



Institute for Research in Cognitive Science

**Finite Model Theory and
Finite Variable Logics
Ph.D. Dissertation**

Eric Rosen

**University of Pennsylvania
3401 Walnut Street, Suite 400C
Philadelphia, PA 19104-6228**

October 1995

**Site of the NSF Science and Technology Center for
Research in Cognitive Science**

IRCS Report 95-28

FINITE MODEL THEORY AND FINITE VARIABLE LOGICS

Eric Rosen

A DISSERTATION

in

Philosophy

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy.

1995

Scott Weinstein

Supervisor of Dissertation

Samuel Freeman

Graduate Group Chairperson

© Copyright 1995

by

Eric Rosen

To my parents

Acknowledgments

I am extremely grateful to my advisor, Scott Weinstein, for all that he has taught me and for the help and encouragement that he has provided so graciously over the last four years. It has been an immense pleasure working with him; he made this project fun and exciting. Scott originally suggested that I pursue finite model theory, a subject which has developed into an active and fertile area of research, and he continues to guide me in the right direction. This dissertation owes a great deal to his ability to find and ask the most interesting questions. Many of my results are attempts to answer problems that he originally posed. I could not have asked for a better advisor.

I would like to thank Steven Lindell for the many ways in which he helped me with this work. Scott, Steven, and I have met regularly to discuss a variety of topics in finite model theory, during which time I learned much of what I know about descriptive complexity theory. Steven also listened patiently to many of my ideas, and provided useful feedback. Finally, as a member of my dissertation committee, he read my work carefully and suggested many improvements.

I am also grateful to Maria Bonet, Anuj Dawar, Yuri Gurevich, and James Rogers for valuable discussions on topics covered in my dissertation. Marco Hollenberg's detailed comments on the final chapter helped me improve the presentation and eliminate a number of errors.

I owe a special debt of gratitude to Prof. Johan van Benthem. In March of 1995 he visited Penn to give a series of lectures on his research on modal logic. At that time, I had a number of very interesting discussions with both him and Scott, during which they suggested looking at this work on modal logic from the perspective of finite model theory. Since then, Prof. van Benthem has made more many useful recommendations which have

been very helpful. Chapter 5 contains some results in this area.

Long before my adventures in logic began, I entered the Penn philosophy department, intending to pursue my interest in Wittgenstein and the philosophy of language. Although things have turned out differently, I would like to thank Professors Thomas Ricketts and Gary Ebbs for what they taught me about analytic philosophy, and for what they showed me about how to do philosophy. Thank you also to my friends here who made graduate school more interesting and enjoyable, Peter Dillard, Amy Herzel, Scott Kimbrough, Rukesh Korde, Shari Rudavsky, Peter Schwartz, Lisa Shabel, and Larry Shapiro. Alisa and Alan, my sister and brother, have also been good friends. From the very beginning, my parents instilled in me a love of learning. For all their love and support, I dedicate this dissertation to them.

ABSTRACT

FINITE MODEL THEORY AND FINITE VARIABLE LOGICS

Eric Rosen

Supervisor: Scott Weinstein

In this dissertation, I investigate some questions about the model theory of finite structures. One goal is to better understand the expressive power of various logical languages, including first-order logic (FO), over this class. A second, related, goal is to determine which results from classical model theory remain true when relativized to the class, $\mathcal{F}$, of finite structures. As it is well-known that many such results become false, I also consider certain weakened generalizations of classical results.

I prove some basic results about the languages $L^k(\exists)$ and $L_{\infty\omega}^k(\exists)$, the existential fragments of the finite variable logics L^k and $L_{\infty\omega}^k$. I show that there are finite models whose $L^k(\exists)$ -theories are not finitely axiomatizable. I also establish the optimality of a normal form for $L_{\infty\omega}^k(\exists)$, and separate certain fragments of this logic. I introduce a notion of a ‘generalized preservation theorem’, and establish certain partial positive results. I then show that existential preservation fails for the language $L_{\infty\omega}^\omega$, both over $\mathcal{F}$ and over the class of all structures. I also examine other preservation properties, e.g. for classes closed under homomorphisms.

In the final chapter, I investigate the finite model theory of propositional modal logic. I show that, in contrast to more expressive logics, modal logic is ‘well-behaved’ over $\mathcal{F}$. In particular, I establish that various theorems that are true over the class of all structures also hold over $\mathcal{F}$. I prove that, over $\mathcal{F}$, a class of models is FO-definable and closed under bisimulations iff it is defined by a modal FO sentence. In addition, I prove that, over $\mathcal{F}$, a class is defined by a modal sentence and closed under extensions iff it is defined by a $\Diamond$ -modal sentence.

Contents

Acknowledgments	iv
Abstract	vi
1 Introduction	1
1.1 Preservation theorems	3
1.2 Preliminaries	5
1.3 Logical equivalence and Ehrenfeucht-Fraisse games	7
2 Basic finite model theory for $L^k(\exists)$ and $L^k_{\infty\omega}(\exists)$	13
3 Existential preservation	26
3.1 Generalized preservation theorems	26
3.2 The failure of existential preservation for $L^\omega_{\infty\omega}$	30
4 Other (generalized) preservation theorems	36
4.1 The class HOM	37
4.2 Generalized preservation theorems for HOM and MON	49
5 Modal logic over finite structures	55
5.1 Background	57
5.2 Preservation theorems	62
6 Conclusion	70
Bibliography	72

Chapter 1

Introduction

Finite model theory investigates the model theory of finite structures. This subject interacts with a variety of fields from math, logic, and computer science, including classical model theory, graph theory, and complexity theory. The different areas and aspects of finite model theory are unified by an interest in the expressive power of logical languages. In this dissertation, I pursue model theoretic questions pertaining to definability, paying particular attention to preservation theorems. Some of these problems are just finite versions of results from classical model theory. That is, we can ask whether a classical theorem remains true when restricted to the class, $\mathcal{F}$, of finite models. Other questions are variations on standard ideas. Chapters 3 and 4, for example, examine preservation theorems involving languages other than first-order logic.

It is well known that many theorems from classical model theory become false over the class of finite models (see [14]). For example, the Los-Tarski theorem states that a first-order sentence defines a class of models that is closed under extensions if and only if it is equivalent to an existential sentence. Tait [23] showed that this proposition becomes false when relativized to the class $\mathcal{F}$. That is, there is a sentence, φ , such that $\text{Mod}_{\mathcal{F}}(\varphi)$, the class of finite models of φ , is closed under extensions but φ is not equivalent over $\mathcal{F}$ to any existential sentence. As a result of these kinds of ‘failures’, it would be interesting to find classical theorems that remain true over $\mathcal{F}$. But ‘negative’ results can also be viewed as raising new problems pertaining to what we call *generalized preservation theorems*. For example, Tait’s example suggests that we look for some alternative characterization of the

first-order definable classes of models that are closed under extensions. Chapter 3 contains some results in this direction. I also investigate preservation theorems for other logics prominent in finite model theory.

The remainder of this introduction provides some information about more general aspects of finite model theory that provide a setting for what follows. Below, I briefly discuss the importance of logical languages other than first-order logic. In Section 1, I then describe preservation theorems in more detail, and briefly summarize the topics covered in the remaining chapters. Section 2 contains notation, background information, preliminary definitions, and some basic results. Section 3 describes the connection between logical equivalence and Ehrenfeucht-Fraïssé games.

Over the class of finite models, central results of classical model theory either become obviously false, such as the Compactness theorem, or meaningless, like the Lowenheim-Skolem theorem. The failure of compactness, in particular, means that most standard proofs of classical results are invalid over $\mathcal{F}$. Furthermore, it has been shown that, when relativized to the class $\mathcal{F}$, many of these results actually become false, including the Los-Tarski theorem, the Beth definability theorem, Craig’s interpolation theorem (see [14]), and Lyndon’s lemma (see [2]). In addition, many natural and computationally simple properties, such as parity and graph connectedness, are not expressible in FO. As a consequence, first-order logic (FO) is not as natural and attractive, over $\mathcal{F}$, as it is in the general case.

A central motivation for the investigation of other logics has been the desire to find logical characterizations of computational complexity classes. An important early result from Fagin [10] says that a property (that is, a class of models) is in NP iff it is definable by an existential second-order sentence. Since then, Immerman and others have shown that, over the class of ordered finite structures, many other complexity classes are *captured*, in this sense, by different logics. This research has highlighted the interest of a variety of fixed point logics, which extend FO by adding some sort of recursion operator.

Barwise [6] showed that, over a fixed structure, every formula in least fixed point logic is equivalent to a formula in $L_{\infty\omega}^\omega$, infinitary finite variable logic, which is defined below. Kolaitis and Vardi [19] observed that this remains true over the class $\mathcal{F}$. Although finite variable logic looks rather strange because of the way in which variables are reused, it has

been useful for proving results about the expressive power of fixed point logics, since there is a nice algebraic characterization of logical definability for the language. From a very different point of view, others (see [4]) have argued for the relevance of finite variable logic to modal logic. Because of these connections, as well as my belief in the intrinsic interest of this logic, it has been afforded considerable attention in this dissertation. Chapter 2, in particular, is devoted to basic questions about the model theory of the existential fragments of L^k and $L_{\infty\omega}^k$.

Various kinds of questions arise about the expressive power of logical languages. As mentioned above, Fagin and Immerman have established close connections between the complexity of describing a property of finite structures in a logical language and the complexity of computing the property on a Turing machine, or some other abstract model of computation. A major open problem is to determine whether there is a logic that can express exactly those properties that are in P. Given two logics, L and L' , we can also ask about their relative expressive power, that is, is every sentence φ in L equivalent to some sentence θ in L' ? Finally, given a single property, such as graph planarity, and a logic, L , we can ask whether there is a sentence φ in L that expresses the property.

To show that a property can be defined in L , it suffices to exhibit a sentence that expresses it. On the other hand, negative results require a more general method. Over the class of all structures, one generally uses compactness; over $\mathcal{F}$, these kinds of results are most often established using Ehrenfeucht-Fraïssé type games. This technique, which also works in the classical setting, plays an important role in finite model theory, and has been applied to logics other than FO, including, especially, finite variable logics and fragments thereof. Some of these games are defined in Section 3.

1.1 Preservation theorems

Classical preservation theorems establish a connection between syntactic and semantic properties of first-order logic. In particular, they are propositions of the following form.

A class of models, $\mathcal{C}$, is FO-definable and closed under [‘preserved under’] some specified algebraic operation iff $\mathcal{C}$ is defined by a FO-sentence of some specified syntactic form.

Thus, the Los-Tarski theorem relates classes closed under extensions to existential sentences. The Homomorphism preservation theorem states that a class $\mathcal{C}$ is FO-definable and closed under homomorphisms iff it is defined by a positive existential sentence.

We remarked above that one aspect of finite model theory has been the attempt to determine which classical theorems remain valid over the class of finite structures. It was also noted that essentially every known answer is negative. A fundamental motivation for this dissertation has been to try to find positive model theoretic results that hold over $\mathcal{F}$. To this end, we introduce a generalization of the notion of a preservation theorem in order to formulate certain weaker versions of classical theorems that we would like to show remain true over $\mathcal{F}$. The starting point for our investigation is Tait’s result that the Los-Tarski theorem fails finitely. This led us to ask whether there is a natural logic, stronger than FO, such that every FO-definable class that is closed under extensions is defined by an ‘existential’ sentence of this logic.

This question also suggests that we investigate preservation theorems for these stronger logics. For example, if there is a logic L that contains FO and has an existential preservation theorem over $\mathcal{F}$, then the answer to the previous question must be yes. One of the main results of this dissertation is that existential preservation does not hold for $L_{\infty\omega}^\omega$, either over $\mathcal{F}$ or over all structures.

Chapter 2 contains some basic results about the model theory of the languages $L^k(\exists)$ and $L_{\infty\omega}^k(\exists)$, the existential fragments of the finite variable logics L^k and $L_{\infty\omega}^k$. We show that there are finite structures whose $L^k(\exists)$ -theories are not finitely axiomatizable. We also establish the optimality of a normal form for $L_{\infty\omega}^k(\exists)$, due to Kolaitis and Vardi, and separate certain fragments of this language.

Chapter 3 discusses preservation theorems for classes closed under extensions. Section 1 establishes some generalized preservation theorems for fragments of first-order logic. In Section 2, we prove that existential preservation fails for $L_{\infty\omega}^\omega(\exists)$. In Chapter 4, we examine generalized preservation theorems for other classes of models, including those that are ‘monotone’ and those that are closed under homomorphisms.

Chapter 5 initiates the investigation of the finite model theory of modal logic, which, it is well known, can be viewed as a fragment of FO. The results here indicate that, in contrast to stronger languages, modal logic is ‘well-behaved’ over $\mathcal{F}$. Thus, we prove that

some preservation theorems, due to van Benthem and his collaborators, remain true over $\mathcal{F}$. A somewhat open-ended question raised by this work is the extent to which these arguments can be generalized to apply to stronger fragments of FO, especially those considered in [5]. Recently, connections have emerged between modal logic and certain areas of theoretical computer science. We hope that some of our results, and the techniques developed here, will be of interest to researchers in these fields.

1.2 Preliminaries

Let $\mathcal{F}_\sigma$ be the collection of finite structures of signature σ . We will assume that the universe of any $A \in \mathcal{F}_\sigma$ is an initial segment of $N = \{0, 1, 2, \dots\}$. We will often use $A, B, \dots$ etc. to denote both a structure and its universe when no confusion is likely to result. We assume that the signature σ is finite and contains no function symbols; we suppress mention of σ when no confusion is likely to result. A *boolean query* $\mathcal{C} \subseteq \mathcal{F}$ is a class of finite structures that is closed under isomorphisms. We use $\mathcal{C}$ to range over boolean queries. In Chapters 2 and 3, we focus on boolean queries which are closed under extensions.

Definition 1 $\text{EXT} = \{\mathcal{C} \subseteq \mathcal{F} \mid \forall A, B \in \mathcal{C}, \text{ if } A \in \mathcal{C} \text{ and } A \subseteq B, \text{ then } B \in \mathcal{C}\}.$

Let L be a logical language and let φ be a sentence of L . $\text{Mod}(\varphi) = \{A \mid A \models \varphi\}$ is the L -class determined by φ and $\text{Mod}_f(\varphi) = \{A \in \mathcal{F} \mid A \models \varphi\}$ is the boolean query expressed by φ . We say that $\mathcal{C}$ is L -definable, just in case it is the boolean query expressed by some sentence $\varphi \in L$. We will often use L to denote the set of L -definable boolean queries. We let FO denote first-order logic, $L_{\infty\omega}$, the usual infinitary extension of first-order logic which allows conjunction and disjunction over arbitrary sets of formulas, L^k , the fragment of FO consisting of those formulas all of whose variables both free and bound are among $x_1, \dots, x_k$, and similarly $L_{\infty\omega}^k$, the k -variable fragment of $L_{\infty\omega}$; $L_{\infty\omega}^\omega = \bigcup_{k \in \omega} L_{\infty\omega}^k$. We let $\text{FO}(\exists)$ denote the set of existential formulas of FO, that is, those formulas obtained by closing the set of atomic formulas and negated atomic formulas under the operations of conjunction, disjunction, and existential quantification. We define $L_{\infty\omega}(\exists)$, the set of existential formulas of $L_{\infty\omega}$, similarly, but require, in addition, closure under infinitary conjunction and disjunction. We let $L^k(\exists)$ consist of the formulas common to $\text{FO}(\exists)$ and L^k and we define $L_{\infty\omega}^k(\exists)$ and $L_{\infty\omega}^\omega(\exists)$ similarly.

A Datalog($\neq, \neg$) program P is a collection of rules of the form

$$\eta_0 \leftarrow \eta_1, \dots, \eta_k.$$

Such a rule has a *head*, η_0 , and a *body*, $\eta_1, \dots, \eta_k$. Each of the η_i is either an inequality or a literal over the signature $\sigma \cup \tau$ where σ and τ are disjoint; σ consists of the *extensional* relations and constants of P and τ consists of the *intensional* relations of P . The heads of all rules are built from intensional relations and intensional relations occur only positively throughout P . The program contains a distinguished intensional relation R of arity $n \geq 0$ and determines an n -ary query over structures in $\mathcal{F}_\sigma$. The value of this query for a given $A \in \mathcal{F}_\sigma$ is the value of R when the program is viewed as determining least-fixed points for each of the intensional relations with respect to a simultaneous induction associated with the program. The reader may consult [1, 18] for further details and discussion. As with logics, we use Datalog($\neq, \neg$) to refer to the class of queries computed by Datalog($\neq, \neg$) programs as well as to the class of programs themselves. Datalog programs are defined similarly except that all the η_i are restricted to be positive literals, even those built from extensional relations. Observe that Datalog($\neq, \neg$) is contained in the least fixed-point extension of first-order logic (LFP).

In our current notation, the failure of the Los-Tarski Theorem over finite structures may be expressed as:

$$\text{FO} \cap \text{EXT} \not\subseteq \text{FO}(\exists).$$

This raises the question of whether $\text{FO} \cap \text{EXT}$ is contained in the existential fragment of some stronger logic. The following proposition completely characterizes the relative expressive power of the existential fragments of the logics in which we are interested.

Proposition 1

$$\text{FO}(\exists) \subset \text{Datalog}(\neq, \neg) \subset L_{\infty\omega}^\omega(\exists) \subset L_{\infty\omega}(\exists) = \text{EXT}.$$

Proof. It is easy to see that every query in $\text{FO}(\exists)$ can be expressed by a program in Datalog($\neq, \neg$) which makes use of no recursion. It is well-known that this inclusion is strict, for example, the query (s, t) -connectivity is expressible in Datalog but not in FO. The inclusion of Datalog($\neq, \neg$) in $L_{\infty\omega}^\omega(\exists)$ has been noted by Afrati, Cosmadakis, and Yannakakis [1] (see also [18]); the argument to show this is a variant of the proof that

least fixed-point logic is contained in $L_{\infty\omega}^\omega$ over the class of finite structures (see [19]). Afrati, Cosmadakis, and Yannakakis [1] also exhibit queries which witness the separation of Datalog($\neq, \neg$) and $L_{\infty\omega}^\omega(\exists)$, even over the class of polynomial time computable queries. The identity between $L_{\infty\omega}^\omega(\exists)$ and EXT has been noted by Kolaitis and independently by Lo (see [1] and [20]). Finally, it is easy to construct polynomial time computable boolean queries in EXT which are not in $L_{\infty\omega}^\omega$. For example, let $\mathcal{C}$ be the query over the signature $\{E, s, t\}$ of source-target graphs that says that there is an E -path from s to t whose length is less than half the cardinality of the structure. It is clear that $\mathcal{C} \in \text{EXT}$. It is also easy to verify that $\mathcal{C}$ is not in $L_{\infty\omega}^\omega$ (and therefore not in $L_{\infty\omega}^\omega(\exists)$) by a straightforward application of the k -pebble Ehrenfeucht-Fraïssé game which we review below. ■

The above proposition together with the failure of the Los-Tarski Theorem in the finite case suggests the following questions.

1. Is $\text{FO} \cap \text{EXT} \subseteq L_{\infty\omega}^\omega(\exists)$?
2. Is $\text{FO} \cap \text{EXT} \subseteq \text{Datalog}(\neq, \neg)$?
3. Is $L_{\infty\omega}^\omega \cap \text{EXT} \subseteq L_{\infty\omega}^\omega(\exists)$?

Clearly a positive answer to the second or third question would imply a positive answer to the first. In Chapter 3, we provide partial positive answers to the first and second questions, and a negative answer to the third question. Recently, Martin Grohe [13] has proved that the answer to question 1 is no.

1.3 Logical equivalence and Ehrenfeucht-Fraïssé games

Let L be one of the logical languages we have defined above. Given a structure A , the L -theory of A is the collection of sentences of L which are satisfied by A . We say that A is L -equivalent to B , if and only if, the L -theory of A is equal to the L -theory of B and we say that A is L -compatible with B , if and only if, the L -theory of A is contained in the L -theory of B . Note that if L is closed under negation, then the relations of L -equivalence and L -compatibility coincide, whereas for languages like $L^k(\exists)$ and $L_{\infty\omega}^k(\exists)$ these relations are distinct. We use the notations $\equiv^k$, $\equiv_{\infty\omega}^k$, $\preceq^k$, and $\preceq_{\infty\omega}^k$ for L^k -equivalence, $L_{\infty\omega}^k$ -equivalence, $L^k(\exists)$ -compatibility, and $L_{\infty\omega}^k(\exists)$ -compatibility, respectively. More generally,

if $\bar{a}$ and $\bar{b}$ are j -tuples of elements from A and B , then we write $(A, \bar{a}) \preceq^k (B, \bar{b})$ iff for all formulas $\varphi(\bar{x}) \in L^k(\exists)$, if $A \models \varphi[\bar{a}]$, then $B \models \varphi[\bar{b}]$.

The main tool for studying these relations are refinements of the Ehrenfeucht-Fraïssé game. Barwise [6] characterized $L^k_{\infty\omega}$ -equivalence in terms of partial isomorphisms, while Immerman [17] and Poizat [21] provided related pebble game characterizations of L^k -equivalence. Kolaitis and Vardi [18] characterized compatibility in the negation free fragment of $L^k_{\infty\omega}(\exists)$ both in terms of collections of partial *homomorphisms* as well as in terms of a one-sided, positive version of the pebble game. Below we use a minor variant of the approach in [18] to characterize $L^k_{\infty\omega}(\exists)$ -compatibility.

A set I of partial isomorphisms from A to B is said to have the k -[back-and-]forth property if for all $f \in I$ such that the domain of f has cardinality $< k$, and all $a \in A$ [$b \in B$], there is a function $g \in I$ such that $f \subseteq g$ and $a \in \text{dom}(g)$ [$b \in \text{rng}(g)$]. (That is, the k -forth property is the one-sided version, going forth from A , of the k -back-and-forth property.)

Barwise [6] proved the following proposition which gives an algebraic characterization of $L^k_{\infty\omega}$ -equivalence.

Proposition 2 (Barwise [6]) *Let A and B be structures of signature σ and let h be the map with $\text{dom}(h) = \{c^A \mid c \in \sigma\}$ such that $h(c^A) = c^B$ for all $c \in \sigma$. The following conditions are equivalent.*

1. $A \equiv^k_{\infty\omega} B$.
2. *There is a non-empty set I of partial isomorphisms from A to B such that*
 - (a) *I is closed under subfunctions;*
 - (b) *I has the k -back-and-forth property;*
 - (c) *for all $f \in I$, $f \cup h$ is a partial isomorphism from A to B .*

In a similar spirit, Kolaitis and Vardi [18] gave an algebraic characterization of the compatibility relation for the negation free fragment of $L^k_{\infty\omega}(\exists)$ in terms of collections of partial *homomorphisms* with the k -forth property. We adapt their approach to the case of $L^k_{\infty\omega}(\exists)$ in the following theorem.

Proposition 3 (Kolaitis and Vardi [18]) *Let A and B be structures of signature σ and let h be the map with $\text{dom}(h) = \{c^A \mid c \in \sigma\}$ such that $h(c^A) = c^B$ for all $c \in \sigma$. The following conditions are equivalent.*

1. $A \preceq_{\infty\omega}^k B$.
2. *There is a non-empty set I of partial isomorphisms from A to B such that*
 - (a) *I is closed under subfunctions;*
 - (b) *I has the k -forth property;*
 - (c) *for all $f \in I$, $f \cup h$ is a partial isomorphism from A to B .*

Both Propositions 2 and 3 can be expressed more colorfully in terms of pebble games. This approach to L^k -equivalence was introduced by Immerman [17] and Poizat [21] and as an approach to $L_{\infty\omega}^k(\exists)$ -compatibility by Kolaitis and Vardi [18]. In order to state the relevant results in a suitably refined form, we require the notion of the *quantifier rank* of a formula. We state this definition for formulas of $L_{\infty\omega}$ since all the languages we consider are fragments of it.

Definition 2 *The quantifier rank of $\varphi \in L_{\infty\omega}$, $qr(\varphi)$, is defined by the following induction.*

1. $qr(\varphi) = 0$ if φ is atomic;
2. $qr(\neg\varphi) = qr(\varphi)$;
3. $qr(\bigwedge \Phi) = qr(\bigvee \Phi) = \sup(\{qr(\varphi) \mid \varphi \in \Phi\})$;
4. $qr(\exists x\varphi) = qr(\forall x\varphi) = qr(\varphi) + 1$.

The n -round, k -pebble Ehrenfeucht-Fraïssé game on A and B is played between two players, Spoiler and Duplicator, with k pairs of pebbles, $(\alpha_1, \beta_1), \dots, (\alpha_k, \beta_k)$. The Spoiler begins each round by choosing a pair of pebbles (α_i, β_i) that may or may not be in play on the boards A and B . He (by convention, the Spoiler is male, the Duplicator female) either places α_i on an element of A , or β_i on an element of B . The Duplicator then plays the remaining pebble on the other model. The Spoiler wins the game if after any round $m \leq n$ the function f from A to B , which sends the element pebbled by α_i to the element

pebbled by β_i and preserves the denotations of constants, is not a partial isomorphism; otherwise, the Duplicator wins the game. The n -round $\exists^k$ -game is the one-sided version of the n -round, k -pebble Ehrenfeucht-Fraïssé game in which the Spoiler is restricted to play a pebble α_i into A at every round while the Duplicator responds by playing β_i into B ; the winning condition remains the same. Both the k -pebble Ehrenfeucht-Fraïssé game and its one-sided variant have infinite versions, which we call the *eternal* k -pebble Ehrenfeucht-Fraïssé game and the *eternal* $\exists^k$ -game. In these games, the play continues through a sequence of rounds of order type ω . The Spoiler wins the game, if and only if, he wins at the n^{th} -round for some $n \in \omega$ as above; otherwise, the Duplicator wins. In describing the play of pebble games below, we will often use S to refer to the Spoiler and D to refer to the Duplicator. We will also often use α_i, β_i , etc. to refer to both pebbles and the elements they pebble at a given round of play.

The foregoing n -round games may be used to characterize equivalence and compatibility of structures with respect to L^k sentences and $L^k(\exists)$ sentences of quantifier rank n , and the eternal games may be used to characterize equivalence and compatibility of structures with respect to $L_{\infty\omega}^k$ sentences and $L_{\infty\omega}^k(\exists)$ sentences. Given structures A and B we let $A \equiv^{k,n} B$, if and only if, A and B satisfy the same sentences of L^k of quantifier rank $\leq n$ and we let $A \preceq^{k,n} B$, if and only if, every sentence of $L^k(\exists)$ of quantifier rank $\leq n$, which is true in A , is also true in B . The following two propositions use the n -round pebble games to characterize these relations. The first is due to Immerman [17] and Poizat [21] and the second is essentially due to Kolaitis and Vardi [18].

Proposition 4 (Immerman [17], Poizat [21]) *For all structures A and B , the following conditions are equivalent.*

1. $A \equiv^{k,n} B$.
2. *The Duplicator has a winning strategy for the n -round, k -pebble Ehrenfeucht-Fraïssé game on A and B .*

Proposition 5 (Kolaitis and Vardi [18]) *For all structures A and B , the following conditions are equivalent.*

1. $A \preceq^{k,n} B$.

2. *The Duplicator has a winning strategy for the n -round $\exists^k$ -game on A and B , with the Duplicator playing on B .*

The next proposition gives a characterization of the infinitary equivalence and compatibility relations in terms of the eternal games. It is essentially due to Kolaitis and Vardi [19, 18].

Proposition 6 (Kolaitis and Vardi [19, 18]) 1. *For all structures A and B , the following conditions are equivalent.*

- (a) $A \equiv_{\infty\omega}^k B$.
 - (b) *The Duplicator has a winning strategy for the eternal k -pebble Ehrenfeucht-Fraïssé game on A and B .*
2. *For all structures A and B , the following conditions are equivalent.*

- (a) $A \preceq_{\infty\omega}^k B$.
- (b) *The Duplicator has a winning strategy for the eternal $\exists^k$ -game on A and B , with the Duplicator playing on B .*

Kolaitis and Vardi [19, 18] observed that over finite structures infinitary equivalence and compatibility coincide with their finitary analogs.

Proposition 7 (Kolaitis and Vardi [19, 18]) 1. *Let A or B be a finite structure. Then, the following conditions are equivalent.*

- (a) $A \equiv_{\infty\omega}^k B$.
- (b) $A \equiv^k B$.

2. *Let B be a finite structure. Then, the following conditions are equivalent.*

- (a) $A \preceq_{\infty\omega}^k B$.
- (b) $A \preceq^k B$.

The foregoing propositions yield the following corollaries concerning definability.

Proposition 8 (Kolaitis and Vardi [18]) *For all $\mathcal{C} \subseteq \mathcal{F}$, the following conditions are equivalent.*

1. $\mathcal{C}$ is $L^k_{\infty\omega}(\exists)$ -definable.
2. For all $A \in \mathcal{C}$ and $B \notin \mathcal{C}$, $A \not\leq^k_{\infty\omega} B$.
3. For all $A \in \mathcal{C}$ and $B \notin \mathcal{C}$, $A \not\leq^k B$.
4. For all $A \in \mathcal{C}$ and $B \notin \mathcal{C}$, there is an $n \in \omega$ such that the Spoiler has a winning strategy for the n -round $\exists^k$ -game on A and B with the Spoiler playing on A .

Chapter 2

Basic finite model theory for $L^k(\exists)$ and $L_{\infty\omega}^k(\exists)$

In this chapter, we present some basic model theory for L^k , $L_{\infty\omega}^k$, $L^k(\exists)$, and $L_{\infty\omega}^k(\exists)$, answering questions concerning finite axiomatizability and normal forms. Let L and L' be logical languages and let T be a collection of sentences of L . We say that T is *finitely axiomatizable in L'* , if and only if, there is a sentence $\varphi \in L'$ such that $\text{Mod}_f(T) = \text{Mod}_f(\varphi)$. Dawar, Lindell and Weinstein [9] prove that the $L_{\infty\omega}^k$ -theory of any finite model is finitely axiomatizable in L^k . As a corollary, they obtain a simple normal form for $L_{\infty\omega}^k$ over $\mathcal{F}$, in particular, they show that every sentence of $L_{\infty\omega}^k$ is equivalent to a countable disjunction of sentences of L^k and is also equivalent to a countable conjunction of sentences of L^k . In contrast, we show below that there are finite models whose $L^k(\exists)$ -theories are not finitely axiomatizable in $L^k(\exists)$. Building on this result, we prove that the normal form for $L_{\infty\omega}^k$ over $\mathcal{F}$ (every sentence of $L_{\infty\omega}^k$ is equivalent over $\mathcal{F}$ to a countable disjunction of countable conjunctions of sentences of L^k) exhibited by Kolaitis and Vardi [19] is optimal when considered as a normal form for $L_{\infty\omega}^k(\exists)$ sentences over $L^k(\exists)$.

We begin by proving that there are models whose $L^k(\exists)$ -theories are not finitely axiomatizable in $L^k(\exists)$. Our argument exploits the *k-extension axioms*, which we now describe briefly. Let σ be a purely relational, finite signature. A *basic k-type* π over the signature σ is a maximal consistent set of literals over σ in the variables $x_1, \dots, x_k$. A *k-extension axiom* of signature σ is a sentence of the form $\forall x_1 \dots x_{k-1} \exists x_k (\bigwedge \pi \rightarrow \bigwedge \pi')$, where π is a

basic $(k - 1)$ -type of signature σ , π' is a basic k -type of signature σ , and $\pi \subseteq \pi'$. Over a fixed signature σ , the *k-Gaifman theory*, Γ_k , is the set of all k -extension axioms of signature σ . It is easy to see that, for each k , there are only finitely many k -extension axioms. Gaifman [12] showed that the theory $T = \bigcup_k \Gamma_k$ axiomatizes an ω -categorical model called the *random structure*. Fagin [11] proved the 0-1 law for first-order logic by showing that every extension axiom is *almost surely true* over $\mathcal{F}$. Fagin's result implies that almost every $A \in \mathcal{F}$ satisfies the k -Gaifman theory. Immerman [17] showed that any two models of the k -Gaifman theory are L^k -equivalent and Kolaitis and Vardi [19] made use of the k -Gaifman theory in their proof of the 0 – 1 law for $L_{\infty\omega}^\omega$. We make the following easy observation.

Proposition 9 *Let $A \models \Gamma_k$, and let B be any (finite or infinite) model. Then $B \preceq_{\infty\omega}^k A$. Equivalently, for all $\varphi \in L_{\infty\omega}^k(\exists)$, if φ is satisfiable, then $A \models \varphi$.*

Proof. The proof follows easily from Proposition 6 by considering the eternal $\exists^k$ -game on B and A with the Duplicator playing on A . The k -Gaifman axioms essentially say that D can extend a partial isomorphism with domain of size $< k$ in every possible way. Therefore, she has a winning strategy for the game. ■

We observe that this result yields a compactness theorem over finite structures and a finitary analog of the Löwenheim-Skolem Theorem for $L_{\infty\omega}^k(\exists)$.

Corollary 1 *For every $k \in \omega$, there is an $n_k \in \omega$ such that for every set Φ of sentences of $L_{\infty\omega}^k(\exists)$, Φ is satisfiable, if and only if, every finite subset of Φ is satisfiable, if and only if, Φ is satisfied in a model of size n_k .*

The next proposition establishes that there are finite structures whose $L^k(\exists)$ -theory is not finitely axiomatizable in $L^k(\exists)$.

Proposition 10 *For all $k \geq 2$, there is a model $A_k \in \mathcal{F}$ such that the $L^k(\exists)$ -theory of A_k is not finitely axiomatizable in $L^k(\exists)$.*

Proof. Let A_k be any finite model of the k -Gaifman theory over the language of graphs. We show that for any $n \in \omega$, there is a B_k^n such that $A_k \preceq^{k,n} B_k^n$ and $A_k \not\preceq^{k,n+1} B_k^n$. This

implies that the theory of A_k cannot be axiomatized by $L^k(\exists)$ sentences of quantifier rank $\leq n$ and, therefore, that it is not finitely axiomatizable in $L^k(\exists)$.

For the purpose of defining the models B_k^n , we require the following notion and notation. A basic k -type π satisfies the *distinctness condition* if for every $l < k$, the formula $x_l \neq x_k \in \pi$. Let $\{\pi_1, \dots, \pi_s\}$ be a set of basic $(k-1)$ -types such that

1. every basic $(k-1)$ -type is equivalent to some π_i and
2. if $i \neq j$, then π_i is not equivalent to π_j .

Similarly, for each $1 \leq i \leq s$, let $\{\pi_{i,1}, \dots, \pi_{i,n(i)}\}$ be a set of basic k -types each of which extends π_i and satisfies the distinctness condition such that

1. every basic k -type which extends π_i and satisfies the distinctness condition is equivalent to some $\pi_{i,j}$ and
2. if $j \neq j'$, then $\pi_{i,j}$ is not equivalent to $\pi_{i,j'}$.

We proceed to define the models B_k^n . Let B_k^1 be the graph on two vertices with exactly one loop and no other edges. Thus B_k^1 realizes both basic 1-types. Given that B_k^n has been defined, we now define B_k^{n+1} as an extension of B_k^n . For each $(k-1)$ -tuple $\bar{b}$ of elements of B_k^n , let $\tau(\bar{b})$ be the unique i such that $B_k^n \models \pi_i[\bar{b}]$, and let $X_{\bar{b}} = \{b_{\bar{b},j}^{n+1} \mid 1 \leq j \leq n(\tau(\bar{b}))\}$ be a set of distinct objects disjoint from B_k^n . We suppose that for any distinct pair of $(k-1)$ -tuples $\bar{a}$ and $\bar{b}$ of elements of B_k^n , $X_{\bar{a}} \cap X_{\bar{b}} = \emptyset$. Let X be the union of all the sets $X_{\bar{b}}$. We let the universe of $B_k^{n+1} = B_k^n \cup X$. The edge relation of B_k^{n+1} is obtained from that of B_k^n by adding the minimal number of edges so that each k -tuple $\bar{b} * b_{\bar{b},j}^{n+1}$ satisfies $\pi_{\tau(\bar{b}),j}$. It is easy to see that each B_k^{n+1} is well-defined. We say that the *height* of an element b introduced in this construction is the least n such that $b \in B_k^n$.

We first show that $A_k \preceq^{k,n} B_k^n$. By Proposition 5, it suffices to describe a winning strategy for D in the n -round $\exists^k$ -game with D playing on B_k^n and S playing on A_k . The strategy we describe for D will allow her to play her m^{th} move on some $b \in B_k^m$, for each $m \leq n$. In round 1, D answers the first move of S by playing her pebble on the appropriate element of $B_k^1 \subseteq B_k^n$ to create a partial isomorphism. Suppose that D has played only onto elements of B_k^m through round m , where $m < n$. Let S choose pebble pair (α_l, β_l) to play in round

$(m + 1)$. We consider two cases. If S plays α_l on the same element as some $\alpha_{l'}$, for $l \neq l'$, then D must play β_l onto the element pebbled by $\beta_{l'}$. Doing so, she obviously maintains a partial isomorphism and succeeds in playing within B_k^{m+1} . On the other hand, suppose that S plays α_l on a distinct element such that the elements pebbled by $\bar{\alpha} * \alpha_l$ on A after the round satisfy $\pi_{i,j}$ (we may need to pad the tuple pebbled by $\bar{\alpha}$ to a tuple of length $(k - 1)$ by repeating its last element, if all the pebbles are not in play at this round). Before D plays her $(m + 1)^{st}$ move, the pebbles $\bar{\beta}$ are on a tuple $\bar{b}$ (similarly padded, if necessary) that satisfies π_i . She then plays β_l on the element $b_{b,j}^{m+1} \in B_k^{m+1}$, thereby maintaining a partial isomorphism. This strategy enables her to win the n -round game.

Next, we show that $A_k \not\preceq^{k,n+1} B_k^n$. By Proposition 5, it suffices to show that S can win the $(n + 1)$ -round game with D playing on B_k^n and S playing on A_k . We describe a strategy for play by S which forces D to pebble an element of height at least m by the end of round m to avoid losing at that round. It follows that S wins the $(n + 1)$ -round game since all elements of B_k^n have height $\leq n$. S plays as follows. He first places his k -pebbles on a set of k distinct elements which form a k -clique, that is, for every pair of distinct pebbled elements a and a' , $A_k \models E(a, a')$. S may play in this way since $A_k \models \Gamma_k$. By our construction above, if $b, b' \in B_k^n$ are distinct elements of the same height, $B_k^n \not\models E(b, b')$. It follows immediately that any r -clique in B_k^n contains an element of height at least r . Therefore, if S has not won by round k , D has pebbled an element of height at least k by the end of that round. Note that in case $(n + 1) \leq k$, we are done, since at round $(n + 1)$, D will be unable to play onto an element of height at least $(n + 1)$ to form an $(n + 1)$ -clique.

We proceed to describe the strategy for S's continuing play under the assumption that $k < (n + 1)$. Suppose that through round $m, k \leq m < (n + 1)$, D has played a pebble onto an element of height at least m , and that the k pebbles S has played lie on distinct elements of A_k which form a k -clique. We show how S can play to ensure that D must play onto an element of height at least $(m + 1)$ at round $(m + 1)$, if she is to prevent S from winning at this round, and leave the round with a k -clique pebbled. Suppose that β_i is pebbling an element b of height greater than the height of any other element pebbled in B_k^n at round m . By our hypothesis, the height of b is at least m . Pick $j \neq i$ (recall that $2 \leq k$) and let $a \in A_k$ be the element pebbled by α_j . S picks up α_j and places it on an $a' \in A_k$ such that

1. $A_k \models E(a, a) \leftrightarrow \neg E(a', a')$ and
2. for every $a'' \in A_k$ on which one of the remaining $(k - 1)$ pebbles lies, $a' \neq a''$ and $A_k \models E(a', a'') \wedge E(a'', a')$.

The existence of such an a' follows from the fact that $A_k \models \Gamma_k$. We claim that to avoid losing at this round, D must play her pebble β_j onto an element b' of height greater than the height of b , and hence of height at least $(m + 1)$. Let b'' be the element pebbled by β_j at round m . By our construction, each element of B_k^n is connected to at most $(k - 1)$ elements of lesser height. Therefore, from the hypotheses that S had pebbled a k -clique at round m , and that b is an element of maximal height pebbled by D at that round, we may conclude that the only element of height $\leq$ the height of b adjacent to b onto which D could play β_j is b'' itself. But this play would fail to maintain a partial isomorphism with the elements S has now pebbled at round $(m + 1)$ by the first condition we have imposed on the choice of a' above. Therefore, to avoid losing at round $(m + 1)$, D must pebble an element of height at least $(m + 1)$. ■

The next result follows immediately.

Corollary 2 *There are infinitely many formulas of $L^k(\exists)$ which are pairwise inequivalent over $\mathcal{F}$.*

We now consider $L_{\infty\omega}^k(\exists)$ -theories and normal forms for $L_{\infty\omega}^k(\exists)$ sentences over $\mathcal{F}$. We let $\text{Th}_{\exists}^k(A)$ denote the $L_{\infty\omega}^k(\exists)$ -theory of A . Before proceeding, we define the following fragments of $L_{\infty\omega}^k(\exists)$.

1. Let $\bigwedge L^k(\exists) = \{\theta \mid \theta = \bigwedge \Phi, \text{ for some } \Phi \subseteq L^k(\exists)\}$.
2. Let $\bigvee L^k(\exists) = \{\theta \mid \theta = \bigvee \Phi, \text{ for some } \Phi \subseteq L^k(\exists)\}$.
3. Let $\bigwedge(\bigvee L^k(\exists)) = \{\theta \mid \theta = \bigwedge \Phi, \text{ for some countable } \Phi \subseteq \bigvee L^k(\exists)\}$.
4. Let $\bigvee(\bigwedge L^k(\exists)) = \{\theta \mid \theta = \bigvee \Phi, \text{ for some countable } \Phi \subseteq \bigwedge L^k(\exists)\}$.

Proposition 11 *For all finite structures A , there is a $\theta \in \bigwedge L^k(\exists)$ such that $\text{Mod}_f(\theta) = \text{Mod}_f(\text{Th}_{\exists}^k(A))$.*

Proof. Observe that $\text{Mod}_f(\text{Th}_{\exists}^k(A)) = \{B \in \mathcal{F} \mid A \preceq_{\infty\omega}^k B\}$. Let $\mathcal{C}_A = \mathcal{F} - \text{Mod}_f(\text{Th}_{\exists}^k(A))$. By Proposition 7, for each $B \in \mathcal{C}_A$, there is a sentence $\varphi_B \in L^k(\exists)$ such that $A \models \varphi_B$ and $B \not\models \varphi_B$. Let $\theta = \bigwedge_{B \in \mathcal{C}_A} \varphi_B$. It is easy to verify that $\text{Mod}_f(\theta) = \text{Mod}_f(\text{Th}_{\exists}^k(A))$. ■

Kolaitis and Vardi [18] obtained a normal form for the negation free fragment of $L_{\infty\omega}^k(\exists)$ over $\mathcal{F}$. It is easy to extend their result to $L_{\infty\omega}^k(\exists)$ and to provide a dual normal form as well. We codify these normal forms in the next proposition.

Proposition 12 (Kolaitis and Vardi [18]) *For every sentence $\varphi \in L_{\infty\omega}^k(\exists)$, there is a $\theta \in \bigvee(\bigwedge L^k(\exists))$ and a $\zeta \in \bigwedge(\bigvee L^k(\exists))$ such that $\text{Mod}_f(\varphi) = \text{Mod}_f(\theta) = \text{Mod}_f(\zeta)$.*

Proof. Let $\mathcal{C} = \text{Mod}_f(\varphi)$. By Proposition 8, for each $A \in \mathcal{C}, B \in \mathcal{F} - \mathcal{C}$, there is a sentence $\theta_{A,B} \in L^k(\exists)$ such that $A \models \theta_{A,B}$ and $B \not\models \theta_{A,B}$. Let $\theta = \bigvee_{A \in \mathcal{C}} (\bigwedge_{B \notin \mathcal{C}} \theta_{A,B})$ and let $\zeta = \bigwedge_{B \notin \mathcal{C}} (\bigvee_{A \in \mathcal{C}} \theta_{A,B})$. It is easy to verify that the proposition holds for this choice of θ and ζ . ■

Next we show that the fragments $\bigwedge L^k(\exists)$ and $\bigvee L^k(\exists)$ are closed under finite conjunction, finite disjunction, and existential quantification over $\mathcal{F}$. This means that if an $L_{\infty\omega}^k(\exists)$ -definable query cannot be expressed in either $\bigwedge L^k(\exists)$ or $\bigvee L^k(\exists)$, then it is only definable using both an infinitary conjunction and an infinitary disjunction.

Proposition 13 *The languages $\bigwedge L^k(\exists)$ and $\bigvee L^k(\exists)$ are both closed under finite conjunction, finite disjunction, and existential quantification over $\mathcal{F}$.*

Proof. Let $\Phi = \{\varphi_i(x, \overline{y}) \mid i \in \omega\}$ be a set of formulas of $L^k(\exists)$. We show that if $\theta(\overline{y}) = \exists x \bigwedge \Phi$, then $\theta(\overline{y})$ is equivalent over $\mathcal{F}$ to some formula $\theta'(\overline{y}) \in \bigwedge L^k(\exists)$. (The other closure conditions may be easily verified.) Let $\psi_m = \bigwedge_{0 \leq l \leq m} \varphi_l(x, \overline{y})$ and let $\theta'(\overline{y}) = \bigwedge_{m \in \omega} \exists x \psi_m$. We show θ' is equivalent to θ . It is obvious that θ implies θ' . Let $A \in \mathcal{F}$ and $\overline{a} \in A$ be such that $A \models \theta'[\overline{a}]$. Because A is finite, there is some $a' \in A$ such that for arbitrarily large m , $A \models \psi_m[a', \overline{a}]$. Therefore $A \models \bigwedge_{m \in \omega} \psi_m[a', \overline{a}]$, and θ' implies θ . ■

Below we show that the query classes $\bigwedge L^k(\exists)$ and $\bigvee L^k(\exists)$ are proper subsets of $\bigwedge(\bigvee L^k(\exists))$ and that neither of $\bigwedge L^k(\exists)$ and $\bigvee L^k(\exists)$ is a subset of the other. We first give necessary and sufficient conditions for classes to be definable in $\bigwedge L^k(\exists)$ and $\bigvee L^k(\exists)$, and prove a lemma from Kolaitis and Vardi [18] that we need below.

Proposition 14 1. A class $\mathcal{C}$ is definable in $\bigwedge L^k(\exists)$ iff for all $B \notin \mathcal{C}$, there is a $\varphi_B \in L^k(\exists)$ such that $B \not\models \varphi_B$ and for all $A \in \mathcal{C}$, $A \models \varphi_B$.

2. A class $\mathcal{C}$ is definable in $\bigvee L^k(\exists)$ iff for all $A \in \mathcal{C}$, there is a $\varphi_A \in L^k(\exists)$ such that $A \models \varphi_A$ and for all $B \notin \mathcal{C}$, $B \not\models \varphi_A$.

Proof. To prove 1., suppose that $\mathcal{C}$ is defined by the sentence $\bigwedge_{n \in \omega} \psi_n$, and that $B \notin \mathcal{C}$. Then there is some ψ_m such that $B \not\models \psi_m$. Let φ_B be this ψ_m . In the other direction, observe that the sentence $\varphi = \bigwedge_{B \notin \mathcal{C}} \varphi_B$ defines $\mathcal{C}$. The proof of 2. is similar. ■

Lemma 1 (Kolaitis and Vardi [18]) The relation $\preceq^k$ is polynomial time computable.

Proof. Let A and B be models of signature σ . We define $\langle A, B \rangle$ to be the following model, with signature $\sigma \cup \{Qx\}$, where Qx is a unary predicate not in σ . It is the disjoint union of A and B , with the extension of the predicate Qx interpreted as the universe of B . It is easy to see that, given a standard encoding of the models A and B on Turing machines (e.g. see [9]), an encoding of $\langle A, B \rangle$ can be produced in polynomial time.

Modifying an idea from Dawar, Lindell, and Weinstein [9], we now show that there is an LFP sentence θ such that for all A and B , $\langle A, B \rangle \models \theta$ iff $A \preceq^k B$. It is well known (see [14]) that every LFP query can be computed in polynomial time. Composing the function that outputs a representation of $\langle A, B \rangle$ with the function that computes the truth value of θ then yields the desired algorithm.

Let $R(x_1, \dots, x_k, y_1, \dots, y_k)$ be the $2k$ -ary relation on models $\langle A, B \rangle$ such that $\langle A, B \rangle \models R(a_1, \dots, a_k, b_1, \dots, b_k)$ iff each a_i is in A , each b_i is in B , and $(A, \bar{a}) \not\preceq^k (B, \bar{b})$. We first show that $R(\bar{x}, \bar{y})$ can be expressed in LFP. Let $\Psi = \{\psi_1, \dots, \psi_t\}$ be the set of all atomic formulas over σ with free variables among $x_1, \dots, x_k$. Given any k -tuple $\bar{a} \subseteq A$ and k -tuple $\bar{b} \subseteq B$, $(A, \bar{a}) \not\preceq^{k,0} (B, \bar{b})$ iff there is a $\psi \in \Psi$ such that $A \models \psi[\bar{a}]$ iff $B \not\models \psi[\bar{b}]$. In general, $(A, \bar{a}) \not\preceq^{k,n+1} (B, \bar{b})$ iff there is an $a' \in A$ and an $i \leq k$ such that for all $b' \in B$, $(A, \bar{a}') \not\preceq^{k,n} (B, \bar{b}')$, where $\bar{a}'$ and $\bar{b}'$ are the k -tuples obtained from $\bar{a}$ and $\bar{b}$ by replacing the i^{th} component by a' and b' , respectively. Then the least fixed point of the following formula defines the desired relation, $R(\bar{x}, \bar{y})$.

$$\gamma(\bar{x}, \bar{y}) = \mathbf{lfp} \left[\bigwedge_{i \leq k} (\neg Qx_i \wedge Qy_i) \wedge \left(\bigvee_{\psi \in \Psi} (\psi[\bar{x}] \leftrightarrow \neg \psi[\bar{y}]) \vee \bigvee_{i \leq k} \exists x_i \forall y_i (\neg Qx_i \wedge (Qy_i \rightarrow R(\bar{x}, \bar{y}))) \right) \right]$$

Observe that $A \preceq^k B$ iff for each k -tuple $\bar{a} \subseteq A$, there is a k -tuple $\bar{b} \subseteq B$ such that $(A, \bar{a}) \preceq^k (B, \bar{b})$. If $A \not\preceq^k B$, then there is a sentence $\varphi \in L^k(\exists)$ such that $A \models \varphi$ and $B \not\models \varphi$. Then for any $\bar{a} \subseteq A$ and $\bar{b} \subseteq B$, $(A, \bar{a}) \models \varphi[\bar{a}]$ and $(B, \bar{b}) \not\models \varphi[\bar{b}]$. In the other direction, assume that $A \preceq^k B$ and $\bar{a} \subseteq A$. Suppose that for each $\bar{b} \subseteq B$, there is a formula $\zeta_{\bar{b}}(\bar{x})$ such that $(A, \bar{a}) \models \zeta_{\bar{b}}(\bar{x})$ and $(B, \bar{b}) \not\models \zeta_{\bar{b}}(\bar{x})$. Then $\zeta = \exists \bar{x} \bigwedge_{\bar{b} \subseteq B} \zeta_{\bar{b}}(\bar{x})$ is an $L^k(\exists)$ sentence such that $A \models \zeta$ and $B \not\models \zeta$, a contradiction. Finally, let θ be the following sentence.

$$\theta = \forall x_1 \dots x_k \left(\bigwedge_{i \leq k} \neg Qx_i \rightarrow \exists y_1 \dots y_k \left(\bigwedge_{i \leq k} Qy_i \wedge \neg \gamma(\bar{x}, \bar{y}) \right) \right)$$

This completes the proof. ■

Proposition 15 *For each $k \geq 2$, there is a polynomial time computable boolean query $\mathcal{C} \in \bigwedge L^k(\exists) - \bigvee L^k(\exists)$.*

Proof. Let $k \geq 2$ be given and let the graph A_k be a model of the k -Gaifman theory. Let T be the $L^k(\exists)$ -theory of A_k and let $\theta = \bigwedge T$. Clearly, $\theta \in \bigwedge L^k(\exists)$. Let $\mathcal{C} = \text{Mod}_f(\theta)$. It is easy to see that $\mathcal{C} = \{B \in \mathcal{F} \mid A_k \preceq^k B\}$; thus, by Lemma 1, $\mathcal{C}$ is polynomial time computable. In the proof of Proposition 10, we showed that for every satisfiable $\varphi \in L^k(\exists)$, $\text{Mod}_f(\varphi) \not\subseteq \mathcal{C}$. This implies that for every $\psi \in \bigvee L^k(\exists)$, $\mathcal{C} \neq \text{Mod}_f(\psi)$. ■

Proposition 16 *There is a polynomial time computable boolean query $\mathcal{C} \in \bigvee L^2(\exists)$ such that for all $k \in \omega$, $\mathcal{C} \not\subseteq \bigwedge L^k(\exists)$. In consequence, for each $k \geq 2$, there is a class $\mathcal{C} \in \bigvee L^k(\exists) - \bigwedge L^k(\exists)$.*

Proof. Over the signature $\sigma = \{E, s, t\}$, let $\mathcal{C} = \{A \mid \text{there is a path from } s \text{ to } t\}$, the class of (s, t) -connected graphs. This class is clearly in $\bigvee L^2(\exists)$. As noted earlier, it is in Datalog, and, hence, polynomial time computable. From Proposition 14, to show that $\mathcal{C} \not\subseteq \bigwedge L^k(\exists)$, it suffices to show that there is a $B \notin \mathcal{C}$ such that for all $n \in \omega$, there is an $A_n \in \mathcal{C}$ such that $A_n \preceq^{k,n} B$. This latter condition is equivalent to D's possessing a winning strategy for the n -round $\exists^k$ -game on A_n and B . We construct B to give her the greatest possible freedom in choosing her moves. Let M be any graph such that $M \models \Gamma_{k+1}$, and let M_s (resp. M_t) be obtained from M by requiring that s (resp. t) denote a loop-free element. We define B to be the disjoint union of M_s and M_t , thus insuring that $B \notin \mathcal{C}$.

For each n , let A_n be the simple chain from s to t of length 2^{n+2} . The basic idea is that by choosing the chain to be long enough, S will not be able to witness the existence of a path from s to t in only n moves. Let $d(x, y)$ be the natural distance function on A_n .

We now describe D's strategy. In each round m , D chooses to play on an element of M_s iff S just played a pebble on $a \in A_n$ such that either (i) $d(s, a) \leq 2^{(n+2)-m}$; or (ii) there is a j such that β_j is on an element of M_s and $d(\alpha_j, a) \leq 2^{(n+2)-m}$. She then plays her pebble on an element of the appropriate component of B so that she maintains a partial isomorphism among the pebbles on that component. It is easy to see that this is possible because M_s and M_t are models of Γ_{k+1} .

In order to establish that this is a winning strategy, it suffices to verify the following two claims.

1. In each round $l \leq n$, if D plays a pebble β_i on M_s , then α_i is not adjacent to t on A_n . Similarly for M_t and s .
2. After each round l , for all pairs of pebbles $\{\alpha_i, \alpha_j\}$, if $A_n \models E(\alpha_i, \alpha_j)$, then β_i and β_j are on the same component of B .

We argue, by induction, that if D plays β_i on M_s in round m , then $d(s, \alpha_i) \leq (2^{(n+2)-1} + 2^{(n+2)-2} + \dots + 2^{(n+2)-m}) < 2^{n+2} - 1$. Since $d(s, t) = 2^{n+2}$, this establishes that $A_n \not\models E(\alpha_i, t)$. In round 1, D plays β_i on M_s iff $d(s, \alpha_i) \leq 2^{(n+2)-1}$. Suppose that in round $m+1$ D plays β_i on M_s . Then either $d(s, \alpha_i) \leq 2^{(n+2)-m}$ or there is an α_j such that β_j is on M_s , $d(\alpha_i, \alpha_j) \leq 2^{(n+2)-(m+1)}$, and, by induction hypothesis, $d(s, \alpha_j) \leq (2^{(n+2)-1} + 2^{(n+2)-2} + \dots + 2^{(n+2)-m})$. In both cases, the induction condition is maintained. The second part of Claim 1 follows from the fact that in round m , if D plays β_i on M_t , then S must have played α_i such that $d(s, \alpha_i) > 2^{(n+2)-m} > 1$.

To prove Claim 2, observe that at each round m , if $\beta_i \in M_s$, and $\beta_j \in M_t$, then $d(\alpha_i, \alpha_j) \geq 2^{(n+2)-m} > 1$. The details are similar to the previous argument. $\blacksquare$

The next result shows that the normal form for $L_{\infty\omega}^k(\exists)$ over $\mathcal{F}$ given in Proposition 12 is optimal.

Proposition 17 *For all $k \geq 2$, there is a class $\mathcal{C} \subseteq \mathcal{F}$ such that $\mathcal{C} \in \bigvee(\bigwedge L^k(\exists)) - (\bigwedge L^k(\exists) \cup \bigvee L^k(\exists))$.*

Proof. The proof of this proposition is a synthesis of the proofs of the preceding two results. We define a set of models $\{A_1, A_2, \dots\}$ which are pairwise $L^k(\exists)$ -incompatible such that for each i , the $L^k(\exists)$ -theory of A_i is not finitely axiomatizable in $L^k(\exists)$. We then let $\mathcal{C} = \{B \mid \exists i(A_i \preceq^k B)\}$. The arguments to show that this class is neither in $\bigvee L^k(\exists)$ nor in $\bigwedge L^k(\exists)$ are variants of the proofs of Propositions 15 and 16.

We define each model A_i as an expansion of a homeomorphic image of a graph which is a model of the $(k+1)$ -Gaifman theory. To clarify the exposition, we also add a unary predicate V to the signature to label the original vertices of the graph. Let R be a finite graph that verifies Γ_{k+1} ; observe that R also verifies Γ_k . Each A_i is obtained from R by replacing all edges which are not loops by pairwise disjoint paths of length i . (Where there is a two-way, undirected edge, a single undirected path is inserted, rather than two directed paths.) If $i = 1$, then A_1 is just the expansion of R , with signature $\{Exy, Vx\}$, such that $V^{A_1} = R$. If $i > 1$, then the universe of A_i is the set $R \cup \{\{a, b, j\} \mid a, b \in R, R \models Eab, \text{ and } 1 \leq j < i\}$. (We have labeled each new vertex by a set of size 3.) If a and b are connected in R , then in A_i there is a path of length i from a to b along the vertices $\{a, b, 1\}, \dots, \{a, b, i-1\}$. Again, we set $V^{A_i} = R$. Observe that each vertex $\{a, b, j\}$ is connected to exactly two other vertices. Also, if $a, b \in V^{A_i}, a \neq b$, then $d(a, b) \geq i$.

To verify that $\mathcal{C}$ is not in $\bigvee L^k(\exists)$, it suffices to show that there is a model $A \in \mathcal{C}$ and a sequence $B^1, B^2, \dots$, disjoint from $\mathcal{C}$, such that for each $n, A \preceq^{k,n} B^n$. Let A be A_1 , and let each B^n be obtained from the model B_k^n from the proof of Proposition 10 by putting every vertex into the extension of the predicate V . From that proof it is immediate that, for all $n, A_1 \preceq^{k,n} B^n$ but $A_1 \not\preceq^k B^n$. For each $i \geq 2, A_i \models \exists x \neg Vx$ and, consequently, $A_i \not\preceq^k B^n$. This establishes that each B^n is not in $\mathcal{C}$.

In order to show that $\mathcal{C} \not\subseteq \bigwedge L^k(\exists)$, we now define a single $B \notin \mathcal{C}$ such that for all n , there is an $A_{f(n)}$ such that $A_{f(n)} \preceq^{k,n} B$. By Proposition 14, this will establish that $\mathcal{C} \not\subseteq \bigwedge L^k(\exists)$. Let R^+ be an expansion of R obtained by letting $V^{R^+} = \{a\}$ for some $a \in R$ such that $R \models Eaa$. Let R^- be an expansion of R obtained by letting $V^{R^-} = \{a\}$, for some $a \in R$ such that $R \models \neg Eaa$. We say that an element a *contains a loop*, or *is looped*, iff Eaa . Otherwise, it is *loop free*. Likewise, we say that each R^+ is looped and that R^- is loop free. We define B to be the disjoint union of k copies of both R^+ and R^- . A *component* of B is any submodel that is one of the copies of R^+ or R^- . Observe that the

components are exactly the maximal connected submodels. Here the predicate V plays a role like the constants s and t in the proof of Proposition 16.

It is easy to see that B is not in $\mathcal{C}$. For each i , A_i has the property, expressible in L^3 , that there are two distinct vertices, both in the relation V , that are connected by a path of length i . On the other hand, no component of B contains two distinct elements in V ; thus, for all i , $A_i \not\preceq^k B$.

Let $f(x) = 2^{x+1} + 1$. It remains to establish that, for each n , $A_{f(n)} \preceq^{k,n} B$. As in the proof of Proposition 16, the Duplicator can win the n -move $\exists^k$ -game on $A_{f(n)}$ and B because the vertices of $A_{f(n)}$ that are in V are too far apart for the S to distinguish the models by witnessing that they are actually connected. In order to describe the D's winning strategy, we define an auxiliary *matching* partial function, $\mu(x, m)$, that assigns to each vertex $x \in A_{f(n)}$ that is pebbled in round m a vertex $a' \in A_{f(n)}$ such that $A_{f(n)} \models Va'$. We will write $\mu_m(x)$ for $\mu(x, m)$, or even omit the subscript when it is unnecessary. Let $a_j \in A_{f(n)}[b_j \in B]$ denote the vertex pebbled by the S [D] in round j ; let R_j denote the component of B that contains b_j . For all $a \in A^{f(n)}$, say that a is *live* in round m iff Va or a is occupied by a pebble at the end of the round. The function $\mu_m(x)$ will satisfy the following conditions, for all $a, a' \in \{b \mid b \text{ is live in round } m\}$:

1. If the S does not replay the pebble on a in round m , then $\mu_{m+1}(a) = \mu_m(a)$.
2. For all $m \leq n$, if Va , then $\mu_m(a) = a$.
3. If $a \neq a'$ and $\mu_m(a) \neq \mu_m(a')$, then $d(a, a') > 2^{n-(m+1)}$. In particular, if $\mu_m(a) \neq \mu_m(a')$, then there is no edge connecting a and a' .

The D will also maintain the following ‘modularity’ condition.

4. In each round m , if the pebbles on a_i and a_j , $i < j$, have not been replayed between rounds i and m , then $\mu_m(a_i) = \mu_m(a_j)$ iff b_i and b_j are on the same component of B .

In round 1 of the game, let the S play on $a_1 \in A_{f(n)}$. Let $\mu_1(a_1)$ be the element $a' \in V$ that is closest to a_1 ; observe that this is well-defined and that $d(a_1, \mu_1(a_1)) \leq 2^n$. Since the distance between any two elements in V is greater than $f(n) = 2^{n+1}$, this implies that for all $v \in V - \{a\}$, $d(a, v) > 2^n$, as required by condition 3. The D then chooses a component

R_1 of B such that R_1 is looped iff $\mu(a_1)$ is looped; let v_1 be the unique element in R_1 such that Vv_1 . She then plays on an element $b_1 \in R_1$ such that the 2-tuple (b_1, v_1) has the same atomic type as $(a_1, \mu(a_1))$, which immediately implies that the pebbles in play determine a partial isomorphism. Since $R_1 \models \Gamma_{k+1}$, it is easy to see that there is such an element.

In round $m+1 \leq n$, the S plays on some a_{m+1} . We describe the D's response by considering two cases. One, for all $a \neq a_{m+1}$ that are live in round $m+1$, $d(a, a_{m+1}) \geq 2^{n-((m+1)+1)}$. In this case, let $\mu_{m+1}(a_{m+1})$ be any $v \in V$ such that, for all live $a \neq v$, $\mu_m(a) \neq v$. The D now chooses an unpebbled component of B , which we call R_{m+1} , that is looped iff v is looped. Since there are k copies of each of R^+ and R^- , and only k pebbles, there is always such a component. She then plays on some $b_{m+1} \in R_{m+1}$ such that the atomic type of (b_{m+1}, v_{m+1}) is the same as that of $(a_{m+1}, \mu(a_{m+1}))$, where v_{m+1} is the unique element of R_{m+1} such that Vv_{m+1} . Note that for all live, pebbled $a_j, j \leq m$, $\mu_{m+1}(a_j) \neq \mu_{m+1}(a_{m+1})$, which implies, by conditions 2 and 4 above, that $a_j[b_j]$ is not adjacent to $a_{m+1}[b_{m+1}]$. Therefore the D has succeeded in preserving a partial isomorphism; it is easy to verify that conditions 1–4 are also maintained.

Second if there is an element $a \in \{b \mid b \text{ is live}\}$ such that $d(a, a_{m+1}) \leq 2^{n-(m+2)}$, then let $\mu_{m+1}(a_{m+1})$ be $\mu_m(a)$. In order to see that $\mu_{m+1}(a_{m+1})$ is well-defined, suppose that there are two such elements, a and a' . Observe that $d(a, a') \leq d(a, a_{m+1}) + d(a_{m+1}, a') \leq 2^{n-(m+1)}$. By condition 3, we have that $\mu_m(a) = \mu_m(a')$, as desired. Note that $\mu_{m+1}(x)$ satisfies the above conditions 1–3. The D then plays on some element $b_{m+1} \in R_{m+1}$ such that $(b_{l_1}, \dots, b_{l_j}, b_{m+1}, v_{m+1})$ has the same atomic type as $(a_{l_1}, \dots, a_{l_j}, a_{m+1}, \mu_{m+1}(a_{m+1}))$, where $\mu_{m+1}(a_{l_i}) = \mu_{m+1}(a_{m+1})$ and $R_{l_i} = R_{m+1}$, for all $i \leq j$. Again, this is possible because $R_{m+1} \models \Gamma_{k+1}$. Note that the D also maintains condition 4. This establishes that the D has a winning strategy. $\blacksquare$

Finally, we prove the following separation.

Proposition 18 *Over $\mathcal{F}$, for $k \geq 3$, $L^k(\exists) \subset (\bigwedge L^k(\exists)) \cap (\bigvee L^k(\exists))$.*

Proof. Let $\text{Path}(x, y)$ express the binary query ‘there is an E -path from x to y .’ For signature $\sigma = \{E, s\}$, we define $\mathcal{C} = \{A \mid \exists x (\text{Path}(s, x) \text{ and } \text{Path}(x, x))\}$. Let $\theta_n(x, y)$ be an $L^3(\exists)$ formula that defines the binary query ‘there is a path of length n from x to y .’ It is easy to see that $\mathcal{C}$ is in $\bigvee L^k(\exists)$. Also observe that $\varphi = \bigwedge_{n \in \omega} \exists x \exists y (s = x \wedge \theta_n(x, y))$ defines

$\mathcal{C}$. Finally, there are arbitrarily large minimal models in $\mathcal{C}$, that is, models $A \in \mathcal{C}$ such that for all proper submodels $B \subset A$, $B \notin \mathcal{C}$. This immediately implies that $\mathcal{C} \notin \text{FO}(\exists)$ and, *a fortiori*, not in $L^k(\exists)$. ■

Chapter 3

Existential preservation

3.1 Generalized preservation theorems

In this section, we prove some generalized preservation theorems for fragments of FO. Our results are of the form

$$L \cap \text{EXT} \subseteq L'$$

for certain quantifier prefix classes $L \subset \text{FO}$, and $L' = L_{\infty\omega}^\omega(\exists)$ or $\text{Datalog}(\neq, \neg)$. We introduce the following notation.

Definition 3 *Let w be a regular expression over the alphabet $\{\forall, \exists\}$, in the sense of formal language theory. $\text{FO}[w]$, with square brackets, is the set of prenexed sentences φ such that the quantifier prefix of φ is a word in the regular language determined by the regular expression w . (For example, $\text{FO}[\forall\exists^*]$ is the set of sentences whose quantifier prefix is a single $\forall$ followed by a string of $\exists$'s.)*

Recall that Tait [23] showed $\text{FO} \cap \text{EXT} \not\subseteq \text{FO}(\exists)$. Gurevich and Shelah [14, 15] give examples witnessing that

$$\text{FO}[\forall\exists^*] \cap \text{EXT} \not\subseteq \text{FO}(\exists)$$

and Compton observed that

$$\text{FO}[\exists^*\forall^*] \cap \text{EXT} \subseteq \text{FO}(\exists),$$

showing that these examples are best possible in terms of quantifier alternation prefix (see [14]). Kolaitis and Vardi (see [3]) observed that the example of Gurevich and Shelah [14]

can be defined in $\text{Datalog}(\neq, \neg)$. Theorem 2 below establishes that

$$\text{FO}[\exists^*\forall\exists] \cap \text{EXT} \subseteq \text{Datalog}(\neq, \neg).$$

It follows that the above mentioned examples in the literature witnessing the failure of the Los-Tarski theorem in the finite case are definable in $\text{Datalog}(\neq, \neg)$, since all these examples are in $\text{FO}[\exists^*\forall\exists]$. The next theorem establishes a slightly more general result with $L_{\infty\omega}^\omega(\exists)$ in place of $\text{Datalog}(\neq, \neg)$.

Theorem 1 $\text{FO}[\exists^*\forall\exists^*] \cap \text{EXT} \subseteq L_{\infty\omega}^\omega(\exists)$.

Proof. Let $\varphi \in \text{FO}[\exists^*\forall\exists^*] \cap \text{EXT}$. That is, $\varphi \in \text{FO}[\exists^*\forall\exists^*]$ and $\text{Mod}_f(\varphi) \in \text{EXT}$. Let $\mathcal{C} = \text{Mod}_f(\varphi)$. We proceed to show that $\mathcal{C} \in L_{\infty\omega}^\omega(\exists)$. By Proposition 8, it suffices to show that there is a k such that, for each $A \in \mathcal{C}$ and $B \notin \mathcal{C}$, there is a $\theta_{A,B} \in L_{\infty\omega}^k(\exists)$ such that $A \models \theta_{A,B}$ and $B \not\models \theta_{A,B}$.

Let $\varphi = \exists x_1 \dots x_i \forall y \exists z_1 \dots z_j \psi(\bar{x}, y, \bar{z})$, where ψ is quantifier free, and let $k = i + j + 1$ (we suppose, without loss of generality, that $i > 0$). We now describe a winning strategy for S in the eternal $\exists^k$ -game on A and B , for $A \in \mathcal{C}$ and $B \notin \mathcal{C}$, which establishes, by Proposition 6, the existence of $\theta_{A,B} \in L_{\infty\omega}^k(\exists)$ with the desired properties. There are two stages. Let $\bar{a} = (a_1, \dots, a_i)$ be a sequence of elements of A such that $A \models \forall y \exists \bar{z} \psi(\bar{a}, y, \bar{z})$. If D has not lost after h rounds, for $h < i$, S plays pebble α_{h+1} on element a_{h+1} . If S has not won after i moves, and D has played her pebbles on $\bar{b} = (b_1, \dots, b_i)$, then $B \models \exists y \forall \bar{z} \neg \psi(\bar{b}, y, \bar{z})$ (since $B \not\models \varphi$).

The goal of the second part of S's strategy is to force D to play a pebble on some element b' such that $B \models \forall \bar{z} \neg \psi(\bar{b}, b', \bar{z})$, without removing any of the pebbles $\alpha_1, \dots, \alpha_i$ which 'fix the interpretation' of the variables $x_1, \dots, x_i$ on both A and B . Regardless of the element a' on which S will have played his corresponding pebble, $A \models \exists \bar{z} \psi(\bar{a}, a', \bar{z})$, so that he can then win easily. In order to describe S's strategy, we first define a sequence of subsets of the universe of B . Let $\Gamma_0 = \{b' \mid b' \in B \text{ and } B \models \forall \bar{z} \neg \psi(\bar{b}, b', \bar{z})\}$. Observe that $B \models \exists y \forall \bar{z} \neg \psi(\bar{b}, y, \bar{z})$, and therefore Γ_0 is non-empty. Given $\Gamma_0, \dots, \Gamma_m$, if $(\bigcup_{l \leq m} \Gamma_l) \cap \bar{b} = \emptyset$, then let B_{m+1} be the submodel of B whose universe is $(B - \bigcup_{l \leq m} \Gamma_l)$. Let $\Gamma_{m+1} = \{b' \mid b' \in B_{m+1} \text{ and } B_{m+1} \models \forall \bar{y} \neg \psi(\bar{b}, b', \bar{y})\}$. For each B_m , since $B_m \subseteq B$, we have that $B_m \models \forall \bar{x} \exists y \forall \bar{z} \neg \psi(\bar{x}, y, \bar{z})$. In particular, $B_m \models \exists y \forall \bar{z} \neg \psi(\bar{b}, y, \bar{z})$ and thus, as above, Γ_{m+1}

is non-empty. Since B is finite, there is some n such that $\Gamma_n \cap \bar{b} \neq \emptyset$, and some element $b_f \in \Gamma_n \cap \bar{b}$ pebbled by β_f . Then B is partitioned into the sets $\Gamma_0, \dots, \Gamma_{n-1}, B_n$. We also have that $A \models \exists \bar{z} \psi(\bar{a}, a_f, \bar{z})$, and $B_n \models \forall \bar{z} \neg \psi(\bar{b}, b_f, \bar{z})$.

The Spoiler can win by executing a substrategy that compels D to play in sets Γ_m of successively smaller index. Let $\bar{c}$ be a sequence of elements of length j such that, $A \models \psi(\bar{a}, a_f, \bar{c})$. S plays his next j moves on this sequence, until D makes a losing move or plays a pebble β_g onto an element in Γ_m , for $m \leq n-1$. We claim that one of these two possibilities must occur. For suppose that D plays on a sequence $\bar{d} \subseteq B_n$. Then $B_n \models \neg \psi(\bar{b}, b_f, \bar{d})$, and $\psi(\bar{a}, y, \bar{z})$ witnesses that the function that takes $\bar{a} * a_f * \bar{c}$ to $\bar{b} * b_f * \bar{d}$ and preserves the denotations of constants is not a partial isomorphism.

Suppose that D has played some pebble β_g into some set Γ_m . By the same argument as above, reusing pebbles $\{\alpha_{i+1}, \dots, \alpha_k\} - \{\alpha_g\}$, S can either win or force D to play into some $\Gamma_{m'}$, for some $m' < m$. Iterating this procedure, S can force D to play into Γ_0 , and then win by using the same procedure one more time. $\blacksquare$

We remark the following two refinements of the foregoing theorem.

1. For each $B \notin \mathcal{C}$, there is a number m_B such that for all $A \in \mathcal{C}$, S wins the m_B -round $\exists^k$ -game on A and B . (Here, m_B is determined by the maximum number of sets Γ that get defined on B , for any choice of D's first i moves.) It follows easily from Proposition 5 that this condition is equivalent to there being a $\theta_B \in L^k(\exists)$, with quantifier rank $\leq m_B$, such that for all $A \in \mathcal{C}$, $A \models \theta_B$, and $B \not\models \theta_B$. Then $\theta' = \bigwedge_{B \notin \mathcal{C}} \theta_B$ is equivalent to φ and is a single infinite conjunction of $L^k(\exists)$ sentences. We know by Proposition 16 that not all sentences of $L_{\infty\omega}^k(\exists)$ can be expressed in this form. Indeed, it follows from Theorem 2 below that if $\varphi \in \text{FO}[\exists^* \forall \exists] \cap \text{EXT}$, then φ is equivalent to a formula in $\bigwedge L^k(\exists) \cap \bigvee L^k(\exists)$ for some k .
2. Suppose that φ is an L^k sentence with quantifier type $\forall \exists^*$, that is, no $\forall$ occurs in φ in the scope of another quantifier. In this case, we can show, by a modification of the proof of Theorem 1, that φ is equivalent to an $L_{\infty\omega}^k(\exists)$ sentence. This contrasts with Proposition 19 below which establishes that for all k , there is a sentence $\varphi_k \in L^3$ such that $\text{Mod}_f(\varphi_k) \in \text{EXT}$, but φ_k is not equivalent over $\mathcal{F}$ to any sentence in $L_{\infty\omega}^k(\exists)$.

Theorem 2 $\text{FO}[\exists^* \forall \exists] \cap \text{EXT} \subseteq \text{Datalog}(\neq, \neg)$.

Proof. Let $\varphi = \exists x_1 \dots x_j \forall y \exists z \beta(\bar{x}, y, z)$, with $\beta(\bar{x}, y, z)$ quantifier free. Let $\bar{c} = (c_1, \dots, c_p)$ be the sequence of constants in the signature of φ and let $\mathcal{C} = \text{Mod}_f(\varphi)$. For $a \in A$, we say that a *closes* with parameters $\bar{a}$ iff there is a sequence $a_0(= a), a_1, \dots, a_n$ such that for all $l < n$, $A \models \beta(\bar{a}, a_l, a_{l+1})$ and there is an $m \leq n$ such that $A \models \beta(\bar{a}, a_n, a_m)$. Note that this is equivalent to there being an a' such that there is a $\beta(\bar{a}, y, z)$ -path from a to a' , and a $\beta(\bar{a}, y, z)$ -cycle including a' .

We claim that $A \models \varphi$ iff there is a j -tuple $\bar{a}$ such that every element of $\bar{a} \cup \bar{c}$ closes with parameters $\bar{a}$. Suppose that A does not satisfy these conditions. We prove that $A \models \forall \bar{x} \exists y \forall z \neg \beta(\bar{x}, y, z)$ where the latter sentence is equivalent to $\neg \varphi$. Let $\bar{a} \subseteq A$ be a sequence of length j . By hypothesis, there is an $a' \in \bar{a} \cup \bar{c}$ such that a' does not close with parameters $\bar{a}$. Since A is finite, this implies that there is an $m \geq 0$ and a sequence $a' = a'_0, \dots, a'_m$ such that for all $l < m$, $A \models \beta(\bar{a}, a'_l, a'_{l+1})$ and $A \models \forall z \neg \beta(\bar{a}, a'_m, z)$, as desired.

In the other direction, let $\bar{a}$ be such that every member of $\bar{a} \cup \bar{c}$ closes with parameters $\bar{a}$. Let $\bar{s}_h = \langle a_{h0}(= a_h), \dots, a_{hm_h} \rangle$ and $\bar{t}_h = \langle e_{h0}(= c_h), \dots, e_{hn_h} \rangle$ be sequences witnessing that each element of $\bar{a} \cup \bar{c}$ closes with parameters $\bar{a}$. Let B be the submodel of A with universe $\bigcup_i \bar{s}_i \cup \bigcup_j \bar{t}_j$. Then it is easy to verify that $B \models \varphi$ and, since $\text{Mod}_f(\varphi) \in \text{EXT}$, it follows that $A \models \varphi$.

The following program, with $\bar{x} = (x_1, \dots, x_j)$, computes φ :

$$P(\bar{x}, y, z) \leftarrow \beta(\bar{x}, y, z)$$

$$P(\bar{x}, y, z) \leftarrow P(\bar{x}, y, w), P(\bar{x}, w, z)$$

$$Q \leftarrow P(\bar{x}, x_1, y_1), P(\bar{x}, y_1, y_1), \dots, P(\bar{x}, x_j, y_j), P(\bar{x}, y_j, y_j),$$

$$P(\bar{x}, c_1, w_1), P(\bar{x}, w_1, w_1), \dots, P(\bar{x}, c_p, w_p), P(\bar{x}, w_p, w_p)$$

This can be easily converted into a Datalog($\neq, \neg$) program. Let $\beta(\bar{x}, y, z) = \bigvee_i \delta_i$, where each δ_i is a conjunction of literals. Replace the clause $P(\bar{x}, y, z) \leftarrow \beta(\bar{x}, y, z)$ with the clauses $P(\bar{x}, y, z) \leftarrow \delta_i$, for all i . ■

3.2 The failure of existential preservation for $L_{\infty\omega}^\omega$

In this section we prove that $L_{\infty\omega}^\omega \cap \text{EXT} \not\subseteq L_{\infty\omega}^\omega(\exists)$. Indeed, we establish that there is a sentence $\theta \in L_{\infty\omega}^\omega$ such that $\text{Mod}(\theta)$ is closed under extensions, but there is no $\psi \in L_{\infty\omega}^\omega(\exists)$ such that $\text{Mod}_f(\theta) = \text{Mod}_f(\psi)$. Thus, θ witnesses the failure of existential preservation for $L_{\infty\omega}^\omega$ simultaneously over the class of finite structures and over the class of all structures. The central lemma on which this result relies is of interest in itself. It says that for all $k \geq 3$, the finitary language L^k fails in a strong way to satisfy an existential preservation property. Andreka, van Benthem, and Nemeti [4] showed that for every $k \geq 3$, there is a sentence $\varphi_k \in L^k$ which is preserved under extensions, but which is not equivalent to any sentence of $L^k(\exists)$. For $k \geq 3$, the sentence φ_k they construct uses a relation symbol of arity $k - 1$ and has the property that it is equivalent to a sentence of $L^{k+1}(\exists)$. They state the following open problems.

- For any $k \geq 3$ and $n \in \omega$, is there a sentence $\varphi_n \in L^k$ which is preserved under extensions, but which is not equivalent to any sentence of $L^{k+n}(\exists)$?
- For $k > 3$, is there a formula of L^k containing only (one) binary relation symbols which is preserved under extensions, but is not equivalent to any sentence of $L^k(\exists)$?

The next proposition settles both these open problems. The main result of the section follows easily from the proof of this proposition.

Proposition 19 *For each $k < \omega$, there is a sentence $\theta_k \in L^3$, containing a single binary relation, such that*

1. $\text{Mod}(\theta_k)$ is closed under extensions, but
2. $\text{Mod}_f(\theta_k) \neq \text{Mod}_f(\varphi)$ for all $\varphi \in L^k(\exists)$.

Proof. Before presenting the full proof, we sketch the basic outline. Let the k -pyramid of B , $\mathcal{P}^k(B)$, be the smallest class of (finite and infinite) models containing B that is closed under substructures and L^k -equivalence. For each $k \geq 3$, we define finite structures A_k and B_k with the following properties:

1. $A_k \preceq_{\infty\omega}^k B_k$;

2. $\mathcal{P}^3(B_k)$ is L^3 -definable;

3. $A_k \notin \mathcal{P}^3(B_k)$.

Let $\varphi_k \in L^3$ be such that $\text{Mod}(\varphi_k) = \mathcal{P}^3(B_k)$, and let $\theta_k = \neg\varphi_k$. It is obvious that $\text{Mod}(\theta_k)$ is closed under extensions, that $A_k \models \theta_k$, and that $B_k \not\models \theta_k$. Suppose $\varphi \in L^k(\exists)$ is such that $A_k \models \varphi$. Since $A_k \preceq_{\infty\omega}^k B_k$, this implies that $B_k \models \varphi$, and therefore that φ is not equivalent to θ_k .

We define structures A_k and B_k in terms of simpler submodels. For $f \leq t$, let the $[t, f]$ -flag, $F[t, f]$, be the directed chain of length t with one additional vertex attached to the f^{th} link. That is, the vertex set of $F[t, f]$ is $\{0, 1, \dots, t, t+1\}$, and the edge relation is $\{(i, i+1) \mid i < t\} \cup \{(f, t+1)\}$. A_k is the disjoint union of the $k+1$ flags— $F[2k+2, k+1], F[2k+2, k+2], \dots, F[2k+2, 2k+1]$. Let the $[k, j]$ -tree, $T[k, j]$, be the tree obtained from A_k by fusing the i^{th} nodes of each flag, for all $i \leq j$. This tree has height $2k+2$ and the node at height j has outdegree $k+1$. Then B_k is the disjoint union of the k trees— $T[k, 0], T[k, 1], \dots, T[k, k-1]$.

First we show that $A_k \preceq_{\infty\omega}^k B_k$ by describing a winning strategy for D in the eternal $\exists^k$ -game on A_k and B_k . A *component* of a model is a maximal connected submodel. Observe that every component of A_k is embeddable in every component of B_k . Call a component of either A_k or B_k *vacant* at round n if there is no pebble located on any element of that component before the players make their n^{th} moves. We consider two cases of moves for S. First, suppose that in some round n , S plays pebble α_i on a vacant component A^n of A_k . Since there are only k pairs of pebbles, and since pebble β_i is not on the board, there is a vacant component B^n of B_k , and an isomorphic injection $h_n : A^n \hookrightarrow B^n$. D will play pebble β_i on $h_n(\alpha_i)$. In the other case, S plays on a non-vacant component A^n . There is some $m < n$ such that A^n has been occupied continuously since round m and either $m = 1$ or A^n was vacant at round $m-1$. Thus $A^n = A^m$, and there are previously defined B^m and h_m . D now plays β_i on $h_m(\alpha_i)$. By this condition, every pair of pebbles (α_l, β_l) on components A^m and B^m satisfies the condition that $h_m(\alpha_l) = \beta_l$. In both cases, it is clear that D has maintained a partial isomorphism. By Proposition 6, it now follows immediately that $A_k \preceq_{\infty\omega}^k B_k$.

Next, we show that $\mathcal{P}^3(B_k)$ is definable in L^3 . Consider the following properties:

1. A contains no chains of length $\geq 2k + 2$.
2. A contains no cycles of length $\leq 2k + 2$.
3. No element $a \in A$ has indegree ≥ 2 , that is, $A \models \neg \exists x \exists y \exists z (x \neq y \wedge Exz \wedge Eyz)$.

It is easy to show that each property is expressible in L^3 , is closed under substructures, and holds of B_k . From this it follows immediately that each $B' \in \mathcal{P}^3(B_k)$ possesses all three properties. Consequently, every member of $\mathcal{P}^3(B_k)$ is a forest consisting of directed trees of height $\leq 2k + 2$.

Next we note the following facts:

Lemma 2 *Let A and B be the disjoint unions of components $A_1, \dots, A_m$, and $B_1, \dots, B_n$, respectively. For $k \geq 3$, $A \equiv_{\infty\omega}^k B$ if and only if for each component A_i $[B_i]$, either the number of components of A that are L^k -equivalent to it is equal to the number of components of B that are L^k -equivalent to it or both numbers are $\geq k$.*

This result can be proved by a simple pebble game argument.

Lemma 3 *For each h , and each $k \geq 3$, up to equivalence in L^k there are only finitely many trees of height $\leq h$.*

The proof proceeds by induction on h . The case where $h = 1$ is obvious. Given a tree T , call a proper subtree that contains a node t of height 1 and all of its descendants a *1-tree* of T . For $h > 1$, we claim that two trees T_1 and T_2 of height at most h are L^k -equivalent if and only if for each 1-tree $T' \subset T_i$, the number of 1-trees of T_1 that are L^k -equivalent to T' equals the number of 1-trees of T_2 that are L^k -equivalent to T' , or both numbers are $\geq k$. The argument is just like the proof of the preceding lemma. From the claim, the lemma follows immediately.

Corollary 3 *For each h , and each $k \geq 3$, up to equivalence in L^k there are only finitely many forests of height $\leq h$.*

This is an immediate consequence of the preceding lemmas.

These observations establish that there are only finitely many complete L^k -theories that are satisfiable in $\mathcal{P}^3(B_k)$. Moreover, each such theory has a finite model. By [9], every

such theory is axiomatized by a single L^k sentence. Hence, if we let φ_k be the disjunction of these sentences, we have $\text{Mod}(\varphi_k) = \mathcal{P}^3(B_k)$ as desired.

Finally, we argue that $A_k \notin \mathcal{P}^3(B_k)$. By the definition of $\mathcal{P}^3(B_k)$, for every $B' \in \mathcal{P}^3(B_k)$, there is an $m \in \omega$ and a sequence $(E_0, D_1, E_1, \dots, D_m, E_m)$ of structures, with $B_k = E_0$ and $B' = E_m$, such that:

1. For all $1 \leq i \leq m$, $D_i \subseteq E_{i-1}$.
2. For all $1 \leq i \leq m$, $D_i \equiv^3 E_i$.

It suffices to show that for any such sequence, A_k cannot be embedded in any E_i . Let $g : \mathcal{P}^3(B_k) \mapsto \{0, 1, \dots, k+1\}$ be the function such that $g(D)$ is the maximum number of components of A_k that can be embedded in D pairwise disjointly. We show that for each $i \leq m$, $g(E_i) < k+1$. In fact, we show that g is monotonically decreasing on the aforementioned sequence. Because each D_i is a submodel of E_{i-1} , it is clear that $g(D_i) \leq g(E_{i-1})$. It remains to establish that $g(B_k) < k+1$ and that $g(E_i) \leq g(D_i)$.

Observe that any embedding of a flag $F[2k+2, f]$ into a component C of any $B' \in \mathcal{P}^3(B_k)$ must map the root of the flag to the root of C . This implies that no two flags of A_k can be disjointly embedded into any such component and, since B_k has only k components, that $g(B_k) < k+1$.

From Lemma 2, it follows that every E_i can be obtained from D_i by repeated application of the following three operations. First, replace some component with a component that is L^3 -equivalent to it. Second, add a disjoint copy of a tree that is L^3 -equivalent to at least 3 components. Third, remove a component that is L^3 -equivalent to at least 3 other components. Thus, it suffices to argue that no such operation performed on some $B' \in \mathcal{P}^3(B_k)$ can yield a B'' such that $g(B'') > g(B')$. It is obvious that removing a component cannot increase the value of g .

We claim that it suffices to consider the effect of the other two operations on components of height $= 2k+2$. If trees T and T' are L^3 -equivalent, then they have the same height. Also, no component $F[2k+2, f]$ of A_k can be embedded in any tree of height $< 2k+2$. This establishes that the presence of shorter components in a model B does not affect the value of $g(B)$.

Observe that for all trees T and T' such that $T \equiv^3 T'$, $F[t, f]$ can be embedded in T iff it can be embedded in T' . This is because the following property can be expressed in L^3 : there is an element x such that (i) there is a y such that there is a path of length f from y to x ; (ii) x has outdegree 2; (iii) there is a y such that there is a path of length $t - f$ from x to y . Over trees, this property says that the model embeds $F[t, f]$. Consequently the operation of replacement cannot increase the value of g .

It remains to establish that adding an additional component to a model $B' \in \mathcal{P}^3(B_k)$ does not change the value of g . We observe that B_k has the following properties:

1. For each $(2k + 2)$ -chain contained in B_k there is at most one $j, 0 \leq j \leq k - 1$, such that the j^{th} link of the chain has outdegree > 1 .
2. For each $(2k + 2)$ -chain contained in B_k there is at most one $j, k + 1 \leq j \leq 2k + 1$, such that the j^{th} link of the chain has outdegree > 1 .

These properties are closed under substructures and L^3 -equivalence; consequently, they hold of every model $B' \in \mathcal{P}^3(B_k)$. Let C_1, C_2 , and C_3 be L^3 -equivalent components of B' of height $2k + 2$. The above argument establishes that each C_i is either some $F[2k + 2, f]$, or the simple $(2k + 2)$ -chain. Let B'' be the extension of B' obtained by adding a component C_4 . Observe that, in fact, all four components must be isomorphic, and embed at most one isomorphism type of flag. Therefore, the image of any embedding $h : A_k \mapsto B''$ can contain vertices from at most one of these four components. This demonstrates that $g(B') = g(B'')$, and completes the proof. $\blacksquare$

The following result establishes the failure of existential preservation for $L_{\infty\omega}^\omega$.

Theorem 3 *There is a sentence $\theta \in L_{\infty\omega}^\omega$ such that both*

1. $\text{Mod}(\theta)$ *is closed under extensions.*
2. *For all $\varphi \in L_{\infty\omega}^\omega(\exists), \text{Mod}_f(\theta) \neq \text{Mod}_f(\varphi)$.*

Proof. We claim that it suffices to show that for each $k \in \omega$ there is a sentence $\theta_k \in L^3$ and a pair of finite models A_k and B_k such that

1. $\text{Mod}(\theta_k)$ is closed under extensions.

2. $A_k \models \theta_k$ and $B_k \not\models \theta_k$.

3. $A_k \preceq_{\infty\omega}^k B_k$.

4. For all j , $A_j \models \theta_k$.

Let $\theta = \bigwedge_k \theta_k$. It is clear that θ is closed under extensions and that it has finite models, since it is true in each A_k . Suppose that φ is a sentence in $L_{\infty\omega}^k(\exists)$ such that θ implies φ . Then $A_k \models \varphi$, and therefore $B_k \models \varphi$. But for all l , $B_l \not\models \theta$. Therefore, $\text{Mod}_f(\theta) \neq \text{Mod}_f(\varphi)$.

The sentences θ_k and the models A_k and B_k from the proof of Proposition 19 fail to meet condition 4 because for $j < k$, $A_j \not\models \theta_k$. To see this, observe that A_j will always be a submodel of B_k . To fix this defect, it suffices to construct A'_k, B'_k , and θ'_k as in the proof of Proposition 19 that also satisfy the additional condition that, for all j and k , $A'_j \notin \mathcal{P}^3(B'_k)$. In order to accomplish this, we add simple ‘gadgets’ to the models. Let the k -cycle, C_k , be the graph on k vertices whose edge relation forms a simple, directed cycle of length k . Then let A'_k and B'_k be obtained from A_k and B_k , respectively, by adding a disjoint copy of C_k . By slightly modifying the proof of Proposition 19, we can show that $A'_k \preceq_{\infty\omega}^k B'_k$, and that there is a $\theta'_k \in L^3$ satisfied by exactly the models in the complement of $\mathcal{P}^3(B'_k)$ such that $A'_k \models \theta'_k$. Finally, it is easy to verify that for $j \neq k$, the j -cycle cannot be embedded in any $B \in \mathcal{P}^3(B'_k)$ and, therefore, $A'_j \models \theta'_k$. ■

Chapter 4

Other (generalized) preservation theorems

In previous chapters, we investigated ‘existential logics’ and definability over the class EXT of sets of structures closed under extensions. We now turn our attention to some other natural classes of structures and examine the status of (generalized) preservation theorems in connection with these classes. Recall that a *homomorphism* $h(x)$ from A to B is a function from A to B such that for all n -ary relation symbols $R(\bar{x})$ in the signature of A , and all n -tuples $\bar{a}$ in A^n , if $A \models R(\bar{a})$, then $B \models R(h(\bar{a}))$. Let HOM be the class consisting of all sets of finite models that are closed under homomorphisms. A model B is an *enrichment* of A over the relations $R_1, \dots, R_t$ iff the universe of B is equal to the universe of A , and for all n -ary relations $R_i(\bar{x})$ $i \leq t$, and all n -tuples $\bar{a} \in A^n [= B^n]$, if $A \models R_i(\bar{a})$, then $B \models R_i(\bar{a})$. A class $\mathcal{C}$ of models is *monotone* in relations $R_1, \dots, R_t$ iff for all $A \in \mathcal{C}$, if B is an enrichment of A over the relations $R_1, \dots, R_t$, then $B \in \mathcal{C}$. Below we will be interested in sets of structures that are monotone in every relation of their signature. Let MON denote the class of such sets of finite models.

Preservation theorems from classical model theory provide exact characterizations of the FO-definable classes that are closed under homomorphisms and that are monotone. The Homomorphism preservation theorem says that a FO-definable class is closed under homomorphisms iff it is defined by a (*purely*) *positive existential sentence*, i.e. an existential sentence in which every relation symbol, and equality, occurs only positively. Lyndon’s

lemma states that a FO-definable class of models is monotone in relations $\overline{R} = \{R_1, \dots, R_k\}$ iff it is defined by a sentence θ in which each relation in $\overline{R}$ occurs only positively, i.e. in the scope of no negations. It is still an open problem whether the Homomorphism preservation theorem fails over $\mathcal{F}$. We discuss this question in depth in Section 1, in which we present some partial positive results. Ajtai and Gurevich [2] showed that Lyndon's lemma fails over the class of finite models. More recently, Stolboushkin [22] has constructed a simpler counterexample. Below, we give a slight simplification of Stolboushkin's example that is also monotone in every relation symbol. This result, and generalized preservation theorems over HOM and MON, are discussed in Section 2.

4.1 The class HOM

We investigate the status of preservation theorems over the class HOM. Although it is unknown whether the Homomorphism preservation theorem remains true over $\mathcal{F}$, below we present some partial positive results, answering the question for certain fragments of FO. In particular, we show that every sentence in $\text{FO}[\forall^* \exists^* \forall^*] \cap \text{HOM}$ is equivalent to a positive existential sentence. In contrast to earlier results for EXT, we resolve (affirmatively) the homomorphism preservation theorem only for the finite variable language L^2 . We then discuss the class IHOM of sets closed under injective homomorphisms. Finally, we establish a preservation theorem for identity free FO sentences over $\mathcal{F}$.

We introduce the following notation. Let $\text{FO}(+, \neq)$ denote the fragment of FO containing exactly those sentences in which no relation symbol occurs in the scope of a negation. Thus, the negation symbol may only bind equalities. We use $\text{FO}(\exists, +)$ to denote the purely positive existential fragment of FO. Adding inequalities to this fragment, we get $\text{FO}(\exists, +, \neq)$. In this terminology, the major open problem is whether $\text{FO} \cap \text{HOM} = \text{FO}(\exists, +)$.

4.1.1 The homomorphism preservation property for FO

In this section, we consider various fragments of FO. We first show that if $\varphi \in \text{FO} \cap \text{HOM}$ is either existential or positive, then it is equivalent to a positive existential sentence. Recall that A is a minimal model of a class $\mathcal{C}$ iff for all proper submodels B of A , $B \notin \mathcal{C}$. Also, the

positive diagram of a model A of cardinality n is the conjunction of all atomic formulas with free variables among $\{x_1, \dots, x_n\}$ that are true in A under some fixed injective assignment of these variables onto the universe of A . The next lemma is straightforward.

Lemma 4 *Let φ be an existential FO sentence such that $\text{Mod}_f(\varphi)$ is closed under homomorphisms. Then there is a $\theta \in \text{FO}(\exists, +)$ that is equivalent to φ .*

Proof. Let $\mathcal{C} = \text{Mod}_f(\varphi)$. Since $\mathcal{C}$ is in $\text{FO}(\exists)$, it has finitely many minimal models. Let θ be the disjunction of the existential closures of the positive diagrams of each minimal model in $\mathcal{C}$. It is easy to verify the equivalence of θ and φ . ■

Below we establish the complementary result for the positive fragment of FO. Our proof requires the following Ehrenfeucht-Fraisse game, played on a single structure. Unlike games played on two structures, in each round only one player makes a move.

Definition 4 *Let φ be a prenexed FO sentence, $\varphi = Q_1x_1 \dots Q_nx_n\psi(x_1, \dots, x_n)$, where each Q_i is a quantifier and ψ is quantifier free. The φ -game is played as follows. In each round $m, 1 \leq m \leq n$, the D plays if Q_m is an $\exists$; otherwise, the S plays. As usual, a move consists of placing a pebble, α_m , on some element of A . After n rounds, the D wins if $A \models \psi[\alpha_1, \dots, \alpha_n]$, and the S wins otherwise.*

The following proposition characterizes satisfaction of a sentence in a model game theoretically.

Proposition 20 *For all prenexed sentences, φ , and all structures A , $A \models \varphi$ iff the D has a winning strategy in the φ -game.*

Proposition 21 *For all satisfiable sentences φ , $\varphi \in \text{FO}(+, \neq) \cap \text{HOM}$ iff there is an equivalent $\theta \in \text{FO}(\exists, +)$. [If φ is unsatisfiable, and thus in HOM , then it is equivalent to the existential sentence $\exists x(x \neq x)$.]*

Proof. Let φ be satisfiable and in $\text{FO}(+, \neq) \cap \text{HOM}$. Let $\mathcal{C} = \text{Mod}_f(\varphi)$. If φ is valid, we let θ be $\exists x(x = x)$. Otherwise, by Corollary 6, proved below, we can assume that φ is identity free, i.e. contains no equalities or inequalities. Furthermore, we can also assume that φ has been prenexed and that its matrix is in conjunctive normal form, i.e. is a conjunction of

disjunctions of atomic formulas. We say that an atomic formula, $R\bar{x}$, is a $\forall$ -formula of φ iff there is a variable $x' \in \bar{x}$ that is bound by a universal quantifier in φ . Let θ be the sentence obtained from φ by deleting all occurrences of all $\forall$ -formulas and all universal quantifiers. For example, if φ were $\exists x_1 \forall x_2 \exists x_3 ((Rx_1 x_2 \vee Px_1 \vee Px_3) \wedge (Rx_1 x_3 \vee Rx_2 x_3 \vee Px_2))$, then θ would be $\exists x_1 \exists x_3 ((Px_1 \vee Px_3) \wedge (Rx_1 x_3))$. We claim that θ is equivalent to φ . (Observe that θ is obtained effectively from an identity free φ .)

First we show that every conjunct of φ contains a non- $\forall$ -formula. Suppose, for contradiction, that γ is a conjunct of φ that contains only $\forall$ -formulas. Let A be any model in $\mathcal{C}$, and let A' be the extension of A obtained by adding one element, a' , without altering any of the relations. Observe that A' is in $\mathcal{C}$, since the class is closed under homomorphisms. We claim that the S has a winning strategy in the φ -game on A' , which implies that $A' \models \varphi$, a contradiction. In order to win, it suffices for the S to always play each of his moves on the element a' , regardless of the D's play. Every variable assignment extending the assignment determined by S's moves falsifies the conjunct γ , and hence also the formula φ . Therefore the S wins the φ -game, as desired.

We now show that θ implies φ . Let $\varphi = Q_1 x_1 \dots Q_n x_n \bigwedge_{1 \leq j \leq k} \gamma_j$, where each γ_j is a disjunction of atomic formulas. Let $\theta = Q_{s_1} x_{s_1} \dots Q_{s_m} x_{s_m} \bigwedge_{1 \leq j \leq k} \delta_j$, where each Q_{s_l} is $\exists$, and each δ_j is obtained from γ_j by deleting all $\forall$ -formulas. Suppose that $A \models \theta$; let $\bar{a} = (a_{s_1}, \dots, a_{s_m}) \subseteq A$ be such that $A \models \bigwedge_{1 \leq j \leq k} \delta_j[\bar{a}]$. We now describe the D's winning strategy in the φ -game on A . In each round $s_l \leq n$, she plays a pebble on a_{s_l} . Any variable assignment for $\{x_1, \dots, x_m\}$ that is determined by such a game verifies each δ_j , hence also each γ_j . Therefore $A \models \varphi$, thereby establishing that θ implies φ .

Next we prove the opposite direction. Let $A \models \varphi$ and, again, let A' be the extension of A obtained by adding an 'isolated' element, a' . Since $A' \models \varphi$, the D has a winning strategy for the φ -game on A' . In particular, she can win a game in which the S plays every one of his pebbles on a' . Since a' is not a member of any tuple that is in any relation, $R^{A'}$, and since every atomic formula occurs only positively, we can assume that the D does not play any pebble on a' . Let $\bar{a} = (a_{s_1}, \dots, a_{s_m}) \subseteq A$ be some tuple in $A \subseteq A'$ such that the D wins the φ -game on A' in which the S always plays on a' and, in each round s_l , the D plays on a_{s_l} . Observe that each $\forall$ -formula is falsified by this variable assignment. Therefore each disjunction, γ_j , must contain a non- $\forall$ -formula, η_j , that is satisfied by this

variable assignment in A' . Since each η_j occurs in the disjunction δ_j , it is easy to see that $A \models \bigwedge_{1 \leq j \leq k} \delta_j[\bar{a}]$. Therefore, $A \models \theta$. This establishes that φ implies θ , and completes the proof. $\blacksquare$

Next, we establish another partial positive result over a fragment of FO, defined in terms of quantifier prefix structure. The proof uses the following version of the Ehrenfeucht-Fraisse game. The $\forall^l \exists^m \forall^n$ -game on A and B is a 3-round game played with $l + m + n$ labeled pebble pairs such that:

1. The Spoiler plays l pebbles, $\bar{\beta}^1 = (\beta_1, \dots, \beta_l)$, on B . The Duplicator then puts l pebbles, $\bar{\alpha}^1 = (\alpha_1, \dots, \alpha_l)$, on A .
2. In round 2, the S plays m pebbles $\bar{\alpha}^2 = (\alpha_{l+1}, \dots, \alpha_{l+m})$ on A . The D then puts m pebbles, $\bar{\beta}^2 = (\beta_{l+1}, \dots, \beta_{l+m})$, on B .
3. In round 3, the S plays n pebbles, $\bar{\beta}^3 = (\beta_{l+m+1}, \dots, \beta_{l+m+n})$ on B . The D then puts n pebbles, $\bar{\alpha}^3 = (\alpha_{l+m+1}, \dots, \alpha_{l+m+n})$, on A .

Of course, the D wins just in case the pebbles determine a partial isomorphism from A to B . The following lemma is easy to verify.

Lemma 5 *The following two conditions are equivalent.*

1. For all $\varphi \in \text{FO}[\forall^l \exists^m \forall^n]$, if $A \models \varphi$, then $B \models \varphi$.
2. The D has a winning strategy in the $\forall^l \exists^m \forall^n$ -game on A and B .

Proposition 22 $\text{FO}[\forall^* \exists^* \forall^*] \cap \text{HOM} = \text{FO}(\exists, +)$. Furthermore, given $\varphi \in \text{FO}[\forall^* \exists^* \forall^*] \cap \text{HOM}$, there is an effective procedure for finding an equivalent sentence $\theta \in \text{FO}(\exists, +)$.

Proof. Let φ be in $\text{FO}[\forall^l \exists^m \forall^n] \cap \text{HOM}$, and let σ be the signature of φ . We show that there is an $s \in \omega$ that bounds the size of every minimal model of $\mathcal{C} = \text{Mod}_f(\varphi)$. This implies that $\mathcal{C}$ is defined by a sentence in $\text{FO}(\exists)$ and thus, by Lemma 4, that it is actually definable in $\text{FO}(\exists, +)$. In fact, we can calculate s as a function of m and σ , which establishes that there is an effective procedure for finding a sentence equivalent to φ in $\text{FO}(\exists, +)$. Let r be the number of models, up to isomorphism, of signature σ and cardinality m , and let $s = r \cdot m$. Also let $t = l + m + n$.

Let A be a minimal model of $\mathcal{C}$. We want to show that there is a $B \in \mathcal{C}$, of cardinality $\leq s$, such that there is a homomorphism from B into A . By the minimality of A , the homomorphism must be onto, implying that the cardinality of A is also $\leq s$. Let $\{M_1, \dots, M_q\}$ be the set of submodels of A of cardinality $= m$, again up to isomorphism. We use $C \oplus D$ to denote the model which is the disjoint union of C and D , and $p \cdot D$ to denote the disjoint union of p copies of D . Let $G = (t \cdot M_1) \oplus \dots \oplus (t \cdot M_q)$, and let $B = M_1 \oplus \dots \oplus M_q$. It is obvious that there are homomorphisms from G onto B , and from B into A . Observe also that the cardinality of B is $q \cdot m \leq s$. Since $\mathcal{C}$ is closed under homomorphisms, it suffices to show that $G \in \mathcal{C}$.

To establish this fact, we define an extension A' of A , and describe the D's winning strategy for the $\forall^l \exists^m \forall^n$ -game on A' and G . Since $A' \models \varphi$, this implies that $G \models \varphi$. Let $A' = A \oplus G$, and let $f(x)$ be the obvious injection from G into A' . In Round 1, the S plays l pebbles, $\bar{\beta}^1$, on some l -tuple in G . The D then plays on the l -tuple $f(\bar{\beta}^1)$, in A' . In Round 2, the S plays some pebbles, $\bar{\alpha}^{2,0}$, on $A \subseteq A'$, and plays his other pebbles, $\bar{\alpha}^{2,1}$, on $G \subseteq A'$. Conceptually, the D makes her move in two stages. She first plays her pebbles, $\bar{\beta}^{2,1}$, on $f^{-1}(\bar{\alpha}^{2,1})$. She then chooses an unpebbled component M'_p of G , one of the copies of M_p , such that there is an embedding, $h(x)$, from M'_p into G that contains the tuple $\bar{\alpha}^{2,0}$ in its range. There must be such a component since G contains t copies of each M_p . The D then plays her pebbles, $\bar{\beta}^{2,0}$, on the preimage of $\bar{\alpha}^{2,0}$ under $h(x)$. It is clear that the D succeeds in maintaining a partial isomorphism. Now, let $f'(x)$ be the embedding of G into A' that equals $h(x)$ on M'_p , and equals $f(x)$ on $G - M'_p$. In Round 3, the D plays her pebbles, $\bar{\alpha}^3$, on the image, $f'(\bar{\beta}^3) \subseteq A'$, of the pebbles played by the S. It is easy to see that this is indeed a winning strategy for the D. ■

4.1.2 L^2 has the homomorphism preservation property

In this section, we show that L^2 has the homomorphism preservation property over $\mathcal{F}$ and over the class of all structures. That is, we show that $L^2 \cap \text{HOM} = L^2(\exists, +)$, where $L^2(\exists, +)$ is the set of sentences in $L^2 \cap \text{FO}(\exists, +)$. Recall that it is unknown whether L^2 has the existential preservation property, though the corresponding negative result has been established for all $L^k, k \geq 3$ (see [4]). As L^2 only contains two variables, we assume, without

loss of generality, that the signature, σ , does not contain any relation of arity ≥ 3 . Elements a and b are *adjacent* iff there is a binary relation $Rxy \in \sigma$ such that $A \models Rab \vee Rba$.

Proposition 23 *The homomorphism preservation theorem holds for L^2 over $\mathcal{F}$ and over the class of all structures. In fact, for all $\varphi \in L^2 \cap \text{HOM}$, an equivalent $\theta \in L^2(\exists, +)$ such that $qr(\theta) \leq qr(\varphi)$ can be found effectively.*

Proof. Let $\varphi \in L^2 \cap \text{HOM}$, and let $qr(\varphi) = n$. Let $\mathcal{C} = \text{Mod}_f(\varphi)$. (The same argument also establishes the claim in the infinite case.) For each model $A \in \mathcal{C}$, we define a sentence $\theta_A \in L^2(\exists, +)$, with $qr(\theta_A) \leq n$, such that $A \models \theta_A$ and $\text{Mod}_f(\theta_A) \subseteq \mathcal{C}$. From the construction, it will be clear that, although $\mathcal{C}$ is infinite, there are only finitely many distinct θ_A . Letting $\theta = \bigvee_{A \in \mathcal{C}} \theta_A$, it is immediate that $\text{Mod}_f(\theta) = \text{Mod}_f(\varphi)$.

For each model A , and elements $a, b \in A$, let $\rho^{a,b}(x, y)$ be the atomic type of (a, b) in A , i.e. the conjunction of all atomic formulas ψ , with free variables among x, y , such that $A \models \psi[a, b]$. For all $a \in A$, and all $m \leq n$, we also define a formula $\pi_a^m(x) \in L^2(\exists, +)$, with $qr(\pi_a^m(x)) \leq m$, such that $A \models \pi_a^m[a]$. Let $N(a)$, the *neighbors* of a , denote the set of $b \neq a$ such that a is adjacent to b . $\pi_a^0(x)$ is just the atomic type of a in A . For all $m + 1$, we essentially want $\pi_a^{m+1}(x)$ to be $\pi_a^m(x) \wedge \bigwedge_{b \in N(a)} \exists y (\rho^{a,b}(x, y) \wedge \pi_a^m(y))$, except that we eliminate redundant, identical conjuncts. (Here, $\pi_a^m(y)$ denotes the formula obtained from $\pi_a^m(x)$ by exchanging all occurrences of x and y .) This guarantees that, for fixed m , there are only finitely many formulas of the form $\pi_a^m(x)$. Finally, let $\theta_A = \bigwedge_{a \in A} \exists x \pi_a^{n-1}(x)$, again eliminating redundant conjuncts.

To show that θ_A implies φ , we define a model M such that (i) $M \models \theta_A$; (ii) $M \in \mathcal{C}$; (iii) for all B such that $B \models \theta_A$, there is a homomorphism from M to B . Since $\mathcal{C}$ is closed under homomorphisms, these conditions imply that every model of θ_A is in $\mathcal{C}$, as desired. Given θ_A , let $Q = \{q_1, \dots, q_i\}$ be the set of occurrences of (existential) quantifiers in θ_A . For definiteness, we stipulate that if $i < j$, then q_i occurs to the left of q_j in θ_A . The universe of M is Q . The interpretation of the relations on M is determined straightforwardly from θ_A , as follows. $M \models Eq_i q_j$ iff there is an occurrence of an atomic formula, Evw , such that ‘ v ’ and ‘ w ’ are bound by q_i and q_j , respectively. Similarly for unary predicates. Every formula occurs only positively, so M is well-defined. It is easy to see that M satisfies conditions (i) and (iii). Indeed, for all B , an assignment of variables that verifies that

$B \models \theta_A$ determines a homomorphism from M to B .

To prove that $M \in \mathcal{C}$, it suffices to show that there are A' and N such that (i) $A \subseteq A'$; (ii) $A' \equiv^{2,n} N$ and (iii) there is a homomorphism from N to M . Here, $A' \equiv^{2,n} N$ means that for all $\psi \in L^2$ with $qr(\psi) \leq n$, $A' \models \psi$ iff $N \models \psi$. Since $\mathcal{C}$ is closed under homomorphisms and $L^{2,n}$ -equivalence, (i) – (iii) imply that $M \in \mathcal{C}$. Let $N = 2 \cdot M$ and $A' = A \oplus N$. It is immediate that (i) and (iii) are satisfied.

We define the following supplementary relation on M , and hence also on N . For all $q_i, q_j \in M$, $Sq_i q_j$ iff q_j occurs in the scope of q_i and there is an occurrence of a binary atomic formula in θ_A that contains variables bound by both q_i and q_j . Observe that q_i and q_j in Q are adjacent in M iff they are adjacent in the model (Q, S) . We claim that (Q, S) is a directed forest, i.e. the disjoint union of directed trees. (Alternatively, G is a directed forest iff it is acyclic and every element a has indegree ≤ 1 .) The acyclicity of (Q, S) follows immediately from the definition of Sxy . To establish the claim, it suffices to prove the following lemma.

Lemma 6 *Let ψ be a formula of L^2 , and let q_j be an occurrence of a quantifier in ψ . Then there is at most one quantifier occurrence, q_i , such that (i) q_j is in the scope of q_i , and (ii) there is an atomic formula, Evw , in ψ such that ‘ v ’ and ‘ w ’ are bound by both q_i and q_j .*

Proof. Let q_j occur in ψ , and let $q_j x(\eta(x))$ be a subformula of ψ , such that the scope of q_j is $\eta(x)$. Every occurrence of a binary relation symbol that contains two distinct variables, contains the two variables, x and y , since $\psi \in L^2$. Suppose that q_i satisfies conditions (i) and (ii) of the lemma. Then q_i must bind every free occurrence of the variable y in the subformula $\eta(x)$. Therefore no other quantifier in ψ can satisfy this pair of conditions. ■

Since (Q, S) is a forest, there is a well-defined function, $\nu(x)$, on $M[N]$, such that $\nu(q_i)$ is the height of q_i in (Q, S) . The height of the model $M[N]$ equals $qr(\theta_A) - 1 \leq n - 1$.

We now establish that $A' \equiv^{2,n} N$ by describing the D’s winning strategy in the n -round 2-pebble game on A' and N . We claim that if the S can win the game, then he can do so playing according to the following ‘normal form’.

1. In each round $m + 1$, he plays the pebble pair that was not played in the previous round, and does not replay it on the same element it just occupied.

2. In round 1, he plays a pebble α_i on the A -component of A' .
3. In each round $m + 1$, he plays a pebble $\alpha_i[\beta_i]$ on an element adjacent to $\alpha_{1-i}[\beta_{1-i}]$.

Condition 1 is obvious. To see that 2 does not hinder the S, suppose that he does not play his first move on the A -component of A' . The D will then play all of her moves according to the bijection between N and the N -component of A' , until the S plays on the A -component. To win, the S must eventually play some α_i on the A -component. The D will then play β_i on the vacant M -component of N . At the start of the next round, pebbles α_j and $\beta_j, i \neq j$ are removed from the board; so the S could have reached the same position sooner by playing on $A \in A'$ in round 1.

Consider Condition 3. Suppose that in some round $m + 1$, the S plays β_i on an element of N not adjacent to β_{1-i} . The D then plays α_i on the corresponding element of a vacant M -component in A' . Since the pebbles $\alpha_j, \beta_j, i \neq j$ will be replayed in the next round, for the same reasons as above the S has not made any progress. Likewise, if the S plays on A' , again the D can respond by playing on the vacant M -component of N .

We now describe the D's winning strategy assuming that the S always plays in accord with the above conditions. The S begins by playing on some $a \in A \subseteq A'$. The D then plays on the q_k (of either M -component) such that q_k occurs in the formula θ_A as the quantifier that binds the formula $\pi_a^{n-1}(x)$. Observe that q_k is the root of an S -tree. In all later rounds, the β pebbles are always played adjacently, so the D can play so that these pebbles 'climb up the S -tree'. To win, she maintains the condition that the pebbles, α_i, β_i , are played in round m so that $\nu(\beta_i) = m - 1$ and β_i is located on the element q_l that binds the formula $\rho^{\alpha_j, \alpha_i}(x, y) \wedge \pi_{\alpha_i}^{n-m}(y)$. Suppose that the D has maintained this condition through m rounds, $m < n$, and that, in round $m + 1$, the S plays $\alpha_j \in N(\alpha_i)$. The D will then play pebble β_j on the quantifier occurrence that binds the formula $\rho^{\alpha_i, \alpha'_j}(x, y) \wedge \pi_{\alpha'_j}^{n-(m+1)}(y)$, which is a conjunct of $\pi_{\alpha_i}^{n-m}(y)$. The argument for the case where the S plays on N is similar. Because the D always plays 'up the tree', in every round $m > 2$, the β pebble of lesser height will be replayed. The S is thereby prevented from moving down the tree, as doing so would violate Condition 1. This establishes that $A' \equiv^{2,n} N$. Thus, θ_A implies φ , and $\theta = \bigvee_A \theta_A$ is equivalent to φ .

Lastly, we argue that θ can be found effectively. By induction on m , it is easy to show that one can effectively generate all possible formulas of the form $\pi_a^m(x)$. Thus one can also enumerate the (finite) set, Θ , of formulas of the form θ_A , with $qr(\theta_A) \leq n$, for fixed n . Let d be the maximal number of quantifiers occurring in any formula $\theta_i \in \Theta$. Every minimal model of $\text{Mod}_f(\theta_i)$ has cardinality $\leq d$, so φ is equivalent to θ_i iff the sentences are equivalent on all models of cardinality $\leq d$, which is decidable. Since Θ is finite, one can effectively find the $\theta \in \Theta$ that is equivalent to φ . ■

4.1.3 Injective homomorphisms

In this section, we briefly discuss a class that lies between EXT and HOM. Recall that a map $h(x)$ is *injective* iff for all $a, b \in \text{dom}(h)$, if $f(a) = f(b)$, then $a = b$. Let IHOM be the class consisting of exactly those sets of finite models which are closed under injective homomorphisms. Observe that $\text{HOM} \subseteq \text{IHOM} \subseteq \text{EXT}$, and that each inclusion is proper. Over the class of all structures, the Injective homomorphism theorem says that a FO-definable class of models is closed under injective homomorphisms iff it is definable by a FO($\exists, +, \neq$) sentence. Minor modifications of the proofs of Proposition 19 and Theorem 3 yield the following results.

Proposition 24 *For each $k < \omega$, there is a sentence $\theta_k \in L^3$, containing a single binary relation, such that*

1. $\text{Mod}(\theta_k)$ is closed under injective homomorphisms, but
2. $\text{Mod}_f(\theta_k) \neq \text{Mod}_f(\varphi)$ for all $\varphi \in L^k(\exists)$.

(Sketch) Alter the proof of Proposition 19 by defining the k – pyramid of B , $\mathcal{P}^k(B)$, to be the smallest class of (finite and infinite) models containing B that is closed under substructures, L^k -equivalence, and *impoverishments*. (A is an impoverishment of B iff B is an enrichment of A .) The proof then proceeds as before.

The next theorem follows from the previous proposition as Theorem 3 follows from Proposition 19.

Theorem 4 *There is a sentence $\theta \in L_{\infty\omega}^\omega$ such that both*

1. $\text{Mod}(\theta)$ is closed under injective homomorphisms.

2. For all $\varphi \in L_{\infty\omega}^\omega(\exists)$, $\text{Mod}_f(\theta) \neq \text{Mod}_f(\varphi)$.

Beyond these two propositions, other results relating to preservation properties for EXT and HOM do not seem to generalize easily to yield analagous results for IHOM. For example, we do not see how to adapt any of the examples witnessing the failure of the Existential preservation theorem, due to Tait, Gurevich-Shelah, and Grohe, to define non-trivial FO-classes in IHOM. Furthermore, our proofs of partial positive results concerning FO-definability over HOM appear to rely essentially on the stronger closure properties of HOM. There are thus various open questions regarding injective homomorphism preservation properties over $\mathcal{F}$. We pose the following problem.

- Does the Injective homomorphism preservation theorem hold over $\mathcal{F}$?

By Lemma 4, an affirmative answer to this question immediately implies the Homomorphism preservation theorem over $\mathcal{F}$, though it is uncertain whether the reverse implication holds.

This brief section indicates that the class IHOM is rather different than EXT and HOM, while still sharing features with both classes. Resolving the status of the Injective homomorphism preservation theorem in either way would yield additional information about older questions and results. Thus, a negative answer would clearly strengthen Tait's result. More generally, we believe that further understanding of the relationship between definability over EXT, IHOM, and HOM will provide insight into FO-definability and (generalized) preservation properties.

4.1.4 Identity free FO

The following preservation theorem characterizes the expressive power gained from adding the identity sign to the language of FO. As the proof uses a modified Ehrenfeucht-Fraisse game, it is simultaneously a proof over $\mathcal{F}$ and over the class of all structures. A map $h(x)$ from A to B is a *strict surjection* iff it is a homomorphism of A onto B such that for all k -ary relations, $R(\bar{x})$, in the signature of A , and all k -tuples $\bar{a} \subseteq A$, $A \models R(\bar{a})$ iff $B \models R(h(\bar{a}))$. A class $\mathcal{C}$ is *closed under reverse strict surjections* iff for all A , and all $B \in \mathcal{C}$,

if there is a strict surjection from A to B , then $A \in \mathcal{C}$. For the rest of this section, to avoid trivialities, we restrict our attention to languages with non-empty signatures.

Definition 5 *The n -round ('identity free') i.f.-game on A and B is played according to the same rules as the standard n -round Ehrenfeucht-Fraïssé game on A and B , but has different winning conditions. The S wins at some round m iff there is a k -ary relation symbol, $R(\bar{x})$, and a sequence, $\bar{p} = (p_1, \dots, p_k)$, $p_i \leq m$, such that $A \models R(\alpha_{p_1}, \dots, \alpha_{p_k})$ iff $B \not\models R(\beta_{p_1}, \dots, \beta_{p_k})$. The D wins the game if the S does not win at any round $m \leq n$.*

Observe that the D does not have to play so that the pebbles determine a bijection between the models. This reflects the absence of identity in the language under consideration. We omit the obvious equivalent algebraic characterization. The following proposition and corollary are stated without proof.

Proposition 25 *Given models A and B , the following two conditions are equivalent.*

1. *For all identity free sentences, φ , with quantifier rank $\leq n$, $A \models \varphi$ iff $B \models \varphi$.*
2. *The D has a winning strategy in the n -round i.f.-game on A and B .*

Corollary 4 *For all classes $\mathcal{C}$, the following two conditions are equivalent.*

1. *$\mathcal{C}$ is defined by an identity free sentence of quantifier rank $\leq n$.*
2. *For all $A \in \mathcal{C}, B \notin \mathcal{C}$, the S wins the n -round i.f.-game on A and B .*

We now state and prove the preservation theorem.

Proposition 26 *Let $\mathcal{C}$ be a class of models. Then $\mathcal{C}$ is FO-definable and closed under strict surjections and reverse strict surjections iff it is definable by an identity free sentence.*

Proof. Let $\mathcal{C}$ be defined by a sentence $\varphi \in \text{FO}$, with $qr(\varphi) = n$. We argue that for all $A \in \mathcal{C}, B \notin \mathcal{C}$, the S wins the n -round i.f.-game on A and B . By the preceding Corollary, this implies that $\mathcal{C}$ is definable by an identity free sentence with the same quantifier rank as φ .

Suppose that the D wins the n -round i.f.-game on some $A \in \mathcal{C}$ and $B \notin \mathcal{C}$. Let $A \equiv_n B$ mean that A and B are equivalent on all FO sentences of quantifier rank $\leq n$. We will

define models $A' \in \mathcal{C}$, $B' \notin \mathcal{C}$ such that $A' \equiv_n B'$, a contradiction. Given a model A , and elements $a_0, a_1 \in A$, we say that a_0 is a copy of a_1 iff the permutation $(a_0 a_1)$, permuting a_0 and a_1 , is an automorphism of A . Let $A' [B']$ be the extension of $A [B]$ obtained by adding $n - 1$ copies of every element of the structure. (For example, if A were the graph with universe $\{a_1, b_1\}$, and edge relation $E^A = \{(a_1, b_1)\}$, then A' would be the graph with universe $\{a_1, \dots, a_n; b_1, \dots, b_n\}$ and edge relation, $E^{A'} = \{(a_i, b_j) \mid i, j \leq n\}$.) There are obvious strict surjections from A' and B' to A and B , respectively, and hence $A' \in \mathcal{C}$, $B' \notin \mathcal{C}$.

We now show that the D's winning strategy for the i.f.-game on A and B can be easily adapted to provide a winning strategy for the standard n -round Ehrenfeucht-Fraïssé game on A' and B' , demonstrating that $A' \equiv^n B'$. The basic idea is that the presence of n copies of every element enables the D to modify her strategy from the i.f.-game as follows. Whenever she is in a position where she would have played on a previously pebbled element, she instead plays on a vacant copy of that element. That is, she maintains the condition that if, through round m , pebbles have been placed on $\bar{a}' = (a'_1, \dots, a'_m) \subseteq A'$, $\bar{b}' = (b'_0, \dots, b'_m) \subseteq B'$, then she would win the i.f.-game on A and B with the pebbles on $\bar{a} = (a_1, \dots, a_m)$, $\bar{b} = (b_1, \dots, b_m)$, where each $a'_i [b'_i]$ is a copy of $a_i [b_i]$. Suppose that she maintains this condition through m rounds, $m < n$, and that in round $m + 1$, the S plays on some unpebbled $a'_{m+1} \in A'$. The D then plays on an unpebbled copy $b'_{m+1} \in B'$ of some $b_{m+1} \in B$ such that $(\bar{a} * a_{m+1}, \bar{b} * b_{m+1})$ is a winning position for the D in the i.f.-game on A and B . Similarly if the S plays on B' . ■

The next corollary follows immediately.

Corollary 5 *For all classes $\mathcal{C}$, if $\mathcal{C} \in \text{FO} \cap \text{HOM}$, then it is definable by an identity free sentence.*

The above argument demonstrates the existence of an equivalent identity free sentence θ with the same quantifier rank as φ . More generally, the idea of the above proof yields significant information about the desired identity free sentences. For example, if φ is in prenex form, we can establish that there is an identity free θ with the same quantifier prefix. The next corollary was needed in the proof of Proposition 21, above. It can be

proved using the positive Ehrenfeucht-Fraïssé game from [22], described in Section 2, and the *positive i.f.-game*. $\mathcal{C}$ is *non-trivial* iff it is neither empty nor the class of all structures.

Corollary 6 *Let $\mathcal{C}$ be a non-trivial class definable in $\text{FO}(+, \neq)$ and closed under strict surjections and reverse strict surjections. Then $\mathcal{C}$ is defined by an identity free sentence of $\text{FO}(+)$. [If $\mathcal{C} = \mathcal{F}$, (resp. $\emptyset$), then it is defined by the sentence $\exists x(x = x)$, (resp. $\exists x(x \neq x)$)].*

4.2 Generalized preservation theorems for HOM and MON

In this section we discuss generalized preservation properties for the classes HOM and MON. We pose various open problems in the same spirit as the question of whether $\text{FO} \cap \text{EXT}$ is contained in the existential fragment of some stronger logic, from Chapter 1. One purpose of this investigation is to try to understand better the connection between definability in FO and in ‘resource bounded’ fragments of $L_{\infty\omega}^\omega$. For example, the fact that $\text{FO} \cap \text{Datalog}(\neg) \neq \text{FO}(\exists)$ means that there are classes in $\text{EXT} - \text{FO}(\exists)$ that are definable in two different extensions of $\text{FO}(\exists)$ that are obtained by adding different ‘orthogonal’ mechanisms to the language— $\forall$ to get FO, and recursion to get $\text{Datalog}(\neg)$.

The languages $L_{\infty\omega}^\omega(\exists, +)$ and $L_{\infty\omega}^\omega(+, \neq)$ are defined in the obvious manner. We view Datalog as $\text{LFP}(\exists, +)$, and $\text{Datalog}(\neg)$ as $\text{LFP}(\exists)$. Positive LFP, $\text{LFP}(+, \neq)$, extends Datalog in allowing any $\text{FO}(+, \neq)$ formula to occur in the body of a clause. The intensional predicates are computed in the obvious manner. The following proposition is proved by arguments analagous to those in the proof of Proposition 1.

Proposition 27 1. $\text{Datalog} \subseteq L_{\infty\omega}^\omega(\exists, +) \subseteq \text{HOM}$

2. $\text{LFP}(+, \neq) \subseteq L_{\infty\omega}^\omega(+, \neq) \subseteq \text{MON}$

In strict analogy to open problems posed above for EXT, we ask the following questions.

1. Is $\text{FO} \cap \text{HOM} \subseteq L_{\infty\omega}^\omega(\exists, +)$?
2. Is $\text{FO} \cap \text{HOM} \subseteq \text{Datalog}$?
3. Is $\text{FO} \cap \text{MON} \subseteq L_{\infty\omega}^\omega(+, \neq)$?
4. Is $\text{FO} \cap \text{MON} \subseteq \text{LFP}(+, \neq)$?

Ajtai and Gurevich [3] showed that $\text{FO} \cap \text{Datalog} = \text{FO}(\exists, +)$. Consequently, a positive answer to the second question would imply the truth of the Homomorphism preservation theorem over $\mathcal{F}$. Observe also that their result contrasts with the known fact that $\text{FO} \cap \text{Datalog}(\neg) \neq \text{FO}(\exists)$. We now show that $\text{FO} \cap \text{LFP}(+) \neq \text{FO}(+, \neq)$. The class that we define is based on a construction from Stolboushkin [22], which gives a simple proof that Lyndon's lemma fails finitely.

Proposition 28 *There is a FO-definable class $\mathcal{C} \in \text{MON}$ such that*

1. $\mathcal{C}$ is definable in $\text{LFP}(+)$.
2. $\mathcal{C}$ is not definable in $\text{FO}(+, \neq)$.

Therefore, $\text{FO} \cap \text{LFP}(+) \neq \text{FO}(+, \neq)$.

Proof. We define a class $\mathcal{C}$ which includes a class of structures that we call P, Q -orders, based on the “grids” from [22]. A P, Q -order, A , consists of two disjoint linear orders (with some additional relations) of sets P^A and Q^A , where P and Q are monadic predicates in the signature. We verify directly that $\mathcal{C}$ is definable in FO and in $\text{LFP}(+)$. Using a modification of Stolboushkin's idea, and the appropriate Ehrenfeucht-Fraisse game for $\text{FO}(+, \neq)$, we then show that $\mathcal{C}$ is not $\text{FO}(+, \neq)$ -definable.

Definition 6 *Let $\sigma = \{Px, Qx, x < y, Sxy, Txy, c, d\}$. A is a P, Q -order iff*

1. Every element $a \in A$ is in exactly one of the relations P and Q .
2. The relation $x < y$ linearly orders the submodels P^A and Q^A , and for all $a \in P^A$, $b \in Q^A$, $A \models \neg(a < b \vee b < a)$.
3. Sxy is the successor relation on the submodels P^A and Q^A , and for all $a \in P^A$, $b \in Q^A$, $A \models \neg(Sab \vee Sba)$.
4. c is the $<$ -minimal element in P^A , and d is the $<$ -minimal element in Q^A .
5. $\forall xy(Txy \rightarrow (Px \wedge Qy))$
6. $Tcd \wedge \forall x_1 x_2 x_3 x_4 ((Px_1 \wedge Px_2 \wedge Qx_3 \wedge Qx_4 \wedge Tx_1 x_3) \rightarrow \{(x_3 < x_4 \rightarrow Tx_1 x_4) \wedge ((Sx_1 x_2 \wedge Sx_3 x_4) \rightarrow Tx_2 x_4)\})$

$$7. \exists uvw(u < v \wedge Tuw \wedge Tvw \wedge \forall x(x < w \rightarrow \neg Tux))$$

It is obvious that the class of finite P, Q -orders is defined by some $\varphi \in \text{FO}$. We now define a sentence $\theta \in \text{FO}(\exists, +)$, and let $\mathcal{C} = \text{Mod}_f(\varphi \vee \theta)$. We define $\theta_i, 0 \leq i \leq 5$, as follows, and let $\theta = \bigvee_{i \leq 5} \theta_i$.

$$\theta_0 = \exists x(Px \wedge Qx) \vee Qc \vee Pd$$

$$\theta_1 = \exists xy(Txy \wedge (Py \vee Qx))$$

$$\theta_2 = \exists xy(x < y \wedge ((Px \wedge Qy) \vee (Qx \wedge Py)))$$

$$\theta_3 = \exists xy(x < y \wedge y < x)$$

$$\theta_4 = \exists xyz(x < y \wedge y < z \wedge z < x)$$

$$\theta_5 = \exists xyz(Sxy \wedge (x = y \vee y < x \vee (x < y \wedge y < z) \vee (Px \wedge Qy) \vee (Py \wedge Qx)))$$

We claim that $\mathcal{C}$ is monotone. First, suppose that $A \in \mathcal{C}$ satisfies θ . Since θ is positive, every enrichment of A also satisfies θ . Now suppose that A is a P, Q -order. By considering expansions of the different relations in the signature, it is easy to see that every enrichment of A is also in $\mathcal{C}$. For example, if B is obtained from A by expanding P^A , then there is a $b \in B$ such that $B \models Pb \wedge Qb$. Thus, $B \models \theta$. Similar considerations show that expanding Qx, Sxy , or $x < y$ also produces a model of θ . Finally, there are enrichments B of A such that only the relation Txy is expanded and $B \not\models \theta$, but it is easy to verify that any such enrichment is a P, Q -order.

We now show that $\mathcal{C}$ is LFP(+)-definable. (Observe that, by Proposition 27, this also provides an alternative proof that $\mathcal{C}$ is monotone.) We define in LFP(+) a relation Rxy such that, over the class of P, Q -orders, Rab holds iff $Pa \wedge Qb \wedge \text{height}(b) < \text{height}(a)$. Here, $\text{height}(x)$ is the height of x in the linear order. Rxy is like Txy , except that it consists of edges from P 's to Q 's 'pointing' in the other direction. Let Rxy be the relation computed by the clause, $Rxy \leftarrow (Px \wedge y = d) \vee \exists wz(Rwz \wedge Swx \wedge Szy)$.

The following sentences are components of the LFP(+) program to be defined below. Roughly, any model that satisfies their conjunction is either a model of θ or looks very much like a P, Q -order.

Let $\psi_0 = \forall x(Px \vee Qx) \wedge \forall xy((Qx \vee Qy \vee x < y \vee y < x \vee x = y) \wedge (Px \vee Py \vee x < y \vee y < x \vee x = y))$. The second conjunct says that every pair of elements in $P^A[Q^A]$ is

connected by the relation $x < y$. For assume that no element of A is in both P and Q . (Otherwise, we simply have that $A \models \theta$.) Then ψ_0 says that if x and y are distinct and in P^A , i.e. not in Q^A , then there is an $<$ -edge connecting them. Likewise for x and y in Q^A .

Let $\psi_1 = \forall xy \exists z (Qx \vee Qy \vee x = y \vee y < x \vee Sxz) \wedge \forall xy \exists z (Px \vee Py \vee x = y \vee y < x \vee Sxz)$. This sentence says that every element that is not $<$ -maximal has a ‘successor’. For example, suppose that x and y are in $P^A - Q^A$, such that $x < y$. In particular, x is not maximal. Then, by the first conjunct, x has a successor.

Let $\psi_2 = \forall xy (Qx \vee Py \vee (Rxy \vee Txy)) \wedge \exists vw (Tvw \wedge Rvw)$. This says that Txy behaves in the appropriate way, by using the (inductively defined) relation Rxy as its ‘complement’.

The following LFP(+) program defines $\mathcal{C}$.

$$Rxy \longleftarrow (Px \wedge y = d) \vee \exists wz (Rwz \wedge Swx \wedge Szy)$$

$$B \longleftarrow \theta \vee (\psi_0 \wedge \psi_1 \wedge \psi_2)$$

Here, B is the distinguished Boolean predicate. Suppose that $A \in \mathcal{C}$. Either $A \models \theta$, in which case A is obviously in the class computed by the above program, or it is a P, Q -order. In the latter case, it is easy to verify that ψ_0, ψ_1 , and ψ_2 are each satisfied in A , and hence the value of the Boolean predicate B , on A , is true. To establish the containment in the other direction, suppose that A is in the class computed by the program. If $A \not\models \theta$, but $A \models \psi_0 \wedge \psi_1 \wedge \psi_2$, then it is straightforward to verify that A is a P, Q -order.

The definition of the preceding program exploits the following idea. Since every element in a P, Q -order, A , is either in P or Q , and since $x < y$ linearly orders P^A and Q^A , negation can essentially be expressed in a positive way. For example, $A \models \neg Pa$ iff $A \models Qa$. Also, roughly, $\neg x < y$ iff $x = y \vee y < x$. We also observe that, by making minor changes, one can eliminate the symbols c, d , and Sxy from the vocabulary. For example, Sxy is actually positively definable over P, Q -orders.

It remains to show that $\mathcal{C}$ is not definable by a FO(+, $\neq$) sentence. We adapt a proof of Stolboushkin’s, from [22], in which the *positive n -round Ehrenfeucht-Fraisse game* is introduced. The rules of this game are identical to those of the standard n -round Ehrenfeucht-Fraisse game, except that the D wins iff the function, $f(x)$, from A to B that sends each pebble α_i to pebble β_i is a partial injective homomorphism between the induced submodels of A and B . That is, the D must maintain the condition that for all k -ary relations, R ,

and all k -tuples of pebbles $\bar{\alpha}'$, if $A \models R(\bar{\alpha}')$, then $B \models R(f(\bar{\alpha}'))$. We define $A \preceq^{+,n} B$ to mean that for all sentences $\varphi \in \text{FO}(+, \neq)$, with $qr(\varphi) \leq n$, if $A \models \varphi$, then $B \models \varphi$.

Proposition 29 (Stolboushkin[22]) *For all A and B , and all n , the following conditions are equivalent.*

1. $A \preceq^{+,n} B$.
2. *The D has a winning strategy in the positive n -round game on A and B .*

To prove that $\mathcal{C} \notin \text{FO}(+, \neq)$, it suffices to show that for all n , there are $A \in \mathcal{C}, B \notin \mathcal{C}$ such that $A \preceq^{+,n} B$. We define A to be a P, Q -order, and B to be an impoverishment of A , obtained by removing a single T -edge. Let the universe of A be the set of ordered pairs $\{(w, h) \mid w \in \{0, 1\} \text{ and } 0 \leq h \leq 2^{n+2}\}$, with P^A [resp. Q^A] the set of elements of the form $(0, h)$ [resp. $(1, h)$]. $A \models (w, h) < (w', h')$ iff $w = w'$ and $h < h'$. The relation $x < y^A$ uniquely determines the interpretation of Sxy ; $c^A = (0, 0)$, $d^A = (1, 0)$. Finally, $T^A = \{((w, h), (w', h')) \mid w = 0, w' = 1, \text{ and } h \leq h'\} \cup \{((0, 2^{n+2}), (1, 2^{n+2} - 1))\}$. B is identical to A , except that $T^B = T^A - \{((0, 2^{n+1}), (1, 2^{n+1}))\}$. It is easy to verify that $A \in \mathcal{C}$, and that $B \notin \mathcal{C}$.

We now describe a winning strategy for the D in the n -round positive game on A and B . In each round, she either plays on the same element, on the other structure, as the S, or she plays on its S -predecessor or successor. Roughly, on P or far from the midpoints, $(0, 2^{n+1})$ and $(1, 2^{n+1})$, the D copies the S's moves, and near the midpoints and in Q , she plays so that the β pebbles, on B , are shifted one higher than their α counterparts. In any round m , if S plays on $(0, h)$, in either A or B , then the D plays on $(0, h)$ in the other model. In round 1, if the S plays on $(1, h)$ and $d(h, 2^{n+1}) = |h - 2^{n+1}| > 2^n$, then the D also plays on $(1, h)$. Otherwise, the D plays a 'shift' so that $\alpha_1 = (1, h')$ and $\beta_1 = (1, h' + 1)$, with $h' = h$ or $h - 1$, depending on whether or not the S played on A . In this case, we say that the pebble pair is shifted. In each round $m + 1$, let I_{m+1} be the smallest interval $[s_{m+1}, t_{m+1}]$, $0 < s_{m+1} \leq t_{m+1} < 2^{n+2}$ that contains 2^{n+1} and the h -component, h_l , of any shifted pebble $\beta_l, l \leq m$. If no pebbles have been shifted through round m , then I_{m+1} is the degenerate interval $[2^{n+1}, 2^{n+1}]$. In round $m + 1$, the D copies the S's move, $\alpha_{m+1} = (1, h_{m+1})$ if the distance from h_{m+1} to the interval I_{m+1} is greater than $2^{(n+1)-m}$.

Otherwise, the D plays a shift, exactly as described for the first round. It is easy to see that this provides the D with a winning strategy. ■

Chapter 5

Modal logic over finite structures

In this chapter, we discuss the finite model theory of the language of propositional modal logic, PM. Modal logic has been studied extensively in connection with philosophical logic. More recently, connections have emerged between modal logic and computational linguistics and certain areas of computer science. Below we will be interested in the ‘classical model theory’ of modal logic, an approach taken by van Benthem and others. For example, PM satisfies certain preservation theorems that are analogous to classical theorems for FO. We show that, in contrast to more expressive logics, PM remains well-behaved over $\mathcal{F}$ as various classical results remain true over the class of finite models.

In order to make this chapter self-contained, we briefly describe the syntax and semantics of PM. Most of this material is well-known, and more detailed descriptions can be found in many places, (e.g. see [8]). The syntax of PM is obtained from that of simple sentential logic by adding the two modal operators $\Box\varphi$, *necessarily* φ , and $\Diamond\varphi$, *possibly* φ . Over a signature of *proposition symbols*, $\sigma = \{p_1, \dots, p_k\}$, the class of sentences of $\text{PM}(\sigma)$ is the smallest class containing each atomic sentence p_i and closed under negation, conjunction, disjunction, and the operators $\Box$ and $\Diamond$. We will always assume that the signature is finite and non-empty. A (Kripke) model of $\text{PM}(\sigma)$ is a directed graph A with additional unary predicates $\{P_1, \dots, P_k\}$, corresponding to each proposition symbol. The edge relation Rxy is often called the ‘accessibility relation’, and we will say that b is accessible from a just in case Rab .

Definition 7 *Satisfaction for sentences of PM at a node (or world) is defined inductively.*

1. $(A, a) \models^{PM} p_i$ iff $A \models P_i(a)$.
2. The Boolean operations are given their standard interpretations.
3. For the modal operator necessarily, $(A, a) \models^{PM} \Box q$ iff for all $b \in A$ such that $A \models Rab$, $(A, b) \models^{PM} q$. Possibly is defined dually, $(A, a) \models^{PM} \Diamond q$ iff there is some $b \in A$ such that $A \models Rab$ and $(A, b) \models^{PM} q$.

This semantics suggest a natural interpretation of PM into FO. In fact, by reusing variables we can translate PM into the language L^2 . Since sentences of PM are evaluated at a node of the Kripke model, they naturally translate into FO-formulas with one free variable. In order to keep the image of the translation in L^2 , we will simultaneously define two functions, $\mu_0(\varphi)$ and $\mu_1(\varphi)$ such that (i) $\mu_d(\varphi)$ contains x_d free; and (ii) for all $\varphi \in \text{PM}$, $\mu_1(\varphi)$ is obtainable from $\mu_0(\varphi)$ by replacing every occurrence of x_0 by x_1 , and vice-versa. The functions $\mu_d(\varphi)$ from sentences of PM to formulas of L^2 are defined inductively as follows:

$$\begin{aligned}
\mu_d(p_j) &= P_j(x_d) \\
\mu_d(q_1 \wedge q_2) &= \mu_d(q_1) \wedge \mu_d(q_2) \\
\mu_d(\neg q) &= \neg \mu_d(q) \\
\mu_d(\Box q) &= \forall x_{1-d} (R x_d x_{1-d} \rightarrow \mu_{1-d}(q)) \\
\mu_d(\Diamond q) &= \exists x_{1-d} (R x_d x_{1-d} \wedge \mu_{1-d}(q))
\end{aligned}$$

To simplify the exposition, we add a single constant c to our FO-signature, to convert each formula with one free variable into a sentence. Let $\mu(\varphi)$ be the function from PM to L^2 such that for all $\varphi \in \text{PM}$, $\mu(\varphi)$ is obtained from $\mu_0(\varphi)$ by replacing each *free* occurrence of x_0 by c . Then each model is viewed as having a distinguished node, at which modal sentences are evaluated. Let FO^M , the *modal fragment of first order logic*, be the image of PM under the mapping $\mu(\varphi)$.

In his dissertation [7], van Benthem gave an algebraic characterization of FO-definable classes that are definable by a modal sentence. He introduced the following important notion.

Definition 8 *Given two models A and B (with distinguished nodes c^A and c^B), a bisimulation between A and B , is a binary relation, $\sim$, contained in $A \times B$, such that*

1. $c^A \sim c^B$
2. *For all a, b such that $a \sim b$, if $A \models Raa'[B \models Rbb']$, then there is a $b' \in B[a' \in A]$ such that $a' \sim b'$*
3. *For all a, b such that $a \sim b$, and all P_j , $A \models P_j(a)$ iff $B \models P_j(b)$.*

We say that A bisimulates with B iff there is a bisimulation between the two models. We also write $(A, a) \sim (B, b)$ if there is a bisimulation $\sim$ between A and B such that $a \sim b$.

Bisimulation is an equivalence relation on structures, which can be seen as a modified, weak kind of partial isomorphism. It is easy to see that if there is a bisimulation between a pair of models, then they satisfy the same modal sentences.

Van Benthem proved the following preservation theorem: a FO-definable class of models is closed under bisimulations iff it can be defined by a sentence in FO^M . Below we prove that this result remains true over $\mathcal{F}$. We then show that an ‘existential’ preservation theorem, due to van Benthem and Visser (see [5]), also holds over the class of finite structures. Finally, we give an alternative proof, which does not use the compactness theorem, of Andreka, van Benthem, and Nemeti’s result [5] establishing the modal analog of the Craig interpolation theorem.

5.1 Background

In this section, we present background information needed for the proofs of the main results that appear in Section 2. Our development of this material closely parallels analogous results for both FO and for the various finite variable logics. We first define an infinite game to characterize full bisimulation. We then introduce finite versions of the game, and the notion of ‘ n -bisimulation’, and determine their connection to modal definability.

In the (*eternal*) *modal Ehrenfeucht-Fraisse game* the Spoiler and the Duplicator play a modified two pebble Ehrenfeucht-Fraisse game, with pebble pairs $(\alpha_0, \beta_0), (\alpha_1, \beta_1)$. At the start of the game, pebbles α_0 and β_0 are on c^A and c^B , respectively. In round 1, the S

either places α_1 on some element of A such that $A \models R\alpha_0\alpha_1$ or places β_1 on some element of B such that $B \models R\beta_0\beta_1$. The D then does the same on the other structure. In each subsequent round $n + 1$, the Spoiler chooses a pair (α_i, β_i) of pebbles, already in play, and replays either α_i on A such that $A \models R\alpha_{1-i}\alpha_i$ or β_i on B such that $B \models R\beta_{1-i}\beta_i$. The D then plays the other pebble on the other structure in accordance with the same restriction. Each player loses immediately if he or she cannot make a legal move. The Spoiler wins at round n if there is P_m such that $A \models P_m\alpha_i$ iff $B \not\models P_m\beta_i$. (Observe that the Duplicator does not have to play so that the partial mapping from A to B induced by the pebbles is a partial isomorphism—e.g. in some round, she could play β_1 on the same element as β_0 in B , even if S had not just played α_1 on α_0 in A . This is because sentences of FO^M do not contain equality.) The Duplicator wins the game, just in case, in every round the Spoiler does not win. The following proposition is straightforward.

Proposition 30 *For all A and B of signature σ , the following conditions are equivalent:*

1. *There is a bisimulation between A and B .*
2. *The Duplicator has a winning strategy in the modal game on A on B .*

We turn our attention now to modal definability.

Definition 9 *The quantifier rank of a formula, $qr(\varphi)$, is defined inductively.*

1. $qr(P_i) = 0$
2. $qr(\neg\varphi) = qr(\varphi)$
3. $qr(\varphi_1 \wedge \varphi_2) = qr(\varphi_1 \vee \varphi_2) = \max(qr(\varphi_1), qr(\varphi_2))$
4. $qr(\Diamond\varphi) = qr(\Box\varphi) = qr(\varphi) + 1$

Of course, there are no genuine quantifiers in PM; the choice of terminology emphasizes the connection between PM and FO. In particular, for all $\varphi \in \text{PM}$, $qr(\varphi)$ equals the quantifier rank of the FO-sentence, $\mu(\varphi)$. Let PM^n be the set of sentences of quantifier rank $\leq n$. Given a model A , the PM^n -theory of A is then the set of sentences, of quantifier rank $\leq n$, satisfied by A .

Lemma 7 *Let σ be a fixed signature.*

1. *For all n , up to logical equivalence, there are finitely many sentences of PM^n .*
2. *There is a recursive function $f(n)$ that generates a (finite) list of all sentences, up to logical equivalence, of quantifier rank $\leq n$.*
3. *For all A , the PM^n -theory of A is finitely axiomatizable.*

Proof. We prove Part 1 by induction on n . The case $n = 0$ is obvious. For $n + 1$, observe that every sentence of quantifier rank $\leq n + 1$ is a Boolean combination of sentences of the form $\Diamond\theta$, with $qr(\theta) \leq n$. Parts 2 and 3 follow easily from Part 1. ■

Definition 10 *We say that there is an n -bisimulation between A and B , written $A \sim_n B$, iff there is a sequence of relations $\sim_0, \dots, \sim_n$, each on $A \times B$, such that*

1. $c^A \sim_0 c^B$
2. *For all $m < n$, if $a \sim_m b$, and $A \models Raa'$ then there is a $b' \in B$ such that $B \models Rbb'$ and $a' \sim_{m+1} b'$ [and vice-versa].*
3. *For all $m \leq n$, if $a \sim_m b$, then for all P_j , $A \models P_j(a)$ iff $B \models P_j(b)$.*

Intuitively, $A \sim_n B$ means that A and B bisimulate ‘up to depth n ’. Observe that $A \sim B$ implies $A \sim_n B$, for all n , and that $\sim_n$ also defines an equivalence relation on classes of structures. By fixing a bound on the number of rounds in a game, we get the n -round modal Ehrenfeucht-Fraïssé game. Then the following proposition can be proved by straightforward modification of standard results connecting Ehrenfeucht-Fraïssé games to logical expressibility.

Proposition 31 *For all n , and A and B over some σ , the following conditions are equivalent:*

1. *There is an n -bisimulation between A and B .*
2. *The Duplicator has a winning strategy in the n -round modal game on A on B .*
3. *For all modal formulas θ of quantifier rank $\leq n$, $A \models \theta$ iff $B \models \theta$.*

The next proposition follows easily from Proposition 31 and Lemma 7.

Proposition 32 *Let $\mathcal{C}$ be any class of models, closed under isomorphism. Let $\mathcal{C}'$ be any subclass of $\mathcal{C}$, also closed under isomorphism. Then, for all n , the following conditions are equivalent:*

1. *For all $A \in \mathcal{C}'$, $B \in \mathcal{C} - \mathcal{C}'$, $A \not\sim_n B$.*
2. *For all $A \in \mathcal{C}'$, $B \in \mathcal{C} - \mathcal{C}'$, the S wins the n -round modal game on A and B .*
3. *There is a modal sentence of quantifier rank $\leq n$ that defines the class $\mathcal{C}'$ over $\mathcal{C}$.*

Bisimulation and n -bisimulation are rather weak equivalence relations, in the sense that they determine relatively large equivalence classes. In other words, for every model A there are many other models with the same modal theory. Our proofs will exploit this feature repeatedly.

We fix the following terminology.

Definition 11 *The children of a in A are those b such that $A \models Rab$. We say that b is a descendent of a iff there is a directed path from a to b . For all n , b is an n -descendent of a if there is a path of length $\leq n$ from a to b . The family of a , written F^a is the submodel of A with universe $\{a\} \cup \{b \mid b \text{ is a descendent of } a\}$. For all a and b , we say that a and b are disjoint iff $F_a \cap F_b = \emptyset$.*

The r -neighborhood of a point a , denoted $N_r(a)$, is defined inductively. $N_0(a)$ is the submodel of A with universe $\{a\}$. For all $r + 1$, $b \in N_{r+1}(a)$ iff $b \in N_r(a)$ or there is an $a' \in N_r(a)$ such that $A \models Ra'b \vee Rba'$. An r -tree is a directed tree rooted at c of height $\leq r$. An r -pseudotree is a model such that $N_r(c)$ is a tree such that all distinct pairs of its leaves are disjoint, as defined above.

We now describe certain operations on models that produce either bisimilar or n -bisimilar models. For A and a , we say that A' is obtained from A by adding a copy of the family of a iff A' is the extension of A with universe the disjoint union of A and of F^a such that for all $a \in A$ and $a'_1 \in F^a$, the ‘copy’ of F^a in A' , $A' \models Raa'_1[Ra'_1a]$ iff $A \models Raa_1[Ra_1a]$, where a'_1 is the copy of $a_1 \in F^a$. The binary relation $\{(a, a') \mid a \in A, a' \in A' \text{ and } a = a' \text{ or } a' \text{ is a copy of } a\}$ witnesses that $A \sim A'$.

Another concept from modal logic is that of *unraveling* a structure to produce another structure with which it bisimulates. Before defining this notion, we give a simple illustration. Let A be the graph on one vertex with a loop, and let A' be the directed chain on $c = 0, 1, \dots, n$ such that for all $m < n$, $A' \models Rm, m+1$ and $A' \models Rnn$. We can view A' as having been obtained from A by unraveling, or unwinding, the loop n times. The set $A \times A'$ is itself a bisimulation between A and A' . In general, any model A can be n -unraveled, so that the n -descendents of c form an n -tree. By ω -unraveling F^c in A we obtain a (possibly infinite) tree. Every unraveling of A bisimulates with A .

To simplify the definition, we assume that every element of A is a descendent of c , i.e. $A = F^c$. The n -unraveling of A will be an n -pseudotree, which we call A' . We first describe the tree portion of A' , that is, $N_n(c^{A'})$. The root of the tree will be c itself and, for each path in A of length $s \leq n$ starting at c , there is a node of height s in the tree. Thus, each node is indexed by a path $\bar{a} = (c = a_0, a_1, \dots, a_s)$ [that is, a sequence of length $s+1$] such that for all $q < s$, $A \models Ra_q a_{q+1}$. Given a path $\bar{a}$ and an element $a' \in A$, let $\bar{a} * a'$ denote their concatenation, that is, $(a_0, a_1, \dots, a_s, a')$. For each such $\bar{a}$, $A' \models P_j(\bar{a})$ iff $A \models P_j(a_s)$. In A' , there is an edge from $\bar{a}$ to $\bar{a}_1$ iff $\bar{a}_1 = \bar{a} * a'$, for some $a' \in A$. This completes the description of the n -tree which is the n -neighborhood of c in A' . We now attach copies of families to the leaves of this tree of height n , to obtain the n -pseudotree A' . That is, at each node $\bar{a} = (c = a_0, a_1, \dots, a_n)$, we attach a copy of F^{a_n} , identifying the elements $\bar{a}$ and a_n . There may be many copies of any family, but each pair of families is disjoint. It is now easy to construct a bisimulation between A and A' . The ω -unraveling is defined similarly, except that no families are attached to any nodes.

We collect together some easy to verify facts for later use.

Proposition 33 *For all A , 1. $A \sim F_A^c$. 2. A bisimulates with a tree rooted at c , its ω -unraveling. 3. A bisimulates with an n -pseudotree, its n -unraveling. 4. A n -bisimulates with an n -tree, a submodel of its n -unraveling. 5. Over a fixed signature σ , there is a recursive function $f(x)$ such that for all modal sentences φ of quantifier rank $\leq n$, if φ is satisfiable, by a finite or infinite model, then it is satisfiable by an n -tree of cardinality $\leq f(n)$. 6. For all finite A , the modal theory of A is finitely axiomatizable iff F^c is acyclic.*

Proof. We provide proofs of Facts 5 and 6. From Fact 4 and Proposition 31, it is clear that

for all $\varphi \in \text{PM}^n$, φ is satisfiable iff it is satisfied by an n -tree. Given a fixed finite signature σ , we now define an effective procedure that maps each natural number n into a finite set of n -trees $\mathcal{T}^n$ such that for all $\varphi \in \text{PM}(\sigma)$ of quantifier rank $\leq n$, if φ is satisfiable, then it is satisfied in some $A \in \mathcal{T}^n$. This will suffice to establish the claim. The sets $\mathcal{T}^n$ are defined inductively. $\mathcal{T}^1$ contains every model, up to isomorphism, with exactly one element, and has cardinality $= 2^{|\sigma|}$. For $n + 1$, $A \in \mathcal{T}^{n+1}$ iff $A \in \mathcal{T}^n$ or A is an n -tree rooted at c with children $a_1, \dots, a_k$ satisfying the following properties: (i) for all $i \leq k$, the family F^{a_i} is isomorphic to some tree $B \in \mathcal{T}^n$; and (ii) for all $i \neq j \leq k$, $F^{a_i} \not\cong F^{a_j}$. It is easy to verify both that there is a recursive bound on the size of models in each $\mathcal{T}^n$ and that every n -tree bisimulates with an n -tree in $\mathcal{T}^n$. This establishes Fact 5.

We now prove Fact 6. Suppose that F^c is acyclic. We show, by induction on the height n of F^c , that A is axiomatized by a sentence of quantifier rank $= n + 1$. For $n = 0$, let $\theta = (\bigwedge_{P \in \tau} P \wedge \bigwedge_{Q \in \sigma - \tau} \neg Q) \wedge (\neg \Diamond P' \wedge \Box P')$, where τ is the set of proposition symbols satisfied at c , and P' is any proposition symbol in σ . For $n \geq 1$, and each child a_i of c , let θ_i be a sentence that axiomatizes the family F^{a_i} . Then let $\theta = (\bigwedge_{P \in \tau} P \wedge \bigwedge_{Q \in \sigma - \tau} \neg Q) \wedge (\bigwedge_i \Diamond \theta_i) \wedge (\Box \bigvee_i \theta_i)$. It is clear that θ axiomatizes the modal theory of A . In the other direction, let A be such that F^c contains a cycle, and let θ be a modal sentence of quantifier rank n . Let B be an n -tree that verifies θ . It is easy to show that there is a modal sentence, ψ , of quantifier rank $= n + 1$ true in A but not in B . For example, let $\psi = \Diamond(\dots \Diamond(P \vee \neg P) \dots)$ contain a string of $n + 1$ $\Diamond$'s, for any $P \in \sigma$. Therefore the modal theory of A is not axiomatized by any sentence of quantifier rank n , and hence is not finitely axiomatizable. ■

Observe that Fact 5 implies some well-known results. One, a modal formula is satisfiable iff it is satisfiable by a finite Kripke model. Two, it is decidable whether a formula is satisfiable, both over the class of all structures and over $\mathcal{F}$.

5.2 Preservation theorems

In this section, we show that two modal preservation theorems remain valid over the class $\mathcal{F}$. The arguments do not use finiteness in any essential way; therefore they also give alternative proofs of the theorems in the general case that do not rely on the Compactness

theorem. Finally, we show how these methods can be used to reprove the modal version of the Craig interpolation theorem without employing compactness.

Proposition 34 *The bisimulation preservation theorem for modal sentences remains true in the finite case. That is, a class $\mathcal{C}$ is FO-definable and closed under bisimulations iff it is definable by a modal sentence.*

Proof. Let $\mathcal{C}$ be a FO-definable class that is closed under bisimulations. Suppose that $\mathcal{C}$ is not definable by a modal formula. By Proposition 32, this implies that for all n , there are $A \in \mathcal{C}$ and $B \notin \mathcal{C}$ such that $A \sim_n B$. (Of course, since $\mathcal{C}$ is closed under bisimulations, we have that $A \not\sim B$.) We will show that this condition implies that for all n , there are actually $A \in \mathcal{C}$ and $B \notin \mathcal{C}$ such that $A \equiv_n B$. (Recall that $A \equiv_n B$ means that for all $\varphi \in \text{FO}$, with $qr(\varphi) \leq n$, $A \models \varphi$ iff $B \models \varphi$.) This immediately implies that $\mathcal{C}$ is not FO-definable, a contradiction.

More specifically, we show that there is a function $l(x)$ such that, for all n , if $A \sim_{l(n)} B$, then there are A' and B' such that $A \sim A'$, $B \sim B'$ and $A' \equiv_n B'$. By choosing $A \in \mathcal{C}$ and $B \notin \mathcal{C}$, we get $A' \in \mathcal{C}$ and $B' \notin \mathcal{C}$. Given A and B , we find A' and B' by modifying A and B in a sequence of steps, as described in the following lemmas.

Lemma 8 *Let A and B be such that $A \sim_t B$. Then there are t -pseudotrees A' and B' such that $A \sim A'$, $B \sim B'$, and $A' \sim_t B'$.*

Let A' and B' be the t -unravelings of A and B . Then A' and B' are t -pseudotrees such that $A \sim A'$ and $B \sim B'$. By the transitivity of $\sim_t$, this implies that $A' \sim_t B'$.

Lemma 9 *Let A and B be t -pseudotrees such that $A \sim_t B$. Then there are t -pseudotrees A' and B' such that $A \sim A'$, $B \sim B'$, and $N_t(c^{A'}) \cong N_t(c^{B'})$.*

The proof describes an algorithm for modifying the two models in a sequence of steps that yields models with isomorphic t -neighborhoods of c . After each step s , $s \leq t$, we have models A_s and B_s such that $A \sim A_s$ and $B \sim B_s$, and c^{A_s} and c^{B_s} have isomorphic s -neighborhoods. At each step $s + 1$, A_{s+1} [resp. B_{s+1}] is obtained from A_s by adding copies of families of nodes of distance $s + 1$ from c .

Let $\{a_1, \dots, a_l, b_1, \dots, b_m\}$ be the set of the children of c in A and B . The relation $\sim_{t-1}$ induces an equivalence relation on this set such that each equivalence class has at least one member in each of A and B . To obtain A_1 and B_1 with isomorphic 1-neighborhoods of c that bisimulate with A and B , it suffices to add enough copies of families of the c -children a_i and b_j such that each equivalence class has an equal number of members in A_1 and B_1 . For example, renumbering the indices of c -children if necessary, suppose that $\{a_1, \dots, a_i; b_1, \dots, b_j\}$ is one such equivalence class. Also, without loss of generality, assume that $i \leq j$. Then A_1 will contain $j - i$ additional copies of the family F^{a_i} . Let $g_1(x)$ be a bijection between the c -children in A_1 and B_1 such that for all a_i , $(A_1, a_i) \sim_{t-1} (B_1, g_1(a_i))$. By iterating this procedure, at each step $s + 1$, we obtain A_{s+1} and B_{s+1} , and a bijection g_{s+1} between nodes of distance $s + 1$ from c^A and c^B with the following properties. For all nodes a_i in A_s of distance s from c , the bijection g_{s+1} maps the children of a_i to those of $g_s(a_i)$, and for all $a \in \text{dom}(g_{s+1})$, $(A_{s+1}, a) \sim_{t-(s+1)} (B_{s+1}, g_{s+1}(a))$. Finally, we choose A' and B' to be the models A_t and B_t .

Together, these lemmas establish that there are models $A \in \mathcal{C}$ and $B \notin \mathcal{C}$ that look rather similar. In particular, for all t , there are t -pseudotrees $A \in \mathcal{C}$ and $B \notin \mathcal{C}$ such that $N_t(c^A) \cong N_t(c^B)$. Although these models have isomorphic t -neighborhoods of c , we still know nothing about the other part of each model, which might make A and B ‘look very different’ in FO. The final step of the proof takes care of this by using a version of Hanf’s lemma.

Proposition 35 (Hanf [16]) *For each signature σ , there is a function $f(x)$ with the following property. For all n , A and B , if there is a bijection $h : A \mapsto B$ such that for all $a \in A$, $N_{f(n)}(a) \cong N_{f(n)}(h(a))$, (with a and $h(a)$ distinguished), then $A \equiv_n B$.*

Lemma 10 *Let A and B be $(3f(n))$ -pseudotrees with $N_{3f(n)}(c^A) \cong N_{3f(n)}(c^B)$, where $f(x)$ is the Hanf function. Then there are A' and B' such that $A \sim A'$, $B \sim B'$, and $A' \equiv_n B'$.*

Each of A' and B' will be obtained from A and B , respectively, by extending the original model by adding disjoint components in such a way that it will be obvious that A' and B' possess the same $f(n)$ -nbhds. It is clear that extending models in such a way does not

affect bisimulations. Let $A_0 [B_0]$ be the submodel of $A [B]$ with universe $A - N_{f(n)}(c^A)$ [$B - N_{f(n)}(c^B)$]. We define $A' [B']$ to be the disjoint union of A and B_0 [B and A_0]. We've added to A the part of B that may look very different from it, and vice-versa, so that A' will look the same 'locally' as B' . In particular, for example, it is easy to see that $\text{card}(A') = \text{card}(B')$. We now define a bijection between these models in 3 parts. Let $g(x)$ be an isomorphism between $N_{3f(n)}(c^A)$ in A and $N_{3f(n)}(c^B)$ in B . Define $h_1(x)$ to be the bijection between $N_{2f(n)}(c^A)$ and $N_{2f(n)}(c^B)$ that is a restriction of the isomorphism $g(x)$. Let A_1 be the submodel of A' whose universe is those elements of B_0 that are in $N_{2f(n)}(c^B)$ (viewing B_0 here as a submodel of B .) We define B_1 similarly. Let h_2 be the bijection between A_1 and B_1 that is also a restriction of the isomorphism $g(x)$. Let h_3 be the bijection between the remaining pieces of A' and B' that takes the 'A-part' of A' to the 'A-part' of B' , and the B -part of A' to the B -part of B' . It is then easy to verify that $h = h_1 \cup h_2 \cup h_3$ is a bijection from A' to B' that 'preserves $f(n)$ -nbhds'. (This is perhaps easier to see if one draws a picture.) Thus $A' \equiv_n B'$ as desired.

To complete the proof, all that remains is to combine the above results. Suppose that $\mathcal{C}$ is FO-definable, closed under bisimulations, but not definable by a modal formula. Then by Lemmas 8, 9, and 10, for all n , there are $A \in \mathcal{C}$ and $B \notin \mathcal{C}$ such that $A \equiv_n B$. But this implies that $\mathcal{C}$ is not FO-definable, a contradiction. This proves the proposition. ■

The next preservation theorem that we consider characterizes those sentences whose classes of models are closed under extensions. Before stating the main result, we define some terminology and prove a few preliminary lemmas.

Definition 12 1. A $\Diamond$ -sentence is a modal sentence built up from atomic propositions and negated atomic propositions using $\wedge$, $\vee$, and $\Diamond$.

2. For all A and B , we write $A \sim_\Diamond B$ iff for all $\Diamond$ -sentences φ , if $A \models \varphi$, then $B \models \varphi$.

3. Given a model A , the $\Diamond$ -theory of A is the set of $\Diamond$ -sentences satisfied by A .

Observe that the $\Diamond$ -sentences are precisely those $\varphi \in \text{PM}$ such that $\mu(\varphi)$ is an existential FO sentence. In particular, the class of models of any $\Diamond$ -sentence is closed under extensions.

Lemma 11 Let A be an n -tree, rooted at c .

1. For all $\Diamond$ -sentences, φ , of quantifier rank $\geq n + 1$, $A \not\models \varphi$.
2. The $\Diamond$ -theory of A is axiomatized by a sentence of quantifier rank $= n$.

Proof. Part 1 is obvious, since A does not contain any paths of length $n + 1$. By Lemma 7, let $\theta_1, \dots, \theta_k$ be the set of all $\Diamond$ -sentences of quantifier rank $\leq n$, up to equivalence, satisfied in A . By Part 1, it is clear that $\theta = \bigwedge \theta_i$ axiomatizes the $\Diamond$ -theory of A . ■

Lemma 12 *Given a fixed signature, there is a finite set of n -trees, $\mathcal{T}^n = \{B_1, \dots, B_v\}$ such that for all A , there is a $u \leq v$ such that $A \sim_n B_u$. Furthermore, $\mathcal{T}^n$ can be obtained effectively.*

Proof. This result follows easily from Fact 5 of Proposition 33. Let $\mathcal{T}^n$ be the same set that was defined in the proof of this Fact, such that every satisfiable sentence φ of quantifier rank $\leq n$ is satisfied by some $B \in \mathcal{T}^n$. Let A be any model, and let $\theta_n \in \text{PM}^n$ axiomatize its PM^n -theory, again using Lemma 7. By Fact 5, there is a $B \in \mathcal{T}^n$ such that $B \models \theta_n$. This now implies that $A \sim_n B$. ■

The next result can be viewed as the modal version of the Los-Tarski theorem for finite structures.

Proposition 36 *The existential preservation theorem for modal logic remains true over $\mathcal{F}$. That is, for all φ , if $\text{Mod}_f(\varphi) \in \text{EXT}$, then φ is equivalent to a $\Diamond$ -sentence θ . Moreover, there is an effective procedure for finding the equivalent $\Diamond$ -sentence.*

Proof. Let $\mathcal{C} \in \text{EXT}$ be defined by some sentence φ , with quantifier rank n . Let $\mathcal{C}^n = \mathcal{C} \cap \mathcal{T}^n = \{D_1, \dots, D_k\}$. For each $D_i, i \leq k$, let θ_i axiomatize the $\Diamond$ -theory of D_i . By Lemma 11, $qr(\theta_i) = n$. Let $\theta = \bigvee_{i \leq k} \theta_i$. We claim that φ is equivalent to θ .

First we show that φ implies θ . Suppose that $A \models \varphi$. We claim that there is a $D \in \mathcal{C}^n$ such that $A \sim_n D$. By Lemma 12, there is a $B \in \mathcal{T}^n$ such that $A \sim_n B$. Since $\mathcal{C}$ is closed under $\sim_n$ -equivalence, B must actually be in $\mathcal{C}$, and hence in $\mathcal{C}^n$. Let $D = B$. There is some θ_i , as defined above, such that $D \models \theta_i$. Since $qr(\theta_i) \leq n$, this implies that $A \models \theta_i$, and hence $A \models \theta$.

Now we prove the opposite direction, θ implies φ . Suppose that $A \models \theta$. Then $A \models \theta_i$, for some $i \leq k$. By Lemma 12, there is a $B \in \mathcal{T}^n$ such that $A \sim_n B$. Observe that

$D_i \rightsquigarrow_\diamond B$. We want to show that there is an A' such that (i) $B \sim A'$, and hence $A \sim_n A'$; and (ii) $D_i \subseteq A'$. As $D_i \in \mathcal{C}$, and $\mathcal{C} \in \text{EXT}$, (i) and (ii) imply that $A' \in \mathcal{C}$. Since $\mathcal{C}$ is closed under $\sim_n$ -equivalence, $A \in \mathcal{C}$, as desired. Thus, it suffices to establish the following lemma.

Lemma 13 *Let B, D be trees such that $D \rightsquigarrow_\diamond B$. Then there is a m -tree A' , $m \leq n$, such that $B \sim A'$ and $D \subseteq A'$.*

By induction, on the height n of D . For $n = 0$, it is obvious that $D \subseteq B$, since D is just the single node c^D , and for all predicate symbols p , $D \models p$ iff $B \models p$. Let $A' = B$.

Consider $n > 0$. Let $\{d_1, \dots, d_s\}$ and $\{b_1, \dots, b_t\}$ be the children of c^D and c^B , respectively. We claim that for each d_p , there is a b_r such that $F^{d_p} \rightsquigarrow_\diamond F^{b_r}$. Let ψ , with $qr(\psi) \leq n$, axiomatize the $\diamond$ -theory of F^{d_p} . Then $D \models \diamond\psi$, and therefore $B \models \diamond\psi$. Thus there is a b_r such that $F^{b_r} \models \psi$, as desired.

By adding extra copies of families of the children of c^B to B , if necessary, we get B^0 such that $B \sim B^0$ and there is an injection $h : \{d_1, \dots, d_s\} \longrightarrow \{b_1^0, \dots, b_t^0\}$, $b_j^0 \in B^0$, such that $F^{d_i} \rightsquigarrow_\diamond F^{h(d_i)}$. By the induction hypothesis, each such $F^{h(d_i)}$ bisimulates with an $(n-1)$ -tree, $T^{h(d_i)}$, such that $F^{d_i} \subseteq T^{h(d_i)}$. Let A' be obtained from B^0 by replacing each subtree $F^{h(d_i)} \subseteq B^0$, with the tree $T^{h(d_i)}$. It is easy to see that $B \sim A'$ and $D \subseteq A'$.

This also completes the proof of the proposition. ■

Corollary 7 *For every modal sentence φ , there is a decision procedure that determines whether $\text{Mod}_f(\varphi)$ [$\text{Mod}(\varphi)$] is closed under extensions. Therefore the set of sentences that define such classes is recursive.*

Proof. By the proof of the previous proposition, if $\text{Mod}_f(\varphi)$ [$\text{Mod}(\varphi)$] $\in \text{EXT}$, then it is equivalent to a $\diamond$ -sentence of quantifier rank $\leq qr(\varphi)$. By Lemma 7, one can effectively list, up to logical equivalence, all such sentences, $\psi_1, \dots, \psi_l$. Then it suffices to test the validity of each sentence, $\varphi \leftrightarrow \psi_i$, which is decidable. ■

We now turn to an interpolation theorem, due to Andreka, van Benthem, and Nemeti. It will be convenient to introduce briefly a fragment of second order propositional modal logic, which allows quantification over propositions. We often use $\overline{P}$, etc., as shorthand

for sequences, $(P_1, \dots, P_n)$. We write $\psi(\overline{P})$ to indicate that the set of proposition symbols that occur in ψ equals $\overline{P}$. Also, by $\exists \overline{P} \psi(\overline{P}, \overline{Q})$ we mean the sentence $\exists P_1 \dots \exists P_n \psi(\overline{P}, \overline{Q})$.

Definition 13 *Let $\varphi(\overline{P}, \overline{Q})$ be a sentence of PM, such that $\overline{P} \cap \overline{Q} = \emptyset$. Then $\exists \overline{Q} \varphi(\overline{P}, \overline{Q})$ is a Σ_1^1 modal sentence; for all A , with signature $\sigma = \overline{P}$, $A \models \exists \overline{Q} \varphi(\overline{P}, \overline{Q})$ iff there is a B , an expansion of A with signature $\tau = \overline{P} \cup \overline{Q}$, such that $B \models \varphi(\overline{P}, \overline{Q})$. Π_1^1 modal sentences, of the form $\forall \overline{Q} \varphi(\overline{P}, \overline{Q})$, are defined similarly.*

For all A, B , and n , we write $A \sim_n^{\overline{P}} B$ iff for all sentences φ , $qr(\varphi) \leq n$, that only contain proposition symbols from $\overline{P}$, $A \models \varphi$ iff $B \models \varphi$. Recall that every satisfiable modal sentence is satisfied by a finite model; hence φ implies θ over the class of all models iff φ implies θ over $\mathcal{F}$. By this fact, the truth of the interpolation theorem in the general case immediately yields its truth over $\mathcal{F}$.

Proposition 37 (Andreka, van Benthem, and Nemeti [5]) *Let φ and θ be modal sentences, with signatures σ_φ and σ_θ , such that $\sigma_\varphi \cap \sigma_\theta$ is non-empty. If φ implies θ (over $\mathcal{F}$), then there is a sentence ψ , with $\sigma_\psi \subseteq \sigma_\varphi \cap \sigma_\theta$, such that φ implies ψ and ψ implies θ . Furthermore, $qr(\psi) \leq \max(qr(\varphi), qr(\theta))$.*

Proof. Suppose that $\varphi(\overline{P}, \overline{Q})$ implies $\theta(\overline{P}, \overline{R})$, where $\overline{P}, \overline{Q}$, and $\overline{R}$ are pairwise disjoint sequences of propositions symbols. Equivalently, $\exists \overline{Q} \varphi(\overline{P}, \overline{Q})$ implies $\forall \overline{R} \theta(\overline{P}, \overline{R})$. Thus, we consider models over the signature $\sigma = \overline{P}$. Let $n = \max(qr(\varphi), qr(\theta))$. Recall that, by Lemma 7 or 12, there are only finitely many $\sim_n^{\overline{P}}$ equivalence classes. We claim that it suffices to show that for any $\sim_n^{\overline{P}}$ class $\mathcal{C}$, if there is an $A \in \mathcal{C}$ such that $A \models \exists \overline{Q} \varphi(\overline{P}, \overline{Q})$, then for all $B \in \mathcal{C}$, $B \models \forall \overline{R} \theta(\overline{P}, \overline{R})$. If this is true, for each $\sim_n^{\overline{P}}$ class $\mathcal{C}$ containing an A that satisfies $\exists \overline{Q} \varphi(\overline{P}, \overline{Q})$, let θ_i be a sentence with signature $\overline{P}$, $qr(\theta_i) \leq n$, that defines the class. (Here we use that $\overline{P}$ is non-empty, since no sentence contains no proposition symbols.) Then $\theta = \bigvee \theta_i$ is an interpolant.

Suppose, towards a contradiction, that there are A and B such that $A \sim_n^{\overline{P}} B$, $A \models \exists \overline{Q} \varphi(\overline{P}, \overline{Q})$ and $B \models \exists \overline{R} \neg \theta(\overline{P}, \overline{R})$. Let A' and B' be expansions of A and B such that $A' \models \varphi(\overline{P}, \overline{Q})$ and $B' \models \neg \theta(\overline{P}, \overline{R})$. By Lemma 12, there are n -trees A'' and B'' that are $\sim_n$ -equivalent to A' and B' , respectively. Finally, let A_1 and B_1 be the σ -reducts of A''

and B'' . It is clear that $A_1 \models \exists \overline{Q}\varphi(\overline{P}, \overline{Q})$ and $B_1 \models \exists \overline{R}\neg\theta(\overline{P}, \overline{R})$. We now want to find a D such that $D \models \exists \overline{Q}\varphi(\overline{P}, \overline{Q}) \wedge \exists \overline{R}\neg\theta(\overline{P}, \overline{R})$. This will establish the contradiction.

D is constructed by extending A_1 and B_1 ‘simultaneously’ by iteratively adding copies of families of elements. First we show that for any model M , if M' is obtained from M by adding a copy of a family F^m , for any $m \in M$, then every Σ_1^1 sentence satisfied in M is also satisfied in M' . Suppose that $M \models \exists \overline{P}\psi(\overline{P}, \overline{Q})$. Let N be an expansion of M that verifies the (first-order) modal sentence $\psi(\overline{P}, \overline{Q})$; and let N' be obtained from N by adding a copy of the family of m . It is clear that $N \sim N'$; thus $N' \models \psi(\overline{P}, \overline{Q})$. Since N' is an expansion of M' , $M' \models \exists \overline{P}\psi(\overline{P}, \overline{Q})$, as desired.

We now describe the construction of D . As in the proof of Lemma 9, $\sim_{n-1}$ induces an equivalence relation on the set of children of c^{A_1} and c^{B_1} such that every equivalence class has at least one member in each model. Let A_2 and B_2 be obtained from A_1 and B_1 by adding enough copies of families of these children so that there is a bijection $g_1(x)$ from the children of c^{A_2} to those of c^{B_2} such that for all a_i , $F^{a_i} \sim_{n-1} F^{g_1(a_i)}$. Observe that $N_1(c^{A_2}) \cong N_1(c^{B_2})$. Repeat this procedure at each level $m \leq n$ of the trees, on pairs of subtrees in A_m and B_m determined by the bijection $g_{m-1}(x)$ at the previous level. By the argument of the preceding paragraph, for all m , $A_m \models \exists \overline{Q}\varphi(\overline{P}, \overline{Q})$ and $B_m \models \exists \overline{R}\neg\theta(\overline{P}, \overline{R})$. Furthermore, $N_m(c^{A_{m+1}}) \cong N_m(c^{B_{m+1}})$. This construction yields trees A_{n+1} and B_{n+1} such that $A_1 \sim A_{n+1}$, $B_1 \sim B_{n+1}$, and $A_{n+1} \cong B_{n+1}$. Let $D = A_{n+1}$. ■

Chapter 6

Conclusion

In this dissertation, we have investigated the prospects for the development of positive model theoretic results over the class of finite structures. Regarding preservation theorems, these prospects appear to be somewhat mixed. The positive results that we establish have been for weaker, less expressive, languages, such as propositional modal logic, L^2 , and ‘low’ quantifier prefix classes of FO, and for classes with strong closure conditions, e.g. HOM. In particular, results from Chapter 5 indicate that modal logic remains well-behaved over $\mathcal{F}$. One way to try to extend this work would be to try achieve similar results for stronger levels of the bounded quantifier hierarchy introduced in [5]. It is also unknown whether existential preservation holds for L^2 , both over $\mathcal{F}$ and over the class of all structures (see [4]), and whether homomorphism preservation holds for $L^k, k \geq 3$. We have also given a partial positive answer to what is perhaps the major open question in this area, does the Homomorphism preservation theorem hold over $\mathcal{F}$?

The situation regarding generalized preservation theorems for the class EXT is now fairly well understood. Grohe’s proof that $\text{FO} \cap \text{EXT} \not\subseteq L_{\infty\omega}^\omega(\exists)$ essentially yields a definitive, negative, answer to the questions that appear at the end of Section 1.2. One may still ask, for which quantifier prefix classes, w , is $\text{FO}[w] \cap \text{EXT} \subseteq L_{\infty\omega}^\omega(\exists)$? We can also raise analogous questions for the class HOM.

1. Is $\text{FO} \cap \text{HOM} \subseteq L_{\infty\omega}^\omega(\exists, +)$?
2. Is $\text{FO} \cap \text{HOM} \subseteq \text{Datalog}$?
3. Is $L_{\infty\omega}^\omega \cap \text{HOM} \subseteq L_{\infty\omega}^\omega(\exists, +)$?

Observe that, by Ajtai and Gurevich's result, a positive answer to question 2 would imply the truth of the Homomorphism preservation theorem over $\mathcal{F}$.

Finally, I would like to mention several questions that arise in connection with the results from Chapter 2. Recall that for all k , there is a single finite model that satisfies the set, $\exists^k$, of all consistent sentences of $L^k(\exists)$. By Proposition 15, $\text{Mod}_f(\exists^k)$ is not definable in $\bigvee L^k(\exists)$; of course, it may be definable in $\bigvee L^{k'}(\exists)$, for some $k' > k$. We ask, instead, the following related question.

1. For $k \geq 2$, is $\text{Mod}_f(\exists^k)$ in $\text{FO}(\exists)$?

Recall also that for any model A^k of the k -Gaifman theory, Γ_k , $A^k \models \exists^k$. Let $\mathcal{C}^k = \{A \mid \text{there is a } B \text{ such that } B \subseteq A \text{ and } B \models \Gamma_k\}$, the ‘upward closure’ of $\text{Mod}_f(\Gamma_k)$. It is clear that for all k , $\mathcal{C}^k \subseteq \text{Mod}_f(\exists^k)$, but we do not know whether the classes are equal.

2. For $k \geq 2$, is $\mathcal{C}^k = \text{Mod}_f(\exists^k)$?
3. For $k \geq 2$, is $\mathcal{C}^k$ in $\text{FO}(\exists)$?

The final question can be reformulated as a problem in combinatorics. It is equivalent to asking whether there is a finite set, $\{B_1, \dots, B_n\}$, of models with the ‘ k -extension property’ such that for every model with this property, there is some such B_i embedded in it.

Bibliography

- [1] F. Afrati, S. Cosmadakis, and M. Yannakakis. On datalog vs. polynomial time. In *Proceedings of the 10th ACM Symposium on Principles of Database Systems*, 1991.
- [2] M. Ajtai and Y. Gurevich. Monotone versus positive. *Journal of the ACM*, 34:1004–1015, 1987.
- [3] M. Ajtai and Y. Gurevich. Datalog vs. first-order logic. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 142–146, 1989.
- [4] H. Andreka, J. van Benthem, and I. Nemeti. Submodel preservation theorems in finite variable fragments. In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra*. Cambridge University Press, 1994.
- [5] H. Andreka, J. van Benthem, and I. Nemeti. Back and forth between modal logic and classical logic. Technical Report ML-95-04, University of Amsterdam, 1995.
- [6] J. Barwise. On Moschovakis closure ordinals. *Journal of Symbolic Logic*, 42:292–296, 1977.
- [7] J. van Benthem. *Modal correspondence theory*. PhD thesis, University of Amsterdam, 1976.
- [8] J. van Benthem. *Modal logic and classical logic*. Bibliopolis, 1985.
- [9] A. Dawar, S. Lindell, and S. Weinstein. Infinitary logic and inductive definability over finite structures. *Information and Computation*, 119:160–175, 1995.

- [10] R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In R. M. Karp, editor, *Complexity of Computation, SIAM-AMS Proceedings, Vol 7*, pages 43–73, 1974.
- [11] R. Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41(1):50–58, March 1976.
- [12] H. Gaifman. Concerning measures in first-order calculi. *Israel Journal of Mathematics*, 2:1–18, 1964.
- [13] M. Grohe. Existential least fixed-point logic and its relatives. Manuscript, 1995.
- [14] Y. Gurevich. Toward logic tailored for computational complexity. In M. Richter et al., editors, *Computation and Proof Theory*, pages 175–216. Springer Lecture Notes in Mathematics, 1984.
- [15] Y. Gurevich. On finite model theory (extended abstract). In S. R. Buss and P. J. Scott, editors, *Feasible Mathematics*, pages 211–219. Birkhauser, 1990.
- [16] W. Hanf. Model-theoretic methods in the study of elementary logic. In J. Addison, L. Henkin, and A. Tarski, editors, *The Theory of Models*, pages 132–145. North Holland, 1965.
- [17] N. Immerman. Upper and lower bounds for first-order expressibility. *Journal of Computer and System Sciences*, 25:76–98, 1982.
- [18] Ph. G. Kolaitis and M. Y. Vardi. On the expressive power of datalog: tools and a case study. In *Proceedings of the 9th ACM Symposium on Principles of Database Systems*, pages 61–71, 1990.
- [19] Ph. G. Kolaitis and M. Y. Vardi. Infinitary logics and 0-1 laws. *Information and Computation*, 98(2):258–294, 1992.
- [20] Libo Lo. Preservation theorems of finite models (abstract). *Journal of Symbolic Logic*, 58:376, 1993.
- [21] B. Poizat. Deux ou trois choses que je sais de L_n . *Journal of Symbolic Logic*, 47(3):641–658, 1982.

- [22] A. Stolboushkin. Finitely monotone properties. In *Proceedings of the 10th IEEE Symposium on Logic in Computer Science*, pages 324–330, 1995.
- [23] W. Tait. A counterexample to a conjecture of Scott and Suppes. *Journal of Symbolic Logic*, 24(1):15–16, 1959.