Learning Modal Formulas via Dualities

MSc Thesis *(Afstudeerscriptie)*

written by

Raoul Koudijs
(born 24 February 1997 in Dirksland, Netherlands)

under the supervision of **Balder ten Cate**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:
18 May 2022

Members of the Thesis Committee:
Prof. Benno van den Berg (chair)
Prof. Balder ten Cate (supervisor)
Dr. Malvin Gattinger
Prof. Nick Bezhanishvili

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Abstract

We initiate the study of finite characterisations and exact learnability of modal languages. A finite characterisation of a modal formula w.r.t. a class of formulas is a finite set of finite models (labelled either positive or negative) which distinguishes this formula from every other formula from that class. A language is finitely characterisable if every formula in it has a finite characterisation w.r.t. it. We show that normal modal logics are finitely characterisable if and only if they are locally tabular. Further, we define the category of pointed Kripke models and weak simulations and show that the existence of dualities in this category relate to finite characterisability of the positive modal language without the truth-constants $\top$ and $\bot$. In fact, we show that our techniques apply to a larger class of uniform formulas. Moreover, our results are essentially optimal as we show that allowing any kind of non-uniformity makes the language non-characterisable. Throughout, we indicate what exact learning algorithms can be obtained from these characterisations.

Acknowledgement

I will take some space here to thank all who have supported me. First of all, thank you Balder. Your masterful advice and clarity of thought always kept me on track and brought my aberrations down to earth. Simultaneously, your intellectual stature elevated me high enough to see new pathways opening in the world of logic and beyond. Thank you for navigating me through this thesis, I hope we will continue collaborating in the future. Also, special thanks to Nick Bezhanishvili and Frank Wolter for helpful conversations about the results from this thesis, and to Alexandru Baltag and Johan van Benthem for guiding me through other research projects during my studies and ever feeding my curiosity. Finally, I want to thank my friends and family who have often had to deal with the less pretty side of it all. Thank you for loving me despite my absence, too often I was lost in thought . . .

Contents

Chapter 1

Introduction

In this thesis, we initiate the study of *finite characterisations* and *exact learnability* of modal languages. Exact learning is a framework for learning problems that involve learner-teacher interaction that is well-studied in computational learning theory [2]. In particular, it is applied to the problem of learning or 'reverse engineering' a hidden goal concept from data examples. The teacher is formalised in terms of various kinds of oracles, but we will focus almost exclusively on one kind of oracle-query; membership queries. Finite characterisations are finite sets of data examples that suffice to distinguish the goal concepts from a given class of concepts under consideration. Note that finite characterisations are thus always defined with respect to (w.r.t.) some class of concepts under consideration. Finite characterisations are a necessary precondition for exact learning with membership queries only, but they also always yield some learning algorithm with membership queries (albeit a highly inefficient one).

An example of a concept class is the set of model-classes of formulas from some language. Recently, it has been elegantly shown that a large class of formulas in the positive-existential-conjunctive fragment of FO (also known as 'conjunctive queries') admits polynomial time-computable finite characterisations via the existence of frontiers and dualities in the homomorphism lattice of finite structures [9]. Moreover, an even larger class of formulas within the positive-existential fragment of FO (also known as 'unions of conjunctive queries') admits exponential time-computable characterisations via homomorphism dualities. We make the latter result explicit for the first time, as it was implicit in [1, 9]. It is well-known that modal logic can be seen as a fragment of first order logic, and hence the above-mentioned characterisability and learnability results for positive-existential FO transfer to positive-existential modal logic (i.e. the class of modal formulas that use only $\Diamond, \wedge, \vee$ and positive atoms). First, we make these corrolaries explicit and then improve upon their results in the context of modal languages.

Yet also independently of the line pursued there, we embark on the study of finite characterisations of modal formulas *per sé*. We aim to give a complete picture. In chapter 2, we introduce the necessary background for the results and methods used in this thesis. First, 2.1 introduces the basics of modal logic, in 2.2 we formally define finite characterisations and exact learning algorithms, in 2.3 we introduce the frontiers and dualities from a categorical perspective and finally 2.4 reviews the results from [9] for unions of conjunctive queries. Chapter 3 studies the question when a normal modal logic is finitely characterisable. We show this to be the case iff the logic is 'locally tabular' (i.e. contains only finitely many formulas up to equivalence). Next, chapter 4 introduces *weak simulations* and develops a categorical framework similar to [9, 23] for studying finite characterisations via dualities. This is in anticipation of the next chapter, where our main result is that this fragment is finitely characterisable. Thus in chapter 5, we study the question which *syntactic fragments* (understood as restricting the set of base connectives) of the modal language are finitely characterisable (over all frames). We exploit the categorical framework established in chapter 4 to show that the large class of *uniform* formulas is finitely characterisable (that is, each uniform formula is finitely characterisable w.r.t. the class of all of them). Moreover, we show that our results are essentially optimal. Finally, in chapter 6 we discuss whether more efficient learning algorithms could be obtained via these methods.

Chapter 2

Preliminaries

In this preliminary section, we will review the basics of modal logic, define finite characterisability and exact learnability and introduce dualities and frontiers in a categorical setting. Finally, we will review recent work on finite characterisations and exact learnability for fragments of first order logic (FO) [9]. The main ideas from this thesis are adapted from their setting, and their results even imply results in the modal setting via standard translation of modal logic into FO. We make some of these corollaries implicit in their work explicit, and present the present work as a continuation of their efforts. The section on modal logic is based on [8], so for references please consult this book. All other references have been indicated in the text.

2.1 Modal Logic

First, we define what is a modal formula. Note that the language does not have a 'binding' construct (for 'unfreeing' variables).

Definition 2.1. (Full Modal Language)
The *full modal language* over a set Prop of propositional variables is recursively generated by

$$\varphi ::= \ p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \Box\varphi \mid \Diamond\varphi \mid \top \mid \bot$$

where $p \in$ Prop. Note the omission of the $\rightarrow$ connective is justified as $p \rightarrow q \equiv \neg p \vee q$. A *modal language* is just a collection of modal formulas, i.e. a subset of the full modal language.

We make no assumption on the size of the set of propositional variables Prop. But note that, even if Prop is finite, there are already infinitely modal formulas up to logical equivalence (over the class of all frames).

Definition 2.2. (Variables)
For a modal formula φ, let $var(\varphi)$ be the set of propositional variables occurring in φ. We write $\varphi(p_1, \ldots, p_n)$ to denote that $var(\varphi) \subseteq \{p_1, \ldots, p_n\}$.

Definition 2.3. (Positive and Negative Formulas)
Let $pos(\varphi)$ denote the set of propositional variables occurring *positively* in φ (i.e. within the scope of an even number of negations). Similarly, let $neg(\varphi)$ denote the set of propositional variables occurring negatively in φ (i.e.within the scope of an odd number of negations). Note that $pos(\varphi)$ and $neg(\varphi)$ are not necessarily disjoint. A formula φ is called *positive* if $var(\varphi) = pos(\varphi)$ and *negative* if $var(\varphi) = neg(\varphi)$.

Definition 2.4. (Modal Fragments)
For $C \subseteq \{\wedge, \vee, \Diamond, \Box, \top, \bot\}$ a set of connectives, we denote by $\mathcal{L}_C$ the set of modal formulas generated from literals (i.e. positive or negated propositional variables from Prop) using the connectives in C. We assume all formulas to be in negation normal form, so that *negations may only occur in front of propositional variables.* Thus, $\mathcal{L}_{\Box,\Diamond,\wedge,\vee}$ is the full modal language in negation normal form. We denote by $\mathcal{L}^+_C, \mathcal{L}^-_C$ the set of positive, respectively negative $\mathcal{L}_C$ formulas.

Definition 2.5. (Modal Depth)
For a formula φ, its *modal depth* $d(\varphi)$ is the length of the longest sequence of nested modal operators (i.e. $\Box$ or $\Diamond$) in φ.

Definition 2.6. (Kripke model)
A *Kripke model* is a triple $M = (dom(M), R, v)$ where $dom(M)$ is a set of 'possible worlds', $R \subseteq dom(M) \times dom(M)$ a binary 'accessibility' relation and $v : W \to \mathcal{P}(\text{Prop})$ is a colouring.[1]. A *Kripke frame* (or simply 'frame') is a Kripke model without a colouring. A *pointed model* is a pair (M, s) of a Kripke model M together with a state $s \in dom(M)$.

Definition 2.7. (Path)
A *finite path* through a model M is a finite sequence $(t_0, \dots, t_n)$ of states in $dom(M)$ such that Rt_it_{i+1} holds for all $i < n - 1$. We also write $t_0R \dots Rt_n$ for this path. Similarly, an *infinite ascending path* through M is an infinite sequence $(t_0, t_1, t_2, \dots)$ of states in $dom(M)$ such that $R(t_i, t_{i+1})$ holds for all natural numbers i. Dually, an *infinite descending path* through M is a sequence $(t_0, t_1, t_2, \dots)$ of states in $dom(M)$ such that $R^{-1}(t_it_{i+1})$ (i.e. $R(t_{i+1}, t_i)$) holds for all natural numbers i.

Definition 2.8. (Rooted Models)
A pointed model M, s is *rooted* if every state $t \in dom(M)$ is reachable by some finite path from the root s, i.e. if there are finitely many states $t_1, \dots, t_n$ such that $sRt_1R \dots Rt_nRt$ is path through M.

We will frequently write "consider a model (M, s)" or "consider a rooted model (M, s)", where of course it is clear from notation that (M, s) is a *pointed* model. Next, we define what it means for a pointed model to be acyclic, via the notion of height of a pointed model.

Definition 2.9. (Height)
The height of a pointed model (M, s) is the length of the longest path in M starting at s, or ∞ if there is no finite upper bound.

Definition 2.10. (Cylic Models)
A model M is *cyclic*, i.e. if there is a path in M of the form $(t, \dots, t)$, i.e. a directed cycle $tR \dots Rt$ for some $t \in dom(M)$. A model is *acyclic* if it is not cyclic.

Definition 2.11. (Tree Models)
A pointed model M, s is a *tree* if it is rooted and acyclic.[2]

Observation. For all pointed models M, s, $height(M, s) \leq |dom(M)|$ or $height(M, s) = \infty$

[1]An ordinary valuation function $V : \text{Prop} \to \mathcal{P}(W)$ induces a 'colouring' $col(V) : W \to \mathcal{P}(\text{Prop})$ as its *transpose* $col(V)(w) := \{p \in \text{Prop} \mid w \in V(p)\}$.

[2]Note that being a tree is only defined for pointed models since rootedness is defined relative to a distinguished element. Also note that our definition of rootedness subsumes the connectedness requirement which is commonly included in the definition of a tree.

Proof. Suppose that $height(M,s) \neq \infty$, i.e. $height(M,s) = n$ for some n. Then there is some path $(s, t_1 \ldots, t_{n-1})$ of length n through M starting at s while there is no path of length $n+1$ starting at this point. Then it cannot be that M, s is cyclic, otherwise there would be paths of arbitrary length starting at s (i.e. then $height(M,s) = \infty$). Hence M, s is acyclic but then all the elements in the path $(s, t_1 \ldots, t_{n-1})$ must be pairwise distinct. It follows that $height(M,s) = n \leq |dom(M)|$. □

Definition 2.12. (Semantics)
The semantic clauses for modal logic are as follows:

$$\begin{array}{rcl}
M,s \models p & \text{iff} & p \in v(s) \\
M,s \models \top & & \text{always} \\
M,s \models \bot & & \text{never} \\
M,s \models \neg\varphi \wedge & \text{iff} & \text{not } M,s \models \varphi \\
M,s \models \varphi \wedge \psi & \text{iff} & M,s \models \varphi \text{ and } M,s \models \psi \\
M,s \models \varphi \vee \psi & \text{iff} & M,s \models \varphi \text{ or } M,s \models \psi \\
M,s \models \Diamond\varphi & \text{iff} & \exists t \in R[s] \text{ with } M,t \models \varphi \\
M,s \models \Box\varphi & \text{iff} & \forall t \in R[s] \text{ it holds that } M,t \models \varphi
\end{array}$$

When we write $M \models \varphi$ we mean that $M, s \models \varphi$ for all $s \in dom(M)$.

There are two ways of thinking about modal logic; syntactically as a set of formulas derivable in some proof system, or semantically as the set of validities of some frame class. We adopt the former perspective.

Definition 2.13. (Normal Modal Logics)
A *normal modal logic* is a set of modal formulas containing all instances of the K-axiom

$$(K) \qquad \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$$

and closed under uniform substitution, modus ponens and generalisation.

Terminological note: we will use regular letters L to refer to modal *logics* (i.e. sets of validities) and calligraphic letters $\mathcal{L}$ to refer to modal *languages* (i.e. arbitrary sets of modal formulas).

Definition 2.14. (Bisimulation)
A relation $Z \subseteq M \times M'$ is a *bisimulation* (notation $M, s \underline{\leftrightarrow} M', s'$) if all $(t, t') \in Z$ satisfy the following clauses:

(atom) $v^M(t) = v^{M'}(t')$

(forth) $R^M tu$ implies $\exists u' \in M'$ with $R^{M'} t'u'$ and $(u, u') \in Z$

(back) $R^{M'} t'u'$ implies $\exists u \in M$ with $R^M tu$ and $(u, u') \in Z$

Theorem 2.1. *Modal formulas are invariant under bisimulations, i.e. if $M, s \models \varphi$ and $M, s \underline{\leftrightarrow} M', s'$ then $M', s' \models \varphi$ (and vice versa) for all modal formulas φ.*

Call a relation $Z \subseteq A \times B$ is *total* if $\forall a \in A. \exists b \in B (a, b) \in Z$ and $\forall b \in B. \exists a \in A (a, b) \in Z$.

Proposition 2.1. *[8] Every bisimulation between rooted models is total.*

A converse to bisimulation-invariance is well-known as the *Van Benthem Characterisation Theorem*, which says that every first order formula that is invariant under bisimulations is equivalent to a modal formula under the standard translation.[3]

2.2 Finite Characterizations and Exact Learning

We will be interested in concept-learning, which means learning a concept from a given *concept class*.

Definition 2.15. (Concept Classes)
A *concept* C over a domain D is a subset of the domain $C \subseteq D$. Consequently, a *concept class* $\mathcal{C}$ over some domain D is a collection of subsets of D, i.e. $\mathcal{C} \subseteq \mathcal{P}(D)$.

Learning a concept C from some concept class $\mathcal{C}$ thus means learning a subset $C \subseteq D$ of the domain by distinguishing it from all others subsets $C' \subseteq D$ with $C' \in \mathcal{C}$.

Definition 2.16. (Examples and Instances)
An *instance* is an element $E \in D$ of the domain D. An *example* is an element $E \in D$ of the domain that is labelled either positive (+) or negative (-). We will not make these labels explicit in notation, instead we will simply refer to them as positive or negative examples.

A concept C *fits* a positive example e if $e \in C$, while C fits a negative example e if $e \notin C$. A concept fits a set of examples if it fits each of them. Further, we say that an example e *distinguishes* between two concepts C and C' if C fits e while C' does not, or if C' fits e while C does not. This leads us to the central concept of this thesis.

Definition 2.17. (Finite Characterizations)
A *finite characterization* of a concept $C \in \mathcal{C}$ w.r.t a concept class $\mathcal{C}$ is a pair of finite sets of examples (E^+, E^-) (where the examples in E^+ are labelled positive and the examples in E^- are labelled negative) such that (i) C fits these examples and (ii) C is the only concept from $\mathcal{C}$ which fits these examples. Further, we say that a concept class $\mathcal{C}$ is *finitely characterisable* if each concept $C \in \mathcal{C}$ has a finite characterisation w.r.t. $\mathcal{C}$.

Finite characterisations, besides being interesting in their own right, are important as a precondition for exact learning from examples (with 'membership queries' only, to be defined shortly). Exact learnability was introduced by Dana Angluin [2] as an interactive form of learning that includes both a learner and a teacher aspect. The idea is that the learning algorithm tries to learn some hidden target concept by asking certain oracle queries. We will be looking at two types of oracles (but there are more, see [2]).

A *membership oracle* is an oracle that when presented with an instance of the domain, (truthfully) says "yes" if the instance belongs to the goal concept and "no" if not. An *equivalence oracle* is an oracle that, when presented with (some representation of) a concept $C \in \mathcal{C}$, returns "yes" if the goal concept is identical to C, and otherwise returns "no" with a counterexample witnessing their non-identity.

But besides learnability, the generation of a finite exhaustive set of data examples consistent with a given logical specification, can be useful for illustration, interactive specification, and debugging purposes (e.g., [27] for relational database queries, [**AlexeCKT2011**] for schema mappings, and [31] for XML queries). The exhaustive nature of the examples is useful in these settings, as they essentially display all 'ways' in which the specification can be satisfied or falsified.

[3] The *standard translation* is the embedding of modal logic into FO. We give a definition in chapter 5.

Definition 2.18. (Exact Learning Algorithm)
An *exact learning algorithm* with membership and/or equivalence queries for a concept class $\mathcal{C}$ is an algorithm that takes no input but has access to the membership oracle and/or equivalence oracle for an unknown goal concept $C \in \mathcal{C}$. The algorithm must terminate after a finite amount of time and output (some representation of)[4] the goal concept C. A concept class $\mathcal{C}$ is *exactly learnable* if there exists a learning algorithm that can identify a concept $C \in \mathcal{C}$ in finite time.

Finite characterisability is a necessary precondition for exact learnability with membership queries only. For if infinitely many examples are needed to distinguish certain concepts from each other, then it is impossible to distinguish between them in finite time by asking membership queries. But the converse also holds, effective finite characterisability (by which we mean the effective computability, not merely the existence, of finite characterisations) of some countable concept class $\mathcal{C}$ induces a naive learning algorithm with membership queries (albeit a very inefficient one). Namely, the algorithm simply enumerates all the concepts $C_1, C_2, \ldots.$ in $\mathcal{C}$ and goes along this list, checking for each C_i whether the hidden goal concept C_{goal} is equivalent to C_i via the finite characterisations by asking membership queries, and halts when it has uniquely identified C_{goal} from this list.

2.2.1 Relationship to Formal Learning Theory

Here, we relate the notion of finite characterisability from computational learning theory to the learnability notion of finite identifiability from formal learning theory. This will not be relevant for the rest of the thesis but serves as contextualization of the quite novel concept of finite characterisations. Our presentation here is based on the dynamic topological framework for learning theory recently developed in [4]. For further background, see [22]. Formal learning theory is concerned with the problem of tracking the truth by making conjectures based on observations. Examples are language-learning (inferring the grammar of a language by observing well-formed sentences) and scientific inquiry (inferring a theory of a phenomenon by observing its effects).

Formally, this is captured by a hypothesis space X containing all (typically infinitely many) hypotheses under consideration, together with a set of observable properties or observations $\mathcal{O} \subseteq \mathcal{P}(X)$ that is closed under finite intersections. That is, observations $O \subseteq X$ are formalised as the set of hypotheses consistent with these observations. The topology on X generated by $\mathcal{O}$ as a topological basis is called the *observational topology*. Various strategies for learning are then formalized as *learners*, i.e. a functions $L : \mathcal{O} \to \mathcal{P}(X)$ from observations to conjectures.

Definition 2.19. (Intersection Space) An *intersection space* is a tuple $(X, \mathcal{O})$, where X is a space of hypotheses and $\mathcal{O} \subseteq \mathcal{P}(X)$ is a family of subsets of X that is closed under finite intersections (i.e. the observations). Further, a triple $(X, \mathcal{O}, L)$ where $(X, \mathcal{O})$ is an intersection space and L is a learner, is called a *learning frame*.

The central notions of formal learning theory are *finite identifiability* and *identifiability in*

[4] We are assumed that concepts are specified using some representation system, so the length of the specification of a concept is well-defined. Formally, a representation system for $\mathcal{C}$ is a string language $\mathcal{L}$ over some finite alphabet Σ (i.e. $\mathcal{L} \subseteq \Sigma^*$) together with a surjective function $r : \mathcal{L} \to \mathcal{C}$. Then by the size $|C|$ of a concept $C \in \mathcal{C}$ we mean the length of the smallest representation of C (i.e. the shortest string in $s \in \mathcal{L}$ such that $r(s) = C$). Similarly, we assume such a representation system, with a corresponding notion of length, for the examples. When there is no risk of confusion, we may conflate concepts (and examples) with their representations.

the limit.[5] Note that all learnability notions are defined relative to a "real world" $x \in X$.[6] A property $P \subseteq X$ is *finitely identifiable* in state $x \in X$ in a learning frame $(X, \mathcal{O}, L)$ if there is some $O \in \mathcal{O}$ with $x \in O$ such that $O \subseteq P$. In other words, it there is some *true* observation that entails P. Note that finite identifiability is *learner-independent* (i.e. it does not make reference to conjectures of the form $L(O)$) and hence depends only on the underlying intersection space $(X, \mathcal{O})$.

By contrast, identifiability in the limit *is* dependent on the learner. The intuition can be described as follows. Certain properties, although true, cannot be known with certainty by any finite sequence of observations $O_1, \dots O_n \in \mathcal{O}$ (which induces a single observation $O_1 \cap \dots \cap O_n \in \mathcal{O}$). In these scenarios, the learner has to go 'beyond' the observations, making conjectures she is not certain of in order to converge to the truth in the limit. Note that still, identifiability in the limit is a non-introspective notion. That is, the learner never becomes aware (or certain) that the property is true, although evidence accumulates over time.

Say that a learner L *believes* P based on an observation $O \in \mathcal{O}$ if $L(O) \subseteq P$. Further, a learner L has an *undefeated belief* in P based on observation $O \in \mathcal{O}$ if L believes P (i.e. $L(O) \subseteq P$) and no further future observation $O \in \mathcal{O}$ may refute this belief, i.e. $\forall O' \in \mathcal{O}\ L(O \cap O') \subseteq P$. Note these notions of belief are not defined relative to a state, because beliefs are not required to be true (whereas knowledge *is* factive). Finally, we say that a learner L *identifies a property P in the limit* in a state $x \in X$ if there is a true observation $O \in \mathcal{O}$ with $x \in O$ such that L has an undefeated belief in P based on O which is true, i.e. with $x \in L(O)$.

Finite characterisations and exact learning as discussed in section 2.2 are related to these notions in the following way. First of all, note that an exact learning algorithm tries to learn a concept with certainty (in contrast to statistical approaches in computational learning theory such as PAC-learning), purely based on observations. So an exact learning algorithm is not a 'learner' in the sense above (or a very trivial one) as it *does not go beyond the data* by making conjectures. This is reflected in the observation that concept classes can be seen as intersection spaces rather than learning frames.

A concept class $\mathcal{C}$ over a domain D is a subset of $\mathcal{P}(D)$ and hence gives rise to an intersection space in the following way. First, we view $\mathcal{C}$ as our hypothesis space. Note that this obfuscates the structure of concepts as subsets of some uniform domain D, instead rendering them as abstract elements of some set. However, the concrete nature of these concepts is used to define the observations. Intuitively, we may observe an instance $d \in D$ of the domain either as a positive example of a concept, or as a negative example, giving rise to the observations $[d]^+ = \{C \in \mathcal{C} \mid d \in C\}$ and $[d]^- = \{C \in \mathcal{C} \mid d \in C\}$ respectively. Let $\mathcal{O}_D$ be the closure of $\{[d]^+ \mid d \in D\} \cup \{[d]^- \mid d \in D\}$ under finite intersections, then $(\mathcal{C}, \mathcal{O}_D)$ is an intersection space. because $\mathcal{O}_D \subseteq \mathcal{P}(\mathcal{C})$ and closed under finite intersections.[7] The converse does not hold: not every intersection space $(X, \mathcal{O})$ can be seen as a concept class since the hypotheses in X need not possess any further internal structure.

Consequently, given the above correspondence between concept classes and special types of intersection spaces, we can show the following. Say that an intersection space $(X, \mathcal{O})$ is *learnable* if every hypothesis $x \in X$ is an isolated point in the observational topology, i.e. if each singleton $\{x\}$ is finitely identifiable.

[5] These are called *learnability with certainty* and *inductive learnability* respectively in [4].

[6] In analogy with dynamic epistemic logic where notions of knowledge and belief are also defined relative to a "real" world. This is relevant since a multi-agent epistemic Kripke model may contain multiple disjoint 'information cells' for agents, representing incompatible information states.

[7] Alternatively, taking $\{[d]^+ \mid d \in D\} \cup \{[d]^- \mid d \in D\}$ as a subbasis generates the observational topology.

Proposition 2.2. *A concept class $\mathcal{C}$ over a domain D is finitely characterisable iff the corresponding intersection space $(\mathcal{C}, \mathcal{O}_D)$ is learnable.*

Proof. From left to right, suppose that $\mathcal{C}$ is finitely characterisable and let $C \in \mathcal{C}$. We show that $\{C\}$ is an intersection of finitely many observations $O_1, \ldots, O_n \in \mathcal{O}$ and hence open in the observational topology. By hypothesis, C has a finite characterisation $\mathbb{E} = \{d_0^+, \ldots, d_n^+, e_0^-, \ldots, e_m^-\}$ w.r.t $\mathcal{C}$, where the + and - labels denote being a positive or negative example respectively and each $d_i \in D$. It immediately follows that

$$[d_0]^+ \cap \ldots \cap [d_n]^+ \cap [e_0]^- \cap \ldots \cap [e_m]^- = \{C\}$$

i.e. $\{C\}$ is a finite intersection of basic open sets $[d_i]^+, [e_j]^- \in \mathcal{O}_D$. Conversely, if $\{C\}$ is open it must be a finite intersection of opens $[d_0]^+, \ldots, [d_n]^+, [e_0]^-, \ldots, [e_m]^- \in \mathcal{O}_D$. It follows that $(\{d_0^+, \ldots, d_n^+\}, \{e_0^-, \ldots, e_m^-\})$ is a finite characterisation on C w.r.t. $\mathcal{C}$. □

2.3 Categories and Dualities

In this section, we introduce the bare category theory needed to get to the crucial concepts of frontier and duality.

Definition 2.20. (Category)
A category $\mathbb{C}$ consists of a collection $Ob(\mathbb{C})$ of objects and a collection $Ar(\mathbb{C})$ of morphisms such that every morphism f has a domain $dom(f)$ and codomain $cod(f)$ in $Ob(\mathbb{C})$. Moreover, these have to satisfy the following conditions:

- given two morphisms f, g with $cod(f) = dom(g)$, their *composition* $g \circ f$ (also abbreviated as gf) exists as a morphism where $dom(gf) = dom(f)$ and $cod(gf) = cod(g)$
- for every object A there is a unique morphism $1_A \in Ar(\mathbb{C})$ (with $dom(1_A) = cod(1_A) = A$) that is the *identity* on A, i.e. given morphisms f, g with $cod(f) = A = dom(g)$, we have $1_A \circ f = f$ and $g \circ 1_A = g$.
- composition is associative, i.e. for every $f, g, h \in Ar(\mathbb{C})$, we have that $(f \circ g) \circ h = f \circ (g \circ h)$

We will simply write $A \to B$ to denote the existence of a morphism from object A to object B. For every category $\mathbb{C}$, its *opposite category* $\mathbb{C}^{op}$ is defined as the category with the same objects as $\mathbb{C}$, but with all morphisms reversed.

Classes of mathematical objects with an appropriate notion of homomorphism or 'structure-preserving map' give rise to categories. In particular, there is a category of all categories **Cat**. The morphisms in this category are called *functors*. Thus, a functor is a morphism between categories, and hence is specified by an 'object part' and a 'morphism part'.

Definition 2.21. (Functors)
A *functor* F from a category $\mathbb{C}$ to $\mathbb{D}$ (notation $F : \mathbb{C} \to \mathbb{D}$) is a map $F_{ob} : Ob(\mathbb{C}) \to Ob(\mathbb{D})$ together with a map $F_{Ar} : Ar(\mathbb{C}) \to Ar(\mathbb{D})$ such that (i) for every morphism $C \stackrel{f}{\to} C'$ in $\mathbb{C}$, $FC \stackrel{F(f)}{\to} FC'$ is a morphism in $\mathbb{D}$. Moreover, we also require that $F(f \circ g) = F(f) \circ F(g)$ and $F(1_C) = 1_{FC}$. For every functor $F : \mathbb{C} \to \mathbb{D}$, the *opposite functor* $F^{op} : \mathbb{D} \to \mathbb{C}$ exists (as a morphism in **Cat**).

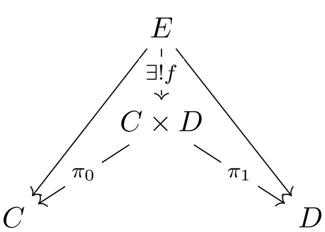
The identity functor $1_{\mathbb{C}}$, which is just the identity on both objects and morphisms of $\mathbb{C}$, is the identity map in **Cat** for a category $\mathbb{C}$. A functor F is a *contravariant* functor if $F : \mathbb{C} \to \mathbb{D}^{op}$ or $F : \mathbb{C}^{op} \to \mathbb{D}^{op}$. A functor $F : \mathbb{C} \to \mathbb{C}$ with the same domain and codomain is what is known as an *endofunctor*.

Definition 2.22. An *isomorphism* f in a category $\mathbb{C}$ is a morphism $f : C \to D$ (where C, D objects in $\mathbb{C}$) that has an inverse $g : D \to C$ such that $gf = 1_C$ and $fg = 1_D$.

In particular, an isomorphism in **Cat** (an "isomorphism of categories") is a functor $F : \mathbb{C} \to \mathbb{D}$ with an inverse functor $G : \mathbb{D} \to \mathbb{C}$ such that $GF = 1_{\mathbb{C}}$ and $FG = 1_{\mathbb{D}}$.[8] A *dual isomorphism* of categories $\mathbb{C}, \mathbb{D}$ is a contravariant functor $F : \mathbb{C} \to \mathbb{D}^{op}$ that is an isomorphism.

Definition 2.23. (Initial and Final Objects)
The *initial object* in a category $\mathbb{C}$ is an object 0 in $\mathbb{C}$ such that there is a *unique* morphism $! : 0 \to C$ to every object C in $\mathbb{C}$. A *weakly initial object* is an object that has at least one morphism to every object in $\mathbb{C}$. The *terminal object* or a *weakly terminal object* in $\mathbb{C}$ is the initial object or a weakly terminal object in $\mathbb{C}^{op}$.

Definition 2.24. (Product and Coproduct)
The product of two objects C and D in a category $\mathbb{C}$ is an object $C \times D$ in $\mathbb{C}$ together with morphisms $\pi_0 : C \times D \to C, \pi_1 : C \times D \to D$ (the projection maps) such that for every object E with morphisms $E \to C, E \to D$, the following diagram commutes



A category $\mathbb{C}$ *has products* if for every two objects C, D, the product exists in $\mathbb{C}$. The *coproduct* of C and D in $\mathbb{C}$ is the product of C and D in $\mathbb{C}^{op}$. Similarly, a category $\mathbb{C}$ *has coproducts* iff $\mathbb{C}^{op}$ has products.

Note that we only defined binary products, but having binary products implies having all finite products. The following notion is only defined for categories that have products.

Definition 2.25. (Exponentials)
Let $\mathbb{C}$ be a category that has products. Then the *exponential* of two objects C and D in a category $\mathbb{C}$ is an object C^D such that for all objects A;

$$A \times D \to C \qquad \text{iff} \qquad A \to C^D$$

That is, exponentiation with an object D is defined as being right adjoint to taking the product with D. If a category has exponentials, we say it is *cartesian closed*. The following notions are not necessarily studied by category theorists, but they nevertheless make sense for arbitrary categories.

[8]Note that this is strictly stronger than what is known as an *equivalence of categories*, where there only have to be natural transformations (which are morphisms between functors) from the endofunctors GF, FG to the identity functors $1_{\mathbb{C}}, 1_{\mathbb{D}}$ respectively.

Definition 2.26. (Dualities)
A (finite) *duality* in a category $\mathbb{C}$ is a pair $(\mathcal{F}, \mathcal{D})$ of finite sets of objects such that

$$\mathcal{F} \rightarrow = Ob(\mathbb{C}) \setminus \rightarrow \mathcal{D}$$

where $\mathcal{F} \rightarrow := \{A \mid \exists F \in \mathcal{F} \text{ such that } F \rightarrow A\}$ and dually $\rightarrow \mathcal{D} := \{A \mid \exists D \in \mathcal{D} \text{ such that } A \rightarrow D\}$.

Hence, a duality in $\mathcal{C}$ partitions $Ob(\mathbb{C})$ into the collection of objects 'above $\mathcal{F}$' and those 'beneath $\mathcal{D}$'. We will refer to $\mathcal{D}$ as the *duals* for $\mathcal{F}$. Note that this is called a *generalised duality* in [18].If $\mathcal{F} = \{A\}$ consists of a single object and is the left-hand side of a duality, then we say that A *has finite duality* [9]. Dualities can be viewed as a generalisation of the notion of a *splitting* of a lattice [28]. Also closely related to dualities is the notion of a *frontier* [9].

Definition 2.27. (Frontiers)
A (finite) *frontier* of an object A is a finite set of objects $\mathcal{F}$ such that (i) for each $F \in \mathcal{F}$, $F \rightarrow A$ and $A \not\rightarrow F$ and (ii) for every B with $B \rightarrow A$ and $A \not\rightarrow B$, $B \rightarrow F$ for some $F \in \mathcal{F}$.

Frontiers are a generalisation of '*gap pairs*' as studied in [23], which is just a pair of objects (A, F) such that F is 'strictly below' A (i.e. $F \rightarrow A$ and $A \not\rightarrow F$) while there is no other object B 'strictly between' A and F (i.e. no $B \neq F$ such that $F \rightarrow B \rightarrow A$). Thus if (A, F) is a gap pair then the ambient order of morphism is not 'dense' just below A, so F is a *minimal weakening* of A. Frontiers generalise this to a *set* of minimal weakenings of an object (cf. definition 2.27). So if $\{F\}$ is a singleton frontier for an object A then (A, F) is a gap pair.

Observe that dualities are defined for sets of object while frontiers are only defined for single objects. Yet the two notion are related in the case of single objects. The following results are from [9]. Although they only show this for a specific category their proof is completely general for arbitrary categories, as it just uses the universal properties of products and exponential object.[9]

Proposition 2.3. *([9]) If a category $\mathbb{C}$ has products and $(\{A\}, \mathcal{D})$ is a duality, then $\{A \times D \mid D \in \mathcal{D}\}$ is a frontier for A.*

Proposition 2.4. *([9]) In a cartesian closed category $\mathbb{C}$, if $\mathcal{F}$ is a frontier for an object A then $\{F^A \mid F \in \mathcal{F}\}$ is a duality for A.*

Proof. Suppose that $\mathcal{F}$ is a frontier for an object A. We have to show that for all B $B \rightarrow F^A$ for some $F \in \mathcal{F}$ iff $A \not\rightarrow B$. From left to right, note that if $B \rightarrow F^A$ then $A \times B \rightarrow F$. As F is a frontier, $A \not\rightarrow A \times B$ (otherwise $A \rightarrow F$ by composition) and hence as $A \rightarrow A$, by properties of the product it must be that $A \not\rightarrow B$. Conversely, if $A \not\rightarrow B$ then $A \not\rightarrow A \times B$. But as $A \times B \rightarrow A$ via the projection and $\mathcal{F}$ is a frontier, there must be some $F \in \mathcal{F}$ such that $A \times B \rightarrow F$. Hence $B \rightarrow F^A$. □

Note that the construction of the frontier from the duality is polynomial while the construction of the duality from the frontier requires an exponential blow-up in size [9].

[9] In fact, their case is a little different because the role of the exponential is played by a *set* of objects. This is a recurring theme through this thesis.

2.4 Finite Characterisations for Conjunctive Queries

In this thesis, we will study a particular kind of concept classes, namely logical languages. Whereas we will concentrate on *modal* languages, finite characterisations and exact learnability have been studied for fragments of first order languages. In particular, the *positive-existential fragment* of FO, also known as *unions of conjunctive queries*, was recently shown to admit finite characterisations [9]. Interestingly, they show this via the existence of frontiers and dualities in the category **Hom** of finite structures and homomorphisms. In the first order case, the examples will be structures with a tuple of designated elements rather than pointed Kripke models. For A a set, let A^* denote the set of all tuples of finite length consisting only of elements of A. A *schema* S (or relational signature[10]) is a tuple $(R_1, \ldots, R_n, c_1, \ldots, c_m)$ of relation symbols R_i (marked with their arities) and constant symbols c_j.

Definition 2.28. (Relational Structure)
Given a schema $S = (R_1, \ldots, R_n, c_1, \ldots, c_m)$, a ($S$-)*structure* is a tuple $(dom(A), R_1^A, .., R_n^A, c_1^A, \ldots, c_m^A)$ where $dom(A)$ is a set (the domain of the structure), for each relational symbol R_i, R_i^A is a set of tuples of the same length as the arity of R_i, and each c_i^A denotes a distinguished element in $dom(A)$.

When the schema contains constant symbols, we write $(A, \bar{a})$ for a structure together with the tuple of 'distinguished elements' $\bar{a} \in dom(A)^*$ picked out by the constants. We refer to the tuples in the sets R_i^A as *facts*. We say that a structure is *safe* if every distinguished element occurs in some fact. A structure A, a with one distinguished element is a *tree* if viewed as a Kripke model it satisfies the definition of a tree (i.e. definition 2.11).

Definition 2.29. (Conjunctive Queries)
A *conjunctive query* (CQ) is a formula of the form

$$q(\bar{x}) := \exists \mathbf{y} \phi(\bar{x}, \mathbf{y})$$

where $\phi(\bar{x}, \bar{y})$ is a conjunction of atomic formulas $R\bar{z}$ such that every $z \in \{\bar{x, y}\}$ occurs in some atomic conjunct. For a CQ $q(\bar{x})$, the variables $\bar{x}$ are the free variables, and so if $|\bar{x}| = k$ then we say that q is a k*-ary* CQ.

CQs can be seen as a normal form for the positive-existential-conjunctive (i.e. no negation, disjunction or universal quantification) fragment of FO.

Definition 2.30. (Homomorphisms)
An (S-)*homomorphism* $h : (A, \bar{a}) \to_{hom} (B, \bar{a})$ between S-structures is a function $h : dom(A) \to dom(B)$ such that for all tuples $\bar{a'} \in dom(A)^*$, $A \models R(\bar{a})$ implies $B \models R(h(\bar{a}))$.

We write $\leftrightarrow_{hom}$ for the relation of *homomorphic equivalence*, i.e. $A, \bar{a} \leftrightarrow_{hom} B, \bar{b}$ means that there is a homomorphism in both directions between these structures $A, \bar{a} \to_{hom} B, \bar{b}$ and $B, \bar{b} \to_{hom} A, \bar{a}$. It turns out that every structure contains a unique smallest substructure (up to isomorphism), called its *core*, to which it is homomorphically equivalent [23, 9].

Theorem 2.2. *([23]) Every structure $A, \bar{a}$ has a unique smallest substructure $core(A), \bar{a}$ (up to isomorphism) such that $A, \bar{a} \leftrightarrow_{hom} core(A), \bar{a}$.*

[10] We will not consider function symbols here.

Theorem 2.3. *([25]) Positive-existential FO (i.e. the class of all first order formulas using only $\wedge, \vee, \exists$ and positive atoms $R\bar{x}$) is preserved under homomorphisms, i.e. if $A, \bar{a} \models \varphi$ and $A, \bar{a} \rightarrow_{hom} B, \bar{b}$ then $B, \bar{b} \models \varphi$ for all positive-existential φ.*[11]

The above theorem of Lyndon's justifies our thinking of homomorphism as 'structure preserving'. Above, homomorphisms are closed under composition, and this composition is associative (because it is just functional composition). So each schema S induces a category $\mathbf{Hom}_S$ (we will suppress the schema-subscript mostly) of finite S-structures and S-homomorphisms.[12] The properties of these categories, in particular the cases of directed and undirected graphs, have been extensively studied in [23]. These categories **Hom** have been shown to have products, coproducts and a final object.

Direct products of structures fulfills the role of product, but the coproduct is not disjoint union and a little more involved [9]. Moreover, the single point structure M_{final} with $dom(M_{\text{final}}) = \{\cdot\}$ and all facts being true of that point (i.e. $R^{M_{\text{final}}} := \{\cdot\}^k$ for all k-ary relations R) serves as a final object [9]. We observe that, depending on whether we allow the empty structure to be well-defined, the initial object is either the empty structure (like $\emptyset$ is the initial object in the category **Set** of sets and (total) functions), or the single point structure $M_{initial}$ with $dom(M_{\text{initial}}) = \{\cdot\}$ and no facts holding of that point (i.e. $R^{M_{\text{initial}}} = \emptyset$ for all relations R). There is a close correspondence between CQs and finite structures that is well known as the *Chandra-Merlin theorem* (here we state it as a generalisation to the case with constants in [9]).

Definition 2.31. (Canonical Structure)
The *canonical structure* $\widehat{q}$ of a CQ q is the finite structure $(\widehat{q}, \bar{x})$ (which we also simply write as $\widehat{q}$) with domain the set of variables occurring in φ (which is a superset of the free variables) and a constant symbol x for each free variable $x \in \bar{x}$ that is interpreted as itself. Finally, a fact $R\bar{y}$ holds in $\widehat{q}$ iff it is a conjunct of q.

Conversely, every safe finite structure is the canonical structure of some CQ. That is, for a safe structure $A, \bar{a}$, let $q_{A,\bar{a}}$ be the *canonical query* of that structure, for which we have $\widehat{q_{A,\bar{a}}} = A, \bar{a}$.[13]

Theorem 2.4. *(Chandra-Merlin)([9, 12]) For a CQ $q(\bar{x})$,*

$$A, \mathbf{a} \models q \quad \textit{iff} \quad (\widehat{q}, \bar{x}) \rightarrow_{hom} (A, \mathbf{a})$$

Note that it follows that for any two CQs q, q' over the same free variables, $\widehat{q} \rightarrow_{hom} \widehat{q'}$ iff $q \rightarrow q'$ and hence $\widehat{q}, \widehat{q'}$ are *homomorphically equivalent* iff q, q' are logically equivalent. Dalmau and ten Cate have shown that there is an intimate connection between the existence of finite characterisations for CQs and the existence of frontiers for finite structures in **Hom**. For the following theorems to make sense, we first formally introduce CQs as a concept class.

[11] In fact, [25] also proves the converse, i.e. that a homomorphism-preserved first order formula is equivalent to a positive-existential one.

[12] Dalmau and ten Cate refer to it the *Homomorphism Lattice* [9]. The restriction to finite models is motivated by applications in e.g. database theory. Allowing structures of arbitrary size still results in a category, but it is not known whether the categorical constructs that have been shown to exist in this category transfer even to the countable case.

[13] As noted in [9], the restriction of this correspondence to safe structures is inconsequential when considering duals and frontiers (cf lemmas 4.3 and 4.4).

Definition 2.32. Parametric on an arity k and a schema S, the class of all k-ary CQs using only symbols from the schema S forms a concept class. Formally, let $\mathcal{C}_{S,k}^{CQ}$ be the class of all concepts

$$mod(q) := \{(A, \bar{a}) \mid A, \bar{a} \models q\}$$

where $q(\bar{x})$ is a k-ary CQ in schema S. Note that each $(A, \bar{a}) \in mod(q)$ has exactly k distinguished elements.

Theorem 2.5. *([9]) A k-ary CQ has a finite characterisation w.r.t. the class of all k-ary CQs iff it has a frontier.*

Proof. If (E^+, E^-) is such a characterisation then $F = \{\widehat{q} \times (A, \bar{a}) \mid (A, \bar{a}) \in E^-\}$ is a frontier for $\widehat{q}$, while if $\mathcal{F}$ is a frontier for $\widehat{q}$ then $(\{\widehat{q}\}, \mathcal{F})$ is a finite characterisation of q. □

Building on work in structural graph theory on the existence of frontiers [17], the authors give a complete characterisation of the finitely characterisable CQs (w.r.t. the class of all CQs of the same arity). As it turns out, not all of them are finitely characterisable, only the ones satisfying the following criterion which is defined in terms of the canonical structure.

Definition 2.33. (c-acyclicity)
A CQ $q(\bar{x})$ is *c-acyclic* if its canonical structure $\widehat{q}$ is. That is, if every cycle through the incidence graph[14] of $\widehat{q}$ passes through some element $x_i \in \bar{x}$ that is named by a constant symbol. Note that this precludes an element that is not named by a constant symbol to occur twice in the same fact. Note that c-acyclicity is strictly weaker than acyclicity.

Theorem 2.6. *([17, 1]) A k-ary CQ has a frontier iff it is logically equivalent to a c-acyclic CQ.* [15]

One of the main results of [9] is a new proof of the right-to-left direction of theorem 2.6 which (unlike the original) provides a polynomial-time construction of a frontier from a c-acyclic structure. Hence, concatenating theorems 2.5 and 2.6, we obtain the following theorem.

Theorem 2.7. *([9]) A k-ary c-acyclic CQ has a finite characterisation w.r.t. the class of all k-ary CQs iff it is logically equivalent to a c-acyclic CQ. Moreover, for fixed k, the frontier of a k-ary c-acyclic CQ can be computed in polynomial time.* [16]

Through clever use of the frontiers obtained via theorem 2.7, the authors describe learning algorithms that can learn k-ary CQs that are logically equivalent to a c-acyclic one in polynomial time, that is learn to distinguish it from all other (i.e. inequivalent) k-ary CQs.

Theorem 2.8. *([9]) The class of c-acyclic k-ary CQs is polynomial-time exactly learnable with membership queries (w.r.t. the class of all k-ary CQs).*

[14] The incidence graph of a structure $(A, \mathbf{a})$ with k distinguished elements $\mathbf{a}$ is the bipartite multi-graph consisting of all elements and all facts of A, where an edge holds between an element and a fact if that element occurs in that fact.

[15] Note that this is a strictly weaker condition that being c-acyclic.

[16] Note that this does not imply that the frontier has polynomial *size* (i.e. only consist of polynomially many structures). Indeed, in general the size of the smallest frontier is already exponential in the number of distinguished elements [9].

The algorithm maintains a c-acyclic hypothesis CQ that implies the hidden goal CQ. Via the Chandra-Merlin correspondence, this exact learning algorithm that uses only membership queries is able to encode hypotheses or 'guesses' as membership queries. Namely, the (safe) structure it present to the membership oracle is the the canonical structure of the CQ that is the new hypothesis or guess. The initial guess is the final object in **Hom**, which satisfies all CQs. At each iteration $i+1$, the algorithm does the following.

It computes the frontier $\mathcal{F}$ of the previous guess (i.e. the previous structure presented to the oracle) and for each $(F, \bar{f}) \in \mathcal{F}$ asks membership queries. If it exhausts all structures in the frontier without receiving a "yes" answer (as in "yes, the hidden goal CQ is true on this structure), it outputs the previous guess. Otherwise, the oracle answers "yes" on some $(F, \bar{f}) \in \mathcal{F}$. This structure may not be c-acyclic but there is a polynomial time algorithm for turning $(F, \bar{f})$ into a homomorphically equivalent structure that is c-acyclic, and this c-acyclic structure is its next guess (i.e. the next structure presented to the oracle). This algorithm is polynomial because at each iteration, the size of the hypothesis strictly increases [9].

Moreover, they also present a polynomial time construction of frontiers for the special class of unary c-connected acyclic CQs. Although we will not make us of this construction in our proofs, we mention it here to comment on it later. Recall that a CQ q is acyclic and connected iff its canonical structure $\widehat{q}$ is.

Theorem 2.9. *([9]) For every unary connected acyclic CQ q, a frontier w.r.t the class of all unary CQs, consisting solely of connected acyclic structures with one distinguished element, can be constructed in polynomial time.*

2.4.1 Unions of Conjunctive Queries

Although the work in [9] focuses on conjunctive queries, results in it and in earlier work on learning schema mappings [1] imply characterisation and learnability results for the larger class of *unions of conjunctive queries* (UCQs).

Definition 2.34. A *union of conjunctive queries* (UCQ) is a disjunction $q_0 \vee \ldots \vee q_n$ of CQs such that $q_i(\bar{x})$, i.e. $var(q_i) = var(q_j) = \{\bar{x}\}$ for all $i, j \leq n$.

UCQs can be seen as normal forms for positive-existential first-order logic. If ϕ is positive-existential, rewrite it to prenex normal form by pulling out all existential quantifiers $\exists$ and distribute $\wedge$'s over $\vee$'s via the propositional distributive law $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$; the result is a UCQ. Similarly to the case of CQs, the collection of all k-ary UCQs in a schema S also induces a concept class $\mathcal{C}^{UCQ}_{S,k}$, defined by taking all concepts

$$mod(q_0 \vee \ldots \vee q_n) := \{(A, \bar{a}) \mid A, \bar{a} \models q_0 \vee \ldots q_n\} = \bigcup_{i \leq n} mod(q)$$

where $q_1 \vee \ldots \vee q_n$ is a k-ary UCQ in schema S. Moreover, to each UCQ $q_1 \vee \ldots \vee q_n$ we can associate the finite set $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ of canonical structures of its disjunct CQs. The following generalisation of the Chandra-Merlin theorem to UCQs is an immediate consequence of the original result.

Theorem 2.10. *For a k-ary UCQ $Q = q_1 \vee \ldots \vee q_n$ and structure $A, \bar{a}$ with k distinguished elements*

$$A, \bar{a} \models Q \qquad \textit{iff} \qquad \widehat{q_i} \rightarrow_{hom} A, \bar{a} \qquad \textit{for some } \widehat{q_i} \in \{\widehat{q_1}, \ldots, \widehat{q_n}\}$$

We get the following analogue of theorem 2.5 relating characterisations for UCQs with the existence of *dualities* rather than frontiers. We call a duality in **Hom** a *homomorphism duality.*

Theorem 2.11. *(implicit in [1]) A k-ary UCQ $q_1 \vee \ldots \vee q_n$ is finitely characterisable w.r.t. the class of all k-ary UCQs iff $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ is the left-hand side of a homomorphism duality.*

Proof. **From left to right**, we show that if $(\{\widehat{q_1}, \ldots, \widehat{q_n}\}, \mathcal{D})$ is a homomorphism duality then it is also a finite characterisation of Q w.r.t. the class of all UCQs. For readability set $E^+ := \{\widehat{q_1}, \ldots, \widehat{q_n}\}$ and $E^- := \mathcal{D}$. In fact, we show that for a UCQ $Q' = q'_1 \vee \ldots \vee q'_m$ (over the same free variables):

(i) If Q' fits E^+ then $Q \to Q'$

(ii) If Q' fits E^- then $Q' \to Q$

Clearly, if (i) and (ii) hold and Q' fits (E^+, E^-), then $Q \equiv Q'$. (i) If Q' fits E^+ then for each disjunct q_i of Q, we have $\widehat{q_i} \models Q'$, i.e. $\widehat{q_i} \models q'_j$ for some disjunct q'_j of Q'. By theorem 2.4 it follows that $\widehat{q'_j} \to_{hom} \widehat{q_i}$ and hence $q_i \to q'_j$ and thus $q_i \to Q'$. Because this holds for each disjunct q_i, it follows that $Q \to Q'$. (ii) Now suppose that Q' fits E^-. That is, for all $D \in \mathcal{D}$, $D \not\models Q'$. To show that $Q' \to Q$, let $A, \bar{a} \models Q'$ be a model of Q'. Then observe that it cannot be that $A, \bar{a} \to_{hom} D$ for some $D \in \mathcal{D}$, otherwise $D \models Q \wedge \neg Q'$ by preservation of UCQ under homomorphism (2.3). But as $\mathcal{D}$ is the dual of $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$, there must be some disjunct q_i of Q such that $\widehat{q_i} \to_{hom} A, \bar{a}$. Hence, as $\widehat{q_i} \models Q$, by homomorphism-preservation (2.3) $A, \bar{a} \models Q$.

From right to left, suppose that (E^+, E^-) is a finite characterisation of Q w.r.t. the class of all UCQs. The claim is that $(\{\widehat{q_1}, \ldots, \widehat{q_n}\}, E^-)$ is a homomorphism duality. We have to show that for any structure $A, \bar{a}$, $\widehat{q_i} \not\to_{hom} A, \bar{a}$ for all disjuncts q_i of Q iff $A, \bar{a} \to_{hom} E, e$ for some $E, e \in E^-$. [From right to left] If $A, \bar{a} \to_{hom} E, \bar{e}$ for some negative example $(E, \bar{e}) \in E^-$ then it cannot be that $A, \bar{a} \models Q$ for then $E, e \models Q \wedge \neg Q$ would follow by preservation, but this is a contradiction. Hence $A, a \not\models Q$ so it is a negative example for Q. But then by theorem 2.10 it follows that $\widehat{q_i} \not\to_{hom} A, \bar{a}$ for all disjuncts q_i of Q. [From left to right] Suppose that $\widehat{q_i} \not\to_{hom} A, \bar{a}$ for all disjuncts q_i of Q then by theorem 2.10 $A, \bar{a} \not\models Q$, i.e. $A, \bar{a} \not\models q_i$ for some disjunct q_i of Q. We want to show that $A, \bar{a} \to_{hom} E, \bar{e}$ for some $(E, \bar{e}) \in E^-$. By lemma 4.4 of [9], we may suppose that $(A, \bar{a})$ is a safe structure and hence that it corresponds to a canonical CQ $q_{A,\bar{a}}$ with $\widehat{q_{A,\bar{a}}} \leftrightarrow_{hom} A, \bar{a}$. Consider the UCQ $Q^+ = q_1 \vee \ldots \vee q_n \vee q_{A,\bar{a}}$. Note that Q^+ fits $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ and hence $Q \to Q^+$ by theorem 2.10. By extension, it follows that Q^+ fits E^+. But as $\widehat{q_{A,\bar{a}}} \models Q^+ \wedge \neg Q$ we have $Q \not\equiv Q^+$. Hence, as (E^+, E^-) is a finite characterisation of Q w.r.t. the class of all UCQs, there must be some negative example $(E, \bar{e}) \in E^-$ such that $E, \bar{e} \models Q^+ \wedge \neg Q$, i.e. $E, \bar{e} \models q_{A,\bar{a}} \wedge \neg q_1 \wedge \ldots \wedge \neg q_n$. By theorem 2.4 it follows that $\widehat{q_{A,a}} = A, \bar{a} \to_{hom} E, \bar{e}$ where $(E, \bar{e}) \in E^-$, so we have a duality. □

In fact, we can prove the analogue of theorem 2.6 and the full theorem 2.7 for UCQs. Say that UCQ is c-acyclic if each CQ-disjunct is c-acyclic.

Theorem 2.12. *A k-ary UCQ is finitely characterizable w.r.t. the class of all k-ary UCQs iff it is logically equivalent to a c-acyclic UCQ. Moreover, for a c-acyclic UCQ this characterisation can be computed in single exponential time.*

Proof. **From right to left:** Let $Q = q_1 \vee \ldots \vee q_n$ be a c-acyclic UCQ, i.e. each disjunct q_i is c-acyclic. By theorem 2.6, each $\widehat{q_i}$ has a frontier that can be computed in polynomial time. It follows that $\{\widehat{q_i}\}$ is the left-hand side of a homomorphism duality [9]. However, note that their construction of the dual from the frontier is slightly different (namely, bigger) because in **Hom**, only a *set* of objects can play the role of the exponential. Let $\mathcal{D}_i$ be the set of duals for each $\{\widehat{q_i}\}$, it follows that

$$\mathcal{D} := \{(D_1, \bar{d_1}) \times \ldots \times (D_n, \bar{d_n}) \mid (D_i, \bar{d_i}) \in \mathcal{D}_i \text{ for all } 1 \leq i \leq n\}$$

can serve as the set of duals for $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$. So we need to show that $\widehat{q_i} \not\to_{hom} A, \bar{a}$ for all disjuncts q_i of Q iff $A, \bar{a} \to_{hom} D, \bar{d}$ for some $(D, \bar{d}) \in \mathcal{D}$. From left to right, if $\widehat{q_i} \not\to_{hom} A, \bar{a}$ for all disjuncts q_i then by the dualities $(\{\widehat{q_i}\}, \mathcal{D}_i)$ there are examples $(D_i, \bar{d_i}) \in \mathcal{D}_i$ for all $1 \leq i \leq n$ such that $A, \bar{a} \to_{hom} D_i, \bar{d_i}$. Hence $A, \bar{a} \to_{hom} (D_1, \bar{d_1}) \times \ldots \times (D_n, \bar{d_n}) \in \mathcal{D}$ by properties of the product. Conversely, $A, \bar{a} \to_{hom} (D_1, \bar{d_1}) \times \ldots \times (D_n, \bar{d_n}) \in \mathcal{D}$ implies that $A, \bar{a} \to D_i, \bar{d_i}$ and hence $\widehat{q_i} \not\to_{hom} A, \bar{a}$ for each i. We have just shown that $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ is the left-hand side of a homomorphism duality, so by theorem 2.11 Q is characterised by $(\{\widehat{q_1}, \ldots, \widehat{q_n}\}, \mathcal{D})$. It follows from the results in [9] that the dual $\mathcal{D}_i$ of a each $\widehat{q_i}$ be computed in time singly exponential in $|q_i|$, first performing a polynomial time construction of the frontier of $\widehat{q_i}$ from and then performing an exponentiation construction for structures with designated elements (which is in general a *set* of structures with designated elements). Thus, if we set $|Q| = |q_1 \vee \ldots \vee q_n| = |q_1| + \ldots + |q_n|$, the set of duals for $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ $\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_n = \{(D_1, \bar{d_1}) \times \ldots \times (D_n, \bar{d_n}) \mid (D_i, \bar{d_i}) \in \mathcal{D}_i \text{ for all } 1 \leq i \leq n\}$ can also computed in single exponential time as $2^{|q_1|} \cdot \ldots \cdot 2^{|q_n|} = 2^{|q_1| + \ldots + |q_n|}$.

From left to right: Let $\mathcal{D}$ be the right-hand side of the duality. If $q_i \to q_j$ for some disjuncts q_i, q_j of Q, let Q' be the UCQ obtained from Q by removing the disjunct q_j, and note that $Q \equiv Q'$. Hence we may assume w.l.o.g. that none of the disjuncts of Q imply each other and hence that $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ is an antichain in **Hom** via theorem 2.4. It suffices to construct frontier for each disjunct CQ q_i, for then by theorem 2.6 each disjunct CQ q_i and hence the UCQ $q_1 \vee \ldots \vee q_n$ will be c-acyclic. Set

$$\mathcal{F}_i := \{\widehat{q_i} \times (D, \bar{d}) \mid (D, \bar{d}) \in \mathcal{D}\}$$

and show that $\mathcal{F}_i$ is a frontier for $\widehat{q_i}$. Let $(F, \bar{f}) := \widehat{q_i} \times (D, \bar{d})$. First of all, observe that $\pi_0 : (F, \bar{f}) \to_{hom} \widehat{q_i}$ for all $(F, \bar{f}) \in \mathcal{F}_i$ via the projections. Moreover, $\widehat{q_i} \not\to_{hom} (D, \bar{d})$ by the fact that $(\{\widehat{q_1}, \ldots, \widehat{q_n}\}, \mathcal{D})$ is a duality and hence $\widehat{q_i} \not\to_{hom} (F, \bar{f})$ by properties of the product. Finally, let $A, \bar{a}$ be a structure 'strictly below' $\widehat{q_i}$ in the homomorphism order, i.e. $A, \bar{a} \to_{hom} \widehat{q_i}$ and $\widehat{q_i} \not\to_{hom} A, \bar{a}$. But it cannot be that $\widehat{q_j} \to_{hom} A, \bar{a}$ because then it would follow that$\widehat{q_j} \to_{hom} \widehat{q_i}$, contrary to our hypothesis that $\{\widehat{q_1}, \ldots, \widehat{q_n}\}$ is an antichain. Hence it must be that $A, \bar{a} \to_{hom} D, \bar{d}$ for some $(D, \bar{d}) \in \mathcal{D}$ by the duality $(\{\widehat{q_1}, \ldots, \widehat{q_n}\}, \mathcal{D})$. But note that $(F, \bar{f}) := \widehat{q_i} \times (D, \bar{d}) \in \mathcal{F}_i$ with $A, \bar{a} \to_{hom} F, \bar{f}$, so each $\mathcal{F}_i$ is a frontier for each $\widehat{q_i}$.[17] □

Because the characterisation of a UCQ Q can be computed in time single exponential in $|Q|$, the naive learning algorithm (with membership queries only) that this induces already gives us a singly exponential time upper bound.

Theorem 2.13. *The class of c-acyclic UCQs is exact learnable in exponential time.*

Proof. As noted in the preliminaries, finite characterisability always induces a naive learning algorithm with membership queries. It works as follows. First, we list all the UCQs

[17] The proof essentially uses same insights as the connection between finite maximal antichains and generalised dualities (recall this is what we cal a duality) in **Hom** established in [18]. However, we create the right-hand side of a generalised duality for a set in terms of *generalised dualities* of its element, while [18] generates the right-hand side of a generalised duality in terms of *duality pairs* of its elements, where a duality pair is generalised duality where both the left and right-hand side are singletons.

$Q_0, Q_1, Q_2, \ldots$ in order of increasing size, i.e. $|Q_i| < |Q_j|$ whenever $i < j$. Then, for each Q_i in this list, we compute the finite characterisation of Q_i in time exponential time in $|Q_i|$ via theorem 2.12. It follows that for each i with $|Q_i| < |Q|$ (i.e. Q_i comes before Q in our list), we can test whether $Q \equiv Q_i$ in time exponential in $|Q_i|$ and hence in time exponential in $|Q|$. Moreover, note that there are at most $2^{\mathcal{O}(|Q|)}$ UCQs smaller than Q.[18] It follows that we can learn in Q in time $2^{\mathcal{O}(|Q|)}$. □

Allowing equivalence queries has much impact; it enables faster learning algorithms as well as the possibility of learning non c-acyclic UCQs. This does not contradict theorem 2.12 because finite characterisability is only a necessary precondition for exact learning with membership queries only.

Theorem 2.14. *([1, 9]) The class of all UCQs is polynomial time exact learnable with membership and equivalence queries.*

[18]There are 2^n bit of length n so at most $k^n = 2^{n \log k}$ strings of length n in our representation system, where $k = |\Sigma|$ is the size of the alphabet of the representation.

Chapter 3

Characterising the Full Modal Language over Frame Classes

In this chapter, we will study finite characterisations (and exact learnability) in the context of modal logic. Here, our concepts are (pointed) model-classes of modal formulas.

Definition 3.1. (Modal Concept Class)
A normal modal logic L over a finite set of propositional variables Prop induces a *modal concept class* $\mathcal{C}_{L,\text{Prop}}$, which is the collection of all concepts

$$mod^L(\varphi) := \{(M, s) \text{ pointed model} \mid M, s \models \varphi \text{ and } M \text{ is based on a } L\text{-frame}\}$$

where φ is a modal formula with $var(\varphi) \subseteq \text{Prop}$. We will suppress the superscript L when the ambient logic is clear from context.

We will be sloppy in notation and write, for a normal modal logic L, that L is finitely characterisable rather than that $\mathcal{C}_L$ is finitely characterisable. Modal concepts are concepts over the domain of pointed Kripke models. and hence we can only identify formulas up to logical equivalence through these concepts as $\varphi \equiv \psi$ iff $mod(\varphi) = mod(\psi)$. Thus, we obtain the following definition of what an *example* is in the context of modal concept classes.

Definition 3.2. (Modal Examples)
An *example* is a pointed model M, s that is labelled either positively (+) or negatively (-) (exclusive or). A modal concept $mod(\varphi)$ fits a positive example $(M, s)^+$ if $M, s \models \varphi$ and fits a negative example $(M, s)^-$ if $M, s \not\models \varphi$.

We will suppress the labels in notation for readability, and refer to 'positive and negative examples' instead. Moreover, we will be sloppy and refer to unlabelled models as examples as well. In principle, examples can be finite or infinite. However, in view of applications we would like to think of examples as finite objects, i.e. *finite models*. Let

$$mod^L_{\text{fin}}(\varphi) := \{M, s \text{ pointed model} \mid M, s \models \varphi \text{ and } M \text{ is based on a } \textit{finite } L\text{-frame}\}$$

and consider a normal modal logic L with the finite model property. Then we clearly havee

$$\varphi \equiv_L \psi \qquad \text{iff} \qquad mod^L_{\text{fin}}(\varphi) = mod^L_{\text{fin}}(\psi)$$

so it follows that $\mathcal{C}_L$ and $\mathcal{C}_L^{fin}$ are essentially the same concept classes.[1] All normal modal logics we consider in this thesis have the finite model property, and hence we will simply write e.g. $K, S4$ for the concept classes $\mathcal{C}_K^{\text{fin}}, \mathcal{C}_{K4}^{\text{fin}}$.

We show that few modal concept classes are finitely characterisable. First, we give explicit counterexample via formulas that can force arbitrary height. Then we show that a concept class C_L (or rather, C_L^{fin}) is finitely characterisable iff it contains finitely many concepts. That is, a normal modal logic L is finitely characterisable iff it is locally tabular. These negative results motivates our move to considering fragments of the full modal language in subsequent chapters.

3.1 Counterexamples Via Forcing Height

We show that the modal concept class induced by basic modal logic K is not finitely characterisable. We do this by showing that it contains formulas that can force arbitrary height. Our method is rather general, and we give a few corollaries.

Lemma 3.1. *There is a modal formula $height_n$ for each n such that $M, s \models height_n$ iff $height(M, s) = n$ for all pointed models M, s.*

Proof. Let $height_n := \Box^{n+1}\bot \wedge \Diamond^n\top$. Suppose that $M, s \models height_n$. Because M, s is rooted and $M, s \models \Box^{n+1}\bot$, it must be that M, s is acyclic, otherwise $M, s \models \Diamond^{n+1}\top$. Moreover, there cannot be paths of length $n+1$ starting from s so $height(M, s) \leq n$. But as $M, s \models \Diamond^n\top$, there is some path of length n starting at s and ending in some state $t \in dom(M)$, and hence $height(M, s) \geq n$. It follows that $height(M, s) = n$.

Conversely, suppose that $height(M, s) = n$. This means that there are no paths of length $n+1$ starting at s and hence $M, s \models \Box^{n+1}\bot$. On the other hand, there must be some path of length n starting at s and hence $M, s \models \Diamond^n\top$, so $M, s \models height_n$. ☐

Theorem 3.1. *Basic modal logic K is not finitely characterisable.*

Proof. It suffices to give one counterexample, say $\Box\bot$. Suppose for contradiction that $\Box\bot$ has a finite characterisation (E^+, E^-) w,r.t the full modal language. Set $n > \max\{height(E, e) \mid (E, e) \in E^+ \cup E^-\}$, and let $\psi := \Box\bot \vee height_n$, i.e. $\psi = \Box\bot \vee (\Box^{n+1}\bot \wedge \Diamond^n\top)$. We claim that ψ fits (E^+, E^-) yet $\Box\bot \not\equiv \psi$.

To see that ψ is fitting, note that it fits E^+ by properties of disjunction. But by choice of n, E^- cannot contain a tree of exactly height n (either the height is ∞ or some number $< n$), and hence $E, e \not\models height_n$ for all $(E, e) \in E^-$, so ψ also fits E^-. Finally, the n-length path model (pointed at the initial state) with empty valuation satisfies $\neg\Box\bot \wedge height_n$, and hence distinguishes ψ from $\Box\bot$. ☐

In fact, a far more radical statement holds.

Theorem 3.2. *No formula has a finite characterisation w.r.t. K.*

Proof. Let φ be a formula and suppose it has a finite characterisation (E^+, E^-) w.r.t. K. Observe that either (i) $\varphi \models \Box^k\bot$ for some k or (ii) $\varphi \wedge \Diamond^k$ is satisfiable for each k. In case (i), by contraposition we get that $\Diamond^k\top \models \neg\varphi$. Let $n > \max(\{|dom(E)| \mid (E, e) \in E^-\} \cup \{k\})$. Now take $\varphi' := \varphi \vee height_n$. By the same construction as before, φ' fits (E^+, E^-). Moreover,

[1] Alternatively, for a normal modal logic L that has the finite model property, there exist a finite characterisation of φ w.r.t. L iff there exists a finite characterisation of φ w.r.t. L consisting only of finite models.

the n-length path model P_n (pointed at the initial state) has height n so $P_n \models height_n$. But $k \leq n$ which means that $height_n \models \Diamond^k \top$ so by the first entailment $P_n \models \neg\varphi$, hence $P_n \models \neg\varphi \wedge \varphi'$, distinguishing φ from ψ.

In case (ii), let $n > \max(\{|dom(E)| \mid (E, e) \in E^-\} \cup \{d(\varphi)\})$, and consider the formula $\varphi' := \varphi \wedge \neg height_n$. Clearly φ' fits E^- by properties of conjunction. Further, by choice of n and lemma 3.1 above, all negative examples in E^- satisfy $\neg height_n$. But by hypothesis there is some pointed model M, s satisfying $\varphi \wedge \Diamond^n \top$. Now let $M \restriction n$ be the submodel of M induced the by the set of all n-successor of s (i.e. all states that can be reached by a path of length n). Clearly $height(M \restriction n, s) = n$ and thus $M \restriction n, s \models height_n$ (and hence $M \restriction n, s \models \neg\varphi'$) by 3.1. But we also have $M, s \leftrightarroweq_n M \restriction n, s$ and $d(\varphi) < n$, so it follows that $M \restriction n, s \models \varphi$, distinguishing φ from φ'. □

Remark. In fact, by inspection of the above proof we see that only frame formulas (i.e. formulas not containing any propositional variables but only the truth constants $\top$ and $\bot$) are used. Hence, it follows that no frame formula has a finite characterisation w.r.t. the frame language $C_{K,\emptyset}$.

3.2 Restricted Frame Classes

First of all, we observe that our previous method for the class of all frames still works over certain restricted frame classes.

Observation. All the steps in the proof of theorem 3.2 also go through when restricting to transitive frames. In particular, in case (ii) of the proof we cut-off the model at a designated height but transitivity is preserved under taking submodels. It follows that no formula has a finite characterisation w.r.t. $K4$. However, this reasoning does not extend to e.g. $S4$ in the presence of reflexivity.

More structurally, it turns out that there is a close connection between local tabularity and finite characterisability.

Definition 3.3. A normal modal logic L is *locally tabular* if for every finite set of propositional variables Prop, there are only finitely many formulas φ up to L-equivalence with $var(\varphi) \subseteq$ Prop.[2]

Thus if a normal modal logic L is locally tabular then $\mathcal{C}_L$ (as well as $\mathcal{C}_L^{\text{fin}}$) will only contains finitely many distinct concepts $mod(\varphi)$ (or $mod_{\text{fin}}(\varphi)$). But then a finite characterisation of one of these concepts would only have to distinguish it from finitely many others, which surely can be done with only finitely many examples! In fact, we show that the converse holds as well for normal modal logics. For the proof, we will need the following lemma.

Lemma 3.2. *Every locally tabular normal modal logic L has the finite model property.*

Proof. Let L be a locally tabular normal modal logic. Then for each finite subset $P \subseteq_{\text{fin}}$ Prop there are only finitely many formulas $\varphi_0, \ldots, \varphi_n$ with free variables contained in P (henceforth called P-formulas) up to L-equivalence. It follows immediately that there are only finitely many maximally consistent sets of P-formulas (where consistency means L-consistency), because there are only 2^n ways of satisfying some subset of $\{\varphi_0, .., \varphi_n\}$ and hence all P-formulas on L-frames. So let MCS_P be the *finite* set of all maximal consistent

[2]While we assume to be working with the full modal language here, this definition also makes sense for modal fragments as in definition 2.4.

sets of P-formulas. Then we the *weak canonical model* to be $M_P = (MCS_P, R^c, v^c)$, where the canonical relation and colouring are defined as usual, i.e. $R^c := \{(\Delta, \Sigma) \mid \forall \Box\varphi \in \Delta,\ \varphi \in \Sigma\}$ and $v^c(\Delta) := \{p \in \text{Prop} \mid p \in \Delta\}$. Note that $v^c(\Delta) \subseteq P \subseteq_{\text{fin}} \text{Prop}$ because each $\Delta \in MCS_P$ only contains formulas with free variables contained in P. The weak canonical model has the following properties:

(i) for each $\Delta \in dom(M_P)$, $M_p, \Delta \models \varphi$ iff $\varphi \in \Delta$

(ii) $M_P \models \varphi$ iff $\varphi \in L$

Clearly each L-satisfiable formula φ can be satisfied on this *finite* model M_P. So it rests to show that M_P is based on an L-frame, i.e. $(MCS_P, R^c) \models L$. Recall that a frame satisfies a formula if for all valuations on this frame, the corresponding model satisfies the formula. Let $v : MCS_P \to \text{Prop}$ be a valuation on the canonical frame, and $M_v := (MCS_P, R^c, v)$ the corresponding model. We want to show that for every $\varphi \in L$, $M_v \models \varphi$. We will need the following argument.

Claim: Every subset of the canonical frame (MCS_p, R^c) can be defined by a formula, relative to the canonical valuation.

Let $\{\Delta_0, \ldots, \Delta_1\}$ be a subset of the finite set MCS_P. Now for every pair of distinct $\Delta, \Sigma \in MCS_P$, there is a formula $\chi_{\Delta,\Sigma} \in (\Delta \setminus \Sigma) \cup (\Sigma \setminus \Delta)$ which by (i) means that either $\Delta \models \chi_{\Delta,\Sigma}$ and $\Sigma \not\models \chi_{\Delta,\Sigma}$ or vice versa. In either case, $\chi_{\Delta,\Sigma}$ distinguishes Δ from Σ in the weak canonical model M_P. It follows that $\psi_\Delta := \bigwedge_{\Sigma \in dom(M_P), \Sigma \neq \Delta} \chi_{\Delta,\Sigma}$ defines the singleton subset $\{\Delta\}$ in M_P and consequently $\psi := \bigvee_{i \leq m} \psi_{\Delta_i}$ defines the subset $\{\Delta_0, .., \Delta_m\}$ of M_P.

For every free variable $q_i \in \{q_1, \ldots, q_n\} = var(\varphi)$, consider the set $||q_i||_v = \{\Delta \in MCS_P \mid q_i \in v(\Delta)\}$, i.e. the set of states that make q_i true under v. By the claim above, this set is definable by a formula ψ_{q_i} relative to the canonical valuation, i.e. $||q||_v = ||\psi_{q_i}||_{v^c}$. Let $\varphi' := \varphi[\psi_{q_1}/q_1] \ldots [\psi_{q_n}/q_n]$. It follows that $M_P = (MCS_P, R^c, v^c) \models \varphi'$ iff $(MCS_P, R^c, v) \models \varphi$. But as L is a normal modal logic is closed under uniform substitution, $\varphi' \in L$ as well. But then it must be that φ' is contained in every maximal L-consistent set $\Delta \in MCS_P$, i.e. $M_P \models \varphi'$. □

Theorem 3.3. *A normal modal logic L is finitely characterisable iff it is locally tabular.*

Proof. From left to right, suppose that L is finitely characterisable. Then in particular $\bot$ must have a finite characterisation (E^+, E^-) w.r.t. the class of all modal formulas over L-frames. Note that this implies that all the examples must be based on L-frames. As $\bot$ is unsatisfiable, it must be that $E^+ = \emptyset$ and thus every formula that is satisfiable on an L-frame must be satisfied at some $(E_i, e_i) \in E^- = \{(E_1, e_1), \ldots, (E_n, e_n)\}$. But then there cannot be infinitely many pairwise inequivalent formulas $(\psi_i)_{i \in \omega}$ (that is, pairwise inequivalent over the class of L-frames) because that would require the existence of at least infinitely many distinct examples in E^-. To see this, note that there are only 2^n ways of being true/false on a subset of E^-. Hence there will be infinitely many formulas $\psi_{i_1}, \psi_{i_2}, \ldots$ that hold of the same subset of E^-, which means that e.g. the formula $(\psi_{i_1} \wedge \neg\psi_{i_2}) \vee (\neg\psi_{i_1} \wedge \psi_{i_2})$ is unsatisfiable on L-frames (for recall that every L-satisfiable formula must be satisfied at some example in E^-). But this contradicts the fact that $\varphi_{i_1} \not\equiv_L \varphi_{i_2}$, hence L must be locally tabular.

In the other direction, let L be a locally tabular logic. Then for every finite subset $P \subseteq_{\text{fin}}$ Prop there are only finitely many P-formulas $\varphi_0, \ldots, \varphi_n$ up to L-equivalence. Whenever $i \neq j$, we have $\varphi_i \not\equiv_L \varphi_j$ and hence the formula $(\varphi_i \wedge \neg\varphi_j) \vee (\neg\varphi_i \wedge \varphi_j)$ must be satisfiable on an L-frame. But by lemma 3.2, L has the finite model property and hence there is a *finite* pointed model E_j, e_j based on an L-frame that distinguishes φ_i from φ_j. Clearly then, the set $\{(E_j, e_j) \mid i \neq j, j \leq n\}$, appropriately labelled positive and negative, is a finite characterisation of φ_i w.r.t L. Hence L is finitely characterisable. □

This shows that a concept class C_L is finitely characterisable iff it contains only finitely many concepts. In fact, by lemma 3.2 it follows that any finitely characterisable C_L is 'equivalent' to C_L^{fin} in the sense that $mod_{\text{fin}}(\varphi)$ already contains all relevant information of $mod(\varphi)$. Given the above characterisation of finitely characterisable normal modal logics in terms of local tabularity, we can import the theory of syntactic characterisations of local tabularity to our setting.

Theorem 3.4. *[30, 26] A normal modal logic $L \supseteq K4$ is locally tabular iff it is of finite height, i.e. if it contains B_h for some h, where*

$$B_1 := p_1 \to \Box\Diamond p_1$$
$$B_{i+1} := p_{i+1} \to \Box(\Diamond p_{i+1} \vee B_i)$$

Corollary 3.1. *A normal modal logic $L \supseteq K4$ is finitely characterisable iff it is of finite height.*

Corollary 3.2. *$S5$ is finitely characterisable, and $S4, K4, KT$ are not finitely characterisable.*

Chapter 4

Weak Simulations

In the previous chapter, we completely characterised the finitely characterisable normal modal logics in terms of local tabularity. In subsequent chapters, we will be concerned with syntactic fragments of the full modal language obtained by restricting the base of connectives, i.e. modal fragments as defined in definition 2.4. In particular, given the role of negation in the construction of counterexamples for the full modal language, we will be focusing on *positive* modal languages. In this chapter, we will develop a categorical framework in analogy to the one in [23, 9] in order to obtain characterisability and learnability results for positive fragments (and beyond) in chapter 5.

4.1 Positive Modal Logic

Positive modal logic may mean various things. Originally, Dunn [15] proposed looking at the syntactic fragments $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\bot,\top}$ (and also $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\bot}$,$\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top}$ separately) of the full modal language and proved its completeness w.r.t. standard Kripke semantics as well as a representation theorem for the corresponding algebras (i.e. distributive lattices for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, and bounded distributive lattices for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\bot,\top}$).[1] However, in the absence of negation and implication (on which the classical and intuitionistic interpretation disagree), this fragment also allows for an intuitionistic semantics based on Kripke models with an additional ordering $\leq$ that coheres in a certain way with the accessibility relation [10].

The axiomatic system from [15] can also be extended by modal formulas expressing certain properties of the accessibility relation. For instance, adding the sequents $\Box\varphi \vdash \Box\Box\varphi$ and $\Diamond\Diamond\varphi \vdash \Diamond\varphi$ corresponds to restricting to transitive models. However, as noted in [15], only adding one of the duals of an ordinary modal axiom results in an incomplete logic. This difficulty is overcome by a move to the intuitionistic semantics, where the two dual formulations $\Box\Box\varphi \vdash \Box\varphi, \Diamond\Diamond\varphi \vdash \Diamond\varphi$ correspond to the transitivity of the $R_\Box := R\circ \leq, R_\Diamond := R\circ \leq^{-1}$ respectively. Furthermore, duality and correspondence and Sahlqvist theory [11], as well as a coalgebraic approach [29] have been developed for positive modal logic. However, this line of work is orthogonal to ours, as we will consider $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ and $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ as fragments of the modal language under the classical semantics.

[1] For a recent investigation into the theory of non-distributive positive modal logic, see [14].

4.2 Simulations

We will be interested in the classical version of positive modal logic, i.e. the one obtained as a syntactic restriction of the language over the usual classical semantics. In this context, $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ has been studied from a semantic perspective by Kurtonina and De Rijke in [24]. They propose the notion of *simulation* as a directed weakening of bisimulations, and prove a corresponding Van Benthem-style characterisation theorem.[2] Since positive-existential FO and hence $\mathcal{L}^+_{\Diamond,\wedge,\vee}$ is preserved by homomorphisms (2.3), one can view simulations as an appropriate generalization of homomorphisms to guarded universal quantification.

Definition 4.1. (Simulation)
Given two pointed models $(M, s), (M', s')$, a *simulation* between them is a relation $Z \subseteq M \times M'$ such that for all $(t, t') \in Z$, the following conditions hold.

(atom) $v^M(t) \subseteq v^{M'}(t')$

(forth) If $R^M tu$, then there is a u' with $R^{M'} t'u'$ and $(u, u') \in Z$

(back) If $R^{M'} t'u'$, then there is a u with $R^M tu$ and $(u, u') \in Z$

Theorem 4.1. *([24]) $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ is preserved under simulations, i.e. $M, s \models \varphi$ an M', s' simulates M, s then $M', s' \models \varphi$ as well for all $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$*

In fact, they prove a converse stating that a simulation-preserved first order formula is equivalent to the standard translation of a $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ formula.

4.3 Weak Simulations

We will be looking at the other positive fragment of modal logic identified by Dunn, i.e. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ without $\top$ and $\bot$. As we will see in the next chapter, $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ and even $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\bot}$ are not finitely characterisable, while our main result is that $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ is. Here, we lay the theoretical foundation for these results by proposing a further weakening of simulations, called *weak simulations* which we show characterises $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ as a fragment of FO. Moreover, we show that the class of pointed models with simulations or weak simulations as morphisms naturally forms a category with an interesting automorphism whose properties we heavily exploit. Before giving the definition of weak simulations, we introduce some special notation.

Definition 4.2. (Empty and Full Loopstates)
The *empty loopstate*, denoted by $\circlearrowleft_\emptyset$, is the pointed model consisting of a single reflexive point with am empty valuation. Dually, the *full loopstate*, denoted by $\circlearrowleft_{\mathrm{Prop}}$, is the pointed model consisting of a single reflexive point with a full valuation.

Definition 4.3. (Weak Simulation)
Given two pointed models $(M, s), (M', s')$, a *weak simulation* between them (notation: $M, s \underline{\rightarrow} M', s'$)[3] is a relation $Z \subseteq M \times M'$ such that for all $(t, t') \in Z$;

[2]Kurtonina and De Rijke persistently use 'directed simulations' to emphasize their directedness as opposed to the symmetry of bisimulations [24]. However, the present author feels as this is already emphasized enough by the omission of the prefix 'bi-' when compared to bisimulation. Moreover, we will introduce 'weak simulations' shortly, and these would otherwise get the misnomer of 'weak directed simulation', while we don't want this to evoke the impression that the directedness would somehow be weakened.

[3]It would be natural to use $\underline{\rightarrow}$ for simulations (as is done in [24]) and $\underline{\rightarrow}_w$ for weak simulations. However, we will not be concerned much with simulations as such so we use the more succinct notation for weak simulations instead.

(atom) $v^M(t) \subseteq v^{M'}(t')$

(forth') If $R^M tu$, either $M, u \leftrightarrows \circlearrowright_\emptyset$ or there is a u' with $R^{M'}t'u'$ and $(u, u') \in Z$

(back') If $R^{M'}t'u'$, either $M, u \leftrightarrows \circlearrowright_{\text{Prop}}$ or there is a u with $R^M tu$ and $(u, u') \in Z$

we refer to the extra 'loopstate-conditions' as the *escape clauses*.

Clearly, $\circlearrowright_\emptyset$ and $\circlearrowright_{\text{Prop}}$ play an important role in these weak simulations. The following 2 lemmas explicate this role (in fact they are closely connected) and the second one plays a vital role in our characterisation theorem.

Lemma 4.1. *Every pointed model weakly simulates $\circlearrowright_\emptyset$ and is weakly simulated by $\circlearrowright_{\text{Prop}}$.*

Proof. Let M, s be a pointed model. We claim that $Z := dom(\circlearrowright_\emptyset) \times dom(M)$ is a weak simulation $Z : \circlearrowright_\emptyset \;\underline{\rightarrow}\; M, s$.[4] Let $(r, t) \in Z$ be arbitrary. [atom] trivial since $v(r) = \emptyset$. [forth'] trivial since we can always use the escape clause. [back'] Whenever $Rt'u'$ we can always choose r itself as matching successor as Rrr holds and $(r, u') \in Z$. A dual argument shows that $dom(M) \times dom(\circlearrowright_{\text{Prop}})$ is the unique weak simulation $M, s \;\underline{\rightarrow}\; \circlearrowright_{\text{Prop}}$. □

In particular, the deadlock model (i.e. the single point model with no successors) *weakly* simulates $\circlearrowright_\emptyset$ and is *weakly* simulated by $\circlearrowright_{\text{Prop}}$, whereas it does not simulate $\circlearrowright_\emptyset$ nor is simulated by $\circlearrowright_{\text{Prop}}$. The following lemma is the crucial insight that makes the characterisation theorem work.

Lemma 4.2. *The empty loopstate $\circlearrowright_\emptyset$ satisfies no formula in $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, while the full loopstate $\circlearrowright_{\text{Prop}}$ satisfies every formula in $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$.*

Proof. Clearly $\circlearrowright_\emptyset \not\models p$ and $\circlearrowright_{\text{Prop}} \models p$ for every atomic formula p. Now suppose that $\circlearrowright_\emptyset \not\models \varphi$ and $\circlearrowright_{\text{Prop}} \models \varphi$. It follows that $\circlearrowright_\emptyset \not\models \Diamond\varphi \vee \Box\varphi$ and $\circlearrowright_{\text{Prop}} \models \Diamond\varphi \wedge \Box\varphi$ because in the former case the root has exactly one successor, and φ, respectively $\neg\varphi$ holds there. □

An immediate corollary is that the syntactic restriction to $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ thus trivialises the problems of determining satisfiability and validity.

Corollary 4.1. *All $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ formulas are satisfiable, and none of them is valid.*

Theorem 4.2. *(Preservation of $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ under Weak Simulations)*
Every formula $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ is preserved under weak simulations, i.e. if $M, s \models \varphi$ and $M, s \;\underline{\rightarrow}\; M', s'$ then $M', s' \models \varphi$.

Proof. Let $Z : M, s \;\underline{\rightarrow}\; M', s'$ be a weak simulation. By formula induction on φ. Booleans are immediate by inductive hypothesis.

(atom) If $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ is atomic then $\varphi = p$ for some $p \in$ Prop. Then $M, s \models p$ implies $M', s' \models p$ as $v(s) \subseteq v(s')$.

($\Diamond$) Suppose that $M, s \models \Diamond\varphi$ where $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, i.e. there is some $t \in R^M[s]$ with $M, t \models \varphi$. By lemma 4.2, $M, t \not\leftrightarrows \circlearrowright_\emptyset$ and hence by [forth'] there must be some $t' \in R^{M'}[s']$ such that $(t, t') \in Z$ and hence $M', t' \models \varphi$ by inductive hypothesis. It follows by the semantics that $M', s' \models \Diamond\varphi$.

[4] There is not a *unique* such weak simulation, not even between rooted models. A counterexample is when $M = (\{s, t\}, \{(s, t), (t, t)\}, v)$ where $v(s) = v(t) =$ Prop}. Then both $\{(r, s)\}$ and $\{(r, s), (r, t)\}$ are weak simulations $\circlearrowright_\emptyset \;\underline{\rightarrow}\; M, s$ (where $dom(\circlearrowright_\emptyset) = \{r\}$ is the root of the empty loop model).

($\Box$) Suppose that $M, s \models \Box\varphi$ with $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. We show that $M', s' \models \Box\varphi$. If $R^{M'}[s'] = \emptyset$ this holds vacuously. For an arbitrary $t' \in R^{M'}[s']$, by the [back'] clause either $M', t' \underline{\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ or there is a successor $t \in R^M[s]$ with $(t, t') \in Z$. In the former case, by lemma 4.2 $M', t' \models \varphi$. Else, as $M, s \models \Box\varphi$ we have $M, t \models \varphi$ and hence by inductive hypothesis $M', t' \models \varphi$.

$\Box$

This validates our thinking of weak simulations as 'structure-preserving', for we are interested in the structure of Kripke models as expressible in $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. Note that the converse does not hold as $\top$ and $\bot$ are also preserved under weak simulations (actually these are always preserved under any relation) but not equivalent to a formula in our language $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ (observe that no $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ is valid or unsatisfiable by corollary 4.1). We leave it as an open problem whether the slight extension $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee} \cup \{\top\}$ does satisfy a van Benthem-style characterisation theorem. [5]

4.4 The Category of Pointed Models and Weak Simulations

In this section, we show that our non-standard notion of weak simulation is well-behaved in the sense that it gives rise to a category over the class of pointed models. In particular, we show that weak simulations are closed under composition, and that this composition is associative. Finally, we show that this category is dually isomorphic to itself, like the category **Rel** of sets and binary relations [5]. We will need the following definitions and lemmas.

Definition 4.4. (Operations on Binary Relations)
For a binary relation $Z \subseteq A \times B$, set $\pi_0(Z) = \{a \in A \mid \exists b \in B\ (a, b) \in Z\}$ and $\pi_1(Z) = \{b \in B \mid \exists a \in A\ (a, b) \in Z\}$. Furthermore, given two binary relations $Z_0 \subseteq A \times B, Z_1 \subseteq B \times C$, let $Z_0 \circ Z_1 = \{(a, c) \in A \times C \mid \exists b \in B\ (a, b) \in Z_0\ \&\ (b, c) \in Z_1\}$ be their (relational) *composition*. Thirdly, given a binary relation $Z \subseteq A \times B$, let $Z^{-1} = \{(b, a) \in B \times A \mid (a, b) \in Z\}$ denote the *converse* of Z. A binary relation $Z \subseteq A \times B$ is *total* if $\pi_0(Z) = A$ and $\pi_1(Z) = B$ (note that this is equivalent to the earlier definition in the preliminaries).

Lemma 4.3. *Every weak simulation $Z : M, s \underline{\rightarrow} M', s'$ between rooted models is total on the set of all 'non-loopstates', i.e. $\{t \in dom(M) \mid M, t \underline{\not\leftrightarrow} \circlearrowleft_\emptyset\} \subseteq dom(Z)$ and $\{t' \in dom(M') \mid M', t' \underline{\not\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}\} \subseteq cod(Z)$.*

Proof. Let $Z : M, s \underline{\rightarrow} M', s'$ be a weak simulation where M, s and M', s' are rooted. Now take some $t \in dom(M)$ such that $M, t \underline{\not\leftrightarrow} \circlearrowleft_\emptyset$. By rootedness, there is a path $(s = t_0, \ldots, t_n = t)$ of some finite length n from the root s to t in M. This implies that $M, t_i \underline{\not\leftrightarrow} \circlearrowleft_\emptyset$. Hence by repeated application of the [forth'] clause of Z we get a path $(s' = t'_0, \ldots, t'_n)$ of length n in M' with $(t_j, t'_j) \in Z$ for all $j \leq n$. In particular $(t_n, t'_n) \in Z$ so $t_n \in dom(Z)$. A dual argument shows that if $M', t' \underline{\not\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ then $t' \in dom(Z)$. $\Box$

Lemma 4.4. *Let $M, s \underline{\rightarrow} M', s'$. If $M', s' \underline{\leftrightarrow} \circlearrowleft_\emptyset$, then it must be that $M, s \underline{\leftrightarrow} \circlearrowleft_\emptyset$. Similarly, if $M, s \underline{\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ then $M', s' \underline{\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$.*

[5] It seems that proof of the analogous result in [24] might adapt to our case. Subsequently, one could also wonder whether these characterisations theorems still hold 'in the finite', as in Rosen's refinement of the original van Benthem characterisation theorem.

Proof. Let $Z : M, s \underline{\rightarrow} M', s'$ and $Z' : M', s' \underline{\leftrightarrow} \circlearrowright_\emptyset$. Without loss of generality, assume that M, s and M', s' are rooted. Otherwise, we may pass to the submodels of M, M' generated by s, s' respectively. The restrictions of Z and Z' to these submodels will still be a weak simulation an bisimulation respectively.[6] Then Z' must be total (cf. proposition 2.1) so it must be that $Z' = dom(M') \times dom(\circlearrowright)$ as this the only total relation on the domains of these models. Let

$$Z^* := (Z \circ Z') \cup \{(t, r) \in dom(M) \times dom(\circlearrowright_\emptyset) \mid M, t \underline{\leftrightarrow} \circlearrowright_\emptyset\}$$

we show that this is a bisimulation $Z^+ : M, s \underline{\leftrightarrow} \circlearrowright_\emptyset$. First of all, note that by lemma 4.3, it must be that $dom(Z^*) = dom(M)$ and hence Z^* must be total. So let $(t, r) \in Z^*$ be arbitrary. Note that if $(t, r) \in Z^* \setminus (Z \circ Z')$, then all the clauses are trivially satisfied because $M, t \underline{\leftrightarrow} \circlearrowright_\emptyset$ by definition. So suppose that $(t, r) \in Z \circ Z'$. Then there is some $t' \in dom(M)$ with $(t, t') \in Z$ and $(t', r) \in Z'$.

[atom] By the directed [atom] clause of Z that $v(t) \subseteq v(t')$ but we already know that $v(t') = \emptyset$, hence $v(t) = \emptyset = v(r)$. [forth] If $R^M tu$ and $M, u \underline{\leftrightarrow} \circlearrowright_\emptyset$ by definition $(u, r) \in Z^*$ and r is its own successor. If $M, u \underline{\not\leftrightarrow} \circlearrowright_\emptyset$ by the [forth'] clause of Z there must be a matching successor u' of t' in M' and so by the [forth] clause of Z' also a matching successor of r in $\circlearrowright_\emptyset$. But r is its own only successor so it must be that $(u, u') \in Z$ and $(u', r) \in Z$ and hence $(u, r) \in Z \circ Z'$. Again, this is matching successor because r is its own successor.

Before showing that the [back] clause holds, let us observe that M must be serial. For suppose otherwise, then there is some $t \in dom(M)$ which is a deadlock, i.e. $R^M[t] = \emptyset$. Then $M, t \underline{\not\leftrightarrow} \circlearrowright_\emptyset$ so by lemma 4.3 there is some $t' \in dom(M')$ such that $(t, t') \in Z$. Then by the [back'] clause of Z either t' is also a deadlock or has only full loopstate-successors. But the latter cannot happen because every state in M' has empty colour. Thus $R^{M'}[t'] = \emptyset$, but then it cannot be that (t', r) is in the bisimulation Z' because a deadlock state cannot be bisimilar to $\circlearrowright_\emptyset$. But (t', r) *must* be in Z' by totality of Z'. Hence we may suppose that M is serial. Now we can continue to show that the [back] clause holds. [back] The only successor of r is r and since M is serial $\exists u \in R^M[t] \neq \emptyset$. Finally, by totality of Z^* we get $(u, r) \in Z^*$. □

With these lemmas in hand, we can proceed to show that weak simulations can be naturally seen as morphisms between pointed models in the categorical sense.

Theorem 4.3. *Weak simulations are closed under composition.*

Proof. Suppose that $M_0, s_0 \overset{Z_1}{\underline{\rightarrow}} M_1, s_1 \overset{Z_2}{\underline{\rightarrow}} M_2, s_2$ are weak simulations. I claim that $Z_1 \circ Z_2 : M_0, s_0 \underline{\rightarrow} M_2, s_2$ is a weak simulation. So let $(t_0, t_2) \in Z_1 \circ Z_2$ be arbitrary, then there is some t_1 with $(t_0, t_1) \in Z_1$ and $(t_1, t_2) \in Z_2$. [atom] Clearly $v(t_0) \subseteq v(t_1) \subseteq v(t_2)$ by the [atom] clauses of Z_1 and Z_2. [forth'] If Rt_0u_0, by [forth'] of Z_1 either $M_0, u_0 \underline{\leftrightarrow} \circlearrowright_\emptyset$ or there is a matching successor u_1 of t_1 with $(u_0, u_1) \in Z_1$. In the former case, we can use the escape clause for the pair (t_0, t_2) and we are done. Else, we apply the [forth'] clause on u_1 and get that either $M_1, u_1 \underline{\leftrightarrow} \circlearrowright_\emptyset$ or there is a successor u_2 of t_2 with $(u_1, u_2) \in Z_2$ and hence $(u_0, u_2) \in Z_1 \circ Z_2$. In the latter case, we would be done. In the former case we have $M_0, u_0 \underline{\rightarrow} M_1, u_1 \underline{\leftrightarrow} \circlearrowright_\emptyset$ so by lemma 4.4 $M_0, u_0 \underline{\leftrightarrow} \circlearrowright_\emptyset$ and we can use the escape clause on $(t_0, t_2) \in Z_1 \circ Z_2$.

[back'] If Rt_2u_2 by [back'] of Z_2 either $M_2, u_2 \underline{\leftrightarrow} \circlearrowright_{\mathrm{Prop}}$ or there is a matching successor u_1 of t_1 with $(u_1, u_2) \in Z_2$. In the former case, we can use the escape clause for the pair (t_0, t_2)

[6] It is worth noting here that the semantic properties of a pointed model M, s are completely determined by the set of all *finite* paths in M starting at s, which is also the notion of *unravelling* a model from the modal literature [8].

and we are done. Else, we apply the [back'] clause on u_1 and get that either $M_1, u_1 \underline{\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ or there is a successor u_0 of t_0 with $(u_0, u_1) \in Z_1$ and hence $(u_0, u_2) \in Z_1 \circ Z_2$. In the latter case, we would be done. In the former case we have $\circlearrowleft_{\mathrm{Prop}} \underline{\leftrightarrow} M_1, u_1 \underline{\rightarrow} M_2, u_2$ so by lemma 4.4 $M_2, u_2 \underline{\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ and we can use the escape clause on $(t_0, t_2) \in Z_1 \circ Z_2$. □

Proposition 4.1. *Composition of weak simulations is associative.*

Proof. This follows from the fact that relational composition is associative. Let $M_0, s_0 \overset{Z_1}{\underline{\rightarrow}} M_1, s_1 \overset{Z_2}{\underline{\rightarrow}} M_2, s_2 \overset{Z_2}{\underline{\rightarrow}} M_3, s_3$ be weak simulations. Observe that:

$$\begin{aligned}(Z_1 \circ Z_2) \circ Z_3 =& \{(t_0, t_3) \mid \exists t_2 (t_0, t_2) \in Z_1 \circ Z_2 \text{ and } (t_2, t_3) \in Z_3\} \\ =& \{(t_0, t_3) \mid \exists t_2, t_3 (t_0, t_1) \in Z_1,\ (t_1, t_2) \in Z_2 \text{ and } (t_2, t_3) \in Z_3\} \\ =& \{(t_0, t_3) \mid \exists t_1 (t_0, t_1) \in Z_1 \text{ and } \ (t_1, t_3) \in Z_2 \circ Z_3\} = Z_1 \circ (Z_2 \circ Z_3)\end{aligned}$$

□

It rests to observe that the diagonal $1_{M,s} = \{(t, t) \mid t \in dom(M)\}$ can serve as the identity morphism for each pointed model M, s.

Corollary 4.2. *The class of pointed models with weak simulations as morphisms forms a category* **wSim** *with weakly initial object* $\circlearrowleft_{\emptyset}$ *and weakly final object* $\circlearrowleft_{\mathrm{Prop}}$.

Although not relevant to the results in this thesis, simulations also give rise to a category.

Theorem 4.4. *The class of pointed models with simulations as morphisms forms a category* **Sim**.

Proof. This follows from the fact that frame bisimulations and hence simulations are closed under composition, where this composition (which is just relational composition) is moreover associative. □

Note that $\circlearrowleft_{\emptyset}$ and $\circlearrowleft_{\mathrm{Prop}}$ are not even weakly initial and terminal respectively in **Sim**, because e.g. the deadlock model does not simulate $\circlearrowleft_{\emptyset}$ nor is it simulated by $\circlearrowleft_{\mathrm{Prop}}$.

4.5 Some Further Properties of Weak Simulation

Simulations and weak simulations are directed weakenings of bisimulations. The following propositions provides an interesting insight into the consequences of this directedness. First, we define a 'flipping' operation on models

Definition 4.5. Given a model $M = (dom(M), R, v)$, let $M^{\neg} = (dom(M), R, v^{\neg})$ be obtained from M by 'flipping' all the valuations, i.e. $v^{\neg}(t) = \mathrm{Prop} - v(t)$ for all $t \in dom(M)$.[7]

Proposition 4.2. *Whenever* $Z : E, e \underline{\rightarrow} M, s$ *also* $Z^{-1} : M^{\neg}, s \underline{\rightarrow} E^{\neg}, e$

Proof. Let $(t, f) \in Z^{-1}$ be arbitrary, then $(f, t) \in Z$. [atom] By the [atom] clause of Z we get $v(f) \subseteq v(t)$ and hence $v^{\neg}(t) \subseteq v^{\neg}(f)$. [forth'] If Rtu and $M^{\neg}, u \underline{\not\leftrightarrow} \circlearrowleft_{\emptyset}$ then $M, u \underline{\not\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ and hence by the [back'] clause of Z there must be some successor g of f with $(g, u) \in Z$ and hence $(u, g) \in Z^{-1}$. [back'] Similarly, if Rfg and $E^{\neg}, g \underline{\not\leftrightarrow} \circlearrowleft_{\mathrm{Prop}}$ then $E, g \underline{\not\leftrightarrow} \circlearrowleft_{\emptyset}$ so by [forth'] of Z there must be some successor u of t with $(g, u) \in Z$ and hence $(u, g) \in Z^{-1}$. □

[7] Equivalently, with ordinary valuations we get $V^{\neg}(p) = dom(M) - V(p)$

Note that the analogue proposition also holds for simulations. In fact, using the above proposition we can define an endofunctor on **wSim** as follows (this is also a well-defined on **Sim** by the analogue of proposition 4.2 for **Sim**).

Definition 4.6. Define a contravariant functor $(\cdot)^\neg : \mathbf{wSim} \to \mathbf{wSim}^{op}$ by setting $(M, s)^\neg := M^\neg, s$ on objects and $Z^\neg := Z^{-1}$. Clearly this is well-defined by proposition 4.2.

It follows that the contravariant *opposite functor* $(\cdot)^{\neg^{op}} : \mathbf{wSim}^{op} \to \mathbf{wSim}$ also exist. We claim that this is the inverse functor of $(\cdot)^\neg$, and thus that $(\cdot)^\neg$ is an isomorphism.

Theorem 4.5. $(\cdot)^{\neg^{op}} : \mathbf{wSim}^{op} \to \mathbf{wSim}$ *is dual isomorphism of categories*

Proof. Observed that $(\cdot)^\neg$ is *idempotent* on objects since $v^{\neg\neq}(p) = \text{Prop} \setminus (\text{Prop} \setminus V(p)) = V(p)$ (note that this is identity not some weaker notion of equivalence). But clearly also $Z^{\neg\neg} = (Z^{-1})^{-1} = Z$. Also, $(1_{M,s})^\neg = \{(t,t) \mid t \in \mathit{dom}(M)\} = 1_{M^\neg,s} = 1_{(M,s)^\neg}$ and thus $(\cdot)^\neg \circ (\cdot)^{\neg^{op}} = 1_{\mathbf{wSim}}$ and $(\cdot)^{\neg^{op}} \circ (\cdot)^\neg = 1_{\mathbf{wSim}^{op}}$. □

In fact, the exact same argument shows that $(\cdot)^\neg : \mathbf{Sim} \to \mathbf{Sim}^{op}$ is also an isomorphism of categories.[8] Note that this implies that the co-limit of a diagram always needs to be the $(\cdot)^\neg$ image of the limit of the diagram (if it exists) and vice versa. This is analogous to the situation in the category **Rel** of sets and binary relations [5].

Next, we can show that our functor also coheres nicely with the following syntactic operation on formulas.

Definition 4.7. Given a modal formula $\varphi(p_1, \ldots, p_n)$, let $\varphi^\neg := \varphi[\neg p_1/p_1] \ \ldots \ [\neg p_n/p_n]$. Note that $\varphi^{\neg\neg} \equiv \varphi$, i.e. our operation is idempotent up to logical equivalence.[9]

Theorem 4.6. *For all formulas φ and pointed models M, s*

$$M, s \models \varphi \qquad \textit{iff} \qquad M^\neg, s \models \varphi^\neg$$

Proof. By induction on the complexity of φ. The boolean case are immediate by inductive hypothesis.

(atomic) Note that $p^\neg = \neg p$ and clearly: $M, s \models p$ iff $p \in v(s)$ iff $p \notin v^\neg(s)$ iff $M^\neg, s \models \neg p$.

($\Diamond$) $M, s \models \Diamond\varphi$ iff there is some $t \in R^M[s]$ such that $M, t \models \varphi$ iff (by inductive hypothesis) there is some $t \in R^{M^\neg}[s]$ such that $M^\neg, t \models \varphi^\neg$ iff $M^\neg, s \models \Diamond\varphi^\neg$ (and $(\Diamond\varphi)^\neg = \Diamond\varphi^\neg$).

($\Box$) $M, s \models \Box\varphi$ iff for all $t \in R^M[s]$ we have $M, t \models \varphi$ iff (by inductive hypothesis) for all $t \in R^{M^\neg}[s]$ we have $M^\neg, t \models \varphi^\neg$ iff $M^\neg, s \models \Box\varphi^\neg$ (and $(\Box\varphi)^\neg = \Box\varphi^\neg$).

□

In fact, we can generalize these operations as follows. This will be relevant in chapter 5 when we will consider extending our positive results to richer languages that use $\neg$. Let $Q \subseteq \text{Prop}$, then $M^{\neg Q}$ is obtained from M by flipping all the valuation of propositional variables $q \in Q$. Similarly, we define $\varphi^{\neg Q}$ as the restriction of the $\neg$ operation on formulas to propositional variables $q \in Q$. That is, if $Q = \{q_1, \ldots, q_n\}$ then we have $\varphi^{\neg Q} := \varphi[\neg q_1/q_1] \ldots [\neg q_n/q_n]$. These cohere in the same way as before

Theorem 4.7. *For all formulas φ and pointed models M, s,*

$$M, s \models \varphi \qquad \textit{iff} \qquad M^{\neg Q}, s \models \varphi^{\neg Q}$$

[8] Perhaps shedding light on the nature of simulations, for it seems to be an intimate connection between simulations from M, s to M', s' and pairs of homomorphisms $M, s \to_{hom} M', s'$,$M'^\neg, s' \to_{hom} M^\neg, s$

[9] In fact, φ can be easily retrieved from $\varphi^{\neg\neg}$ by eliminating double negations.

Chapter 5

Characterising Modal Fragments

In this section, we will study finite characterisability for *fragments* of the full modal language. We will only be looking at these fragments as interpreted over the class of *all finite* frames. That is, for $\mathcal{L}$ a modal fragment in some finite set of propositional variables Prop, let

$$\mathcal{C}_{\mathcal{L}} := \{mod^{K}_{\text{fin}}(\varphi) \mid \varphi \in \mathcal{L}\}$$

be the modal concept class induced by this fragment. Hence all the concept classes $\mathcal{C}_{\mathcal{L}}$ are subclasses of $\mathcal{C}^{\text{fin}}_{K}$ (which is 'equivalent' to $\mathcal{C}_K$ because K has the finite model property) and contain only model classes $mod_{\text{fin}}(\varphi)$ over the class of all finite frames.[1] Therefore, henceforth we will just write $mod(\varphi)$ instead of $mod^{L}_{\text{fin}}(\varphi)$. Although not reflected in our notation, such concept class thus depends on a choice of a finite set of propositional variables Prop. Hence, a finite characterisation of a concept $mod(\varphi) \in \mathcal{C}_{\mathcal{L}}$ is a *finite* set of *finite* examples (i.e. finite pointed models with a label + or -) that distinguishes $mod(\varphi)$ from all other concepts $mod(\psi) \in \mathcal{C}_{\mathcal{L}}$ where $\varphi \not\equiv \psi$ (i.e. non-equivalent over the class of all finite frames).

In chapter 3, we explored finite characterisability for normal modal logics (i.e. the full modal language over various restricted frame-classes). There, we saw that the fact that the full modal language is closed under negation plays an important role in the construction of counterexamples and in 3.3. Moreover, the positive results in [9] are also restricted to positive fragments of FO. Therefore, it seems reasonable to hope for finitely characterisable *positive* fragments of the modal language. However, inspecting the proof of 3.2, we see that they also imply negative results for fragments.

Theorem 5.1. *$\mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ is not finitely characterisable. In fact, no formula in it has a finite characterisation w.r.t. this fragment.*

Proof. Observe that $\Box\bot$ and $height_n$ are in $\mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ for each n and this fragment is closed under disjunction and conjunction. Hence, we can redo the proof of theorem 3.2 in this fragment from which we can conclude that no $\varphi \in \mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ has a finite characterisation w.r.t. this fragment. □

As it turns out, we can also force the characterisations to contain models of arbitrary height without using $\top$.

Theorem 5.2. *$\mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee,\bot}$ is not finitely characterisable, assuming* Prop $\neq \emptyset$.

[1] As we will see in the discussion, in view of applications it is also interesting to look at concept classes for fragments of the modal language over restricted classes of models.

Proof. Note that $M, s \models \Box^{n+1}\bot \wedge \Diamond^n p$ implies that $height(M, s) \geq n$ for all pointed models M, s.[2] Hence if we suppose that this fragment is finitely characterisable then in particular the formula $\Box\bot \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\bot}$ has a finite characterisation (E^+, E^-) w.r.t. this fragment. Let $n < \max\{|dom(E)| \mid (E, e) \in E^+ \cup E^-\}$, then by almost the same argument as in 3.1 it follows that $\varphi := \Box\bot \vee (\Box^{n+1}\bot \wedge \Diamond^n p)$ fits (E^+, E^-) yet $\Box\bot \not\equiv \varphi$. Next, clearly φ fits E^+ by properties of disjunction. Moreover, φ also fits E^- by construction, for take any negative example $E, e \in E^-$ for $\Box\bot$, then we have $E, e \not\models \Box\bot$. However, note that $M, s\Box^{n+1}\bot \wedge \Diamond^n p$ implies that $height(M, s) = n$, because the right conjunct expresses the existence of an n-length path while by the first conjunct makes it impossible for this path to have cycles, as well as the non-existence of a path of length $> n$. Finally, note that $\Box^{n+1}\bot \wedge \Diamond^n p) \to \Diamond\top$ and since $\neg\Box\bot = \Diamond\top$ this means that $\Box\bot \not\equiv \varphi$ (because the right disjunct $(\Box^{n+1}\bot \wedge \Diamond^n p)$ of φ is satisfiable on the n-length path model ending in a p-state). □

Also note that the fragment $\mathcal{L}_{\Box,\Diamond,\wedge,\vee}$ (which does not restrict the polarity of the propositional variables) is not finitely characterisable because it can express $\top$ and $\bot$ via the equivalences $p \wedge \neg p \equiv \bot, p \vee \neg p \equiv \top$, assuming $\text{Prop} \neq \emptyset$.

5.1 Positive-Existential Modal Logic

It is well-known that modal logic embeds into FO via the *standard translation*. Hence, the foregoing results following from [9] imply characterisability (as well as learnability) results for fragments of modal logic. For the sake of being self-contained we first define the standard translation of modal logic into a fragment of FO.

Observe that a Kripke model $M = (dom(M), R, v)$ can be seen as a structure with $M^{str} = (dom(M), R, \{P \mid p \in \text{Prop}\}, s)$ where each propositional variable $p \in \text{Prop}$ becomes a unary relation P with interpretation $P^{M'} = \{m \in dom(M) \mid p \in v(m)\}$. The inverse of $(\cdot)^{str}$ is the transformation $(\cdot)^{krip}$ that maps a structure over such a schema to the corresponding Kripke structure, i.e. turning every propositional variable p into a unary predicate. This mapping between models then coheres together with the following translation between formulas.

Definition 5.1. (Standard Translation)
Let x be a first order variable. We define the standard translation $ST_x(\varphi)$ in free variable x of a modal formula φ as follows:

$$\begin{aligned} ST_x(p) &:= Px \\ ST_x(\varphi \wedge \psi) &:= ST_x(\varphi) \wedge ST_x(\psi) \\ ST_x(\neg\varphi) &:= \neg ST_x(\varphi) \\ ST_x(\Diamond\varphi) &:= \exists y(Rxy \wedge ST_y(\varphi)) \end{aligned}$$

We call the schema of the first order language into which the modal language translates the *correspondence schema*. There is one-to-one correspondence between (pointed) Kripke models and structures over the correspondence schema (with one distinguished element).[3]

[2]Interestingly, note that the $\bot$ and $\top$-free formula $\bigwedge_{i\leq n} \Box\neg p \wedge \Diamond^{n+1} p$ also forces the height of a pointed model satisfying it to be $\geq n$.

[3]We can define a variant of the standard translation where we translate a modal formula φ to the first order *sentence* $ST_x(\varphi)[c/x]$ where c is a fresh constant. If S is the schema of the first order language which is the image of this translation, which thus contains a single constant, then $\mathbf{Hom}_S$ is the category of *pointed* Kripke models and homomorphisms.

We transform a Kripke model into a structure by turning all propositional variables $p \in$ Prop into unary predicates P, and conversely transform a structure over the correspondence schema into a Kripke models by turning all unary predicates into propositional variables. Henceforth we will blur the distinction between (pointed) Kripke models and structures with one designated element. For a structure M and a first order formula $\varphi(x)$ with one free variable, we will frequently write $M, s \models \varphi$ rather than $M \models \varphi(s)$ or $M \models \varphi(x)[s/x]$. This is warranted because free variables are interpreted similarly to constants anyway under the standard Tarskian semantics.

Theorem 5.3. *(see e.g. [8]) For all pointed models M, s and modal formulas φ;*

$$M, s \models \varphi \qquad \textit{iff} \qquad M, s \models ST_x(\varphi)$$

Observe that for a formula $\varphi \in \mathcal{L}^+_{\Diamond,\wedge}$, $ST_x(\varphi)$ is equivalent to a unary CQ q for which $\widehat{q}$ is a c-acyclic (and in fact acyclic) structure. Because a characterisation w.r.t. all unary CQs (or all unary connected acyclic CQs) is also a characterisation w.r.t. $\mathcal{L}^+_{\Diamond,\wedge}$ (considered as a subclass through the standard translation), we get the following result as an immediate corollary of theorem 2.7.[4]

Corollary 5.1. *$\mathcal{L}^+_{\Diamond,\wedge}$ is finitely characterisable. Moreover, this characterisation can be computed in polynomial time.*

Moreover, note that every formula $\varphi \in \mathcal{L}^+_{\Diamond,\wedge,\vee}$ can be written as a disjunction of formulas in $\mathcal{L}^+_{\Diamond,\wedge}$ by Kripke normality and the propositional distributive law [5], and for formulas in this normal form $ST_x(\varphi)$ is a c-acyclic UCQ. Hence we can *identify* $\mathcal{L}^+_{\Diamond,\wedge,\vee}$ with a subclass of (unary) UCQs. Therefore, the following is an immediate corollary of theorem 2.12.

Corollary 5.2. *$\mathcal{L}^+_{\Diamond,\wedge,\vee}$ is finitely characterisable. Moreover, this characterisation can be computed in singly exponential time.*

However, in order to obtain learning algorithms for these modal fragments we have to put in some extra work. Namely, the output of the foregoing learning algorithms in section 2.4 is a CQ or a UCQ, and not a modal formula per sé. We show that there is polynomial time algorithm for computing, from a c-acyclic CQ that is equivalent to a modal formula, a modal formula witnessing this equivalence. Let $Sub(\varphi)$ denote the set of subformulas of φ. Set the *size* of a modal formulas to mean the number of its subformulas, i.e. $|\varphi| := |Sub(\varphi)|$.

Theorem 5.4. *$\mathcal{L}^+_{\Diamond,\wedge}$ is polynomial-time exact learnable with membership queries only.*

Proof. Take some modal formula $\varphi_{goal} \in \mathcal{L}^+_{\Diamond,\wedge}$. Then $ST_x(\varphi_{goal})$ is a c-acyclic (in fact acyclic) and connected CQ. This means that the learning algorithm from theorem 2.8 can learn to find a core unary c-acyclic CQ $q(x)$ such that $q(x) \equiv ST_x(\varphi)$ and hence $q \equiv \varphi_{goal}$ (modulo the distinction between structures over the correspondence schema and Kripke models). We may assume that the output CQ q of the learning algorithm of theorem 2.8 is a core, which means that $\widehat{q}$ is a core [9]. By theorem 2.4, as $q \equiv ST_x(\varphi)$ it follows that $\widehat{q} \leftrightarrow_{hom} \widehat{ST_x(\varphi)}$. But note that $\widehat{ST_x(\varphi)}$ is a tree structure, and since $\widehat{q}$ is core it must be the core of this

[4] Alternatively, we could also get such characterisations via the frontiers obtained through theorem 2.9, which is arguably closer to modal spirit. Though note these are strictly weaker characterisations, as they characterise the formula in question only w.r.t. the class of (unary) acyclic connected CQs. On the other hand, the frontier are much simpler; for instance they are connected and tree-shaped.

[5] That is, by the equivalences $\Diamond \bigvee \Phi \equiv \bigvee_{\varphi \in \Phi} \Diamond \varphi$ and $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$

tree. We show that it follows that $\widehat{q}$ must also be a tree. Let $h : \widehat{q} \to_{hom} \widehat{ST_x(\varphi)}$ and $h' : \widehat{ST_x(\varphi)} \to_{hom} \widehat{q}$ be homomorphisms witnessing this equivalence. Note that we may take h' to be such that $h' \circ h$ is the identity on $\widehat{q}$. Otherwise, let $h' \circ h$ not be the identity. As $\widehat{q}$ is core, $g := h' \circ h$ is an automorphism of $\widehat{q}$. But as $g^{-1} \circ g$ must be the identity of $\widehat{q}$, setting $h'' := g^{-1} \circ h' : \widehat{ST_x(\varphi)} \to_{hom} \widehat{q}$ we get that $h'' \circ h = (g^{-1} \circ h') \circ h = g^{-1} \circ (h' \circ h) = g^{-1} \circ g$ is the identity on $\widehat{q}$.

We know that $\widehat{q}$ must be acyclic (in the sense of definition 2.10), otherwise this cycle in $\widehat{q}$ would homomorphically projects via h onto a cycle in the acyclic structure $\widehat{ST_x(\varphi)}$, which is impossible. It also follows that $\widehat{q}$ is rooted. For take some $t \in dom(\widehat{q})$, we know that $h(t) \in dom(\widehat{ST_x(\varphi)})$ so by rootedness of this structure there must be a path $xR \dots Rh(t)$ from the root x to this element (recall that the domain of this structure is the set of variables occurring in $ST_x(\varphi)$). But then this path homomorphically projects onto a path $h'(x)R \dots Rh' \circ h(t)$ in $\widehat{q}$. However, this homomorphism maps distinguished elements to each other hence $h'(x) = x \in dom(\widehat{q})$ where x is the single distinguished element of $\widehat{q}$ since $free(q) = \{x\}$. it follows that $xR \dots Rt$ since $h' \circ h$ is the identity, and hence $\widehat{q}$ is rooted. Note that this also follows that $\widehat{q}$ is generated by the root x in the sense of modal logic. Hence we have shown that $\widehat{q}$ is a tree.

Further, this tree has the special property that all the leaves satisfy some unary predicate. We know this is the case for $ST_x(\varphi)$ because $\varphi \in \mathcal{L}^+_{\Diamond,\wedge}$ and hence does not contain $\top$. But if t is a leaf of the tree $\widehat{q}$ then $h(t)$ must be a leaf of the other structure and thus satisfy some unary predicate, i.e. $\widehat{ST_x(\varphi)} \models Ph(t)$ for some P. But $h'h(t) = t$ and hence $\widehat{q} \models Pt$ for some P. Now we show that the canonical query of every tree satisfying this special 'coloured leaves'-condition can be defined by a modal $\mathcal{L}^+_{\Diamond,\wedge}$ formula. By induction on tree-depth. If M, s is a tree of depth 0, s is a deadlock and we set $\varphi_{M,s} := \bigwedge_{M \models Ps} p$, and if the depth > 0 and $t_0, \dots, t_n$ are the children of s, then set $\varphi_{M,s} := \bigwedge_{i \leq n} \Diamond \varphi_{M,t_i}$.

Claim: $M', s' \models \varphi_{M,s}$ iff $M, s \to_{hom} M', s'$

Proof. From left to right, suppose that $M', s' \models \varphi_{M,s}$. We define a homomorphism $h : M, s \to_{hom} M', s'$ by setting $h(t) := \{t' \in dom(M') \mid M', t' \models \varphi_{M,t}\}$. Conversely, let $M, s \to_{hom} M', s'$. We show that $M', s' \models \varphi_{M,s}$ by induction on tree-depth. If $t \in dom(M)$ is a leaf of M, s then $h(t) \in dom(M')$ is a leaf of M', s' and satisfies all unary predicates that hold at t. Hence $M', h(t) \models \varphi_{M,t}$. For the inductive step, if $M', h(t_i) \models \varphi_{M,t_i}$ for all successors t_i of a node $t \in dom(M)$ then $M', h(t) \models \varphi_{M,t}$ since each $h(t_i)$ is a successor of $h(t)$ in M', s' and hence $M', h(t) \models \Diamond \varphi_{M,t_i}$ for each i. □

It follows that the canonical query of $\widehat{q}$ can also be defined by a modal formula $\varphi_{\widehat{q}} \in \mathcal{L}^+_{\Diamond,\wedge}$. But q is the canonical query of this structure, hence for every pointed model M, s

$$M, s \models q \qquad \text{iff} \qquad \widehat{q} \to_{hom} M, s \qquad \text{iff} \qquad M, s \models \varphi_{\widehat{q}}$$

hence $q \equiv \varphi_{\widehat{q}}$ (modulo the distinction between structures over the correspondence schema and Kripke models). But then we have $\varphi_{goal} \equiv q \equiv \varphi_{\widehat{q}}$. Thus the modal formula $\varphi_{\widehat{q}}$ is the desired output of this learning algorithm. Finally, note that the construction construction here is polynomial in $|ST_x(\varphi_{goal})|$ because constructing $\varphi_{\widehat{q}}$ from q only requires computing $\varphi_{\widehat{q},t}$ for each $t \in dom(\widehat{q})$. Hence for any measure of size of first order formulas such that

$|\varphi| \leq |ST_X(\varphi)|$, it follows that the algorithm also runs in time polynomial in $|\varphi|$.[6] □

By contrast, obtaining a learning algorithm for $\mathcal{L}^+_{\Diamond,\wedge,\vee}$ as a corollary of theorem 2.13 does not require extra work because we used the naive algorithm.

Theorem 5.5. $\mathcal{L}^+_{\Diamond,\wedge,\vee}$ *is exponential time exact learnable with membership queries only.*

Proof. Enumerate all formulas $\varphi_0, \varphi_1, \varphi_2, \ldots$ from $\mathcal{L}^+_{\Diamond,\wedge,\vee}$ in order of increasing size (i.e. assuming measures of size of modal formulas and CQs satisfying $|\varphi| \leq |ST_x(\varphi)|$). By corollary 5.2, each formula φ_i in this list has a finite characterisation w.r.t. $\mathcal{L}^+_{\Diamond,\wedge,\vee}$ which can be computed in time $2^{\mathcal{O}(|\varphi_i|)}$. Then by the same argument as in theorem 2.13, we can learn φ with membership queries only in time $2^{\mathcal{O}(|\varphi|)}$. □

5.2 Monotone Modal Logic

In this section, we extend the above corollaries for positive-existential modal logic by showing that we may also allow $\Box$ as a connective while staying finitely characterisable. Hence our main result is that *monotone modal logic*, i.e. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ is finitely characterisable, but we show how our result also extends to the larger class of uniform formulas $\mathcal{L}^u_{\Box,\Diamond,\wedge,\vee}$. We give a compositional construction of positive examples for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ formulas in *Fine normal form* [16], and show how the negative examples can be obtained from these. In fact, the characterisations $(E^+_\varphi, E^-_\varphi)$ we construct are *weak simulation dualities*, i.e. dualities in **wSim** that partition the class of pointed models *exactly* as the models and non-models of φ.

5.2.1 Normal Forms

Now we'll give a normal form for our language $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, which is obtained as a restriction of the normal form for the full modal language once proved by Kit Fine.

Definition 5.2. (Fine Normal Form)
A *basic normal form* of level 0 is a non-empty[7] conjunction of *positive* literals.[8] A basic normal form of level $n+1$ is a non-empty conjunction of formulas

$$\pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)$$

where π is a (possibly empty) conjunction of positive (i.e. unnegated) propositional variables p, and each φ_i, ψ_j is a basic normal form of level at most n (this ensures that a normal form of level n is also a normal form of level n' for all $n' > n$). A *normal form* of level m is a non-empty[9] disjunction of basic normal forms of level m.

Intuitively, basic normal forms are modal formulas in negation normal form that *may have disjunctions under the scope of boxes but not under the scope of any diamonds.*

Theorem 5.6. *([16]) For every formula* $\xi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$*, there is a normal* $nf(\xi)$ *of level* $d(\xi)$ *such that* $\xi \equiv nf(\xi)$ *which can be effectively computed.*

[6] For instance, taking the size of a first order formula (and hence of CQs) to also be the number of its subformulas, then we clearly have $|\varphi| = |ST_x(\varphi)|$.
[7] Because $\top$ is the empty conjunction.
[8] The original normal form in [16] is formulated for *all* modal formulas.
[9] Because $\bot$ is the empty disjunction.

Proof. By induction on $d(\xi)$. If $d(\xi) = 0$ we can just use the disjunctive normal form for propositional logic, i.e. we simply distribute $\wedge$'s over $\vee$'s via the validity $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$. If $d(\xi) = n + 1$, observe that ξ is a Boolean combination of formulas $\Diamond\varphi$, $\Box\psi$ and propositional variables p, where each $d(\varphi), d(\psi) \leq n$. By the inductive hypothesis, we may assume that each such φ and ψ are normal forms of level n. This means that $\varphi \equiv \bigvee \Phi$ and $\psi \equiv \bigvee \Psi$ for some nonempty finite sets Φ, Ψ of basic normal forms of level n. By Kripke normality we get $\Diamond\varphi = \Diamond \bigvee \Phi \equiv \bigvee_{\varphi\in\Phi} \Diamond\varphi$ and thus ξ is equivalent to a Boolean combination of formulas of the form $\Diamond\varphi_i$, and $\Box\psi_j$ and (unnegated) propositional variables p, where each φ_i and ψ_j is a basic normal form of level n. Applying the propositional distribute law we obtain the disjunctive normal form of this Boolean combination, that is we get a disjunction of basic normal forms of the form

$$\pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n \wedge \wedge\Box(\psi_0 \vee \ldots \vee \psi_m)$$

where π is a (possibly empty) conjunction of positive atoms and each φ_i, ψ_j is a basic normal forms of level at most n. Hence, ξ is equivalent to a normal form of level $n + 1$. □

Corollary 5.3. *Every basic normal form of level n is of one of the following forms:*

$$\begin{array}{lll} (i) & p_1 \wedge \ldots \wedge p_n & \text{if } n = 0 \\ (ii) & \pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n & \text{if } n > 0 \\ (iii) & \pi \wedge \Box(\psi_0 \vee \ldots \vee \psi_m) & \text{if } n > 0 \\ (iv) & \pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n \wedge \Box(\psi_0 \vee \ldots \vee \psi_m) & \text{if } n > 0 \end{array}$$

where each φ_i, ψ_j is a basic normal form of level $n-1$ and π is a (possibly empty conjunction) of positive atoms.

5.2.2 Construction of Examples

Now we will give a recursive definition for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ formulas in Fine normal form. In particular, we will define the positive examples and show how to derive the negative examples from them exploiting the symmetry $(\cdot)^\neg$ on our category **wSim**. We define the following '*gluing*' operator (which we denote by $\blacktriangledown$ to suggest the connection with the syntactic ∇-operator) for building the positive examples.

Definition 5.3. Given a set[10] of pointed models $\mathbb{E}$ and a set of propositional variables $P \subseteq$ Prop let $\blacktriangledown_P(\mathbb{E})$ be the model obtained by gluing all the examples in $\mathbb{E}$ to a new root r with $v(r) = P$. That is:

- Domain $dom(\blacktriangledown_P(\mathbb{E})) = \biguplus_{(E,e)\in\mathbb{E}} dom(E) \uplus \{r\}$
- Relation $R^\blacktriangledown[r] = \{e \mid (E, e) \in \mathbb{E}\}$ and $R^\blacktriangledown \restriction dom(E) = R^E$ for all $(E, e) \in \mathbb{E}$
- Colouring $v^\blacktriangledown \restriction dom(E) = v^E$ for all $(E, e) \in \mathbb{E}$ and $v^\blacktriangledown(r) = P$

We write $\blacktriangledown_P(E, e)$ rather than $\blacktriangledown_P(\{(E, e)\})$ for singletons. By theorem 5.6, every $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ is equivalent to a normal form $nf(\varphi)$ of level $d(\varphi)$. In fact, we observed there (corollary 5) that every normal form is of the form (i)-(iv). Hence, we will define the positive

[10] Although all the models we work with will be finite, and will only glue together finite sets of them, note that this model-building operation is well-defined also for infinite sets.

examples for these four cases separately. Note that for a conjunction of literals π, $pos(\pi)$ denotes the set of positively occurring propositional variables in π. Further, we suppress the distinguished element for readability, writing E instead of (E, e).

Definition 5.4. (Positive Examples)
For every formula in $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ in Fine normal form, let:

$$\begin{aligned}
E^+_\pi &:= \{\blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset)\} \\
E^+_{\pi\wedge\Diamond\varphi_0\wedge\ldots\wedge\Diamond\varphi_n} &:= \{\blacktriangledown_{pos(\pi)}(\{E_1,\ldots,E_n\}\cup\{\circlearrowright_\emptyset\}) \mid E_i \in E^+_\varphi \text{ for all } i \leq n\} \\
E^+_{\pi\wedge\Box(\psi_0\vee\ldots\vee\psi_m)} &:= \{\blacktriangledown_{pos(\pi)}(\mathbb{E}) \mid \mathbb{E} \subseteq \bigcup_{j\leq m} E^+_{\psi_j}\} \\
E^+_{\pi\wedge\Diamond\varphi_0\wedge\ldots\wedge\Diamond\varphi_n\wedge\Box(\psi_0\vee\ldots\vee\psi_m)} &:= \{\blacktriangledown_{pos(\pi)}(\{E_1,\ldots,E_n\}\cup\mathbb{E}) \mid E_i \in \bigcup_{j\leq m} E^+_{\varphi\wedge\psi_j} \text{ for all } i\leq n \text{ and} \\
&\qquad\qquad \emptyset \neq \mathbb{E} \subseteq \bigcup_{\psi\in\Psi} E^+_\psi\} \\
E^+_{\bigvee_{i\leq n}\varphi_i} &:= \bigcup_{i\leq n} E^+_{\varphi_i}
\end{aligned}$$

Note that in the one but last line, $E^+_{\varphi\wedge\psi}$ is already defined by inductive hypothesis.[11] As it turns out, we can define the negative examples for formulas in $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ in terms of the positive ones via the substitution operator $(\cdot)^\neg$ we saw in chapter 3. We will do this in terms of the operator $\triangleleft$, which was introduced in [21] as a modality in the intuitionistic semantics of positive modal logic and can be seen as a weak form of negation.[12] However, in the classical interpretation its semantics can be defined in terms of the other operators, i.e. it reduces to a *syntactic operation*. Intuitively, $\triangleleft\varphi$ is the formula φ with all logical symbols in it 'dualized' while it does nothing on propositional variables. That is;

$$\begin{aligned}
\triangleleft(\Box\varphi) &= \Diamond \triangleleft (\varphi) \\
\triangleleft(\Diamond\varphi) &= \Box \triangleleft (\varphi) \\
\triangleleft(\varphi\wedge\psi) &= \triangleleft(\varphi) \vee \triangleleft(\psi) \\
\triangleleft(\varphi\vee\psi) &= \triangleleft(\varphi) \wedge \triangleleft(\psi) \\
\triangleleft(p) &= p
\end{aligned}$$

First of all, note that $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ implies that $\triangleleft(\varphi) \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ because $\triangleleft(p) = p$. Moreover, observe that $\neg\varphi \equiv \triangleleft(\varphi)^\neg$. To see this, note that rewriting $\neg\varphi$ into negation normal form by pushing the negation $\neg$ over all other logical symbols, 'dualizing' them and adding a negation to every literal encountered is precisely the same procedure as first dualizing all operators with $\triangleleft$ and then substituting $[\neg p/p]$ wherever possible with $(\cdot)^\neg$.

Definition 5.5. For each $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ set $\mathbb{E}_\varphi = (E^+_\varphi, E^-_\varphi)$ where

$$E^+_\varphi := E^+_{nf(\varphi)} \qquad\qquad E^-_\varphi := \{(E^\neg, e) \mid (E, e) \in E^+_{nf(\triangleleft(\varphi))}\}$$

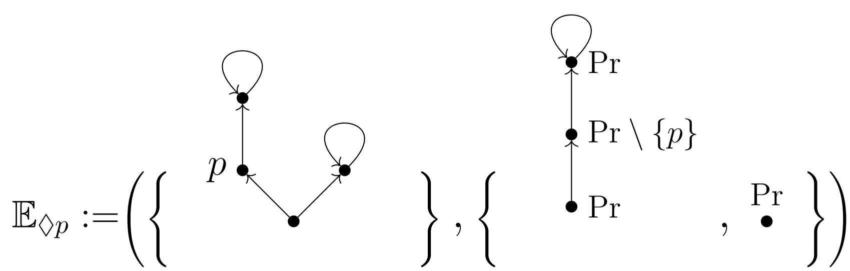
[11] Note the closeness of the sets $E^+_{\Diamond p} = \{\Diamond\widehat{(p\wedge q)}\}$ and $E^+_{\Box p} = \{\blacktriangledown_\emptyset(\emptyset), \blacktriangledown_\emptyset(\widehat{p})\}$ to their translation with the ∇ modality; $\Box p \equiv \nabla(p) \vee \nabla(\emptyset)$, $\Diamond p \equiv \nabla(p, \top)$.

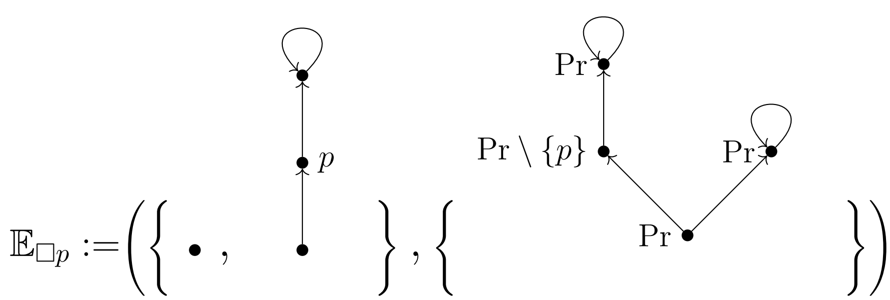
[12] The authors call this extension of (the intuitionistic version of) positive modal logic 'distributive modal logic'.

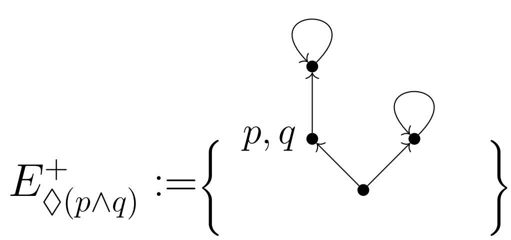
5.2.3 Examples

Each positive atom p has exactly one positive example $\widehat{p} := \blacktriangledown_{\{p\}}(\circlearrowleft_\emptyset)$ and one negative example $\blacktriangledown_{\mathrm{Prop}\setminus\{p\}}(\circlearrowleft_{\mathrm{Prop}})$. Note that $\blacktriangledown_{\mathrm{Prop}\setminus\{p\}}(\circlearrowleft_{\mathrm{Prop}}) = (\widehat{p})^\neg$. Similarly, conjunctions of positive atoms π also have a single positive example but more than one positive example, e.g. $E^+_{p\wedge q} = \{\widehat{pq}\}$ where $\widehat{pq} := \blacktriangledown_{\{p,q\}}(\circlearrowleft_\emptyset)$ but $E^-_{p\wedge q} = E^-_p \cup E^-_q = \{(\widehat{p})^\neg, (\widehat{q})^\neg\}$. Actually, it follows from definition 5.4 above that every $\varphi \in \mathcal{L}^+_{\Diamond,\wedge}$ has a single positive example $\widehat{\varphi}$, and hence by duality (cf. definition 5.5) that every $\psi \in \mathcal{L}^+_{\Box,\vee}$ has a single negative example. We will give some examples of the characterisations obtain this way, first via the $\blacktriangledown$-syntax and then visually. We will write Pr instead of Prop for spacing.

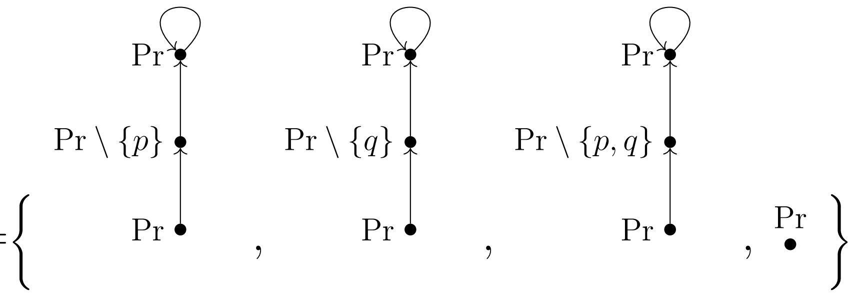
$$
\begin{aligned}
\mathbb{E}_{\Diamond p} &= (\{\blacktriangledown_\emptyset(\widehat{p}, \circlearrowleft_\emptyset)\}, \{\blacktriangledown_{\mathrm{Prop}}((\widehat{p})^\neg), \blacktriangledown_{\mathrm{Prop}}(\emptyset)\})\\
\mathbb{E}_{\Box p} &= (\{\blacktriangledown_\emptyset(\emptyset), \blacktriangledown_\emptyset(\widehat{p})\}, \{\blacktriangledown_{\mathrm{Prop}}((\widehat{p})^\neg, \circlearrowleft_{\mathrm{Prop}})\})\\
E^+_{\Diamond(p\wedge q)} &= \{\blacktriangledown_\emptyset(\widehat{pq}, \circlearrowleft_\emptyset)\}\\
E^-_{\Diamond(p\wedge q)} &= \{\blacktriangledown_{\mathrm{Prop}}(\emptyset), \blacktriangledown_{\mathrm{Prop}}((\widehat{p})^\neg), \blacktriangledown_{\mathrm{Prop}}((\widehat{q})^\neg), \blacktriangledown_{\mathrm{Prop}}((\widehat{p})^\neg, (\widehat{q})^\neg)\}\\
E^+_{\Box(p\vee q)} &= \{\blacktriangledown_\emptyset(\emptyset), \blacktriangledown_\emptyset(\widehat{p}), \blacktriangledown_\emptyset(\widehat{q}), \blacktriangledown_\emptyset(\widehat{p}, \widehat{q})\}\\
E^-_{\Box(p\vee q)} &= \{\blacktriangledown_{\mathrm{Prop}}((\widehat{pq})^\neg, \circlearrowleft_{\mathrm{Prop}})\}\\
E^+_{\Box(p\vee q)\wedge\Diamond r} &= \{\blacktriangledown_\emptyset(\widehat{pr}, \widehat{p}), \blacktriangledown_\emptyset(\widehat{pr}, \widehat{q}), \blacktriangledown_\emptyset(\widehat{pr}, \widehat{p}, \widehat{q}), \blacktriangledown_\emptyset(\widehat{qr}, \widehat{p}), \blacktriangledown_\emptyset(\widehat{qr}, \widehat{q}), \blacktriangledown_\emptyset(\widehat{qr}, \widehat{p}, \widehat{q})\}\\
E^-_{\Box(p\vee q)\wedge\Diamond r} &= E^-_{\Box(p\vee q)} \cup E^-_{\Diamond r} = \{\blacktriangledown_{\mathrm{Prop}}((\widehat{pq})^\neg, \circlearrowleft_{\mathrm{Prop}}), \blacktriangledown_{\mathrm{Prop}}((\widehat{r})^\neg), \blacktriangledown_{\mathrm{Prop}}(\emptyset)\}
\end{aligned}
$$



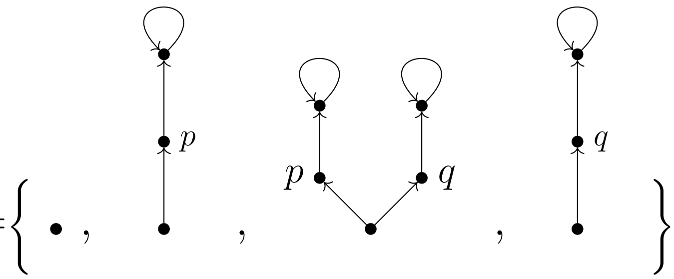




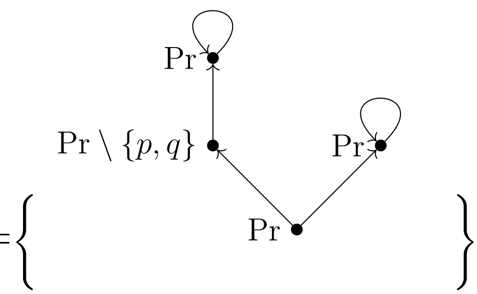
$E^{-}_{\Diamond(p\wedge q)} :=$



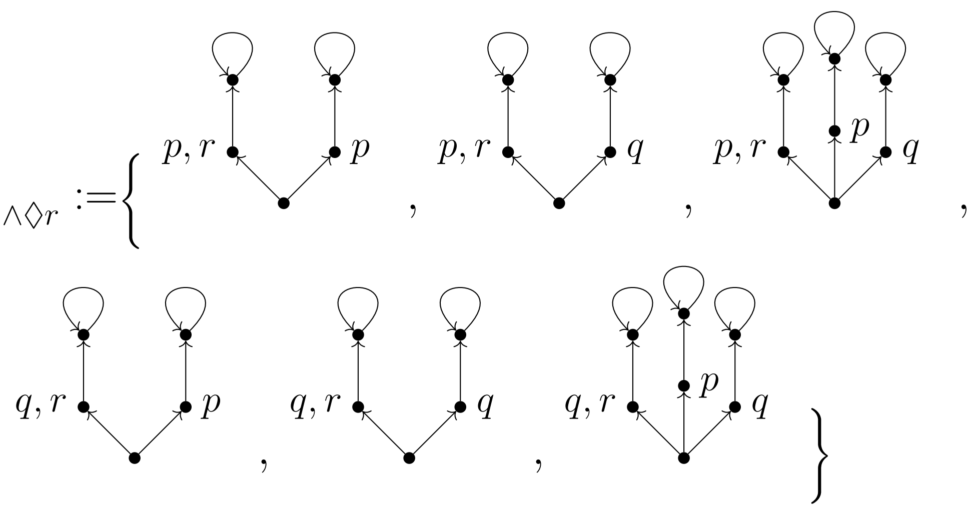
$E^{+}_{\Box(p\vee q)} :=$



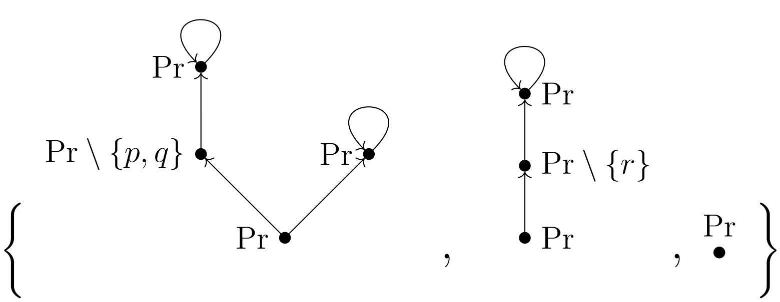
$E^{-}_{\Box(p\vee q)} :=$



$E^{+}_{\Box(p\vee q)\wedge\Diamond r} :=$



$E^{-}_{\Box(p\vee q)\wedge\Diamond r} :=$



5.2.4 Correctness of Characterisations

Now we will show the correctness of the previous construction, i.e. that for each $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, $\mathbb{E}_\varphi$ actually characterises φ w.r.t. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$.

Proposition 5.1. *Every $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ fits $\mathbb{E}_\varphi$.*

Proof. Let $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. We only have to check the cases (i)-(iv) from corollary 5.3. Case (i) is obvious from the definition of $\blacktriangledown_P(\cdot)$ operator. (ii) Let $\varphi = \pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n$. Then every example $\blacktriangledown_{pos(\pi)}(\{E_1, \ldots, E_n\} \cup \{\circlearrowright_\emptyset\}) \in E^+_\varphi$ is positive for φ since each $E_i \in E^+_{\varphi_i}$ is by inductive hypothesis a positive example for φ_i because each conjunct $\Diamond\varphi_i$ is witnessed by some successor e_i and the root satisfies all $p \in pos(\pi)$.[13] (iii) Let $\varphi = \pi \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)$ and consider an example $\blacktriangledown_{pos(\pi)}(\mathbb{E}) \in E^+_\varphi$. Note that by inductive hypothesis all $(E.e) \in \mathbb{E}$ are positive examples for some ψ_j. Hence this example is positive for φ as the root of this model satisfies all $p \in pos(\pi)$ and only has successors in $\mathbb{E}$. (iv) Let $\varphi = \pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \varphi_n \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)$ and consider an example $\blacktriangledown_{pos(\pi)}(\{(E_1, e_1), \ldots, (E_n, e_n)\} \cup \mathbb{E}) \in E^+_\varphi$. By inductive hypothesis, each (E_i, e_i) is a positive example for $\varphi_i \wedge \psi_j$ for some j and each $(E, e) \in \mathbb{E}$ is a positive example for some ψ_j. It follows that $\blacktriangledown_{pos(\pi)}(\{(E_1, e_1), \ldots, (E_n, e_n)\} \cup \mathbb{E})$ is a positive example for φ because each existential conjunct $\Diamond\varphi_i$ has a witness e_i and each successor e satisfies some ψ_k. Finally, the root of this model satisfies each $p \in pos(\pi)$ by construction.

For the negative examples, observe that $(E, e) \in E^-_\varphi$ means that $(E^\neg, e) \in E^+_{nf(\lhd(\varphi))}$ where $nf(\lhd(\varphi)) \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, so by the above argument $E^\neg, e \models nf(\lhd(\varphi))$. It follows from theorem 4.6 that $E, e \models nf(\lhd(\varphi))^\neg$. But we already noted that $\neg\varphi \equiv \lhd(\varphi)^\neg$ and hence $E, e \not\models \varphi$. $\Box$

Theorem 5.7. *For every $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, every positive example for φ weakly simulates some example from E^+_φ.*

Proof. By induction on formula complexity. Consider a $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ in normal form, i.e. φ is a disjunction of basic normal forms of level $d(\varphi)$. It suffices to show that the claim holds for all basic normal form disjuncts of φ, for if $M, s \models \varphi$ then M, s satisfies some disjunct, and the positive examples for φ are the union of all the positive examples for its disjuncts.

(i) Let $\varphi = \pi$ be a conjunction of positive atoms and observe that $E^+_{pos(\pi)} = \{\blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset)\}$. We look at 2 cases: (i) $R^M[s] = \emptyset$ or (ii) $R^M[s] \neq \emptyset$.

(i) The claim we make here is that the relation with a single pair $\{(r, s)\}$ is a weak simulation $\blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset) \underline{\rightarrow} M, s$. [atom] $v(r) = pos(\pi)$ and as $M, s \models \pi$ clearly $v(r) \subseteq v(s)$. [forth'] trivial since we can always use the escape clause. [back'] trivial since $R^M[s] = \emptyset$. (ii) Now we claim that $Z := \{(r, s)\} \cup (\{\circlearrowright_\emptyset\} \times dom(M)\}$ is a weak simulation $Z : \blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset) \underline{\rightarrow} M, s$. It follows straight from the definition of weak simulation that each pair of the form $(\circlearrowright_\emptyset, t)$ (suggestive notation, where $\circlearrowright_\emptyset$ denotes a state in our example whose generated submodel is bisimilar to $\circlearrowright_\emptyset$) satisfies all the clauses. Hence it suffices to show that all clauses hold of the root-pair $(r, s) \in Z$. [atom] $v(r) = pos(\pi)$ and as $M, s \models \pi$ clearly $v(r) \subseteq v(s)$. [forth'] the only successor of r is $\circlearrowright_\emptyset$ so we can just use the escape clause. [back'] If $R^M st$ we clearly have the empty loopstate successor of r with is Z-related to $t \in dom(M)$.

[13] Recall that we only denoted pointed models (E_i, e_i) by a single letter E_i in definition 5.4 for the sake of readability.

(ii) Let $M, s \models \pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n$. By inductive hypothesis, this means that there are examples $(E_i, e_i) \in E^+_{\varphi_i}$ for each $1 \leq i \leq n$ with weak simulations $Z_i : E_i, e_i \underline{\rightarrow} M, t_i$, where $t_i \in R^M[s]$, as well as a simulation $Z_\pi : \blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset) \underline{\rightarrow} M, s$. It follows that $\blacktriangledown_{pos(\pi)}(\{(E_1, e_1), \ldots, (E_n, e_n)\}) \in E^+_{\pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n}$. The claim is that

$$Z := \bigcup_{1 \leq i \leq n} Z_i \cup (\{\circlearrowright_\emptyset\} \times dom(M)) \cup \{(r, s)\}$$

is a weak simulation $\blacktriangledown_{pos(\pi)}(\{(E_1, e_1), \ldots, (E_n, e_n)\}) \underline{\rightarrow} M, s$. We already know that all the Z_i are weak simulations. Moreover, every pair of the form $(\circlearrowright_\emptyset, t)$ satisfies all the clauses (cf. lemma 4.1). Hence we only have to show that the root-pair (r, s) satisfies all the clauses. [atom] By the [atom] clause of Z_π (note that $pos(\pi)$ may be empty here) [forth'] For the empty loopstate successor of r we can use the escape clause, and every other successor of r is of the form e_i for some $i \leq n$. But then there is a matching successor t_i of s with $(e_i, u_i) \in Z_i \subseteq Z$. [back'] We can match every successor u of t with the empty loopstate-successor or r since $(\circlearrowright_\emptyset, u) \in Z$ for all $u \in dom(M)$.

(iii) Let $M, s \models \pi \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)$. By inductive hypothesis, this means that for each $t \in R^M[s]$ there is some example $(E_t, e_t) \in \bigcup_{j \leq m} E^+_{\psi_j}$ with a weak simulation $Z_t : E_t, e_t \underline{\rightarrow} M, t$, as well as simulation $Z_\pi : \blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset) \underline{\rightarrow} M, s$. Let $\mathbb{E} := \{(E_t, e_t) \mid t \in R^M[s]\}$, then $\blacktriangledown_{pos(\pi)}(\mathbb{E}) \in E^+_{\pi \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)}$. The claim is that

$$Z := \bigcup_{t \in R^M[s]} Z_t \cup \{(r, s)\}$$

is a weak simulation $Z : \blacktriangledown_{pos(\pi)}(\mathbb{E}) \underline{\rightarrow} M, s$. We already know each Z_t is a weak simulation, so it suffices to show that (r, s) satisfies all the clauses. [atom] By the [atom] clause of Z_π (note that $pos(\pi)$ may be empty here). [forth'] every successor of r is of the form e_t and hence there is some successor t of s with $(e_t, t) \in Z_t \subseteq Z$. [back'] For every successor t of s we have by construction a successor e_t of r with $(e_t, t) \in Z_t \subseteq Z$.

(iv) Let $M, s \models \pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)$. By inductive hypothesis, this means that for each $t \in R^M[s]$ there is some example $(E_t, e_t) \in \bigcup_{j \leq m} E^+_{\psi_j}$ with a weak simulation $Z_t : E_t, e_t \underline{\rightarrow} M, t$. Moreover, $R^M[s] \neq \emptyset$ as e.g. $M, s \models \Diamond\varphi_0$ and $\varphi_0 \not\equiv \bot$. Set $\mathbb{E} := \{(E_t, e_t) \mid t \in R^M[s]\}$ and note that $\mathbb{E} \neq \emptyset$ as e.g. $M, s \models \Diamond\varphi_0$ and $\varphi_0 \not\equiv \bot$. Moreover, by inductive hypothesis there is also a weak simulation $Z_\pi : \blacktriangledown_{pos(\pi)}(\circlearrowright_\emptyset) \underline{\rightarrow} M, s$ and for each $i \leq n$ there are examples $(E_i, e_i) \in \bigcup_{j \leq m} E^+_{\varphi_i \wedge \psi_j}$ such that there is a simulation $Z_i : E_i, e_i \underline{\rightarrow} M, t_i$ where t_i is some successor of s. Set $\mathbb{E}' := \{(E_i, e_i) \mid 0 < i \leq n\}$. It follows that $\blacktriangledown_{pos(\pi)}(\mathbb{E} \cup \mathbb{E}') \in E^+_{\pi \wedge \Diamond\varphi_0 \wedge \ldots \wedge \Diamond\varphi_n \wedge \Box(\psi_0 \vee \ldots \vee \psi_m)}$. We claim that the relation

$$Z := \bigcup_{i \leq n} Z_i \cup \bigcup_{t \in R^M[s]} Z_t \cup \{(r, s)\}$$

is a weak simulation $Z : \blacktriangledown_{pos(\pi)}(\mathbb{E} \cup \mathbb{E}') \underline{\rightarrow} M, s$. Again, since we already know each Z_i and Z_t to be weak simulations, it suffices to show that the root-pair (r, s) satisfies all the clauses. [atom] By the [atom] clause of Z_π (note that $pos(\pi)$ may be empty here). [forth'] Every successor of the root is either (i) of the form e_t or (ii) of the form e_i. If (i) then we have a matching successor t of s with $(e_t, t) \in Z_t \subseteq Z$, and if (ii) then we have a matching successor t_i of s with $(e_i, t_i) \in Z_i \subseteq Z$. [back'] every successor t of s has a matching successor e_t of r with $(e_t, t) \in Z_t \subseteq Z$.

□

Theorem 5.8. *For all* $\varphi, \psi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, ψ *fits* E^+_φ *iff* $\varphi \models \psi$.

Proof. Let $\varphi, \psi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. From left to right, suppose that ψ fits E^+_φ and let $M, s \models \varphi$. Then M, s is a positive example for φ and hence by theorem 5.7 there is a positive example $(E, e) \in E^+_\varphi$ with $E, e \underline{\rightarrow} M, s$. But as ψ fits E^+_φ, in particular $E, e \models \psi$ from which follows that $M, s \models \psi$ by preservation of $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ formulas under weak simulation (i.e. theorem 4.2). Conversely, suppose that $\varphi \models \psi$. By soundness (proposition 5.1 φ fits E^+_φ so by the entailment ψ fits these as well. □

Theorem 5.9. *For every* $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, *every negative example of* φ *is weakly simulated by some example in* E^-_φ.

Proof. Suppose that $M, s \not\models \varphi$ for some $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. Then $M, s \models \neg\varphi$ and thus $M^\neg, s \models (\neg\varphi)^\neg$ by theorem 4.6. But as $\neg\varphi \equiv \triangleleft(\varphi)^\neg$ also $(\neg\varphi)^\neg \equiv \triangleleft(\varphi)$ because $(\cdot)^\neg$ on formulas is also idempotent. Hence $M^\neg, s \models \triangleleft(\varphi)$ where $\triangleleft\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ so by theorem 5.7 there is some positive example $(E, e) \in E^+_{\triangleleft(\varphi)}$ and a weak simulation $Z : E, e \underline{\rightarrow} M^\neg, s$. But then by proposition 4.2, this implies that $Z^{-1} : M, s \underline{\rightarrow} E^\neg, e$ since $M^{\neg\neg} = M$. But as $(E, e) \in E^+_{\triangleleft(\varphi)}$ by definition $(E^\neg, e) \in E^-_\varphi$. So every negative example of φ is weakly simulated by some example in E^-_φ. □

Theorem 5.10. *For all* $\varphi, \psi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, ψ *fits* E^-_φ *iff* $\psi \models \varphi$.

Proof. Let $\varphi, \psi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ and note that $E^-_\varphi = \{(E, e)^\neg \mid (E, e) \in E^+_{\triangleleft(\varphi)}\}$ by definition 5.5 (where we assume φ and hence $\triangleleft(\varphi)$ to in Fine normal form). It follows that ψ fits E^-_φ (i.e. ψ is false on all $(E, e) \in E^-_\varphi$) iff $\neg\psi$ is true on all $(E, e) \in E^-_\varphi$ iff $(\neg\psi)^\neg$ is true on all $(E, e) \in E^+_{\triangleleft(\varphi)}$ (by theorem 4.6) iff $(\neg\psi)^\neg$ fits $E^+_{\triangleleft(\varphi)}$. But $(\neg\psi)^\neg$ fits $E^+_{\triangleleft(\varphi)}$ iff $\triangleleft(\varphi) \models (\neg\psi)^\neg$ by theorem 5.8. Finally, we noted before that $\triangleleft(\chi)^\neg \equiv \neg\chi$ and $\chi^{\neg\neg} \equiv \chi$ for all formulas χ. Moreover, it follows from theorem 4.6 that $\chi \models \xi$ iff $\chi^\neg \models \xi^\neg$. Bringing everything back together, we see that the entailment $\triangleleft(\varphi) \models (\neg\psi)^\neg$ holds iff $\triangleleft(\varphi)^\neg \models \neg\psi$ iff $\neg\varphi \models \neg\psi$ iff $\psi \models \varphi$. □

Theorem 5.11. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ *is finitely characterisable. Moreover, the characterisations can be effectively constructed.*

Proof. Let $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. By theorem 5.6, we can effectively compute the Fine normal form $nf(\varphi) \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ of φ. Definition 5.4 together with definition 5.5 give us an algorithm for computing the finite characterisation $(E^+_\varphi, E^-_\varphi)$ of φ (we will come back to complexity issues in the next chapter). By soundness (proposition 5.1), φ fits these examples, so it rests to show that it is the *only* formula from $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ (up to logical equivalence) that fits these examples. So suppose $\psi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ also fits these examples. By theorem 5.8 it follows that $\varphi \models \psi$ and by theorem 5.10 $\psi \models \varphi$, hence $\varphi \equiv \psi$. □

5.3 Language Extensions

Now we will apply the techniques and results from above to characterise larger classes of modal formulas. We begin with a simple observation.

Theorem 5.12. $\mathcal{L}^-_{\Box,\Diamond,\wedge,\vee}$ *is finitely characterisable. Moreover, the characterisations can be effectively constructed*

Proof. Let $\varphi \in \mathcal{L}^-_{\Box,\Diamond,\wedge,\vee}$, then $\neg\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ so by theorem 5.11 it has a finite characterization $\mathbb{E}_{\neg\varphi} = (E^+_{\neg\varphi}, E^-_{\neg\varphi})$ w.r.t. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ which can be efficiently constructed. It follows that $(E^-_{\neg\varphi}, E^+_{\neg\varphi})$ is a finite characterisation of φ w.r.t. $\mathcal{L}^-_{\Box,\Diamond,\wedge,\vee}$ because every $\psi \in \mathcal{L}^-_{\Box,\Diamond,\wedge,\vee}$ fits $(E^-_{\neg\varphi}, E^+_{\neg\varphi})$ iff $\neg\psi$ fits $(E^+_{\neg\varphi}, E^-_{\neg\varphi})$ iff $\neg\psi \equiv \neg\varphi$ iff $\psi \equiv \varphi$. □

In fact, the preceding argument is completely general.

Corollary 5.4. *If a language* $\mathcal{L}$ *is finitely characterisable, so is* $\mathcal{L}^\neg = \{\neg\varphi \mid \varphi \in \mathcal{L}\}$.

Recall from the preliminaries that $pos(\varphi), neg(\varphi)$ denote the set of variables occurring positively, respectively negatively in φ. The terminology for the following definition is taken from [8].

Definition 5.6. (Uniform Formulas)
A formula φ is called *uniform* if $pos(\varphi) \cap neg(\varphi) = \emptyset$. That is, each propositional variable $p \in var(\varphi)$ either occurs only positively in φ (whence $p \in pos(\varphi) \setminus neg(\varphi)$), or only negatively (whence $p \in neg(\varphi) \setminus pos(\varphi)$).

Note that if φ is uniform then $pos(\varphi)$ together with $neg(\varphi)$ partitions $var(\varphi)$ into two cells. Hence we will write $\varphi(p_1, \ldots, p_n; q_1, \ldots, q_m)$ or $\varphi(P; Q)$ to denote that $pos(\varphi) \subseteq \{p_1, \ldots, p_n\} = P$ and $neg(\varphi) \subseteq \{q_1, \ldots, q_m\} = Q$, where $P \cap Q = \{p_1, \ldots, p_n\} \cap \{q_1, \ldots, q_m\} = \emptyset$ are disjoint because φ is uniform. For a class of modal formulas $\mathcal{L}$, let $\mathcal{L}^u$ denote the set of all uniform $\mathcal{L}$-formulas. Note that uniform formulas thus may contains negations, but only in front of propositional variables.

Theorem 5.13. $\mathcal{L}^u_{\Box,\Diamond,\wedge}$ *is finitely characterisable. Moreover, the characterisations can be effectively constructed*

Proof. Consider a uniform formula $\varphi(P; Q) \in \mathcal{L}^u_{\Box,\Diamond,\wedge}$ where $P = \{p_1, \ldots, p_n\}$ and $Q = \text{Prop}\setminus P$ Observe that $\varphi^{\neg Q} \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ making every negative occurrence of some q in φ positive in $\varphi^{\neg Q}$. By theorem 5.11, $\varphi^{\neg Q}$ has a finite characterisation $\mathbb{E}_{\varphi^{\neg Q}} = (E^+_{\varphi^{\neg Q}}, E^-_{\varphi^{\neg Q}})$ w.r.t. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. Set

$$\mathbb{E}_\varphi := (\{\circlearrowright_P\} \cup \{(E^{\neg Q}, e) \mid (E, e) \in E^+_{\varphi^{\neg Q}}\},\ \{(E^{\neg Q}, e) \mid (E, e) \in E^-_{\varphi^{\neg Q}}\})$$

Now let $\psi \in \mathcal{L}^u_{\Box,\Diamond,\wedge}$ be some formula fitting these examples. Since $\circlearrowright_P \models \psi$ and ψ does not contain disjunctions, it must be that $pos(\psi) \subseteq P$ and $neg(\psi) \subseteq Q$ since $neg(\psi) \cap P = \emptyset$.[14] Hence $\psi(P; Q)$ is of a compatible 'uniform type' as φ (both have positive variables contained in P and negative variables disjoint from P, i.e. contained in Q). Now consider a (positive or negative) example $(E, e) \in \mathbb{E}_{\varphi^{\neg Q}}$ and note that $\psi^{\neg Q} \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. By theorem 4.6, ψ fits $E^{\neg Q}, e$ iff $\psi^{\neg Q}$ fits E, e. It follows that $\psi^{\neg Q}$ fits all the examples in $\mathbb{E}_{\varphi^{\neg Q}}$, but as this is a characterisation of $\varphi^{\neg Q}$ w.r.t. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, it must be that $\varphi^{\neg Q} \equiv \psi^{\neg Q}$ and hence $\varphi \equiv \psi$ by theorem 4.6. □

In fact, we can redo much of the categorical theory for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ developed in chapter 4 for uniform formulas. Defining for each 'uniform type' (P, Q) (with $P \cap Q = \emptyset$ and $P \cup Q = \text{Prop}$)

[14] The proof of this follows the same steps as the proof of lemma 4.2, and in fact can be seen as a generalisation of it.

the notions of weak (P, Q)-simulation, (P, Q)-normal forms and obtain the category of pointed models and weak (P, Q)-simulations. However, disjunction of uniform formulas with partially overlapping types are a problem for extending our results to the class of uniform formulas *with* disjunction $\mathcal{L}^u_{\Box,\Diamond,\wedge,\vee}$. However, we keep finite characterisability if allow disjunction but restrict to a *single* uniform type.

Theorem 5.14. *For every* $P, Q \subseteq \mathrm{Prop}$ *with* $P \cap Q = \emptyset$ *and* $P \cup Q = \mathrm{Prop}$*, the fragment* $\mathcal{L}^{(P,Q)}_{\Box,\Diamond,\wedge,\vee} := \{\varphi \in \mathcal{L}^u_{\Box,\Diamond,\wedge,\vee} \mid pos(\varphi) \subseteq P, neg(\varphi) \subseteq Q\}$ *is finitely characterisable.*

Proof. Let $\varphi(P; Q) \in \mathcal{L}^{(P,Q)}_{\Box,\Diamond,\wedge,\vee}$ (so φ may contain disjunctions this time). Then $\varphi^{\neg Q} \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ so by theorem 5.11 it has a finite characterisation $\mathbb{E}_{\varphi^{\neg Q}} = (E^+_{\varphi^{\neg Q}}, E^-_{\varphi^{\neg Q}})$ w.r.t. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$. Now set $\mathbb{E}_\varphi := (\{(E^{\neg Q}, e) \mid (E, e) \in E^+_{\varphi^{\neg Q}}\}, \{(E^{\neg Q}, e) \mid (E, e) \in E^-_{\varphi^{\neg Q}}\}$; we claim this characterises φ w.r.t. $\mathcal{L}^u_{\Box,\Diamond,\wedge,\vee}$. By theorem 4.7 it follows that $\varphi^{\neg Q}$ fits $\mathbb{E}_\varphi$. Moreover, if $\psi \in \mathcal{L}^u_{\Box,\Diamond,\wedge,\vee}$ fits $\mathbb{E}_\varphi$ then $\psi^{\neg Q} \in \mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ fits $\mathbb{E}_{\varphi^{\neg Q}}$ so as the latter is a characterisation w.r.t $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$, $\psi^{\neg Q} \equiv \varphi^{\neg Q}$ and hence $\psi \equiv \varphi$ by theorem 4.7. □

Our finite characterisability results for fragments in theorems 5.11,5.12,5.13 and 5.14 above give rise to the following learnability results via the naive learning algorithm.

Corollary 5.5. $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}, \mathcal{L}^-_{\Box,\Diamond,\wedge,\vee}, \mathcal{L}^u_{\Box,\Diamond,\wedge}$ *and* $\mathcal{L}^{(P,Q)}_{\Box,\Diamond,\wedge,\vee}$ *are all exact learnable with membership queries.*

In the next chapter, we will give a quick summary of our results and provide further discussion, as well as indicate directions for future research.

Chapter 6

Discussion

Before entering into discussion proper, we give a short summary of the contributions of this thesis.

- We situated the relatively novel notion of finite characterisations in the context of formal learning theory, showing that finite characterisability of a concept x means that the $\{x\}$ is an isolated point in the corresponding intersection space. This has the benefit of bridging the gap between two relatively disjoint research communities (i.e. computational learning theory and formal learning theory).
- We spelled out in detail the proof of finite characterisability of the class of UCQs, which was left implicit in the literature. Moreover, we were able to show that UCQs are exactly learnable with membership queries in exponential time via the naive learning algorithm.
- We have shown that no formula in the full modal language is characterisable. More generally, we have shown how the full modal language is rarely ever fully finitely characterisable; this only happens when there are finitely many formulas up to logical equivalence.
- We identified a seemingly odd fragment $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ of the positive fragment $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee,\top,\bot}$ of basic modal logic, and a corresponding weakening of simulations and proved a semantic preservation theorem for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ w.r.t. these 'weak simulations'.
- Subsequently, we showed that these weak simulations and pointed models give rise to a category **wSim**. We showed that this category has a initial and final object, and moreover has an interesting symmetry $(\cdot)^\neg$.
- We related dualities in **wSim** to finite characterisations of $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$-formulas, in analogy with the connection between frontiers (dualities) in the category **Hom** (in fact it is a lattice) and finite characterisations of CQs (UCQs) exploited in [9].
- Then, we give an explicit construction of finite characterisations for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$-formulas in Fine normal form by showing these sets of examples form weak simulation dualities. Interestingly, the symmetry $(\cdot)^\neg$ was used to generate the negative examples out of the positive ones. Although this construction is non-elementary, it gives rise to a naive exact learning algorithm with membership queries for $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$.
- We showed how our results generalised to various uniform fragments that allow for limited form of negation.

The rest of this section will be devoted to discussion of topics that would have distracted from the story line developed in the main text. In particular, we indicate promising directions for future research.

6.1 Complexity

The time complexity of our algorithm for computing the characterisation $\mathbb{E}_\varphi$ of a formula in $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ cannot be bounded by a fixed tower of exponentials. For it is easily seen from definition 5.4 that $|E^+_{\Box\varphi}| = 2^{|E^+_\varphi|}$. It follows that $|\mathbb{E}_{\Box^n p}| \geq |E^+_{\Box^n p}| = itexp(n,1)$, where $itexp(0,k) = n$ and $itexp(i+1,k) = 2^{itexp(i,k)}$. This can be taken to show that we cannot give a sensible *upper bound* to our construction; even a simple formula like $\Box^n p$ already gets *non-elementary many positive examples*. Consequently, the same holds for the time complexity of the naive learning algorithms from corollary 5.5. We also did not manage to prove matching *lower bounds*.

However, we can show that interesting facts follows from the existence of polynomial size characterisations. Namely, if a modal fragment $\mathcal{L}$ has polynomial-sized characterisations, then it follows that equivalence testing for $\mathcal{L}$ is in NP.[1] For given $\varphi, \psi \in \mathcal{L}$, we can non-deterministically *guess* a polynomial size characterisation (E^+, E^-) of φ w.r.t. $\mathcal{L}$, and then model check ψ on all these examples. This can be done in polynomial time because model checking modal formulas can be done in polynomial time [8]. But we clearly have $\varphi \equiv \psi$ iff ψ fits (E^+, E^-) because (E^+, E^-) is a characterisation, hence equivalence for this fragment is in NP. In fact, if a characterisation can be computed in polynomial time from any formula $\varphi \in \mathcal{L}$, then it follows by the same argument that equivalence testing can be done in polynomial time. By the contrapositive, if we can show that equivalence testing for some fragment is *not* in NP, then it follows that fragment cannot have polynomial size characterisations.

Question: Is $\mathcal{L}^+_{\Box,\Diamond,\wedge,\vee}$ polynomial-time exact learnable?

6.2 Learning via Frontiers versus Learning via Dualities

We have essentially seen two approaches to learning formulas via examples; via *frontiers* and via *dualities*. The difference between these two approaches is best explained in terms of the characterisations that they induce. Theorem 2.5 demonstrated the correspondence between frontiers in **Hom** and finite characterisations of CQs, i.e. taking the frontier as the set of negative examples. However *disjunctions* of CQs (i.e. UCQs) in general require a *set* of positive examples. Consequently, the notion of a frontier does not make sense in this context, but dualities do. Theorem 2.11 demonstrated the correspondence between left-hand sides of dualities in **Hom** and finite characterisations of UCQs. We have also seen that, while frontiers could be computed in polynomial time, computing dualities in general requires exponential time.[2] Moreover, the transformation from dual to frontier (for single objects on the left-hand side) in proposition 2.3 is polynomial while the transformation from frontier to dual in proposition 2.4 is exponential. Thus, the lesson here is that dualities and hence disjunctions are costly. Therefore, in view of the above question, we want to see whether an approach via frontiers is

[1]Equivalence testing (for $\mathcal{L}$) is the problem of checking, on input formulas φ, ψ (from $\mathcal{L}$), whether $\varphi \equiv \psi$.

[2]It follows that these frontier are also of *polynomial size* while the dualities are of *exponential size*.

possible for $\mathcal{L}^+_{\Box,\Diamond,\wedge}$.

Question: Is there a learning algorithm for $\mathcal{L}^+_{\Box,\Diamond,\wedge}$ going via frontiers?

The polynomial time learning algorithm for c-acyclic CQs in [9] (that we discussed in some detail in section 2.4) relies on three key ingredients: (i) a final object, (ii) products and (iii) that a single positive example suffices for characterizing a CQ (cf. 2.4). It seems that all but the first one are missing. However, we will discuss the existence of products in **wSim** next.[3]

6.3 Products and Coproducts via Bisimulation Products

It is a recurring theme in the categories we consider that there are no single objects fulfilling the role of common categorical limits and colimits, but a *set* of objects fulfilling this role.[4] For instance, the exponential of two structures in **Hom** with at least 2 distinguished elements is in general a *set* of structures [9]. Consequently, we believe that the product of two pointed models in **wSim** in the sense of a categorical limit can only be fulfilled by a *set* of pointed models. The intuition is that there are more than one mutually exclusive ways of weakly simulating two models. For instance, one can simulate the positive example $\blacktriangledown_\emptyset(\widehat{p},\widehat{q})$ for $\Box(p \vee q)$ and the positive example $\blacktriangledown_\emptyset(\widehat{r},\circlearrowright_\emptyset)$ for $\Diamond r$ in at least two ways; i.e. witnessing r at a p-successor or at a q-successor (cf. $E^+_{\Box(p\vee q)\wedge\Diamond r}$ in subsection 5.2.3). A similar comment can be made about **Sim**.

Hence, we believe that the generalised set-product in **Sim** should be an appropriate generalisation of *bisimulation products* to models. For **wSim**, we would also need to define an appropriate notion of 'weak frame bisimulation' underlying weak simulations. Moreover, like for **Rel**, the symmetry imposed by $(\cdot)^\neg$ requires that the coproduct is the $(\cdot)^\neg$ image of the product. Thus in **Rel**, disjoint union of sets plays the role of product as well as coproduct.[5] Observe for instance that $(\circlearrowright_\emptyset)^\neg = \circlearrowright_{\mathrm{Prop}}$ so that in **wSim** the weak final object is indeed the $(\cdot)^\neg$ image of the weak initial object. Thus we believe that the set-coproducts in **Sim** and **wSim** should also be some form of bisimulation products, since $(\cdot)^\neg$ does not act on the frame.

6.4 Characterisations for Positive Existential Modal Logic

In this thesis, we have demonstrated two ways of obtaining finite characterisations for $\mathcal{L}^+_{\Diamond,\wedge,\vee}$, one via homomorphism dualities and the other via weak simulation dualities. While the former is only singly exponential, the latter is non-elementary but results in rather simple models called *looptrees*. Looptrees are a simple generalisation of trees that allows for loops at leaf-nodes.

[3] Also, note that it follows from definition 5.4 that every $\varphi \in \mathcal{L}^+_{\Box,\Diamond,\wedge}$ only gets a single *serial* positive example under our construction.

[4] One may wonder whether restricting the underlying class of objects of our category (in such a way that all distinct concepts will still be distinguished by some object) will result in stronger categorical constructs (such as an actual initial object). We believe that there might exist *weak simulations cores* in **wSim** (as well as simulation cores in **Sim**), in analogy to the graph cores in **Hom** from [23] and bisimulation cores (a.k.a. bisimulation contractions) in modal logic.

[5] In fact, we note that $(\cdot)^\neg$ satisfies almost all axioms of a *dagger*, except that it is not the identity on objects. Just like the isomorphism of categories 'dagger' $\dagger : \mathbf{Rel} \to \mathbf{Rel}$, the operation on morphisms, which are binary relations Z, is just the converse, i.e. $\dagger(Z) = (Z)^\neg = Z^{-1}$.

6.5 Remaining Modal Fragments

We have seen in section 5 that adding $\bot$ as a connective to $\mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee}$ already makes the language non-characterisable. However, we conjecture that $\mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee,\top}$ is still finitely characterisable.

Open Problem: Are $\mathcal{L}_{\Diamond,\wedge,\vee}, \mathcal{L}_{\Box,\wedge,\vee}, \mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee,\top}, \mathcal{L}^{u}_{\Box,\Diamond,\wedge,\vee}$ finitely characterisable?

Apart from this few questions left, we have completely solved the problem of characterising which *syntactically defined* modal fragments $\mathcal{L}^{+}_{C}, \mathcal{L}^{-}_{C}, \mathcal{L}_{C}$ or $\mathcal{L}^{u}_{C}$ are finitely characterisable (where $C \subseteq \{\Box, \Diamond, \wedge, \vee, \top, \bot\}$ a set of base connectives). Beyond this setting, one might to consider add e.g. $\rightarrow$ as a connective (under the classical semantics). This brings us into the larger arena of determining which *semantically defined* modal fragments are finitely characterisable, by which we mean the class of all fragments that can be obtained by allowing any modal formula $\varphi(p_1, \ldots, p_n)$ as an n-ary connective. We hope to gain something like Dalmau's dichotomy theorem for fragments of propositional logic [13]. However, this results relies strongly on a known characterisation of *Post's lattice* while it there is known equivalent of that characterisation in the modal setting.

6.6 Relationship to Description Logic

Modal logic is intimately related to *description logics*. In fact, description logics are just polymodal versions of modal languages. Some description logics correspond to *fragments* of the full modal language, while others correspond to *extensions* of the modal language with e.g. counting modalities (a.k.a. graded modalities) or backwards modalities (a.k.a. backward-looking modalities), i.e. modalities for the inverse relation R^{-1}. For instance, the description logic $\mathcal{EL}$ corresponds to the polymodal version of the fragment $\mathcal{L}^{+}_{\Diamond,\wedge}$ we considered here. It was shown in [9] that the description logic $\mathcal{ELI}$ (the extension of $\mathcal{EL}$ with backwards modalities) is polynomial-time exactly learnable with membership queries only. Subsequently, it has been shown that this logic remains efficiently learnable under various ontologies [19, 20], and characterisation results have been obtained for temporal extensions (formulated as a fragment of LTL) of this description logic and variants. Moreover, the characterisation result for XML path queries from [31] can also be interpreted as characterisation results on a certain model language that includes the $\Diamond^*$ connective, the modality for the reflexive-transitive closure R^* of the accessibility relation R.

We believe that, although more tedious, the results in this thesis should easily extend to polymodal versions of the modal languages considered here. In particular, we conjecture that the polymodal version of $\mathcal{L}^{+}_{\Box,\Diamond,\wedge,\vee}$, which is an extension of $\mathcal{EL}$ with 'value restriction' (this is just $\Box$) and disjunctions [3], is finitely characterisable. It seems worthwhile to also extend our results to the case with inverse, counting or backward-looking modalities (or any combination of these). Another interesting direction is to look at the problem of learning modal concepts under *semantic constraints* (called 'T-boxes' in the description logic community [3]) in analogy with the restricted frame classes we considered in chapter 3. However, instead of restricting the class of examples by their underlying frame, the class of examples is restricted by the constraint, which is a finite theory (i.e. a formula). There is already some work [20, 19] on such applications that build on the polynomial construction of frontiers for connected acyclic CQs from [9], i.e. theorem 2.9.

6.7 Characterising Models by Formulas

Recall from section 2.2 that a concept class $\mathcal{C} \subseteq \mathcal{P}(D)$ over a domain D gave rise to an intersection space $(\mathcal{C}, \mathcal{O}_D)$. However, concept classes $\mathcal{C} \subseteq \mathcal{P}(D)$ that are closed under finite intersections invite another natural topological interpretation, namely where the concepts generate a topology as a subbasis, i.e. then $(D, \mathcal{C})$ is an intersection space. In such a setting, one may ask the converse question of finite characterisations of elements in the domain, i.e. whether an element $d \in D$ from the domain can be characterised by finitely many concepts up to some notion of equivalence? That is, are there finitely many concepts $C_0, \ldots, C_n, C'_0, \ldots, C'_m \in \mathcal{C}$ such that the point $d \in D$ is the only point from D (perhaps up to some notion of equivalence) contained in $\bigcap_{i \leq n} C_i$ and disjoint from $\bigcup_{j \leq m} C'_j$?

For instance, consider the concept class C_K^{fin}, i.e. the class of all $mod_{\text{fin}}(\varphi)$ for φ a modal formula and note that the full modal language is closed under conjunction and negation. One can wonder whether such model-class concepts can characterize elements of the domain M, s (i.e. pointed models) up to bisimulation. Observe that there are formulas $\varphi_0, \ldots, \varphi_n, \varphi'_0, \ldots, \varphi'_m$ characterising M, s up to bisimulation in the sense above (i.e. in the sense that M, s is the only pointed model contained in $\bigcap_{i \leq n} mod(\varphi_i)$ and disjoint from $\bigcup_{j \leq m} mod(\varphi'_j)$) iff there is a single formula $\psi := \varphi_0 \wedge \ldots \wedge \varphi_n \wedge \neg\varphi'_0 \wedge \ldots \wedge \varphi'_m$ such that ψ characterises M, s up to bisimulation. This is because

$$
\begin{aligned}
& mod(\varphi_0) \cap \ldots \cap mod(\varphi_n) \cap mod(\neg\varphi'_0) \cap \ldots \cap mod(\neg\varphi'_m) \\
=& mod(\varphi_0 \wedge \ldots \wedge \varphi_n \wedge \neg\varphi'_0 \wedge \ldots \wedge \neg\varphi'_m) = mod(\psi)
\end{aligned}
$$

Thus in the case of modal logic, a pointed M, s can be characterised up to bisimulation by a set of modal concepts iff it can be characterised by a single modal concept. In fact, the class of pointed models for which such a characterisation exists (in the standard full modal language) has already been identified.[6] We say that a Kripke model M is *well-founded* if there are no *strict* infinite descending paths (cf. definition 2.7) in M, i.e. no infinite descending paths $(t_0, t_1, t_2, \ldots)$ such that $t_i \neq t_j$ for all natural numbers i, j.[7]

Theorem 6.1. *(Baltags Theorem),([6]) A pointed model is characterised up to bisimulation by a modal formula iff it is well-founded.*

[6] Extending the language allows for characterising more models. For instance, it can be shown that propositional dynamic logic (PDL) can characterise *all* finite models up to bisimulation [7].

[7] This characterisation of well-foundedness assumes dependent choice.

Bibliography

[1] Bogdan Alexe, Phokion Kolaitis, and Wang-chiew Tan. "Characterizing schema mappings via data examples". In: Jan. 2010, pp. 261–272. DOI: 10.1145/1807085.1807120.

[2] Dana Angluin. "Queries and Concept Learning". In: *Mach. Learn.* 2.4 (1988), pp. 319–342. ISSN: 0885-6125. DOI: 10.1023/A:1022821128753.

[3] Franz Baader et al. *The Description Logic Handbook: Theory, Implementation, and Applications.* Jan. 2007.

[4] Alexandru Baltag et al. "A Dynamic Logic for Learning Theory". English. In: *Dynamic Logic. New Trends and Applications.* Ed. by Alexandre Madeira and Mario Benevides. Vol. 10669. Lecture Notes in Computer Science. DaLi - Dynamic Logic: new trends and applications 2017, DALI ; Conference date: 23-09-2017 Through 24-09-2017. Springer, 2018, pp. 35–54. DOI: 10.1007/978-3-319-73579-5_3.

[5] Michael Barr and Charles Wells. *Category theory for computing science.* Vol. 1. Prentice Hall New York, 1990.

[6] K. Jon Barwise and Lawrence S. Moss. "Vicious circles - on the mathematics of non-wellfounded phenomena". In: *CSLI lecture notes series.* 1996.

[7] Johan van Benthem. *Dynamic Odds & Ends.* Amsterdam, 1998.

[8] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic.* Vol. 53. Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press, 2001.

[9] Balder ten Cate and Victor Dalmau. "Conjunctive Queries: Unique Characterizations and Exact Learnability". In: *24th International Conference on Database Theory (ICDT 2021).* Ed. by Ke Yi and Zhewei Wei. Vol. 186. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 9:1–9:24. ISBN: 978-3-95977-179-5. DOI: 10.4230/LIPIcs.ICDT.2021.9. URL: https://drops.dagstuhl.de/opus/volltexte/2021/13717.

[10] S. Celani and R. Jansana. "A New Semantics for Positive Modal Logic". In: *Notre Dame Journal of Formal Logic* 38.1 (1997), pp. 1–18. DOI: 10.1305/ndjfl/1039700693. URL: https://doi.org/10.1305/ndjfl/1039700693.

[11] Sergio Celani, Facultad Exactas, and Ramon Jansana. "Priestley Duality, a Sahlqvist Theorem and a Goldblatt-Thomason Theorem for Positive Modal Logic". In: *Logic Journal of the IGPL* 7 (Dec. 1999).

[12] Ashok K. Chandra and Philip M. Merlin. "Optimal implementation of conjunctive queries in relational data bases". In: *STOC '77.* 1977.

[13] Víctor Dalmau. "Boolean Formulas are Hard to Learn for most Gate Bases". In: *ALT.* 1999.

[14] Anna Dmitrieva. *Positive modal logic beyond distributivity: duality, preservation and completeness*. URL: https://eprints.illc.uva.nl/id/eprint/1811/.

[15] J. Michael Dunn. "Positive Modal Logic". In: *Studia Logica: An International Journal for Symbolic Logic* 55.2 (1995), pp. 301–317. ISSN: 00393215, 15728730. URL: http://www.jstor.org/stable/20015820 (visited on 04/07/2022).

[16] Kit Fine. "Normal Forms in Modal Logic". In: *Notre Dame Journal of Formal Logic* 16.2 (1975), pp. 229–237. DOI: 10.1305/ndjfl/1093891703.

[17] Jan Foniok, Jaroslav Nesetril, and Claude Tardif. "Generalised dualities and maximal finite antichains in the homomorphism order of relational structures". In: *Eur. J. Comb.* 29 (2008), pp. 881–899.

[18] Jan Foniok, Jaroslav Nešetřil, and Claude Tardif. "Generalised Dualities and Maximal Finite Antichains in the Homomorphism Order of Relational Structures". In: *Eur. J. Comb.* 29.4 (2008), pp. 881–899. ISSN: 0195-6698. DOI: 10.1016/j.ejc.2007.11.017. URL: https://doi.org/10.1016/j.ejc.2007.11.017.

[19] Maurice Funk, Jean Christoph Jung, and Carsten Lutz. "Actively Learning Concepts and Conjunctive Queries under ELr-Ontologies". In: *IJCAI*. 2021.

[20] Maurice Funk, Jean Christoph Jung, and Carsten Lutz. "Actively Learning ELI Queries under DL-Lite Ontologies". In: *Description Logics*. 2021.

[21] Mai Gehrke, Hideo Nagahashi, and Yde Venema. *A Sahlqvist theorem for distributive modal logic*. Tech. rep. Annals of Pure and Applied Logic 131, Issues, 2002.

[22] Nina Gierasimczuk. *Knowing one's limits: logical analysis of inductive inference, PhD thesis*. 2010.

[23] Pavol Hell and Jaroslav Nesetril. *Graphs and homomorphisms*. 2004.

[24] Natasha Kurtonina and Maarten De Rijke. "Simulating without Negation". In: *Journal of Logic and Computation* 7.4 (Aug. 1997), pp. 501–522.

[25] Roger C. Lyndon. "Properties preserved under homomorphism." In: *Pacific Journal of Mathematics* 9.1 (1959), pp. 143–154. DOI: pjm/1103039459.

[26] Larisa Lvovna Maksimova. "Modal logics of finite layers". In: *Algebra i logika* 14.3 (1975), pp. 304–319.

[27] Heikki Mannila and Kari-Jouko Räihä. "Test Data for Relational Queries". In: *PODS*. 1986, pp. 217–223.

[28] Ralph McKenzie. "Equational bases and nonmodular lattice varieties". In: *Transactions of the American Mathematical Society* 174 (1972), pp. 1–43.

[29] Alessandra Palmigiano. "A coalgebraic view on positive modal logic". In: *Theoretical Computer Science* 327.1 (2004). Selected Papers of CMCS '03, pp. 175–195. ISSN: 0304-3975. DOI: https://doi.org/10.1016/j.tcs.2004.07.026.

[30] Krister Segerberg. *An Essay in Classical Modal Logic*. Uppsala, Sweden: Uppsala, Filosofiska Föreningen Och Filosofiska Institutionen Vid Uppsala Universitet, 1971.

[31] Slawek Staworko and Piotr Wieczorek. "Characterizing XML Twig Queries with Examples". In: *18th International Conference on Database Theory, ICDT 2015, March 23-27, 2015, Brussels, Belgium*. Ed. by Marcelo Arenas and Martın Ugarte. Vol. 31. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015, pp. 144–160. DOI: 10.4230/LIPIcs.ICDT.2015.144. URL: https://doi.org/10.4230/LIPIcs.ICDT.2015.144.