, as was necessary to prove.

> Let $\mathcal{C}_{\text{sym}}$ be the class of symmetric frames. The only case for KB that we need to prove is that $\Gamma \vdash_{\mathsf{KB}} \varphi$ implies $\Gamma \vDash_{\mathcal{C}_{\text{sym}}} \varphi$ when $\varphi = \psi \to \Box\Diamond\psi$ is an instance of axiom **B**. Let $\mathcal{M}$ be any model based on a symmetric frame and w a world of $\mathcal{M}$. Suppose that $\mathcal{M}, w \vDash \psi$. If w reaches no other worlds, then $\mathcal{M}, w \vDash \Box\Diamond\psi$ vacuously. On the other hand, if w reaches some world v, then also v reaches w, since $\mathcal{M}$ is symmetric. Therefore $\mathcal{M}, v \vDash \Diamond\psi$, since $\mathcal{M}, w \vDash \psi$. But as v was arbitrary, we see that $\mathcal{M}, w \vDash \Box\Diamond\psi$. Therefore $\mathcal{M}, w \vDash \psi \to \Box\Diamond\psi$, as was necessary to prove. ■

We can easily use Proposition 2.6 to combine the soundness theorems and get soundness theorems for the logics $\mathsf{S4}$, $\mathsf{S5}$ and $\mathsf{KD45}$. This is the idea behind the following proposition.

Proposition 5.21

If Λ and Λ' are normal logics, Λ is sound with respect to frame class $\mathcal{C}_\Lambda$, Λ' is sound with respect to frame class $\mathcal{C}'_\Lambda$ and Λ'' is the normal logic consisting of all axioms and derivation rules of Λ and Λ' combined, then Λ'' is sound with respect to the frame class $\mathcal{C}_\Lambda \cap \mathcal{C}'_\Lambda$. ◁

5.3 EXERCISES

Exercise 5.1. Provide proofs for the following statements:

a) $\Box\varphi \vdash_{\mathsf{K}} \Box(\varphi \vee \psi)$
b) $\Box(\varphi \vee \psi), \Box(\varphi \to \chi), \Box(\psi \to \chi) \vdash_{\mathsf{K}} \Box\chi$
c) $\Box\varphi \wedge \Box\psi \vdash_{\mathsf{K}} \Box(\varphi \wedge \psi)$
d) $\Box(\varphi \to \Diamond\psi), \Box\Box\neg\psi \vdash_{\mathsf{K}} \Box\neg\varphi$
e) $\Diamond(\varphi \vee \psi) \vdash_{\mathsf{K}} \Diamond\varphi \vee \Diamond\psi$
f) $\neg\Diamond\varphi \vdash_{\mathsf{K}} \Box(\psi \to \neg\varphi)$

Exercise 5.2. In this exercise we prove Lemma 5.16. Use counter models to show that:

a) $\nvdash_{\mathsf{K}} \Box\varphi \to \varphi$
b) $\nvdash_{\mathsf{K}} \varphi \to \Box\Diamond\varphi$
c) $\nvdash_{\mathsf{K}} \Box\varphi \to \Diamond\varphi$
d) $\nvdash_{\mathsf{K}} \Box\varphi \to \Box\Box\varphi$
e) $\nvdash_{\mathsf{K}} \Diamond\varphi \to \Box\Diamond\varphi$

Exercise 5.3. Provide proofs for the following statements:

a) $\Box\Diamond\Box\varphi \vdash_{\mathsf{KTB}} \varphi$
b) $\vdash_{\mathsf{S4}} \Box\varphi \leftrightarrow \Box\Box\varphi$
c) $\Box\varphi \vdash_{\mathsf{S4}} \varphi \wedge \Box\Box\Box\varphi$
d) $\vdash_{\mathsf{KD5}} \Box\varphi \to \Box\Diamond\varphi$
e) $\vdash_{\mathsf{K4}} \Diamond\Diamond\varphi \to \Diamond\varphi$
f) $\vdash_{\mathsf{S5}} \Diamond\Box\Box\Diamond\varphi \leftrightarrow \Diamond\varphi$

Exercise 5.4. Prove the following relationships between normal modal logics:

a) If $\vdash_{\mathsf{KT}} \varphi$ then $\vdash_{\mathsf{KD}} \varphi$
b) $\vdash_{\mathsf{KT5}} \varphi$ if and only if $\vdash_{\mathsf{KDB5}} \varphi$ if and only if $\vdash_{\mathsf{KDB4}} \varphi$

Exercise 5.5. Prove that $\vdash_{\mathsf{K}} \Box(\varphi \to \psi)$ implies $\vdash_{\mathsf{K}} \Diamond\varphi \to \Diamond\psi$.

Exercise 5.6. Prove that the following derivation rule can be derived in K:

(Hypothetical Syllogism) From $\vdash_{\mathsf{K}} \Box(\varphi \to \psi)$ and $\vdash_{\mathsf{K}} \Box(\psi \to \chi)$ infer $\vdash_{\mathsf{K}} \Box(\varphi \to \chi)$.

Exercise 5.7. In this exercise we give some alternative requirements for being a normal logic by giving some alternative Hilbert systems for the logic K. Let H be the Hilbert system for classical propositional logic, as mentioned in Theorem 5.4. Consider the following axioms and derivation rules:

(Axiom **NT**)	$\Box\top$
(Axiom **C**)	$\Box\varphi \wedge \Box\psi \to \Box(\varphi \wedge \psi)$
(Strong Necessitation rule)	From $\Gamma \vdash_{\mathsf{K}'} \varphi$ infer $\Box\Gamma \vdash_{\mathsf{K}'} \Box\varphi$.
(K rule)	From $\vdash_{\mathsf{K}''} \varphi \to \psi$ infer $\vdash_{\mathsf{K}''} \Box\varphi \to \Box\psi$.
(Deduction rule)	From $\Gamma \cup \{\psi\} \vdash_{\mathsf{K}'} \varphi$ infer $\Gamma \vdash_{\mathsf{K}'} \psi \to \varphi$.

Here $\Box\Gamma$ denotes the set $\Box\Gamma = \{\Box\psi \mid \psi \in \Gamma\}$. In the following questions you may assume that H can prove any classical tautology.

a) Show that each of these axioms and derivation rules are allowed in K.
b) Let K' be the Hilbert system that extends H with the strong necessitation rule and the deduction rule. Prove that K' is a normal modal logic.
c) Let K'' be the Hilbert system that extends H with the axioms **NT** and **C**, as well as with the K rule. Prove that K'' is a normal modal logic.

Exercise 5.8. Let Γ, Δ be sets of formulas and Λ a Hilbert system for a normal modal logic. Prove that if $\Gamma \vdash_\Lambda \varphi$ and $\Delta \vdash_\Lambda \psi$, then $\Gamma \cup \Delta \vdash_\Lambda \varphi \wedge \psi$.

Exercise 5.9. Prove that if Λ and Λ' are Hilbert systems, Λ is sound with respect to the frame class $\mathcal{C}$ and Λ' is sound with respect to the frame class $\mathcal{C}'$, then the logic $\Lambda \cup \Lambda'$ consisting of the axioms and derivation rules of both Λ and Λ' is sound with respect to the frame class $\mathcal{C} \cap \mathcal{C}'$.

6 MODAL TABLEAUX

In this section we will introduce a variant of analytic tableaux that can be used as a proof system for modal logic. Before we can define the tableaux, we need to borrow some preliminary notions from graph theory, in particular we will use the terminology of **trees**.

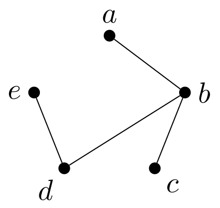
6.1 TREES

A mathematical tree is a special kind of **graph**. Of course this begs the question what an graph then is.

Definition 6.1 — *Undirected graph*
An **(undirected) graph** is a pair $G = \langle V, E \rangle$, where V is a set of **vertices** (singular: vertex) and $E \subseteq V \times V$ is a relation that is *symmetric* and *irreflexive* (see Example 2.2). An **edge** is a pair $\{x, y\}$ such that $x \; E \; y$. We usually denote an edge $\{x, y\}$ simply as xy.

If xy is an edge of G, then the vertices x and y are called **adjacent** to each other. ◁

A natural way to view a graph is as a selection of dots connected with lines. The dots represent the vertices and the lines represent the edges. In this context, vertices are adjacent when they are connected by a line. For example, the graph with vertices a, b, c, d and e and edges ab, bc, bd and de can be drawn as follows:



We can see that in this graph it is possible to *"walk"* from a to e by following a *path* from a to b to d to e via the edges. This notion is made precise in the following definition.

Definition 6.2 — *Path*
A **path** through a graph $G = \langle V, E \rangle$ is a (finite or infinite) sequence $\langle x_0, \ldots, x_i, \ldots \rangle$ of vertices such that all x_i are distinct and each x_i is adjacent to x_{i+1}.

If $n \in \mathbb{N}$ and $p = \langle x_0, \ldots, x_n \rangle$ is a finite path, then we say that the **length** of p is n. If $p = \langle x_0, \ldots, x_i, \ldots \rangle$ is a path, let $p(i) = x_i$ be the i-th vertex on the path. ◁

The path from the example above has thus length 3. There is something special about the graph from the example, namely that for any two distinct vertices, there is only one path between them. This property is what makes a graph a **tree**. Another (equivalent) way to define trees is as a connected graph that does not contain any cycles, that is, there is no path of positive length from any vertex to itself.

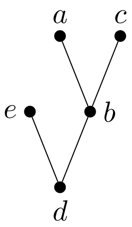
To describe the structure of trees we use analogies with biological trees (such as roots, branches and leaves) and with family trees (to describe the relationship between vertices).

Definition 6.3 — *Tree*
A **tree** is a graph $T = \langle V, E\rangle$ where there exists a *unique* path between any two distinct vertices. A **rooted tree** is a pair $\langle T, r\rangle$ where $T = \langle V, E\rangle$ is a tree and $r \in V$ is a fixed vertex called the **root** of T.

If $x \in V$ is a vertex of a rooted tree $\langle T, r\rangle$, the **ancestry** of x be the unique path p_x starting at r and ending at x. If $y \neq x$ is an element of the ancestry p_x, then we call y an **ancestor** of x.

If $p_x = \langle r, x_1, \ldots, x_{n-1}, x\rangle$ is the ancestry of $x \neq r$, then the vertex x_{n-1} is called the **parent** of x, the vertex x_{n-2} the grandparent, and so on. If y is the (grand)parent of x, then x is a (grand)**child** of y. A vertex with no children is called a **leaf**. A **branch** of a rooted tree $\langle T, r\rangle$ is a path p starting at r that is *maximal*, i.e. there is no path that extends p. The ancestry of any leaf is a branch. ◁

For example, if we take the graph from before, and set d to be the root, then we have a rooted tree. By moving around the vertices a little, we can get it to look more *tree-shaped.* Of course moving around the vertices is purely cosmetic, and does not change anything about the graph.



In this example a has parent b, and its ancestry is the path $\langle d, b, a\rangle$. Since a is a leaf, the path $\langle d, b, a\rangle$ is also a branch of the tree. The root d has two children, namely e and b.

6.2 MODAL TABLEAUX

We will first give a general definition of a system of modal tableaux. It is recommended to read the definition once without pondering too much about it, then go through the examples for modal tableaux that follow, and go through the definition a second time afterwards.

Definition 6.4 — *System of modal tableaux*
A system of tableaux S consists of a set of **derivation rules**. It has no axioms.

A **modal tableau** (plural: **tableaux**) is a rooted tree in which every vertex consists of a pair $\langle s, \varphi\rangle$, where s is a sequence of natural numbers and φ is a modal formula. In the literature this is usually called a **system of labelled tableaux**, where the sequence s is called a **label**. Every vertex either is an **assumption** or follows from one of its ancestors by a derivation rule.

Furthermore, if a vertex is an assumption it can only be preceded by other assumptions, and the root of the tree is always an assumption. Finally, if two vertices occur in the same branch of the tree, then the two vertices are distinct (we will sometimes violate this requirement for vertices to be distinct for the sake of clarity, but as far as correctness of tableaux is concerned it will not be a problem to relax this requirement).

Two vertices $\langle s, \varphi\rangle$ and $\langle s', \varphi'\rangle$ are **conjugates** if $s = s'$ and $\varphi = \neg\varphi'$ or $\neg\varphi = \varphi'$. A branch of a tableau is **closed** if it contains two vertices that are conjugates, otherwise it is **open** . A tableau is **closed** if all of its branches are closed and otherwise is open.

A **proof** that a formula φ follows from a set of formulas Γ in a system of modal tableaux S, consists of a closed tableau with root vertex $\langle (), \neg\varphi\rangle$ and all vertices that are assumptions are vertices of the form $\langle (), \psi_i\rangle$ for some $\psi_i \in \Gamma$. If there exists a proof that φ follows from Γ in system S, we write $\Gamma \vdash_{\mathsf{S}} \varphi$ ◁

We will now give a set of derivation rules for a system of modal tableau for the logic K. The vertices $\langle s, \varphi\rangle$ are written in two columns, the left column for s and the right column for φ. Each s is a sequence of natural numbers $\langle n_0, \dots, n_{k-1}\rangle$, and we use the notation $s^\frown n_k$ to denote the sequence $\langle n_0, \dots, n_{k-1}, n_k\rangle$. The edges of the tree are only drawn when a parent has multiple children, and when a parent vertex has a single child the parent vertex is placed directly above the child. As such, the root of the tree is the vertex at the top of the tableau (in this sense, we draw our trees upside down compared to the biological trees).

The rules should be interpreted as a part of a tableau, in the sense that the vertices below the dashed line are only allowed if all of the vertices above the dashed line occur in their ancestry. Except for the closure rules, the part below the line only depends on a single vertex being present in their ancestry. From the perspective of constructing a tableau from scratch, we can therefore see those rules as being *applied* on the ancestor vertex, and see the vertices below the line as the *conclusion* of applying the rule.

Definition 6.5 — *Modal tableaux for* K

We have the following derivation rules for the system K:

closure:

$$\begin{array}{cc} s & \neg\varphi \\ s & \varphi \\ \hdashline \multicolumn{2}{c}{\times} \end{array} \qquad\qquad \begin{array}{cc} s & \varphi \\ s & \neg\varphi \\ \hdashline \multicolumn{2}{c}{\times} \end{array}$$

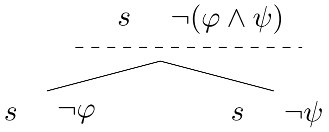
double negation elimination:

$$\begin{array}{cc} s & \neg\neg\varphi \\ \hdashline s & \varphi \end{array}$$

positive conjunction:

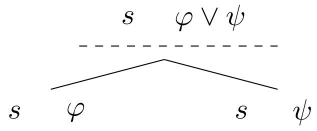
$$\begin{array}{cc} s & \varphi \wedge \psi \\ \hdashline s & \varphi \\ s & \psi \end{array}$$

negative conjunction:



$$\begin{array}{cccc} & s \quad \neg(\varphi \wedge \psi) & & \\ \hdashline s & \neg\varphi & s & \neg\psi \end{array}$$

positive disjunction:



$$\begin{array}{cccc} & s \quad \varphi \vee \psi & & \\ \hdashline s & \varphi & s & \psi \end{array}$$

negative disjunction:

$$\begin{array}{cc} s & \neg(\varphi \vee \psi) \\ \hdashline s & \neg\varphi \\ s & \neg\psi \end{array}$$

positive implication:

$$\begin{array}{cc} \multicolumn{2}{c}{s \quad \varphi \to \psi} \\ \hdashline s \quad \neg\varphi & s \quad \psi \end{array}$$

negative implication:

$$\begin{array}{l} s \quad \neg(\varphi \to \psi) \\ \hdashline s \quad \varphi \\ s \quad \neg\psi \end{array}$$

positive equivalence:

$$\begin{array}{cc} \multicolumn{2}{c}{s \quad \varphi \leftrightarrow \psi} \\ \hdashline s \quad \varphi & s \quad \neg\varphi \\ s \quad \psi & s \quad \neg\psi \end{array}$$

negative equivalence:

$$\begin{array}{cc} \multicolumn{2}{c}{s \quad \neg(\varphi \leftrightarrow \psi)} \\ \hdashline s \quad \varphi & s \quad \neg\varphi \\ s \quad \neg\psi & s \quad \psi \end{array}$$

positive box:

$$\begin{array}{c} s \quad \Box\varphi \\ \hdashline s^\frown n \quad \varphi \end{array}$$

negative box:

$$\begin{array}{c} s \quad \neg\Box\varphi \\ \hdashline s \quad \Diamond\neg\varphi \end{array}$$

where $s^\frown n$ must already occur as the sequence of some ancestor vertex of $\langle s^\frown n, \varphi \rangle$.

positive diamond:

$$\begin{array}{c} s \quad \Diamond\varphi \\ \hdashline s^\frown n \quad \varphi \end{array}$$

negative diamond:

$$\begin{array}{c} s \quad \neg\Diamond\varphi \\ \hdashline s \quad \Box\neg\varphi \end{array}$$

where $s^\frown n$ does not yet occur as the sequence of any ancestor vertex of $\langle s^\frown n, \varphi \rangle$. ◁

We can see a tableau intuitively as a search for a countermodel to the assumptions. Each branch corresponds to one possible construction of a Kripke model, with the worlds being defined by the sequences of the vertices. A vertex $\langle s, \varphi \rangle$ represents that the formula φ is true in world s, and the accessibility relation can be deduced as $s \; R \; s'$ if and only if $s' = s^\frown n$ for some number n. We can see a branch as all the vertices describing a true formula in their respective world simultaneously, while two different branches give two different possible models.

The closure rule makes sense in this context, since it closes a branch where a world s simultaneously makes both φ and $\neg\varphi$ true, something that is impossible in a Kripke model. A closed branch implies that the search for a countermodel has failed. Therefore, if all branches close, this means that a countermodel does not exist, and thus a closed tableau forms a proof of the validity of the entailment.

The classical rules now become clear as well. For example, if $\varphi \wedge \psi$ hold in a world s, then both φ and ψ also hold in world s. This corresponds to the positive conjunction rule. On the other hand, if $\varphi \vee \psi$ holds in s, then there are two ways this could be true, namely φ could be true or ψ could be true in s. Therefore, the search for a countermodel splits into two searches, one for each of these possibilities. In the negated rules, we make use of the De Morgan laws to get a similar construction, for example $\neg(\varphi \wedge \psi)$ is equivalent to $\neg\varphi \vee \neg\psi$, and thus the negative conjunction rule has a similar structure to the positive disjunction rule.

The rules for implication and equivalence can be explained by rewriting $\varphi \to \psi$ as $\neg\varphi \vee \psi$ and applying the disjunction rules, while $\varphi \leftrightarrow \psi$ can be explained by rewriting it as $(\varphi\wedge\psi)\vee(\neg\varphi\wedge\neg\psi)$ and first apply the disjunction rule, and then the conjunction rule in the resulting two branches.

The negative box and negative diamond rules are simply a rewriting of $\neg\Box\varphi$ to $\Diamond\neg\varphi$, and of $\neg\Diamond\varphi$ to $\Box\neg\varphi$. The positive box and positive diamond rules are the most delicate rules of the system, and are the only rules that have requirements to their conclusions.

The positive diamond rule can be seen as creating a new world $s^\frown n$ reachable from the original s. In the new world $s^\frown n$ we need φ to be true, to witness that $\Diamond\varphi$ is true in the old world s. However, as we do not know anything about this new world $s^\frown n$, apart form that φ is true in it, we must make sure that $s^\frown n$ did not yet appear earlier in the branch, to avoid making false assumptions about the new world.

The positive box rule has to ascertain that in every world reachable from s (so that means every $s^\frown n$ that has occurs somewhere on the branch) makes φ true, since in s it holds that $\Box\varphi$. However, we can not assume that there exists such a world $s^\frown n$, therefore the rule can only be applied on those sequences $s^\frown n$ that are already present when the rule is applied.

Except for the positive box rule, all the rules can be applied exactly once on a relevant vertex in each branch, since applying the same rule twice will result in the same new vertices, and vertices are not allowed to appear more than once on a branch. Usually during the construction of a tableau, it is helpful to mark those vertices on which the rule has been applied, to keep track of the vertices that can still be used to expand the tree. The positive box rule is the only rule that can be applied more than once on a certain vertex, as it can be applied once for every n such that $s^\frown n$ appears in the branch, so we have to take care not to mark vertices on which the positive box rule can be applied.

Example 6.6
We will prove that $\Diamond q, \Box(q \to \neg p) \vdash_{\mathsf{K}} \neg\Box p$

1.	() $\neg\neg\Box p$		assumption
2.	() $\Diamond q$		assumption
3.	() $\Box(q \to \neg p)$		assumption
4.	() $\Box p$		1. $\neg\neg$
5.	(1) q		2. $\Diamond^+$
6.	(1) $q \to \neg p$		3. $\Box^+$
7.	(1) p		4. $\Box^+$
8.	(1) $\neg q$ $\times$	(1) $\neg p$ $\times$	6. $\to^+$

The numbers on the left denote the level of the tree, as an index for the sequence of formulas in each branch. On the right are comments explaining which of the derivation rules is being used to obtain the step. Of course, these comments are not part of the proof itself, and merely serve as an explanation. Since this proof is still rather simple, the single line of comments is enough. In more complex proofs, with multiple branches in which different things will happen, it might become necessary to give individual comments for each branch.

The centre column consists of the vertices of the tree. We see that line 1 contains the negation $\neg\neg\Box p$ of the formula $\Box p$ that is to be proved. Lines lines 2 and 3 contain the formulas that are the assumptions. Line 4 is the result of the double negation elimination rule on line 1, which is denoted by the comment "1. $\neg\neg$". All of the top four vertices are formulas paired with the empty sequence.

Line 5 is the result of the positive diamond rule applied to line 2. Since the diamond rule makes the sequence longer, we see that the vertex of line 5 is a formula paired with the sequence (1). We will comment on the positive and negative operator rules by using the notation $\triangle^{\pm}$, where $\triangle$ is the operator, and $\pm$ is either $+$ or $-$ depending on if the rule is positive or negative.

In lines 6 and 7 we have the result of applying the positive box rule to lines 3 and 4, since the box rules can now be applied for the new sequence (1) from line 5. Finally line 8 is the result of applying the positive implication rule to line 6, resulting in two separate branches. Both branches close: the left one since it contains both $\langle(1), \neg q\rangle$ and $\langle(1), q\rangle$ (on line 8 and 5 respectively) and the right one since it contains both $\langle(1), \neg p\rangle$ and $\langle(1), p\rangle$ (on line 8 and 7 respectively). $\triangleleft$

Example 6.7
We will prove that $\Box(p \to \Box q), \Diamond\Diamond r \vdash_{\mathsf{K}} \Diamond\Diamond(q \wedge r) \vee \Diamond\neg p$:

1.	$()$	$\neg(\Diamond\Diamond(q \wedge r) \vee \Diamond\neg p)$			assumption
2.	$()$	$\Box(p \to \Box q)$			assumption
3.	$()$	$\Diamond\Diamond r$			assumption
4.	$()$	$\neg\Diamond\Diamond(q \wedge r)$			1. $\vee^-$
5.	$()$	$\neg\Diamond\neg p$			1. $\vee^-$
6.	$()$	$\Box\neg\Diamond(q \wedge r)$			4. $\Diamond^-$
7.	$()$	$\Box\neg\neg p$			5. $\Diamond^-$
8.	(1)	$\Diamond r$			3. $\Diamond^+$
9.	(1)	$p \to \Box q$			2. $\Box^+$
10.	(1)	$\neg\Diamond(q \wedge r)$			6. $\Box^+$
11.	(1)	$\neg\neg p$			7. $\Box^+$
12.	(1)	$\Box\neg(q \wedge r)$			10. $\Diamond^-$
13.	(1)	$\neg p$	(1)	$\Box q$	9. $\to^+$
14.		$\times$	$(1,1)$	r	8. $\Diamond^+$
15.			$(1,1)$	q	13. $\Box^+$
16.			$(1,1)$	$\neg(q \wedge r)$	12. $\Box^+$
17.	$(1,1)$	$\neg q$	$(1,1)$	$\neg r$	16. $\wedge^-$
		$\times$		$\times$	

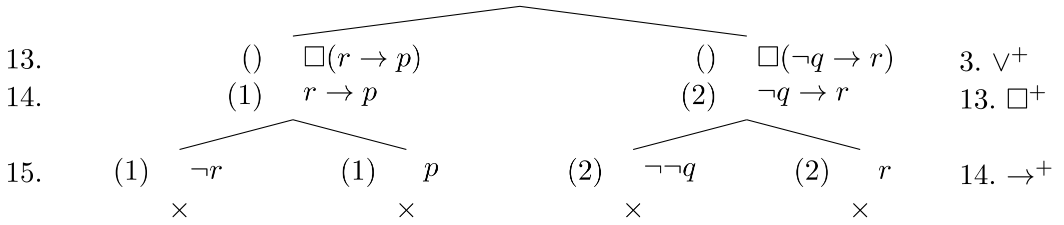
In this proof we see a positive diamond rule being applied multiple times within a branch. The first time in step 8, to create the sequence (1), and the second time in step 14, to create $(1, 1)$. These two rules operate on a different level, since step 8 appends a new number to the empty sequence, whereas step 14 appends a new number to the sequence (1). $\triangleleft$

In the next example we will see two positive diamond rules operating on the same level.

Example 6.8
We will prove that $\Diamond r \wedge \Diamond\neg r, \Box(r \to p) \vee \Box(\neg q \to r) \vdash_{\mathsf{K}} \Diamond(p \vee q)$.

In this proof, two different positive diamond rules are used in the same branch, so two different sequences (1) and (2) are created. The positive box rule can be applied once for each of those sequences, such as happens with the positive box being applied on line 4 to obtain both line 8 and line 11.

1.	()	$\neg\Diamond(p \vee q)$	assumption
2.	()	$\Diamond r \wedge \Diamond\neg r$	assumption
3.	()	$\Box(r \to p) \vee \Box(\neg q \to r)$	assumption
4.	()	$\Box\neg(p \vee q)$	1. $\Diamond^-$
5.	()	$\Diamond r$	2. $\wedge^+$
6.	()	$\Diamond\neg r$	2. $\wedge^+$
7.	(1)	r	5. $\Diamond^+$
8.	(1)	$\neg(p \vee q)$	4. $\Box^+$
9.	(1)	$\neg p$	8. $\vee^-$
10.	(2)	$\neg r$	6. $\Diamond^+$
11.	(2)	$\neg(p \vee q)$	4. $\Box^+$
12.	(2)	$\neg q$	11. $\vee^-$



◁

SOUNDNESS

Before we prove the soundness of modal tableaux for K, we will introduce the concept of satisfiability for branches.

Definition 6.9 — *Satisfiability of branches*
A branch $\langle s_1, \varphi_1\rangle, \ldots, \langle s_n, \varphi_n\rangle$ of a tableau is **satisfiable** if there exists a model $\mathcal{M}$ with worlds $w_1, \ldots, w_n$ such that:

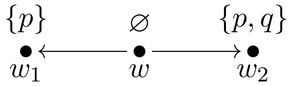
- $w_i = w_j$ whenever $s_i = s_j$,
- $w_i \; R \; w_j$ whenever there is some integer k such that $s_i{}^\frown k = s_j$,
- $\mathcal{M}, w_i \vDash \varphi_i$ for every i.

A tableau is satisfiable if at least one of its branches is satisfiable. ◁

Example 6.10
Consider the single branch of the following tableau:

1.	()	$\neg\Box(p \to q)$	assumption
2.	()	$\Box p$	assumption
3.	()	$\Diamond q$	assumption
4.	()	$\Diamond\neg(p \to q)$	1.$\Box^-$
5.	(1)	$\neg(p \to q)$	4.$\Diamond^+$
6.	(1)	$\neg q$	5. $\to^-$
7.	(1)	p	5. $\to^-$
8.	(2)	q	3.$\Diamond^+$
9.	(2)	p	2.$\Box^+$

To show that this branch is satisfiable, we have to find w, w_1 and w_2, where w corresponds to the sequence $()$, w_1 to (1) and w_2 to (2). The words must be related such that $w\ R\ w_1$ and $w\ R\ w_2$, since $()^\frown 1 = (1)$ and $()^\frown 2 = (2)$. Finaly we need each world to make all formulas true that correspond to their associated sequence. This is indeed possible, as we can check in the following model:



◁

Lemma 6.11 — *Satisfiable branch is open*
If a branch is satisfiable, then it is open. Consequently, if a tableau is satisfiable, it contains an open branch. ◁

Proof. This is Exercise 6.5. ■

The proof of the soundness theorem depends heavily on the following lemma. It implies that if a tableau is satisfiable, then we can not extend it to become a closed tableau. This means that if we start with a satisfiable tableau, it is impossible to find a proof using that tableau, since none of the branches will become closed after the application of derivation rules.

Lemma 6.12 — *Extension of satisfiable tableau is satisfiable*
If a tableau is satisfiable, then the tableau that results from applying any derivation rule will still be satisfiable. ◁

Proof. Let T be a tableau with the assumptions $\neg\varphi, \psi_1, \ldots, \psi_n$. Then Lemma 6.11 gives us that T is open if it is satisfiable.

Now suppose that T is a tableau that contains a satisfiable branch. Then we will show that T will still have a satisfiable branch after we apply any of the derivation rules for the system of modal tableaux for K. In this proof we will just discuss the cases for applying the $\vee^+$, $\to^-$, $\Box^+$ and $\Diamond^+$ rules. The other cases are left as Exercise 6.4

Positive disjunction rule. Let T be a tableau with a satisfiable branch B with vertices $\langle s_1, \varphi_1\rangle, \ldots, \langle s_n, \varphi_n\rangle$, and let $\langle s_i, \varphi_i\rangle$ be a vertex on which the $\vee^+$ rule can be applied. Then we see that $\varphi_i = \psi \vee \chi$. After we apply the $\vee^+$ rule on $\langle s_i, \varphi_i\rangle$, we get a tableau T' that contains two branches B_1' with vertices $\langle s_1, \varphi_1\rangle, \ldots, \langle s_n, \varphi_n\rangle, \langle s_i, \psi\rangle$ and B_2' with vertices $\langle s_1, \varphi_1\rangle, \ldots, \langle s_n, \varphi_n\rangle, \langle s_i, \chi\rangle$.

Since B is satisfiable, let $\mathcal{M}$ be a model for the branch B with worlds $w_1, \ldots, w_n$ as in Definition 6.9. Then $\mathcal{M}, w_i \vDash \varphi_i$, which was equal to $\mathcal{M}, w_i \vDash \psi \vee \chi$, and thus $\mathcal{M}, w_i \vDash \psi$ or $\mathcal{M}, w_i \vDash \chi$. In the first case we see that B_1' is satisfiable, since the $(n+1)$-th vertex of B_1' is $\langle s_i, \psi\rangle$, and thus we can take $w_{n+1} = w_i$ to see that $\mathcal{M}, w_{n+1} \vDash \psi$. And in the second case we see similarly that B_2' is satisfiable.

Therefore T' contains a satisfiable branch (B_1' or B_2').

Negative implication rule. Let T be a tableau with a satisfiable branch B with vertices $\langle s_1, \varphi_1\rangle, \ldots, \langle s_n, \varphi_n\rangle$, and let $\langle s_i, \varphi_i\rangle$ be a vertex on which the $\to^-$ rule can be applied, i.e. $\varphi_i = \neg(\psi \to \chi)$. Applying the $\to^-$ rule on $\langle s_i, \varphi_i\rangle$ gives a tableau T' that contains a branch B' with vertices $\langle s_1, \varphi_1\rangle, \ldots, \langle s_n, \varphi_n\rangle, \langle s_i, \psi\rangle, \langle s_i, \neg\chi\rangle$.

Let $\mathcal{M}$ be a model for the B with worlds $w_1, \ldots, w_n$ as in Definition 6.9, then $\mathcal{M}, w_i \vDash \neg(\psi \to \chi)$, and thus $\mathcal{M}, w_i \vDash \psi$ and $\mathcal{M}, w_i \vDash \neg\chi$. Therefore B' is satisfiable as well, by letting $w_{n+1} = w_{n+2} = w_i$, and thus T' contains a satisfiable branch.

Positive box rule. Let T be a tableau with a satisfiable branch B with vertices $\langle s_1, \varphi_1 \rangle, \ldots, \langle s_n, \varphi_n \rangle$, and let $\langle s_i, \varphi_i \rangle$ be a vertex on which the $\Box^+$ rule can be applied. Then $\varphi_i = \Box\psi$ and there are some numbers k and j such that $s_j = s_i{}^\frown k$. Applying the $\Box^+$ rule on $\langle s_i, \varphi_i \rangle$ gives a tableau T' that contains a branch B' with vertices $\langle s_1, \varphi_1 \rangle, \ldots, \langle s_n, \varphi_n \rangle, \langle s_i{}^\frown k, \psi \rangle$.

Let $\mathcal{M}$ be a model with worlds $w_i, \ldots, w_n$ as in Definition 6.9, then we see that $w_i \mathrel{R} w_j$ by the second bullet point of Definition 6.9. Since $\mathcal{M}, w_i \vDash \Box\psi$, we see that $\mathcal{M}, w_j \vDash \psi$, which shows that B' is satisfiable.

Positive diamond rule. Let T be a tableau with a satisfiable branch B with vertices $\langle s_1, \varphi_1 \rangle, \ldots, \langle s_n, \varphi_n \rangle$, and let $\langle s_i, \varphi_i \rangle$ be a vertex on which the $\Diamond^+$ rule can be applied. Then $\varphi_i = \Diamond\psi$ and there is some number k such that $s_i{}^\frown k$ does not occur in the branch yet. Applying the $\Diamond^+$ rule on $\langle s_i, \varphi_i \rangle$ gives a tableau T' that contains a branch B' with vertices $\langle s_1, \varphi_1 \rangle, \ldots, \langle s_n, \varphi_n \rangle, \langle s_i{}^\frown k, \psi \rangle$.

Let $\mathcal{M}$ be a model with worlds $w_i, \ldots, w_n$ as in Definition 6.9, then $\mathcal{M}, w_i \vDash \Diamond\psi$ implies that there is a world v such that $w_i \mathrel{R} v$ and $\mathcal{M}, v \vDash \psi$. Define $w_{n+1} = v$, then we see that B' is also satisfiable. ■

Theorem 6.13 — *Soundness theorem*
The system of modal tableaux for K is sound with respect to Kripke semantics. ◁

Proof. We have to prove that $\Gamma \vdash_{\mathsf{K}} \varphi$ implies $\Gamma \vDash \varphi$. Let $\psi_1, \ldots, \psi_n \in \Gamma$ and let T be the following tableau:

$$\begin{array}{rcll}
1. & () & \neg\varphi & \text{assumption} \\
2. & () & \psi_1 & \text{assumption} \\
 & & \vdots & \\
n+1. & () & \psi_n & \text{assumption}
\end{array}$$

By Lemma 6.12 we know that if T is satisfiable, then every tableau extending T is satisfiable. This implies that if there exists a model $\mathcal{M}$ and world w such that $\mathcal{M}, w \vDash \psi$ for every $\psi \in \Gamma$ and $\mathcal{M}, w \vDash \neg\varphi$, then this model shows that the tableau T above is satisfiable. In fact, this $\mathcal{M}$ shows that any tableau with only assumptions from $\Gamma \cup \{\neg\varphi\}$ is satisfiable. By Lemma 6.11 we then see that any tableau with assumptions from $\Gamma \cup \{\neg\varphi\}$ is open, and therefore there exists no closed tableau with such assumptions. It follows that $\Gamma \nvdash_{\mathsf{K}} \varphi$.

By contraposition we then see that $\Gamma \vdash_{\mathsf{K}} \varphi$ implies that if $\mathcal{M}, w \vDash \psi$ for every $\psi \in \Gamma$, then $\mathcal{M}, w \nvDash \neg\varphi$, and thus $\mathcal{M}, w \vDash \varphi$, which is exactly what $\Gamma \vDash \varphi$ means. ■

Completeness

There are two main ways that completeness of the tableau method is usually proved. In the next chapter we will give one of the ways, namely by constructing a *canonical model*, which is the same method we use for the completeness of Hilbert systems. The other way is more closely connected and perhaps more natural for tableau methods, and this is what will be treated in this section. The only caveat is that we will not be able to prove *strong* completeness without knowledge of some particular model theoretic tools, therefore this section will only prove *weak* completeness. That is, we will show that if φ is a general validity (so, $\vDash \varphi$), then there exists a tableau for $\neg\varphi$ (so, $\vdash_{\mathsf{K}} \varphi$).

The way the proof of completeness happens is very similar to the way we proved soundness. If there is no closed tableau for a certain set of assumptions, then there is no proof, and thus every tableau with those assumptions will have an open branch. The open branch will provide us with a counter model if it is satisfiable.

We already saw that if a tableau is satisfiable, then the result from applying any derivition rule will still be satisfiable. We can use this to keep applying derivation rules until the tableau is as large as it can possibly get: we keep going until there are no vertices left on which a rule can be applied.

For this we need to make precise what we mean by that it is possible to apply a rule on a vertex or not.

Definition 6.14 — *Active vertices*
A vertex $\langle s, \varphi \rangle$ in a tableau is **active** in an open branch if there are no rules that could be applied to the vertex. In particular, $\langle s, \varphi \rangle$ is active if ...

- ... φ is not of the form $\Box\psi$ or $\Diamond\psi$ and there is a derivation rule that can be applied to the vertex with a result that does not occur on the branch.
- ... φ is of the form $\Diamond\psi$ and for any vertex $\langle s^\frown n, \psi \rangle$ on the branch there is some ancestor $\langle s^\frown n, \chi \rangle$ of $\langle s^\frown n, \psi \rangle$ that also uses the sequence $s^\frown n$.
- ... φ is of the form $\Box\psi$ and for some n there is a vertex with sequence $s^\frown n$ on the branch, but there is no vertex $\langle s^\frown n, \psi \rangle$ on the branch.

A vertex that is not active or is part of a closed branch is **inactive**. ◁

Example 6.15
Consider the following tableau:

1.	()	$\neg(\Box r \to \Box p)$	assumption
2.	()	$\Diamond(p \wedge \Diamond q)$	assumption
3.	()	$\Diamond r$	assumption
4.	()	$\neg\Box p$	
5.	()	$\Box r$	
6.	()	$\Diamond\neg p$	
7.	(1)	$p \wedge \Diamond q$	
8.	(1)	r	
9.	(2)	$\neg p$	

We will investigate which vertices are active and which are inactive.

- Line 1 is inactive, since only the $\to^-$ rule could be applied to it, but the conclusions $\langle (), \neg\Box p \rangle$ and $\langle (), \Box r \rangle$ are already present on the branch in lines 4 and 5.
- Line 2 is also inactive, since line 7 contains a vertex $\langle (1), p \wedge \Diamond q \rangle$, but there is no ancestor of this vertex that also uses the sequence (1).

- Line 3 is active: there is a vertex $\langle (n), r\rangle$ on the branch, namely $\langle (1), r\rangle$ on line 8, but there is an ancestor of this vertex that already uses the sequence (1) (namely line 7).
- Line 4 is inactive, since the $\Box^-$ rule has been applied to it, giving its conclusion $\langle (), \Diamond\neg p\rangle$ in line 6.
- Line 5 is active, since there are vertices with the sequence (2) on the branch, but there is no vertex $\langle (2), r\rangle$. Note that line 5 is active even though its conclusion $\langle (1), r\rangle$ already appears in line 8: the positive box rule can be applied multiple times on the same vertex.
- Line 6 is inactive, since there is the vertex $\langle (2), \neg p\rangle$ on line 9 and no ancestor of this branch uses the sequence (2).
- Line 7 is active, since the $\wedge^+$ rule has not been applied to it: its conclusions $\langle (1), p\rangle$ and $\langle (1), \Diamond q\rangle$ do not occur on the branch.
- Line 8 is inactive, since there are no rules that can be applied to vertices with atomic variables.
- Line 9 is inactive, since there are no rules that can be applied to vertices with the negation of an atomic variable. ◁

Definition 6.16 — *Exhausted tableau*
A tableau is **exhausted** if all of its vertices are inactive. ◁

We are especially interested in exhausted tableaux, since open branches in exhausted tableaux will contain all the information we need to construct a counter model. Ideally there is a way to construct an exhausted tableau for any given set of assumptions, since this would mean that we either find a proof if all branches will become closed, or we find an open branch from which we can construct a counter model. Luckily, for the logic K we can give such a method of constructing exhausted tableaux.

Lemma 6.17 — *Existence of exhausted tableaux in* K
For any set of assumptions $\{\neg\varphi, \psi_1, \ldots, \psi_n\}$ there exists an exhausted tableau in the system K. ◁

Proof. Let T be a non-exhausted finite tableau. We will only allow the $\Diamond^+$ rule to be applied if no other derivation rules can be applied on any vertex in the same branch. Furthermore we only allow the $\Diamond^+$ rule to be applied to those vertices with a sequence of minimal length compared with the sequences of all vertices on which the $\Diamond^+$ rule can be.

The claim is that following this method will result in an exhausted tableau. This follows from the following observations:

1. Except for the $\Box^+$ rule, any of the derivation rules can be applied exactly once on an active vertex, after which the vertex becomes (and stays) inactive.
2. The $\Diamond^+$ rule is the only rule that allows new sequences to be created. Because of this, if a vertex $\langle s, \Box\varphi\rangle$ is inactive it can only become active after applying the $\Diamond^+$ rule on some other vertex $\langle s, \Diamond\psi\rangle$. After the $\Diamond^+$ rule is applied on such a vertex, the vertex $\langle s, \Box\varphi\rangle$ becomes inactive again after applying the $\Box^+$ rule.
3. For each of the derivation rules, the number of logical symbols in the formulas of the conclusion are less than or equal to the number of logical symbols in the formula of the vertex on which the rule was applied.

Given a non-exhausted finite tableau, we can apply derivation rules following our method to get a tableau in which the only active vertices are vertices on which the $\Diamond^+$ rule is applicable. This can be done in finitely many steps, since by point

3. the complexity of the formula (roughly, this means the number of symbols) can not increase. There are only finitely many sequences present in the tableau, and there exist only finitely many formulas of a given complexity using the same atomic variables, therefore the number of vertices that can be added to the tree without introducing new sequences is finite as well.

Next, let d be the least length of a sequence s in any active vertex $\langle s, \Diamond\varphi\rangle$. We will apply the $\Diamond^+$ rule to those active vertices that have a sequence of length d. This will create new vertices that possibly make some of the vertices on which the $\Box^+$ rule can be applied active once again. However, each of these new vertices will be of the form $\langle s^\frown n, \varphi\rangle$, and therefore has a sequence of length $d+1$. Thus the only vertices on which the $\Box^+$ rule can be applied are those vertices with a sequence of length d.

Let j be the least length of a path from the root to a vertex with a sequence of length d. Then at this stage, none of the vertices that have a distance less than j to the root are active, and they will never become active after applying new rules. It follows that we can repeat this process to make every vertex eventually inactive. ■

In an exhausted tableau we cannot apply any rules to make it larger. Now, if the assumptions of the tableau are satisfiable, then the exhausted tableau has to be satisfiable by Lemma 6.12.

Lemma 6.18 — *Counter model method for* K
If T is an open exhausted tableau with assumptions $\{\neg\varphi, \psi_1, \dots, \psi_n\}$, then $\{\psi_1, \dots, \psi_n\} \nvDash \varphi$. ◃

Proof. This follows, since an open branch in an exhausted tableau T can be used to build a model in which both $\psi_1, \dots, \psi_n$ is true and φ is false. The proposed model is built exactly as in Definition 6.9.

Let $B = (\langle s_0, \varphi_0\rangle, \langle s_1, \varphi_1\rangle, \dots, \langle s_i, \varphi_i\rangle, \dots)$ be an open branch in T. Let the model $\mathcal{M} = \langle W, R, V\rangle$ be defined such that:

- $W = \{s_i \mid s_i \text{ is a sequence for some vertex in } b\}$,
- $s_i \, R \, s_j$ if and only if $s_j = s_i{}^\frown n$ for some $n \in \mathbb{N}$, and
- for any atom p we have $p \in V(s_i)$ if and only if $\langle s_i, p\rangle$ is a vertex in B.

We can see that since B is open, if $\langle s, \neg p\rangle$ appears on the branch, then $p \notin V(s)$.

We will prove that for any $\langle s_i, \varphi_i\rangle$ in the branch B we have that $\mathcal{M}, s_i \vDash \varphi_i$. In particular, we will see that this holds for the assumptions $\{\neg\varphi, \psi_1, \dots, \psi_n\}$ of the tableau T, and thus $\mathcal{M}, s_0 \vDash \psi_1 \wedge \cdots \psi_n \wedge \neg\varphi$ gives us the needed counter model. The proof is done by induction on the complexity of φ_i. We will give a few of the steps in the induction.

Base case. If $\varphi_i = p$ is an atomic variable, then by definition of the valuation V we have $p \in V(s_i)$, and thus $\mathcal{M}, s_i \vDash p$. Similarly, if $\varphi_i = \neg p$ is the negation of an atomic variable, $\mathcal{M}, s_i \nvDash p$, and thus $\mathcal{M}, s_i \vDash \neg p$.

For the induction cases, assume as induction hypothesis that for any $j \in \mathbb{N}$ if $\langle s_j, \psi\rangle$ and $\langle s_j, \chi\rangle$ are vertices on B, then $\mathcal{M}, s_j \vDash \psi$ and $\mathcal{M}, s_j \vDash \chi$.

Positive disjunction. If $\varphi_i = \psi \vee \chi$, then because the tableau is exhausted, the $\vee^+$ rule has been applied on $\langle s_i, \psi \vee \chi\rangle$. Therefore either $\langle s_i, \psi\rangle$ or $\langle s_i, \chi\rangle$ is also a vertex on the branch B. By the induction hypothesis we see that $\mathcal{M}, s_i \vDash \psi$ or $\mathcal{M}, s_i \vDash \chi$, therefore $\mathcal{M}, s_i \vDash \psi \vee \chi$.

Negative disjunction. If $\varphi_i = \neg(\psi \vee \chi)$, then because the tableau is exhausted, the $\vee^-$ rule has been applied on $\langle s_i, \neg(\psi \vee \chi)\rangle$. Therefore both $\langle s_i, \neg\psi\rangle$ and $\langle s_i, \neg\chi\rangle$ are vertices on the branch B. By the induction hypothesis we see that $\mathcal{M}, s_i \vDash \neg\psi$ and $\mathcal{M}, s_i \vDash \neg\chi$, therefore $\mathcal{M}, s_i \vDash \neg(\psi \vee \chi)$.

Positive necessitation. If $\varphi_i = \Box\psi$, then because the tableau is exhausted, for any $n \in \mathbb{N}$ for which the sequence $s_i{}^\frown n$ appears on B we see that the $\Box^+$ rule has been applied on $\langle s_i, \Box\psi\rangle$, and thus that the vertex $\langle s_i{}^\frown n, \psi\rangle$ is on B. By the induction hypothesis we see that $\mathcal{M}, s_i{}^\frown n \vDash \psi$ for any such $n \in \mathbb{N}$. By how we defined the accessibility relation R, we know that $s_i \; R \; s_j$ if and only if $s_j = s_i{}^\frown n$ for some $n \in \mathbb{N}$ such that s_j appears on the branch B, and thus for all worlds s_j reachable from s_i it holds that $\mathcal{M}, s_j \vDash \psi$. Therefore $\mathcal{M}, s_i \vDash \Box\psi$. ■

To see that this implies weak completeness, we need to show that $\vDash \varphi$ implies $\vdash_{\mathsf{K}} \varphi$. So by contraposition we could assume that $\nvdash_{\mathsf{K}} \varphi$, which means that every tableau with assumption $\neg\varphi$ is open. By Lemma 6.17 we can assume that there is an open exhausted tableau with assumption $\neg\varphi$. Then Lemma 6.18 shows that there is a model for $\neg\varphi$, and thus $\nvDash \varphi$.

Theorem 6.19 — *Weak completeness theorem for tableaux in* K
The tableau method for K is weakly complete with respect to Kripke semantics. $\lhd$

To finish this section, we will give an example of the construction presented in Lemma 6.17. The tableau will be open, thus we will also use it to demonstrate the construction of a counter model presented in Lemma 6.18.

Example 6.20
We will show that $\Diamond\Diamond\varphi \to \Diamond\varphi \nvdash_{\mathsf{K}} \Box\varphi \to \Box\Box\varphi$. We therefore start with a tableau with assumptions $\neg(\Box p \to \Box\Box\Diamond p)$ and $\Diamond\Diamond p \to \Diamond p$ and apply derivation rules until the only active vertices are those with a formula $\Diamond\psi$ for some ψ:

1.	()	$\neg(\Box p \to \Box\Box p)$	assumption
2.	()	$\Diamond\Diamond p \to \Diamond p$	assumption
3.	()	$\Box p$	1. $\to^-$
4.	()	$\neg\Box\Box p$	1. $\to^-$
5.	()	$\Diamond\neg\Box p$	4. $\Box^-$

6.	()	$\neg\Diamond\Diamond p$	2. $\to^+$	()	$\Diamond p$	2. $\to^+$
7.	()	$\Box\neg\Diamond p$	6. $\Diamond^-$			

It is easy to see that only the vertex in line 5 and the vertex in line 6 on the right branch are active. We can now apply the $\Diamond^+$ rule on all of these vertices to get the following tree. Note that this made the vertex in line 3 and the vertex in line 7 of the left branch active, even though they were previously inactive.

1.	()	$\neg(\Box p \to \Box\Box p)$	assumption
2.	()	$\Diamond\Diamond p \to \Diamond p$	assumption
3.	()	$\Box p$	1. $\to^-$
4.	()	$\neg\Box\Box p$	1. $\to^-$
5.	()	$\Diamond\neg\Box p$	4. $\Box^-$

6.	()	$\neg\Diamond\Diamond p$	2. $\to^+$	()	$\Diamond p$	2. $\to^+$
7.	()	$\Box\neg\Diamond p$	6. $\Diamond^-$	(1)	$\neg\Box p$	5. $\Diamond^+$
8.	(1)	$\neg\Box p$	5. $\Diamond^+$	(2)	p	6. $\Diamond^+$

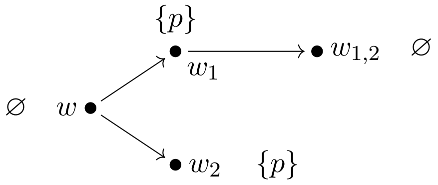
Now we once again apply derivation rules until the only active vertices those on which $\Diamond^+$ can be applied, and then apply the $\Diamond^+$ rule on those vertices. After this it turns out we will be left with the following exhausted tableau:

1.	$()$	$\neg(\Box p \to \Box\Box p)$	assumption
2.	$()$	$\Diamond\Diamond p \to \Diamond p$	assumption
3.	$()$	$\Box p$	1. $\to^-$
4.	$()$	$\neg\Box\Box p$	1. $\to^-$
5.	$()$	$\Diamond\neg\Box p$	4. $\Box^-$

6.	$()$	$\neg\Diamond\Diamond p$	2. $\to^+$	$()$	$\Diamond p$	2. $\to^+$
7.	$()$	$\Box\neg\Diamond p$	6. $\Diamond^-$	(1)	$\neg\Box p$	5. $\Diamond^+$
8.	(1)	$\neg\Box p$	5. $\Diamond^+$	(2)	p	6. $\Diamond^+$
9.	(1)	p	3. $\Box^+$	(1)	p	3. $\Box^+$
10.	(1)	$\Diamond\neg p$	8. $\Box^-$	(1)	$\Diamond\neg p$	7 $\Box^-$
11.	$(1, 2)$	$\neg p$	10. $\Diamond^+$	$(1, 2)$	$\neg p$	10. $\Diamond^+$

Both branches are open, since there are no two vertices with the same sequence that have both a formula ψ and its negation $\neg\psi$. This means that we can construct a counter model using Lemma 6.18. We will use the right branch to build a counter model, but the left branch would work just as fine.

We have four different sequences appearing on the right branch, namely $()$, (1), (2) and $(1, 2)$. Therefore we let $W = \{w, w_1, w_2, w_{1,2}\}$ represent these sequences. The accessibility relation then has the relations $w\ R\ w_1$ and $w\ R\ w_2$ and $w_1\ R\ w_{1,2}$ and no other relations. Finally the valuation is defined such that $p \in V(w_1)$ and $p \in V(w_2)$, as both the vertices $\langle (1), p\rangle$ and $\langle (2), p\rangle$ appear on the right branch, and we have $V(w_{1,2}) = V(w) = \varnothing$, as there are no vertices $\langle (1, 2), p\rangle$ and $\langle (), p\rangle$ on the right branch. This model can be drawn as follows:



We leave it as an exercise to verify that $\mathcal{M}, w \vDash \Diamond\Diamond p \to \Diamond p$ while also $\mathcal{M}, w \nvDash \Box p \to \Box\Box p$. ◁

6.3 NORMAL MODAL LOGICS

We can formulate additional rules to create tableau systems for other normal logics. In the following definition we give rules that can be added to the tableau system for K to get tableau systems for the logics in Definition 5.17.

Definition 6.21 — *Modal tableau rules*
We have the following new rules

T rule:

$$\begin{array}{cc} s & \Box\varphi \\ \hdashline s & \varphi \end{array}$$

D rule:

$$\begin{array}{cc} s & \Box\varphi \\ \hdashline s & \Diamond\varphi \end{array}$$

B rule:

$$\begin{array}{cc} s^\frown n & \Box\varphi \\ \hdashline s & \varphi \end{array}$$

4 rule:

$$\begin{array}{cc} s & \Box\varphi \\ \hdashline s^\frown n & \Box\varphi \end{array}$$

where $s^\frown n$ must already occur as the sequence of some ancestor vertex of $\langle s^\frown n, \Box\varphi\rangle$.

5 rule:

$$\begin{array}{cc} s & \Diamond\varphi \\ \hdashline s^\frown n & \Diamond\varphi \end{array}$$

where $s^\frown n$ must already occur as the sequence of some ancestor vertex of $\langle s^\frown n, \Diamond\varphi\rangle$. ◁

Note that similar to the positive box rule, the 4 and 5 rules can only be applied if the sequence $s^\frown n$ was already present in the branch. Also note that the B rule is special in that it is the only rule that has a conclusion in a shorter sequence than the original vertex has.

In the tableaux system for K, one can show that every tableau is finite (see Exercise 6.8). This does not stay true if we add some of the new derivation rules. Consider for example the following attempt to prove that $\Box\varphi$ does not imply $\Box\Box\varphi$ in the system K45:

1.	$()$	$\neg\Box\Box\varphi$	assumption
2.	$()$	$\Box\varphi$	assumption
3.	$()$	$\Diamond\neg\Box\varphi$	1. $\Box^-$
4.	(1)	$\Diamond\neg\Box\varphi$	3. **5** rule
5.	$(1,1)$	$\Diamond\neg\Box\varphi$	4. **5** rule
6.	$(1,1,1)$	$\Diamond\neg\Box\varphi$	5. **5** rule
7.	$(1,1,1,1)$	$\Diamond\neg\Box\varphi$	6. **5** rule
		$\vdots$	

Even though this tableau is infinitely long, it still has an open branch (because it does not contain both $\langle s, \psi\rangle$ and $\langle s, \neg\psi\rangle$ for any sequence s and formula ψ). Therefore you might be tempted to immediately conclude that $\Box\varphi \nvdash_{\mathsf{K45}} \Box\Box\varphi$. This conclusion is wrong, however! We know that $\Box\varphi \vdash_{\mathsf{K45}} \Box\Box\varphi$ is correct, since we have the following closed tableau:

1.	$()$	$\neg\Box\Box\varphi$	assumption
2.	$()$	$\Box\varphi$	assumption
3.	$()$	$\Diamond\neg\Box\varphi$	1. $\Box^-$
4.	(1)	$\neg\Box\varphi$	3. $\Diamond^+$
5.	(1)	$\Box\varphi$	4. **4** rule
	$\times$		

This shows the problem with the first tableau: it is not exhausted. This means that the search for a closed tableau needs to be done a little more carefully than we did in the system K. The method that was proposed in Lemma 6.17 fortunately still works, and we could check that the proof for an analogue of Lemma 6.17 for K45 or any of the logics defined in Definition 5.17 is exactly as for the case in K.

We will give one example of a soundness proof for the rules from Definition 6.21 with respect to the usual frame classes. The proofs for the other rules will be left as an exercise, since the process is very similar for each of them.

Theorem 6.22 — *Soundness theorem*
- the system of modal tableaux for KT is sound with respect to reflexive frames.
- the system of modal tableaux for KB is sound with respect to symmetric frames.
- the system of modal tableaux for KD is sound with respect to serial frames.
- the system of modal tableaux for $\mathsf{K4}$ is sound with respect to transitive frames.
- the system of modal tableaux for $\mathsf{K5}$ is sound with respect to Euclidean frames. ◁

Proof. We will show that the **B** rule is sound with respect to symmetric frames. As with the soundness theorem for K, the proof is based on the claim that if a tableau is satisfiable, then it will remain satisfiable after one of the rules is applied. In this case we restrict our attention to symmetric frames, thus for this section let a tableau T be called B-*satisfiable* if it has a branch that is satisfiable on a symmetric model. In other words, the model described in Definition 6.9 has as additional condition that R should be symmetric.

Suppose that T is B-satisfiable and that $\langle s, \varphi \rangle$ is an active vertex, then we have to prove that one of the branches of T is still B-satisfiable after application of a derivation rule on $\langle s, \varphi \rangle$. We will only show that this is the case for the **B** rule, as the proofs from Lemma 6.12 will work here as well.

B rule. Let T be a B-satisfiable tableau and let B be a B-satisfiable branch with vertices $\langle s_1, \varphi_1 \rangle, \ldots, \langle s_n, \varphi_n \rangle$, and let $\langle s_i, \varphi_i \rangle$ be a vertex on which the **B** rule can be applied, then $s_i = s_j^\frown n$ for some j and $\varphi_i = \Box\psi$. After applying the **B** rule, we get a tableau T' that contains a branch B' with vertices $\langle s_1, \varphi_1 \rangle, \ldots, \langle s_n, \varphi_n \rangle, \langle s_j, \psi \rangle$.

Let $\mathcal{M}$ be a model based on a symmetric frame with worlds $w_1, \ldots, w_n$ as in Definition 6.9. Then we see that $\mathcal{M}, w_i \vDash \Box\psi$, and since $s_i = s_j^\frown n$, we also see that $w_j \; R \; w_i$. Because R is symmetric we have $w_i \; R \; w_j$ as well, and thus $\mathcal{M}, w_i \vDash \Box\psi$ implies $\mathcal{M}, w_j \vDash \psi$. We see that $\mathcal{M}$ is also a model witnessing the B-satisfiability of B', and thus T' is also B-satisfiable. ■

We can also prove weak completeness for each of these logics with respect to their frame classes, since Lemma 6.17 translates perfectley for each of these logics. The only thing that we have to prove is an analogue for Lemma 6.18 for each of the logics. It follows from the soundness theorem that the model that is define in Lemma 6.18 will have the necessary frame property. We will leave the proof as Exercise 6.9.

6.4 EXERCISES

Exercise 6.1. Use the method of tableaux to find either a proof or a counter model for the following statements:

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

b) $\Diamond\Box\neg\varphi \vdash_{\mathsf{K}} \Box\Diamond\varphi$

c) $\Box(\varphi \to \psi), \neg\Diamond\psi, \Diamond\neg\varphi \to \Box\psi \vdash_{\mathsf{K}} \Box(\varphi \wedge \psi)$

d) $\Diamond\Diamond\neg\varphi, \Box\Diamond\varphi \vdash_{\mathsf{K}} \Diamond\Box\varphi \to \Box\Diamond\neg\varphi$
e) $\vdash_{\mathsf{K}} \Box\varphi \vee \Box\psi \to \Box(\varphi \vee \psi)$
f) $\vdash_{\mathsf{K}} \Diamond\varphi \vee \Diamond\psi \to \Diamond(\varphi \vee \psi)$
g) $\Box\Diamond\varphi \to \Diamond(\psi \vee \neg\varphi) \vdash_{\mathsf{K}} \neg\Box\Diamond\psi$

Exercise 6.2. Prove the following using the method of tableaux:
a) $\Diamond\Diamond\Diamond\Diamond\varphi \vdash_{\mathsf{K4}} \Diamond\varphi$
b) $\Box\Diamond\Box\varphi \vdash_{\mathsf{KTB}} \varphi$
c) $\Diamond\varphi \vdash_{\mathsf{KB5}} \Diamond\Diamond\Box\varphi$
d) $\Diamond\Box\Diamond\Diamond\Box\varphi \vdash_{\mathsf{S5}} \Diamond\varphi$
e) $\Box\varphi \vdash_{\mathsf{KD4}} \Diamond\Diamond\Diamond\varphi$

Exercise 6.3. Build a counter models for each of following statements by building an exhausted tableau and reading of the model from an open branch:
a) $\Diamond\varphi, \Diamond\Diamond\neg\varphi \vdash_{\mathsf{KTB}} \Box\Diamond\varphi$
b) $\vdash_{\mathsf{K5}} \Diamond\varphi \wedge \Diamond\psi \to \Diamond\Diamond(\psi \vee \varphi)$
c) $\Box\varphi \vdash_{\mathsf{KD4}} \Diamond\neg\varphi$

Exercise 6.4. Complete the proof of Lemma 6.12, by showing the cases for the other derivation rules.

Exercise 6.5. Prove Lemma 6.11.

Exercise 6.6. Prove that the Deduction theorem (see Theorem 5.9) holds for any tableaux systems that contains the positive and negative implication rules.

∗ **Exercise 6.7.** If Λ is a logic, then the **compactness theorem** states that given set of formulas Γ, if for any finite subset $\Delta = \{\psi_1, \ldots, \psi_n\} \subseteq \Gamma$ there is exists a model $\mathcal{M}_\Delta$ with world w such that $\mathcal{M}_\Delta, w \vDash \psi_1 \wedge \cdots \wedge \psi_n$ for any finite subset, then there is a model $\mathcal{M}_\Gamma$ with world w such that $\mathcal{M}_\Gamma, w \vDash \psi$ for all $\psi \in \Gamma$.

Suppose that Λ is weakly complete and satisfies the compactness theorem. Prove that Λ is strongly complete.[*]

∗ **Exercise 6.8.** Show that any tableau in the system K is finite.

Exercise 6.9. Prove Lemma 6.18 for the logic $\mathsf{KD4}$:
a) Let B is an open branch in an exhausted tableau T in the logic $\mathsf{KD4}$, then give an appropriate definiton for a model $\mathcal{M}$ constructed from B (as in the proof of Lemma 6.18) such that $\mathcal{M}$ is based on a serial transitive frame. Prove that your defined model is indeed a serial transitive model.
b) Do the induction case for positive necessitation: assume as induction hypothesis that if $\langle s_j, \psi\rangle$ is a vertex on B for any $j \in \mathbb{N}$, then $\mathcal{M}, s_j \vDash \psi$. Prove that if $\langle s_i, \Box\psi\rangle$ is a vertex on B, then $\mathcal{M}, s_i \vDash \Box\psi$.

---

[*] Since we can prove that K satisfies the compactness theorem using techniques from model theory, this exercise implies that by Theorem 6.19 the tableau system for K is strongly complete

7 Completeness

In this chapter we will show that the Hilbert systems from Chapter 5 are strongly complete. The same method can be adapted to prove strong completeness for the tableau method, but we will go through this chapter with Hilbert systems in mind.

Definition 7.1 — *Completeness*
Let $\mathcal{C}$ be a frame class and Λ a normal modal logic. Λ is **strongly complete** with respect to $\mathcal{C}$ if $\Gamma \vDash_{\mathcal{C}} \varphi$ implies $\Gamma \vdash_{\Lambda} \varphi$. Λ is **weakly complete** with respect to $\mathcal{C}$ if $\vDash_{\mathcal{C}} \varphi$ implies $\vdash_{\Lambda} \varphi$. ◁

We will prove the strong completeness of the system K and the logics from Definition 5.17 with respect to their frame classes. Since the proof will for the most part not depend on the logic Λ and the frame class $\mathcal{C}$, we will leave out the subscripts $\mathcal{C}$ for the frame class and Λ for the logic until these become relevant again.

The way we prove it is by building consistent sets of formulas and creating a very special model out of them. A consistent set of formulas is intuitively speaking a set of formulas that do not *provably* contradict each other. Formally we define consistent sets of formulas as follows.

Definition 7.2 — *Consistent*
A set of formulas Γ is **consistent** if $\Gamma \nvdash \neg\varphi$ for any $\varphi \in \Gamma$. Otherwise Γ is **inconsistent**. ◁

Inconsistent sets are special in that they can prove anything. This is a basic consequence of the *ex falso quodlibet* law from classical logic, that is, the law that $\bot \to \varphi$ is a tautology. In fact, inconsistent sets are the *only* sets that can prove anything.

Lemma 7.3
Γ is inconsistent if and only if $\Gamma \vdash \psi$ for every formula ψ. ◁

> *Proof.* ($\Rightarrow$) If Γ is inconsistent, let $\varphi \in \Gamma$ such that $\Gamma \vdash \neg\varphi$. Then $\Gamma \vdash \varphi$ as well, because $\varphi \in \Gamma$, so therefore $\Gamma \vdash \varphi \wedge \neg\varphi$. But $\varphi \wedge \neg\varphi \to \psi$ is a classical tautology, so we can derive $\Gamma \vdash \psi$.
>
> ($\Leftarrow$) If $\Gamma \vdash \psi$ for any ψ, then certainly $\Gamma \vdash \neg\varphi$ for any $\varphi \in \Gamma$. ■

Before we go any further with our proof of completeness, it might be good to give an intuitive idea of how it will be structured.

> *Sketch of a proof.* As said, the idea is to create a model based on consistent sets of formulas. To be precise, we will create a model whose worlds *are* consistent sets of formulas. This model will be called the **canonical model**. Remember that the set of worlds W in some model is just a set; it is not specified what kind of elements it has. We have already seen examples where the worlds were abstract symbols, such as w, v or u, and we have seen examples where the worlds were numbers, such as in Example 1.11. In this case the worlds of our model are consistent sets of formulas.

> The reason why we let the worlds be consistent sets of formulas, is that we want to define a valuation on the canonical model that makes all the formulas in a consistent set of formulas *valid* in the world that it equates to. Note we are talking about two different things here: if A is a consistent set of formulas, and φ is a formula, then we want that $\varphi \in A$ implies that $A \vDash \varphi$. The antecedent of this implication says that φ is an element in our set, and the consequent says that the φ is true in the world.
>
> This distinction makes it possible to connect the notion of provability to the notion of validity. To be more explicit, the fact that a world is a **consistent set** of formulas, means that it is impossible to prove a contradiction using just the formulas in the set. On the other hand, the fact that a world makes all its formulas **valid**, means that there is a model with a world where all the formulas in the consistent set are valid.
>
> Now remember that we are trying to show that $\Gamma \vDash \varphi$ implies $\Gamma \vdash \varphi$. By contraposition, this is the same as proving that $\Gamma \nvdash \varphi$ implies $\Gamma \nvDash \varphi$. We know from Lemma 7.3 that $\Gamma \nvdash \varphi$ means that Γ is consistent and that $\varphi \notin \Gamma$. If we could then somehow show that the formula φ is not valid in the world Γ of the canonical model, then we have found a model with a world in which all the formulas of Γ are true, but not the formula φ. In other words, we have found a counter model for $\Gamma \vDash \varphi$.
>
> The trick to showing that φ is not valid, is by making Γ as large as possible. We try to find a consistent set of formulas that is so large, that for any formula ψ it contains either ψ or $\neg\psi$. The main part of the proof is that we can indeed always find such a large consistent set for which $\Gamma \nvdash \varphi$ still holds. Since Γ is consistent, we can conclude that $\varphi \notin \Gamma$. Because Γ is as large as possible, we get that $\neg\varphi \in \Gamma$, and this is precisely what we need: in the canonical model we then see that $\neg\varphi$ is valid in the world Γ, and thus that $\Gamma \nvDash \varphi$ (both in the sense that the world Γ does not model the formula φ and that in general there exists a model $\mathcal{M}, w \vDash \Gamma$ for which $\mathcal{M}, w \nvDash \varphi$).

Let us first make concrete what we mean by *as large as possible*. We do this by introducing two equivalent definitions.

Definition 7.4 — *Complete set*
A set of formulas Γ is **complete** if it is consistent and for every formula φ either $\varphi \in \Gamma$ or $\neg\varphi \in \Gamma$. ◁

Definition 7.5 — *Maximal consistent set*
A set of formulas Γ is a **maximal consistent set** (MCS) if it is consistent and for any $\varphi \notin \Gamma$ the set $\Gamma \cup \{\varphi\}$ is inconsistent. ◁

We will now show that these two definitions are indeed equivalent to each other.

Lemma 7.6
A set of formulas Γ is complete if and only if it is an MCS. ◁

> *Proof.* ($\Rightarrow$) If Γ is complete, and φ is a formula such that $\varphi \notin \Gamma$, then $\neg\varphi \in \Gamma$. But then $\Gamma \cup \{\varphi\} \supseteq \{\varphi, \neg\varphi\}$ is not consistent.
>
> ($\Leftarrow$) If Γ is consistent, but not complete, then there is a formula φ such that $\varphi \notin \Gamma$ and $\neg\varphi \notin \Gamma$. If $\Gamma \vdash \varphi$ and $\Gamma \vdash \neg\varphi$, then Γ would be inconsistent. Therefore we can

without loss of generality assume that $\Gamma \nvdash \neg\varphi$, which makes $\Gamma \cup \{\varphi\}$ consistent. But that means that Γ is not an MCS. ■

It is clear from the structure of proofs in a Hilbert system that $\Gamma \vdash \varphi$ for any formula $\varphi \in \Gamma$. However, with maximal consistent sets, the converse is true as well.

Lemma 7.7 — *Deductive closure of maximal consistent sets*
If Γ is an MCS and $\Gamma \vdash \varphi$, then $\varphi \in \Gamma$. ◁

Proof. Suppose Γ is an MCS and $\Gamma \vdash \varphi$. If $\varphi \notin \Gamma$, then $\Gamma \cup \{\varphi\}$ would be inconsistent by maximality of Γ, so from Lemma 7.3 we see that $\Gamma \cup \{\varphi\} \vdash \neg\varphi$. But using the deduction theorem (Theorem 5.9) we then have $\Gamma \vdash \varphi \to \neg\varphi$. Together with $\Gamma \vdash \varphi$ we can then derive that $\Gamma \vdash \varphi \wedge \neg\varphi$, which makes Γ inconsistent. This is a contradiction, therefore $\varphi \in \Gamma$. ■

The main question remains of course whether consistent sets actually exist. For the proof of the completeness theorem we will need an even stronger claim, that for any consistent set of formulas Γ such that $\Gamma \nvdash \varphi$ we can find an MCS $\Gamma' \supseteq \Gamma$ such that $\Gamma' \nvdash \varphi$ as well. This is guaranteed by the following lemma, due to logician Adolf Lindenbaum.

Lemma 7.8 — *Lindenbaum's lemma*
If Γ is a consistent set, then there exists an MCS $\Gamma' \supseteq \Gamma$. ◁

Proof. Our language $\mathcal{L}_{\text{ML}}$ contains finitely many logical symbols and countably many atomic variables, therefore $\mathcal{L}_{\text{ML}}$ is countable. It is an elementary result of set theory that the set of finite strings over a countable alphabet is countable (see Exercise 7.4), and since formulas are finite strings over $\mathcal{L}_{\text{ML}}$, we see that the set of all modal formulas Fml_{ML} is countable. This means that we can fix an enumeration $\langle \varphi_n \mid n \in \mathbb{N} \rangle$ of the set Fml_{ML}.

We can now recursively construct the MCS Γ', by letting:

$$\begin{aligned} \Gamma_0 &:= \Gamma \\ \Gamma_{n+1} &:= \begin{cases} \Gamma_n \cup \{\varphi_n\} & \text{if } \Gamma_n \cup \{\varphi_n\} \text{ is consistent} \\ \Gamma_n & \text{if } \Gamma_n \cup \{\varphi_n\} \text{ is inconsistent} \end{cases} \\ \Gamma' &:= \bigcup_{n \in \mathbb{N}} \Gamma_n \end{aligned}$$

We have to check that Γ' is indeed an MCS. We will do this by showing Γ' is complete, from which it follows that Γ' is an MCS by Lemma 7.6.

First note that Γ_n is consistent for every $n \in \mathbb{N}$ by construction. Let ψ be any formula, then $\psi = \varphi_n$ and $\neg\psi = \varphi_m$ for some $n, m \in \mathbb{N}$. Without loss of generality we can assume that $n < m$. Since Γ_n is consistent we know that either $\Gamma_n \nvdash \varphi_m$ or $\Gamma_n \vdash \varphi_m$.

If $\Gamma_n \nvdash \varphi_m$, then $\Gamma_n \cup \{\varphi_n\}$ is consistent (because $\varphi_m \equiv \neg\varphi_n$), so $\varphi_n = \psi \in \Gamma_{n+1}$.

On the other hand, if $\Gamma_n \vdash \varphi_m$, then also $\Gamma_m \vdash \varphi_m$, since $\Gamma_m \supseteq \Gamma_n$ (see Exercise 5.8). But because Γ_m is consistent, this means that $\Gamma_m \cup \{\varphi_m\}$ is consistent. Therefore $\varphi_m = \neg\psi \in \Gamma'$.

We have now shown that for any formula ψ, at least one of ψ or $\neg\psi$ is in Γ'. What is left to do, is showing Γ' is consistent. Suppose it was not, then there is a formula

$\psi \in \Gamma'$ such that $\Gamma' \vdash \neg\psi$. Because proofs are finite we see that there is a finite set $\Delta := \{\psi_1, \ldots, \psi_n\} \subseteq \Gamma'$ such that $\Delta \vdash \neg\psi$. Since $\psi, \psi_1, \ldots, \psi_n$ are formulas, they must appear in our enumeration $\langle \varphi_n \mid n \in \mathbb{N} \rangle$, therefore we can find natural numbers $m, m_1, \ldots, m_n$ such that $\psi = \varphi_m$, and $\psi_i = \varphi_{m_i}$ for all $1 \leq i \leq n$. Let k be a natural number larger than all of $m, m_1, \ldots, m_n$. Then $\psi, \psi_i \in \Gamma_k$ for all $1 \leq i \leq n$. But this means that Γ_k contains ψ and $\Gamma_k \vdash \neg\psi$, so Γ_k is inconsistent. This is a contradiction, therefore Γ' is consistent. ■

In other words, Lindenbaum's lemma tells us that if $\Gamma \nvdash \varphi$, then we can extend the consistent set $\Gamma \cup \{\varphi\}$ to a maximal consistent set Γ'. It follows that $\Gamma' \nvdash \varphi$ as well. The canonical model that we will define in the lemma will have the property that the world Γ' does not make φ valid either, that is, $\Gamma' \nvDash \varphi$ either.

Definition 7.9 — *Canonical model*

The **canonical model** of a normal logic Λ is the model $\mathcal{M}_\Lambda = \langle W_\Lambda, R_\Lambda, V_\Lambda \rangle$ with:

- $W_\Lambda = \{\Gamma \subseteq \text{Fml}_{\text{ML}} \mid \Gamma \text{ is an MCS}\}$.
- $\langle \Gamma, \Delta \rangle \in R_\Lambda$ if and only if $\varphi \in \Delta$ for every formula φ such that $\Box\varphi \in \Gamma$.
- For all $p \in \mathsf{At}$ we have $p \in V_\Lambda(\Gamma)$ if and only if $p \in \Gamma$. ◁

As discussed before, the worlds of the canonical model are maximal consistent sets. The last piece of the puzzle is to show that the worlds in the canonical model make exactly their own formulas valid.

Since we want each world to make all of its own formulas true, we certainly need it to make all propositional atoms true that it contains, and all other propositional atoms false. This is reflected in how the valuation is defined, since the valuation of the world contains exactly those propositional atoms that are contained in the world.

Finally we need to make sure that also the modal formulas contained in the MCS are valid in the world. This is ensured by how the accessibiliy relation is defined, as a world Δ is accessible from Γ if and only if Δ makes all formulas φ true for which $\Box\varphi$ is true in Γ.

Lemma 7.10 — *Truth lemma*

$\varphi \in \Gamma$ if and only if $\mathcal{M}_\Lambda, \Gamma \vDash \varphi$. ◁

Proof. The proof is by induction on the complexity of the formula φ.

If $\varphi = p$ is an atom, then $p \in \Gamma$ if and only if $p \in V_\Lambda(\Gamma)$ if and only if $\mathcal{M}_\Lambda, \Gamma \vDash p$.

Assume for the induction hypothesis that φ and ψ are formulas such that $\varphi \in \Gamma'$ if and only if $\mathcal{M}_\Lambda, \Gamma' \vDash \varphi$ and $\psi \in \Gamma'$ if and only if $\mathcal{M}_\Lambda, \Gamma' \vDash \psi$ for every MCS Γ'.

Let Γ be an MCS and $\varphi \wedge \psi \in \Gamma$, then $\varphi \in \Gamma$ and $\psi \in \Gamma$ by Lemma 7.7, thus by induction hypothesis $\mathcal{M}_\Lambda, \Gamma \vDash \varphi$ and $\mathcal{M}_\Lambda, \Gamma \vDash \psi$. This in turn implies $\mathcal{M}_\Lambda, \Gamma \vDash \varphi \wedge \psi$. On the other hand, if $\mathcal{M}_\Lambda, \Gamma \vDash \varphi \wedge \psi$, then $\mathcal{M}_\Lambda, \Gamma \vDash \varphi$ and $\mathcal{M}_\Lambda, \Gamma \vDash \psi$, therefore by induction hypothesis $\varphi \in \Gamma$ and $\psi \in \Gamma$. Once again by Lemma 7.7 we see that $\varphi \wedge \psi \in \Gamma$.

The cases for $\neg\varphi$, $\varphi \vee \psi$, $\varphi \to \psi$ and $\varphi \leftrightarrow \psi$ are exactly as the above case.

Finally, let Γ be an MCS and $\Box\varphi \in \Gamma$, then by definition of the relation R_Λ we see that every MCS Δ such that $\Gamma \; R_\Lambda \; \Delta$ contains φ, and thus by induction hypothesis $\mathcal{M}_\Lambda, \Delta \vDash \varphi$ for all such Δ. It follows that $\mathcal{M}_\Lambda, \Gamma \vDash \Box\varphi$. On the other hand, suppose that $\mathcal{M}_\Lambda, \Gamma \not\models \Box\varphi$, then there is some Δ such that $\Gamma\ R_\Lambda\ \Delta$ and $\mathcal{M}_\Lambda, \Delta \not\models \varphi$. By induction hypothesis we then see that $\varphi \notin \Delta$, and thus by how R_Λ is defined it follows that $\Box\varphi \notin \Gamma$.

The case for $\Diamond\varphi$ is similar the the previous case. ■

We will now finally prove the completeness theorem for Hilbert system K.

Theorem 7.11 — *Strong completeness theorem*
The logic K is strongly complete with respect to the class of all Kripke frames. In other words, $\Gamma \models \varphi$ implies $\Gamma \vdash_{\mathsf{K}} \varphi$. ◁

Proof. We prove this by contraposition. Suppose that $\Gamma \nvdash_{\mathsf{K}} \varphi$, then also $\Gamma \nvdash_{\mathsf{K}} \neg\neg\varphi$, so $\Gamma \cup \{\neg\varphi\}$ is a consistent set. From Lindenbaum's lemma we know that there is an MCS $\Delta \supseteq \Gamma \cup \{\neg\varphi\}$. In the canonical model $\mathcal{M}_{\mathsf{K}}$ we then see that $\mathcal{M}_{\mathsf{K}}, \Delta \models \Gamma$ and $\mathcal{M}_{\mathsf{K}}, \Delta \models \neg\varphi$ using the Truth lemma. Therefore the canonical model provides a countermodel to show that $\Gamma \not\models \varphi$. ■

We can also prove the completeness theorem for the other systems defined in Definition 5.17. The proofs have the same structure, but we have to restrict our canonical model to the corresponding frame class. Suppose that Λ is a proof system corresponding with a certain frame class $\mathcal{C}_\Lambda$. For completeness we wish to show that if $\Gamma \nvdash_\Lambda \varphi$, then $\Gamma \not\models_{\mathcal{C}_\Lambda} \varphi$. Therefore, it is not enough to find just any counter model where Γ holds in a world, but φ does not: we need to make sure that the counter model is based on a frame in $\mathcal{C}_\Lambda$.

The missing step is therefore to show that the canonical model $\mathcal{M}_\Lambda$ is based on a frame in $\mathcal{C}_\Lambda$. Before we will state the completeness theorem for the other normal logics, we will prove a lemma that will come in handy in the proof of the theorem. The lemma gives us an alternative method to define the accessibility relation of the canonical model.

Lemma 7.12
Let $\mathcal{M}_\Lambda$ be the canonical model for the logic Λ, and let Γ and Δ be MCS's of Λ. Then $\Gamma\ R_\Lambda\ \Delta$ if and only if $\Diamond\varphi \in \Gamma$ for every formula φ such that $\varphi \in \Delta$. ◁

Proof. ($\Rightarrow$) Suppose that $\Gamma\ R_\Lambda\ \Delta$ and that $\varphi \in \Delta$. Since Γ is an MCS, either $\Diamond\varphi \in \Gamma$ or $\neg\Diamond\varphi \in \Gamma$ by maximal consistent sets being complete.

Suppose that $\neg\Diamond\varphi \in \Gamma$. Because $\neg\Diamond\varphi \vdash_\Lambda \Box\neg\varphi$ for normal logics Λ, we see by Lemma 7.7 that $\Box\neg\varphi \in \Gamma$. But this implies that $\neg\varphi \in \Delta$, since we had assumed $\Gamma\ R_\Lambda\ \Delta$. This is a contradiction, because if φ and $\neg\varphi$ are both elements of Δ, then Δ is not consistent. Therefore $\neg\Diamond\varphi \notin \Gamma$, and thus $\Diamond\varphi \in \Gamma$.

($\Leftarrow$) We leave the other direction as Exercise 7.5. ■

We are now ready to prove the completeness theorem for the other systems from Definition 5.17. We will only give a proof for the completeness of the logics KT and K5, and leave the other proofs as exercises.

Theorem 7.13 — *Completeness theorem*

- The logic KT is strongly complete with respect to the class of reflexive frames.
- The logic KB is strongly complete with respect to the class of symmetric frames.
- The logic KD is strongly complete with respect to the class of serial frames.
- The logic K4 is strongly complete with respect to the class of transitive frames.
- The logic K5 is strongly complete with respect to the class of Euclidean frames. ◁

Proof. For KT, as was mentioned before, it is enough to show that the presence of Axiom **T** makes the canonical model reflexive.

Suppose that Γ is an MCS in the logic KT, and $\Box\varphi \in \Gamma$ for some formula φ. We know that Γ is deductively closed by Lemma 7.7, therefore if ψ is a formula such that $\Box\varphi \vdash_{\mathsf{KT}} \psi$, then $\psi \in \Gamma$ as well. In particular, we have $\Box\varphi \vdash_{\mathsf{KT}} \varphi$ by the following proof:

1.	$\Box\varphi$	Assumption
2.	$\Box\varphi \to \varphi$	Ax. **T**
3.	φ	MP $1, 2$

Therefore $\varphi \in \Gamma$ for every formula φ such that $\Box\varphi \in \Gamma$. Looking at Definition 7.9, we see that the accessibility relation for the canonical model contains $\langle \Gamma, \Gamma \rangle$. In other words, any MCS of KT reaches itself in the canonical model, therefore the canonical model is reflexive.

For K5, we have to show that the canonical model is Euclidean. Suppose that Γ, Δ and Θ are MCS's for the logic K5 and that $\Gamma \; R_{\mathsf{K5}} \; \Delta$ and $\Gamma \; R_{\mathsf{K5}} \; \Theta$. We have to prove that $\Delta \; R_{\mathsf{K5}} \; \Theta$ as well to show that the canonical model is Euclidean. Suppose that $\Box\varphi \in \Delta$, then by Lemma 7.12 we know that $\Diamond\Box\varphi \in \Gamma$. The following proof shows that $\Diamond\Box\varphi \vdash_{\mathsf{K5}} \Box\varphi$:

1.	$\Diamond\Box\varphi$	Assumption
2.	$\Diamond\neg\varphi \to \Box\Diamond\neg\varphi$	Ax. **5**
3.	$(\Diamond\neg\varphi \to \Box\Diamond\neg\varphi) \to (\neg\Box\Diamond\neg\varphi \to \neg\Diamond\neg\varphi)$	Taut. $(A \to B) \to (\neg B \to \neg A)$
4.	$\neg\Box\Diamond\neg\varphi \to \neg\Diamond\neg\varphi$	MP $2, 3$
5.	$\Diamond\neg\Diamond\neg\varphi \to \neg\Diamond\neg\varphi$	Sub. $\neg\Box A \equiv \Diamond\neg A$
5.	$\Diamond\Box\varphi \to \Box\varphi$	Sub. $\neg\Diamond\neg A \equiv \Box A$
6.	$\Box\varphi$	MP $1, 5$

Therefore, by Lemma 7.7 we see that $\Box\varphi \in \Gamma$. Since $\Gamma \; R_{\mathsf{K5}} \; \Theta$, we see that $\varphi \in \Theta$. This shows that $\varphi \in \Theta$ for any φ such that $\Box\varphi \in \Delta$, and thus $\Delta \; R_{\mathsf{K5}} \; \Theta$, as was needed to prove. ■

Combining the completeness proofs for KT and K5 gives us a completeness proof for KT5 = S5 with respect to the class of equivalence frames. Similarly we can prove completeness for S4:

Corollary 7.14

- The logic S4 is strongly complete with respect to the class of preordered frames.
- The logic S5 is strongly complete with respect to the class of equivalence frames. ◁

You might be tempted to think that any normal logic is strongly complete with respect to their corresponding frame class. This is not the case, as there are normal logics that are only weakly complete, and even normal logics that are incomplete with respect to any frame class. We will investigate one of these cases in Exercise 7.7. We can however state the following theorem, although its proof is beyond the scope of this course.

Theorem 7.15

If Λ is a normal logic that corresponds to a first-order definable frame class $\mathcal{C}$, then Λ is strongly complete with respect to $\mathcal{C}$. ◁

7.1 EXERCISES

Exercise 7.1. Let Γ and Δ be MCSs of a normal logic. Prove that if $\Gamma \neq \Delta$, then there is a formula $\varphi \in \Gamma$ such that $\neg\varphi \in \Delta$.

Exercise 7.2. Determine for each of the following Γ if it is consistent in the given logic. If it is inconsistent, give a proof that $\Gamma \vdash_\Lambda \neg\varphi$ for some $\varphi \in \Gamma$. If it is consistent, give a model for Γ based on a frame in the appropriate frame class.

a) $\Gamma = \{\Box\Diamond\varphi, \Diamond\Box\psi, \neg\Box\Box(\varphi \vee \psi)\}$ in the logic K.
b) $\Gamma = \{\Diamond\varphi, \Diamond\Box\neg\varphi\}$ in the logic K.
c) $\Gamma = \{\Diamond\varphi, \Diamond\Box\neg\varphi\}$ in the logic $\mathsf{K5}$.
d) $\Gamma = \{\Box(\Box\varphi \to \varphi), \neg\varphi\}$ in the logic KT.
e) $\Gamma = \{\Diamond\Box\varphi, \Box\Diamond\Diamond\Box\psi, \Box(\psi \to \neg\varphi)\}$ in the logic $\mathsf{S5}$.

* **Exercise 7.3.** A logic Λ has the **disjunctive property** when for any formulas φ and ψ we have that $\vdash_\Lambda \Box\varphi \vee \Box\psi$ implies $\vdash_\Lambda \varphi$ or $\vdash_\Lambda \psi$.

a) Prove that $\mathsf{S4}$ has the disjunctive property.
(Hint: use completeness)
b) Prove that $\mathsf{S5}$ does not have the disjunctive property.

* **Exercise 7.4.** Remember that a set X is countable if and only if there is an injective function $f : X \to \mathbb{N}$ that sends each element of X to a unique natural number.

a) Prove that there is an injection from $\mathbb{N} \times \mathbb{N}$ to $\mathbb{N}$.
(Hint: the number of $(a, b) \in \mathbb{N} \times \mathbb{N}$ such that $a + b = k$ for any $k \in \mathbb{N}$ is finite)
b) Prove that if $\mathcal{L}$ is a countable alphabet, then the set $\mathcal{L}^n$ of a strings of length n with $n \in \mathbb{N}$ is countable.
(Hint: use induction on n)
c) Use the previous two parts to show that the set $\mathcal{L}^{<\omega}$ of all strings over $\mathcal{L}$ of finite length is countable.

Exercise 7.5. Finish the proof of Lemma 7.12. That is, prove that if $\Diamond\varphi \in \Gamma$ for every $\varphi \in \Delta$, then $\Gamma \; R_\Lambda \; \Delta$.

Exercise 7.6. Let a logic Λ be strongly complete with respect to a frame class $\mathcal{C}$. Prove the **compactness theorem** for Λ:

> If Γ is a set of formulas and for every finite subset $\Delta \subseteq \Gamma$ there is a model $\mathcal{M}_\Delta$ based on a frame in $\mathcal{C}$ and world w of $\mathcal{M}_\Delta$ such that $\mathcal{M}_\Delta, w \vDash_\mathcal{C} \Delta$, then there is a model $\mathcal{N}$ based on a frame in $\mathcal{C}$ and world v of $\mathcal{N}$ such that $\mathcal{N}, v \vDash_\mathcal{C} \Gamma$.

(Hint: prove by contraposition, use soundness, completeness and the fact that proofs are finite)

* **Exercise 7.7.** The logic GL is the extension of K with the Gödel-Löb axiom $\Box(\Box\varphi \to \varphi) \to \Box\varphi$. We saw in Theorem 2.10 that GL corresponds to the class of transitive conversely well-founded frames. In this exercise we will see that GL is not strongly complete with respect to this frame class, nor with any other frame class.

Consider the set $\Gamma = \{\Diamond\varphi_0\} \cup \{\Box(\varphi_i \to \Diamond\varphi_{i+1}) \mid i \in \mathbb{N}\}$.

a) Prove that Γ is consistent in GL.
(Hint: if Γ were inconsistent, let $\psi \in \Gamma$ such that $\Gamma \vdash_{\mathsf{GL}} \neg\psi$. Use that a proof is finite, thus there would be a finite $\Delta \subseteq \Gamma$ such that $\Delta \vdash_{\mathsf{GL}} \neg\psi$ and use soundness to show that this is a contradiction)
b) Prove that a model where Γ is true is not conversely well-founded.
c) Prove that GL is not strongly complete with respect to any frame class.

PART III

EPISTEMIC LOGIC

COMING SOON...

APPENDIX

HOW TO LEARN MATHEMATICS

A.1 HOW TO READ MATHEMATICAL TEXT

Anyone who has seen a mathematical text before has probably noticed that it is not structured like a novel. Apart from mathematical jargon, there are plenty of symbols, and the main text is usually separated by blocks of definitions, lemmas and proofs. It should therefore not come as a surprise that reading a mathematical text requires a very different approach than reading a novel. There are some common pitfalls that the uninitiated will stumble into without a proper strategy on how to tackle the mathematical text. This section serves as a small guide to the strategy that should be used, and hopefully it will help you to avoid these pitfalls yourself in the future.

First of all let's look at the structure of a mathematical text. Often the goal of the text will be to present an important result, called a **theorem**. The theorem states an important property of the thing that is being studied. In order to truly understand a theorem, there are two essential requirements: an intuitive grasp of what the theorem means, and complete certainty about the truth of the theorem. Usually knowing why a theorem is true will also help in building intuition for what the theorem says. To reach this goal of true understanding, you need to first remove any ambiguity about the meaning of the theorem and the relevant terms that are used in a theorem, which is done through precise and unambiguous **definitions**. Next you need to prove the theorem to make absolutely sure that the theorem is correct. Unless the theorem is elementary, proving the theorem is aided by proving intermediate results, called **lemmas**. Some results are not used as an intermediate result, but also not significant enough to be called a *theorem*, and therefore they are often called **propositions**. Propositions that easily follow from a lemma or theorem that has just been proved, are called **corollaries**, and are quite often stated without a proof. Claims which have strong evidence to be true, but have yet to be proved or disproved, are called **conjectures**.

Note that theorems rely on the validity of previous results. But what do these previous results rely on? We need some basic building blocks for proofs to be able to start things off. These basic building blocks are propositions without a proof that are *assumed to be true*. Such propositions are called **axioms**. For an example: most mathematics is based on the rules of first-order logic and the Zermelo-Fraenkel axioms. Another example of axioms are the postulates of Euclides used in classical Greek geometry, such as is taught in a typical high school.

Building towards a proof of a theorem is done *from the ground up*. If you want to make sure a theorem is correct, it is essential that you are convinced of the truth of each of the intermediate steps that lead to the proof: without it, the whole proof will collapse. For this reason *you should not rush through a mathematical text.* Take care to make sure you understand each sentence as it is supposed to be understood. This takes time and effort, but it is the only way to understand what is written. Although reading a novel can be done leisurely, reading mathematics is very much a *labourious* activity. Make sure you ask yourself with every statement *what* the statement means and *why* the statement is true. If you don't understand what the meaning of a statement is, try to think of examples and counterexamples, and work them out. Work along with the steps, work out steps that are not written down explicitly if necessary, and investigate any statement that you are uncertain about.

Mathematical texts are notoriously compact. You will notice that almost any word in a mathematical statement has a precise and specific mathematical meaning. Sometimes words that are synonymous in English have distinct meanings in a mathematical context. For example *'arbitrary'* (does not depend on what you choose) and *'random'* (chosen by a certain stochastic process) have a distinct meaning in mathematics. Make sure *you understand the meaning of all the terms in a mathematical text.*

Some specific expressions are often used, sometimes in a slightly different way than in natural language. You will get used to these expressions through exposure, but I will highlight a few of the most common ones:

- **Let − be a −**: used to define objects.
  For example: *"Let x be a positive real number"*.
- **Such that** (s.t.): used to define a property a definition has to satisfy.
  For example: *"Let x be a real number s.t. $x^2 + x$ is even"*.
- **If and only if** (iff): used as a bi-implication.
  For example: *"x is a rational number iff there exists integers p and q such that $x = \frac{p}{q}$"*.
- **The following are equivalent** (t.f.a.e.): used to define a list of equivalent properties.
  For example: *"The following are equivalent:*
  1. x is a rational number,
  2. there are integers p and q s.t. $x = \frac{p}{q}$,
  3. the decimal expansion of x is repeating."
- **Obviously / Clearly / It follows easily that**: used to leave out parts of the explanation that do not require new ideas to solve, such as computations that are not particularly insightful or things that have been explained in the past. Note that it does not actually have to be easy or obvious at first glance, but that the writer assumes the reader can work out the details without too much creativity.
  For example: *"It follows easily that $x^5 + 2x^4 - 21x^3 - 2x^2 + 116x - 120$ has three distinct roots, all of which are integers"*. This doesn't mean that finding the roots is easy, but it does mean that it can be done with tools that are considered understood.
- **Without loss of generality** (w.l.o.g.): used when there are several cases to consider, but the proof of each of these cases is so similar that it is sufficient to prove only a single case.
  For example: *"We will consider the case where a and b are distinct real numbers. W.l.o.g. let $a < b$."* The proof that will follow can easily be transformed to prove the other case that $b < a$, for example by renaming a and b to each other.

To summarise this section, keep the following bullet points in the back of your head while reading mathematics:

- Build from the ground up (Axioms + Definitions → Lemmas → Theorem),
- Understand each intermediate step,
- Actively work along and investigate any uncertainties,
- Read statements word by word and take note of the precise meaning of all the terms.

A.2 HOW TO WRITE PROOFS OF MATHEMATICAL STATEMENTS

Now that we have discussed how you should read a mathematical text, the next step is naturally how to write mathematical text yourself. When you start proving things it will be difficult to understand how one should do it. It is important to understand the two goals of writing proofs: a proof serves to

1. **eliminate any doubt** on the validity of a statement, and
2. **communicate** your reasoning to others.

Note that the purpose of communication makes it necessary that your proof is clearly understandable for other people. Of course this means that you use *unambiguous* language with *proper sentences* and that you *introduce new terminology* properly before using it (or make sure that the terminology that you use is understood by your audience). It should be made clear that **a computation without explanation does not constitute a proof**.

In general the use of first and second person pronouns (*'I'*, *'we'* and *'you'*) are discouraged in scientific writing. However, you will be surprised to discover that in mathematics it is standard practice to write proofs in plural first person perspective (that is, the *'we'* perspective). This highlights the fact that reading mathematics requires *active participation* and that proofs are ideally more like a *conversation* communicating an idea than like a statement of facts.

If you use symbols in your proof, make sure you use them *consistently* and *introduce* them properly. An average mathematical text will use many symbols, usually a lot more than the 26 letters of the English alphabet. For this reason symbols are often distinguished by their case and their typographic style, and additionally sometimes symbols are borrowed from other alphabets, such as Greek or Hebrew. As an example, it is not unlikely that one would encounter a sentence such as the following in a text:

> *Consider the real number* $r \in \mathbb{R}$*, the set of relations* $\mathcal{R} \subseteq \mathcal{P}(\mathbb{R} \times \mathbb{R})$ *and suppose that* $R \in \mathcal{R}$ *is one of these relations such that* $r \; R \; r$*.*

In fact most of these renditions of the letter **r** can be considered relatively common: real numbers are usually represented by lower case letters (often $a, b, c, d, p, q, r, s, t, x, y$ or z), relations by uppercase letters (often R, S, T or E), families of sets by cursive uppercase letters (such as $\mathcal{R}$, $\mathcal{S}$ or $\mathcal{T}$) and the set of reals is almost always what is meant when an uppercase blackboard boldface $\mathbb{R}$ is encountered.

Unfortunately not every author uses the same convention regarding these symbols. It is therefore pertinent to clarify which conventions you adhere to and clearly explain the purpose of each symbol at the beginning of a text. For this reason the introduction of this reader summarises some of the mathematical and logical notation that will be used in this reader, as well as repeat some of the prerequired knowledge.

Of course the hardest part of writing a proof is usually not explicitly writing it down, but thinking up the proof itself in the first place. Luckily there are a few heuristics and methods that could be employed to aid the search of a proof. Some of these can be found in the next and last section.

A.3 PROOF STRATEGY

In this section we will explore several heuristics that could aid you in proving statements. As with everything, by practicing these methods it will become clear when and how they should be used, so if any of these seem unfamiliar or difficult to you, it is advised to do some of the exercises at the end of this chapter.

Proof by contraposition

If we are given a statement *"A implies B"*, we could try proving it directly, by assuming A and showing that B follows. However, in some cases it is easier to prove that if B were false, then A is also false. Since logically $A \to B$ and $\neg B \to \neg A$ are equivalent, the method of direct proof and proof by contraposition are logically equivalent.

Example A.1
If the square x^2 of an integer x is even, then x is also even. $\triangleleft$

The statement we want to prove is the implication *"x^2 is even $\Rightarrow$ x is even"*, then we can use that an integer is not even if and only if it is odd, and formulate the contrapositive of this statement as *"x is odd $\Rightarrow$ x^2 is odd"*. We will give two proofs, one by contraposition, and one by direct proof. You could judge yourself which proof is the easier one.

> *Proof by contraposition.* If x is odd, then we can write it as $x = 2k + 1$ for some integer k. Then $x^2 = (2k + 1)^2 = 4k^2 + 4k + 1$. Now $4k^2 + 4k$ is divisible by 4, so it is even. But then x^2 is one more than an even number, and hence x^2 is odd. ■
>
> *Proof by direct proof.* Let x^2 be even, then it can be written as $x^2 = 2k$ for some integer k. We then see that $x = \pm\sqrt{2k} = \pm\sqrt{2}\sqrt{k}$. We know that x is an integer, and $\sqrt{2}$ is not, hence $\sqrt{k}$ should be equal to some integer n divided by $\sqrt{2}$, to get rid of the non-integer factor $\sqrt{2}$. But if $\sqrt{k} = \frac{n}{\sqrt{2}}$, then $n = \frac{\sqrt{k}}{\sqrt{2}} = \sqrt{\frac{k}{2}}$. Since n is an integer, and only square numbers have integer roots, we see that $\frac{k}{2}$ is a square number. So let $\frac{k}{2} = m^2$, then $k = 2m^2$. Now we can substitute this back into $x = \pm\sqrt{2k}$ to get $x = \pm\sqrt{2 \cdot 2m^2} = \pm\sqrt{4m^2} = \pm 2m$, so x is even. ■

Proof by contradiction

Related to proof by contraposition, but slightly different, is proof by contradiction. In this case we start with a statement A that we want to prove, but instead of proving A directly, we assume the negation $\neg A$ is true and then try to show this leads to a contradiction. Logically we then have shown that $\neg A \to \bot$, which is equivalent (by contraposition) to $\neg\bot \to A$. But $\neg\bot = \top$ is always true, so this gives us a proof of A.

For the next example, remember that a rational number is a number that can be written as a fraction $\frac{a}{b}$, where a and b are any integers and $b \neq 0$. An irrational number is any number that is not a rational number.

Example A.2
$\sqrt{2}$ is irrational. $\triangleleft$

> *Proof.* Let's assume that $\sqrt{2}$ is rational, then there are integers a and b such that $\sqrt{2} = \frac{a}{b}$. If both a and b are divisible by 2, then we see that $\frac{a}{b} = \frac{a/2}{b/2}$ where $a/2$ and $b/2$ are also integers. By doing this until no factors of 2 remain in the denominator or in the numerator of the fraction, we can be sure that there is a fraction $\frac{a}{b} = \sqrt{2}$ with either a or b an odd integer.
>
> We now square both sides of the equation to get $\frac{a^2}{b^2} = 2$, and thus $a^2 = 2b^2$, or in other words a^2 is even. By the last example we know that if a^2 is even, then a is also even, so therefore we can say $a = 2m$ for some integer m. We then get $(2m)^2 = 4m^2 = 2b^2$. Dividing this by the factor 2, we get that $b^2 = 2m^2$. But now we see that b^2 is also even, so b must be even. However, we had assumed that either a or b is odd! This is a contradiction, so we can conclude that $\sqrt{2}$ is irrational. ■

Direct proof versus indirect proof

Proof by contradiction and proof by contrapositive are examples of indirect proof methods. Note that it is not always possible to give a direct proof of statements that you can prove indirectly. For example, if you want to prove that an object *exists*, then a direct proof would consist of giving an example of the object, whereas an indirect proof might not give you any example, but simply shows that the object must exists regardless of what it is. Often showing that an object exists is a lot easier than giving an explicit example. In the worst cases the latter might not even be possible.

Example A.3
There exist two irrational numbers a and b such that a^b is a rational number. ◁

> *Proof.* By the last example we know that $\sqrt{2}$ is irrational. Consider the number $\sqrt{2}^{\sqrt{2}}$. If this number is rational, then we have found the numbers $a = b = \sqrt{2}$ such that a^b is irrational. On the other hand, if $\sqrt{2}^{\sqrt{2}}$ is irrational, then $(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^{2} = 2$ is certainly rational, and hence $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$ give us our answer. ■

Note that in the last proof we have shown that the numbers a and b exist, but we still don't know what they are: the proof doesn't tell us whether or not $\sqrt{2}^{\sqrt{2}}$ is rational. Although we have proved the statement about the *existence* of two irrational numbers, we have not *explicitly* found two irrational numbers that work.

Proof by induction

When you want to show that a property is true for every object in a large or infinite set of things, this might seem like a very daunting task. Sometimes the property is directly provable from the nature of the object, but sometimes this is impossible or very difficult. Proof by induction gives a way by first proving an easy base case, and then reshaping the proof of this base case recursively into a proof of the other cases.

For example, if we want to prove that a statement is true for all natural numbers, then we could start by proving the property for the number 0, and give a method of transforming a proof for the number n into a proof for the number $n + 1$.

Example A.4
Every integer n is either divisible by 3, or it is 1 more than a number divisible by 3, or it is 2 more than a number divisible by 3. ◁

> *Proof.* We prove this just for natural numbers n. The cases where $n < 0$ could be done similarly and are left as an exercise.
>
> As a base case, we can observe that 0 is divisible by 3 (since $0 = 0 \cdot 3$. This makes 1 exactly 1 more than a multiple of 3, and it makes 2 exactly 2 more than a multiple of 3.
>
> Suppose we already have a proof of the statement for all numbers less than or equal to n and we haven't proved the statement for $n + 1$ yet, then $n + 1 \geq 3$ (as we have the above proofs for $n = 0, 1, 2$), so $n - 2 \geq 0$ is a natural number. By our

assumption we have a proof of the statement for $n - 2$, so either $n - 2 = 3k$ or $n - 2 = 3k + 1$ or $n - 2 = 3k + 2$ for some natural number k. But then $n + 1 = 3k + 3$ or $n + 1 = 3k + 3 + 1$ or $n + 1 = 3k + 3 + 2$. Clearly all of these possibilities satisfy the statement, so the statement also holds for $n + 1$. ■

The method of proof by induction is not only useful for proving statements about natural numbers. In logic it is a particularly strong tool for proving general statements about formulas. Since formulas are in the end combinations of atomic formulas with logical connectives, we could start with proving the statement for just the atomic formulas, and then give a method for transforming proofs for a formula φ into a proof of a sligtly more complex formula ψ. We will encounter this type of proof a few times during this course.

Proof by cases

Sometimes a statement can be split up in several easier to prove statements. This method of proof is called proof by cases. It consists of two main parts: first one shows that a statement A is equivalent to some other statements $A_1 \wedge \ldots \wedge A_n$, and then each of the A_i is proved individually. Once again we see that this is logically a valid way of reasoning.

Example A.5
There is no square number x^2 that is 1 less than a multiple of 3. ◁

Proof. Let x be an integer, then either x is divisible by 3, it is 1 more than a number divisible by 3, or it is 2 more than a number divisible by 3, as we have proved in the last example. We consider each of these three cases.

If x is divisible by 3, then $x = 3n$ for some integer n. We then see that $x^2 = 9n^2$, which is not 1 less than a multiple of 3 (since it is a multiple of 3 itself).

If x is 1 more than a number divisible by 3, then $x = 3n + 1$ for some integer n, so then $x^2 = (3n + 1)^2 = 9n^2 + 6n + 1$, and $9n^2 + 6n$ is divisible by 3, so we see that x^2 is 1 more than a number divisible by 3.

If x is 2 more than a number divisible by 3, then $x = 3n + 2$ for some integer n, so then $x^2 = (3n + 2)^3 = 9n^2 + 12n + 4 = 9n^2 + 12n + 3 + 1$, and $9n^2 + 12n + 3$ is divisible by 3, so we see that x^2 is 1 more than a number divisible by 3. ■

A.4 EXERCISES

Exercise A.1. In this exercise, let x and y be integers. Prove the following statements by contraposition:

a) Prove that if $7x^2 - 3$ is even, then x is odd.
b) Prove that if $3x - 5y^2$ is odd, then only one of x and y is odd.
c) Prove that x^2 is divisible by 5 if and only if x is divisible by 5.
(Hint: use that if x is not divisible by 5, then it is equal to $5n + m$ with n and m integers and $0 < m \leq 4$. You might want to use proof by cases as well.)

Exercise A.2. Prove the following statements by contradiction:

a) Prove that $\sqrt{5}$ is irrational.
b) Prove that if x and y are positive real numbers, then $\sqrt{x + y} \neq \sqrt{x} + \sqrt{y}$.

c) If A, B and C are sets such that $A \cap B = A \cap C$ and $A \cup B = A \cup C$, then $B = C$.

Exercise A.3. Prove the following statements by induction:

a) Prove that $1 + 2 + 3 + \cdots + n = \frac{n^2+n}{2}$.

b) Remember that $n! = 1 \cdot 2 \cdot \cdots \cdot n$. Prove that $n! > 2^n$ for all $n \geq 4$.

c) An L-tromino is the shape that is built out of three unit squares and looks like this: . Show that for any integer $n \geq 1$, if we take a grid of $2^n \times 2^n$ unit squares and colour one of the squares black, we can fill the remaining squares of the board using only L-trominos (and without overlapping or going outside of the grid).

Exercise A.4. Prove the following statements by cases:

a) Let x and y be integers. If $x \cdot y$ is odd, then $x^2 + y^2$ is even.

b) If n is an integer, then n^3 is divisible by 9, or 1 more than a number divisible by 9 or 1 less than a number divisible by 9.

c) There exists no real number x such that $x > x^2$ and $x < x^3$.
(Hint: consider different intervals in which x could be.)

Exercise A.5. Prove the following statements:

a) There exist two irrational numbers a and b such that a^b is rational and $a \neq b$.

b) Prove Exercise A.3.a without using induction.

c) Prove Exercise A.4.c by contradiction.
(Hint: square numbers are nonnegative and the product of two nonnegative numbers is nonnegative.)

d) Prove that $100 \cdot 2^n + 10 \cdot 2^{n+1} + 2^{n+3} = 2^{n+7}$ for all $n \in \mathbb{N}$.

e) A prime number is a natural number x such that $x > 1$ and there is no natural number y such that $1 < y < x$ and x is divisible by y. Prove there is no largest prime number.

f) The Fibonacci numbers are defined as following: the first two Fibonnacci numbers F_1 and F_2 are equal to 1, and for any $n \geq 3$ we define $F_n = F_{n-1} + F_{n-2}$. Prove that F_n is even if and only if n is divisible by 3.

BIBLIOGRAPHY

[1] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic.* Cambridge University Press, 2001.

[2] Patrick Blackburn, Johan van Benthem, and Frank Wolter. *Handbook of Modal Logic.* Elsevier, 2007.

[3] Alexander Chagrov and Michael Zakharyaschev. *Modal Logic.* Oxford University Press, 1997.

[4] Gary Chartrand, Albert D. Polimeni, and Ping Zhang. *Mathematical Proofs: A Transition to Advanced Mathematics.* Pearson, second edition, 2008.

[5] Marcello d'Agostino, Hänle Reiner Gabbay, Dov, and Joachim Posegga. *Handbook of Tableau Methods.* Springer, 1999.

[6] Herbert B. Enderton. *A Mathematical Introduction to Logic.* Harcourt Academic Press, second edition, 2001.

[7] Melvin Fitting. *Proof Methods for Modal and Intuitionistic Logics.* Springer, 1983.

[8] Melvin Fitting and Richard L. Mendelsohn. *First-Order Modal Logic.* Springer, 1998.

[9] Rod Girle. *Modal Logics and Philosophy.* McGill-Queen's University Press, 2000.

[10] Robert Goldblatt. *Logics of Time and Computation.* Center for the Study of Language and Information, second edition, 1992.

[11] Rosja Mastop. Modal logic for artificial intelligence, 2011. Lecture Notes.

[12] Eric Pacuit. *Neighborhood Semantics for Modal Logic.* Springer, 2017.

[13] Graham Priest. *An Introduction to Non-Classical Logic.* Cambridge University Press, second edition, 2008.

[14] Raymond M. Smullyan. *First-Order Logic.* Springer-Verlag, second edition, 1971.

[15] Johan van Benthem. *Modal Logic for Open Minds.* Center for the Study of Language and Information, 2010.

[16] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic.* Springer, 2008.

INDEX