Extension Preservation in the Finite and Prefix Classes of First Order Logic

Anuj Dawar
Department of Computer Science and Technology, University of Cambridge, UK
anuj.dawar@cl.cam.ac.uk

Abhisekh Sankaran
Department of Computer Science and Technology, University of Cambridge, UK
abhisekh.sankaran@cl.cam.ac.uk

Abstract

It is well known that the classic Łoś-Tarski preservation theorem fails in the finite: there are first-order definable classes of finite structures closed under extensions which are not definable (in the finite) in the existential fragment of first-order logic. We strengthen this by constructing for every n, first-order definable classes of finite structures closed under extensions which are not definable with n quantifier alternations. The classes we construct are definable in the extension of Datalog with negation and indeed in the existential fragment of transitive-closure logic. This answers negatively an open question posed by Rosen and Weinstein.

2012 ACM Subject Classification Theory of computation → Finite Model Theory

Keywords and phrases finite model theory, preservation theorems, extension closed, composition, Datalog, Ehrenfeucht-Fraisse games

Digital Object Identifier 10.4230/LIPIcs.CSL.2021.18

Funding Research supported by the Leverhulme Trust through a Research Project Grant on "Logical Fractals"

1 Introduction

The failure of classical preservation theorems of model theory has been a topic of persistent interest in finite model theory. In the classical setting, preservation theorems provide a tight link between the syntax and semantics of first-order logic (FO). For instance, the Łoś-Tarski preservation theorem (see [8]) implies that any sentence of first-order logic whose models are closed under extensions is equivalent to an existential sentence. This, like many other classical preservation theorems, is false when we retrict ourselves to finite structures. Tait [17] and Gurevich [7] provide examples of sentences whose finite models are closed under extensions, but which are not equivalent, over finite structures, to any existential sentence. Many other classical preservation theorems have been studied in the context of finite model theory (e.g. [12, 14]), but our focus in this paper is on extension-closed properties.

The failure of the Łoś-Tarski theorem in the finite opens a number of different avenues of research. One line of work has sought to investigate restricted classes of structures on which a version of the preservation theorem holds (see [2, 4]). Another direction is prompted by the question of whether we can identify some proper syntactic fragment of FO, beyond the existential, which contains definitions of all extension-closed FO-definable properties. For instance, the examples from Tait and Gurevich are both Σ_3 sentences. Could it be that every FO sentence whose finite models are closed under extensions is equivalent to a Σ_3 sentence? Or, indeed, a Σ_n sentence for some constant n? We answer these questions negatively in this paper. That is, we show that we can construct, for each n, a sentence φ whose finite models are extension closed but which is not equivalent in the finite to a Σ_n sentence.

© Anuj Dawar and Abhisekh Sankaran;
licensed under Creative Commons License CC-BY

29th EACSL Annual Conference on Computer Science Logic (CSL 2021).
Editors: Christel Baier and Jean Goubault-Larrecq; Article No. 18; pp. 18:1–18:13
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

A related question is posed by Rosen and Weinstein [13]. They observe that the constructions due to Tait and Gurevich both yield classes of finite structures that are definable in $\texttt{Datalog}(\neg)$, the existential fragment of fixed-point logic in which only extension-closed properties can be expressed. They ask if it might be the case that $\mathrm{FO} \cap \texttt{Datalog}(\neg)$ is contained in some level of the first-order quantifier alternation hierarchy, be it not the lowest level. That is, could it be that every property that is first-order definable and also definable in $\texttt{Datalog}(\neg)$ is definable by a Σ_n sentence for some constant n? Our construction answers this question negatively as we show that the sentences we construct are all equivalent to formulas of $\texttt{Datalog}(\neg)$.

Our result also greatly strengthens a previous result by Sankaran [15] which showed that for each k there is an extension-closed property of finite structures definable in FO but not in Π_2 with k leading universal quantifiers. Indeed, our result answers (negatively) Problem 2 in [15].

In Section 2 we give the necessary background definitions. We construct the sentences in Section 3 and show that they can all be expressed in $\texttt{Datalog}(\neg)$. Section 4 contains the proof that the sequence of sentences contains, for each n, a sentence that is not equivalent to any Σ_n sentence. We conclude with some suggestions for further investigation.

2 Preliminaries

We work with logics: first-order logic (FO) and extensions of $\texttt{Datalog}$ over finite relational vocabularies. We assume the reader is familiar with the basic definitions of first-order logic (see, for instance [10]). A *vocabulary* τ is a set of predicate and constant symbols. In the vocabularies we use, all predicate symbols are either unary or binary. We denote by $\mathrm{FO}(\tau)$ the set of all FO formulas over the vocabulary τ. A sequence $(x_1, \ldots, x_k)$ of variables is denoted by $\bar{x}$. We use $\psi(\bar{x})$ to denote a formula ψ whose free variables are among $\bar{x}$. A formula without free variables is called a *sentence*. A formula which begins with a string of quantifiers that is followed by a quantifier-free formula, is said to be in *prenex normal form (PNF)*. The string of quantifiers in a PNF formula is called the *quantifier prefix* of the formula. It is well known that every formula is equivalent to a formula in PNF. We denote by Σ_n, the collection of all formulas in PNF whose quantifier prefix contains at most n blocks of quantifiers beginning with a block of existential quantifiers. Equivalently, a PNF formula is in Σ_n if it starts with a block of existential quantifiers and contains at most $n-1$ alternations in its quantifier prefix. Similarly, a formula is Π_n if it begins with a block of universal quantifiers and contains at most $n-1$ alternations in its quantifier prefix. We write $\Sigma_{n,k}$ for the subclass of Σ_n consisting of those formulas in which every quantifier block has at most k quantifiers. Similarly, $\Pi_{n,k}$ is the subclass of Π_n where each block has at most k quantifiers. Thus $\Sigma_n = \bigcup_{k \geq 1} \Sigma_{n,k}$ and $\Pi_n = \bigcup_{k \geq 1} \Pi_{n,k}$.

We use standard notions concerning τ-structures as defined in [3]. We denote τ-structures as $\mathfrak{A}, \mathfrak{B}$ etc., and refer to them simply as structures when τ is clear from the context. We denote by $\mathfrak{A} \subseteq \mathfrak{B}$ that $\mathfrak{A}$ is a substructure of $\mathfrak{B}$, and by $\mathfrak{A} \cong \mathfrak{B}$ that $\mathfrak{A}$ is isomorphic to $\mathfrak{B}$.

We now introduce some notation with respect to the classes of formulas $\Sigma_{n,k}$ and $\Pi_{n,k}$.

▶ **Definition 1.** *We say $\mathfrak{A} \Rrightarrow_{n,k} \mathfrak{B}$ if every $\Sigma_{n,k}$ sentence true in $\mathfrak{A}$ is also true in $\mathfrak{B}$.*
*We say $\mathfrak{A}$ and $\mathfrak{B}$ are $\equiv_{n,k}$-*equivalent*, and write $\mathfrak{A} \equiv_{n,k} \mathfrak{B}$, if $\mathfrak{A} \Rrightarrow_{n,k} \mathfrak{B}$ and $\mathfrak{B} \Rrightarrow_{n,k} \mathfrak{A}$.*

By extension, for tuples $\bar{a}$ and $\bar{b}$ of elements of $\mathfrak{A}$ and $\mathfrak{B}$ respectively, we also write $(\mathfrak{A}, \bar{a}) \Rrightarrow_{n,k} (\mathfrak{B}, \bar{b})$ to indicate that every formula φ which is satisfied in $\mathfrak{A}$ when its free variables are instantiated with $\bar{a}$ is also satisfied in $\mathfrak{B}$ when they are instantiated by $\bar{b}$, and similarly for $\equiv_{n,k}$. Note that $\mathfrak{A} \Rrightarrow_{n,k} \mathfrak{B}$ holds if every $\Pi_{n,k}$ sentence true in $\mathfrak{B}$ is also true in $\mathfrak{A}$. The following useful fact is now immediate from the definition.

▶ **Lemma 2.** *$\mathfrak{A} \Rrightarrow_{n+1,k} \mathfrak{B}$ if, and only if, for every k-tuple $\bar{a}$ of elements of $\mathfrak{A}$, there is a k-tuple $\bar{b}$ of elements of $\mathfrak{B}$ such that $(\mathfrak{B}, \bar{b}) \Rrightarrow_{n,k} (\mathfrak{A}, \bar{a})$.*

We assume the reader is familiar with the standard Ehrenfeucht-Fraïssé game characterizing the equivalence of two structures with respect to sentences of a given quantifier nesting depth (see for example [10, Chapter 3]). In this paper, we use a "prefix" variant of this Ehrenfeucht-Fraïssé game. For $n, k \geq 1$, the (n, k)*-prefix Ehrenfeucht-Fraïssé game* on a pair $(\mathfrak{A}, \mathfrak{B})$ of structures, is the usual Ehrenfeucht-Fraïssé game on $\mathfrak{A}$ and $\mathfrak{B}$ but with two restrictions: (i) In every odd round, the Spoiler plays on $\mathfrak{A}$ and in every even round, on $\mathfrak{B}$, and (ii) in each round the Spoiler chooses a k-tuple of elements from the relevant structure (as opposed to a single element in the usual Ehrenfeucht-Fraïssé game). The winning condition for the Duplicator is the same as that in the usual Ehrenfeucht-Fraïssé game: Duplicator wins at the end of n rounds, if when $\bar{a}_1, \ldots, \bar{a}_n$ are the k-tuples picked in $\mathfrak{A}$ and $\bar{b}_1, \ldots, \bar{b}_n$ are the k-tuples picked in $\mathfrak{B}$, then the map taking the nk-tuple $(\bar{a}_1 \cdots \bar{a}_n)$ to $(\bar{b}_1 \cdots \bar{b}_n)$ pointwise is a partial isomorphism from $\mathfrak{A}$ to $\mathfrak{B}$. Entirely analogously to the usual Ehrenfeucht-Fraïssé game theorem (see [10, Theorem 3.18]), we have the following.

▶ **Theorem 3.** *Duplicator has a winning strategy in the* (n, k)-prefix Ehrenfeucht-Fraïssé game *on a pair* $(\mathfrak{A}, \mathfrak{B})$ *of structures if, and only if,* $\mathfrak{A} \Rrightarrow_{n,k} \mathfrak{B}$

Note in particular that, given the fact that any two linear orders of length $\geq 2^n$ are equivalent with respect to all sentences of quantifier nesting depth n [10, Theorem 3.6], it follows that there is a winning strategy for the Duplicator in the (n, k)-prefix Ehrenfeucht-Fraïssé game on any pair of linear orders, each of length $\geq 2^{n \cdot k}$ and any two such linear orders are $\equiv_{n,k}$-equivalent.

Where it causes no confusion, we still use $\equiv_m$ to denote the usual equivalence up to *quantifier rank* m. Note that $\mathfrak{A} \equiv_m \mathfrak{B}$ implies $\mathfrak{A} \equiv_{n,k} \mathfrak{B}$ whenever $m \geq nk$.

Formulas in Σ_1 are said to be *existential.* A Σ_1 formula that also contains no occurrences of the negation symbol is said to be *existential positive.* It is easy to see that the class of models of any Σ_1 sentence φ is closed under extensions: if $\mathfrak{A} \models \varphi$ and $\mathfrak{A} \subseteq \mathfrak{B}$, then $\mathfrak{B} \models \varphi$. Dually, the class of models of any Π_1 sentence is closed under taking substructures. Similarly, the class of models of any existential positive sentence is closed under homomorphisms.

`Datalog` is a database query language which can be seen as an extension of existential positive first-order logic with a recursion mechanism. Equivalently, it can be seen as the existential positive fragment of the logic of least fixed points `LFP` (see [10, Chapter 10]). We briefly review the definitions of this language, along with its extension `Datalog`$(\neg)$.

A *Datalog program* is a finite set of rules of the form $T_0 \leftarrow T_1, \ldots, T_m$, where each T_i is an atomic formula. T_0 is called the *head* of the rule, while the right-hand side is called the *body.* These atomic formulas use relational symbols from a vocabulary $\sigma \cup \tau$, where the symbols in σ are called *extensional* predicates and those in τ are *intensional* predicates. Every symbol that occurs in the head of a rule is an intensional predicate, while both intensional and extensional predicates can occur in the body of a rule. The semantics of such a program is defined with respect to a σ-structure $\mathfrak{A}$. Say that a rule $T_0 \leftarrow T_1, \ldots, T_m$ is satisfied in a $\sigma \cup \tau$ expansion $\mathfrak{A}'$ of $\mathfrak{A}$ if $\mathfrak{A}' \models \forall \bar{x}\big((\bigwedge_{1 \leq i \leq m} T_i) \rightarrow T_0\big)$, where $\bar{x}$ enumerates all the variables occurring in the rule. The interpretation of a `Datalog` program in $\mathfrak{A}$ is the smallest expansion of $\mathfrak{A}$ (when ordered by pointwise inclusion of the relations interpreting τ) satisfying all the rules in the program. This is uniquely defined as it is obtained as the simultaneous least fixed-point of the existential closure of the right-hand side of the rules. We distinguish one intensional predicate G and call it the *goal predicate.* Then, the query computed by a program π is the interpretation of G in the interpretation of π in $\mathfrak{A}$. In particular, if G is a 0-ary predicate symbol (i.e. a Boolean variable), π defines a Boolean query, i.e. a class of structures.

Since the interpretation of π is obtained as the least fixed-point of an existential positive formula, it is easily seen that the query defined is closed under homomorphisms and hence also under extensions. We can understand `Datalog` as the existential positive fragment of the least-fixed point logic LFP, though it is known that there are homomorphism-closed properties definable in LFP that are not expressible in `Datalog` (see [5]).

We get more general queries by allowing limited forms of negation. Specifically, in `Datalog`$(\neg)$, in a rule $T_0 \leftarrow T_1, \ldots, T_m$, each T_i on the right-hand side is either an atom *or* a negated atom involving an *extensional* predicate symbol or equality. In short, we allow negation on the predicate symbols in σ and on equalities but the fixed-point variables (i.e. the predicate symbols in τ) still only appear positively, so the least fixed-point is still well defined. As it is still the least-fixed point of existential formulas, the formula still defines a property closed under extensions. For more on the extensions of `Datalog` with negation, see [1, 9].

3 The Extension-Closed Properties

We now construct a family of properties, each of which is definable in first-order logic and closed under extensions. Indeed, we show that each of the properties is definable in `Datalog`$(\neg)$.

3.1 First-Order Definitions

To begin, we define, for each $n \in \mathbb{N}$, a vocabulary σ_n. These are defined by induction on n. The vocabulary σ_1 consists of three binary relation symbols $\leq, S, R$. For all $n > 1$, $\sigma_n = \sigma_{n-1} \cup \{S_n, R_n, P_n\}$ where S_n and R_n are binary relation symbols and P_n is a unary relation symbol.

Consider first the sentence NLO of FO which asserts that $\leq$ is *not* a linear order. This is easily seen to be an existential sentence and so also definable in `Datalog`$(\neg)$. Suppose now that φ is any sentence whose models, restricted to ordered structures (i.e. those structures which interpret $\leq$ as a linear order), are extension closed. Then, it follows that NLO $\vee\, \varphi$ defines an extension-closed class of structures. Moreover, this class is FO or `Datalog`$(\neg)$ definable if φ is in the respective logic. Also, if φ is a Σ_n sentence, then so is NLO $\vee\, \varphi$, and if $n > 1$ and φ is a Π_n sentence then so is NLO $\vee\, \varphi$. Thus, in what follows, we restrict our attention to the class of ordered structures. We construct our sentences on the assumption that structures are ordered, and show that they define extension-closed classes on ordered structures.

With this in mind, we use some convenient notational abbreviations. We write $x < y$ as short-hand for $x \leq y \wedge x \neq y$. We also write "y is the successor of x", "x is the minimum element", etc. with their obvious meanings. Also, let φ be any formula of FO, and x and y be variables not occuring in φ. We write $\varphi^{[x,y]}$ for the formula $x \leq y \wedge \varphi^{\star}$ where $\varphi^{\star}$ is the formula obtained by relativizing every quantifier in φ to the interval $[x, y]$. That is to say, inductively, every subformula $\exists z \theta$ is replaced by $\exists z(x \leq z \wedge z \leq y) \wedge \theta^{\star}$ and every subformula $\forall z \theta$ by $\forall z(x \leq z \wedge z \leq y) \rightarrow \theta^{\star}$. Where the variables x and y do appear in φ, the formula $\varphi^{[x,y]}$ is defined by first renaming variables in φ to avoid clashes and then applying the relativization.

Next, consider the sentence PartialSucc defined as follows.

$$\text{PartialSucc} \;:=\; \forall x \forall y \; S(x,y) \rightarrow \text{“}y \text{ is the successor of } x\text{”}.$$

This is a Π_1 sentence, asserting that the relation S is a "partial successor" relation. Its negation is an existential sentence and hence closed under extensions. Thus, if φ defines an extension-closed class of structures when restricted to ordered structures in which S is a partial successor relation, then $\mathrm{NLO} \vee \neg\mathrm{PartialSucc} \vee \varphi$ defines an extension-closed class.

We write $\mathrm{Total}(x, y)$ for the formula that asserts, on structures in which PartialSucc is true, that in the interval $[x, y]$, S is, in fact, total. That is:

$$\mathrm{Total}(x, y) \ := \ x < y \wedge \forall z\big(x \leq z \wedge z < y\big) \rightarrow \exists w\big(z < w \wedge w \leq y \wedge S(z, w)\big).$$

Now, we can define the sentence $\mathrm{SomeTotalR}_1$.

$$\mathrm{SomeTotalR}_1 \ := \ \neg\mathrm{PartialSucc} \vee \exists x \exists y \ \big(R(x, y) \wedge \mathrm{Total}(x, y)\big).$$

Note that $\mathrm{Total}(x, y)$ is a Π_2 formula, and $\mathrm{SomeTotalR}_1$ is a Σ_3 sentence. The latter is the first in our family of sentences. To see why this sentence is closed under extensions on ordered structures, suppose $\mathfrak{A}$ is an ordered model of $\mathrm{SomeTotalR}_1$ on which S is a partial successor. Thus, there is an interval $[x, y]$ in $\mathfrak{A}$ on which S is total. Let $\mathfrak{B}$ be an extension of $\mathfrak{A}$. If $\mathfrak{B}$ contains no additional elements in the interval $[x, y]$, then $\mathrm{Total}(x, y)$ still holds in $\mathfrak{B}$ and therefore $\mathfrak{B}$ is a model of $\mathrm{SomeTotalR}_1$. On the other hand, suppose $\mathfrak{B}$ contains an additional element w in the interval $[x, y]$. Let a and b be the two successive elements of $\mathfrak{A}$ between which w appears. Since S is total in the interval $[x, y]$ in $\mathfrak{A}$, we know that $S(a, b)$ holds in $\mathfrak{A}$ and, by extension, in $\mathfrak{B}$. Since b is not the successor of a in $\mathfrak{B}$, we conclude that $\neg$PartialSucc is true in $\mathfrak{B}$ and therefore the structure is a model of $\mathrm{SomeTotalR}_1$.

The sentence $\mathrm{SomeTotalR}_1$ is essentially the example constructed by Tait that exhibits an existential-closed first-order property that is not expressible by an existential sentence. We now define σ_n-sentences $\mathrm{SomeTotalR}_n$, for $n > 0$ by induction.

First, we define a formula $\mathrm{Succ}_n(x, y)$ as follows.

$$\mathrm{Succ}_n(x, y) \ := \ P_n(x) \wedge P_n(y) \wedge S_n(x, y) \wedge \mathrm{SomeTotalR}_{n-1}^{[x,y]}.$$

We further define the formula $\mathrm{PartialSucc}_n$ which asserts that Succ_n is a partial successor relation when restricted to the elements in the relation P_n. That is,

$$\mathrm{PartialSucc}_n \ := \ \forall x \forall y \ \mathrm{Succ}_n(x, y) \rightarrow \forall z(P_n(z) \rightarrow z \leq x \vee y \leq z).$$

We can now define the formula $\mathrm{Total}_n(x, y)$ which defines, in those structures in which $\mathrm{PartialSucc}_n$ is true, those intervals $[x, y]$ where the successor defined by Succ_n is total. That is,

$$\mathrm{Total}_n(x, y) \ := \ x < y \wedge \forall z\big(P_n(z) \wedge x \leq z \wedge z < y\big) \rightarrow \exists w\big(z < w \wedge w \leq y \wedge \mathrm{Succ}_n(z, w)\big).$$

Finally, we define the sentence

$$\mathrm{SomeTotalR}_n \ := \ \neg\mathrm{PartialSucc}_n \vee \exists x \exists y \ \big(R_n(x, y) \wedge \mathrm{Total}_n(x, y)\big).$$

Note that, $\mathrm{SomeTotalR}_n$ is a Σ_{2n+1} sentence. This can be established by induction on n. Indeed, as we noted, $\mathrm{SomeTotalR}_1$ is a Σ_3 sentence. Assuming $\mathrm{SomeTotalR}_n$ is a Σ_{2n+1} sentence for some n, we note that Succ_{n+1} is a Σ_{2n+1} formula, and so is $\neg\mathrm{PartialSucc}_{n+1}$. Then Total_{n+1} is a Π_{2n+2} formula and $\mathrm{SomeTotalR}_{n+1}$ is Σ_{2n+3}.

3.2 Datalog Definitions

Next, we show that these formulas also admit a definition in `Datalog`($\neg$), which establishes, in particular, that they define extension-closed classes. We use the same names for formulas in `Datalog`($\neg$) as we used for FO formulas above, when they define the same property. As we noted, the sentences NLO and $\neg$PartialSucc are both Σ_1 sentences and we therefore assume they are available as `Datalog`($\neg$) predicates. We now define Total by the following rules.

$$\begin{aligned}\mathrm{Total}(x,y) &\longleftarrow S(x,y)\\ \mathrm{Total}(x,y) &\longleftarrow S(x,z),\mathrm{Total}(z,y)\end{aligned}$$

This just defines Total as the transitive closure of S. It is clear that, in ordered structures where S is a partial successor relation, the pair (x,y) is in the transtive closure of S precisely when $x<y$ and S is total in the interval $[x,y)$. Thus, we can now define:

$$\mathrm{RTotal}_1(x,y) \longleftarrow x\le u, v\le y, R(u,v),\mathrm{Total}(u,v)$$

This defines those pairs (x,y) such that for some u,v in the interval $[x,y]$, $R(u,v)$ holds and the successor relation is total. In other words, it defines $\mathrm{SomeTotalR}_1^{[x,y]}$. We can obtain $\mathrm{SomeTotalR}_1$ as the existential closure of this. For the inductive definition, the predicate RTotal_n is useful.

Inductively, we define the relation, Succ_n as follows.

$$\mathrm{Succ}_n(x,y) \longleftarrow P_n(x),P_n(y),S_n(x,y),\mathrm{RTotal}_{n-1}(x,y)$$

The negation of $\mathrm{PartialSucc}_n$ is now defined by the following

$$\mathrm{NotPartialSucc}_n \longleftarrow \mathrm{Succ}_n(x,y),P_n(z),x\le z,z\le y,x\ne z,y\ne z$$

Now, entirely analogously to Total above, we can give a definition of Total_n as the transitive closure of Succ_n and this is equivalent to the FO definition given above on ordered structures on which $\mathrm{PartialSucc}_n$ is true.

$$\begin{aligned}\mathrm{Total}_n(x,y) &\longleftarrow \mathrm{Succ}_n(x,y)\\ \mathrm{Total}_n(x,y) &\longleftarrow \mathrm{Succ}_n(x,z),\mathrm{Total}_n(z,y)\end{aligned}$$

Inductively we define the relation RTotal_n, and its existential closure, giving the sentence $\mathrm{SomeTotalR}_n$.

$$\mathrm{RTotal}_n(x,y) \longleftarrow x\le u,v\le y,R_n(u,v),\mathrm{Total}_n(u,v)$$

It should be noted that the only use of the recursive features of `Datalog`($\neg$) that we made use in writing the formulas above was to define the transitive closure of the relations Total and Total_n. Thus, the definitions could equally well be formalized in the existential fragment of transitive closure logic.

4 Proof of the Main Result

In this section, we establish our main result. We establish that $\mathrm{SomeTotalR}_n$, which we noted is a Σ_{2n+1} sentence, is not equivalent to a Π_{2n+1} sentence. To do this, we construct ordered structures $\mathfrak{M}_{n,k}$ and $\mathfrak{N}_{n,k}$ for every k such that $\mathfrak{M}_{n,k}$ is a model of $\mathrm{SomeTotalR}_n$, $\mathfrak{N}_{n,k}$ is not a model of $\mathrm{SomeTotalR}_n$ but $\mathfrak{N}_{n,k} \Rrightarrow_{2n+1,k} \mathfrak{M}_{n,k}$. The main lemma establishing this is Lemma 5 below. Here we state the theorem that is a consequence.

▶ **Theorem 4.** *For every n, there is a Σ_{2n+1} sentence whose finite models are closed under extensions and which is equivalent to a* `Datalog`$(\neg)$ *program, but which is not equivalent over finite structures to any Π_{2n+1} sentence.*

Proof. The sentence is $\mathrm{NLO} \vee \mathrm{SomeTotalR}_n$ which we have already noted is a Σ_{2n+1} sentence, expressible as a `Datalog`$(\neg)$ program and its models are extension-closed. Suppose it were expressible as a Π_{2n+1} sentence. Then, since it is satisfied in $\mathfrak{M}_{n,k}$ as we show in Section 4.1 and since $\mathfrak{N}_{n,k} \Rrightarrow_{2n+1,k} \mathfrak{M}_{n,k}$ by Lemma 5 we have that the sentence is true in $\mathfrak{N}_{n,k}$. But, as we show in Section 4.1, $\mathfrak{N}_{n,k}$ is not a model of $\mathrm{SomeTotalR}_n$, yielding a contradiction. ◀

4.1 Construction of the Structures

We describe the construction of structures $\mathfrak{M}_{n,k}$ and $\mathfrak{N}_{n,k}$ for each n and k. The construction is by induction on n, simultaneously for all k. In the course of the construction we also define, for all n and k structures $\mathsf{Tot}_{n,k}$ and $\mathsf{Gap}_{n,k}$ which we use as auxilliary structures. For all n and k, $\mathfrak{M}_{n,k}, \mathfrak{N}_{n,k}, \mathsf{Tot}_{n,k}$ and $\mathsf{Gap}_{n,k}$ are structures over the vocabulary σ_n.

All structures we consider interpret the relation symbol $\leq$ as a linear order of the universe and S as a partial successor relation. It is useful to formally define the notion of an ordered sum of structures. For a pair $\mathfrak{A}$ and $\mathfrak{B}$ of ordered structures the *ordered sum* $\mathfrak{A} \oplus \mathfrak{B}$ is a structure whose universe is the disjoint union of the universes of $\mathfrak{A}$ and $\mathfrak{B}$ *except* that the maximum element of $\mathfrak{A}$ is identified with the minimum element of $\mathfrak{B}$. The relation $\leq$ is interpreted in $\mathfrak{A} \oplus \mathfrak{B}$ by taking the union of its interpretations in $\mathfrak{A}$ and $\mathfrak{B}$ and letting $a \leq b$ for all a in $\mathfrak{A}$ and b in $\mathfrak{B}$. All other relation symbols are interpreted in $\mathfrak{A} \oplus \mathfrak{B}$ by the union of their interpretations in the two structures. The operation of ordered sum is clearly associative and we can thus write $\bigoplus_{i \in I} \mathfrak{A}_i$ for the ordered sum of a sequence of structures indexed by an ordered set I. Note that our use of the term "ordered sum" differs somewhat from its use, say by Ebbinghaus and Flum [6, Sec. 1.A.3]. The key difference is that in their definition we do not identify the maximum element of $\mathfrak{A}$ with the minimum element of $\mathfrak{B}$ but rather simply take the disjoint union of the two universes.

The structure $\mathsf{Tot}_{1,k}$ has $m = 6(k+2)^2$ elements which we identify with the initial segment of the positive integers $[1, \ldots, m]$ with $\leq$ the natural linear order on these, S the successor relation and the relation R containing just the pair $(1, m)$. The structure $\mathsf{Gap}_{1,k}$ is obtained from $\mathsf{Tot}_{1,k}$ by removing from the relation S the *central* pair of elements, i.e. $(m/2, m/2+1)$.

We now obtain $\mathfrak{N}_{1,k}$ as the ordered sum of $4(k+3)^3 + 2k + 1$ copies of $\mathsf{Gap}_{1,k}$. That is $\mathfrak{N}_{1,k} = \bigoplus_{i \in [4(k+3)^3+2k+1]} \mathfrak{G}_i$ where each $\mathfrak{G}_i$ is isomorphic to $\mathsf{Gap}_{1,k}$. We also let $\mathfrak{M}_{1,k} = \bigoplus_{i \in [2(k+3)^3+k]} \mathfrak{G}_i \oplus \mathsf{Tot}_{1,k} \oplus \bigoplus_{i \in [2(k+3)^3+k]} \mathfrak{G}_i$. In short, $\mathfrak{M}_{1,k}$ is obtained from $\mathfrak{N}_{1,k}$ by replacing the central copy of $\mathsf{Gap}_{1,k}$ with a copy of $\mathsf{Tot}_{1,k}$.

Let now $n \geq 2$ and suppose we have defined the σ_{n-1}-structures $\mathfrak{N}_{n-1,k}, \mathfrak{M}_{n-1,k}, \mathsf{Tot}_{n-1,k}$ and $\mathsf{Gap}_{n-1,k}$. Write $\mathfrak{N}^+_{n-1,k}$ and $\mathfrak{M}^+_{n-1,k}$ for the σ_n-structures that are obtained from $\mathfrak{N}_{n-1,k}$ and $\mathfrak{M}_{n-1,k}$ respectively by interpreting P_n as the two element set $\{\min, \max\}$ containing the minimum and maximum elements of the structure and S_n as the relation containing the single pair $(\min, \max)$ (R_n is empty in both these structures). Now, $\mathsf{Tot}_{n,k}$ is the structure obtained from $\bigoplus_{i \in [4(k+3)^{2n}+2k+1]} \mathfrak{M}^+_{n-1,k}$ (i.e. the ordered sum of $4(k+3)^{2n} + 2k + 1$ copies of $\mathfrak{M}^+_{n-1,k}$) by adding to the relation R_n the pair relating the minimum and maximum elements of the linear order. Similarly $\mathsf{Gap}_{n,k}$ is obtained from $\bigoplus_{i \in [2(k+3)^{2n}+k]} \mathfrak{M}^+_{n-1,k} \oplus \mathfrak{N}^+_{n-1,k} \oplus \bigoplus_{i \in [2(k+3)^{2n}+k]} \mathfrak{M}^+_{n-1,k}$ by adding to the relation R_n the pair relating the minimum and maximum elements of the linear order. Equivalently, $\mathsf{Gap}_{n,k}$ is obtained from $\mathsf{Tot}_{n,k}$ by replacing the central copy of $\mathfrak{M}_{n-1,k}$ by a copy of $\mathfrak{N}_{n-1,k}$.

Finally, we can define $\mathfrak{N}_{n,k}$ as the ordered sum of $4(k+3)^{2n+1}+2k+1$ copies of $\mathsf{Gap}_{n,k}$ and $\mathfrak{M}_{n,k}$ as the structure obtained from $\mathfrak{N}_{n,k}$ by replacing the central copy of $\mathsf{Gap}_{n,k}$ by a copy of $\mathsf{Tot}_{n,k}$. This completes the definition of the structures.

We now argue that for all values of n and k, $\mathfrak{M}_{n,k}$ is a model of $\mathrm{SomeTotalR}_n$ and $\mathfrak{N}_{n,k}$ is not. This is an easy induction on n. For $n=1$, every interval $[x,y]$ of $\mathfrak{N}_{1,k}$ for which $R(x,y)$ holds induces a copy of $\mathsf{Gap}_{1,k}$. By construction S is not a complete successor relation in $\mathsf{Gap}_{1,k}$, and so $\mathfrak{N}_{1,k}$ does not satisfiy $\mathrm{SomeTotalR}_1$. On the other hand, $\mathfrak{M}_{1,k}$ contains an interval $[x,y]$ with $R(x,y)$ that induces a copy of $\mathsf{Tot}_{1,k}$ and so $\mathfrak{M}_{1,k} \models \mathrm{SomeTotalR}_1$.

Inductively, assume that $\mathfrak{M}_{n-1,k} \models \mathrm{SomeTotalR}_{n-1}$ and $\mathfrak{N}_{n-1,k} \not\models \mathrm{SomeTotalR}_{n-1}$. Now, in both $\mathsf{Tot}_{n,k}$ and $\mathsf{Gap}_{n,k}$, the relation S_n relates successive elements that are in P_n. If x,y is a pair of such successive elements then in $\mathsf{Tot}_{n,k}$ the interval $[x,y]$ always induces a structure whose σ_{n-1}-reduct is a copy of $\mathfrak{M}_{n-1,k}$ and therefore satisfies $\mathrm{SomeTotalR}_{n-1}$. Hence $\mathrm{Succ}_n(x,y)$ is satisfied in $\mathsf{Tot}_{n,k}$ for all such pairs. On the other hand, in $\mathsf{Gap}_{n,k}$ there is an interval $[x,y]$ with $S_n(x,y)$ which induces a structure whose σ_{n-1}-reduct is a copy of $\mathfrak{N}_{n-1,k}$ and therefore fails to satisfy $\mathrm{SomeTotalR}_{n-1}$. Hence $\mathrm{Total}_n(x_0,y_0)$ is true in $\mathsf{Tot}_{n,k}$ and false in $\mathsf{Gap}_{n,k}$ when x_0 and y_0 are interpreted as the minimum and maximum elements in the structure respectively. Since in $\mathfrak{N}_{n,k}$ all intervals $[x,y]$ for which $R_n(x,y)$ holds induce a copy of $\mathsf{Gap}_{n,k}$ and in $\mathfrak{M}_{n,k}$ there is such an interval which induces a copy of $\mathsf{Tot}_{n,k}$, we conclude that $\mathfrak{M}_{n,k} \models \mathrm{SomeTotalR}_n$ and $\mathfrak{N}_{n,k} \not\models \mathrm{SomeTotalR}_n$.

4.2 The Game Argument

Our aim in this section is to establish the following lemma using an Ehrenfeucht-Fraïssé game argument:

▶ **Lemma 5.** *For each n,k, $\mathfrak{N}_{n,k} \Rrightarrow_{2n+1,k} \mathfrak{M}_{n,k}$.*

Our development of the Duplicator winning strategy in the game follows the inductive construction of the structures themselves. For this, we first develop some tools for constructing strategies on ordered sums and expansions of structures from strategies on their component parts. First, we introduce some useful notation.

For any ordered structure $\mathfrak{A}$, write $\mathfrak{A}^*$ for the expansion of $\mathfrak{A}$ with constants min and max interpreted by the minimum and maximum elements of the structure. The main reason for introducing these is that we generally want to restrict attention to Duplicator strategies that respect the minimum and maximum elements and a notationally convenient way to do this is to have constants for these elements.

It is a standard fact that the equivalence relation $\equiv_m$ is a congruence with respect to various ways of combining structures. In particular, it is so with respect to the notion of ordered sum defined by Ebbinghaus and Flum [6, Prop. 2.3.10]. The same method of composition of strategies can be used to show the following.

▶ **Lemma 6.** *If $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{B}_1, \mathfrak{B}_2$ are ordered structures and $\bar{a}_1, \bar{a}_2, \bar{b}_1, \bar{b}_2$ tuples of elements from $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{B}_1$ and $\mathfrak{B}_2$ respectively, such that $(\mathfrak{A}_1, \bar{a}_1)^* \Rrightarrow_{n,k} (\mathfrak{B}_1, \bar{b}_1)^*$ and $(\mathfrak{A}_2, \bar{a}_2)^* \Rrightarrow_{n,k} (\mathfrak{B}_2, \bar{b}_2)^*$, then*

$$(\mathfrak{A}_1 \oplus \mathfrak{A}_2, \bar{a}_1\bar{a}_2)^* \Rrightarrow_{n,k} (\mathfrak{B}_1 \oplus \mathfrak{B}_2, \bar{b}_1\bar{b}_2)^*.$$

Note that it is an immediate consequence that the same is true with $\equiv_{n,k}$ in place of $\Rrightarrow_{n,k}$.

Furthermore, this also extends to ordered sums of sequences. Moreover, we do not have to match the lengths of the sequences as long as they are long enough. Again, this is standard for the equivalence relation $\equiv_m$ [6, Ex. 2.3.13], for sequences of length at least 2^m and a slightly different notion of ordered sum. For our relations we obtain a tighter bound, so we prove this explicitly as the proof is instructive.

Define the following function ρ on pairs of natural numbers by recursion.

$$\begin{aligned}\rho(1,k) &= 2k+2\\ \rho(n+1,k) &= (k+2)(\rho(n,k)+1).\end{aligned}$$

A simple induction on n shows that $2(k+3)^n > \rho(n,k)$ for all $k,n \geq 1$.

▶ **Lemma 7.** *If $\mathfrak{A}$ and $\mathfrak{B}$ are ordered structures, $\mathfrak{A}^* \Rrightarrow_{n,k} \mathfrak{B}^*$ and $s,t \geq \rho(n,k)$, then $\left(\bigoplus_{1\leq i\leq s}\mathfrak{A}_i\right)^* \Rrightarrow_{n,k} \left(\bigoplus_{1\leq j\leq t}\mathfrak{B}_i\right)^*$, where $\mathfrak{A}_i \equiv_{n,k} \mathfrak{A}$ and $\mathfrak{B}_i \equiv_{n,k} \mathfrak{B}$ for all i.*

Proof. The proof is by induction on n. Suppose $n = 1$ and Spoiler plays a move choosing k elements from $\bigoplus_{1\leq i\leq s}\mathfrak{A}_i$. Suppose these are chosen from $\mathfrak{A}_{i_1},\ldots,\mathfrak{A}_{i_l}$ with $1 \leq i_1 < \cdots < i_l \leq s$ for some $l \leq k$. Further, let $i_0 = 0$ and $i_{l+1} = t$. Since $l+2 \leq k+2 < 2k+2 = \rho(1,k) \leq s$, there is some p such that $i_{p+1} > i_p+1$. Duplicator must choose structures $(\mathfrak{B}_{j_q})_{1\leq q\leq l}$ in which to respond. Moreover, since $\mathfrak{A}_i$ and $\mathfrak{A}_{i+1}$ share an element for all i, whenever $i_{q+1} = i_q + 1$, we must choose $j_{q+1} = j_q + 1$.

Duplicator chooses values $0 = j_0 < j_1 < \cdots < j_l \leq j_{l+1} = t$ as follows. For all values of q from 0 to $p-1$, choose $j_{q+1} = j_q + 1$ if $i_{q+1} = i_q + 1$ and choose $j_{q+1} = j_q + 2$ otherwise. For all values of q from l down to $p+1$, choose $j_q = j_{q+1} - 1$ if $i_q = i_{q+1} - 1$ and choose $j_q = j_{q+1} - 2$ otherwise. Because $t \geq 2k+2$, this guarantees that $j_{p+1} > j_p + 1$. Thus, Duplicator can respond to the elements picked in $\mathfrak{A}_{i_p}$ with elements in $\mathfrak{B}_{j_p}$ by composing the winning strategies for $\mathfrak{A}_{i_p} \equiv_{1,k} \mathfrak{A} \Rrightarrow_{1,k} \mathfrak{B} \equiv_{1,k} \mathfrak{B}_{j_p}$ and this is a winning response. Moreover, by construction, this strategy maps the minimum and maximum elements of $\bigoplus_{1\leq i\leq s}\mathfrak{A}_i$ to the corresponding elements of $\bigoplus_{1\leq i\leq t}\mathfrak{B}_j$.

Now suppose $n \geq 2$ and the statement has been proved for $n-1$. Let Spoiler play a move choosing k elements from $\bigoplus_{1\leq i\leq s}\mathfrak{A}_i$, and again say these are chosen from $\mathfrak{A}_{i_1},\ldots,\mathfrak{A}_{i_l}$ with $1 \leq i_1 < \cdots < i_l \leq s$ for some $l \leq k$. Further, define $i_0 = 0$ and $i_{l+1} = s$. Since $l+2 \leq k+2$ and $t \geq \rho(n,k) = (k+2)(\rho(n-1,k)+1)$, Duplicator can choose indices $0 = j_0 < j_1 \cdots < j_l < j_{l+1} = t$ so that for all p either $j_{p+1} - j_p = i_{p+1} - i_p$ or $(j_{p+1} - j_p), (i_{p+1} - i_p) \geq \rho(n-1,k)+1$. Now choose, for each p with $1 \leq p \leq l$ a response $\bar{b}_p$ for Duplicator in $\mathfrak{B}_{j_p}$ to the elements $\bar{a}_p$ chosen by Spoiler in $\mathfrak{A}_{i_p}$. We claim that

$$\Big(\bigoplus_{1\leq j\leq t}\mathfrak{B}_j,(\bar{b}_p)_{1\leq p\leq l}\Big)^* \Rrightarrow_{n-1,k} \Big(\bigoplus_{1\leq i\leq s}\mathfrak{A}_i,(\bar{a}_p)_{1\leq p\leq l}\Big)^*.$$

To prove this, it suffices to show for each p with $0 \leq p \leq l$ that

▷ Claim 8.

$$\Big(\bigoplus_{j_p<j\leq j_{p+1}}(\mathfrak{B}_j,\bar{b}_p)\Big)^* \Rrightarrow_{n-1,k} \Big(\bigoplus_{i_p<i\leq i_{p+1}}(\mathfrak{A}_i,\bar{a}_p)\Big)^*,$$

for then the claim follows by l applications of Lemma 6. Note that we have,

1. for all i,j that $\mathfrak{B}_j \Rrightarrow_{n-1,k} \mathfrak{A}_i$ by the assumption that $\mathfrak{A} \Rrightarrow_{n,k} \mathfrak{B}$; and
2. for all p we have $(\mathfrak{B}_{j_p},\bar{b}_p) \Rrightarrow_{n-1,k} (\mathfrak{A}_{i_p},\bar{a}_p)$ by the choice of $\bar{b}_p$ as Duplicator's winning response to Spoiler's choice of $\bar{a}_p$.

Now, for each value of p there are two possibilities:

case (i): $j_{p+1} - j_p = i_{p+1} - i_p$. In this case, the two sides of Claim 8 are the ordered sums of sequences of equal length. The corresponding pieces are all related by $\Rrightarrow_{n-1,k}$, either by 1. above for all except the last piece or by 2. for the last piece. Thus, Claim 8 is established by application of Lemma 6;

case (ii): $(j_{p+1}-j_p), (i_{p+1}-i_p) \geq \rho(n-1,k)+1$. In this case, the structures on the two sides of Claim 8 can be expressed as $\big(\bigoplus_{j_p<j\leq j_{p+1}-1} \mathfrak{B}_j\big) \oplus (\mathfrak{B}_{j_{p+1}}, \bar{b}_{p+1})$ and $\big(\bigoplus_{i_p<i\leq i_{p+1}-1} \mathfrak{A}_i\big) \oplus (\mathfrak{A}_{i_{p+1}}, \bar{a}_{p+1})$, respectively. Since $(j_{p+1}-j_p-1), (i_{p+1}-i_p-1) \geq \rho(n-1,k)$, we have by the induction hypothesis and 1. that $\big(\bigoplus_{j_p<j\leq j_{p+1}-1} \mathfrak{B}_j\big)^* \Rrightarrow_{n-1,k} \big(\bigoplus_{i_p<i\leq i_{p+1}-1} \mathfrak{A}_i\big)^*$. This, together with 2. and Lemma 6 establishes Claim 8, and hence the inductive step of the proof.

◀

Besides ordered sums, another key step in the inductive constructions of our structures is adding unary relations which include the minimum and maximum elements of a structure and adding binary relations which relate the minimum and maximum elements. These operations also behave well with respect to games. To be precise, suppose U is a unary relation symbol and T a binary relation symbol. Let $\mathfrak{A}$ be an ordered structure with minimum and maximum elements min and max respectively. Write $\mathfrak{A}_U$ for the structure obtained from $\mathfrak{A}$ by including min and max in the interpretation of U. Similarly, write $\mathfrak{A}_T$ for the structure obtained from $\mathfrak{A}$ by adding the pair $(\min, \max)$ to the interpretation of T. Note that we do not assume that U or T are in the vocabulary of $\mathfrak{A}$. If they are not, then their interpretations in $\mathfrak{A}_U$ and $\mathfrak{A}_T$ respectively contain nothing other than the elements added.

▶ **Lemma 9.** *Let $n, k \geq q$. If $\mathfrak{A}$ and $\mathfrak{B}$ are ordered structures for which $\mathfrak{A}^* \Rrightarrow_{n,k} \mathfrak{B}^*$ then $\mathfrak{A}_U \Rrightarrow_{n,k} \mathfrak{B}_U$ and $\mathfrak{A}_T \Rrightarrow_{n,k} \mathfrak{B}_T$.*

Proof. This is immediate from the fact that a Duplicator winning strategy between $\mathfrak{A}^*$ and $\mathfrak{B}^*$ must map the minimum elements of the two structures to each other, and similarly for the maximum. ◀

We are now ready to start inductively constructing the Duplicator winning strategy that establishes Lemma 5. We begin with games on some simple structures. For any $m \geq 1$ write L_m for the structure with exactly m elements and two binary relations $\leq$ and S where $\leq$ is a linear order and S the corresponding successor relation.

▶ **Lemma 10.** *If $m_1, m_2 > \rho(n,k)$ then $L^*_{m_1} \Rrightarrow_{n,k} L^*_{m_2}$.*

Proof. Note that L_m is the ordered sum of a sequence of $m-1$ copies of L_2, so the result follows immediately from Lemma 7. ◀

Without loss of generality, assume that the universe of L_m is $\{1,\ldots,m\}$ and write G_m for the structure obtained from L_{m+1} by deleting the element $\lceil \frac{m}{2} \rceil$. Note that G_m is isomorphic to the structure obtained from L_m by removing from the relation S the pair $(\lceil \frac{m}{2} \rceil - 1, \lceil \frac{m}{2} \rceil)$.

▶ **Lemma 11.** *If $m_1, m_2 \geq 2k+2$ then $G^*_{m_1} \Rrightarrow_{1,k} L^*_{m_2}$.*

Proof. Since $G^*_{m_1}$ is a substructure of $L^*_{m_1+1}$, every existential sentence true in the former is also satisfied in the latter. Now, since $L^*_{m_1+1} \Rrightarrow_{1,k} L^*_{m_2}$ by Lemma 10, the result follows. ◀

The next two lemmas give us the base case of the inductive proof of Lemma 5.

▶ **Lemma 12.** $\mathsf{Tot}_{1,k} \Rrightarrow_{2,k} \mathsf{Gap}_{1,k}$

Proof. Recall that the $\{\leq, S\}$-reduct of $\mathsf{Tot}_{1,k}$ is the structure L_m for $m = 6(k+2)^2$. Note that $m > 2\rho(2,k) + (k+2)(2k+4)$. We think of this as composed of three segments: the first $\rho(2,k)$ elements; the last $\rho(2,k)$ elements and a *middle segment* containing the remainder. Suppose now that Spoiler chooses k elements $a_1 < \cdots < a_k$ from $\mathsf{Tot}_{1,k}$ in the first round of the game. Since the middle segment contains more than $(k+2)(2k+4)$ elements, it must contain an interval $[x, y]$ of $2k+2$ consecutive elements which are not chosen. Let us say that $a_i < x$ and $y < a_{i+1}$. Thus, we can write the $\{\leq, S\}$-reduct of $\mathsf{Tot}_{1,k}$ with the chosen elements as

$$(L_{m_1}, a_1, \ldots, a_i) \oplus L_{m_2} \oplus (L_{m_3}, a_{i+1}, \ldots, a_k),$$

where $m_1, m_3 \geq \rho(2, k)$ and $m_2 \geq 2k+2$.

Consider now $\mathsf{Gap}_{1,k}$. The $\{\leq, S\}$-reduct of this structure is G_m, which can also be written as $L_{\frac{m}{2}-k-1} \oplus G_{2k+2} \oplus L_{\frac{m}{2}-k-1}$. Since $\frac{m}{2} - k - 1 > \rho(2,k)$, we have by Lemma 10 that $L^*_{m_1} \Rrightarrow_{2,k} L^*_{\frac{m}{2}-k-1}$ and hence there is a choice of elements $b_1, \ldots, b_i$ such that $(L_{\frac{m}{2}-k-1}, b_1, \ldots, b_i)^* \Rrightarrow_{1,k} (L_{m_1}, a_1, \ldots, a_i)^*$. Similarly, there is a choice of elements $b_{i+1}, \ldots, b_k$ such that $(L_{\frac{m}{2}-k-1}, b_{i+1}, \ldots, b_k)^* \Rrightarrow_{1,k} (L_{m_3}, a_{i+1}, \ldots, a_k)^*$. Further, we know that $G^*_{2k+2} \Rrightarrow_{1,k} L^*_{m_3}$ by Lemma 11. Hence, by Lemma 6 we have that $L^*_m \Rrightarrow_{2,k} G^*_m$. The result now follows by Lemma 9 as $\mathsf{Tot}_{1,k}$ and $\mathsf{Gap}_{1,k}$ are obtained from L_m and G_m respectively by relating the minimum and maximum elements with the relation R. ◀

A similar pattern of argument is repeated in the second base case, and we will be less detailed in spelling it out.

▶ **Lemma 13.** $\mathfrak{N}^*_{1,k} \Rrightarrow_{3,k} \mathfrak{M}^*_{1,k}$.

Proof. Recall that $\mathfrak{N}_{1,k}$ is the ordered sum of $m = 4(k+3)^3 + 2k + 1$ copies of $\mathsf{Gap}_{1,k}$. So $\mathfrak{N}_{1,k} = \bigoplus_{i \in [m]} \mathfrak{G}_i$. Note that $m > 2\rho(3,k) + k + 1$. Suppose now that Spoiler chooses k elements $a_1 < \cdots < a_k$ in the first round of the game. Thus, there is an index i in the middle segment of $[m]$ of length $k+1$ such that $\mathfrak{G}_i$ does not contain a chosen element and we can write $\mathfrak{N}_{1,k}$ with the chosen elements as $(\bigoplus_{i \in [m_1]} \mathfrak{G}_i, a_1, \ldots, a_j) \oplus \mathsf{Gap}_{1,k} \oplus (\bigoplus_{i \in [m_2]} \mathfrak{G}_i, a_{j+1}, \ldots, a_k)$, where $m_1, m_2 > \rho(3, k)$.

On the other side, $\mathfrak{M}_{1,k} = \bigoplus_{i \in [(m-1)/2]} \mathfrak{G}_i \oplus \mathsf{Tot}_{1,k} \oplus \bigoplus_{i \in [(m-1)/2]} \mathfrak{G}_i$. Since $(m-1)/2 > \rho(3,k)$, by Lemma 7 we have $\left(\bigoplus_{i \in [m_1]} \mathfrak{G}_i\right)^* \Rrightarrow_{3,k} \left(\bigoplus_{i \in [(m-1)/2]} \mathfrak{G}_i\right)^*$ and $\left(\bigoplus_{i \in [m_2]} \mathfrak{G}_i\right)^* \Rrightarrow_{3,k} \left(\bigoplus_{i \in [(m-1)/2]} \mathfrak{G}_i\right)^*$. Thus, we can find elements $b_1, \ldots, b_k$ such that

$$\Big(\bigoplus_{i \in [\frac{m-1}{2}]} \mathfrak{G}_i, b_1, \ldots, b_j\Big)^* \Rrightarrow_{2,k} \Big(\bigoplus_{i \in [m_1]} \mathfrak{G}_i, a_1, \ldots, a_j\Big)^*$$

and

$$\Big(\bigoplus_{i \in [\frac{m-1}{2}]} \mathfrak{G}_i, b_{j+1}, \ldots, b_k\Big)^* \Rrightarrow_{2,k} \Big(\bigoplus_{i \in [m_2]} \mathfrak{G}_i, a_{j+1}, \ldots, a_k\Big)^*.$$

Combining this with the fact that $\mathsf{Tot}_{1,k} \Rrightarrow_{2,k} \mathsf{Gap}_{1,k}$ by Lemma 12, we get by Lemma 6 that $(\mathfrak{M}_{1,k}, b_1, \ldots, b_k)^* \Rrightarrow_{2,k} (\mathfrak{N}_{1,k}, a_1, \ldots, a_k)^*$ and the result follows. ◀

These last two lemmas form the base case of the induction that establishes the main result. Where the argument is analogous to the previous ones, we skim over the details.

Proof of Lemma 5. We prove the following two statements by induction for all $n, k \geq 1$.

1. $\mathsf{Tot}_{n,k} \Rrightarrow_{2n,k} \mathsf{Gap}_{n,k}$
2. $\mathfrak{N}^*_{n,k} \Rrightarrow_{2n+1,k} \mathfrak{M}^*_{n,k}$

The case of $n = 1$ is established in Lemmas 12 and 13 respectively. Suppose now $n \geq 2$ and we have established both statements for $n - 1$.

First, recall that $\mathfrak{M}^+_{n-1,k}$ and $\mathfrak{N}^+_{n-1,k}$ are obtained from $\mathfrak{M}_{n-1,k}$ and $\mathfrak{N}_{n-1,k}$ respectively by including their minimum and maximum elements in the unary relation P_n and the binary relation S_n. Thus, by the induction hypothesis and Lemma 9, we have $\mathfrak{N}^+_{n-1,k} \Rrightarrow_{2n-1,k} \mathfrak{M}^+_{n-1,k}$.

$\mathsf{Tot}_{n,k}$ consists of the ordered sum of a sequence of $m = 4(k+3)^{2n}+2k+1 > 2\rho(2n,k)+k+1$ copies of $\mathfrak{M}^+_{n-1,k}$, along with a relation R_n containing just the pair with the minumum and maximum elements. When Spoiler chooses k elements from this structure, we can find a copy of $\mathfrak{M}^+_{n-1,k}$ in the middle $k+1$ copies that has no element chosen and there are $m_1 > \rho(2n,k)$ copies before it and $m_2 > \rho(2n,k)$ copies after it. We can similarly express $\mathsf{Gap}_{n,k}$ as the ordered sum of a sequence of $(m-1)/2 > \rho(2n,k)$ copies of $\mathfrak{M}^+_{n-1,k}$, followed by a copy of $\mathfrak{N}^+_{n-1,k}$ and a further $(m-1)/2 > \rho(2n,k)$ copies of $\mathfrak{M}^+_{n-1,k}$. Lemma 7 tells us that we can find a response to the chosen elements in the first and third parts. This combined with the fact that $\mathfrak{N}^+_{n-1,k} \Rrightarrow_{2n-1,k} \mathfrak{M}^+_{n-1,k}$ and using Lemma 9 to expand to the relation R_n gives us the desired result.

The argument for the second statement is entirely analogous. $\mathfrak{N}_{n,k}$ is the ordered sum of a sequence of $m = 4(k+3)^{2n+1} + 2k + 1 > 2\rho(2n+1,k) + k + 1$ copies of $\mathsf{Gap}_{n,k}$. When Spoiler chooses k elements from this structure, we can find a copy of $\mathsf{Gap}_{n,k}$ in the middle $k+1$ copies that has no element chosen and there are $m_1 > \rho(2n+1,k)$ copies before it and $m_2 > \rho(2n+1,k)$ copies after it. Since $\mathfrak{M}_{n,k}$ has $(m-1)/2 > \rho(2n+1,k)$ copies of $\mathsf{Gap}_{n,k}$ followed by a copy of $\mathsf{Tot}_{n,k}$ and a further $(m-1)/2 > \rho(2n+1,k)$ copies of $\mathsf{Gap}_{n,k}$, by Lemma 7 we can find responses to the chosen elements in the first and third parts. We have already proved that $\mathsf{Tot}_{n,k} \Rrightarrow_{2n,k} \mathsf{Gap}_{n,k}$. We can combine these to complete the proof. ◀

5 Concluding Remarks

We have established in this paper that the extension-closed properties of finite structures that are definable in first-order logic are not contained in any fixed quantifier-alternation fragment of the logic. The construction of the sentences demonstrating this is recursive. It builds on a base of known counter-examples for the Łoś-Tarski theorem in the finite and lifts them up inductively. Also, the argument for showing inexpressibility in fixed levels of the quantifier-alternation hierarchy builds on the game arguments used with previously known examples and builds on them systematically using a form of Feferman-Vaught decomposition (see [11]) for ordered sums of structures.

Our result actually establishes that for all *odd* $n > 1$, there is a Σ_n-definable property that is closed under extensions but not definable by a Π_n sentence. Interestingly, this is not true for even values of n. In particular, it is known that every Σ_2-definable extension-closed property is already definable in Σ_1. This observation is credited to Compton in [7] and may also be found in [16]. We do not know, however, whether for even $n > 2$, the extension closed properties in Σ_n can all be expressed in Π_n or even Σ_{n-1}. On the other hand, we are able to observe that for all even $n > 1$, there is a Π_n-definable extension-closed property that is not in Σ_n. This is a direct consequence of our proof. Indeed, consider the sentence φ_n obtained from $\mathrm{SomeTotalR}_n$ by removing the two outer existential quantifiers and replacing the resulting free variables with new constants a and b. This is a Π_{2n} sentence and is easily seen to define an extension-closed class. By our construction, φ_n is satisfied in the expansion of $\mathsf{Tot}_{n,k}$ where a and b are the minimum and maximum elements respectively. At the same time φ_n is false in the similar expansion of $\mathsf{Gap}_{n,k}$. Since we showed that $\mathsf{Tot}^*_{n,k} \Rrightarrow_{2n,k} \mathsf{Gap}^*_{n,k}$, it follows that φ_n is not equivalent to a Σ_{2n} sentence. It would be interesting to complete the picture of extension-closed properties in the remaining quantifier-alternation fragments, specifically Σ_n for even values of n and Π_n for odd values of n.

It is a feature of our construction that the vocabulary σ_n in which we construct the sentences which separate extension-closed Σ_{2n+1} from Π_{2n+1} grows with n. Could our results be established in a fixed vocabulary? Indeed, does something like Theorem 4 hold for finite graphs?

Another interesting direction left open from our work is the relation with Datalog(¬). All the extension-closed FO-definable properties we construct are also definable in Datalog(¬). Rosen and Weinstein [13] ask whether this is true for all FO-definable extension-closed properties, and this remains open. Indeed, it is conceivable that we have an extension-preservation theorem for least fixed-point logic in the finite, so that even all LFP-definable extension-closed properties are in Datalog(¬).

References

1 Serge Abiteboul, Richard Hull, and Victor Vianu. *Foundations of databases*, volume 8. Addison-Wesley Reading, 1995.

2 Albert Atserias, Anuj Dawar, and Martin Grohe. Preservation under extensions on well-behaved finite structures. *SIAM Journal on Computing*, 38:1364–1381, 2008.

3 Chen C. Chang and Howard J. Keisler. *Model Theory*, volume 73. Elsevier, 1990.

4 Anuj Dawar. Finite model theory on tame classes of structures. In *MFCS*, volume 4708 of *Lecture Notes in Computer Science*, pages 2–12. Springer, 2007.

5 Anuj Dawar and Stephan Kreutzer. On Datalog vs. LFP. In *International Colloquium on Automata, Languages, and Programming*, pages 160–171. Springer, 2008.

6 Heinz-Dieter Ebbinghaus and Jörg Flum. *Finite model theory*. Springer Science & Business Media, 2005.

7 Yuri 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.

8 Wilfrid Hodges. *Model theory*, volume 42. Cambridge University Press, 1993.

9 Phokion G Kolaitis and Moshe Y Vardi. On the expressive power of Datalog: tools and a case study. *Journal of Computer and System Sciences*, 51(1):110–134, 1995.

10 Leonid Libkin. *Elements of Finite Model Theory*. Springer-Verlag, 2004.

11 Johann A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. *Ann. Pure Appl. Log.*, 126:159–213, 2004. doi:10.1016/j.apal.2003.11.002.

12 Eric Rosen. Some aspects of model theory and finite structures. *Bulletin of Symbolic Logic*, 8:380–403, 2002.

13 Eric Rosen and Scott Weinstein. Preservation theorems in finite model theory. In *Logical and Computational Complexity. Selected Papers.*, pages 480–502, 1994. doi:10.1007/3-540-60178-3_99.

14 Benjamin Rossman. Homomorphism preservation theorems. *Journal of the ACM*, 55, 2008.

15 Abhisekh Sankaran. Revisiting the generalized Łoś-Tarski theorem. In *Logic and Its Applications - 8th Indian Conference, ICLA 2019*, pages 76–88, 2019. doi:10.1007/978-3-662-58771-3_8.

16 Abhisekh Sankaran, Bharat Adsul, Vivek Madan, Pritish Kamath, and Supratik Chakraborty. Preservation under substructures modulo bounded cores. In *Logic, Language, Information and Computation - 19th International Workshop, WoLLIC 2012*, pages 291–305, 2012. doi:10.1007/978-3-642-32621-9_22.

17 William W. Tait. A counterexample to a conjecture of Scott and Suppes. *J. Symb. Logic*, 24(1):15–16, 1959.