Inofficial Script for the Lecture "Modal Logic"
Summer term 2025

Held by Lutz Schröder

Abstract

Modal logics are an important tool for talking about relational structures and have many applications, for example, in computer science, philosophy, and linguistics. This course is an introduction to "pure" modal logic covering many fundamental topics such as modal expressiveness, completeness, and computational complexity.

Prerequisites: an acquaintance with the basics of propositional and first-order logic is expected. Some knowledge in complexity theory will also be helpful.

The first three blocks of this course are for the most part based on a previous lecture on modal logics by Prof. Carsten Lutz at the University of Bremen; otherwise, the course is largely based on the standard textbook by Blackburn, de Rijke, and Venema (2001).

The early editions of the course have been designed by Daniel Hausmann. The first versions of the script have been typeset by Johannes Schilling and Daniel Hausmann.

Literature

- Sally Popkorn. *First Steps in Modal Logic*. 314 pages, Cambridge University Press, 1994.
- Patrick Blackburn, Maarten de Rijke, Yde Venema. *Modal Logic*. 554 pages, Cambridge University Press, 2001.
- Alexander Chagrov and Michael Zakharyaschev. *Modal Logic*. 605 pages, Oxford University Press, 1997.
- Julian Bradfield and Colin Stirling. Chapter *Modal μ-calculi* in Patrick Blackburn, Johan van Benthem, Frank Wolter, *Handbook of Modal Logic*. 1260 pages, Elsevier, 2006.
- Yde Venema. *Lectures on the modal μ-calculus* (lecture notes). 134 pages, draft, 2012.
(`https://staff.science.uva.nl/y.venema/teaching/ml/mu/mu20121116.pdf`)
- Bart Jacobs. Introduction to Coalgebra (draft v2.0). 367 pages, unpublished, 2012.
(`http://www.cs.ru.nl/B.Jacobs/CLG/JacobsCoalgebraIntro.pdf`)
- Dirk Pattinson. Expressive Logics for Coalgebras via Terminal Sequence Induction. Notre Dame Journal of Formal Logic 45(1): 19-33 (2004)
- Lutz Schröder. A finite model construction for coalgebraic modal logic. J. Log. Algebr. Program. 73(1-2): 97-110 (2007)

Contents

Chapter 1

Semantic Introduction

Some slogans to outline the meaning of modal logics in sciences:

- Language for talking about relational structures (graphs).
- Family of logics "between" propositional logic and first order logic.
- Family of logics with attractive computational properties.
- Success story in computer science, philosophy, mathematics, linguistics,...

1.1 Basic Definitions

Definition 1.1 (Modal language, modal formulae)**.** Let $\mathcal{A}$ be a countably infinite set of propositional letters. A modal language is based on a non-empty set of *modal indices* I. The set $\mathcal{F}(I)$ of *modal formulae over* I is given by the rule

$$\phi := \bot \mid p \mid \neg\phi \mid \phi \wedge \psi \mid \Box_i\phi$$

where $p \in \mathcal{A}$ and $i \in I$.

We use several abbreviations:

- $\top$ for $\neg\bot$
- $\phi \vee \psi$ for $\neg(\neg\phi \wedge \neg\psi)$,
- $\phi \to \psi$ for $\neg\phi \vee \psi$,
- $\phi \leftrightarrow \psi$ for $\phi \to \psi \wedge \psi \to \phi$,
- $\Diamond_i\phi$ for $\neg\Box_i\neg\phi$.

The operators $\Box_i$ and $\Diamond_i$ are jointly referred to as *modalities*. If $|I| = 1$, then $\mathcal{F}(I)$ is called the *uni-modal* language, otherwise a *multi-modal* language. In the case of uni-modal languages, we drop the indices on modalities. To save brackets, we give the unary operators $\Box_i, \Diamond_i, \neg$ the highest priority, followed by $\wedge$, then $\vee$, then $\to, \leftrightarrow$.

Example 1.2. Diamonds and boxes obtain various meanings depending on the application. The original understanding of $\Box$ is as *necessarily* (So $\Box\phi$ says that ϕ is somehow 'forced' to be true, to distinguish from the situation where ϕ just somehow 'happens' to be true.) Many further readings have bee established subsequently:

Logics of knowledge or belief $\Box_x\phi$ is read"Agent x *knows/believes* that ϕ is true.".

- What is the meaning of $\Diamond_x$ in this case?
- What would presumably be the axiomatic difference between knowledge and belief?
- How would you axiomatize positive introspection ("I know what I know") and negative introspection ("I know what I do not know")?

Temporal logic $\Diamond\phi$ is read"*at some* point in the future, ϕ is true." Correspondingly, $\Box\phi$ is read "at *all* points in the future, ϕ is true." (Briefly, $\Diamond\phi$ is read 'eventually ϕ' and $\Box\phi$ is read 'always ϕ'.)

- What is the meaning of $\Diamond\Box p$?
- What is the meaning of $\Box\Diamond p$?
- One can additionally introduce *past* operators $\Diamond^-$, $\Box^-$ read 'at some point in the past' / 'at all points in the past'. Can you think of formulae relating past and future that you expect to be valid?

Description logic (DL) Modalities are indexed by *roles* that represent relations between individuals. For instance, $\Diamond_{\mathsf{hasPart}}\phi$ is read "the present object has some part that satisfies ϕ." Correspondingly, $\Box_{\mathsf{hasPart}}\phi$ is read"*all* parts of the present object satisfy ϕ." A formula such as $\Box_{\mathsf{hasPart}}\Diamond_{\mathsf{hasPart}}\mathsf{Fragile}$ then says that "all parts of the present object contain a fragile part."

Deontic logic $\Box\phi$ is read "ϕ is obligatory'.

- What would be your understanding of $\Diamond\phi$?

Definition 1.3 (Worlds, Frames, Models)**.** A *frame* for a set of modal indices I is a tuple

$$\mathfrak{F} = (W, (R_i)_{i\in I})$$

where W is a non-empty set of *worlds* or *states*, and $R_i \subseteq W \times W$ for each $i \in I$.
A *model* for I is a pair

$$\mathfrak{M} = (\mathfrak{F}, V)$$

where $\mathfrak{F}$ is a frame for I and V is a *valuation* assigning a set $V(p) \subseteq W$ to each $p \in \mathcal{A}$.

If $\mathfrak{M} = ((W, (R_i)_{i\in I}), V)$, then we say that $\mathfrak{M}$ is *based* on the frame $(W, (R_i)_{i\in I})$, and usually write $\mathfrak{M} = (W, (R_i)_{i\in I}, V)$. For historical reasons, models are also called *Kripke structures* or Kripke models; similarly, frames are also called *Kripke frames*.

Example 1.4. As examples of uni-modal frames, consider the following:

1. A full binary tree of depth 42.
2. The natural numbers with the standard successor relation.

3. The rational numbers with standard strict order"$<$".

Definition 1.5 (Satisfaction of Formulae)**.** Let $\mathfrak{M} = (W, (R_i)_{i\in I}, V)$ and $w \in W$. We inductively define *satisfaction of formulae ϕ in $\mathfrak{M}$ at w* as follows:

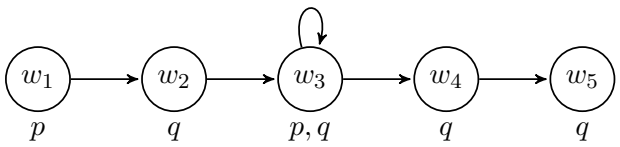
$$\begin{aligned}
&\mathfrak{M}, w \not\models \bot \\
&\mathfrak{M}, w \models p \text{ iff } w \in V(p) \\
&\mathfrak{M}, w \models \neg\phi \text{ iff } \mathfrak{M}, w \not\models \phi \\
&\mathfrak{M}, w \models \phi \wedge \psi \text{ iff } \mathfrak{M}, w \models \phi \text{ and } \mathfrak{M}, w \models \psi \\
&\mathfrak{M}, w \models \Box_i\phi \text{ iff for all } w' \in W \text{ s.t. } (w, w') \in R_i, \text{ we have } \mathfrak{M}, w' \models \phi.
\end{aligned}$$

If $\mathfrak{M}, w \models \phi$, then we say that ϕ is *satisfied* or *true* in $\mathfrak{M}$ at w, and that $\mathfrak{M}, w$ is a *model of* ϕ. If $\mathfrak{M}, w \models \phi$ for all $w \in W$, then we write $\mathfrak{M} \models \phi$ and say that ϕ is *globally true* in $\mathfrak{M}$.

We generalize these notions to (possibly infinite) *sets* Φ of formulae in the obvious way, reading sets as big conjunctions. For instance, $\mathfrak{M}, w \models \Phi$ if $\mathfrak{M}, w \models \phi$ for all $\phi \in \Phi$.

It is easy to derive semantics for the defined symbols $\vee, \rightarrow, \Diamond$, etc. In particular $\mathfrak{M}, w \models \Diamond_i\phi$ iff there is $w' \in W$ such that $(w, w') \in R_i$ and $\mathfrak{M}, w' \models \phi$.

Example 1.6. Consider the uni-modal frame $\mathfrak{F} = (\{w_1, w_2, w_3, w_4, w_5\}, R)$ consisting of five worlds and the relation R that is defined by putting $(w_i, w_j) \in R$ iff $j = i + 1$ or $i = j = 3$. Also consider the valuation $V(p) = \{w_1, w_3\}$, $V(q) = \{w_1, w_2, w_3, w_4, w_5\}$, $V(r) = \emptyset$. The model $(\mathfrak{M}, R, V)$ can be visualized as follows:



We observe that the following statements regarding the satisfaction of formulae are true:

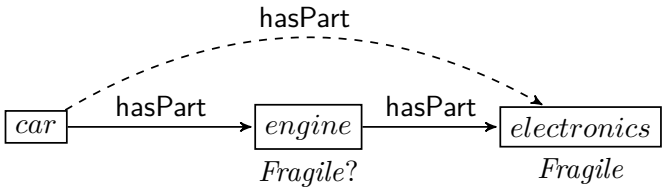
- $\mathfrak{M}, w_1 \models \Diamond\Box p$,
- $\mathfrak{M}, w_2 \not\models \Diamond\Box p$,
- $\mathfrak{M}, w_1 \not\models \Diamond(\Box p \rightarrow p)$ (notice the higher binding of modal operators),
- $\mathfrak{M}, w_2 \models \Diamond(\Box p \rightarrow p)$,
- $\mathfrak{M}, w_4 \models \Diamond\Box\bot$,
- $\mathfrak{M} \models q \rightarrow \Box q$.

When dealing with modal formulae, we shall often restrict ourselves to certain *classes* of frames and models.

Example 1.7. We consider some examples of classes of frames that are suitable for particular logics:

1. In description logics, the relation for the $\Diamond_{\mathsf{hasPart}}$ modality should be transitive:

   Consider the formula $\Diamond_{\mathsf{hasPart}}\Diamond_{\mathsf{hasPart}}\mathit{Fragile}$ and the following model:

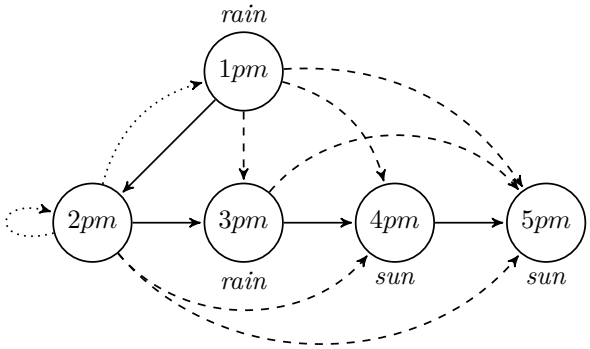


The diagram indicates that an engine is a part of a car and the electronics are a fragile part of the engine. Thus an engine contains a fragile part while it is not necessarily fragile itself; however, we should obtain that a car also contains a fragile part. In other words, the relation R_{hasPart} should be transitive. Formally, we require in this example that,

$$\Diamond_{\mathsf{hasPart}}\Diamond_{\mathsf{hasPart}}\mathit{Fragile} \to \Diamond_{\mathsf{hasPart}}\mathit{Fragile}.$$

Note. For description logics, worlds symbolize objects in the real world and formulae denote properties of such objects.

2. In temporal logic, we want at least irreflexivity, antisymmetry and transitivity, as can be seen by the following example. Assume that we have five points in time, named 1pm up to 5pm, where $i + 1$pm is in the future of $i$pm as indicated by the solid arrows in the following diagram:



Notice how we have $V(\mathit{rain}) = \{1\text{pm}, 3\text{pm}\}$ and $V(\mathit{sun}) = \{4\text{pm}, 5\text{pm}\}$. Now consider the following formulae:

$$\psi_1 = \mathit{rain} \to \Diamond \mathit{sun} \qquad \psi_3 = \mathit{rain} \wedge \Diamond\Box \mathit{sun}$$
$$\psi_2 = \mathit{rain} \to \Box \mathit{sun} \qquad \psi_4 = \mathit{rain} \wedge \Diamond \mathit{rain}$$

As $\Diamond\phi$ has the intuition that "at some point in the future, ϕ holds", ψ_1 should be satisfied at 1pm, as it rains at 1pm and the sun shines at 4pm. To obtain that indeed $1\text{pm} \models \mathit{rain} \to \Diamond \mathit{sun}$, we have to take the transitive closure of the solid arrows in the above diagram, that is, we have to add the dashed transitions, so that e.g. 4pm is defined to be in the future of 1pm. On the other hand, we do not normally want symmetry (e.g. the dotted transitions should not be added to the model) as this would be at odds with the directedness of time, i.e. the distinction between past and future.

For instance, we may choose the set of all linear orders as our class of frames for temporal logics. Other classes are dense linear orders, right-unbounded linear orders, etc.

Note. In temporal logics, worlds represent points in time and propositional letters denote time-dependent propositions.

3. Consider the following formula in the logic of knowledge and belief:

$$\Diamond_x p \wedge \Box_x \Diamond_y \neg p,$$

with the intuitive reading that "agent x cannot exclude p and knows that agent y cannot exclude $\neg p$". In this setting, worlds are possible states of affairs in the real world, and the transition relation for agent x represents epistemic alternative,

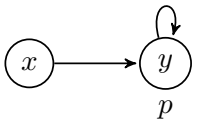
Definition 1.8 (Satisfiability and validity in frames)**.** A frame $\mathfrak{F}$ *satisfies* a formula ϕ (notation: $\mathfrak{F} \models \phi$) if $\mathfrak{M} \models \phi$ for every model $\mathfrak{M}$ based on $\mathfrak{F}$. Then, ϕ is *valid* over a class S of frames if $\mathfrak{F} \models \phi$ for every $\mathfrak{F} \in S$. Dually, ϕ is *satisfiable* over S if there exist a model $\mathfrak{M}$ based on a frame $\mathfrak{F} \in S$ and a world w of $\mathfrak{M}$ such that $\mathfrak{M}, w \models \phi$. More generally, a set Φ of formulae is *satisfiable over* S if there is a model $\mathfrak{M}$ based on a frame $\mathfrak{F} \in S$ and a world w of $\mathfrak{M}$ such that $\mathfrak{M}, w \models \Phi$.

Example 1.9. For instance,

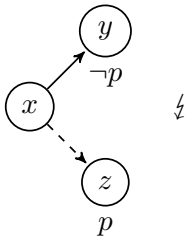
- The formula $\Diamond\Box p$ is satisfiable in the class of all frames; for instance we have

$$\mathfrak{M}, x \models \Diamond\Box p,$$

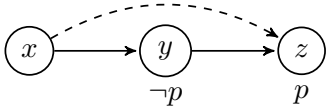
where $\mathfrak{M}$ is given by the following diagram:



- On the other hand, $\Diamond p \wedge \Box\neg p$ is not satisfiable in the class of all frames: If $\Diamond p$ is satisfied at some world x, then x has a successor world at which p is satisfied, and this world is a counterexample to x satisfying $\Box\neg p$.



- The formula $\Diamond\Diamond p \wedge \Box\neg p$ is satisfiable in the class of all frames, but not in the class of transitive frames: The solid arrows in the diagram



constitute a non-transitive model for the formula; on the other hand, by adding the dashed transitive arrow, we obtain a transitive model, in which however the formula is no longer satisfied at x.

1.2 Relation to First-Order-Logic (FOL)

Assume $I = \mathbb{N}$ and $\mathcal{A} = \{p_1, p_2, p_3, \dots\}$. Obviously, each model $\mathfrak{M} = (W, (R_i)_{i\in I}, V)$ can be viewed as a FOL-structure:

$$(W, \underbrace{R_1, R_2, R_3, \dots}_{R_i \subseteq W\times W}, \underbrace{\overbrace{V(p_1)}^{=P_1}, \overbrace{V(p_2)}^{=P_2}, \overbrace{V(p_3)}^{=P_3}, \dots}_{V(p_i)\subseteq W}),$$

where the R_i are binary predicates and the $V(p_i)$ are unary predicates.

Definition 1.10 (Standard translation)**.** Let x be a first-order variable. The *standard translation* ST_x takes modal formulae to FOL-formulae and is defined inductively:

- $ST_x(p_i) = P_i(x)$
- $ST_x(\neg\phi) = \neg ST_x(\phi)$
- $ST_x(\phi \wedge \psi) = ST_x(\phi) \wedge ST_x(\psi)$
- $ST_x(\Box_i\phi) = \forall y.\, R_i(x, y) \to ST_y(\phi)$, where y is distinct from x.

We can derive that

- $ST_x(\phi \vee \psi) = ST_x(\phi) \vee ST_x(\psi)$
- $ST_x(\Diamond_i\phi) \equiv \exists y.\, R_i(x, y) \wedge ST_y(\phi)$.

(and most of the time we assume that the translation of $\Diamond_i$ is defined in the above equivalent way).

Example 1.11. The formula $\Diamond(\Box p \to q)$ translates to FOL as follows:

$$\begin{aligned} ST_x(\Diamond(\Box p \to q)) &= \exists y_1.\, R(x, y_1) \wedge ST_{y_1}(\Box p \to q) \\ &= \exists y_1.\, R(x, y_1) \wedge (ST_{y_1}(\Box p) \to ST_{y_1}(q)) \\ &= \exists y_1.\, R(x, y_1) \wedge ((\forall y_2.\, R(y_1, y_2) \to ST_{y_2}(p)) \to Q(y_1)) \\ &= \exists y_1.\, R(x, y_1) \wedge ((\forall y_2.\, R(y_1, y_2) \to P(p_2)) \to Q(y_1)) \end{aligned}$$

Note. Modal formulae are translated to FOL-formulae with exactly one free variable.

Proposition 1.12 (Relation to FOL)**.** Let $\phi \in \mathcal{F}(I)$ and let $\models_{\text{FOL}}$ denote the satisfaction relation of FOL. Then

1. for all models $\mathfrak{M}$ and all worlds w of $\mathfrak{M}$, we have:

$$\mathfrak{M}, w \models \phi \quad \text{iff} \quad \mathfrak{M}, \eta \models_{\text{FOL}} ST_x(\phi) \text{ where } \eta = [w/x].$$

2. for all models $\mathfrak{M}$, we have:

$$\mathfrak{M} \models \phi \quad \text{iff} \quad \mathfrak{M} \models_{\text{FOL}} \forall x.\, ST_x(\phi)$$

Proof. 1. The proof is left as an exercise.

2. We note that

$$\begin{aligned}\mathfrak{M} \models \phi \quad &\text{iff} \quad \forall w \in W.\, \mathfrak{M}, w \models \phi \\ &\text{iff} \quad \forall w \in W.\, \mathfrak{M}, \eta \models_{\mathrm{FOL}} ST_x(\phi) \text{ where } \eta = [w/x] \\ &\text{iff} \quad \mathfrak{M} \models_{\mathrm{FOL}} \forall x.\, ST_x(\phi),\end{aligned}$$

where the second equivalence is by item 1.

□

This allows us to transfer some results on FOL to modal logic.

Compactness. FOL has the *compactness* property: If Φ is a set of FOL-formulae, and every finite subset of Φ is satisfiable, then the whole set Φ is satisfiable.

We observe that by Proposition 1.12, modal logic also has the compactness property: For sets Φ of modal formulae, put

$$ST_x(\Phi) = \{ST_x(\phi) \mid \phi \in \Phi\}.$$

Let Φ_{ML} be a set of modal formulae such that every finite subset of Φ_{ML} is satisfiable. Then for each $\Phi_f \subseteq \Phi_{\mathrm{ML}}$, $ST_x(\Phi_{\mathrm{ML}})$ and $ST_x(\Phi_f)$ are sets of FOL-formulae with $ST_x(\Phi_f) \subseteq ST_x(\Phi_{\mathrm{ML}})$. As Φ_f is satisfiable, there is a model $\mathfrak{M}$ and a world w of $\mathfrak{M}$ such that $\mathfrak{M}, w \models \phi$ for all $\phi \in \Phi_f$. By Proposition 1.12, for all $\phi \in \Phi_f$, $\mathfrak{M}, \eta \models_{\mathrm{FOL}} ST_x(\phi)$, where $\eta = [w/x]$, showing that $ST_x(\Phi_f)$ is satisfiable. By compactness of FOL, $ST_x(\Phi_{\mathrm{ML}})$ is satisfiable as well so that there is a model $\mathfrak{M}'$ and a world w' of M' such that $\mathfrak{M}', \eta \models_{\mathrm{FOL}} ST_x(\Phi_{\mathrm{ML}})$, where $\eta = [w'/x]$. Using the backwards direction of Proposition 1.12, we obtain that $\mathfrak{M}', w \models \Phi_{\mathrm{ML}}$, showing that modal logic has the compactness property.

Löwenheim-Skolem Property. FOL has the *Löwenheim-Skolem property*: If a set of formulae Φ is satisfiable, then Φ has a *countable* model.

Again, by Proposition 1.12, modal logic has this property:

Let Φ_{ML} be satisfiable. As FOL has the Löwenheim-Skolem property, $ST_x(\Phi_{\mathrm{ML}})$ has a countable model $\mathfrak{M}$ containing a world w, with $\mathfrak{M}, \eta \models_{FOL} ST_x(\Phi_{\mathrm{ML}})$ for $\eta = [w/x]$. By Proposition 1.12, extended to sets of formulae, $\mathfrak{M}, w \models \Phi_{\mathrm{ML}}$, i.e. Φ_{ML} has a countable model.

Also this brings some "modal flavour" to FOL:

1. FOL-formulae that can be obtained by translating modal formulae are of restricted shape:

   - existential quantifiers appear only in the form $\exists y.\, R_i(x, y) \wedge \phi(y)$
   - universal quantifiers appear only in the form $\forall y.\, R_i(x, y) \rightarrow \phi(y)$

   This observation gives rise to the *guarded fragment* of FOL, which is – in contrast to full FOL – decidable.

   Is every guarded FOL-formula equivalent to a modal formula? No, e.g. the guarded FOL-formula

   $$\exists x.\, R(y, x) \wedge R(x, x)$$

   is not equivalent to a modal formula.

2. We can modify the standard translation so that it uses only two variables: simply alternate when "going down" in the induction, e.g. the modal formula $\Diamond(\Box p \to q)$ is translated to $\exists y_1.\, R(x, y_1) \wedge (\forall x.\, (R(y_1, x) \to P(x)) \to Q(y_1))$. Note that x is shadowed inside the universal quantifier.

   The *two-variable fragment* of FOL is also decidable.

   Is every two-variable FOL-formula equivalent to a modal formula? No, e.g. the two-variable FOL-formula

$$\exists x.\, R(y, x) \wedge R(x, y)$$

   is not equivalent to (the standard translation of) a modal formula. We will see why.

Chapter 2

Modal Expressivity

What is the expressive power of modal logic? What is "expressive power"? One possible characterization: find a condition that describes when two worlds in two Kripke structures *cannot* be distinguished by modal formulae.

Definition 2.1. Let $\mathfrak{M}, \mathfrak{M}'$ be models, w a world of M, w' a world of $\mathfrak{M}'$. The *type* of w in $\mathfrak{M}$ is defined as

$$T_{\mathfrak{M}(w)} \coloneqq \{\phi \in \mathrm{Form}(I) \mid \mathfrak{M}, w \models \phi\}$$

If $T_{\mathfrak{M}(w)} = T_{\mathfrak{M}'(w')}$, then we say w and w' are *modally equivalent* and denote this by $\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w'$. We will omit mention of $\mathfrak{M}, \mathfrak{M}'$ when these are clear from the context.

2.1 Invariance Results

We introduce three important ways of constructing new models from old ones, which leave the type of worlds unchanged. These basic constructs are very useful for proving various results. For simplicity, we consider uni-modal logic.

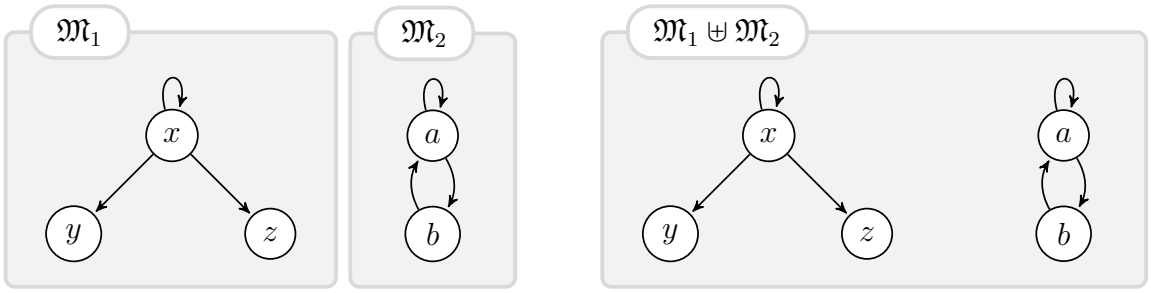
2.1.1 Disjoint Unions

Definition 2.2. Two models are called *disjoint* if their sets of worlds are disjoint. For a family of pairwise disjoint models $\mathfrak{M}_j = ((W_j, R_j, V_j))_{j \in K}$ their *disjoint sum* is the model

$$\biguplus_{j \in K} \mathfrak{M}_j \coloneqq (W, R, V)$$

where $W = \bigcup_{j \in K} W_j, R = \bigcup_{j \in K} R_j, V(p) = \bigcup_{j \in K} V_j(p)$ for each $p \in \mathcal{A}$. Of course, we can *make* any two given models disjoint by just renaming their worlds; we understand the *disjoint sum* of any two models as arising by the above construction applied to suitably renamed models in this sense. (Of course, the disjoint sum is then, again, only determined up to renaming the worlds.)

Example 2.3. The following diagram depicts two individual models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ and their disjoint union $\mathfrak{M}_1 \uplus \mathfrak{M}_2$:



Proposition 2.4. Let $\mathfrak{M}_j = (W_j, R_j, V_j))_{j \in K}$ be a family of pairwise disjoint models. For each $j \in K$ and all worlds $w \in W_j$, we have

$$\mathfrak{M}_j, w \equiv_{ML} (\biguplus_{j \in K} \mathfrak{M}_j), w.$$

Proof. Exercise. Use structural induction on formulae. □

Note. Not all classes of frames/models are closed under disjoint unions, e.g. the class of linear orders is not, as e.g. $<_{\mathbb{N}} \uplus <_{\mathbb{Q}}$ is not a linear order.

Example 2.5. We consider the extension of the basic modal language by a "global diamond" $\mathsf{E}\phi$ whose semantics is

$$\mathfrak{M}, w \models \mathsf{E}\phi \quad \text{iff} \quad \mathfrak{M}, w' \models \phi \text{ for } \textit{some } w' \in W.$$

Its dual, the "global box" is defined as $\mathsf{A}\phi = \neg\mathsf{E}\neg\phi$, i.e.

$$\begin{aligned} \mathfrak{M}, w \models \mathsf{A}\phi \quad &\text{iff} \quad \mathfrak{M}, w' \models \phi \text{ for } \textit{all } w' \in W \\ &\text{iff} \quad \mathfrak{M} \models \phi \end{aligned}$$

E and A add a "global flavour" to modal logic. Can E and A be defined in the basic modal language like $\vee$ or $\Box$, or do they extend it? Answer: These operators can *not* be defined inside the basic modal language:
Suppose there is a modal formula $\alpha(p)$ s.t. for every model $\mathfrak{M}$, we have

$$\mathfrak{M}, w \models \alpha(p) \quad \text{iff} \quad \mathfrak{M} \models p.$$

Then take two models $\mathfrak{M}_1$ and $\mathfrak{M}_2$ s.t. $\mathfrak{M}_1 \models p$ and $\mathfrak{M}_2 \models \neg p$. Let w be a world of $\mathfrak{M}_1$. We have $\mathfrak{M}_1, w \models \alpha(p)$ and thus by Proposition 2.4, $\mathfrak{M}_1 \uplus \mathfrak{M}_2, w \models \alpha(p)$. This is in contradiction to $\mathfrak{M}_1 \uplus \mathfrak{M}_2, w' \models \neg p$ for all w' of $\mathfrak{M}_2$.

In general, invariance results are suitable for showing *un*definability

2.1.2 Generated Submodels

Disjoint unions are useful for constructing bigger models from smaller ones. Generated submodels address the converse direction (picking out smaller models from existing bigger ones).

Definition 2.6. Let $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$ be two models. $\mathfrak{M}'$ is called a *submodel* of $\mathfrak{M}$, if

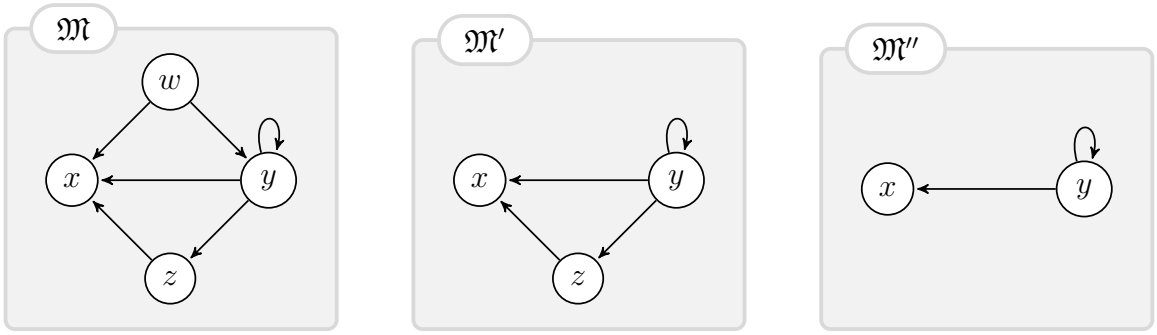
1. $W' \subseteq W$

2. $R' = R \cap (W' \times W')$

3. $V'(p) = V(p) \cap W'$ for all $p \in \mathcal{A}$.

$\mathfrak{M}'$ is a *generated submodel* of $\mathfrak{M}$, if it is a submodel of $\mathfrak{M}$ and also satisfies the closure condition that $w \in W'$ and $(w, w') \in R$ implies $w' \in W'$.

Example 2.7. Consider the following three models $\mathfrak{M}$, $\mathfrak{M}'$ and $\mathfrak{M}''$. While $\mathfrak{M}'$ is a generated submodel of $\mathfrak{M}$, $\mathfrak{M}''$ lacks the *generatedness property*, i.e. $\mathfrak{M}''$ is missing the world z to which y has a transition in $\mathfrak{M}$ so that $\mathfrak{M}''$ is a submodel of $\mathfrak{M}$, but not a generated submodel of $\mathfrak{M}$.



Proposition 2.8. Let $\mathfrak{M}$ be a model, $\mathfrak{M}'$ a generated submodel of $\mathfrak{M}$, and consider w be a world of $\mathfrak{M}'$. Then

$$\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w.$$

Proof. The proof is by structural induction over formulae (Exercise). □

2.1.3 p-Morphisms (also known as bounded morphisms)

Disjoint union and generated submodels are rather "brute force" methods for constructing new models. p-morphisms are a more subtle concept that generalizes both of them.

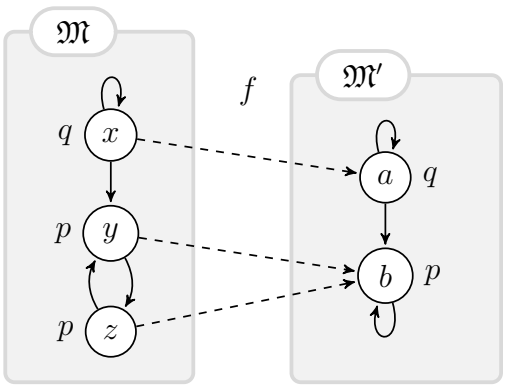
Definition 2.9. Let $\mathfrak{M} = (W, R, V)$ and $\mathfrak{M}' = (W', R', V')$ be two models. A mapping $f : W \to W'$ is a *p-morphism* from $\mathfrak{M}$ to $\mathfrak{M}'$ if it satisfies the following properties, for all $w, w' \in W$:

1. $w \in V(p)$ iff $f(w) \in V'(p)$ for all $p \in \mathcal{A}$,

2. if $(w, w') \in R$, then $(f(w), f(w')) \in R'$,

3. if $(f(w), v') \in R'$, then there is a $v \in W$ s.t. $(w, v) \in R$ and $f(v) = v'$.

Example 2.10. The following diagram depicts two models $\mathfrak{M}$ and $\mathfrak{M}'$ and a p-morphism f from $\mathfrak{M}$ to $\mathfrak{M}'$:



Notice how we have $f(x) = a$, $f(y) = f(z) = b$, as indicated by the dashed arrows. It is easily checked that f indeed fulfils the three properties of p-morphisms.

Proposition 2.11. Let $\mathfrak{M}$ and $\mathfrak{M}'$ be models s.t. there is a p-morphism f from $\mathfrak{M}$ to $\mathfrak{M}'$. Then we have that for all worlds w of $\mathfrak{M}$,

$$\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', f(w).$$

Proof. Exercise. Use structural induction on formulae. □

As an example application of p-morphisms, let us prove that every satisfiable formula has a tree-shaped model.

Definition 2.12. A model $\mathfrak{M} = (\mathcal{W}, \mathcal{R}, \mathcal{V})$ is a *tree-model* for ϕ iff the frame $(\mathcal{W}, \mathcal{R})$ is a tree with root w such that $\mathfrak{M}, w \models \phi$.
For multi-modal languages, we demand that $(\mathcal{W}, \bigcup_{i \in I} \mathcal{R}_i)$ is a tree.

Proposition 2.13. Every satisfiable modal formula has a tree-model.

Proof. Let ϕ be a satisfiable formula, $\mathfrak{M} = (\mathcal{W}, \mathcal{R}, \mathcal{V})$ a model, and w a world of $\mathfrak{M}$ with $\mathfrak{M}, w \models \phi$. Define a new model $\mathfrak{M}' = (\mathcal{W}', \mathcal{R}', \mathcal{V}')$ by *unravelling* $\mathfrak{M}$ at w:

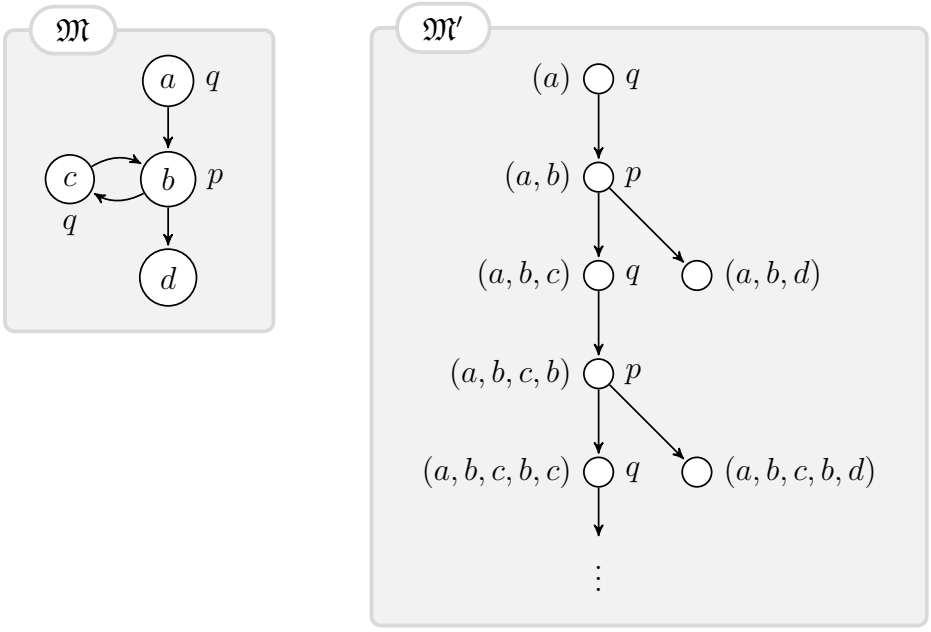
- $\mathcal{W}'$ consists of all finite sequences $(w, u_1, \ldots, u_n)$ such that
  - $n \geq 0$,
  - $(w, u_1) \in \mathcal{R}$ if $n > 0$,
  - $(u_i, u_{i+1}) \in \mathcal{R}$ for $1 \leq i \leq n$.
- $\mathcal{R}' = \{(w, u_1, \ldots, u_n), (w, u_1, \ldots, u_n, v) \mid (u_n, v) \in \mathcal{R}\}$.
- $\mathcal{V}'(p) = \{(w, u_1, \ldots, u_n) \mid u_n \in \mathcal{V}(p)\}$ for all $p \in \mathcal{A}$.

Clearly, $(\mathcal{W}', \mathcal{R}')$ is a (potentially infinite) tree with root (w). The mapping

$$f : (w, u_1, \ldots, u_n) \mapsto u_n$$

is easily seen to be a p-morphism from $\mathfrak{M}$ to $\mathfrak{M}'$ (Exercise: prove this). By Proposition 2.11, $\mathfrak{M}, w \models \phi$ implies $\mathfrak{M}', f(w) \models \phi$. □

Example 2.14. The following diagram shows a model $\mathfrak{M}$ and its unravelling at a:



Note. If a logic has the property formulated in Proposition 2.13, we say it has the *tree model property* (TMP). This property is useful for proving decidability and complexity results.

2.2 Bisimulations

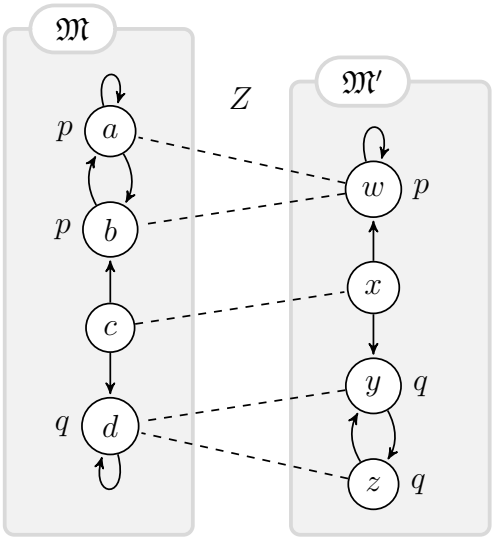
Recall that p-morphisms are *functions* from one model into another. To characterize modal equivalence adequately, we need a *relation* between sets of worlds.

Definition 2.15. Let $\mathfrak{M} = (\mathcal{W}, \mathcal{R}, \mathcal{V})$ and $\mathfrak{M}' = (\mathcal{W}', \mathcal{R}', \mathcal{V}')$ be two models. A relation $Z \subseteq \mathcal{W} \times \mathcal{W}'$ is a *bisimulation* between $\mathfrak{M}'$ and $\mathfrak{M}$ if whenever wZw', then the following conditions hold:

- w and w' satisfy the same propositional letters;
- whenever $(w, v) \in \mathcal{R}$, then there exists a $v' \in \mathcal{W}'$ such that vZv' and $(w', v') \in \mathcal{R}'$ (*forth* condition).
- whenever $(w', v') \in \mathcal{R}'$, then there exists a $v \in \mathcal{W}$ such that vZv' and $(w, v) \in \mathcal{R}$ (*back* condition).

We write $\mathfrak{M}, w \simeq \mathfrak{M}', w'$ if there exists a bisimulation Z between $\mathfrak{M}$ and $\mathfrak{M}'$ with wZw'.

Example 2.16. The following diagram depicts two models $\mathfrak{M}$, $\mathfrak{M}'$ and a bisimulation Z between $\mathfrak{M}$ and $\mathfrak{M}'$:



Notice how we have $Z = \{(a, w), (b, w), (c, x), (d, y), (d, z)\}$, as indicated by the dashed lines. It is easily checked that Z indeed fulfils the three properties of bisimulations. It is also worth noticing that we can neither define a p-morphism from $\mathfrak{M}$ to $\mathfrak{M}'$ nor define a p-morphism from $\mathfrak{M}'$ to $\mathfrak{M}$.

It is intuitively helpful to develop a game-based view of bisimilarity, as follows. The *bisimilarity game* on models $\mathfrak{M}_1 = (W_1, (R_i^1)_{i \in I}, V_1)$, $\mathfrak{M}_2 = (W_2, (R_i^2)_{i \in I}, V_2)$ is played between two players *Abelard* ($\forall$) and *Eloise* ($\exists$); Eloise aims to show that two given worlds are bisimilar, while Abelard aims to refute this. The game is defined as follows:

- *Positions*: Pairs $(w_1, w_2) \in W_1 \times W_2$.

- *Moves*: The game proceeds in rounds; in each round, starting at a current position (w_1, w_2), $\forall$ first picks one of the models, say $\mathfrak{M}_1$, and then selects $i \in I$ and $w_1' \in W_1$ such that $w_1 R_i^1 w_1'$. Then $\exists$ has to reply with a world $w_2' \in W_2$ such that $w_2 R_i^2 w_2'$. The game then continues from the new position (w_1', w_2'). (N.B.: $\forall$ may pick a different side in each round.)

- *Winning conditions:*

  - $\forall$ wins as soon as a position (w_1, w_2) is reached such that for some $p \in \mathcal{A}$, either $w_1 \in V_1(p)$ and $w_2 \notin V_2(p)$ or $w_1 \notin V_1(p)$ and $w_2 \in V_2(p)$.
  - Any player who cannot move (because the required successor worlds do not exist), loses.
  - Infinite plays are won by $\exists$.

While we keep formal definitions of game-theoretic notions mostly implicit, relying instead on standard intuitions, we emphasize the following points:

- We distinguish terminologically between the *game*, defined by its positions, moves etc. as above, and a *play* of the game; a play is an actual sequence of positions reached according to the decisions made by the players.

- We only care about perfect play, that is, when we say that a player *wins* a game, we mean that he or she has a *winning strategy*, i.e. can choose moves so that he or she wins the play no matter what the opponent does.

- We gloss over issues of *determinacy*, i.e. we take it for granted that one of the players will always have a winning strategy. (This is problematic only for infinite games; in our present case, this is not actually a problem because all infinite games are won by the same player.)

The game characterizes bisimilarity in the following sense:

Lemma 2.17. *Let $\mathfrak{M}_1$, M_2 be models. Worlds $w_1 \in W_1$, $w_2 \in W_2$ are bisimilar iff $\exists$ wins the position (w_1, w_2) in the bisimilarity game on $\mathfrak{M}_1$, $\mathfrak{M}_2$.*

Proof. Exercise: Use the principle of describing winning strategies by positional invariants. □

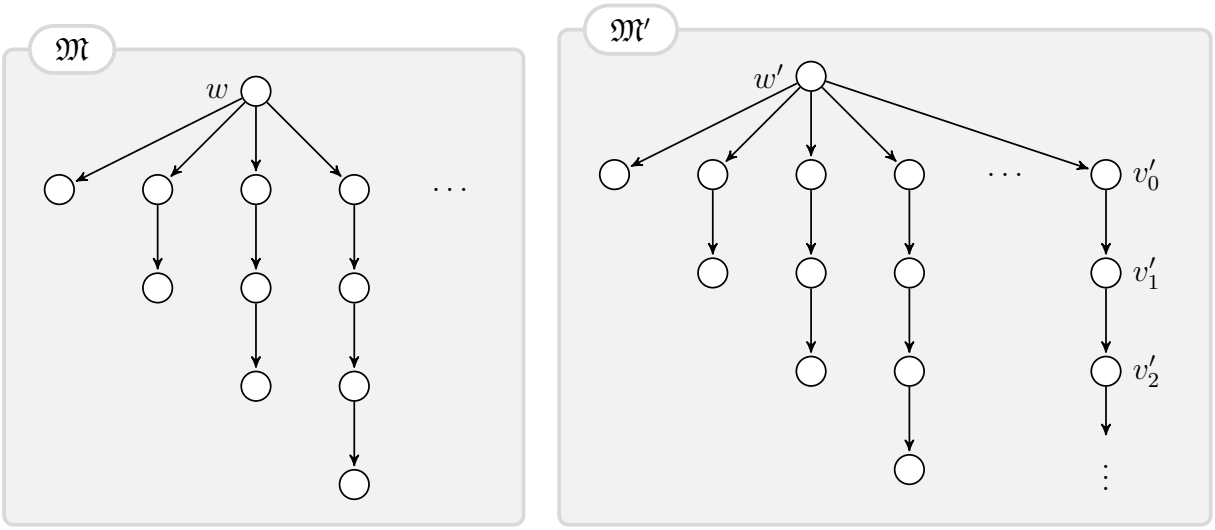
Importantly, modal logic is *invariant* under bisimilarity:

Lemma 2.18. *Let ϕ be a modal formula. If $\mathfrak{M}, w \simeq \mathfrak{M}', w'$, then $\mathfrak{M}, w$ and $\mathfrak{M}', w'$ agree on ϕ, i.e. $\mathfrak{M}, w \models \phi$ iff $\mathfrak{M}', w' \models \phi$.*

Proof. Induction on ϕ. The steps for Boolean constructs ($\bot$, $\neg$, $\wedge$) are trivial; for instance, in the inductive step for negation, we have $\mathfrak{M}.w \models \neg\phi$ iff $\mathfrak{M}.w \not\models \phi$ iff (by the inductive hypothesis) $\mathfrak{M}', w' \not\models \phi$ iff $\mathfrak{M}', w' \models \neg\phi$. The inductive step for propositional letters is immediate from the first clause in the definition of bisimulation. We do the inductive step for the modality $\Box_i$, using the fact that D wins the position (w, w') in the bisimilarity game on $\mathfrak{M}$, $\mathfrak{M}'$. By symmetry, it suffices to prove one implication of the inductive claim. So suppose that $\mathfrak{M}, w \models \Box_i\phi$. We have to show that $\mathfrak{M}', w' \models \Box_i\phi$. So let $w' R_i' v'$; we have to show $v' \models \phi$. Now S can move from v to v' in the bisimilarity game; let v be the winning response of D. Then D wins $(v, v'$. Since $w R_i v$, we have $v \models \phi$; by the inductive hypothesis, it follows that $v' \models \phi$ are required. □

The question thus arises whether bisimulation *characterizes* modal equivalence, i.e. whether the converse implication is true: does $\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w'$ imply $\mathfrak{M}, w \simeq \mathfrak{M}', w'$? Indeed, this is not true in general:

Example 2.19. Let $\mathfrak{M}$ be a tree-model with $|\mathbb{N}|$ paths, where the k-th path has length k, and let w be the "root" world of $\mathfrak{M}$. Let $\mathfrak{M}'$ be a similar model, just with one more path, which is of infinite length, and let w' be the "root" world of $\mathfrak{M}'$, as indicated by the following diagrams:



Then $\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w'$ (why?). On the other hand, $\mathfrak{M}, w$ and $\mathfrak{M}', w'$ fail to be bisimilar (why?).

However, the converse of Lemma 2.18 does hold if we restrict the class of models, disabling the above counterexample:

Definition 2.20. A model $\mathfrak{M} = (\mathcal{W}, \mathcal{R}, \mathcal{V})$ is *image-finite* if for every $w \in \mathcal{W}$, the set

$$R(w) = \{w' \in \mathcal{W} \mid (w, w') \in \mathcal{R}\}$$

is finite. In particular, every finite model is image-finite.

Theorem 2.21 (Hennessy-Milner)**.** Let $\mathfrak{M}$ and $\mathfrak{M}'$ be image-finite models such that there exists a bisimulation between $\mathfrak{M}$ and $\mathfrak{M}'$. Then, for each $w \in \mathcal{W}$ and $w' \in \mathcal{W}'$, we have

$$\mathfrak{M}, w \simeq \mathfrak{M}', w' \quad \text{iff} \quad \mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w'.$$

Proof.

- *'only if'*: by Lemma 2.18.
- *'if'*: We prove that the relation $\equiv_{ML}$ is a bisimulation. So let $\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w'$. We check the conditions of the definition:
  - Agreement on propositional letters: Trivial.
  - Forth condition: Let and $(w, v) \in \mathcal{R}$. To the contrary of what is to be shown, assume that there is *no* v' in $\mathfrak{M}'$ such that $(w', v') \in \mathcal{R}'$ and $\mathfrak{M}, v \equiv_{ML} \mathfrak{M}', v'$. Let $s' = \{u' \in \mathcal{W}' \mid (w', u') \in \mathcal{R}\}$; since $\mathfrak{M}$ is finitely branching, we may write $s' = \{v'_1, \ldots, v'_n\}$, with $n \geq 0$. For every $v'_i \in s'$, we have, by assumption and using negation if needed, a formula ψ_i such that $\mathfrak{M}, v \models \psi_i$ but $\mathfrak{M}', v'_i \not\models \psi_i$. It follows that

$$\mathfrak{M}, w \models \Diamond(\psi_1 \wedge \cdots \wedge \psi_n) \text{ and } \mathfrak{M}', w' \not\models \Diamond(\psi_1 \wedge \cdots \wedge \psi_n),$$

which contradicts $\mathfrak{M}, w \equiv_{ML} \mathfrak{M}', w'$.

– Back condition: analogous to the forth condition. □

Bisimulation can help understand the "modal fragment" of FOL. We have already seen that there are FOL-formulae with one free variable that are *not* equivalent to the standard translation of a modal formula, e.g.

$$\exists y.(R(x, y) \wedge R(y, x)).$$

How can we describe the class of FOL-formulae that are equivalent (the translation of) a modal formula?

Definition 2.22. Let $\phi(x)$ be a first-order formula with one free variable in the signature $(\mathcal{R}, \mathcal{A})$. Then $\phi(x)$ is *invariant for bisimulations* if, for all models $\mathfrak{M}$ and $\mathfrak{M}'$, all states w of $\mathfrak{M}$ and v of $\mathfrak{M}'$, and all bisimulations Z between $\mathfrak{M}$ and $\mathfrak{M}'$ such that wZv, we have

$$\mathfrak{M}, \eta \models_{FOL} \phi \quad \text{iff} \quad \mathfrak{M}', \rho \models_{FOL} \phi,$$

where $\eta = [w/x]$ and $\rho = [v/x]$.

Theorem 2.23 (van Benthem)**.** Let $\phi(x)$ be a first-order formula with one free variable in the signature $(\mathcal{R}, \mathcal{A})$. Then $\phi(x)$ is invariant for bisimulations iff it is equivalent to the standard translation of a modal formula.

Proof. Out of the scope of this lecture. □

Chapter 3

The Syntactic View

We have introduced modal logic using a relational semantics and viewed it as a fragment of FOL. But this is not always appropriate:

For example, in logics of knowledge and belief, it is not always immediately clear which frame class should be used. Such logics are usually defined by saying which statements should be true:

- $\Box\phi \to \phi$ for all formulae ϕ: if the agent knows ϕ, then ϕ is true.
- $\Box\phi \to \Box\Box\phi$ for all formulae ϕ: if the agent knows ϕ then he knows that he knows ϕ (positive introspection).
- $\neg\Box\phi \to \Box\neg\Box\phi$ for all formulae ϕ: if the agent does not know ϕ then he knows that he does not know ϕ (negative introspection).

Using this approach, how can the semantics be defined?

3.1 Modal logics and normal modal logics

The following definition of modal logics is inspired by Hilbert-style proof systems:

Definition 3.1. A *modal logic* (or *logic*, for short) Λ is a set of modal formulae that

1. contains all propositional tautologies,
2. is closed under *modus ponens* (MP), i.e.

   if $\phi \in \Lambda$ and $\phi \to \psi \in \Lambda$, then $\psi \in \Lambda$,

3. and is closed under *uniform substitution*, i.e. if $\phi \in \Lambda$ and ψ can be obtained from ϕ by uniformly replacing propositional letters with arbitrary formulae, then $\psi \in \Lambda$.

If $\phi \in \Lambda$, we say that ϕ is *theorem* of Λ and write $\vdash_\Lambda \phi$.

Definition 3.2. Let ϕ be a formula, $\mathcal{F}$ a frame, and S a class of frames. Then ϕ is *valid* in

- $\mathcal{F}$ (written $\models_{\mathcal{F}} \phi$) iff $\mathfrak{M}, w \models \phi$ for every model $\mathfrak{M}$ based on $\mathcal{F}$ and every world w of $\mathcal{F}$,

- S (written $\models_S \phi$) iff $\models_{\mathcal{F}} \phi$ for every frame $\mathcal{F} \in S$.

We write $\models \phi$ if ϕ is valid in the class of all frames and we use Λ_S to denote the set of all formulae valid in S, i.e. we put $\Lambda_S = \{\phi \mid \models_S \phi\}$.

Example 3.3. We consider the following statements about validity of formulae:

- $\models \Diamond(p \vee q) \to \Diamond p \vee \Diamond q$
- $\models_{\mathrm{K4}} \Box p \to \Box\Box p$, where K4 is the class of all transitive frames.

Note.
- For all formulae ϕ, we have $\models_S \phi$ iff $\neg\phi$ is unsatisfiable in S.
- Validity is defined on the level of frames, whereas global truth is defined on the level of models. More precisely: $\models_{\mathcal{F}} \phi$ implies $\mathfrak{M} \models \phi$ if $\mathfrak{M}$ is based on $\mathcal{F}$ (but not vice versa).

Example 3.4. Some examples for modal logics:

1. The set of all formulae is a logic, the *inconsistent logic*.
2. Let PC be the smallest set containing all propositional tautologies that is closed under uniform substitution. PC is a logic.
3. Let S be a class of frames. Then Λ_S is a logic. It is easily checked (exercise!) that each propositional tautology is valid in every frame and that validity is preserved by modus ponens and uniform substitution (e.g. if $\models_S \phi$ and $\models_S \phi \to \psi$ then $\models_S \psi$).
4. If M is a class of models, then

$$\Lambda_M = \{\phi \mid \mathfrak{M} \models \phi \text{ for all } \mathfrak{M} \in M\}$$

need *not* be a logic. Consider a class M of models containing only one model $\mathfrak{M}$ with $\mathcal{V}(p) = \mathcal{W}$ and $\mathcal{V}(q) \neq \mathcal{W}$. Then $p \in \Lambda_M$, but $q \notin \Lambda_M$ implying that Λ_M is not closed under uniform substitution.

The third example shows that there is a close connection between modal logics and validity in classes of frames, whereas the fourth example shows that we cannot work on the level of models.

What does the logic Λ_S of a class of frames S "look like"?

Definition 3.5. A modal logic Λ is called *normal* if it contains the formulae

$$\begin{aligned} &\text{(K)} && \Box(p \to q) \to (\Box p \to \Box q), \\ &\text{(Dual)} && \Diamond p \leftrightarrow \neg\Box\neg p, \end{aligned}$$

and is closed under *generalization*, i.e. $\vdash_\Lambda \phi$ implies $\vdash_\Lambda \Box\phi$.

Example 3.6. We consider some examples of normal modal logics:

1. The inconsistent logic is a normal modal logic.
2. PC is *not* a normal logic (since (K)$\notin PC$).

3. If $\{\Lambda_i \mid i \in I\}$ is family of normal logics, then $\cap_{i\in I}\Lambda_i$ is a normal logic.

4. If S is any class of frames, then Λ_S is a normal logic.

It is convenient to think of modal logics in terms of Kripke structures, but the second and the fourth examples imply that there exist (non-normal!) modal logics Λ such that $\Lambda \neq \Lambda_S$ for all classes of frames S. Thus not all modal logics can be understood in terms of Kripke structures.

Definition 3.7. Let Γ be a set of formulae (called *axioms* in this context). Due to Examples 3.6.1 and 3.6.3, there is for each Γ a unique smallest normal modal logic containing Γ. This logic is called the *normal modal logic generated by* Γ. The normal modal logic generated by $\emptyset$ is called **K**.

Example 3.8. Consider the following important axioms for the generation of normal modal logics:

(M) or (T)	$p \to \Diamond p$	(reflexivity)
(4)	$\Diamond\Diamond p \to \Diamond p$	(transitivity)
(B)	$p \to \Box\Diamond p$	(symmetry)
(D)	$\Box p \to \Diamond p$	(seriality)
(L) or (G)	$\Box(\Box p \to p) \to \Box p$	(Löb-axiom)

We note that dual representations of axioms that involve implications can be obtained by considering their contrapositive, e.g. for the transitivity axiom, we observe that

$$\begin{aligned}\Diamond\Diamond p \to \Diamond p &= \neg(\Diamond\Diamond p) \vee \Diamond p \\ &= \Box\Box\neg p \vee \neg\Box\neg p \\ &= \Box\neg p \to \Box\Box\neg p.\end{aligned}$$

The notion of normal modal logics is appropriate for describing modal logics of classes of frames:

1. It is not too strong by Example 3.6.4

2. As we will see, the minimal normal logic **K** comprises *precisely* the formulae valid in the class of all frames.

We now introduce the fundamental concepts linking the syntactic and the semantic perspectives.

Definition 3.9 (Soundness)**.** Let S be a class of frames. A modal logic Λ is *sound* with respect to S if, for all formulae ϕ, $\vdash_\Lambda$ implies $\models_S \phi$, i.e. if $\Lambda \subseteq \Lambda_S$.

Soundness proofs for generated modal logics are often simple: suppose that we want to show that a normal modal logic generated by some axiom A is sound w.r.t. some class of frames S. Since we know (from Examples 3.4.3 and 3.6.4) that

- (K), $(Dual)$ and all propositional tautologies are valid in S, and
- modus ponens, uniform substitution and generalization preserve validity,

it suffices to show that $\models_S A$.

Lemma 3.10. The following logics are sound w.r.t. the respective class of frames:

Logic	generating axioms	is sound for the class of
K	$\emptyset$	all frames
K4	$\{(4)\}$	transitive frames
T	$\{(T)\}$	reflexive frames
B	$\{(B)\}$	symmetric frames
S4	$\{(T),(4)\}$	reflexive, transitive frames
S5	$\{(T),(4),(B)\}$	reflexive, transitive, symmetric frames

Proof. The proof is left as exercise! □

Definition 3.11. Let S be a class of frames. A modal logic Λ is *complete* w.r.t. S, if, for all formulae ϕ, $\models_S \phi$ implies $\vdash_\Lambda \phi$, i.e. if $\Lambda_S \subseteq \Lambda$.

Completeness proofs are usually much harder than soundness proofs.

3.2 The canonical model for K

We use the method of canonical models to show that **K** is complete w.r.t. the class of all frames. It is convenient to do this based on consistency.

Definition 3.12. Let Λ be a logic. Then

- a formula ϕ is said to be Λ-*consistent* if $\nvdash_\Lambda \neg\phi$, and Λ-*inconsistent* otherwise.
- a set of formulae Γ is Λ-consistent if, for all finite subsets $\{\psi_1,\dots,\psi_n\} \subseteq \Gamma$, the formula $\psi_1 \wedge \dots \wedge \psi_n$ is Λ-consistent, and Λ-inconsistent otherwise.

Proposition 3.13. Let Λ be a logic and S a class of frames. Then

1. Λ is sound w.r.t. S iff every formula that is satisfiable in S is Λ-consistent.
2. Λ is complete w.r.t. S iff every Λ-consistent formula is satisfiable in S.

Proof. Exercises! □

Definition 3.14 (Λ-MCS)**.** A set of formulae Γ is *maximal* Λ-*consistent* if Γ is Λ-consistent, and any set of formulae properly containing Γ is inconsistent.

Lemma 3.15 (Consistency)**.** Let Λ be a logic, ϕ a formula and Γ be a Λ-consistent set of formulae. Then $\Gamma \cup \{\phi\}$ is Λ-consistent or $\Gamma \cup \{\neg\phi\}$ is Λ-consistent.

Proof. Assume both sets are Λ-inconsistent. Then there must be formulae $\psi_1,\dots,\psi_m \in \Gamma$ and $\psi'_1,\dots,\psi'_n \in \Gamma$ s.t. both $\psi_1 \wedge \dots \wedge \psi_m \wedge \phi$ and $\psi'_1 \wedge \dots \wedge \psi'_n \wedge \neg\phi$ are Λ-inconsistent, i.e.

$$\vdash_\Lambda \neg(\psi_1 \wedge \dots \wedge \psi_m \wedge \phi) \tag{1}$$
$$\vdash_\Lambda \neg(\psi'_1 \wedge \dots \wedge \psi'_n \wedge \neg\phi) \tag{2}$$

We have $\vdash_\Lambda \neg(a \wedge b) \to (\neg(a' \wedge \neg b) \to \neg(a \wedge a'))$, as this is a tautology. Substitute $a \mapsto (\psi_1 \wedge \dots \wedge \psi_m), a' \mapsto (\psi'_1 \wedge \dots \wedge \psi'_n), b \mapsto \phi$. By (1) and (2) in combination with modus ponens (twice), $\vdash_\Lambda \neg(\psi_1 \wedge \dots \wedge \psi_m \wedge \psi'_1 \wedge \dots \wedge \psi'_n)$ which contradicts Λ-consistency of Γ. □

Proposition 3.16 (Hintikka properties)**.** Let Λ be a logic, Γ be a Λ-MCS. Then

1. for all formulae ϕ, we have that $\phi \in \Gamma$ or $\neg\phi \in \Gamma$, but not both
2. for all formulae ϕ and ψ, we have $\phi \wedge \psi \in \Gamma$ iff $\phi \in \Gamma$ and $\psi \in \Gamma$.
3. Γ is closed under modus ponens, i.e. $\{\phi, \phi \to \psi\} \subseteq \Gamma$ implies $\psi \in \Gamma$.
4. $\Lambda \subseteq \Gamma$.

Proof: Exercises!

Lemma 3.17 (Lindenbaum's Lemma)**.** Let Λ be a logic and let Σ be a Λ-consistent set of formulae. Then there is a Λ-MCS Σ^+ s.t. $\Sigma \subseteq \Sigma^+$.

Proof: Let $\phi_0, \phi_1, \ldots$ be an enumeration of all formulae. We define Σ^+ as follows:

$$\begin{aligned} \Sigma^0 &= \Sigma \\ \Sigma^{n+1} &= \begin{cases} \Sigma^n \cup \{\phi_n\} & \text{if this set is consistent,} \\ \Sigma^n \cup \{\neg\phi_n\} & \text{otherwise} \end{cases} \\ \Sigma^+ &= \bigcup_{0 \leq n} \Sigma^n \end{aligned}$$

Clearly, $\Sigma \subseteq \Sigma^+$. The following properties imply that Σ^+ is a Λ-MCS:

1. Σ^+ is Λ-consistent: Assume Σ^+ is Λ-inconsistent. Then a finite subset $\Gamma \subseteq \Sigma^+$ is Λ-inconsistent. But then $\Gamma \subseteq \Sigma^n$ for some n, which is a contradiction since all Σ^n are Λ-consistent by Lemma 3.15.
2. Σ^+ is maximal: Let $\phi \notin \Sigma^+$ and $\phi = \phi_n$. Since $\phi_n \notin \Sigma^+ \supseteq \Sigma^{n+1}, \Sigma^n \cup \{\phi_n\}$ is Λ-inconsistent. Thus so is $\Sigma^+ \cup \{\phi_n\}$.

Definition 3.18. The *canonical model* M^Λ for a normal modal logic Λ is the triple $(W^\Lambda, R^\Lambda, V^\Lambda)$, where

1. W^Λ is the set of all Λ-MCS,
2. R^Λ is defined by

$$(w, u) \in R^\Lambda \text{ iff, for all formulae } \phi,\ \phi \in u \text{ implies } \Diamond\phi \in w.$$

3. $V^\Lambda(p) = \Big\{ w \in W^\Lambda \;\Big|\; p \in w \Big\}$.

Note. There are two very important properties of canonical models:

1. $\mathfrak{M}^\Lambda, w \models \phi$ iff $\phi \in w$ (see Truth Lemma below).
2. Since $\mathfrak{M}^\Lambda$ consists of all Λ-MCSs, by Lindenbaum's Lemma, every Λ-consistent set of formulae Γ is true in some world of $\mathfrak{M}^\Lambda$. This is why $\mathfrak{M}^\Lambda$ is called canonical.

Lemma 3.19. For every normal modal logic Λ and for all formulae $\phi, \psi_1, \ldots, \psi_n$ we have

1. $\vdash_\Lambda \phi \to \psi$ implies $\vdash_\Lambda \Box\phi \to \Box\psi$, and
2. $\vdash_\Lambda (\Box\psi_1 \wedge \cdots \wedge \Box\psi_n) \to \Box(\psi_1 \wedge \cdots \wedge \psi_n)$.

Proof. Follows from a more general statement proved as an exercise. □

Lemma 3.20 (Existence Lemma)**.** For all normal modal logics Λ and all worlds $w \in W^\Lambda$, if $\Diamond\phi \in w$, then there is a world $v \in W^\Lambda$ s.t. $(w, v) \in R^\Lambda$ and $\phi \in v$.

Proof. Let $\Diamond\phi \in w$. Put $u := \{\phi\} \cup \{\psi \mid \Box\psi \in w\}$. We first show that u is Λ-consistent. Assume that this is not the case, i.e. that for some $\psi_1, \ldots, \psi_n \in u$, we have

$$\vdash_\Lambda \neg(\psi_1 \wedge \cdots \wedge \psi_n).$$

Then we also have

$$\vdash_\Lambda \psi_1 \wedge \cdots \wedge \psi_n \to \neg\phi$$

and therefore, by the rule (RK) derived in the exercises,

$$\vdash_\Lambda (\Box\psi_1 \wedge \cdots \wedge \Box\psi_n) \to \Box\neg\phi \qquad (*)$$

Further, $\Box\psi_1, \ldots, \Box\psi_n \in w$ implies by Item 2. of Proposition 3.16 that $\Box\psi_1 \wedge \cdots \wedge \Box\psi_n \in w$, which in turn implies by Items 3. and 4. of Proposition 3.16 together with $(*)$ that $\Box\neg\phi \in w$. Then we have

$$\vdash_\Lambda \Diamond\phi \to \neg\Box\neg\phi \qquad \text{((Dual) + subst.)} \qquad (1)$$

$$\vdash_\Lambda \Box\neg\phi \to \neg\Diamond\phi \qquad \text{(prop.taut. + subst. + MP on (1))} \qquad (2)$$

which implies by Items 3. and 4. of Proposition 3.16 that $\neg\Diamond\phi \in w$ which is – by Item 2. of Proposition 3.16 – a contradiction to $\Diamond\phi \in w$. Hence u is consistent.

By Lindenbaum's Lemma, it follows that $u \subseteq v$ for some Λ-MCS v. It remains to show that $(w, v) \in R^\Lambda$. Assume $(w, v) \notin R^\Lambda$. Then by definition of R^Λ, $\psi \in v$ and $\Diamond\psi \notin w$ for some ψ. It follows by Item 1. of Proposition 3.16 that $\neg\Diamond\psi \in w$ which implies by (Dual) that $\Box\neg\psi \in w$. Hence by definition of u and v, $\neg\psi \in v$ which is by Item 1. of Proposition 3.16 a contradiction to $\psi \in v$. Thus $(w, v) \in R^\Lambda$. □

Lemma 3.21 (Truth Lemma)**.** For all normal modal logics Λ, all formulae ψ and all worlds in $w \in W^\Lambda$,

$$\mathfrak{M}^\Lambda, w \models \psi \text{ iff } \psi \in w.$$

Proof. We proceed by structural induction on ψ.

Base cases

- $\psi = \bot$ or $\psi = \top$: Trivial.
- If $\psi = p$, then $\mathfrak{M}^\Lambda, w \models p$ iff $p \in w$ by definition of $V^\Lambda(p)$.

Inductive step Assume as induction hypothesis that for all $i \in \{1, 2\}$ and all $v \in W^\Lambda$,

$$\mathfrak{M}^\Lambda, v \models \psi_i \quad \text{iff} \quad \psi_i \in v.$$

- If $\psi = \psi_1 \wedge \psi_2$, then $\mathfrak{M}^\Lambda, w \models \psi_1 \wedge \psi_2$ iff $\mathfrak{M}^\Lambda, w \models \psi_1$ and $\mathfrak{M}^\Lambda, w \models \psi_2$ iff (by the induction hypothesis) $\psi_1 \in w$ and $\psi_2 \in w$ iff (by Item 2. of Proposition 3.16) $\psi_1 \wedge \psi_2 \in w$.
- If $\psi = \neg\psi_1$, then $\mathfrak{M}^\Lambda, w \models \neg\psi_1$ iff $\mathfrak{M}^\Lambda, w \not\models \psi_1$ iff (by the induction hypothesis) $\psi_1 \notin w$ iff (by Item 1. of Proposition 3.16) $\neg\psi_1 \in w$.

- $\psi = \Diamond\psi_1$: For the forth direction, note that $\mathfrak{M}^\Lambda, w \models \Diamond\psi_1$ iff $\exists v \in W^\Lambda.(w,v) \in R^\Lambda$ and $\mathfrak{M}^\Lambda, v \models \psi_1$, by induction hypothesis iff $\exists v \in W^\Lambda.(w,v) \in R^\Lambda$ and $\psi_1 \in v$ and by definition of R^Λ, only if $\Diamond\psi_1 \in w$. For the backwards direction, let $\Diamond\psi_1 \in w$. By the above equivalences it suffices to find a Λ-MCS v s.t. $(w,v) \in R^\Lambda$ and $\psi_1 \in v$. Such a v exists by the Existence Lemma.

□

Theorem 3.22. **K** is complete w.r.t. the class of all frames.

Proof. By Proposition 3.13, it suffices to show that every **K**-consistent formula is satisfiable. So let ϕ be **K**-consistent. We have to find a model $\mathfrak{M}$ (based on any frame) and a world w of $\mathfrak{M}$ s.t. $\mathfrak{M}, w \models \phi$. Simply choose $\mathfrak{M}^{\mathbf{K}}$ and let Γ be a **K**-MCS with $\phi \in \Gamma$ (such a Γ exists by Lindenbaum's Lemma). By the Truth Lemma, $\mathfrak{M}^{\mathbf{K}}, \Gamma \models \phi$. □

3.3 Other frame classes

Observe that all relevant Lemmas in the preceding section have been established for *every* normal modal logic, not only for **K**. Thus, (some) other completeness results are now simple:

Theorem 3.23. **K4** is complete w.r.t. the class of all transitive frames.

Proof. Let ϕ be a **K4**-consistent formula. We need to find a model $\mathfrak{M} = (W, R, V)$ s.t. $\mathfrak{M}, w \models \phi$ for some $w \in W$ and R is transitive. By Lindenbaum's Lemma, there is a **K4**-MCS Γ with $\phi \in \Gamma$. By the Truth Lemma, $\mathfrak{M}^{\mathbf{K4}}, \Gamma \models \phi$. It thus remains to show that $R^{\mathbf{K4}}$ is transitive.
Let $w, v, u \in W^{\mathbf{K4}}$ s.t. $\{(u,v),(v,w)\} \subseteq R^{\mathbf{K4}}$. We have to show that $(u,w) \in R^{\mathbf{K4}}$. Let $\psi \in w$. It remains to show that $\Diamond\psi \in u$. Now $(v,w) \in R^{\mathbf{K4}}$ yields $\Diamond\psi \in v$ and $(u,v) \in R^{\mathbf{K4}}$ yields $\Diamond\Diamond\psi \in u$. Since u is a **K4**-MCS, $\Diamond\Diamond\psi \to \Diamond\psi \in u$ and u is closed under modus ponens (by Proposition 3.16), $\Diamond\psi \in u$. □

Two observations

- The proof strategy is simple: just show that the canonical model has the desired property.
- The proof that $R^{\mathbf{K4}}$ is transitive only uses the fact that $\Diamond\Diamond\psi \to \Diamond\psi \in \mathbf{K4}$. Hence the canonical frame of any logic containing (4) is also transitive.

Lemma 3.24. Let Λ be a normal modal logic. If $(4) \in \Lambda$, then $\mathfrak{M}^\Lambda$ is based on a transitive frame.

Proof. See proof of Theorem 3.23. □

Lemma 3.25. Let Λ be a normal modal logic. If $(\mathrm{T}) \in \Lambda$ / $(\mathrm{B}) \in \Lambda$, then $\mathfrak{M}^\Lambda$ is based on a reflexive/symmetric frame.

Proof.

- Let $w \in W^\Lambda$ and $\phi \in w$. Since w is Λ-MCS and $(\mathrm{T}) \in \Lambda$, $\phi \to \Diamond\phi \in w$. By modus ponens we get $\Diamond\phi \in w$. Thus $(w,w) \in R^\Lambda$.

- Let $w, v \in W^\Lambda$, $(w, v) \in R^\Lambda$ and let $\phi \in w$. As $\phi \to \Box\Diamond\phi \in w$, $\phi \in w$ implies by modus ponens that $\Box\Diamond\phi \in w$ which implies by $(w, v) \in R^\Lambda$ and the Truth Lemma that $\Diamond\phi \in v$. By definition of R^Λ, $\phi \in w$ and $\Diamond\phi \in v$ imply $(v, w) \in R^\Lambda$.

□

Theorem 3.26. The logics **T**, **B** and **S4** are complete w.r.t. the class of reflexive, symmetric and reflexive-transitive frames, respectively. The logic **S5** is complete w.r.t. the class of frames (W, R) with R equivalence relation.

Proof. Proceed as in the proof of Theorem 3.23, using Lemma 3.24 and Lemma 3.25. □

Note. There exist natural and complete (normal) modal logics for which the above approach does not work. Such logics are referred to as *non-canonical logics*. One example of a non-canonical logic is *provability logic*, i.e. the logic **L** that is generated by $\{(\mathrm{L})\}$ (where (L) denotes the Löb-axiom $\Box(\Box p \to p) \to \Box p$).

Example 3.27. Coming back to the logic of knowledge and belief, note that for every normal modal logic Λ,

$$\Box p \to p \in \Lambda \text{ iff } p \to \Diamond p \in \Lambda \tag{T}$$
$$\Box p \to \Box\Box p \in \Lambda \text{ iff } \Diamond p \to \Diamond\Diamond p \in \Lambda \tag{4}$$
$$\neg\Box p \to \Box\neg\Box p \in \Lambda \text{ iff } \Diamond p \to \Box\Diamond p \in \Lambda \tag{5}$$

For logics of knowledge and belief, we are thus interested in the class of reflexive and transitive frames, or even in frames based on equivalence relations.

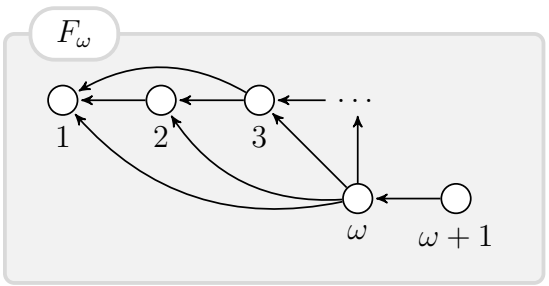
3.4 Incompleteness

In Section 3.1, we have seen that, for every class of frames S, the logic Λ_S is normal. Conversely, is every normal modal logic the logic of some class of frames? Unfortunately this is not the case!

Definition 3.28. Let F_ω be the frame (W, R) with

1. $W = \mathbb{N} \cup \{\omega, \omega + 1\}$
2. $(x, y) \in R$ iff
   (a) $x \neq \omega + 1$ and $y < x$ or
   (b) $x = \omega + 1$ and $y = \omega$

The following diagram depicts F_ω:



Now let S_ω be the class of *models* $\mathfrak{M} = (\mathfrak{F}_\omega, V)$ s.t. for each $p \in \mathcal{A}$, either

1. $V(p)$ is finite and $\omega \notin V(p)$ or
2. $V(p)$ is co-finite (i.e. $W \setminus V(p)$ is finite) and $\omega \in V(p)$.

KvB is the set of formulae that are globally true in all models in S_ω.

Theorem 3.29. KvB is a normal modal logic.

Proof. Closure under modus ponens and necessitation and containment of all propositional tautologies and the K axiom are clear, as these are all sound for satisfaction in *models*. Closure under substitution (which in general is sound only for satisfaction in *frames*) is shown in the exercises. □

Our aim is to show that for all classes of frames S, $\mathbf{KvB} \neq \Lambda_S$. The following formulae play an important role:

$$\vartheta = \Box\Diamond\top \to \Box(\Box(\Box p \to p) \to p)$$
$$\chi = \Box\Diamond\top \to \Box\bot$$

Lemma 3.30. $\vartheta \in \mathbf{KvB}$ and $\chi \notin \mathbf{KvB}$.

Proof. "$\vartheta \in \mathbf{KvB}$": Let $\mathfrak{M} \in S_\omega$; we have to show that $\mathfrak{M} \models \vartheta$. Observe that $\mathfrak{M}, v \models \Box\Diamond\top$ iff $v \in \{0, \omega{+}1\}$. In case $v = 0$, we are done immediately since 0 has no successors. It remains to show that $\mathfrak{M}, \omega + 1 \models \Box(\Box(\Box p \to p) \to p)$, i.e. that $\mathfrak{M}, \omega \models \Box(\Box p \to p) \to p$. So suppose that $\mathfrak{M}, \omega \models \Box(\Box p \to p)$. Then $\mathfrak{M}, n \models \Box p \to p$ for all $n \in \mathbb{N}$. We claim that

$$\mathfrak{M}, n \models p \quad \text{for all } n \in \mathbb{N}. \tag{$*$}$$

From $(*)$, we conclude that $\mathfrak{M}, \omega \models p$, as required: Since $(*)$ implies that $V(p)$ is not finite, it follows from the definition of S_ω that $\omega \in V(p)$.

We prove $(*)$ by course-of-values induction on n, i.e. in the step for n we assume as the inductive hypothesis that the claim already holds for all $k < n$. Since these k are precisely the successors of n, it follows that $\mathfrak{M}, n \models \Box p$, so $\mathfrak{M}, n \models p$ since $\mathfrak{M}, n \models \Box p \to p$.

"$\chi \notin \mathbf{KvB}$": Since ω is the only successor of $\omega + 1$, we have $\mathfrak{M}, \omega + 1 \not\models \chi$ for all $\mathfrak{M} \in S_\omega$. □

Lemma 3.31. For any frame $\mathfrak{F}$, $\models_{\mathfrak{F}} \vartheta$ implies $\models_{\mathfrak{F}} \chi$.

Proof. Let $\mathfrak{F} = (W, R)$ such that $\models_{\mathfrak{F}} \vartheta$. Let $\mathfrak{M} = (\mathfrak{F}, V)$ and $w \in W$. We have to show that $\mathfrak{M}, w \models \chi$. So suppose that $\mathfrak{M}, w \models \Box\Diamond\top$; we have to show that $\mathfrak{M}, w \models \Box\bot$. Assume the contrary, i.e. there is $v \in W$ such that $(w, v) \in R$. Take a new model $\mathfrak{M}' = (\mathfrak{F}, V')$, where $V'(p) = W \setminus \{v\}$ and $V'(q) = V(q)$ for $q \neq p$. Since $\mathfrak{M}, w \models \Box\Diamond\top$, we have $\mathfrak{M}', w \models \Box\Diamond\top$. Thus $\mathfrak{M}', w \models \Box(\Box(\Box p \to p) \to p)$, and hence $\mathfrak{M}', v \models \Box(\Box p \to p) \to p$. Since $\mathfrak{M}', v \not\models p$ it follows that $\mathfrak{M}', v \not\models \Box(\Box p \to p)$. Hence there is $u \in W$ such that $(v, u) \in R$ and $\mathfrak{M}', u \not\models \Box p \to p$, i.e. $\mathfrak{M}', u \models \Box p$ but $\mathfrak{M}', u \not\models p$. Since $V'(p) = W \setminus \{v\}$, the latter implies $u = v$ and thus $(u, u) \in R$. Since $\mathfrak{M}', u \not\models p$, we have a contradiction to $\mathfrak{M}, u \models \Box p$. □

Theorem 3.32. KvB is a normal, consistent modal logic that is *not* complete w.r.t. any class of frames.

Proof.
- **KvB** is normal by Theorem 3.29.

- **KvB** is consistent since S_ω is non-empty and $\mathfrak{M} \not\models \bot$ for every model $\mathfrak{M} \in S_\omega$. Hence $\bot \notin \mathbf{KvB}$.
- Suppose **KvB** is complete w.r.t. a class of frames S. By Lemma 3.31, we have $\models_S \vartheta \to \chi$. By Lemma 3.30 and since modal logics are closed under modus ponens, we have $\nvdash_{\mathbf{KvB}} \vartheta \to \chi$, which is a contradiction to the completeness of **KvB** w.r.t. S.

□

3.5 Frame-definability

We have seen that some modal formulae correspond to first-order properties of *frames*. For instance, the modal formula $\Diamond\Diamond p \to \Diamond p$ corresponds to transitivity of frames, i.e. to the first-order formula $\forall x, y, z.((R(x, y) \land R(y, z) \to R(x, z)))$. Is this the case for every modal formula? No, in general modal formulae correspond to *second-order* properties of frames.

Definition 3.33. Let ϕ be a formula, S a class of frames. We say that ϕ *defines* S if, for all frames $\mathfrak{F}$,

$$\mathfrak{F} \in S \qquad \text{iff} \qquad \models_{\mathfrak{F}} \phi.$$

Example 3.34. We note that

- (T) defines the class of reflexive frames,
- (4) defines the class of transitive frames,
- (L) defines the class of frames (W, R) with R finite transitive tree. We refer to this class as FTT.

Definition 3.35. Let ϕ be a formula, $\mathfrak{F}$ a frame and w a world of $\mathfrak{F}$. We write $\mathfrak{F}, w \models \phi$ if $\mathfrak{M}, w \models \phi$ for all models $\mathfrak{M}$ that are based on $\mathfrak{F}$.

1. Let $\alpha(x)$ be a FOL-formula with one free variable x. Then $\alpha(x)$ *locally corresponds* to a modal formula ϕ if for all frames $\mathfrak{F}$ and worlds w of $\mathfrak{F}$,

$$\mathfrak{F}, w \models \phi \qquad \text{iff} \qquad \mathfrak{F}, \eta \models_{FOL} \alpha, \quad \text{where } \eta = [w/x].$$

2. Let $\beta(x)$ be a second-order formula (MSO). Then $\beta(x)$ *locally corresponds* to ϕ if for all frames $\mathfrak{F}$ and worlds w of $\mathfrak{F}$,

$$\mathfrak{F}, w \models \phi \qquad \text{iff} \qquad \mathfrak{F}, \eta \models_{MSO} \beta, \quad \text{where } \eta = [w/x].$$

Example 3.36.

The axiom (T) (i.e. $p \to \Diamond p$) locally corresponds to $R(x, x)$:

$$\begin{aligned}
\text{Frame } \mathfrak{F} \text{ is reflexive iff } & \models_{\mathfrak{F}} (T) \\
& \text{iff } \forall w \in W. \mathfrak{F}, w \models (T) \\
& \text{iff } \forall w \in W. \mathfrak{F}, \eta \models_{FOL} R(x, x) \text{ for } \eta = [w/x] \\
& \text{iff } \mathfrak{F} \models_{FOL} \forall x. R(x, x)
\end{aligned}$$

The axiom (4) (i.e. $\Diamond\Diamond p \to \Diamond p$) locally corresponds to the FOL-formula

$$\forall y, z.((R(x,y) \wedge R(y,z)) \to R(x,z)).$$

The class of frames (W, R) with R finite transitive tree (FTT) is not first-order definable:

Lemma 3.37. There is no FOL-formula that locally corresponds to (L).

Proof. The proof makes uses of compactness of FOL and is left as exercise. □

Definition 3.38. Let ϕ be a modal formula that mentions n propositional atoms $p_1, \ldots, p_n$. The *second-order translation* of ϕ is given as

$$\forall P_1 \ldots . \forall P_n .\ ST_x(\phi).$$

Proposition 3.39. Let ϕ be a modal formula. Then for all *frames* $\mathfrak{F} = (W, R)$ and all worlds $w \in W$,

1. $\mathfrak{F}, w \models \phi$ iff $\mathfrak{F}, \eta \models_{MSO} \forall P_1 \ldots . \forall P_n .\ ST_x(\phi)$ for $\eta = [w/x]$
2. $\mathfrak{F} \models \phi$ iff $\mathfrak{F} \models_{MSO} \forall P_1 \ldots . \forall P_n . \forall x .\ ST_x(\phi)$.

Proof. Similar to the proof of Proposition 1.6, left as exercise. □

Example 3.40. The second-order translations of (T) and (L) are

(T): $\forall P. ST_x(p \to \Diamond p) = \forall P. P(x) \to \exists y. R(x,y) \wedge P(y)$

(L): $\forall P. ST_x(\Box(\Box p \to p) \to \Box p) =$
$\forall P. \forall y. R(x,y) \to ((\forall z. R(y,z) \to P(z)) \to P(y)) \to (\forall y. R(x,y) \to P(y))$

We now consider a fragment of modal logic that consists of formulae that define first-order properties of frames: The *Sahlqvist fragment* of modal logic.

Definition 3.41. Let ϕ be a formula. A propositional atom p is *positive/negative in* ϕ if every occurrence of p in ϕ is under an even/odd number of negations. A formula ϕ is *positive/negative* if all propositional atoms that occur in ϕ are positive/negative in ϕ.
A *boxed atom* is a formula of the shape $\underbrace{\Box \ldots \Box}_{k \text{ times}} p$ where $k \geq 0$.

Definition 3.42. A *Sahlqvist antecedent* is a formula built from $\bot, \top$, boxed atoms and negative formulae, using $\wedge, \vee$ and $\Diamond$. A *Sahlqvist implication* is an implication $\phi \to \psi$ in which ψ is positive. A *Sahlqvist formula* is a formula that is built up from Sahlqvist implications, using $\Box$ and $\wedge$, and using $\vee$ only between formulae that do not share propositional atoms.

Example 3.43. (T), (4), (B) are Sahlqvist formulae, (L) is not:

(T) $p \to \Diamond p$

(4) $\underbrace{\Diamond\Diamond p}_{\text{Sahlqvist antecedent}} \to \Diamond p$

(L) $\underbrace{\Box(\Box p \to p)}_{\text{not a Sahlqvist antecedent}} \to \Box p$

Theorem 3.44. Let χ be a Sahlqvist formula. Then χ locally corresponds to a *first-order* formula $c_\chi(x)$ on frames. Moreover, c_χ is effectively computable from χ.

Proof. This proof is out of scope of the lecture. □

Chapter 4

Decidability and complexity

Decidability of satisfiability and validity of formulae is, of course, a desirable property of (modal) logics. The following reduction allows us to restrict attention to satisfiability of formulae:

$$\models_S \phi \qquad \text{iff} \qquad \neg\phi \quad \text{is unsatisfiable in } S.$$

Hence decidability of satisfiability implies decidability of validity, and complexity bounds can be transferred.

In this section, we will only consider normal modal logics that are sound and complete w.r.t. some class of frames, namely the classes considered in Section 3.1. The frames from these classes are called Λ-frames, e.g. we will call a frame an **S4**-frame if it is reflexive and transitive. We say that a formula is Λ-satisfiable iff it is satisfiable w.r.t. the class of Λ-frames and write

- SAT(Λ) to denote the satisfiability problem of Λ,
- VAL(Λ) to denote the validity problem of Λ.

4.1 Finite models

Many (but not all) modal logics have the *finite model property* (FMP), which is formulated as follows.

> If a formula is Λ-satisfiable, then it is satisfiable in a *finite* model based on a Λ-frame.

The FMP often allows to prove decidability in a straight-forwarded way.

4.1.1 Filtration

Filtration is one of the most important techniques to convert an infinite model into a finite one.

Definition 4.1. Let $\mathfrak{M} = (W, R, V)$ be a model and Γ a set of formulae. We define a relation $\sim_\Gamma \subseteq W \times W$ by putting $w \sim_\Gamma v$ iff for all $\phi \in \Gamma$, $\mathfrak{M}, w \models \phi$ iff $\mathfrak{M}, v \models \phi$. We write $[w]_\Gamma$ for the equivalence class of w w.r.t. $\sim_\Gamma$, omitting the subscript Γ if understood from the context.

A model $\mathfrak{M}^f = (W^f, R^f, V^f)$ is called a *filtration of* $\mathfrak{M}$ *through* Γ if the following conditions hold:

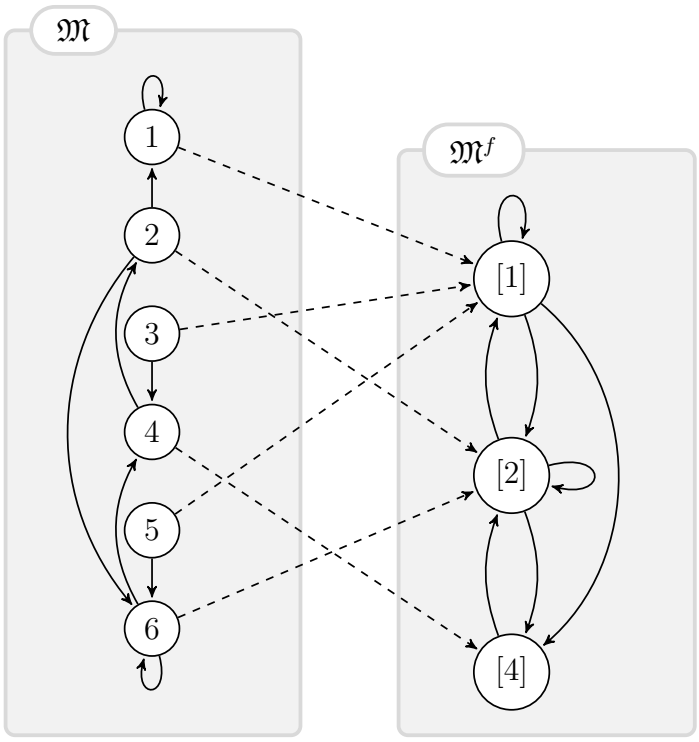
(i) $W^f = \{[w] \mid w \in W\}$,

(ii) for all $w, v \in W$, if $(w, v) \in R$, then $([w], [v]) \in R^f$,

(iii) for all $w, v \in W$ and $\Diamond\phi \in \Gamma$, if $([w], [v]) \in R^f$ and $\mathfrak{M}, v \models \phi$, then $\mathfrak{M}, w \models \Diamond\phi$,

(iv) $V^f(p) = \{[w] \mid w \in V(p)\}$.

Example 4.2. The following diagram depicts a model $\mathfrak{M}$ and a filtration $\mathfrak{M}^f$ of $\mathfrak{M}$ through a set Γ. Without defining Γ explicitly, we assume that we have $1 \sim_\Gamma 3 \sim_\Gamma 5$ and $2 \sim_\Gamma 6$ and the three equivalence classes $[1]_\Gamma$, $[2]_\Gamma$ and $[4]_\Gamma$. The dashed arrows in the diagram indicate the identification of worlds that takes place.



Why is it allowed that there is no loop at [4] in the filtration? Why is it allowed that there is no transition from [4] to [1]? Are all of the existing transitions in the filtration required, or could we omit some of them?

Theorem 4.3. Let Γ be a finite set of formulae that is closed under taking subformulae, let $\mathfrak{M}$ be a model and let $\mathfrak{M}^f$ be a filtration of $\mathfrak{M}$ through Γ. Then

1) $\mathfrak{M}^f$ contains at most $2^{|\Gamma|}$ worlds,

2) for all $\phi \in \Gamma$ and worlds w of $\mathfrak{M}$, we have

$$\mathfrak{M}, w \models \phi \qquad \text{iff} \qquad \mathfrak{M}^f, [w] \models \phi.$$

Proof.

1) Easy by definition of filtration (every world may or may not satisfy any formula from Γ: $2^{|\Gamma|}$ possibilities).

2) By structural induction on ϕ (left as exercise).

□

There is, by definition, no unique filtration of a model, but rather we have some freedom to choose R^f. Let us describe the most prominent choices:

1. The *smallest filtration* is defined by

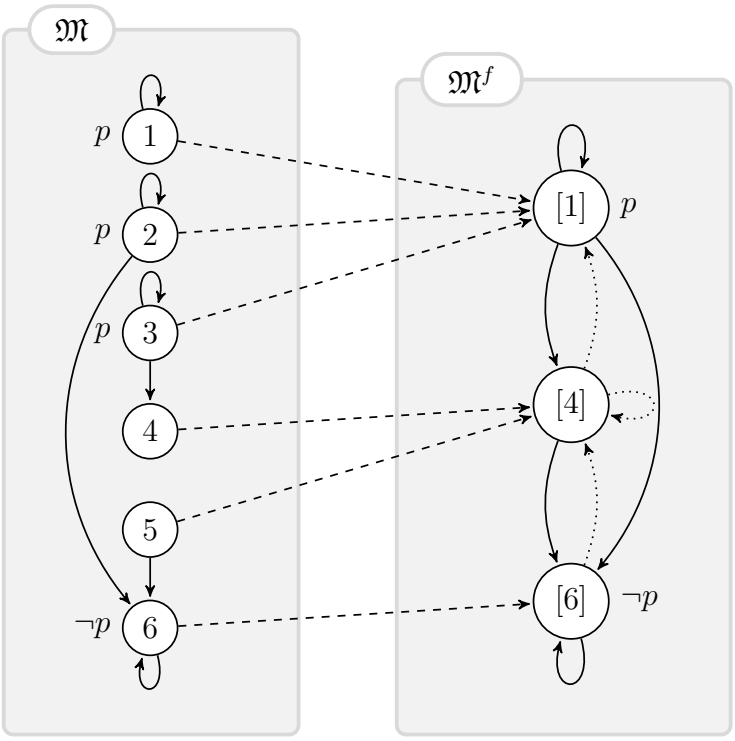
$$R^f := \{([w],[v]) \mid \exists w' \in [w] \text{ and } v' \in [v] \text{ s.t. } (w',v') \in R\}\,.$$

2. The *largest filtration* is defined by

$$R^f := \{([w],[v]) \mid \forall \Diamond\phi \in \Gamma. \mathfrak{M}, v \models \phi \text{ implies } \mathfrak{M}, w \models \Diamond\phi\}\,.$$

It is easily checked (exercise!) that the two choices satisfy the second and third property of filtrations.

Example 4.4. The following diagram depicts a model $\mathfrak{M}$ and two filtrations of $\mathfrak{M}$ through a set Γ. Without defining Γ explicitly, we assume $\sim_\Gamma$ identifies the nodes 1, 2 and 3 as well as the nodes 4 and 5. The dashed arrows in the diagram indicate the identification of worlds that takes place. The atom p is intended to be satisfied only where indicated explicitly; we have emphasized satisfaction of $\neg p$ only where particularly relevant. Two filtrations are depicted by $\mathfrak{M}^f$: The solid arrows in $\mathfrak{M}^f$ constitute (together with the nodes in $\mathfrak{M}^f$) the smallest filtration of $\mathfrak{M}$ through Γ while the model that is obtained by also adding the dotted transitions is the largest filtration of $\mathfrak{M}$ through Γ.



Regarding the largest filtration, we assume that the only $\Diamond$-formula in Γ is $\Diamond p$ so that all the depicted dotted transitions are allowed (also, we assume that Γ is closed under subformulae, in particular contains p). Notice that there is not even a dotted transition from [6] to [1] since $\mathfrak{M}, 1 \models p$ but $\mathfrak{M}, 6 \not\models \Diamond p$,

We now use filtration to prove that **K** has the FMP. Let $|\phi|$ denote the *length* of a formula ϕ, i.e. the number of symbols used to write it.

Theorem 4.5. Every satisfiable formula ϕ has a model containing at most $2^{|\phi|}$ worlds.

Proof. Let ϕ be a satisfiable formula, $\mathfrak{M}$ a model of ϕ and w a world of $\mathfrak{M}$ s.t. $\mathfrak{M}, w \models \phi$. Let Γ be the set of all subformulae of ϕ and let $\mathfrak{M}^f$ be a filtration of $\mathfrak{M}$ through Γ. By Theorem 4.3, $\mathfrak{M}^f$ has at most $2^{|\Gamma|}$ worlds and $\mathfrak{M}^f, [w] \models \phi$. It is easily checked that $|\Gamma| \leq |\phi|$. □

Note. Since Theorem 4.5 gives an explicit bound on the number of worlds, the established model property is also called *bounded model property* (BMP). BMP implies FMP, but not vice-versa!

Theorem 4.5 can be used to prove decidability of SAT(**K**).

Lemma 4.6 (Model checking for **K**)**.** Given a model $\mathfrak{M} = (W, R, V)$, a world $w \in W$ and a formula ϕ, it is decidable in time $\mathcal{O}(|\phi| \cdot (|W| + |R|)$, whether $\mathfrak{M}, w \models \phi$.

Proof. Let $\psi_1, \ldots, \psi_k$ be the subformulae of ϕ, listed in order of length. Then $\psi_k = \phi$ and if ψ_i is subformula of ψ_j, then $i < j$. To decide whether $\mathfrak{M}, w \models \phi$, for each $i = 1, \ldots, k$, label each world $v \in W$ with ψ_i or $\neg\psi_i$, depending on whether $\mathfrak{M}, v \models \psi_i$ or not. Finally, return "yes" if w is labelled with ϕ and "no" otherwise.
The algorithm needs at most $|\phi|$ rounds. In each round, it has to check $|W|$ worlds. Checking formulae $\top, \bot, p, \neg\psi$ and $\psi \wedge \vartheta$ is trivial. To check formulae $\Diamond\psi$, we need to visit all successors of the currently considered world, at most $|R|$. □

Theorem 4.7. SAT(**K**) and VAL(**K**) are decidable.

Proof. By Theorem 4.5, to check satisfiability of a formula ϕ, it suffices to

1) enumerate all models of size up to $2^{|\phi|}$, and

2) for each such model $\mathfrak{M}$, check whether $\mathfrak{M}, w \models \phi$ for some world w of $\mathfrak{M}$ (which is decidable by Lemma 4.6).

Since only propositional letters used in ϕ are relevant, there are only finitely many such models, i.e. the algorithm terminates. □

4.1.2 Selection

Assume that we are interested in the class of frames whose relation is a partial function, henceforth called *functional frames*. This occurs e.g. when modal logics are used for reasoning about programs where functionality corresponds to determinism. We call the logic of this class of frames **KCD**; it is generated by the axiom

$$(CD) \qquad \Diamond p \to \Box p$$

Here, filtration does not work. To establish the FMP for **KCD**, we use a different technique: selection.

Theorem 4.8. If a formula ϕ is **KCD**-satisfiable, then it is satisfiable in a model based on a **KCD**-frame (W, R) such that $|W| \leq |\phi|$.

Proof. Let $\mathfrak{M} = (W, R, V)$ be a model of ϕ. We construct a selection function s that assigns to each world w of $\mathfrak{M}$ and each subformula ψ of ϕ a set $s(\psi, w)$ of worlds; s is

defined by recursion on ψ as follows.

$$\begin{aligned} s(\bot, w) &= \{w\} \\ s(p, w) &= \{w\} \\ s(\neg\psi, w) &= s(\psi, w) \\ s(\psi \wedge \vartheta, w) &= s(\psi, w) \cup s(\vartheta, w) \\ s(\Diamond\psi, w) &= \{w\} \cup \bigcup_{\{v \mid (w,v) \in R\}} s(\psi, v) \end{aligned}$$

Intuitively, $s(\psi, w)$ selects (at most) the worlds that are relevant for evaluating ψ at w. Now let $\mathfrak{M}'$ be the model (W', R', V') defined by

$$\begin{aligned} W' &= s(\phi, w) \\ R' &= R \cap W' \times W' \\ V'(p) &= V(p) \cap W' \quad \text{for all } p \in \mathcal{A}. \end{aligned}$$

It is easily proved (exercise!) by structural induction that for all subformulae ψ of ϕ and all $w \in W'$, if $s(\psi, w) \subseteq W'$, then

$$\mathfrak{M}, w \models \psi \text{ iff } \mathfrak{M}' \models \psi.. \qquad (*)$$

Let $w \in W$ s.t. $\mathfrak{M}, w \models \phi$. By (*), we obtain $\mathfrak{M}', w \models \phi$, noting that $s(\phi, w) \subseteq W'$ by construction. Since (W, R) is functional, it is easy to see that the number of worlds in $s(\phi, w)$ is bounded by the number of diamond-subformulae of ϕ. Hence the number of worlds of $\mathfrak{M}'$ is bounded by $|\phi|$. Finally, R' is obviously functional. □

Corollary 4.9. SAT(**KCD**) and VAL(**KCD**) are decidable.

Proof. Analogous to the proof of Theorem 4.7. □

4.2 NP

For some modal logics, the satisfiability problem is in NP, i.e. not harder than satisfiability of propositional logic.

Theorem 4.10. SAT(**KCD**) is NP-complete.

Proof. "in NP": By Theorem 4.8, it suffices to look for models up to size $|\phi|$, to decide satisfiability of a formula ϕ. Hence a nondeterministic Turing-Machine may

1) "guess" a model of size at most ϕ, and
2) check whether
   a) (W, R) is functional
   b) $\mathfrak{M}, w \models \phi$ for some world $w \in W$.

   Then return "sat" if both conditions are satisfied, "unsat" otherwise.

By Lemma 4.6, the resulting algorithm is in NP.

"NP-hard": Propositional satisfiability can trivially be reduced to satisfiability in **KCD**.

□

The most prominent example of an NP-complete modal logic is **S5**. Recall that **S5** is sound and complete w.r.t. the class of frames based on equivalence relations. However, **S5**is also sound and complete w.r.t. another interesting class of frames: The class of *universal frames* (W, R) with $R = W \times W$.

Theorem 4.11. A uni-modal formula ϕ is satisfiable w.r.t. the class of frames based on equivalence relations iff it is satisfiable w.r.t. the class of universal frames.

Proof. "$\Rightarrow$": Let $\mathfrak{M} = (W, R, V)$ be a model of ϕ with R equivalence relation and let $w \in W$ s.t. $\mathfrak{M}, w \models \phi$. Define a new model $\mathfrak{M}' = (W', R', V')$ by putting

$$\begin{aligned} W' &= \{w' \in W \mid (w, w') \in R\} \\ R' &= R \cap W' \times W' \\ V'(p) &= V(p) \cap W' \quad \text{for all } p \in \mathcal{A}. \end{aligned}$$

Since R is an equivalence relation, R' is universal and $\mathfrak{M}'$ is a generated submodel of $\mathfrak{M}$ (that is, the inclusion $\mathfrak{M}' \hookrightarrow \mathfrak{M}$ is a functional bisimulation). Thus by Proposition 2.8, $\mathfrak{M}', w \models \phi$.

"$\Leftarrow$": This direction is trivial since every universal frame is based on an equivalence relation. □

We can now show that **S5** has a BMP:

Theorem 4.12. A uni-modal formula ϕ is satisfiable w.r.t. the class of universal frames iff ϕ is satisfiable in a model $\mathfrak{M}$ based on a universal frame (W, R) such thath $|W| \leq |\phi|$.

Proof. "$\Rightarrow$": Let $\mathfrak{M} = (W, R, V)$ be a model such that R is universal, and let $w \in W$ such that $\mathfrak{M}, w \models \phi$. Let D_ϕ be the set of all subformulae $\Diamond\psi$ of ϕ s.t. $\mathfrak{M}, w \models \Diamond\psi$. For each $\Diamond\psi \in D_\phi$, fix a world $w_\psi \in W$ such that $\mathfrak{M}, w_\psi \models \psi$. Define a new model $\mathfrak{M}' = (W', R', V')$ (where R' is again universal) as follows:

$$\begin{aligned} W' &= \{w\} \cup \{w_\psi \mid \Diamond\psi \in D_\phi\} \\ R' &= R \cap (W' \times W') = W' \times W' \\ V'(p) &= V(p) \cap W' \quad \text{for all } p \in \mathcal{A}. \end{aligned}$$

It is easily seen that $|W'| \leq \phi$. We show by structural induction that, for all $w' \in W'$ and all subformulae ϑ of ϕ, $\mathfrak{M}, w' \models \vartheta$ iff $\mathfrak{M}', w' \models \vartheta$. We only consider the modal case.

"$\Rightarrow$": Assume $\mathfrak{M}, w' \models \Diamond\psi$. Then $\mathfrak{M}, w \models \Diamond\psi$ since R is universal and thus $\Diamond\psi \in D_\phi$ and $\mathfrak{M}, w_\psi \models \psi$. By the induction hypothesis, $\mathfrak{M}', w_\psi \models \psi$. Since R' is universal, $(w', w_\psi) \in R'$. Thus $\mathfrak{M}', w' \models \Diamond\psi$.

"$\Leftarrow$": Now assume $\mathfrak{M}', w' \models \Diamond\psi$. Then there is a $v \in W'$ such that $(w', v) \in R'$ and $\mathfrak{M}', v \models \psi$. By the induction hypothesis, $\mathfrak{M}, v \models \psi$ and since (W, R) is universal, $(w', v) \in R$. Thus $\mathfrak{M}, w' \models \Diamond\psi$.

"$\Leftarrow$": This direction is again trivial.

□

Corollary 4.13. SAT(**S5**) is NP-complete (and VAL(**S5**) is co-NP-complete).

Proof. Analogous to the proof of Theorem 4.10. □

4.3 PSPACE

The NP-completeness results from the previous section relied on a BMP that provided us with polynomial-sized (in fact: linear-sized) models. In other logics, however, we can enforce much larger models.

4.3.1 Size of Models

In Theorem 4.5 we have seen that for **K**, every satisfiable formula ϕ has a model with at most $2^{|\phi|}$ worlds. The next theorem tells us that we cannot do any better.

Theorem 4.14. For each logic $\Lambda \in \{\mathbf{K}, \mathbf{T}, \mathbf{B}, \mathbf{K4}, \mathbf{S4}\}$ and each $n \in \mathbb{N}$, there is a formula ϕ_n s.t.

1) $|\phi_n| \in \mathcal{O}(n^2)$,

2) ϕ_n is Λ-satisfiable,

3) for all models $\mathfrak{M} = (W, R, V)$ of ϕ_n, we have $|W| \geq 2^n$.

Proof. Idea: Construct formulae ϕ_n whose models are *full binary trees of depth n*. The proof is left as exercise (Sheet 3, Exercise 4). □

4.3.2 PSPACE-Hardness

The prototypical PSPACE-hard (and PSPACE-complete) problem is truth of quantified boolean formulae.

Definition 4.15. A *quantified boolean formula* (QBF) is of the form

$$Q_1 q_1. \ldots . Q_n q_n . \phi(q_1, \ldots, q_n)$$

where $Q_i \in \{\exists, \forall\}$ and ϕ is a propositional formula using only the variables q_1 to q_n.

Truth of QBFs is defined by induction over the length of the quantifier prefix:

Definition 4.16. A QBF $Q_1 q_1. \ldots . Q_n q_n . \phi(q_1, \ldots, q_n)$ is *true* iff

i) $Q_1 = \exists$: $Q_2 q_2. \ldots . Q_n q_n . \phi[\top/q_1]$ or
$Q_2 q_2. \ldots . Q_n q_n . \phi[\bot/q_1]$ is true,

ii) $Q_1 = \forall$: $Q_2 q_2. \ldots . Q_n q_n . \phi[\top/q_1]$ and
$Q_2 q_2. \ldots . Q_n q_n . \phi[\bot/q_1]$ is true.

Example 4.17. Consider the quantified boolean formula $Q = \forall q_1 . \exists q_2 . \forall q_3 . (q_1 \to (q_2 \vee q_3))$. We note that Q is true and has a so-called quantifier tree.

Note. A QBF $Q = Q_1 q_1. \ldots . Q_n q_n . \phi(q_1, \ldots, q_n)$ is true iff it has a quantifier tree s.t.

- level i of the tree corresponds to the quantifier Q_i,
- in $\forall$-levels i, each node has two successors, one for $q_i = \top$ and one for $q_i = \bot$,
- in $\exists$-levels i, each node has one successor: it suffices to explore only one of the possibilities,

- at every leaf, ϕ evaluates to $\top$.

Theorem 4.18. The truth of QBF is PSPACE-complete.

Proof. This proof is out of scope of this lecture, but we note that deciding the truth of a QBF ϕ is equivalent to answering the question whether ϕ has a quantifier tree; quantifier trees in general have exponential size so that an algorithm cannot just "guess" a tree t and verify in polynomial time that t is a quantifier tree for ϕ. □

Theorem 4.19 (Ladner)**.** SAT(Λ) is PSPACE-hard for each $\Lambda \in \{\mathbf{K}, \mathbf{K4}, \mathbf{T}, \mathbf{B}, \mathbf{S4}\}$.

Proof. Idea: For a QBF Q, construct a formula ϕ_Q defining a quantifier tree for Q, s.t. ϕ_Q is satisfiable iff Q is true. The details of the construction are similar to the construction from the proof of Theorem 4.14 and are left as an exercise. □

4.3.3 A PSPACE-algorithm for K

Let us collect some ideas for a PSPACE decision procedure for SAT($\mathbf{K}$).

1) By Proposition 2.13, it suffices to check for the existence of *tree* models.

2) We can assume that the branching width of trees is bounded by the number of $\Diamond\psi$-subformulae of the input formula ϕ (i.e. that the branching width is bounded by $|\phi|$).

3) We will see that it suffices to consider trees whose depth is linear in $|\phi|$.

4) The number of nodes in such trees is exponential, but the number of nodes in each path is polynomial (in fact: linear) in $|\phi|$.

Thus while computing, we keep only paths of the tree in memory. This is achieved by constructing the trees in a depth-first manner.

Let us fix the set of formulae that are "relevant" for a given input:

Definition 4.20. Let ϕ be a formula. We define the *closure* of ϕ as

$$cl(\phi) = \{\psi, \neg\psi \mid \psi \text{ is subformula of } \phi\}$$

and observe that $|cl(\phi)| \leq 2 \cdot |\phi|$.

Definition 4.21. The *tableau-rules* for $\mathbf{K}$ are defined as follows (for all sets of formulae Γ, all formulae $\phi_1, \ldots, \phi_n, \psi, \phi$ and all $p \in \mathcal{A}$):

$$(\bot)\ \frac{\Gamma, \bot}{} \qquad (\lightning)\ \frac{\Gamma, p, \neg p}{} \qquad (\neg\neg)\ \frac{\Gamma, \neg\neg\psi}{\Gamma, \psi}$$

$$(\wedge)\ \frac{\Gamma, \phi \wedge \psi}{\Gamma, \phi, \psi} \qquad (\neg\wedge)\ \frac{\Gamma, \neg(\phi \wedge \psi)}{\Gamma, \neg\phi \qquad \Gamma, \neg\psi}$$

$$(\Diamond_i)\ \frac{\Gamma, \neg\Diamond_i\phi_1, \ldots, \neg\Diamond_i\phi_n, \Diamond_i\phi}{\neg\phi_1, \ldots, \neg\phi_n, \phi}$$

where the $(\Diamond_i)$-rule comes with the side-condition that Γ contains no $\neg\Diamond\phi_j$-formula.

A rule *matches* a set of formulae Λ, if Λ is the premise of the rule.

Algorithm 4.22 (Decide **K**-satisfiability of ϕ)**.**

1. initialize $\Delta = \{\phi\}$.
2. return sat(Δ),

where the procedure sat(Δ) is defined recursively as follows:

```
procedure sat(Δ)
    if Δ = ∅ then return "false"
    for each rule that matches Δ and has conclusion Σ {
            if for each Δ_i ∈ Σ, sat(Δ_i) returns "false" then return "false" }
    return "true"
```

Note: This constructs a tree with nodes that have labels $\Delta \subseteq cl(\phi)$, i.e. trees of size at most $2^{\mathcal{O}(|\phi|)}$.

Theorem 4.23. Algorithm 4.22 returns "true" on input ϕ iff ϕ is **K**-satisfiable; furthermore, the algorithm can be implemented to run in exponential time but using only polynomial space.

Proof. The correctness proof consists of two directions:

"$\Leftarrow$": Let ϕ be **K**-satisfiable, i.e. let ϕ have a (tree) model $\mathfrak{M} = (W, R, V)$ with $w \in W$ s.t. $\mathfrak{M}, w \models \phi$. We show that for all $\Delta \subseteq cl(\phi)$ and all $v \in W$, if $\Delta \subseteq T_{\mathfrak{M}(v)}$, then sat($\Delta$) returns "true". Let $m(\Delta)$ denote the number of diamond-operators in Δ and let $pr(\Delta)$ denote the number of propositional operators in Δ. The proof is by lexicographic induction over $(m(\Delta), pr(\Delta))$, considering all rules that match Δ. We have to show that each such rule has a conclusion Δ_i s.t. sat(Δ_i) returns "true".

We consider just the base case and two cases for the inductive step:

Base case: $m(\Delta) = 0$ and $pr(\Delta) = 0$. Then Δ consists only of propositional atoms and Box-formulae. Since $\Delta \subseteq T_{\mathfrak{M}(v)}$ and there is no $p \in \mathcal{A}$ s.t. $\mathfrak{M}, v \models p \wedge \neg p$, we know that for each $p \in \mathcal{A}$, Δ does not contain p and $\neg p$ at the same time. Thus no rule matches Δ and sat(Δ) returns "true".

Inductive step: If the ($\wedge$)-rule matches Δ, then $\Delta = \Gamma, \phi \wedge \psi$ for some Γ, ϕ and ψ and the rule has the conclusion Γ, ϕ, ψ. As $m(\Gamma, \phi, \psi) = m(\Delta)$ and $pr(\Gamma, \phi, \psi) = pr(\Delta) - 1$ and as $\Gamma, \phi, \psi \subseteq T_{\mathfrak{M}(v)}$ follows from $\Delta \subseteq T_{\mathfrak{M}(v)}$, the induction hypothesis implies that sat(Γ, ψ, ϕ) returns "true".

If the ($\Diamond_i$)-rule matches Δ, then $\Delta = \Gamma, \neg\Diamond_i\phi_1, \ldots, \neg\Diamond_i\phi_n, \Diamond_i\psi$ for some number n and some $\Gamma, \phi_1, \ldots, \phi_n$ and ψ and the rule has the conclusion $\{\neg\phi_1, \ldots, \neg\phi_n, \psi\}$. Then $m(\{\neg\phi_1, \ldots, \neg\phi_n, \psi\}) < m(\Delta)$ and as $\Delta \subseteq T_{\mathfrak{M}(v)}$, $\mathfrak{M}, v \models \Diamond_i\psi$ and $\mathfrak{M}, v \not\models \Diamond_i\phi_i$, i.e. $\mathfrak{M}, v \models \Box_i\neg\phi_i$ for all $1 \leq i \leq n$ so that there is a world $u \in R(v)$ with $\mathfrak{M}, u \models \neg\phi_1 \wedge \cdots \wedge \neg\phi_n \wedge \psi$, i.e. the induction hypothesis again implies that sat($\{\neg\phi_1, \ldots, \neg\phi_n, \psi\}$) returns "true".

"$\Rightarrow$": It is easily possible to extract a model for ϕ from the tableau (i.e. the tree) that is constructed by a successful run of the algorithm. The details of the model extraction are left as an exercise to the reader.

PSPACE: To see that the algorithm uses only a polynomial amount space, notice the recursive style in which the algorithm is formulated. Using a depth-first approach, the algorithm maintains a recursion stack that contains the path in the tableau that leads from the root of the tableau to the current node. Such paths have length at most $|\phi|$. Each node has size at most $|\phi|$. Thus the algorithm can be implemented to run in space $\mathcal{O}(|\phi|^2)$ (and in time $2^{\mathcal{O}(n)}$), showing that the problem of **K**-satisfiability is in PSPACE.

□

As a corollary of Theorem 4.19 and Theorem 4.23, we obtain the following:

Theorem 4.24. SAT(**K**) and VAL(**K**) are PSPACE-complete.

Note. We can adapt the algorithm to obtain PSPACE upper bounds for e.g. **K4**, **T**, **B** and **S4**.

Chapter 5

Propositional Dynamic Logic

Programs $\alpha, \beta, \ldots$ and (PDL) *formulae* $\phi, \psi, \ldots$ are defined by mutual recursion:

$$\alpha, \beta ::= a \mid \phi? \mid \alpha \cup \beta \mid \alpha; \beta \mid \alpha^*$$
$$\phi, \psi ::= \bot \mid p \mid \neg\phi \mid \phi \wedge \psi \mid [\alpha]\phi$$

where p ranges over propositional letters as usual, and a ranges over *atomic* (or *basic*) *programs*. *Regular PDL* is the language without the test construct $\phi?$.

Semantics Given relations $R_a \subseteq W \times W$, programs α are interpreted as relations $R_\alpha \subseteq W \times W$ and formulae are interpreted by a satisfaction relation as usual. These data are defined by mutual recursion by the clauses

$$R_{\phi?} = \{(w, w) \mid w \models \phi$$
$$R_{\alpha\cup\beta} = R_\alpha \cup R_\beta$$
$$R_{\alpha;\beta} = R_\alpha; R_\beta$$
$$R_{\alpha^*} = (R_\alpha)^*.$$

for programs, and the usual clauses for satisfaction of formulae, with $[\alpha]$ being a box modality for the relation R_α. Models where the relations are defined according to the above clauses are called *regular* PDL-models/frames; we will have occasion to consider non-regular models.

We are mostly interested only in regular PDL, except in exercises and examples.

Axioms Normal modal logic generated by

$$[\phi?]\psi \leftrightarrow (\phi \to \psi)$$
$$[\alpha \cup \beta]\psi \leftrightarrow ([\alpha]\psi \wedge [\beta]\psi)$$
$$[\alpha; \beta]\psi \leftrightarrow [\alpha][\beta]\psi$$
$$[\alpha^*]\psi \leftrightarrow (\psi \wedge [\alpha][\alpha^*]\psi)$$
$$[\alpha^*](p \to [\alpha]p) \to p \to [\alpha^*]p$$

The last axioms is Segerberg's induction axiom; soundness is shown in the exercises. Exercise: In regular PDL, the last four axioms define the regular PDL-frames.

Fischer-Ladner closure Restrict to regular PDL from now on.

A set Σ of formulae is *Fischer-Ladner closed* if it is closed under subformulae, and moreover

1. if $[\alpha;\beta]\psi \in \Sigma$, then $[\alpha][\beta]\psi \in \Sigma$
2. if $[\alpha \cup \beta]\psi \in \Sigma$, then $[\alpha]\psi, [\beta]\psi \in \Sigma$
3. if $[\alpha^*]\psi \in \Sigma$, then $[\alpha][\alpha^*]\psi \in \Sigma$.

Write $\mathsf{FL}(\phi)$ for the least Fischer-Ladner closed set containing ϕ. $\mathsf{FL}(\phi)$ is finite, and of linear cardinality in $|\phi|$ (see exercises).
We write $\neg\mathsf{FL}(\phi)$ for the closure of $\mathsf{FL}(\phi)$ under *single negation*; that is, $\neg\mathsf{FL}(\phi)$ consists of all formulae in $\mathsf{FL}(\phi)$ and additionally all formulae $\neg\psi$ where $\psi \in \mathsf{FL}(\phi)$ is not of the form $\neg\psi'$. It is easy to see that $\neg\mathsf{FL}(\phi)$ is still Fischer-Ladner closed, and indeed closed under sigle negations.

Definition 5.1. An *atom* is a maximally consistent subset of $\neg\mathsf{FL}(\phi)$.

Lemma 5.2 (Lindenbaum)**.** *Every consistent subset of $\neg\mathsf{FL}(\phi)$ is contained in an atom.*

Proof?

Lemma 5.3 (Hintikka properties)**.** *Let A be an atom. Then*

1. *if $\neg\psi \in \neg\mathsf{FL}(\phi)$, then $\neg\psi \in A$ iff $A \notin A$*
2. *if $\psi \wedge \chi \in (\neg)\mathsf{FL}(\phi)$, then $\psi \wedge \chi \in A$ iff $\psi, \chi \in A$*
3. *if $[\alpha;\beta]\psi \in (\neg)\mathsf{FL}(\phi)$, then $[\alpha;\beta]\psi \in A$ iff $[\alpha][\beta]\psi \in A$*
4. *if $[\alpha \cup \beta]\psi \in (\neg)\mathsf{FL}(\phi)$, then $[\alpha \cup \beta]\psi \in A$ iff $[\alpha]\psi, [\beta]\psi \in A$*
5. *if $[\alpha^*]\psi \in (\neg)\mathsf{FL}(\phi)$, then $[\alpha^*]\psi \in A$ iff $\phi, [\alpha][\alpha^*]\psi \in A$.*

$\mathsf{At}(\phi)$ = set of all atoms in $\mathsf{FL}(\phi)$. Note we can describe a set $\mathcal{D} \subseteq \mathsf{At}(\phi)$ by the formula

$$\delta = \bigvee_{D \in \mathcal{D}} \hat{D}$$

The following lemma is effectively needed in the proof of the inclusion $S_\alpha \subseteq R_\alpha$ as stated later, see handwritten notes on StudOn:

Lemma 5.4. *1. For $A \neq B \in \mathsf{At}(\phi)$, $\hat{A} \wedge \hat{B}$ is inconsistent (i.e. $\vdash \neg(\hat{A} \wedge \hat{B})$).*

2. $\mathsf{PDL} \vdash \bigvee_{A \in \mathsf{At}(\phi)} \hat{A}$
3. *For $\mathcal{D} \subseteq \mathsf{At}(\phi)$ and $\delta = \bigvee_{D \in \mathcal{D}} \hat{D}$,*

$$\mathsf{PDL} \vdash \neg\delta \leftrightarrow \bigvee_{E \in \mathsf{At}(\phi) \setminus \mathcal{D}} \hat{E}.$$

Proof. 1. Immediate from the Hintikka property for negation.

2. The formula

$$\bigwedge_{\neg\phi \in \neg\mathsf{FL}(\phi)} \phi \vee \neg\phi$$

is clearly derivable. Taking its DNF and removing the PDL-inconsistent conjunctive clauses yields PDL-derivability of $\bigvee_{A \in \mathsf{At}(\phi)} \hat{A}$.

3. Immediate from the first two claims.

□

Small canonical model $\mathfrak{M}^\phi$:

1. Worlds: $\mathsf{At}(\phi)$
2. Valuation: $V^\phi(p) = \{A \in \mathsf{At}(\phi) \mid p \in A\}$
3. Relations: S^ϕ_α:
$$A\, S^\phi_\alpha\, B \quad \text{iff} \quad \hat{A} \wedge \langle\alpha\rangle \hat{B} \text{ consistent}$$
where $\hat{A} = \bigwedge_{\psi \in A} \psi$ for $A \in \mathsf{At}$. (In particular, if $\langle\alpha\rangle\psi \in \mathsf{FL}(\phi)$, then $A\, S^\phi_\alpha\, B$ and $\psi \in B$ imply $\langle\alpha\rangle\psi \in A$; why?)

This model in general fails to be regular, so we generate a regular model from it, with relations R^ϕ_α:

$$\begin{aligned} R^\phi_a &:= S^\phi_a \\ R^\phi_{\alpha;\beta} &:= R^\phi_\alpha ; R^\phi_\beta \\ R^\phi_{\alpha\cup\beta} &:= R_\alpha\phi \cup R^\phi_\beta \\ R^\phi_{\alpha^*} &= (R^\phi_\alpha)^* \end{aligned}$$

- Existence lemma for small canonical model; needs to be proved by selection process.
- Lemma: $S_{\alpha^*} \subseteq (S_\alpha)^*$ (Proof: Express reachable atoms by finite formula δ, prove $\delta \to [\alpha^*]\delta$ by Segerberg)
- Conclude inclusions $S_\alpha \subseteq R_\alpha$ by induction.
  - Sequential composition $\alpha;\beta$: Have $\hat{A} \wedge \langle\alpha\rangle\langle\beta\rangle\hat{B}$ consistent. Define formula δ describing atoms C s.t. $\hat{C} \wedge \langle\beta\rangle\hat{B}$ consistent. Suffices $\hat{A} \wedge \langle\alpha\rangle\delta$ consistent, else $\hat{A} \to [\alpha]\neg\delta$, then $\neg\delta \wedge \langle\beta\rangle\hat{B}$ consistent, contradiction.
- Conclude existence lemma for R_α.
- Truth lemma: Remains to prove "easy" direction for diamonds; induction on programs, using unfolding and induction on paths in the step for $*$.

Chapter 6

Modal fixpoint logics

Many modal logics can only express properties that talk about a fixed number of worlds. By adding fixpoint operators to modal logic, one obtains modal fixpoint logics; such logics often are more expressive than the respective basic modal logic, e.g. they may allow to express the situation that a property holds after *any* number of transitions.

For instance let p be a propositional atom and consider a model $\mathfrak{M}$ with root w and, for each $n \in \mathbb{N}$, exactly one path that starts at w, that reaches – after n transition steps – a node v_n satisfying p, and that loops through v_n afterwards. Also consider models $\mathfrak{M}^i$, for each $i \in \mathbb{N}$, where $\mathfrak{M}^i$ is as $\mathfrak{M}$ with the only difference being that p is not satisfied at v_i on the i-th path in the model $\mathfrak{M}^i$. Can we distinguish $\mathfrak{M}$ from all $\mathfrak{M}^i$ by use of a single formula? No: we cannot distinguish infinitely many worlds using finite formulae. E.g. for the formula

$$\vartheta = \underbrace{p \vee \Box(p \vee \Box(\ldots(p \vee \Box(}_{j \text{ times}} p \ \vee \bot)\ldots))),$$

we have $\mathfrak{M}, w \models \vartheta$ and $\mathfrak{M}^{j+1}, w \models \vartheta$. Adding a fixpoint operator $AF\ p$ (with the intuitive meaning that "on all paths, p holds finally") to the basic modal language, we will obtain that

$$\mathfrak{M}, w \models AF\ p \text{ but } \mathfrak{M}^i, w \not\models AF\ p \text{ for all } i \in \mathbb{N}.$$

6.1 Computation Tree Logic (CTL)

Definition 6.1. The syntax of CTL is defined by the following grammar:

$$\psi, \phi := p \mid \top \mid \neg\psi \mid \psi \wedge \phi \mid \Diamond\phi \mid A(\phi U \psi) \mid E(\phi U \psi)$$

where $p \in \mathcal{A}$.

The formula $A(\phi U \psi)$ comes with the intuition that "on all paths, ϕ holds until eventually ψ holds" while the formula $E(\phi U \psi)$ comes with the intuition that "there is a path on which ϕ holds until eventually ψ holds".

CTL-formulae are interpreted over *serial* models (i.e. models (W, R, V) with the property that $\forall x.\exists y.R(x, y)$).

Definition 6.2. Let $M = (W, R, V)$ be a serial model. The *satisfaction relation* $\models_{CTL}$ that relates worlds of $\mathfrak{M}$ and CTL-formulae (where we omit the subscript, if no confusing

arises) is defined as follows (for all $w \in W$ and all $p \in \mathcal{A}$):

$$\begin{aligned}
&\mathfrak{M}, w \models \top \\
&\mathfrak{M}, w \models p \text{ iff } w \in V(p) \\
&\mathfrak{M}, w \models \neg\phi \text{ iff } \mathfrak{M}, w \not\models \phi \\
&\mathfrak{M}, w \models \phi \wedge \psi \text{ iff } \mathfrak{M}, w \models \phi \text{ and } \mathfrak{M}, w \models \psi \\
&\mathfrak{M}, w \models \Diamond\phi \text{ iff there is a } w' \in W \text{ s.t. } (w, w') \in R_i \text{ and } \mathfrak{M}, w' \models \phi. \\
&\mathfrak{M}, w \models A(\phi U \psi) \text{ iff } \textit{for all } R\text{-paths } w_1 R w_2 R \ldots \text{ with } w_1 = w, \\
&\qquad\qquad \exists i. \mathfrak{M}, w_i \models \psi \text{ and } \forall j < i, \mathfrak{M}. w_j \models \phi. \\
&\mathfrak{M}, w \models E(\phi U \psi) \text{ iff } \textit{there is} \text{ an } R\text{-path } w_1 R w_2 R \ldots \text{ with } w_1 = w, \\
&\qquad\qquad \exists i. \mathfrak{M}, w_i \models \psi \text{ and } \forall j < i, \mathfrak{M}. w_j \models \phi.
\end{aligned}$$

We abbreviate:

$$\begin{array}{lll}
AF\ \phi = A(\top U \phi) & AG\ \phi = \neg EF \neg\phi & A(\phi R\ \psi) = \neg E(\neg\phi U \neg\psi) \\
EF\ \phi = E(\top U \phi) & EG\ \phi = \neg AF \neg\phi & E(\phi R\ \psi) = \neg A(\neg\phi U \neg\psi),
\end{array}$$

Here, the formula $AF\ \phi$ comes with the intuition that "on all paths, ϕ holds *finally*", $AG\ \phi$ comes with the intuition that "on all paths, ϕ holds *globally*" and the *release*-formula $A(\phi R\ \psi)$ has the intuition that "on all paths, ψ holds globally unless ϕ holds eventually".

We observe that CTL-formulae can express *safety* properties (e.g. $AG\ \neg$*deadlock*) as well as *liveness* properties (e.g. AF *finish_computation*).

6.2 Fixpoints

To be able to define the semantics of the μ-calculus, we first have to introduce several basic concepts from the theory of fixpoints.

Definition 6.3. Let X be a set and let $f : \mathcal{P}(X) \to \mathcal{P}(X)$. Then f is called *monotone w.r.t. set inclusion* if for all $A, B \subseteq X$,

$$A \subseteq B \text{ implies } f(A) \subseteq f(B).$$

A set $A \subseteq X$ is a *prefixpoint* of f if $f(A) \subseteq A$ and a *postfixpoint* of f if $A \subseteq f(A)$. The set A is a *fixpoint* of f if $f(A) = A$. We denote the sets of prefixpoints, postfixpoints and fixpoints of f by $PRE(f)$, $POST(f)$ and $FIX(f)$, respectively. If $FIX(f)$ has a least (greatest) element, this element if denoted as $LFP(f)$ ($GFP(f)$).

Theorem 6.4 (Knaster-Tarski)**.** Let $f : \mathcal{P}(X) \to \mathcal{P}(X)$ be monotone w.r.t. set inclusion. Then f has both a least and a greatest fixpoint and these are given as

$$\begin{aligned}
LFP(f) &= \bigcap PRE(f) \\
GFP(f) &= \bigcup POST(f).
\end{aligned}$$

Proof. We consider the least fixpoint case and note that the proof of the greatest fixpoint case is analogous. Put $Q := \bigcap PRE(f)$ so that we have $Q \subseteq Z$ for all $Z \in PRE(f)$. By monotonicity of f, we have $f(Q) \subseteq f(Z)$ for all $Z \in PRE(f)$ and hence by definition of prefixpoints $f(Q) \subseteq Z$ for all $Z \in PRE(f)$. Thus $f(Q) \subseteq Q$, i.e. Q is a prefixpoint of f. It remains to show that Q is also a postfixpoint of f, i.e. that $Q \subseteq f(Q)$. Since $Q \subseteq Z$ for all $Z \in PRE(f)$, it suffices to show that $f(Q) \in PRE(f)$, i.e. that $f(f(Q)) \subseteq f(Q)$. But this follows by monotonicity from $f(Q) \subseteq Q$. □

Note. For any set X and any function $g : \mathcal{P}(X) \to \mathcal{P}(X)$, put

$$g^0(X) = X$$
$$g^{n+1}(X) = g(g^n(X)).$$

For a *finite* set X and a function $f : \mathcal{P}(X) \to \mathcal{P}(X)$ that is monotone w.r.t. set inclusion, we use Kleene's fixpoint theorem to obtain the following alternative characterization of the least and greatest fixpoints:

$$LFP(f) = f^n(\emptyset)$$
$$GFP(f) = f^m(X),$$

where n is the first number s.t. $f^n(\emptyset) = f^{n+1}(\emptyset)$, where m is the first number s.t. $f^m(X) = f^{m+1}(X)$, and where $n \leq |X|$ and $m \leq |X|$.

Definition 6.5. Let X be a set. The *complement* of some set $A \subseteq X$ is defined as

$$\overline{A} = X \setminus A.$$

Let $f : \mathcal{P}(X) \to \mathcal{P}(X)$. A function $g : \mathcal{P}(X) \to \mathcal{P}(X)$ is *complementary to* f, if for all $A \subseteq X$,

$$f(A) = \overline{g(\overline{A})}.$$

Notice that for all $A, B \subseteq X$, we have that $\overline{\overline{A}} = A$, that $A = B$ implies $\overline{A} = \overline{B}$ and that $A \subseteq B$ implies $\overline{A} \supseteq \overline{B}$.

Least and greatest fixpoints are dual concepts in the following sense:

Lemma 6.6. Let $f : \mathcal{P}(X) \to \mathcal{P}(X)$ be monotone w.r.t. set inclusion and let g be complementary to f. Then

$$LFP(f) = \overline{GFP(g)}$$
$$GFP(f) = \overline{LFP(g)}.$$

Proof. First note that since f is monotone and g is complementary to f, g is monotone as well. We consider only the least fixpoint case, the proof of the greatest fixpoint case is analogous. We show that $\overline{GFP(g)}$ is a fixpoint of f that is contained in each prefixpoint of f. Since g is complementary to f and since $GFP(g)$ is a fixpoint of g, we have

$$f(\overline{GFP(g)}) = \overline{g(\overline{\overline{GFP(g)}})} = \overline{g(GFP(g))} = \overline{GFP(g)},$$

which shows that $\overline{GFP(g)}$ is a fixpoint of f. Now let Z be a prefixpoint of f so that $f(Z) \subseteq Z$. Then

$$g(\overline{Z}) = \overline{f(\overline{\overline{Z}})} = \overline{f(Z)} \supseteq \overline{Z},$$

i.e. $\overline{Z}$ is a postfixpoint of g so that $\overline{Z} \subseteq GFP(g)$. But then $Z \supseteq \overline{GFP(g)}$, as required. $\square$

6.3 The modal μ-calculus

The μ-calculus extends CTL by providing general fixpoint operators.

Definition 6.7. Let *Var* be a set of *fixpoint variables*. The syntax of the modal μ-calculus is defined as follows:

$$\psi, \phi \ := \top \mid \bot \mid p \mid \neg p \mid \psi \wedge \phi \mid \psi \vee \phi \mid \Diamond_i \psi \mid \Box_i \psi \mid X \mid \mu X.\psi \mid \nu X.\psi,$$

where $p \in \mathcal{A}$, $i \in I$, $X \in \mathit{Var}$. For $\mu X.\ \psi$ and $\nu X.\ \psi$, X is *bound* in ψ and we require that all formulae are *closed*, that is, we require that formulae do not contain variables that are not bound by an enclosing fixpoint operator.

Definition 6.8. Let $\mathfrak{M} = (W, (R_i)_{i \in I}, V)$ be a model, let $i \in I$ and let ϵ denote the empty substitution. The *extension* (or *truth-set*) of a μ-calculus formula ϕ in $\mathfrak{M}$ is defined as $[\![\phi]\!]_\epsilon$, where for all substitutions $\sigma : \mathit{Var} \to \mathcal{P}(W)$ and all μ-calculus formulae ϕ, ψ,

$$\begin{aligned}
[\![X]\!]_\sigma &= \sigma(X) \\
[\![\top]\!]_\sigma &= W \\
[\![\bot]\!]_\sigma &= \emptyset \\
[\![p]\!]_\sigma &= V(p) \\
[\![\neg p]\!]_\sigma &= W \setminus V(p) \\
[\![\psi \wedge \phi]\!]_\sigma &= [\![\psi]\!]_\sigma \cap [\![\psi]\!]_\sigma \\
[\![\psi \vee \phi]\!]_\sigma &= [\![\psi]\!]_\sigma \cup [\![\psi]\!]_\sigma \\
[\![\Diamond_i \psi]\!]_\sigma &= \{w \in W \mid \exists v \in R_i(w).v \in [\![\psi]\!]_\sigma\} \\
[\![\Box_i \psi]\!]_\sigma &= \{w \in W \mid \forall v \in R_i(w).v \in [\![\psi]\!]_\sigma\} \\
[\![\mu X.\ \psi]\!]_\sigma &= \mathit{LFP}([\![\psi]\!]_\sigma^X) \\
[\![\nu X.\ \psi]\!]_\sigma &= \mathit{GFP}([\![\psi]\!]_\sigma^X),
\end{aligned}$$

and where for all $A \subseteq W$,

$$[\![\psi]\!]_\sigma^X(A) = [\![\psi]\!]_{\sigma[A/X]}.$$

The syntax in Definition 6.7 does not contain explicit negation so that X is by definition positive in ψ; thus the function $[\![\psi]\!]_\sigma^X$ is monotone w.r.t. set inclusion and hence has a least and a greatest fixpoint so that the above indeed constitutes a definition.

We write $\mathfrak{M}, w \models \phi$ if $w \in [\![\phi]\!]_\epsilon$ in $\mathfrak{M}$.

We can define the negation $\neg\phi$ of a μ-calculus formula ϕ as follows:

$$\begin{aligned}
\neg X &= X \\
\neg \top &= \bot \\
\neg \bot &= \top \\
\neg(p) &= \neg p \\
\neg(\neg p) &= p \\
\neg(\phi \wedge \psi) &= \neg\phi \vee \neg\psi \\
\neg(\phi \vee \psi) &= \neg\phi \wedge \neg\psi \\
\neg(\Diamond_i \psi) &= \Box_i \neg\psi \\
\neg(\Box_i \psi) &= \Diamond_i \neg\psi \\
\neg(\mu X.\ \psi) &= \nu X.\ \neg\psi \\
\neg(\nu X.\ \psi) &= \mu X.\ \neg\psi
\end{aligned}$$

Notice that we define $\neg X = X$ so that in e.g. $\neg(\mu X.\ \psi) = \nu X.\ \neg\psi$, X is positive in both ψ and $\neg\psi$.

Lemma 6.9. For all μ-calculus formulae ψ,

$$[\![\psi]\!]_\sigma = \overline{[\![\neg\psi]\!]_\sigma}.$$

Proof. The proof is by induction over ψ, using Lemma 6.6 to obtain

$$[\![\mu X.\ \psi]\!]_\sigma = \overline{[\![\nu X.\ \neg\psi]\!]_\sigma} \text{ and}$$
$$[\![\nu X.\ \psi]\!]_\sigma = \overline{[\![\mu X.\ \neg\psi]\!]_\sigma}.$$

The details of the proof are left as exercise to the reader. □

Example 6.10. Consider the μ-calculus formula $\mu X.\ p \vee \Box X$. As $LFP(f) = f(LFP(f))$, we have

$$\begin{aligned}
[\![\mu X.\ p \vee \Box X]\!]_\epsilon &= LFP[\![p \vee \Box X]\!]_\epsilon^X \\
&= [\![p \vee \Box X]\!]_\epsilon^X(LFP[\![p \vee \Box X]\!]_\epsilon^X) \\
&= [\![p \vee \Box X]\!]_\epsilon^X([\![\mu X.\ p \vee \Box X]\!]_\epsilon) \\
&= [\![p \vee \Box(\mu X.\ p \vee \Box X)]\!]_\epsilon \\
&= [\![(p \vee \Box X)[(\mu X.\ p \vee \Box X)/X]]\!]_\epsilon.
\end{aligned}$$

The above derivation shows that unfolding of fixpoint formulae does not change their semantics. We can iterate this process and obtain for all $j \in \mathbb{N}$ that

$$[\![\mu X.\ p \vee \Box X]\!]_\epsilon = [\![\underbrace{p \vee \Box(p \vee \Box(\ldots(p \vee \Box(}_{j \text{ times}}\ \mu X.\ p \vee \Box X))\ldots))]\!]_\epsilon.$$

If we restrict the models $\mathfrak{M}$ and $\mathfrak{M}^i$ from the beginning of this chapter to have a fixed number of paths n, then we obtain by Kleene's fixpoint theorem that

$$\begin{aligned}
[\![\mu X.\ p \vee \Box X]\!]_\epsilon &= ([\![p \vee \Box X]\!]_\epsilon^X)^m(\emptyset) \\
&= [\![\underbrace{p \vee \Box(p \vee \Box(\ldots(p \vee \Box(}_{m \text{ times}}\ \bot))\ldots))]\!]_\epsilon
\end{aligned}$$

for some finite number m so that $w \in [\![\mu X.\ p \vee \Box X]\!]_\epsilon$ in $\mathfrak{M}$ but $w \notin [\![\mu X.\ p \vee \Box X]\!]_\epsilon$ in $\mathfrak{M}^i$.

As the previous example suggests, CTL is just a fragment of the μ-calculus:

Definition 6.11. The *embedding function* e that maps CTL-formulae to μ-calculus formulae is defined inductively as follows:

$$\begin{aligned}
e(\top) &= \top \\
e(\neg\psi) &= \neg e(\psi) \\
e(p) &= p \\
e(\psi \wedge \phi) &= e(\psi) \wedge e(\phi) \\
e(\Diamond\psi) &= \Diamond e(\psi) \\
e(A(\psi U \phi)) &= \mu X.\ e(\phi) \vee (e(\psi) \wedge \Box X) \\
e(E(\psi U \phi)) &= \mu X.\ e(\phi) \vee (e(\psi) \wedge \Diamond X)
\end{aligned}$$

Notice how CTL-formulae translate to μ-calculus formulae that use only the single fixpoint variable X. Thus CTL is even a fragment of the single-variable μ-calculus.

Lemma 6.12. Let $\mathfrak{M}$ be a serial model and let w be a world of $\mathfrak{M}$. Then for all CTL-formulae ψ, we have

$$\mathfrak{M}, w \models \psi \text{ iff } w \in [\![e(\psi)]\!]_\epsilon \text{ in } \mathfrak{M}.$$

Proof. The proof is by induction over ψ, where the fixpoint operator cases demand particular attention. The details of the proof are left as exercise. □

6.4 Model checking for fixpoint logics

Recall that the model checking problem consists in deciding, whether $\mathfrak{M}, w \models \phi$, for a given model $\mathfrak{M}$, a given world $w \in \mathfrak{M}$ and a given formula ϕ. Model checking for fixpoint logics is considerably more involved than model checking for basic modal logics such as **K**. In fact, we will see that the problem of model checking for the μ-calculus is equivalent to the problem of solving so-called parity games.

Definition 6.13 (Parity games)**.** A *game arena* (V, E) consists of a set of *nodes* $V = V_\exists \uplus V_\forall$, where each node belongs to either ∃loise ($V_\exists$) or ∀belard ($V_\forall$) and a set of *edges* $E \subseteq V \times V$. The edges encode the allowed moves and can be seen as the rules of the game arena. A *parity game* $G = (V, E, \alpha)$ consists of a game arena (V, E) and a *priority function* $\alpha : V \to \mathbb{N}$ that assigns *priorities* $\alpha(v)$ to nodes $v \in V$. A *play* $\rho = v_0 v_1 v_2 \dots$ is a (possibly infinite) sequence of nodes $v_i \in V$ that adheres to the rules of the game in the sense that for all $i \geq 0$, we have $v_i E v_{i+1}$. A (history-free) *strategy* for ∃loise (∀belard) is a function $s : V_\exists \to V$ ($s : V_\forall \to V$) that assigns a move $s(v)$ to each node $v \in V_\exists$ ($v \in V_\forall$). A play ρ *adheres* to a strategy s, if for all $j \geq 0$, $\rho_j \in V_\exists$ ($\rho_j \in V_\forall$) implies that $\rho_{j+1} = s(\rho_j)$. ∃loise (∀belard) *wins* an infinite play ρ, if the highest priority that occurs infinitely often in ρ is even (odd), or more formally, if $\max\{\inf(\alpha \circ \rho)\}$ is even (odd). A finite play ending in a node $v \in V_\exists$ ($v \in V_\forall$) is won by ∀belard (∃loise). A *winning strategy* at $v \in V$ for ∃loise (∀belard) is a strategy s s.t. ∃loise (∀belard) wins every play that starts at v and adheres to s. *Solving* a parity game amounts to computing the *winning regions* for ∃loise and ∀belard, i.e. the sets of nodes $win_\exists$ and $win_\forall$ for which the respective player has a winning strategy.

Example 6.14. Consider the following example of a parity game:

Theorem 6.15 (Jurdzinski)**.** The problem of solving parity games is in UP ∩ Co-UP.

Proof. This proof is out of the scope of this lecture. □

In more detail, the partition $V = win_\exists \uplus win_\forall$ of the set of nodes of a parity game $G = (V, E, \alpha)$ can be computed in deterministic time $n^{\mathcal{O}(k)}$, where $n = |V|$ and $k = \max\{\alpha(v) \mid v \in V\}$.

Note. Parity games with n nodes and a fixed number of priorities k can hence be solved in deterministic polynomial time. However, the question, whether parity games with *unbounded* number of priorities can also be solved in time polynomial in n and k is a long-standing open problem.

Definition 6.16. The *alternation depth* $ad(\psi)$ of a μ-calculus formula ψ is the maximum number of nestings of "entangled" least and greatest fixpoint formulae (i.e. fixpoint formulae using common variables) in ψ, starting with 0 for greatest fixpoint operators and with 1 for least fixpoint operators.

For instance we have

$$ad(\nu X.\, p \wedge \Diamond X) = 0 \qquad ad(\mu X.\, p \vee \Box X) = 1$$
$$ad(\nu X.\nu Y.\, p \vee \Box(X \wedge Y)) = 0 \qquad ad(\mu X.\mu Y.\, p \vee \Box(X \wedge Y)) = 1$$
$$ad(\nu X.\mu Y.\, (p \wedge \Box X) \vee (\neg p \wedge \Box Y)) = 2 \qquad ad(\mu X.\nu Y.\, (p \wedge \Box X) \vee (\neg p \wedge \Box Y)) = 1$$

Theorem 6.17. Model checking for the μ-calculus is linear time equivalent to solving parity games.

Proof. (Sketch) First, we reduce model checking to parity games. So, let $\mathfrak{M}$ be a model, let w a world of $\mathfrak{M}$ and let ψ be a μ-calculus formula. To decide, whether $\mathfrak{M}, w \models \psi$, construct the parity game $G = (V, E, \alpha)$ where $V = cl(\psi) \times W$. We put $(\phi, v) \in V_\exists$ if $\phi = \psi_1 \vee \psi_2$, if $\phi = \Diamond_i \psi_1$, if $\phi = p$ and $v \notin V(p)$ or if $\phi = \neg p$ and $v \in V(p)$; in all other cases, we put $(\phi, v) \in V_\forall$. For all v and all $w \in R(v)$, define E by putting

$$(\psi_1 \vee \psi_2, v)E(\psi_1, v) \qquad (\psi_1 \vee \psi_2, v)E(\psi_2, v)$$
$$(\psi_1 \wedge \psi_2, v)E(\psi_1, v) \qquad (\psi_1 \wedge \psi_2, v)E(\psi_2, v)$$
$$(\Diamond_i \psi_1, v)E(\psi_1, w) \qquad (\Box_i \psi_1, v)E(\psi_1, w)$$
$$(\mu X.\, \psi_1, v)E(\psi_1[\mu X.\, \psi_1/X], v) \qquad (\nu X.\, \psi_1, v)E(\psi_1[\nu X.\, \psi_1/X], v).$$

If ϕ is a fixpoint formula, i.e. if ϕ is of the shape $\mu X.\, \psi_1$ or $\nu X.\, \psi_1$, for some $X \in Var$ and some μ-calculus formula ψ_1, then put $\alpha(\phi, v) = ad(\phi)$; otherwise put $\alpha(\phi, v) = 0$.

Notice that the game G can be constructed from $\mathfrak{M}$ and ψ in linear time. One can show that $\exists$loise has a winning strategy at (w, ψ) in the constructed game G iff $\mathfrak{M}, w \models \psi$.

To reduce parity games to model checking, one proceeds in a similar manner and constructs a model and a formula from a given parity game and shows that again $\exists$loise wins a node in the given parity game iff the according world in the constructed model satisfies the constructed formula. □

Corollary 6.18. The model checking problem of the μ-calculus is in UP ∩ Co-UP. The model checking problem of CTL is in P.

Proof. The first statement follows directly from Theorem 6.15 and Theorem 6.17, while the proof of the second statement requires the additional observation that the embedding function e translates CTL-formulae to *single-variable* μ-calculus formulae and that the model checking problem of CTL thus is (again by Theorem 6.17) equivalent to the problem of solving parity games with just two priorities 0 and 1 (as the alternation depth of single-variable μ-calculus formulae is at most 1). □

6.5 Satisfiability solving for fixpoint logics

We note that the μ-calculus (and hence also CTL) has the FMP:

Theorem 6.19 (Kozen, 1988)**.** Let ψ be a satisfiable μ-calculus formula. Then ψ is satisfiable in a finite model.

Proof. The proof of this theorem is out of scope of the lecture; it involves the use of so-called *quasi-well orders* to construct finite models. □

In fact, the μ-calculus even has a BMP, as one can show that every satisfiable μ-calculus formula is satisfiable in a model of size at most $2^{\mathcal{O}(nk \log n)}$. Thus the satisfiability and validity problems of CTL and the μ-calculus are decidable. Furthermore, these problems are ExpTime-complete:

Theorem 6.20 (Fischer, Emerson, Halpern, Jutla)**.** The satisfiability (and validity) problems of CTL and the μ-calculus are ExpTime-complete.

Proof. (Ideas:)

1. ExpTime-hardness is shown by reducing polynomial-space alternating Turing-machines to CTL-formulae.

2. To see containment in ExpTime, construct a tableau using the tableau-rules from Definition 4.21, extended with the following two rules that unfold fixpoint formulae:

$$(\nu)\ \frac{\Gamma, \nu X.\ \psi}{\Gamma, \psi[\nu X.\ \psi / X]} \qquad\qquad (\mu)\ \frac{\Gamma, \mu X.\ \psi}{\Gamma, \psi[\mu X.\ \psi / X]}$$

   Due to the presence of these two rules, the resulting tableau is a (cyclic) graph. In a second step, construct a parity game that traces fixpoint formulae through this graph; for CTL, the resulting game has two priorities, for the μ-calculus, it has an unbounded number of priorities. The described algorithm runs in time $2^{\mathcal{O}(n^2 k^2 \log n)}$, where k denotes the alternation-depth of the input formula and where n denotes the length of the input formula.

□

6.6 Expressivity of fixpoint logics

We have seen that fixpoint logics are more expressive than basic modal logics. But "how expressive" are fixpoint logics in detail?

Definition 6.21. Let $\mathfrak{M}$ and $\mathfrak{M}'$ be two models and let w and w' be worlds of $\mathfrak{M}$ and $\mathfrak{M}'$, respectively. Put

$$\mathfrak{M}, w \equiv_\mu \mathfrak{M}, w' \qquad \text{iff} \qquad T^\mu_{\mathfrak{M}(w)} = T^\mu_{\mathfrak{M}'(w')},$$

where $T^\mu_{\mathfrak{M}(w)} = \{\psi \mid \psi \text{ is } \mu\text{-calculus formula with } \mathfrak{M}, w \models \psi\}$.

Lemma 6.22. If $\mathfrak{M}$ and $\mathfrak{M}'$ are image-finite models, then for all worlds w of $\mathfrak{M}$ and all worlds w' of $\mathfrak{M}'$,

$$\mathfrak{M}, w \simeq \mathfrak{M}', w' \qquad \text{iff} \qquad \mathfrak{M}, w \equiv_\mu \mathfrak{M}', w'.$$

Proof. See proof of Theorem 2.21. □

It turns out that we can embed the μ-calculus into monadic second-order logic (MSO):

Definition 6.23. The *standard translation* $ST^{\mu}_x(\psi)$ of a μ-calculus formula ψ is defined to be $ST_x(\psi)$, if ψ is not of the shape $\mu X.\ \psi_1$ or $\nu X.\ \psi_1$ and as

$$ST^{\mu}_x(\mu X.\ \psi_1) = \forall X.(\forall y.(ST^{\mu}_x(\psi_1) \to X(y))) \to X(x)$$
$$ST^{\mu}_x(\nu X.\ \psi_1) = \exists X.(\forall y.(X(y) \to ST^{\mu}_x(\psi_1))) \wedge X(x)$$

otherwise.

Lemma 6.24. For all μ-calculus formulae ψ, all models $\mathfrak{M}$ and all worlds w of $\mathfrak{M}$,

$$\mathfrak{M}, w \models \psi \qquad \text{iff} \qquad \mathfrak{M}, \eta \models_{MSO} ST^{\mu}_x(\psi) \text{ for } \eta = [w/x].$$

Proof. The proof of this lemma is similar to the proof of Proposition 1.12, but requires special attention in the fixpoint cases. □

The μ-calculus is the bisimulation-invariant fragment of MSO:

Theorem 6.25. Let $\psi(x)$ be a MSO-formula with one free variable in the signature (R, P, Var). Then $\psi(x)$ is invariant for bisimulations (see Definition 2.22) iff it is equivalent to the standard translation of a μ-calculus formula.

Proof. Out of scope. □

Chapter 7

Coalgebraic modal logic

In the previous chapters we have considered various logics and have shown expressivity, decidability and complexity results for some individual logics. The framework of coalgebraic modal logic generalizes the concept of modal logics and allows for obtaining generic results for all coalgebraic modal logics *at once*. We consider the approach to coalgebraic logics that was established by D. Pattinson and L. Schröder in which the semantics of modal operators is defined by means of so-called predicate liftings. Genericity is added w.r.t. to two concepts:

1. the kind of models over which formulae are interpreted, and
2. the semantics of modal operators.

Genericity in 1. is obtained by taking *coalgebras* that are parametrized by *functors* as models; genericity in 2. is obtained by defining semantics of modal operators using the flexible concept of so-called *predicate liftings*.

Example 7.1. The following are examples of logics that are covered by the coalgebraic framework.

1. **K**, **KD** and their multi-modal versions, coming with the standard modalities $\Diamond_i\phi$ for $i \in I$, having the intuition that "there is an i-successor at which ϕ holds."
2. Graded modal logic (GML), having modal operators $\Diamond_n$, for $n \in \mathbb{N}$, where $\Diamond_n\phi$ comes with the intuition that "there are at least n successors in which ϕ holds".
3. Probabilistic modal logic (PML), having modal operators L_p, for $p \in [0,1] \cap \mathbb{Q}$, where $L_p\phi$ comes with the intuition that "with probability at least p, ϕ holds in the next step".
4. Neighbourhood logic, having the modal operators $\Diamond$ and $\Box$ which however come with a different semantics; here $\Diamond\phi$ has the intuition that "there is a neighbourhood in which ϕ holds somewhere".
5. Coalition logic, having a fixed set of n agents $N = \{1, \ldots, n\}$ and modal operators $\Box_D$ (also written [D]), for each *coalition* $D \subseteq N$, where $\Box_D\phi$ comes with the intuition that "coalition D has a strategy to enforce that ϕ holds in the next step".
6. Conditional logic, coming with the *binary* modal operator $\Rightarrow$, where $\phi \Rightarrow \psi$ has the intuition that "if ϕ then *usually* ψ".

However, the standard coalgebraic approach fails for e.g. the logic **S4**, as transitivity and symmetry do not allow for a straightforward coalgebraic representation.

7.1 Category theoretical notions

In order to be able to define the semantics of coalgebraic modal logic, we first have to introduce some basic concepts from category theory.

Definition 7.2. A *category* consists of a collection $\mathbf{C}$ of *objects* and a collection of *morphisms*, where we require that there is – for each object $A \in \mathbf{C}$ – the *identity morphism* id_A and that we have associative composition of morphisms (i.e. that we have $(h \circ g) \circ f = h \circ (g \circ f)$ for any three compatible morphisms). For instance, **Set**, the *category of sets*, has sets as objects and functions as morphisms. Given two categories $\mathbf{C}$ and $\mathbf{D}$, a (covariant) *functor* $T : \mathbf{C} \to \mathbf{D}$ maps objects $A \in \mathbf{C}$ to objects $TA \in \mathbf{D}$ and morphisms $f : A \to B$ from $\mathbf{C}$ to morphisms $Tf : TA \to TB$ in $\mathbf{D}$ s.t.

- for each object $X \in \mathbf{C}$,

$$T(\mathbf{Id}_X) = \mathbf{Id}_{TX},$$

- for all morphisms $f : X \to Y$ and $g : Y \to Z$,

$$T(g \circ f) = Tg \circ Tf$$

If $\mathbf{C} = \mathbf{D}$, then $T : \mathbf{C} \to \mathbf{D}$ is called *endo-functor*.

Definition 7.3. Let $\mathbf{C}$ be a category and let $T : \mathbf{C} \to \mathbf{C}$ be an endo-functor on $\mathbf{C}$. A *T-coalgebra* $C = (W, \gamma)$ consists of an object $W \in \mathbf{C}$ (the *carrier* of the coalgebra) and a morphism $\gamma : W \to TW$ (the *structure* of the coalgebra). A *T-coalgebra-homomorphism* from T-coalgebra $C = (W_1, \gamma_1)$ to T-coalgebra $D = (W_2, \gamma_2)$ is a morphism $h : W_1 \to W_2$ s.t. the following diagram *commutes*

$$\begin{array}{ccc} W_1 & \xrightarrow{h} & W_2 \\ \downarrow{\scriptstyle \gamma_1} & & \downarrow{\scriptstyle \gamma_2} \\ TW_1 & \xrightarrow{Th} & TW_2 \end{array}$$

i.e. s.t. $\gamma_2 \circ h = Th \circ \gamma_1$.

Note. The category $\mathbf{Coalg}(T)$ has T-coalgebras as objects and T-coalgebra-homomorphisms as morphisms. Exercise: show that $\mathbf{Coalg}(T)$ indeed is a category.

Example 7.4. Consider the following exemplary coalgebras:

- Let $T = \mathcal{P}$, where $\mathcal{P}$ is the powerset (**Set** endo-)functor with

$$\begin{aligned} \mathcal{P}(X) &= \{Y \mid Y \subseteq X\} \\ (\mathcal{P}f)(B) &= f[B], \end{aligned}$$

  where $f[B] = \{f(b) \mid b \in B\}$ is the *image* of B under f. $\mathcal{P}$-coalgebras are Kripke frames. For instance we have the $\mathcal{P}$-coalgebra $C = (W, \gamma)$, where $\gamma : W \to \mathcal{P}(W)$ and

$$\begin{aligned} W &= \{a, b, c\} & \gamma(a) &= \{a, b\} \\ \gamma(b) &= \{a\} & \gamma(c) &= \{b\} \end{aligned}$$

- $\mathcal{P}(A \times Id)$-coalgebras are *Labelled transition systems*: Consider the $\mathcal{P}(A \times Id)$-coalgebra $C = (W, \gamma)$, where $\gamma : W \to \mathcal{P}(A \times W)$, $A = \{a, b\}$ and

$$\begin{aligned} W &= \{1, 2, 3\} & \gamma(1) &= \{(a, 1), (a, 3)\} \\ \gamma(2) &= \{(b, 2)\} & \gamma(3) &= \{(a, 2), (b, 2)\} \end{aligned}$$

Definition 7.5 (Predicate liftings)**.** The *contravariant powerset functor* 2 is defined by putting

$$\begin{aligned} 2(X) &= \mathcal{P}(X) \\ 2(f)(B) &= f^{-1}[B], \end{aligned}$$

where $f : X \to Y$, $B \subseteq Y$ and $f^{-1}[B] = \{b \in X \mid f(x) \in B\}$ is the *preimage* of B under f. Let $\heartsuit$ be a modal operator. A *predicate lifting* (Pattinson, 2003) for a **Set** endo-functor $T : \mathbf{Set} \to \mathbf{Set}$ is a *natural transformation*

$$[\![\heartsuit]\!] : 2 \to 2 \circ T,$$

that is, $[\![\heartsuit]\!]$ denotes a family of mappings

$$([\![\heartsuit]\!]_X : \mathcal{P}(X) \to \mathcal{P}(TX))_{X \in \mathbf{Set}}$$

s.t. the following diagram commutes for each $f : X \to Y$:

$$\begin{array}{ccc} \mathcal{P}(X) & \xrightarrow{[\![\heartsuit]\!]_X} & \mathcal{P}(TX) \\ {\scriptstyle f^{-1}}\uparrow & & \uparrow{\scriptstyle (Tf)^{-1}} \\ \mathcal{P}(Y) & \xrightarrow{[\![\heartsuit]\!]_Y} & \mathcal{P}(TY) \end{array} \qquad \begin{array}{c} X \\ \downarrow{\scriptstyle f} \\ Y \end{array}$$

Thus we require that for each $f : X \to Y$,

$$[\![\heartsuit]\!]_X \circ f^{-1} = (Tf)^{-1} \circ [\![\heartsuit]\!]_Y.$$

7.2 Coalgebraic modal logic

Definition 7.6. Let Δ be a set of modal operators. The set of *coalgebraic modal formulae* is defined by the following grammar:

$$\psi, \phi := \top \mid \psi \wedge \phi \mid \neg\psi \mid \heartsuit\psi \qquad (\heartsuit \in \Delta)$$

Note. Propositional variables can be modelled as nullary modal operators, see Example 7.8 below.

Definition 7.7. Let $C = (W, \gamma)$ be a T-coalgebra and let Δ be a set of modal operators coming with a T-predicate lifting $[\![\heartsuit]\!]$ for each $\heartsuit \in \Delta$. The *extension* (or *truth set*) $[\![\psi]\!]^C$ of a coalgebraic modal formula ψ in C is defined inductively as follows:

$$\begin{aligned} [\![\top]\!]^C &= W \\ [\![\psi \wedge \phi]\!]^C &= [\![\psi]\!]^C \cap [\![\phi]\!]^C \\ [\![\neg\psi]\!]^C &= \overline{[\![\psi]\!]^C} = W \setminus [\![\psi]\!]^C \\ [\![\heartsuit\psi]\!]^C &= \gamma^{-1}[[\![\heartsuit]\!]_W([\![\psi]\!]^C)]. \end{aligned}$$

Example 7.8. We consider several exemplary coalgebraic modal logics.

1. Let $T = \mathcal{P}$, let $\Delta = \{\Diamond, \Box\}$ and recall that $\mathcal{P}$-coalgebras are Kripke frames. Define the following predicate liftings for all sets X and all $B \subseteq X$:

$$\begin{aligned}[\![\Box]\!]_X(B) &= \{A \in \mathcal{P}(X) \mid A \subseteq B\} \\ &= \{A \in \mathcal{P}(X) \mid \forall v \in A.v \in B\}\end{aligned}$$

$$\begin{aligned}[\![\Diamond]\!]_X(B) &= \{A \in \mathcal{P}(X) \mid A \cap B \neq \emptyset\} \\ &= \{A \in \mathcal{P}(X) \mid \exists v \in A.v \in B\}\end{aligned}$$

   We observe that the resulting coalgebraic logic is *exactly* **K** (without propositional atoms). As an example, consider the Kripke frame $C = (W, \gamma)$ with $W = \{a, b, c, d\}$ and $\gamma : W \to \mathcal{P}W$, where

$$\begin{aligned}\gamma(a) &= \{a, b, c\} & \gamma(b) &= \emptyset \\ \gamma(c) &= \{d\} & \gamma(d) &= \emptyset.\end{aligned}$$

   We observe that

$$[\![\bot]\!]^C = [\![\neg\top]\!]^C = \overline{[\![\top]\!]^C} = \overline{W} = \emptyset$$

   and that hence

$$\begin{aligned}[\![\Box\bot]\!]^C &= \gamma^{-1}[[\![\Box]\!]_W([\![\bot]\!]^C)] \\ &= \gamma^{-1}[[\![\Box]\!]_W(\emptyset)] \\ &= \gamma^{-1}[\{\emptyset\}] \\ &= \{w \in W \mid \gamma(w) \in \{\emptyset\}\} \\ &= \{b, d\}.\end{aligned}$$

   Also we have

$$\begin{aligned}[\![\Diamond\Box\bot]\!]^C &= \gamma^{-1}[[\![\Diamond]\!]_W([\![\Box\bot]\!]^C)] \\ &= \gamma^{-1}[[\![\Diamond]\!]_W(\{b, d\})] \\ &= \gamma^{-1}[\{\ \{b\}, \{d\}, \{b, d\}, \{d, c\}, \{b, c\}, \{a, b\}, \{a, d\}, \\ &\qquad \{a, b, d\}, \{a, b, c\}, \{a, c, d\}, \{b, c, d\}, \{a, b, c, d\}\ \}] \\ &= \{a, c\}\end{aligned}$$

2. For multi-modal logics with index set I, put $T = \mathcal{P}(I \times Id)$ and $\Delta = \{\Diamond_i, \Box_i \mid i \in I\}$ and define the following predicate liftings for all $i \in I$, all sets X and all $B \subseteq X$:

$$\begin{aligned}[\![\Box_i]\!]_X(B) &= \{S \in TX \mid \forall (b, x) \in S.b = i \to s \in B\} \\ [\![\Diamond_i]\!]_X(B) &= \{S \in TX \mid \exists (b, x) \in S.b = i \wedge s \in B\}\end{aligned}$$

   The obtained logic is multi-modal **K** (without propositional atoms).

3. To model the propositional atoms given as a set $\mathcal{A}$, move from T to $\mathcal{P}(\mathcal{A}) \times T$ so that $\mathcal{P}(\mathcal{A}) \times T$-coalgebras $C = (W, \gamma)$ assign pairs $\gamma(x) = (Q, y)$ to states $x \in W$. Then put $[\![p]\!]^C = \gamma^{-1}[[\![p]\!]_W]$, where for all sets X and all $p \in \mathcal{A}$,

$$[\![p]\!]_X = \{(Q, y) \in \mathcal{P}(\mathcal{A}) \times TX \mid p \in Q\}.$$

4. Probabilistic modal logic: Put $\Delta = \{L_p \mid p \in [0,1] \cap \mathbb{Q}\}$. The *finite distribution functor* $\mathcal{D}_\omega$ maps sets X to the set of probability distributions on X with finite support. $\mathcal{D}_\omega$-coalgebras are probabilistic type spaces (Markov-chains) with finite branching degree. Put

$$[\![L_p]\!]_X(B) = \{P \in \mathcal{D}_\omega(X) \mid P(B) \geq p\},$$

where $P(B) = \sum_{b \in B} P(b)$.

5. As an example that moves away from relational semantics, consider *coalition logic*. Let $N = \{1, \ldots, n\}$ be a set of agents. Coalitions D are subsets of N. Let $\Delta = \{\Box_D \mid D \subseteq N\}$. Define the functor

$$\mathcal{G}(X) = \{(S_1, \ldots, S_n, f) \mid \emptyset \neq S_i \in \mathbf{Set}, f : \prod_{i \in N} S_i \to X\},$$

assigning to a set X the set of all tuples consisting of n non-empty sets of *allowed moves* S_i of the agents $i \in N$ and of an evaluation function f that evaluates tuples of moves $m_1 \in S_1, \ldots, m_n \in S_n$ as $f(m_1, \ldots, m_n) = x$, where $x \in X$. $\mathcal{G}$-coalgebras are in one-to-one correspondence with so-called *game-frames*.

Put $S_D = \prod_{i \in D} S_i$ (so that e.g. $S_{\{1,3\}} = S_1 \times S_3$) and $S_{\overline{D}} = \prod_{i \notin D} S_i$; for $m_D \in S_D$ and $m_{\overline{D}} \in S_{\overline{D}}$, we have $(m_D, m_{\overline{D}}) \in \prod_{i \in N} S_i$. For each coalition $D \subseteq N$, define the predicate lifting

$$[\![\Box_D]\!]_X(B) = \{(S_1, \ldots, S_n, f) \in \mathcal{G}(X) \mid \exists m_D \in S_D . \forall m_{\overline{D}} \in S_{\overline{D}} . f(m_D, m_{\overline{D}}) \in B\}.$$

Consider the game-frame $C = (W, \gamma)$ with $N = \{1, 2\}$, $W = \{x, y, z\}$ and $\gamma : W \to \mathcal{G}(W)$, where

$$\begin{aligned}
\gamma(x) &= \{\ \{l, r\}, \{u, d\}, \{((l, u), x), ((l, d), y), ((r, u), z), ((r, d), y)\}\ \} \\
\gamma(y) &= \{\ \{l\}, \{d\}, \{((l, d), y)\}\ \} \\
\gamma(z) &= \{\ \{l, r\}, \{d\}, \{((l, d), x), ((r, d), y)\}\ \}
\end{aligned}$$

We observe that

$$\begin{aligned}
[\![\Box_{\{1\}} \top]\!]^C &= \gamma^{-1}[[\![\Box_{\{1\}}]\!]_W(W)] \\
&= \gamma^{-1}[\{\gamma(x), \gamma(y), \gamma(z)\}] \\
&= \{x, y, z\} = W
\end{aligned}$$

and that hence

$$\begin{aligned}
x \in [\![\Box_{\{1,2\}} \Box_{\{1\}} \top]\!]^C &= \gamma^{-1}[[\![\Box_{\{1,2\}}]\!]_W([\![\Box_{\{1\}} \top]\!]^C)] \\
&= \gamma^{-1}[[\![\Box_{\{1,2\}}]\!]_W(W)] \\
&= \gamma^{-1}[\{\gamma(x), \gamma(y), \gamma(z)\}] \\
&= \{x, y, z\} = W
\end{aligned}$$

Note. We take note of the following slogan (by Schröder and Pattinson, 2010):

"Rank-1 axiomatizable modal logics are coalgebraic."

7.3 Complexity and expressivity results

We gather a few results on coalgebraic modal logic. Notice that the upcoming results are generic in the sense that they have to be proven only once and then instances of the results follow for any logic to which the coalgebraic framework can be instantiated, e.g. for all the logics considered in the previous section (with the exception of coalition logic, for which the given predicate liftings fail to be expressive; in more detail: the obtained coalition logic coincides with traditional coalition logic and is PSPACE-complete, but it does not characterize behavioural equivalence of game-frames).

To begin with, coalgebraic modal logic has the FMP:

Theorem 7.9 (Schröder, 2007)**.** Let ψ be a satisfiable coalgebraic modal formula. Then ψ is satisfiable in a finite coalgebra.

Proof. Out of scope. □

Theorem 7.10 (Schröder and Pattinson, 2007)**.** The satisfiability and validity problems of coalgebraic modal logic are PSPACE-complete (under the mild assumption of so-called PSPACE-*tractability*).

Proof. Out of scope. □

The notion of bisimulation ($\simeq$) can be generalized from models to T-coalgebras. The resulting general concept is called *behavioural equivalence*, written $\simeq_T$. Under the assumption that the set of predicate liftings of a coalgebraic logic can be used to distinguish any two states that are not behaviourally equivalent, behavioural equivalence characterizes logical equivalence:

Theorem 7.11 (Schröder and Pattinson, 2007)**.** Let $\equiv_T$ denote logical equivalence w.r.t. coalgebraic modal formulae. If the used set of predicate liftings is *expressive*, then for all coalgebras C and D and all states w of C and w' of D, we have

$$C, w \simeq_T D, w' \qquad \text{iff} \qquad C, w \equiv_T D, w'.$$

A generalized Rosen-van Benthem Theorem has been recently proven (Schröder, Pattinson, Litak, 2015), stating that under mild assumptions, coalgebraic modal logic is the behavioural-equivalence-invariant fragment of so-called coalgebraic *predicate* logic.

7.4 The coalgebraic μ-calculus

Similarly to basic modal logic, coalgebraic modal logic can be extended with fixpoint operators:

Definition 7.12. Let Δ be a set of modal operators. The *coalgebraic μ-calculus* is defined by the following grammar:

$$\psi, \phi := \top \mid \bot \mid \psi \wedge \phi \mid \psi \vee \phi \mid X \mid \heartsuit\psi \mid \mu X.\, \psi \mid \nu X.\, \psi \qquad (X \in \mathit{Var})$$

We require that each $\heartsuit \in \Delta$ comes with a predicate lifting $[\![\heartsuit]\!]$ that is monotone w.r.t. set inclusion and that for every $\heartsuit \in \Delta$, there is a $\overline{\heartsuit} \in \Delta$ s.t. for all sets X and all

$B \subseteq X$, $[\![\heartsuit]\!]_X(B) = \overline{[\![\overline{\heartsuit}_X(\overline{B})]\!]}$. Also let $C = (W, \gamma)$ be a T-coalgebra. The *extension* of a coalgebraic μ-calculus formula ψ is defined as $[\![\psi]\!]^C_\epsilon$, where for all substitutions σ,

$$\begin{aligned}
[\![\top]\!]^C_\sigma &= W \\
[\![\bot]\!]^C_\sigma &= \emptyset \\
[\![\psi \wedge \phi]\!]^C_\sigma &= [\![\psi]\!]^C_\sigma \cap [\![\phi]\!]^C_\sigma \\
[\![\psi \vee \phi]\!]^C_\sigma &= [\![\psi]\!]^C_\sigma \cup [\![\phi]\!]^C_\sigma \\
[\![\heartsuit\psi]\!]^C_\sigma &= \gamma^{-1}[[\![\heartsuit]\!]_W([\![\psi]\!]^C_\sigma)] \\
[\![X]\!]^C_\sigma &= \sigma(X) \\
[\![\mu X.\ \psi]\!]^C_\sigma &= LFP([\![\psi]\!]^X_\sigma)^C \\
[\![\nu X.\ \psi]\!]^C_\sigma &= GFP([\![\psi]\!]^X_\sigma)^C,
\end{aligned}$$

where for all $A \subseteq W$,

$$([\![\psi]\!]^X_\sigma)^C(A) = [\![\psi]\!]^C_{\sigma[A/X]}.$$

Example 7.13. Consider the following instances of the coalgebraic μ-calculus:

- For $T = \mathcal{P}(I) \times Id$ and the predicate liftings from Example 7.8, Item 2. and 3., we obtain the standard relational μ-calculus (containing CTL as a fragment).
- For $T = \mathcal{D}_\omega$ and the predicate liftings from Example 7.8, Item 4., we obtain probabilistic fixpoint logic.
- For $T = \mathcal{G}$ and the predicate liftings from Example 7.8, Item 5., we obtain the alternating-time μ-calculus (AMC), containing Alternating-time temporal logic (ATL) as a fragment. ATL has operators $\langle\langle D\rangle\rangle A\phi U\psi$ having the intuition that "coalition D has a strategy to ensure that ϕ holds until eventually ψ holds."

Theorem 7.14 (Cîrstea, Kupke, Pattinson, 2011)**.** The satisfiability and validity problems of the coalgebraic μ-calculus are in ExpTime (under the relatively mild assumption of ExpTime-*tractability*).

Proof. This proof is out of scope; it involves the use of parity games. □

A fast algorithm to decide the satisfiability and validity problems of the *alternation-free* fragment of the coalgebraic μ-calculus has been developed and recently implemented (Hausmann, Schröder, Egger, 2016) as part of the *Coalgebraic Ontology Logic Reasoner* (COOL):

`http://www8.cs.fau.de/research:software:cool`