Complete Axiomatizations of MSO, $\mathsf{FO(TC^1)}$ and $\mathsf{FO(LFP^1)}$ on Finite Trees *

Amélie Gheerbrant, a.gheerbrant@uva.nl, ILLC, Amsterdam
Balder ten Cate, balder.tencate@uva.nl, ISLA, Amsterdam

October 15, 2008

Abstract

We propose axiomatizations of monadic second-order logic (MSO), monadic transitive closure logic ($\mathsf{FO(TC^1)}$) and monadic least fixpoint logic ($\mathsf{FO(LFP^1)}$) on finite node-labeled sibling-ordered trees. We show by a uniform argument, that our axiomatizations are complete, i.e., in each of our logics, every formula which is valid on the class of finite trees is provable using our axioms. We are interested in this class of structures because it allows to represent basic structures of computer science such as XML documents, linguistic parse trees and treebanks. The logics we consider are rich enough to express interesting properties such as reachability. On arbitrary structures, they are well known to be not recursively axiomatizable.

We develop a uniform method for obtaining complete axiomatizations of fragments of MSO on trees. In particular, we obtain a complete axiomatization for MSO, $\mathsf{FO(TC^1)}$, and $\mathsf{FO(LFP^1)}$ on finite node labeled sibling-ordered trees. We take inspiration from Kees Doets, who proposed in [4] a complete axiomatization of first-order logic (FO) on the class of node-labeled finite trees without sibling-order. A similar result was shown in [1] and [19] for FO on node-labeled finite trees with sibling order. We use the signature of [19] and extend the set of axioms proposed there.

Finite trees are basic and ubiquitous structures which are of interest at least to mathematicians, computer scientists (tree-structured documents) and linguists (parse trees). The logics we study are known to be very well-behaved on this particular class of structures and to have an interestingly high expressive power. In particular, they all allow to express reachability, but at the same time, they have the advantage of being decidable on trees.

As XML documents are tree-structured data, our results are particularly relevant to XML query languages. Query languages are logical languages used to make queries into database and information systems. In [20] and [8], MSO and $\mathsf{FO(TC^1)}$ have been proposed as a yardstick of expressivity on trees for these languages. It is known that $\mathsf{FO(LFP^1)}$ has the same expressive power as MSO on trees, but the translations between the two are non-trivial, and hence it is not clear whether an axiomatization for one language can be obtained from an axiomatization for the other language in any straightforward way.

*We are grateful to Jouko Väänänen for helpful comments on an earlier draft. The authors are supported by a GLoRiClass fellowship of the European Commission (Research Training Fellowship MEST-CT-2005-020841) and by the Netherlands Organization for Scientific Research (NWO) grant 639.021.508, respectively.

In applications to computational linguistics, finite trees are used to represent the grammatical structure of natural language sentences. In the context of *model theoretic syntax*, Rogers advocates in [18] the use of MSO in order to characterize derivation trees of context free grammars. Kepser also argues in [12] that MSO should be used in order to query treebanks. A treebank is a text corpus in which each sentence has been annotated with its syntactic structure (represented as a tree structure). In [13] and [21] Kepser and Tiede propose to consider various transitive closure logics, among which $\mathsf{FO}(\mathsf{TC}^1)$, arguing that they constitute very natural formalisms from the logical point of view, allowing concise and intuitive phrasing of parse tree properties.

The remainder of the paper is organized as follows: in Section 1 we present the concept of finite tree and the logics we are interested in together with their standard interpretation. Section 2 merely states our three axiomatizations. In Section 3, we introduce non standard semantics called *Henkin semantics*, for which our axiomatizations are easily seen to be complete. Section 4 introduces operations on Henkin structures: substructure formation and a general operation of Henkin structures combination. We obtain Feferman-Vaught theorems for this operation by means of Ehrenfeucht-Fraïssé games. In Section 5, we prove *real* completeness (that is, on the restricted class of finite trees). For that purpose, we consider substructures of trees that we call forests and use the general operation discussed in Section 4 to combine a set of forests into one new forest. Our Feferman-Vaught theorems apply to such constructions and we use them in our main proof of completeness, showing that no formula of our language can distinguish Henkin models of our axioms from real finite trees. We also point out that every standard model of our axioms actually *is* a finite tree.

We provide additional proofs in Appendix. Appendix A contains proofs of Henkin completeness theorems for our logics. Appendix B contains the proof a relativization lemma that we use in Section 4 and Section 5.2 in order to show that whenever a property is definable in a substructure of some given structure, then it is also definable in this structure. Appendix C contains the definitions and adequacy proofs of three Ehrenfeucht-Fraïssé games that we use in Appendix D to prove our Feferman-Vaught theorems.

1 Preliminaries

1.1 Finite Trees

A tree is a partially ordered set such that the set of predecessors of any element (or *node*) is well-ordered (a set is well-ordered if all its non-empty subsets have a least element) and there is a unique smallest element called the root. We are interested in *finite node-labeled sibling-ordered trees*: finite trees in which the children of each node are linearly ordered. Also, the nodes can be labeled by unary predicates. We will call these structures *finite trees* for short.

Definition 1 (Finite tree)**.** Assume a fixed finite set of unary predicate symbols $\{P_1, \ldots, P_n\}$. By a finite tree, we mean a finite structure $\mathfrak{M} = (M, <, \prec, P_1, \ldots, P_n)$, where $(M, <)$ is a tree (with $<$ the descendant relation) and $\prec$ linearly orders the children of each node.

1.2 Three Extensions of First-Order Logic

In this section, we introduce three extensions of FO: MSO, FO(TC^1) and FO(LFP^1). In the remaining of the paper (unless explicitly stated otherwise), we will always be working with a fixed purely relational vocabulary σ (i.e. with no individual constant or function symbols) and hence, with σ-structures. We assume as usual that we have a countably infinite set of first-order variables. In the case of MSO and FO(LFP^1), we also assume that we have a countably infinite set of set variables. The semantics defined in this section we will refer to as *standard semantics* and the associated structures, as *standard structures.*

We first introduce monadic second order logic, MSO, which is the extension of first-order logic in which we can quantify over the subsets of the domain.

Definition 2 (Syntax and semantics of MSO)**.** Let At be a first-order atomic formula, x a first-order variable and X a set variable, we define the set of MSO formulas in the following way:

$$\phi := At \mid Xx \mid \phi \wedge \psi \mid \phi \vee \psi \mid \phi \rightarrow \psi \mid \neg\phi \mid \exists x\ \phi \mid \exists X\ \phi$$

We use $\forall X\phi$ (resp. $\forall x\phi$) as shorthand for $\neg\exists X\neg\phi$ (resp. $\neg\exists x\neg\phi$). We define the *quantifier depth* of a MSO formula as the maximal number of first-order and second-order nested quantifiers. We interpret MSO formulas in first-order structures. Like for FO formulas, the truth of MSO formulas in $\mathfrak{M}$ is defined modulo a valuation g of variables as objects. But here, we also have set variables, to which g assigns subsets of the domain. We let $g[a/x]$ be the assignment which differs from g only in assigning a to x (similarly for $g[A/X]$). The truth of atomic formulas is defined by the usual FO clauses plus the following:

$$\mathfrak{M}, g \models Xx \text{ iff } g(x) \in g(X) \text{ for } X \text{ a set variable}$$

The truth of compound formulas is defined by induction, with the same clauses as in FO and an additional one:

$$\mathfrak{M}, g \models \exists X\phi \text{ iff there is } A \subseteq M \text{ such that } \mathfrak{M}, g[A/X] \models \phi$$

The second logic we are interested in is monadic transitive closure logic, FO(TC^1), which extends FO by closing it under the transitive closure of binary definable relations.

Definition 3 (Syntax and semantics of FO(TC^1))**.** Let u, v, x, y be first-order variables, $\phi(x, y)$ a FO(TC^1) formula (which, besides x and y, possibly contains other free variables), we define the set of FO(TC^1) formulas in the following way:

$$\phi := At \mid Xx \mid \phi \wedge \psi \mid \phi \vee \psi \mid \phi \rightarrow \psi \mid \neg\phi \mid \exists x\ \phi \mid [TC_{xy}\phi(x, y)](u, v)$$

We use $\forall x\phi$ as shorthand for $\neg\exists x\neg\phi$. We define the *quantifier depth* of a FO(TC^1) formula as the maximal number of nested first-order quantifiers and TC operators. We interpret FO(TC^1) formulas in first-order structures. The notion of assignation and the truth of atomic formulas is defined as in FO. The truth of compound formulas is defined by induction, with the same clauses as in FO and an additional one:

$$\mathfrak{M}, g \models [TC_{xy}\phi](u, v)$$
$$\text{iff}$$
$$\text{for all } A \subseteq M, \text{ if } g(u) \in A$$
$$\text{and for all } a, b \in M,\ a \in A \text{ and } \mathfrak{M}, g[a/x, b/y] \models \phi(x, y) \text{ implies } b \in A,$$
$$\text{then } g(v) \in A.$$

Proposition 1. *On standard structures, the following semantical clause for the TC operator is equivalent to the one given above:*

$$\mathfrak{M}, g \models [TC_{xy}\phi(x,y)](u,v)$$
$$\textit{iff}$$
$$\textit{there exist } a_1 \dots a_n \in M \textit{ with } g(u) = a_1 \textit{ and } g(v) = a_n$$
$$\textit{and } \mathfrak{M}, g \models \phi(a_i, a_{i+1}) \textit{ for all } 0 < i < n$$

Proof. Indeed, suppose there is a finite sequence of points $a_1 \dots a_n$ such that $g(u) = a_1$, $g(v) = a_n$, and for each $i < n$, $\mathfrak{M}, g[x/a_i; y/a_{i+1}] \models \phi$. Then for any subset A containing a_1 and which is closed under ϕ, we can show by induction on the length of the sequence $a_1 \dots a_n$ that a_n belongs to A. Now, on the other hand, suppose that there is no finite sequence like described above. To show that there is a subset A of the required form, we simply take A to be the set of all points that "can be reached from u by a finite sequence". By assumption, v does not belong to this set and the set is closed under ϕ. □

Intuitively this means that for a formula of the form $[TC_{xy}\phi](u,v)$ to hold on a standard structure, there must be a *finite* "ϕ path" between the points that are named by the variables u and v.

Finally we will also be interested in monadic least fixpoint logic ($\mathsf{FO(LFP^1)}$), which extends FO with set variables and an explicit monadic least fixpoint operator. Consider a $\mathsf{FO(LFP^1)}$ formula $\phi(X,x)$ and a structure $\mathfrak{M}$ together with a valuation g. This formula induces an operator F_ϕ taking a set $A \subseteq dom(\mathfrak{M})$ to the set $\{a : \mathfrak{M}, g[a/X, A/X] \models \phi\}$. $\mathsf{FO(LFP^1)}$ is concerned with *least fixpoints* of such operators. If ϕ is positive in X (a formula is positive in X whenever X only occurs in the scope of an even number of negations), the operator F_ϕ is monotone (i.e. $X \subseteq Y$ implies $F_\phi(X) \subseteq F_\phi(Y)$). Monotone operators always have a least fixpoint $LFP(F) = \bigcap\{X|F(X) \subset X\}$ (defined as the intersection of all their prefixpoints).

Definition 4 (Syntax and semantics of $\mathsf{FO(LFP^1)}$)**.** Let X be a set variable, x, y FO variables, ψ, ξ $\mathsf{FO(LFP^1)}$ formulas and $\phi(x,X)$ a $\mathsf{FO(LFP^1)}$ formula positive in X (besides x and X, $\phi(x,X)$ possibly contains other free variables), we define the set of $\mathsf{FO(LFP^1)}$ formulas in the following way:

$$\psi := At \mid Xy \mid \psi \wedge \xi \mid \psi \vee \xi \mid \psi \rightarrow \xi \mid \neg\psi \mid \exists x\ \psi \mid [LFP_{x,X}\phi(x,X)]y$$

We use $\forall x \psi$ as shorthand for $\neg\exists x \neg\psi$. We define the *quantifier depth* of a $\mathsf{FO(LFP^1)}$ formula as the maximal number of nested first-order quantifiers and LFP operators. Again, we can interpret $\mathsf{FO(LFP^1)}$ formulas in first-order structures. The notion of assignation and the truth of atomic formulas are defined similarly as in the MSO case. The truth of compound formulas is defined by induction, with the same clauses as in FO and an additional one:

$$\mathfrak{M}, g \models [LFP_{x,X}\phi]y$$
$$\text{iff}$$
$$\text{for all } A \subseteq dom(\mathfrak{M}), \text{ if for all } a \in dom(\mathfrak{M}),\ \mathfrak{M}, g[a/x, A/X] \models \phi(x,X) \text{ implies } a \in A,$$
$$\text{then } g(y) \in A.$$

Remark 1. *In practice we will use an equivalent (less intuitive but often more convenient) rephrasing:*

$$\mathfrak{M}, g \models [LFP_{x,X}\phi]y$$
$$\textit{iff}$$
$$\textit{for all } A \subseteq dom(\mathfrak{M}),\ \textit{if } g(y) \notin A,$$
$$\textit{then there exists } a \in dom(\mathfrak{M}) \textit{ such that } a \notin A \textit{ and } \mathfrak{M}, g[a/x, A/X] \models \phi(x,X).$$

1.3 Expressive Power

There is a recursive procedure, transforming any $\mathsf{FO(LFP^1)}$ formula ϕ into a MSO formula ϕ' such that $\mathfrak{M}, g \models \phi$ iff $\mathfrak{M}, g \models \phi'$. The interesting clause is $([LFP_{x,X}\phi(x,X)]y)' = \forall X(\forall x(\phi(x,X)' \rightarrow Xx) \rightarrow Xy)$. (The other ones are all of the same type, e.g. $(\phi \wedge \psi)^* = (\phi^* \wedge \psi^*)$.) This procedure can easily be seen adequate by considering the semantical clause for the LFP operator.

Now there is also a recursive procedure transforming any $\mathsf{FO(TC^1)}$ formula ϕ into a $\mathsf{FO(LFP^1)}$ formula ϕ'' such that $\mathfrak{M}, g \models \phi$ iff $\mathfrak{M}, g \models \phi''$. The interesting clause is $([TC_{xy}\phi](u,v))'' = [LFP_{Xy} y = u \vee \exists x((Xx \wedge \phi(x,y)''))]v$. Let us give an argument for this claim. By Proposition 1 it is enough to show that $[LFP_{Xy} y = u \vee \exists x(Xx \wedge \phi(x,y)'')]v$ holds if and only if there is a finite ϕ'' path from u to v. For the right to left direction, suppose there is such a path $a_1 \dots a_n$ with $g(u) = a_1$ and $g(v) = a_n$. Then, for any subset A of the domain, we can show by induction on i that if for all a_i $(1 \leq i \leq n)$, $a_i = u \vee \exists x((Ax \wedge \phi(x,a_i)'')$ implies $a_i \in A$, then $v \in A$, i.e., $[LFP_{Xy} y = u \vee \exists x((Xx \wedge \phi(x,y)''))]v$ holds. Now for the left to right direction, suppose there is no such ϕ'' path. Consider the set A of all points that can be reached from u by a finite ϕ'' path. By assumption, $\neg Av$ and it holds that $\forall y((y = u \vee \exists x(Ax \wedge \phi(x,y)'')) \rightarrow Ay)$, i.e., $\neg[LFP_{Xy} y = u \vee \exists x(Xx \wedge \phi(x,y)'')]v$.

It is known that on arbitrary structures $\mathsf{FO(TC^1)} < \mathsf{FO(LFP^1)} < \mathsf{MSO}$ (see [5]) and on trees $\mathsf{FO(TC^1)} <_{trees} \mathsf{FO(LFP^1)} =_{trees} \mathsf{MSO}$ (see [20] and [17]). It is also known that the (not FO definable) class of finite trees is already definable in $\mathsf{FO(TC^1)}$ (see for instance [13]), which is the weakest of the logics studied here. We provide additional detail in Section 5.3.

2 The Axiomatizations

As many arguments in this paper equally hold for MSO, $\mathsf{FO(TC^1)}$ and $\mathsf{FO(LFP^1)}$, we let $\Lambda \in \{\mathsf{MSO}, \mathsf{FO(TC^1)}, \mathsf{FO(LFP^1)}\}$ and use Λ as a symbol for any one of them. The axiomatization of Λ on finite trees consists of three parts: the axioms of first-order logic, the specific axioms of Λ, and the specific axioms on finite trees.

To axiomatize FO, we adopt the infinite set of logical axioms and the two rules of inference given in Figure 1 (like in [6], except from the fact that we use a generalization rule). To axiomatize MSO, the axioms and rule of Figure 2 are added to the axiomatization of FO. We call the resulting system $\vdash_{\mathsf{MSO}}$. $COMP.$ stands for "comprehension" by analogy with the comprehension axiom of set theory. $MSO1$ plays a similar role as $FO2$, $MSO2$ as $FO3$ and $MSO3$ as $FO4$. To axiomatize $\mathsf{FO(TC^1)}$, the axiom and rule of Figure 3 are added to the axiomatization of FO. We call the resulting system $\vdash_{\mathsf{FO(TC^1)}}$. To axiomatize $\mathsf{FO(LFP^1)}$, the axiom and rule of Figure 4 are added to the axiomatization of FO. We call the resulting system $\vdash_{\mathsf{FO(LFP^1)}}$. We are interested in axiomatizing Λ on the class of finite trees. For that purpose we restrict the class of considered structures by adding to $\vdash_\Lambda$ the axioms given in Figure 5. We call the resulting system $\vdash_\Lambda^{tree}$. Note that the induction scheme in Figure 5 allows to reason by induction on *properties definable in* Λ only. Also, for technical convenience, we adopt the following convention:

Definition 5. Let Γ be a set of Λ-formulas and ϕ a Λ-formula. By $\Gamma \vdash_\Lambda \phi$ we will always mean that there are $\psi_1, \dots, \psi_n \in \Gamma$ such that $\vdash_\Lambda (\psi_1 \wedge \dots \wedge \psi_n) \rightarrow \phi$.

Now the main result of this paper is that on standard structures, the Λ theory of finite trees is completely axiomatized by $\vdash_\Lambda^{tree}$. In the remaining sections we will progressively build a proof of it.

FO1.	Tautologies of sentential calculus
FO2.	$\vdash \forall x\phi \rightarrow \phi^x_t$, where t is substitutable for x in ϕ
FO3.	$\vdash \forall x(\phi \rightarrow \psi) \rightarrow (\forall x\phi \rightarrow \forall x\psi)$
FO4.	$\vdash \phi \rightarrow \forall x\phi$, where x does not occur free in ϕ
FO5.	$\vdash x = x$
FO6.	$\vdash x = y \rightarrow (\phi \rightarrow \psi)$, where ϕ is atomic and ψ is obtained from ϕ by replacing x in zero or more (but not necessarily all) places by y.
Modus Ponens	if $\vdash \phi$ and $\vdash \phi \rightarrow \psi$, then $\vdash \psi$
FO Generalization	if $\vdash \phi$, then $\vdash \forall x\phi$

Figure 1: Axioms and rules of FO

COMP.	$\vdash \exists X\forall x(Xx \leftrightarrow \phi)$, where X does not occur free in ϕ
MSO1.	$\vdash \forall X\phi \rightarrow \phi[X/T]$, where T (which is either a set variable or a monadic predicate) is substitutable in ϕ for X.
MSO2.	$\vdash \forall X(\phi \rightarrow \psi) \rightarrow (\forall X\phi \rightarrow \forall X\psi)$
MSO3.	$\vdash \phi \rightarrow \forall X\phi$, where X does not occur free in ϕ
MSO Generalization	if $\vdash \phi$, then $\vdash \forall X\phi$

Figure 2: Axiom and inference rule of MSO

FO(TC^1) axiom	$\vdash [TC_{xy}\phi](u,v) \rightarrow ((\psi(u) \wedge \forall x\forall y(\psi(x) \wedge \phi(x,y) \rightarrow \psi(y))) \rightarrow \psi(v))$ where ψ is any FO(TC^1) formula
FO(TC^1) Generalization	if $\vdash \xi \rightarrow ((P(u) \wedge \forall x\forall y(P(x) \wedge \phi(x,y) \rightarrow P(y))) \rightarrow P(v))$, and P does not occur in ξ, then $\vdash \xi \rightarrow [TC_{xy}\phi](u,v)$

Figure 3: Axiom and inference rule of FO(TC^1)

FO(LFP^1) axiom	$\vdash [LFP_{x,X}\phi]y \rightarrow (\forall x(\phi(x,\psi) \rightarrow \psi(x)) \rightarrow \psi(y))$ where ψ is any FO(LFP^1) formula and $\phi(x,\psi)$ is the result of the replacement in $\phi(x,X)$ of each occurrence of X by ψ (renaming variables when needed)
FO(LFP^1) Generalization	if $\vdash \xi \rightarrow (\forall x(\phi(x,P) \rightarrow P(x)) \rightarrow P(y))$, and P positive in ϕ does not occur in ξ, then $\vdash \xi \rightarrow [LFP_{X,x}\phi](y)$

Figure 4: Axiom and inference rule of FO(LFP^1)

T1.	$\forall xyz(x < y \wedge y < z \rightarrow x < z)$	$<$ is transitive
T2.	$\neg\exists x(x < x)$	$<$ is irreflexive
T3.	$\forall xy(x < y \rightarrow \exists z(x <_{imm} z \wedge z \leq y))$	immediate children
T4.	$\exists x \forall y \neg(y < x)$	there is a root
T5.	$\forall xyz(x < z \wedge y < z \rightarrow x \leq y \vee y \leq x)$	linearly ordered ancestors
T6.	$\forall xyz(x \prec y \wedge y \prec z \rightarrow x \prec z)$	$\prec$ is transitive
T7.	$\neg\exists x(x \prec x)$	$\prec$ is irreflexive
T8.	$\forall xy(x \prec y \rightarrow \exists z(x \prec_{imm} z \wedge z \preceq y))$	immediately next sibling
T9.	$\forall x \exists y(y \preceq x \wedge \neg\exists z(z \prec y))$	there is a least sibling
T10.	$\forall xy((x \prec y \vee y \prec x) \leftrightarrow (\exists z(z <_{imm} x \wedge z <_{imm} y) \wedge x \neq y))$	$\prec$ linearly orders *siblings*
T11.	$\forall xy(x = y \vee x < y \vee y < x \vee \exists x'y'(x' < x \wedge y' < y \wedge (x' \prec y' \vee y' \prec x')))$	connectedness
Ind.	$\forall x(\forall y((x < y \vee x \prec y) \rightarrow \phi(y)) \rightarrow \phi(x)) \rightarrow \forall x \phi(x)$	

where

$\phi(x)$ ranges over Λ-formulas in one free variable x

and

$x <_{imm} y$ is shorthand for $x < y \wedge \neg\exists z(z < y \wedge x < z)$,
$x \prec_{imm} y$ is shorthand for $x \prec y \wedge \neg\exists z(x \prec z \wedge z \prec y)$

Figure 5: Specific axioms on finite trees

3 Henkin Completeness

As it is well known, MSO, FO(TC^1) and FO(LFP^1) are highly undecidable on arbitrary standard structures (by arbitrary, we mean any sort of structure: infinite trees, arbitrary graphs, partial orders. . .) and hence not recursively enumerable. So in order to show that our axiomatizations $\vdash_\Lambda^{tree}$ are complete on finite trees, we resort to a special trick, already used by Kees Doets in his PhD thesis [4]. We proceed in two steps. First, we show three Henkin completeness theorems, based on non standard (so called Henkin) semantics for MSO, FO(TC^1) and FO(LFP^1) (on the general topic of Henkin semantics, see [10], the original paper by Henkin and also [16]). Each semantics respectively extends the class of standard structures with non standard (Henkin) MSO, FO(TC^1) and FO(LFP^1)-structures. By the Henkin completeness theorems, our axiomatic systems $\vdash_\Lambda^{tree}$ naturally turn out to be complete on the wider class of their Henkin-models. But by compactness, some of these models are infinite. As a second step, we show in Section 5 that *no Λ-sentence can distinguish between standard and non-standard Λ-Henkin-models among models of our axioms*. This entails that our axioms are complete on the class of (standard) finite trees, i.e., each Λ-sentence valid on this class is provable using $\vdash_\Lambda^{tree}$. Now let us point out that Kees Doets was interested in the completeness of *first-order logic* on finite trees. Thus, he was relying on the FO completeness theorem and if he was working with non-standard models of the FO theory of finite trees, he was not concerned with non standard Henkin-structures in our sense. Hence, what makes the originality of the method developed in this paper is its use of Henkin semantics. So let us begin with the concept of Henkin-structure. Such structures are particular cases among structures called *frames* and it is convenient to define frames before defining Henkin-structures.

Definition 6 (Frames)**.** Let σ be a purely relational vocabulary. A **σ-frame** $\mathfrak{M}$ consists of a non-empty universe $dom(\mathfrak{M})$, an interpretation in $dom(\mathfrak{M})$ of the predicates in σ and a set of admissible subsets $\mathbb{A}_{\mathfrak{M}} \subseteq \wp(dom(\mathfrak{M}))$.

Whenever $\mathbb{A}_{\mathfrak{M}} = \wp(dom(\mathfrak{M}))$, $\mathfrak{M}$ can be identified to a standard structure. Assignments g into $\mathfrak{M}$ are defined as in standard semantics, except that if X is a set variable, then we require that $g(X) \in \mathbb{A}_{\mathfrak{M}}$.

Definition 7 (Interpretation of Λ-formulas in frames)**.** Λ-formulas are interpreted in frames as in standard structures, except for the three following clauses. The set quantifier clause of MSO becomes:

$$\mathfrak{M}, g \models \exists X \phi \text{ iff there is } A \in \mathbb{A}_{\mathfrak{M}_\Gamma} \text{ such that } \mathfrak{M}, g[A/X] \models \phi$$

The TC clause of $\mathsf{FO(TC^1)}$ becomes:

$$\begin{gathered}\mathfrak{M}, g \models [TC_{xy}\phi](u,v)\\ \text{iff}\\ \text{for all } A \in \mathbb{A}_{\mathfrak{M}}, \text{ if } g(u) \in A\\ \text{and for all } a,b \in dom(\mathfrak{M}),\ a \in A \text{ and } \mathfrak{M}, g[x/a, b/y] \models \phi \text{ imply } b \in A,\\ \text{then } g(v) \in A.\end{gathered}$$

And finally the LFP clause of $\mathsf{FO(LFP^1)}$ becomes:

$$\begin{gathered}\mathfrak{M}, g \models [LFP_{x,X}\phi]y\\ \text{iff}\\ \text{for all } A \in \mathbb{A}_{\mathfrak{M}}, \text{ if for all } a \in dom(\mathfrak{M}),\ \mathfrak{M}, g[a/x, A/X] \models \phi(x,X) \text{ implies } a \in A,\\ \text{then } g(y) \in A.\end{gathered}$$

Definition 8 (Λ-Henkin-Structures)**.** A Λ-**Henkin-structure** is a frame $\mathfrak{M}$ that is closed under Λ-definability, i.e., for each Λ-formula φ and assignment g into $\mathfrak{M}$:

$$\{a \in M | \mathfrak{M}, g[a/x] \models \varphi\} \in \mathbb{A}_{\mathfrak{M}}$$

Remark 2. *Note that any finite Λ-Henkin-structure is a standard structure, as every subset of the domain is parametrically definable in a finite structure. Hence, non standard Henkin structures are always infinite.*

Theorem 1. *Λ is completely axiomatized on Λ-Henkin-structures by $\vdash_\Lambda$, i.e., for every set of Λ-formulas Γ and Λ-formula ϕ, ϕ is true in all Λ-Henkin-structures of Γ if and only if $\Gamma \vdash_\Lambda \phi$.*

Proof. The proofs are given in Appendix A (Theorems 5, 6, 7). □

Compactness follows directly from Definition 5 and Theorem 1, i.e., a possibly infinite set of Λ-sentences has a model if and only if every finite subset of it has a model. It also follows directly from Theorem 1 that $\vdash_\Lambda^{tree}$ is complete on the class of its Λ-Henkin-models. Nevertheless, by compactness the axioms of $\vdash_\Lambda^{tree}$ are also satisfied on infinite trees. We overcome this problem by defining a slightly larger class of Henkin structures, which we will call *definably well-founded Λ-quasi-trees.*[1]

Definition 9. A Λ-*quasi-tree* is any Λ-Henkin structure $(T, <, \prec, P_1, \ldots, P_n, \mathbb{A}_T)$ (where $\mathbb{A}_T$ is the set of admissible subsets of T) satisfying the axioms and rules of $\vdash_\Lambda$ and the axioms T1–T11 of Figure 5. A Λ-quasi-tree is *definably well founded* if, in addition, it satisfies all instances of the induction scheme Ind of Figure 5.

Corollary 1. A Λ-Henkin-structure satisfies the axioms of $\vdash_\Lambda^{tree}$ if and only if it is a definably well-founded Λ-quasi-tree.

[1]For a nice picture of a *non* definably well-founded quasi-tree see [1].

4 Operations on Henkin-Structures

Let $\Lambda \in \{\mathsf{MSO}, \mathsf{FO}(\mathsf{TC}^1), \mathsf{FO}(\mathsf{LFP}^1)\}$. As noted in Remark 2, every finite Λ-Henkin structure is also a standard structure. Hence, when working in finite model theory, it is enough to rely on the usual FO constructions to define operations on structures. On the other hand, even though our main completeness result concerns finite trees, inside the proof we need to consider infinite (Λ-Henkin) structures and operations on them. In this context, methods for forming new structures out of existing ones have to be redefined carefully. We first propose a notion of substructure of a Λ-Henkin-structure generated by one of its parametrically definable admissible subsets:

Definition 10 (Λ-substructure)**.** Let $\mathfrak{M} = (dom(\mathfrak{M}), Pred, \mathbb{A}_{\mathfrak{M}})$ be a Λ-Henkin-structure (where $Pred$ is the interpretation of the predicates). We call $\mathfrak{M}_{FO} = (dom(\mathfrak{M}), Pred)$ the FO-structure underlying $\mathfrak{M}$. Given a parametrically definable set $A \in \mathbb{A}_{\mathfrak{M}}$, the Λ-substructure of $\mathfrak{M}$ generated by A is the structure $\mathfrak{M} \restriction A = (\langle A \rangle_{\mathfrak{M}_{FO}}, \mathbb{A}_{\mathfrak{M}\restriction A})$, where $\langle A \rangle_{\mathfrak{M}_{FO}}$ is the FO-substructure of $\mathfrak{M}_{FO}$ generated by A (note that A forms the domain of $\langle A \rangle_{\mathfrak{M}_{FO}}$, as the vocabulary is purely relational) and $\mathbb{A}_{\mathfrak{M}\restriction A} = \{X \cap A | X \in \mathbb{A}_{\mathfrak{M}}\}$.

Proposition 2. *Take $\mathfrak{M}$ and A as previously and consider the structure $(\mathfrak{M} \restriction A)' = (\langle A \rangle_{\mathfrak{M}_{FO}}, \mathbb{A}_{(\mathfrak{M}\restriction A)'})$, where $\mathbb{A}_{(\mathfrak{M}\restriction A)'} = \{X \in \mathbb{A}_{\mathfrak{M}} | X \subseteq A\}$. Whenever $\mathfrak{M}$ is a MSO-Henkin structure or a FO(LFP^1)-Henkin structure, $\mathfrak{M} \restriction A$ and $(\mathfrak{M} \restriction A)'$ are one and the same structure.*

Proof. Indeed, take $B \in \mathbb{A}_{\mathfrak{M}\restriction A}$. So there exists $B' \in \mathbb{A}_{\mathfrak{M}}$ such that $B = B' \cap A$. We want to show that also $B' \cap A \in \mathbb{A}_{(\mathfrak{M}\restriction A)'}$ i.e. $B' \cap A \subseteq A$ (which obviously holds) and $B' \cap A \in \mathbb{A}_{\mathfrak{M}}$. The second condition holds because both B' and A are definable in $\mathfrak{M}$, so their intersection also is ($B' \cap A = \{x \mid \mathfrak{M} \models Ax \wedge B'x\}$). Conversely, consider $B \in \mathbb{A}_{(\mathfrak{M}\restriction A)'}$, so $B \in \mathbb{A}_{\mathfrak{M}}$ (because $B = B \cap A$) and $B \subseteq A$. □

Now, in order to show that Λ-substructures are Henkin-structures, we introduce a notion of *relativization* and a corresponding *relativization lemma*. This lemma establishes that for any Λ-Henkin-structure $\mathfrak{M}$ and Λ-substructure $\mathfrak{M} \restriction A$ of $\mathfrak{M}$ (with A a set parametrically definable in $\mathfrak{M}$), if a set is parametrically definable in $\mathfrak{M} \restriction A$ then it is also parametrically definable in $\mathfrak{M}$. This result will be useful again in Section 5.2.

Definition 11 (Relativization mapping)**.** Given two Λ-formulas ϕ, ψ having no variables in common and given a FO variable x, we define $REL(\phi, \psi, x)$ by induction on the complexity of ϕ and call it the *relativization of ϕ to ψ*:

- If ϕ is an atom, $REL(\phi, \psi, x) = \phi$,
- If $\phi :\approx \phi_1 \wedge \phi_2$, $REL(\phi, \psi, x) = REL(\phi_1, \psi, x) \wedge REL(\phi_2, \psi, x)$ (similar for $\vee, \rightarrow, \neg$),
- If $\phi :\approx \exists y \chi$, $REL(\phi, \psi, x) = \exists y(\psi[y/x] \wedge REL(\chi, \psi, x))$ (where $\psi[y/x]$ is the formula obtained by replacing in ψ every occurrence of x by y),
- If $\phi :\approx \exists Y \chi$, $REL(\phi, \psi, x) = \exists Y((Yx \rightarrow \psi) \wedge REL(\chi, \psi, x))$,
- If $\phi :\approx [TC_{yz}\chi](u, v)$, $REL(\phi, \psi, x) = [TC_{yz} REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]](u, v)$,
- If $\phi :\approx [LFP_{Xy}\chi]z$, $REL(\phi, \psi, x) :\approx [LFP_{Xy}\chi \wedge \psi[y/x]]z$.

Lemma 1 (Relativization lemma)**.** *Let $\mathfrak{M}$ be a Λ-Henkin-structure, g a valuation on $\mathfrak{M}$, ϕ, ψ Λ-formulas and $A = \{x \mid \mathfrak{M}, g \models \psi\}$. If $g(y) \in A$ for every variable y occurring free in ϕ and $g(Y) \in \mathfrak{M} \restriction A$ for every set variable Y occurring free in ϕ, then $\mathfrak{M}, g \models REL(\phi, \psi, x) \Leftrightarrow \mathfrak{M} \restriction A, g \models \phi$.*

Proof. Given in Appendix B (Lemma 13). □

Lemma 2. $\mathfrak{M} \restriction A$ *is a* Λ-Henkin-*structure.*

Proof. Take B parametrically definable in $\mathfrak{M} \restriction A$, i.e., there is a Λ-formula $\phi(y)$ and an assignment g such that $B = \{a \in dom(\mathfrak{M} \restriction A) \mid \mathfrak{M} \restriction A, g[a/y] \models \phi(y)\}$. Now we know that A is also parametrically definable in $\mathfrak{M}$, i.e., there is a Λ-formula $\psi(x)$ and an assignment g' such that $A = \{a \in dom(\mathfrak{M}) \mid \mathfrak{M}, g'[a/x] \models \psi(x)\}$. Assume w.l.o.g. that ϕ and ψ have no variables in common, we define an assignment g^* by letting $g^*(z) = g'(z)$ for every variable z occurring in ψ and $g^*(z) = g(z)$ otherwise. The situation with set variables is symmetric. Now by Lemma 1, $B = \{a \in dom(\mathfrak{M}) \mid \mathfrak{M}, g^*[a/x] \models REL(\phi, \psi, x)\}$ and hence $B \in \mathbb{A}_{\mathfrak{M} \restriction A}$. □

There is in model theory a whole range of methods to form new structures out of existing ones. A standard reference on the matter is [7], written in a very general algebraic setting. Familiar constructions like disjoint unions of FO-structures are redefined as particular cases of a new notion of *generalized product* of FO-structures and abstract properties of such products are studied. In particular, an important theorem now called the Feferman-Vaught theorem for FO is proven. We are particularly interested in one of its corollaries, which establishes that generalized products of FO-structures preserve elementary equivalence. This is related to our work in that we show an analogue of this result for a particular case of generalized product of Λ-Henkin-structures that we call *fusion*, this notion being itself a generalization of a notion of disjoint unions of Λ-Henkin-structures that we also define.

Definition 12 (Disjoint union of Λ-Henkin-structures)**.** Let σ be a purely relational vocabulary and $\sigma^* = \sigma \cup \{Q_1, \ldots, Q_k\}$, with $\{Q_1, \ldots, Q_k\}$ a set of new monadic predicates. For any Λ-Henkin-structures $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$ in vocabulary σ with disjoint domains, define their *disjoint union* $\biguplus_{1 \leq i \leq k} \mathfrak{M}_i$ (or, *direct sum*) to be the σ^*-frame that has as its domain the union of the domains of the structures $\mathfrak{M}_i$ and likewise for the relations, except for the predicates Q_i, whose interpretations are respectively defined as the domain of the structures $\mathfrak{M}_i$ (we will use Q_i to index the elements of M_i). The set of admissible subsets $\mathbb{A}_{\biguplus_{1 \leq i \leq k} \mathfrak{M}_i}$ is the closure under finite union of the union of the sets of admissible subsets of the $\mathfrak{M}_i$. That is:

- $dom(\biguplus_{1 \leq i \leq k} \mathfrak{M}_i) = \bigcup_{1 \leq i \leq k} dom(\mathfrak{M}_i)$
- $P^{\biguplus_{1 \leq i \leq k} \mathfrak{M}_i} = \bigcup_{1 \leq i \leq k} P^{\mathfrak{M}_i}$ (with $P \in \sigma$) and $Q_i^{\biguplus_{1 \leq i \leq k} \mathfrak{M}_i} = dom(\mathfrak{M}_i)$
- $A \in \mathbb{A}_{\biguplus_{1 \leq i \leq k} \mathfrak{M}_i}$ iff $A = \bigcup_{1 \leq i \leq k} A_i$ for some $A_i \in \mathbb{A}_{\mathfrak{M}_i}$

It is shown in Appendix D that disjoint unions of Λ-Henkin-structures are also Λ-Henkin-structures (Corollaries 7, 11, 15).

Definition 13 (f-fusion of Λ-Henkin-structures)**.** Let σ be a purely relational vocabulary and $\sigma^* = \sigma \cup \{Q_1, \ldots, Q_k\}$, with $\{Q_1, \ldots, Q_k\}$ a set of new monadic predicates. Let f be a function mapping each n-ary predicate $P \in \sigma$ to a quantifier-free formula over σ^*

in variables $x_1, \ldots, x_n$. For any Λ-Henkin-structures $\mathfrak{M}_1, \ldots, \mathfrak{M}_k$ in vocabulary σ with disjoint domains, define their f-*fusion* to be the σ-frame $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$ that has the same domain and set of admissible subsets as $\biguplus_{1\leq i\leq k} \mathfrak{M}_i$. For any $P \in \sigma$, the interpretation of P in $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$ is the set of n-tuples satisfying $f(P(x_1 \ldots x_n))$ in $\biguplus_{1\leq i\leq k} \mathfrak{M}_i$.

An easy example of f-fusion on standard structures[2] is the ordered sum of two linear orders $(M_1, <_1), (M_2, <_2)$, where all the elements of M_1 are before the elements of M_2. In this case, σ consists of a single binary relation $<$, the elements of M_1 are indexed with Q_1, those of M_2 with Q_2 and f maps $<$ to $x < y \vee (Q_1 x \wedge Q_2 y)$.

We show preservation results involving f-fusions of Λ-Henkin-structures. Hence we deal with analogues of elementary equivalence for these logics and we refer to Λ-*equivalence*.

Definition 14. Given two Λ-Henkin-structures $\mathfrak{M}$ and $\mathfrak{N}$, we write $\mathfrak{M} \equiv_\Lambda \mathfrak{N}$ and say that $\mathfrak{M}$ and $\mathfrak{N}$ are Λ-*equivalent* if they satisfy the same Λ-sentences. Also, for any natural number n, we write $\mathfrak{M} \equiv^n_\Lambda \mathfrak{N}$ and say that $\mathfrak{M}$ and $\mathfrak{N}$ are n-Λ-*equivalent* if $\mathfrak{M}$ and $\mathfrak{N}$ satisfy the same Λ-sentences of quantifier depth at most n. In particular, $\mathfrak{M} \equiv_\Lambda \mathfrak{N}$ holds iff, for all n, $\mathfrak{M} \equiv^n_\Lambda \mathfrak{N}$ holds.

Now we are ready to introduce our "Feferman-Vaught theorems". Comparable work had already been done by Makowski in [15] for extensions of FO, but a crucial difference is that he only considered standard structures, whereas we need to deal with Λ-Henkin-structures. Our proofs make use of Ehrenfeucht-Fraïssé games (defined in Appendix C: Definitions 24, 25, 26) for each of the logics Λ. The MSO game, that we show to be adequate, is rather straightforward and has already been used by other authors (see for instance [14]). The $\mathsf{FO}(\mathsf{LFP}^1)$ game is borrowed from Uwe Bosse [2]. It also applies to Henkin structures, as careful inspection of its adequacy proof shows. The $\mathsf{FO}(\mathsf{TC}^1)$ game has already been mentioned in passing by Grädel in [9] as an alternative to the game he used and we show that it is adequate. It looks also similar to a system of partial isomorphisms given in [3]. However it is important to note that this game is different from the $\mathsf{FO}(\mathsf{TC}^1)$ game which is actually used in [9]. The two games are equivalent when played on standard structures, but not when played on $\mathsf{FO}(\mathsf{TC}^1)$-Henkin structures. This is so because the game used in [3] relies on the alternative semantics for the TC operator given in Proposition 1, so that only finite sets of points can be chosen by players ; whereas the game we use involves choices of not necessarily finite admissible subsets. These are not equivalent approaches. Indeed, on $\mathsf{FO}(\mathsf{TC}^1)$-Henkin structures a simple compactness argument shows that the semantical clause of Proposition 1 (defined in terms of existence of a *finite* path) is not adequate.

Using these games we show that f-fusions of Λ-Henkin-structures preserve Λ-equivalence.

Theorem 2. *Let $\mathfrak{M}_i$ $\mathfrak{N}_i$ with $1 \leq i \leq k$ be Λ-Henkin-structures. For any f such as described in definition 13, whenever $\mathfrak{M}_i \equiv^n_\Lambda \mathfrak{N}_i$ for all $1 \leq i \leq k$, then also $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i \equiv^n_\Lambda \bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i$.*

Proof. The proofs are given in the second Appendix (Theorem 12 and Corollaries 8 and 12). □

As shown in Appendix D (Theorem 5 and Corollaries 9, 13) analogues of these theorems for disjoint union follow as well.

[2] It is simpler to give an example on standard structures, because then, we do not have to say anything about admissible sets.

Proposition 3. *For any* Λ*-Henkin-structures* $\mathfrak{M}_i$ *with* $1 \leq i \leq k$*,* $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$ *is also a Henkin structure.*

Proof. The proofs are given in Appendix D (Corollaries 6, 10 and 14). □

5 Completeness on Finite Trees

5.1 Forests and Operations on Forests

In Section 5.2, we will prove that no Λ-sentence can distinguish Λ-Henkin-models of $\vdash^{tree}_{\Lambda}$ from standard models of $\vdash^{tree}_{\Lambda}$. More precisely, we will show that for each n, any definably well-founded Λ-quasi-tree is n-Λ-equivalent to a finite tree. In order to give an inductive proof, it will be more convenient to consider a stronger version of this result concerning a class of finite and infinite Henkin structures that we call *quasi-forests*. In this section, we give the definition of quasi-forest and we show how they can be combined into bigger quasi-forests using the notion of fusion from Section 4. Whenever quasi forests are finite, we simply call them *finite forests*. As a simple example, consider a finite tree and remove the root node, then it is no longer a finite tree. Instead it is a finite sequence of trees, whose roots stand in a linear (sibling) order.[3] It does not have a unique root, but it does have a unique *left-most root*. For technical reasons it will be convenient in the definition of quasi forests to add an extra monadic predicate R labelling the roots.

Definition 15 (Λ-quasi-forest)**.** Let $T = (dom(T), <, \prec, P_1, \ldots, \ldots P_n, \mathbb{A}_T)$ be a Λ-quasi-tree. Given a node a in T, consider the Λ-substructure of T generated by the set $\{x \mid \exists z(a \preceq z \wedge z \leq x)\}$, which is the set formed by a together with all its siblings to the right and their descendants. The Λ-quasi-forest T_a is obtained by labeling each root in this substructure with R ($Rx \Leftrightarrow_{def} \neg\exists y\ y < x$). Whenever T is a tree, we simply call T_a a forest.

We will show in our main proof of completeness that for each n and for each node a in a Λ-Henkin definably well-founded quasi-tree, the Λ-quasi-forest T_a is n-Λ-equivalent to a finite forest. Our proof will use a notion of composition of Λ-quasi-forests which is a special case of fusion. Given a single node forest F_1 and two Λ-quasi-forests F_2 and F_3, we construct a new Λ-quasi-forest $\bigoplus^{COMP}(F_1, F_2, F_3)$ by letting the only element in F_1 be the left-most root, the roots of F_2 become the children of this node and the roots of F_3 become its siblings to the right. We then derive a corollary of Theorem 2 for compositions of Λ-quasi-forests and use it in Section 5.2.

Definition 16. Let $\sigma = \{<, \prec, R, P_1, \ldots, P_n\}$, be a relational vocabulary with only monadic predicates except $<$ and $\prec$. Given three additional monadic predicates Q_1, Q_2, Q_3, we define a mapping $COMP$ from σ to quantifier-free formulas over $\sigma \cup \{Q_1, Q_2, Q_3\}$ by letting

- $COMP(x < y) = x < y \vee (Q_1(x) \wedge Q_2(y))$
- $COMP(x \prec y) = x \prec y \vee (Q_1(x) \wedge Q_3(y) \wedge R(y))$
- $COMP(R(x)) = (Q_3(x) \wedge R(x)) \vee Q_1(x))$

Corollary 2. Let F_1 be a single node forest and F_2, F_3 Λ-quasi forests. If $F_2 \equiv^n_\Lambda F'_2$ and $F_3 \equiv^n_\Lambda F'_3$ then $\bigoplus^{COMP}(F_1, F_2, F_3) \equiv^n_\Lambda \bigoplus^{COMP}(F_1, F'_2, F'_3)$.

[3]Note that, as far as roots are concerned, two nodes can be siblings without sharing any parent. This would not happen in a quasi tree.

5.2 Main Proof of Completeness

Lemma 3. *For all $n \in \mathbb{N}$, every definably well-founded Λ-quasi-tree of finite signature is n-Λ-equivalent to a finite tree. In particular, a Λ-sentence is valid on definably well-founded Λ-quasi-trees iff it is valid on finite trees.*

Proof. Let T be a Λ-quasi-tree, w.l.o.g. assume that a monadic predicate R labels its root. During this proof, it will be convenient to work with Λ-quasi-forests. Note that finite Λ-quasi-forests are simply finite forests and finite Λ-quasi-trees are simply finite trees. Let X_n be the set of all nodes a of T for which it holds that T_a is n-Λ-equivalent to a finite forest. We first show that "belonging to X_n" is a property definable in T (Claim 1). Then, we use the induction scheme to show that every node of a definably well-founded Λ-quasi-tree (and in particular, the root) has this property (Claim 2).

Claim 1: X_n is invariant for $n+1$-Λ-equivalence (i.e., $(T,a) \equiv^{\Lambda}_{n+1} (T,b)$ implies that $a \in X_n$ iff $b \in X_n$), and hence is defined by a Λ-formula of quantifier depth $n+1$.

Proof of claim. Suppose that $(T,a) \equiv^{\Lambda}_{n+1} (T,b)$. We will show that $T_a \equiv^{\Lambda}_{n} T_b$, and hence, by the definition of X_n, $a \in X_n$ iff $b \in X_n$. By the definition of Λ-quasi-forests, $dom(T_a) = \{x \mid \exists z(a \preceq z \wedge z \leq x)\}$. Let ϕ be any Λ-sentence of quantifier depth n. We can assume w.l.o.g. that ϕ does not contain the variables z and x (otherwise we can rename in ϕ these two variables). By lemma 1, $(T,a) \models REL(\phi, \exists z(a \preceq z \wedge z \leq x), x)$ iff $T_a \models \phi$. Notice that $REL(\phi, \exists z(a \preceq z \wedge z \leq x), x)$ expresses precisely that ϕ holds in (T,a) within the subforest T_a. Moreover, the quantifier depth of $REL(\phi, \exists z(a \preceq z \wedge z \leq x)$ is at most $n+1$. It follows that $(T,a) \models REL(\phi, \exists z(a \preceq z \wedge z \leq x), x)$ iff $(T,b) \models REL(\phi, \exists z(b \preceq z \wedge z \leq x), x)$, and hence $T_a \models \phi$ iff $T_b \models \phi$.

For the second part of the claim, note that, up to logical equivalence, there are only finitely many Λ-formulas of any given quantifier depth, as the vocabulary is finite. ⊣

Claim 2: If all descendants and siblings to the right of a belong to X_n, then a itself belongs to X_n.

Proof of claim. Let us consider the case where a has both a descendant and a following sibling (all other cases are simpler). Then, by axioms T3, T5, T8, T9 and T10, a has a first child b, and an immediate next sibling c. Moreover, we know that both b and c are in X_n. In other words, T_b and T_c are n-Λ-equivalent to finite forests T'_b and T'_c. Now, we construct a finite Λ-quasi-forest T'_a by taking a $COMP$-fusion of T'_b, T'_c and of the Λ-substructure of T generated by $\{a\}$, which unique element becomes a common parent of all roots of T'_b and a left sibling of all roots of T'_c. So we get $T'_a = \bigoplus^{COMP}(T \upharpoonright \{a\}, T'_b, T'_c))$. It is not hard to see that T'_a is again a finite forest. Moreover, by the fusion lemma, $\bigoplus^{COMP}(T \upharpoonright \{a\}, T_b, T_c)) \equiv^{\Lambda}_{n} T'_a$. Now to show that $\bigoplus^{COMP}(T \upharpoonright \{a\}, T_b, T_c))$ is isomorphic to T_a (which entails $T_a \equiv^{\Lambda}_{n} T'_a$ i.e. T_a is n-Λ-equivalent to a finite forest), it is enough to show $\mathbb{A}_{T_a} = \mathbb{A}_{\bigoplus^{COMP}(T \upharpoonright \{a\}, T_b, T_c)}$. It holds that $\mathbb{A}_{\bigoplus^{COMP}(T \upharpoonright \{a\}, T_b, T_c)} \subseteq \mathbb{A}_{T_a}$ because we can define in T_a each such union of sets by means of a disjunction. Now to show $\mathbb{A}_{T_a} \subseteq \mathbb{A}_{\bigoplus^{COMP}(T \upharpoonright \{a\}, T_b, T_c)}$, take $A \in \mathbb{A}_{T_a}$, so $A = A_1 \cup A_2 \cup A_3$ with $A_1 \in \mathbb{A}_{T \upharpoonright \{a\}}$, $A_2 \in \mathbb{A}_{T_b}$, $A_3 \in \mathbb{A}_{T_c}$. The domain of each of these three structures is definable in T_a, let say ϕ_1 defines $dom(T \upharpoonright \{a\})$, ϕ_2 defines $dom(T_b)$ and ϕ_3 defines $dom(T_c)$. So each A_i component is definable in T_a (just take the conjunction $\phi_i(x) \wedge Ax$). But then A_i was already definable in $\bigoplus^{COMP}(T \upharpoonright \{a\}, T_b, T_c)$ (by construction of this structure). ⊣

It follows from these two claims, by the induction scheme for definable properties, that X_n contains all nodes of the Λ-quasi-tree, including the root, and hence T is n-Λ-equivalent to a finite tree. For the second statement of the lemma, it suffices to note that every Λ-sentence has a finite vocabulary and a finite quantifier depth. □

Theorem 3. *Let $\Lambda \in \{\mathsf{MSO}, \mathsf{FO(TC^1)}, \mathsf{FO(LFP^1)}\}$. The Λ-theory of finite trees is completely axiomatized by $\vdash_\Lambda^{tree}$.*

Proof. Theorem 3 follows directly from Lemma 3 and Corollary 1. □

5.3 The set of $\vdash_\Lambda^{tree}$ consequences *defines* the class of finite trees

Proposition 4 shows together with Theorem 3 that on standard structures, the set of $\vdash_\Lambda^{tree}$ consequences actually *defines* the class of finite trees. That is, $\vdash_\Lambda^{tree}$ has *no infinite standard model* at all.

Proposition 4. *Let $\Lambda \in \{\mathsf{FO(TC^1)}, \mathsf{FO(LFP^1)}, \mathsf{MSO}\}$. On standard structures, there is a Λ-formula which defines the class of finite trees.*

Sketch of the proof. It is enough to show it for $\Lambda = \mathsf{FO(TC^1)}$. It follows by Section 1.3 that it also holds for MSO and $\mathsf{FO(LFP^1)}$.

We merely give a sketch of the proof. For additional details we refer the reader to [13]. It can be shown that on standard structures, the finite conjunction of the axioms T1–T11 in Figure 5 "almost" defines the class of finite trees, i.e. any finite structure satisfying this conjunction is a finite tree. Now we will explain how to construct an other sentence, which together with this one, actually defines on arbitrary standard structures the class of finite trees. Let L be a shorthand for the formula labelling the leaves in the tree ($Lx \Leftrightarrow_{def} \neg\exists y x < y$) and R a shorthand for the formula labelling the root ($Rx \Leftrightarrow_{def} \neg\exists y y < x$). Consider the depth-first left-to-right ordering of nodes in a tree and the $\mathsf{FO(TC^1)}$ formula $\phi(x, y)$ saying "the node that comes after x in this ordering is y":

$$\phi(x,y) :\approx (\neg Lx \wedge x <_{imm} y \wedge \neg\exists z z \prec y) \vee (Lx \wedge x \prec_{imm} y) \vee (Lx \wedge \neg\exists z x \prec z \wedge \exists z(z < x \wedge z \prec_{imm} y \wedge \neg\exists w w < x \wedge z < w \wedge \exists u w \prec_{imm} u))$$

There is also a $\mathsf{FO(TC^1)}$ formula which says that "x is the very last node in this ordering". $\phi(x, y)$ can be combined with this formula into an $\mathsf{FO(TC^1)}$ formula χ expressing that the tree is finite by saying that (we rely here for the interpretation of χ on the alternative semantics for the TC operator given in Proposition 1) "there is a finite sequence of nodes $x_1 \ldots x_n$ such that x_1 is the root, x_{i+1} the node that comes after x_i in the above ordering, for all i, and x_n is the very last node of the tree in the above ordering".

$$\chi :\approx \exists u \exists z(Rz \wedge [TC_{xy}\phi](z, u) \wedge \neg\exists u'(u \neq u' \wedge [TC_{xy}\phi](u, u')))$$

□

Theorem 4. *The set of $\vdash_\Lambda^{tree}$ consequences* defines *the class of finite trees.*

Proof. By Proposition 4 we can express in Λ by means of some formula χ that a structure is a finite tree. So χ is necessarily a consequence of $\vdash_\Lambda^{tree}$ (as it is a Λ-formula valid on the class of finite trees). □

6 Conclusions

In this paper, taking inspiration from Kees Doets [4] we developed a uniform method for obtaining complete axiomatizations of fragments of MSO on finite trees. For that purpose, we had to adapt classical tools and notions from finite model theory to the specificities of Henkin semantics. The presence of admissible subsets called for some refinements in model theoretic constructions such as formation of substructure or disjoint union. Also, we noticed that not every Ehrenfeucht-Fraïssé game that has been used for $\mathsf{FO}(\mathsf{TC}^1)$ was suitable to use on Henkin-structures. We focused on a game which doesn't seem to have been used previously in the literature. We also elaborated analogues of the FO Feferman-Vaught theorem for MSO, $\mathsf{FO}(\mathsf{TC}^1)$ and $\mathsf{FO}(\mathsf{LFP}^1)$. We considered fusions of structures, a particular case of the Feferman-Vaught notion of generalized product and obtained results which might be interesting to generalize and use in other contexts.

We applied our method to MSO, $\mathsf{FO}(\mathsf{TC}^1)$ and $\mathsf{FO}(\mathsf{LFP}^1)$, but it would be worth also examining other fragments of MSO, such as monadic deterministic transitive closure logic ($\mathsf{FO}(\mathsf{DTC}^1)$) or monadic alternating transitive closure logic ($\mathsf{FO}(\mathsf{ATC}^1)$), see also [3].

Finally, an important feature of our main completeness argument is the way we used the inductive scheme of Figure 5. Hence, extending our approach to another class of finite structures would involve finding a comparable scheme. We also know that we should focus on a logic which is decidable on this class, as on finite structures recursive enumerability is equivalent to decidability. This suggests that other natural candidates would be fragments of MSO on classes of finite structures with bounded treewidth.

7 References

[1] Rolf Backofen, James Rogers, and Krishnamurti Vijay-Shankar. A first-order axiomatization of the theory of finite trees. *Journal of Logic, Language and Information*, 4(4):5–39, 1995.

[2] Uwe Bosse. An "Ehrenfeucht-Fraïssé game" for fixpoint logic and stratified fixpoint logic. In *CSL '92: Selected Papers from the Workshop on Computer Science Logic*, pages 100–114, London, UK, 1993. Springer-Verlag.

[3] A. Calo and J. A. Makowsky. The Ehrenfeucht-Fraïssé games for transitive closure. *Lecture Notes in Computer Science*, 620:57–68, 1992.

[4] K. Doets. *Completeness and Definability : Applications of the Ehrenfeucht Game in Second-Order and Intensional Logic*. PhD thesis, Universiteit van Amsterdam, 1987.

[5] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1995.

[6] H. Enderton. *A mathematical introduction to Logic*. Academic Press, New York, NY, USA, 1972.

[7] S. Feferman and R. Vaught. The first-order properties of algebraic systems. *Fundamenta Mathematicae*, 47:pp. 57–103, 1959.

[8] Georg Gottlob and Christoph Koch. Monadic datalog and the expressive power of languages for web information extraction. In *Proceedings of PODS 2002*, pages 17–28, 2002.

[9] Erich Grädel. On transitive closure logic. In *CSL '91: Proceedings of the 5th Workshop on Computer Science Logic*, pages 149–163, London, UK, 1992. Springer-Verlag.

[10] Leon Henkin. Completeness in the theory of types. *The Journal of Symbolic Logic*, 15(2):81–91, 1950.

[11] Wilfrid Hodges. *A shorter model theory.* Cambridge University Press, New York, NY, USA, 1997.

[12] Stephan Kepser. Querying linguistic treebanks with monadic second-order logic in linear time. *J. of Logic, Lang. and Inf.*, 13(4):457–470, 2004.

[13] Stephan Kepser. Properties of binary transitive closure logic over trees. In Giorgio Satta Paola Monachesi, Gerald Penn and Shuly Wintner, editors, *Formal Grammar 2006*, pages 77–89, 2006.

[14] Grégory Lafitte and Jacques Mazoyer. Théorie des modèles et complexité. Technical report, Ecole Normale Suprieure de Lyon, 1998.

[15] J. A. Makowsky. Algorithmic uses of the Feferman-Vaught theorem. *Annals of Pure and Applied Logic*, 126(1–3):159–213, 2004.

[16] María Manzano. *Extensions of First-Order logic.* Cambridge University Press, New York, NY, USA, 1996.

[17] Oliver Matz and Nicole Schweikardt. Expressive power of monadic logics on words, trees, pictures, and graphs. In E. Grädel J. Flum and T. Wilke, editors, *Logic and Automata: History and Perspectives*, pages 531–552. Texts in Logic and Games, Amsterdam University Press, 2007.

[18] James Rogers. *Descriptive Approach to Language - Theoretic Complexity.* CSLI Publications, Stanford, CA, USA, 1998.

[19] Balder ten Cate and Maarten Marx. Axiomatizing the logical core of XPath 2.0. In *ICDT*, pages 134–148, 2007.

[20] Balder ten Cate and Luc Segoufin. XPath, transitive closure logic, and nested tree walking automata. In *PODS '08: Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 251–260, New York, NY, USA, 2008. ACM.

[21] Hans-Jörg Tiede and Stephan Kepser. Monadic second-order logic and transitive closure logics over trees. *Electron. Notes Theor. Comput. Sci.*, 165:189–199, 2006.

A Henkin Completeness Proofs

Let $\Lambda \in \{\mathsf{MSO}, \mathsf{FO}(\mathsf{TC}^1), \mathsf{FO}(\mathsf{LFP}^1)\}$. In this appendix we show that $\vdash_\Lambda$ is complete on the class of Λ-Henkin-structures. We are not yet concerned with $\vdash_\Lambda^{tree}$ and we do not consider the specific axioms on trees listed in Figure 5.

Up to now we have been working with purely relational vocabularies. Here we will be using individual constants in the standard way, but only for the sake of readability (we could dispense with them and use FO variables instead). Also, whenever this is clear from the context, we will use $\vdash$ as shorthand for $\vdash_\Lambda$.

A.1 The MSO-Henkin Completeness Proof

This proof is an adaptation to the case of MSO of the proof of completeness for FO given in [6] and of the proof of completeness for the theory of types given in [16].

Lemma 4 (FO generalization lemma)**.** *If $\Gamma \vdash \phi$ and x does not occur free in Γ, then $\Gamma \vdash \forall x\phi$.*

Proof. (by Enderton) Consider a fixed set Γ and a variable x not free in Γ. We show by induction that for any theorem ϕ of Γ, we have $\Gamma \vdash \forall x\phi$. For this it suffices (by the induction principle) to show that the set

$$\{\phi : \Gamma \vdash \forall x\phi\}$$

includes $\Gamma \cup Ax^{MSO}$ (where Ax^{MSO} is the set of logical axioms given in Figures 1 and 2) and is closed under modus ponens. Notice that x can occur free in ϕ.

Case 1. *ϕ is a logical axiom. Then $\forall x\phi$ is also a logical axiom. And so $\Gamma \vdash \forall x\phi$.*

Case 2. *$\phi \in \Gamma$. Then x does not occur free in ϕ. Hence*

$$\phi \rightarrow \forall x\phi$$

is an instance of $FO4$. Consequently, $\Gamma \vdash \forall x\phi$, as from ϕ (which is in Γ) and $\phi \rightarrow \forall x\phi$ (which is an instance of $FO4$) we can infer by modus ponens that $\forall x\phi$.

Case 3. *ϕ is obtained by modus ponens from ψ and $\psi \rightarrow \phi$. Then by inductive hypothesis we have $\Gamma \vdash \forall x\psi$ and $\Gamma \vdash \forall x(\psi \rightarrow \phi)$. This is just the situation in which axiom group $FO3$ is useful. We have $\Gamma \vdash \forall x\phi$. The proof goes as follows. From $\psi \rightarrow \phi$ we obtain by generalization $\forall x\psi \rightarrow \phi$, which together with $\forall x(\psi \rightarrow \phi) \rightarrow (\forall x\psi \rightarrow \forall x\phi)$ (which is an instance of $FO3$) gives by modus ponens $\forall x\psi \rightarrow \forall x\phi$. Now by generalization from ψ we obtain $\forall x\psi$ and by modus ponens, $\forall x\phi$.*

So by induction $\Gamma \vdash \forall x\phi$ for every theorem ϕ of Γ.

□

Lemma 5 (MSO generalization theorem)**.** *If $\Gamma \vdash \phi$ and X does not occur free in Γ, then $\Gamma \vdash \forall X\phi$.*

Proof. The proof is similar as in the FO case, except that MSO generalization is used instead of FO generalization and $MSO2$ and $MSO3$ are used instead of, respectively, $FO3$ and $FO4$. □

Definition 17. We say that a set of MSO formulas Δ contains MSO-Henkin witnesses if and only if for every formula ϕ, if $\neg\forall x\phi \in \Delta$ (respectively $\neg\forall X\phi \in \Delta$), then $\neg\phi[x/t] \in \Delta$ for some term t (respectively $\neg\phi[X/T] \in \Delta$ with T either a monadic predicate or a set variable).

Lemma 6. *(MSO Lindenbaum lemma) Let $\sigma^* = \sigma \cup \{c_n \mid n \in \mathbb{N}\} \cup \{P_n \mid n \in \mathbb{N}\}$, with $c_i \notin \sigma$ and $P_i \notin \sigma$. If $\Gamma \subseteq FORM(\sigma)$ is consistent, then there exists a maximally consistent set Γ^* such that $\Gamma \subseteq \Gamma^*$ and Γ^* contains MSO-Henkin witnesses.*

Proof. Let Γ be a $\vdash_{MSO}$ consistent set of well formed formulas in a countable vocabulary. We expand the language by adding countably many new constants and countably many new monadic predicates. Then Γ remains consistent as a set of well formed formulas in the new language. For all the sets constituted of one formula in the new language, one FO variable and one MSO variable, we adopt the following fixed exhaustive enumeration:

$$< \phi_1, x_1, X_1 >, < \phi_2, x_2, X_2 >, < \phi_3, x_3, X_3 >, < \phi_4, x_4, X_4 >, \dots$$

(possible since the language is countable), where the ϕ_i are formulas, the x_i are FO variables and the X_i, MSO variables.

- Let θ_{2n-1} be $\neg\forall x_n \phi_n \rightarrow \neg\phi_n[x_n/c_l]$, where c_l is the first of the new constants neither occurring in ϕ_n nor in θ_k with $k < 2n-1$
- Let θ_{2n} be $\neg\forall X_n \phi_n \rightarrow \neg\phi_n[X_n/P_l]$, where P_l is the first of the new monadic predicates neither occurring in ϕ_n nor in θ_k with $k < 2n$

Call Θ the set of all the θ_i.

Claim 1. $\Gamma \cup \Theta$ *is consistent*

If not, then because deductions are finite, for some $m \geq 0$, $\Gamma \cup \{\theta_1, \dots, \theta_m, \theta_{m+1}\}$ is inconsistent. Take the least such m, then by the (derivable) rule of *reductio ad absurdum*, $\Gamma \cup \{\theta_1, \dots, \theta_m\} \vdash \neg\theta_{m+1}$. Now there are two cases:

(1) θ_{m+1} is of the form $\neg\forall x\phi \rightarrow \neg\phi[x/c]$ i.e. either $\Gamma \cup \{\theta_1, \dots, \theta_m\} \vdash \neg\forall x\phi$ and $\Gamma \cup \{\theta_1, \dots, \theta_m\} \vdash \phi[x/c]$. Since c does not appear in any formula on the left, by the FO generalization theorem, $\Gamma \cup \{\theta_1, \dots, \theta_m\} \vdash \forall x\phi$, which contradicts the minimality of m (or the consistency of Γ if $m = 0$)

(2) θ_{m+1} is of the form $\neg\forall X_{2n}\phi_{2n} \rightarrow \neg\phi[X/P_{2n}]$

The reasoning is similar (we use the MSO generalization theorem instead of the FO one).

We now extend in the standard way the consistent set $\Gamma \cup \Theta$ to a maximal consistent set Γ^* which is maximal in the sense that for any well formed formula ϕ either $\phi \in \Gamma^*$ or $\phi \notin \Gamma^*$. □

Definition 18. Let $\Gamma^* \subseteq FORM(\sigma)$ be maximally consistent and contain Henkin witnesses. We define an equivalence relation on the set of FO terms, by letting $t_1 \equiv_{\Gamma^*} t_2$ iff $t_1 = t_2 \in \Gamma^*$. We denote the equivalence class of a term t by $|t|$.

Proposition 5. $\equiv_{\Gamma^*}$ *is an equivalence relation.*

Proof. By $FO5$ and $FO6$. □

We will now show that if Γ^* is maximally consistent and contains Henkin witnesses, then Γ^* has a MSO-Henkin model $\mathfrak{M}_{\Gamma^*}$.

Definition 19. We define $\mathfrak{M}_{\Gamma^*}$ (together with a valuation g_{Γ^*}) out of Γ^*.

- $M = \{|t| : t \text{ is a FO term }\}$
- $\mathbb{A}_{\mathfrak{M}_{\Gamma^*}} = \{A_T : T \text{ is a set variable or a monadic predicate}\}$ where $A_T = \{|t| : Tt \in \Gamma^*\}$
- $(|t_1|, \dots, |t_n|) \in P^{\mathfrak{M}}_{\Gamma^*}$ iff $Pt_1 \dots t_n \in \Gamma^*$
- $c^{\mathfrak{M}_{\Gamma^*}} = |c|$
- $g_{\Gamma^*}(x) = |x|$
- $g_{\Gamma^*}(X) = A_X$

We still need to show that $\mathbb{A}_{\mathfrak{M}_{\Gamma^*}}$ is closed under MSO definability. We will be able to do that after having shown the following truth lemma.

Lemma 7. *(Truth lemma) For any MSO formula ϕ, $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \phi$ iff $\phi \in \Gamma^*$.*

Proof. By induction on ϕ. The base case (for atomic formulas) follows from the definition of $\mathfrak{M}_{\Gamma^*}$ together with the maximality of Γ^*. Now consider the inductive step:

- Boolean connectives: standard (no difference with usual FO Henkin completeness proofs).
- FO quantifier: we want to show that

$$\mathcal{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall x\phi \text{ iff } \forall x\phi \in \Gamma^*$$

  We first show $\mathcal{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall x\phi$ entails $\forall x\phi \in \Gamma^*$.

  $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall x\phi$ entails that for all FO term t, $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[x/|t|] \models \phi$, which entails $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \phi[x/t]$. By induction hypothesis, for all t, $\phi[x/t] \in \Gamma^*$. Now suppose $\neg\forall x\phi \in \Gamma^*$, then by construction of Γ^* there exists a variable x_m such that $\neg\phi[x/x_m] \in \Gamma^*$, but this contradicts what we get by induction hypothesis, so $\neg\forall x\phi \notin \Gamma^*$ and by maximal consistency of Γ^*, $\forall x\phi \in \Gamma^*$.

  Now we show $\forall x\phi \in \Gamma^*$ entails $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall x\phi$. We take the contraposition: $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \not\models \forall x\phi$ entails $\forall x\phi \notin \Gamma^*$. Suppose $\mathcal{M}_{\Gamma^*}, g_{\Gamma^*} \not\models \forall x\phi$, so $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \neg\forall x\phi$ i.e. $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \exists x\neg\phi$. So $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[x/|t|] \models \neg\phi$ for some FO term t, which entails $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \neg\phi[x/t]$. By induction hypothesis $\neg\phi[x/t] \in \Gamma^*$, by $FO2$, $\exists x\neg\phi \in \Gamma^*$ by $FO2$, by maximal consistency of Γ^*, $\neg\forall x\phi \in \Gamma^*$.

- Set quantifier: we want to show that

$$\mathcal{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall X\phi \text{ iff } \forall X\phi \in \Gamma^*$$

  We first show $\mathcal{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall X\phi$ entails $\forall X\phi \in \Gamma^*$.

  $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall X\phi$ entails that for all MSO term T, $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[X/A_T] \models \phi$ and so $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \phi[X/T]$ (because for any set variable X, $g_{\Gamma^*}(X) = A_X$ and for any monadic predicate P, $P^{\mathfrak{M}_{\Gamma^*}} = A_P$.) By induction hypothesis, for all T, $\phi[X/T] \in \Gamma^*$. Now suppose $\neg\forall X\phi \in \Gamma^*$, then by construction of Γ^* there exists a variable X_m such that $\neg\phi[X/X_m] \in \Gamma^*$, but this contradicts what we get by induction hypothesis, so $\neg\forall X\phi \notin \Gamma^*$ and by maximal consistency of Γ^*, $\forall X\phi \in \Gamma^*$.

  Now we show $\forall X\phi \in \Gamma^*$ entails $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \forall X\phi$. We take the contraposition $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \not\models \forall X\phi$ entails $\forall X\phi \notin \Gamma^*$. Suppose $\mathcal{M}_{\Gamma^*}, g_{\Gamma^*} \not\models \forall X\phi$, so $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \neg\forall X\phi$ i.e. $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \exists X\neg\phi$. So $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[X/A_T] \models \neg\phi$ for some MSO term T, which entails $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \neg\phi[X/T]$. By induction hypothesis $\neg\phi[X/T] \in \Gamma^*$, by $MSO1$, $\exists X\neg\phi \in \Gamma^*$, by maximal consistency of Γ^*, $\neg\forall X\phi \notin \Gamma^*$ i.e. $\forall X\phi \notin \Gamma^*$.

□

Proposition 6. *$\mathfrak{M}_{\Gamma^*}$ is a MSO-Henkin structure.*

Proof. Essentially here we will use the fact that $\mathfrak{M}_{\Gamma^*}$ is a model of all the $COMP$ instances. We want to see that all sets which are parametrically definable using our MSO-language are in the set of admissible subsets of $\mathfrak{M}_{\Gamma^*}$. Let ϕ be a MSO formula, x a variable and $x_1, \ldots, x_n, X_{n+1}, \ldots, X_m$ the sequence, ordered by occurrence, of all the free variables of ϕ, apart from x.

Take any set variable X not free in ϕ. By hypothesis, $\mathcal{M}$ is a model of the sentence

$$\forall x_1 \ldots \forall x_n \forall X_{n+1} \ldots \forall X_m [\exists X \forall x (Xx \leftrightarrow \phi)]$$

Therefore for all objects $a_1, \ldots, a_n$ and admissible subsets $A_{n+1}, \ldots, A_m$

$$(\mathcal{M}, a_1, \ldots, a_n, A_{n+1}, \ldots, A_m) \text{ is a model of } \exists X \forall x (Xx \leftrightarrow \phi)$$

So there is an $A \in \mathbb{A}_{\mathcal{M}}$ such that for all $a \in A$

$$a \in A \text{ iff } \mathcal{M}, g[x/a, x_1/a_1, \ldots, x_n/a_n, X_{n+1}/A_{n+1}, \ldots, X_m/A_m] \models \phi$$

This $A \in \mathbb{A}_{\mathcal{M}}$ is precisely the relation parametrically defined by the formula ϕ and the variables mentioned above.[4] □

Theorem 5. *Every $\vdash_{MSO}$ consistent set Γ of MSO sentences is satisfiable in a MSO-Henkin structure.*

Proof. First turn Γ into a maximal consistent set Γ^* in a possibly richer language σ^* with Henkin witnesses. Then build a structure $\mathfrak{M}_{\Gamma^*}$ out of this Γ^*. Then the structure $\mathfrak{M}_{\Gamma^*}$ satisfies Γ^* and hence also the (subset) Γ. □

A.2 The $\mathsf{FO}(\mathsf{TC}^1)$-Henkin Completeness Proof

The following proof is a variation of the proofs in [6] and [16]. The originality of the $\mathsf{FO}(\mathsf{TC}^1)$ case essentially lies in the notion of $\mathsf{FO}(\mathsf{TC}^1)$-Henkin witness of Definition 20. In order to use this notion in the proof of Lemma 9, we also need the following lemma:

Lemma 8. *Let Γ be a consistent set of $\mathsf{FO}(\mathsf{TC}^1)$ formulas and θ a $\mathsf{FO}(\mathsf{TC}^1)$ formula of the form $\forall x(\phi \leftrightarrow Px)$ with P a fresh monadic predicate (i.e. not appearing in Γ). Then $\Gamma \cup \{\theta\}$ is also consistent.*

Proof. Suppose $\Gamma \cup \{\forall x(\phi \leftrightarrow Px)\}$ is inconsistent, so there is some proof of $\perp$ from formulas in $\Gamma \cup \{\forall x(\phi \leftrightarrow Px)\}$. We first rename all bound variables in the proof with variables which had no occurrence in the proof or in $\forall x(\phi \leftrightarrow Px)$ (this is possible since proofs are finite objects and we have a countable stock of variables). Also, whenever in the proof the $\mathsf{FO}(\mathsf{TC}^1)$ generalization rule is applied on some unary predicate P, we make sure that this P is different from the unary predicate that we want to substitute by ϕ and which does not appear in the proof; this is always possible because we have a countable set of unary predicates. Now, we replace in the proof all occurrences of Px by ϕ (as we renamed bound variables, there is no accidental binding of variables by wrong quantifiers). Then, every occurrence of $\forall x(\phi \leftrightarrow Px)$ in the proof becomes an occurrence of $\forall x(\phi \leftrightarrow \phi)$ i.e. we have obtained a proof of $\perp$ from $\Gamma \cup \{\forall x(\phi \leftrightarrow \phi)\}$ i.e. from Γ ($\forall x(\phi \leftrightarrow \phi)$ is an axiom, as it can be obtained by FO generalization from a tautology of sentential calculus). It entails that Γ is already inconsistent, which contradicts the consistency of Γ. Now it remains to show that the replacement procedure of all occurrences of Px by ϕ, is correct, that is, we still have a proof of $\perp$ after it. Every time the replacement occurs in an axiom (or its generalization, which is still an axiom as we defined it), then the result

[4] It follows that without the $COMP$ axiom, we get an axiomatization of MSO on arbitrary frames.

is still an instance of the given axiom schema (even for $\mathsf{FO(TC^1)}$ generalizations, because we took care that P is never used in the proof for a $\mathsf{FO(TC^1)}$ generalization). Also, as replacement is applied uniformly in the proof, every application of modus ponens stays correct: consider $\psi \to \xi$ and ψ. Obviously the result ψ^* of the substitution will allow to derive the result ξ^* of the substitution from $\psi^* \to \xi^*$ and ψ^*. Also $\bot^*$ is simply $\bot$, so the procedure gives us a proof of $\bot$. □

Definition 20. We say that a set of $\mathsf{FO(TC^1)}$ formulas Δ contains $\mathsf{FO(TC^1)}$-Henkin witnesses if and only if the two following conditions hold. First, for every formula ϕ, if $\neg\forall x\phi \in \Delta$, then $\neg\phi[x/t] \in \Delta$ for some term t and if $\neg[TC_{xy}\phi](u,v) \in \Delta$, then $Pu \wedge \forall x\forall y((Px \wedge \phi(x,y)) \to Py) \wedge \neg Pv \in \Delta$ for some monadic predicate P. Second, if $\phi \in \Delta$ and x is a free variable of ϕ, then $\forall x(Px \leftrightarrow \phi(x)) \in \Delta$ for some monadic predicates P.

Lemma 9. *($\mathsf{FO(TC^1)}$ Lindenbaum lemma) Let $\sigma^* = \sigma \cup \{c_n \mid n \in \mathbb{N}$ with c_n a new individual constant, $\} \cup \{P_n \mid n \in \mathbb{N}$ with P_n a new monadic predicate$\}$. If $\Gamma \subseteq FORM(\sigma)$ is consistent, then there exists a maximally consistent set Γ^* of σ^* formulas such that $\Gamma \subseteq \Gamma^*$ and Γ^* contains $\mathsf{FO(TC^1)}$-Henkin witnesses.*

Proof. Let Γ be a $\vdash_{\mathsf{FO(TC^1)}}$ consistent set of well formed formulas in a countable vocabulary σ. We expand the language into σ^* by adding countably many new constants and countably many new monadic predicates. Then Γ remains $\vdash_{\mathsf{FO(TC^1)}}$ consistent as a set of well formed formulas in the new language. For all the pairs constituted by one formula and one variable of σ^* and all the pairs constituted by one formula and two terms of σ^*, we adopt the following fixed exhaustive enumeration:

$$< \phi_1, x_1 >, < \phi_2, u_2, v_2 >, < \phi_3, x_3 >, < \phi_4, u_4, v_4 >, \dots$$

(possible since the language is countable), where the ϕ_i are formulas, the x_i are variables and the u_i, v_i are terms.

- Let θ_{3n-2} be $\neg\forall x_{2n-1}\phi_{2n-1} \to \neg\phi_{2n-1}[x_{2n-1}/c_l]$, where c_l is the first of the new constants neither occurring in ϕ_{2n-1} nor in θ_k with $k < 3n-2$
- Let θ_{3n-1} be $\forall x_{2n-1}(\phi_{2n-1} \leftrightarrow P_l x_{2n-1})$, where P_l is the first of the new monadic predicates neither occurring in ϕ_{2n-1} nor in θ_k with $k \leq 3n-1$.
- Let θ_{3n} be $\neg[TC_{xy}\phi_{2n}](u_{2n}, v_{2n}) \to (P_l u_{2n} \wedge \forall x\forall y((P_l x \wedge \phi_{2n}(x,y)) \to P_l y) \wedge \neg P_l v_{2n})$, where P_l is the first of the new monadic predicates neither occurring in ϕ_{2n} nor in θ_k with $k \leq 3n$

Call Θ the set of all the θ_i.

Claim 2. $\Gamma \cup \Theta$ *is consistent*

If not, then because deductions are finite, for some $m \geq 0$, $\Gamma \cup \{\theta_1, \dots, \theta_m, \theta_{m+1}\}$ is inconsistent. Take the least such m, then by the *reductio ad absurdum* rule, $\Gamma \cup \{\theta_1, \dots, \theta_m\} \vdash \neg\theta_{m+1}$. Now there are three cases:

(1) θ_{m+1} is of the form $\neg\forall x\phi \to \neg\phi[x/c]$

(see the MSO case for how to handle this case)

(2) θ_{m+1} is of the form $\neg[TC_{xy}\phi](u,v) \to ((Pu \wedge \forall x\forall y((Px \wedge \phi(x,y)) \to Pv) \wedge \neg Pv)$.

In such a case both $\Gamma \cup \{\theta_1 \dots \theta_m\} \vdash \neg[TC_{xy}\phi](u,v)$ and $\Gamma \cup \{\theta_1 \dots \theta_m\} \vdash (Pu \wedge \forall x\forall y((Px \wedge \phi(x,y)) \to Pv) \to Pv$ hold. Since P does not appear in any formula on the left, by $\mathsf{FO(TC^1)}$ generalization, $\Gamma \cup \{\theta_1 \dots \theta_m\} \vdash [TCx_{xy}\phi](u,v)$, which contradicts the minimality of m (or the consistency of Γ if $m = 0$).

(3) θ_{m+1} is of the form $\forall x(\phi \leftrightarrow Px)$

(by Lemma 8, this is not possible)

We then turn $\Gamma \cup \Theta$ into a maximal consistent set Γ^* in the standard way. $\square$

We now define $\mathfrak{M}_{\Gamma^*}$ and g_{Γ^*} as we did for MSO.

Lemma 10. *(Truth lemma) For any FO(TC^1) formula ϕ, $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \phi$ iff $\phi \in \Gamma^*$.*

Proof. By induction on ϕ.

The base case follows from the definition of $\mathfrak{M}_{\Gamma^*}$ together with the maximality of Γ^*. Now consider the inductive step:

- Boolean connectives and FO quantifier: as in MSO
- TC operator: we want to show that

$$\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [TC_{xy}\phi(x,y)](u,v) \text{ iff } [TC_{xy}\phi(x,y)](u,v) \in \Gamma$$

  - We first show that $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [TC_{xy}\phi(x,y)](u,v)$ implies $[TC_{xy}\phi(x,y)](u,v) \in \Gamma^*$. So suppose $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [TC_{xy}\phi(x,y)](u,v)$ i.e. for all monadic predicates $P_i \in \sigma^*$, if $g_{\Gamma^*}(u) \in A_{P_i}$ and for all $|t_k|, |t_l| \in M$, $|t_k| \in A_{P_i}$ and $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[x/|t_k|, x/|t_l|] \models \phi$ implies $|t_l| \in A_{P_i}$, then $g_{\Gamma^*}(v) \in A_{P_i}$ i.e. $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models P_i u \wedge (((P_i t_k \wedge \phi(t_k,t_l)) \rightarrow P_i t_l) \rightarrow P_i v)$ for all P_i, t_k, t_l and by induction hypothesis $P_i u \wedge (((P_i t_k \wedge \phi(t_k,t_l)) \rightarrow P_i t_l) \rightarrow P_i v) \in \Gamma^*$. And so by the same argument as the one used in the FO quantifier step of the present induction, $P_i u \wedge \forall x \forall y(((P_i x \wedge \phi(x,y)) \rightarrow P_i y) \rightarrow P_i v) \in \Gamma^*$. Now suppose $[TC_{xy}\phi(x,y)](u,v) \notin \Gamma^*$ i.e. $\neg[TC_{xy}\phi(x,y)](u,v) \in \Gamma^*$. Then as Γ^* contains Henkin witnesses, there is a predicate P_m such that $P_m u \wedge \forall x \forall y((P_m x \wedge \phi(x,y)) \rightarrow P_m y) \wedge \neg P_m v \in \Gamma^*$. But that contradicts the maximal consistency of Γ^*. Then $\neg[TC_{xy}\phi(x,y)](u,v) \notin \Gamma^*$ and by maximal consistency of Γ^*, $[TC_{xy}\phi(x,y)](u,v) \in \Gamma^*$.
  - We now show that $[TC_{xy}\phi(x,y)](u,v) \in \Gamma^*$ implies $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [TC_{xy}\phi(x,y)](u,v)$. We consider the contraposition $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \not\models [TC_{xy}\phi(x,y)](u,v)$ implies $[TC_{xy}\phi(x,y)](u,v) \notin \Gamma^*$. So suppose $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \not\models [TC_{xy}\phi(x,y)](u,v)$ i.e. $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \neg[TC_{xy}\phi(x,y)](u,v)$ i.e. there exists $A_{P_i} \in \mathbb{A}_{\mathfrak{M}_{\Gamma^*}}$ such that, $g(u) \in A_{P_i}$ and for all $|t_k|, |t_l| \in M$, $|t_k| \in A_{P_i}$ and $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[x/|t_k|, x/|t_l|] \models \phi$ implies $|t_l| \in A_{P_i}$ and $\neg g_{\Gamma^*}(v) \in A_{P_i}$ and by induction hypothesis $P_i u \wedge (((P_i t_k \wedge \phi(t_k,t_l)) \rightarrow P_i t_l) \wedge \neg P_i v \in \Gamma^*$. And so by the same argument as the one used in the FO quantifier step of the present induction, $P_i u \wedge \forall x \forall y(((P_i x \wedge \phi(x,y)) \rightarrow P_i y) \wedge \neg P_i v \in \Gamma^*$. Now suppose $[TC_{xy}\phi(x,y)](u,v) \in \Gamma^*$. Then by the TC axiom, for every monadic predicate P_m, $(P_m u \wedge \forall x \forall y((P_m x \wedge \phi(x,y)) \rightarrow P_m y)) \rightarrow P_m v \in \Gamma^*$. But that contradicts the maximal consistency of Γ^*. Then $[TC_{xy}\phi(x,y)](u,v) \notin \Gamma^*$ and by maximal consistency of Γ^*, $\neg[TC_{xy}\phi(x,y)](u,v) \in \Gamma^*$.

$\square$

Proposition 7. *$\mathfrak{M}_{\Gamma^*}$ is a FO(TC^1)-Henkin structure.*

Proof. By construction of Γ^* this is immediate (we introduced a monadic predicate for each parametrically definable subset). $\square$

Theorem 6. *Every consistent set Γ of FO(TC^1) formulas is satisfiable in a FO(TC^1)-Henkin structure.*

Proof. First turn Γ into a σ^* maximal consistent set Γ^* with $\mathsf{FO(TC^1)}$-Henkin witnesses in a possibly richer signature σ^* (with extra individual constants and monadic predicates). Then build a σ^* structure $\mathfrak{M}_{\Gamma^*}$ out of this Γ^*. Then the structure $\mathfrak{M}_{\Gamma^*}$ satisfies Γ^* and hence also the (subset) Γ. $\square$

A.3 The FO(LFP^1)-Henkin Completeness Proof

This proof parallels the $\mathsf{FO(TC^1)}$ one. It is a similar variation of the proofs in [6] and [16] and the notion of $\mathsf{FO(LFP^1)}$-Henkin witness in Definition 21 parallels the notion of Henkin witness in Definition 20.

Definition 21. We say that a set of $\mathsf{FO(LFP^1)}$ formulas Δ contains $\mathsf{FO(LFP^1)}$ Henkin witnesses if and only if the two following conditions hold. First, for every formula ϕ, if $\neg\forall x\phi \in \Delta$, then $\neg\phi[x/t] \in \Delta$ for some term t and if $\neg[LFP_{xX}\phi]y \in \Delta$, then $\neg Py \wedge \neg\exists x(\neg Px \wedge \phi(P,x)) \in \Delta$ for some new monadic predicate P. Second, if $\phi \in \Delta$ and x is a free variable of ϕ, then $\forall x(Px \leftrightarrow \phi(x)) \in \Delta$ for some monadic predicates P.

Lemma 11. *(FO(LFP^1) Lindenbaum lemma) Let $\sigma^* = \sigma \cup \{c_n \mid \ \in \mathbb{N}\} \cup \{P_n \mid n \in \mathbb{N}\}$ with $c_i, P_i \notin \sigma$. If $\Gamma \subseteq FORM(\sigma)$ is consistent, then there exists a maximally consistent set Γ^* of σ^* formulas such that $\Gamma \subseteq \Gamma^*$ and Γ^* contains FO(LFP^1)-Henkin witnesses.*

Proof. Let Γ be a consistent set of well formed $\mathsf{FO(LFP^1)}$ formulas in a countable vocabulary. We expand the language by adding countably many new constants and countably many new monadic predicates. Then Γ remains consistent as a set of well formed formulas in the new language. For every pair constituted by one formula and one FO variable of σ^*, we adopt the following fix exhaustive enumeration:

$$< \phi_1, x_1 >, < \phi_2, x_2 >, < \phi_3, x_3 >, < \phi_4, x_4 >, \dots$$

(possible since the language is countable), where the ϕ_i are formulas and the x_i are FO variables.

- Let θ_{3n-2} be $\neg\forall x_n\phi_n \rightarrow \neg\phi[x_n/c_l]$, where c_l is the first of the new constants neither occurring in ϕ_n nor in θ_k with $k < 3n-2$.
- Let θ_{3n-1} be $\neg[LFP_{xX}\phi_n]x_n \rightarrow (\neg P_l x_n \wedge \neg\exists x(\neg P_l x \wedge \phi(P_l, x)))$, where P_l is the first of the new monadic predicates neither occurring in ϕ_n nor in θ_k with $k < 3n-1$.
- Let θ_{3n} be $\forall x_n(\phi_n \leftrightarrow P_l x_n)$, where P_l is the first of the new monadic predicates neither occurring in ϕ_n nor in θ_k with $k < 3n$.

Call Θ the set of all the θ_i.

Claim 3. $\Gamma \cup \Theta$ *is consistent*

If not, then because deductions are finite, for some $m \geq 0$, $\Gamma \cup \{\theta_1, \dots, \theta_m, \theta_{m+1}\}$ is inconsistent. Take the least such m, then by the *reductio ad absurdum* rule, $\Gamma \cup \{\theta_1, \dots, \theta_m\} \vdash \neg\theta_{m+1}$. Now there are three cases:

(1) θ_{m+1} is of the form $\neg\forall x\phi \rightarrow \phi[x/c]$

(see the MSO case for how to handle this case)

(2) θ_{m+1} is of the form $\neg[LFP_{xX}\phi]y \rightarrow (\neg Py \wedge \neg\exists x(\neg Px \wedge \phi(P,x)))$.

In such a case both $\Gamma \cup \{\theta_1 \ldots \theta_m\} \vdash \neg[LFP_{xX}\phi]y$ and $\Gamma \cup \{\theta_1 \ldots \theta_m\} \vdash \neg Py \wedge \neg\exists x(\neg Px \wedge \phi(P,x))$ hold. Since P does not appear in any formula on the left, by FO(LFP^1) generalization, $\Gamma \cup \{\theta_1 \ldots \theta_m\} \vdash [LFP_{xX}\phi]y$, which contradicts the leastness of m (or the consistency of Γ if $m = 0$)

(3) θ_{m+1} is of the form $\forall x(\phi \leftrightarrow Px)$

(see the FO(TC^1) case for how to handle this case, just consider the FO(LFP^1) generalization rule instead of the FO(TC^1) one in Lemma 8)

We then turn $\Gamma \cup \Theta$ into a maximal consistent set Γ^* in the standard way. □

We now define $\mathfrak{M}_{\Gamma^*}$ and g_{Γ^*} as we did for MSO.

Lemma 12. *(Truth lemma) For any FO(LFP^1) formula ϕ, $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \phi$ iff $\phi \in \Gamma^*$.*

Proof. By induction on ϕ.

The base case follows from the definition of $\mathfrak{M}_{\Gamma^*}$ together with the maximality of Γ^*. Now consider the inductive step:

- Boolean connectives and FO quantifier: as in MSO
- LFP operator: we want to show that

$$\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [LFP_{xX}\phi]y \text{ iff } [LFP_{xX}\phi]y \in \Gamma$$

  - We first show that

$$\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [LFP_{xX}\phi]y \text{ implies } [LFP_{xX}\phi]y \in \Gamma^*.$$

    So suppose $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [LFP_{xX}\phi]y$ i.e. for all monadic predicates $P_i \in \sigma^*$, if $g_{\Gamma^*}(y) \notin A_{P_i}$ then there exists $|t_k| \in M$, such that $|t_k| \notin A_{P_i}$ and $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[x/|t_k|, X/A_{P_i}] \models \phi$ i.e. for all P_i such that $\neg P_i y$ there exists t_k such that $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models (\neg P_i t_k \wedge \phi(t_k, P_i))$ and by induction hypothesis $\neg P_i t_k \wedge \phi(t_k, P_i) \in \Gamma^*$. And so by the same argument as the one used in the FO quantifier step of the present induction, $\neg P_i y \rightarrow \exists x(\neg P_i x \wedge \phi(x, P_i)) \in \Gamma^*$. Now suppose $[LFP_{xX}\phi]y \notin \Gamma^*$ i.e. $\neg[LFP_{xX}\phi]y \in \Gamma^*$. Then as Γ^* contains FO(LFP^1) Henkin witnesses, there is a predicate P_m such that $\neg P_m y \wedge \neg\exists x(\neg P_m x \wedge \phi(P_m, x) \in \Gamma^*$. But that contradicts the maximal consistency of Γ^*. Then $\neg[LFP_{xX}\phi]y \notin \Gamma^*$ and by maximal consistency of Γ^*, $[LFP_{xX}\phi]y \in \Gamma^*$.

  - We now show that $[LFP_{xX}\phi]y \in \Gamma^*$ implies $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models [LFP_{xX}\phi]y$. We consider the contraposition

$$\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \not\models \text{ implies } [LFP_{xX}\phi]y \notin \Gamma^*.$$

    So suppose $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \not\models [LFP_{xX}\phi]y$ i.e. $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*} \models \neg[LFP_{xX}\phi]y$ i.e. there exists $A_{P_i} \in \mathbb{A}_{\mathfrak{M}_{\Gamma^*}}$ such that, $g(y) \notin A_{P_i}$ and for all $|t_k| \in M$, $|t_k| \in A_{P_i}$ or $\mathfrak{M}_{\Gamma^*}, g_{\Gamma^*}[x/|t_k|, X/P_i] \models \neg\phi$ and by induction hypothesis for all for all t_k, $\neg P_i y \wedge (P_i t_k \vee \neg\phi(P_i, t_k)) \in \Gamma^*$. And so by the same argument as the one used in the FO quantifier step of the present induction, $\neg P_i y \wedge \forall x(P_i x \vee \neg\phi(P_i, x)) \in \Gamma^*$ i.e. (by maximal consistency) $\neg P_i y \wedge \neg\exists x(\neg P_i x \wedge \phi(P_i, x)) \in \Gamma^*$. Now suppose $[LFP_{xX}\phi]y \in \Gamma^*$. Then by the LFP axiom, for every monadic predicate P_m,

$\neg P_m y \rightarrow \exists x(\neg P_m(x) \wedge \phi(x, P_m)) \in \Gamma^*$. But that contradicts the maximal consistency of Γ^*. Then $[LFP_{xX}\phi]y \notin \Gamma^*$ and by maximal consistency of Γ^*, $\neg[LFP_{xX}\phi]y \in \Gamma^*$.

□

Proposition 8. *$\mathfrak{M}_{\Gamma^*}$ is a FO(LFP^1)-Henkin structure.*

Proof. By construction of Γ^* this is immediate (we introduced a monadic predicate for each parametrically definable subset). □

Theorem 7. *Every consistent set Γ of FO(LFP^1) formulas is satisfiable in $\mathfrak{M}_{\Gamma^*}$.*

Proof. First turn Γ into a FO(LFP^1) maximal consistent set Γ^* with FO(LFP^1)-Henkin witnesses in a possibly richer signature (with extra individual constants and monadic predicates) σ^*. Then build a structure $\mathfrak{M}_{\Gamma^*}$ out of this Γ^*. Then the structure $\mathfrak{M}_{\Gamma^*}$ satisfies Γ^* and hence also the (subset) Γ. □

B Relativization Lemma

Lemma 13 (Relativization lemma)**.** *Let $\mathfrak{M}$ be a Λ-Henkin-structure, g a valuation on $\mathfrak{M}$, ϕ, ψ Λ-formulas and $A = \{x \mid \mathfrak{M}, g \models \psi\}$. If $g(y) \in A$ for every variable y occurring free in ϕ and $g(Y) \in \mathfrak{M} \upharpoonright A$ for every set variable Y occurring free in ϕ, then $\mathfrak{M}, g \models REL(\phi, \psi, x) \Leftrightarrow \mathfrak{M} \upharpoonright A, g \models \phi$.*

Proof. By induction on the complexity of ϕ. Let g be an assignment satisfying the required conditions. Base case: ϕ is an atom and $REL(\phi, \psi, x) = \phi$. So $\mathfrak{M}, g \models \phi \Leftrightarrow \mathfrak{M} \upharpoonright A, g \models \phi$ (by hypothesis, g is a suitable assignment for both models). Inductive hypothesis: the property holds for every ϕ of complexity at most n. Now consider ϕ of complexity $n + 1$.

- $\phi :\approx \phi_1 \wedge \phi_2$ and $REL(\phi_1 \wedge \phi_2, \psi, x) :\approx REL(\phi_1, \psi, x) \wedge REL(\phi_2, \psi, x)$. By induction hypothesis, the property holds for ϕ_1 and for ϕ_2. By the semantics of $\wedge$, it also holds for $\phi_1 \wedge \phi_2$. (Similar for $\vee, \rightarrow, \neg$.)

- $\phi :\approx \exists y \chi$ and $REL(\exists y \chi) :\approx \exists y(\psi[y/x] \wedge REL(\chi, \psi, x))$. By inductive hypothesis, for any node $a \in A$, $\mathfrak{M}, g[a/y] \models REL(\chi, \psi, x) \Leftrightarrow \mathfrak{M} \upharpoonright A, g[a/y] \models \chi$. Hence, by the semantics of $\exists$ and by definition of A, $\mathfrak{M}, g \models \exists y(\psi[y/x] \wedge REL(\chi, \psi, x)) \Leftrightarrow \mathfrak{M} \upharpoonright A, g \models \exists y \chi$.

- $\phi :\approx \exists Y \chi$ and $REL(\exists Y \chi, \psi, x) = \exists Y((Yx \rightarrow \psi) \wedge REL(\chi, \psi, x))$. As every admissible subset of $\mathfrak{M} \upharpoonright A$ is also admissible in $\mathfrak{M}$ (by Proposition 2) it follows by inductive hypothesis that for any $B \in \mathfrak{M} \upharpoonright \mathbb{A}$, $\mathfrak{M}, g[B/Y] \models REL(\chi, \psi, x) \Leftrightarrow \mathfrak{M} \upharpoonright A, g[B/Y] \models \chi$. Hence, by the semantics of $\exists$ and by definition of A, $\mathfrak{M}, g \models \exists Y((Yx \rightarrow \psi) \wedge REL(\chi, \psi, x)) \Leftrightarrow \mathfrak{M} \upharpoonright A, g \models \exists y \chi$.

- $\phi :\approx [TC_{yz}\chi](u, v)$ and $REL([TC_{yz}\chi](u, v), \psi, x) = [TC_{yz} REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]](u, v)$. By definition of TC, the following are equivalent:

  1. $\mathfrak{M} \upharpoonright A, g \models [TC_{yz}\chi](u, v)$,
  2. for all $B \in \mathbb{A}_{\mathfrak{M} \upharpoonright A}$, if $g(u) \in B$ and for all $a, b \in A$, $a \in B$ and $\mathfrak{M} \upharpoonright A, g[a/y, b/z] \models \chi$ implies $b \in B$, then $g(v) \in B$.

By inductive hypothesis, for all $a, b \in A$, $\mathfrak{M}, g[a/y, b/z] \models REL(\chi, \psi, x) \Leftrightarrow \mathfrak{M} \upharpoonright A, g[a/y, b/z] \models \chi$. Hence 2. $\Leftrightarrow$ 3.:

3. for all $B \in \mathbb{A}_{\mathfrak{M}\upharpoonright A}$, if $g(u) \in B$ and for all $a, b \in A$, $a \in B$ and $\mathfrak{M}, g[a/y, b/z] \models REL(\chi, \psi, x)$ implies $b \in B$, then $g(v) \in B$,

By definition of A, 3. $\Leftrightarrow$ 4.:

4. for all $B \in \mathbb{A}_{\mathfrak{M}\upharpoonright A}$, if $g(u) \in B$ and for all $a, b \in dom(\mathfrak{M})$, $a \in B$ and $\mathfrak{M}, g[a/y, b/z] \models REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]$ implies $b \in B$, then $g(v) \in B$,

We claim that 4. $\Leftrightarrow$ 5.:

5. for all $C \in \mathbb{A}_{\mathfrak{M}}$, if $g(u) \in C$ and for all $a, b \in dom(\mathfrak{M})$, $a \in C$ and $\mathfrak{M}, g[a/y, b/z] \models REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]$ implies $b \in C$, then $g(v) \in C$,

which, by the semantics of TC, is equivalent to:

6. $\mathfrak{M}, g \models [TC_{yz} REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]](u, v)$.

It is clear that 5. $\Rightarrow$ 4.. For the 4. $\Rightarrow$ 5. direction, assume 4.. Take any set $C \in \mathbb{A}_{\mathfrak{M}}$ such that $g(u) \in C$ and for all $a, b \in dom(\mathfrak{M})$, $a \in C$ and $\mathfrak{M}, g[a/y, b/z] \models REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]$ implies $b \in C$. Let $B = A \cap C$. By Definition 10, $B \in \mathbb{A}_{\mathfrak{M}\upharpoonright A}$. Now by our assumptions on g and by definition of A, $g[a/y, b/z]$ only assigns points in A. So as $B = A \cap C$, $g(u) \in B$ and for all $a, b \in dom(\mathfrak{M})$, $a \in B$ and $\mathfrak{M}, g[a/y, b/z] \models REL(\chi, \psi, x) \wedge \psi[y/x] \wedge \psi[z/x]$ implies $b \in B$. So by 4., $g(v) \in B$. As $B \subseteq C$, it follows that $g(v) \in C$.

- $\phi :\approx [LFP_{Xy}\chi]z$ and $REL([LFP_{Xy}\chi]z, \psi, x) :\approx [LFP_{Xy}\chi \wedge \psi[y/x]]z$. By definition of LFP, the following are equivalent:

  1. $\mathfrak{M} \upharpoonright A, g \models [LFP_{Xy}\chi]z$,
  2. for all $B \in \mathbb{A}_{\mathfrak{M}\upharpoonright A}$, if for all $a \in A$, $\mathfrak{M} \upharpoonright A, g[a/y, B/X] \models \chi$ implies $a \in B$, then $g(z) \in B$.

  By inductive hypothesis, for all $a \in A$, $B \in \mathfrak{M} \upharpoonright \mathbb{A}$, $\mathfrak{M}, g[a/y, B/X] \models REL(\chi, \psi, x) \Leftrightarrow \mathfrak{M} \upharpoonright A, g[a/y, B/X] \models \chi$. Hence 2. is equivalent to 3.:

  3. for all $B \in \mathbb{A}_{\mathfrak{M}\upharpoonright A}$, if for all $a \in A$, $\mathfrak{M}, g[a/y, B/X] \models REL(\chi, \psi, x)$ implies $a \in B$, then $g(z) \in B$,

  By definition of A, 3. $\Leftrightarrow$ 4.:

  4. for all $B \in \mathbb{A}_{\mathfrak{M}\upharpoonright A}$, if for all $a \in dom(\mathfrak{M})$, $\mathfrak{M}, g[a/y, B/X] \models REL(\chi, \psi, x) \wedge \psi[y/x]$ implies $a \in B$, then $g(z) \in B$,

  We claim that 4. $\Leftrightarrow$ 5.:

  5. for all $C \in \mathbb{A}_{\mathfrak{M}}$, if for all $a \in dom(\mathfrak{M})$, $\mathfrak{M}, g[a/y, C/X] \models REL(\chi, \psi, x) \wedge \psi[y/x]$ implies $a \in C$, then $g(z) \in C$,

  which, by the semantics of LFP, is equivalent to:

  6. $\mathfrak{M}, g \models [LFP_{Xy} REL(\chi, \psi, x) \wedge \psi[y/x]]z$.

It is clear that 5. $\Rightarrow$ 4.. For the 4. $\Rightarrow$ 5. direction, assume 4.. Take any set $C \in \mathbb{A}_{\mathfrak{M}}$ such that for all $a \in dom(\mathfrak{M})$, $\mathfrak{M}, g[a/y, C/X] \models REL(\chi, \psi, x) \wedge \psi[y/x]$ implies $a \in C$. Let $B = A \cap C$. By Definition 10, $B \in \mathbb{A}_{\mathfrak{M}\restriction A}$. Consider $a \in dom(\mathfrak{M})$ such that $\mathfrak{M}, g[a/y, B/X] \models REL(\chi, \psi, x) \wedge \psi[y/x]$. As $REL(\chi, \psi, x)$ is positive in X and X doesn't occur in ψ, $\mathfrak{M}, g[a/y, C/X] \models REL(\chi, \psi, x) \wedge \psi[y/x]$. Also by hypothesis $a \in C$. Now as $\mathfrak{M}, g[a/y] \models \psi[y/x]$, by definition of A, $a \in A$. So $a \in A \cap C$, i.e, $a \in B$ and since we proved it for arbitrary $a \in dom(\mathfrak{M})$, by 4., $g(z) \in B$. As $B \subseteq C$, it follows that $g(z) \in C$.

□

C Ehrenfeucht-Fraïssé Games on Henkin-Structures

Let $\Lambda \in \{\mathsf{MSO}, \mathsf{FO}(\mathsf{TC}^1), \mathsf{FO}(\mathsf{LFP}^1)\}$. In this appendix, we survey Ehrenfeucht-Fraïssé games for FO, MSO, $\mathsf{FO}(\mathsf{TC}^1)$, and $\mathsf{FO}(\mathsf{LFP}^1)$ which are suitable to use on Henkin structures. We also provide adequacy proofs for the MSO game and for the $\mathsf{FO}(\mathsf{TC}^1)$ game.

Let us first introduce basic notions connected to these games. One, rather trivial, sufficient condition for Λ equivalence is the existence of an *isomorphism*. Clearly isomorphic structures satisfy the same Λ-formulas. A more interesting sufficient condition for elementary equivalence is that of Duplicator having a winning strategy in all Λ Ehrenfeucht-Fraïssé games of finite length. To define this, we first need this notion:

Definition 22 (Finite partial isomorphism)**.** A *finite partial isomorphism* between structures $\mathfrak{M}$ and $\mathfrak{N}$ is a finite relation $\{(a_1, b_1), \ldots, (a_n, b_n)\}$ between the domains of $\mathfrak{M}$ and $\mathfrak{N}$ such that for all atomic formulas $\phi(x_1, \ldots, x_n)$, $\mathfrak{M} \models \phi\ [a_1, \ldots, a_n]$ iff $\mathfrak{N} \models \phi\ [b_1, \ldots, b_n]$. Since equality statements are atomic formulas, every finite partial isomorphism is (the graph of) a *injective partial function.*

We will also need the following lemma:

Lemma 14 (Finiteness lemma)**.** *Fix any set $x_1, \ldots, x_k, X_{k+1}, \ldots, X_m$. In a finite relational vocabulary, up to logical equivalence, with these free variables, there are only finitely many Λ-formulas of quantifier depth $\leq n$.*

Proof. This can be shown by induction on k. In a finite relational vocabulary, with finitely many free variables, there are only finitely many atomic formulas. Now, any Λ-formula of quantifier depth $k+1$ is equivalent to a Boolean combination of atoms and formulas of quantifier depth k prefixed by a quantifier. Applying a quantifier to equivalent formulas preserves equivalence and the Boolean closure of a finite set of formulas remains finite, up to logical equivalence. □

Now, as we are concerned with extensions of FO, every Λ game will be defined as an extension of the classical FO game, that we recall here:

Definition 23 (FO Ehrenfeucht-Fraïssé game)**.** The FO Ehrenfeucht-Fraïssé game of length n on structures $\mathfrak{M}$ and $\mathfrak{N}$ (notation: $EF^n_{FO}(\mathfrak{M}, \mathfrak{N})$) is as follows. There are two players, Spoiler and Duplicator. The game has n rounds, each of which consists of a move of Spoiler followed by a move of Duplicator. Spoiler's moves consist of picking an element from one of the two structures, and Duplicator's responses consist of picking an element in the other structure. In this way, Spoiler and Duplicator build up a finite binary relation between the domains of the two structures: initially, the relation is empty; each round, it is extended with another pair. The winning conditions are as follows: if at some point of the game the constructed binary relation is not a finite partial isomorphism, then Spoiler wins immediately. If after each round the relation is a finite partial isomorphism, then the game is won by Duplicator.

Theorem 8 (Adequacy (Ehrenfeucht-Fraïssé)). *Assume a finite relational first-order language. Duplicator has a winning strategy in the game $EF^n_{FO}(\mathfrak{M}, \mathfrak{N})$ iff $\mathfrak{M} \equiv^n_{FO} \mathfrak{N}$. In particular, Duplicator has a winning strategy in all EF-games of finite length between $\mathfrak{M}$ and $\mathfrak{N}$ if and only if $\mathfrak{M} \equiv_{FO} \mathfrak{N}$.*

Proof. The proof for the first order case is classic. We refer the reader to the proof given by Flum and Ebbinghaus in [5]. □

For technical convenience in the course of inductive proofs, we extend the notion of FO parameter by considering set parameters, i.e., instead of interpreting a set variable as a name of the set A, we can add a new monadic predicate A to the signature. The new predicates and the sets they name are called set parameters. (This is similar to the FO notion which can be found in [11].) We will work with parametrized structures, i.e., the assignment is possibly non empty at the beginning of the game, which can begin with some "handicap" for Duplicator, which is some preliminary set of already "distinguished objects and sets" (for distinguished objects, think, for instance, about the situation where we would allow individual constants in the language).

C.1 Ehrenfeucht-Fraïssé Game for MSO

We define a necessary and sufficient condition for MSO equivalence by extending Ehrenfeucht-Fraïssé games from FO to MSO. This game has already been defined in the literature, see for instance [14]. For the sake of the induction, we will work with expanded structures (i.e. structures considered together with partial valuations).

Definition 24 (MSO Ehrenfeucht-Fraïssé game). Consider $\mathfrak{M}$ together with $\bar{A} \in \mathbb{A}^r_{\mathfrak{M}}$, $\bar{a} \in dom(\mathfrak{M})^s$, $\mathfrak{N}$ together with $\bar{B} \in \mathbb{A}^r_{\mathfrak{N}}$, $\bar{b} \in dom(\mathfrak{N})^s$ and $r \geq 0$, $s \geq 0$, $n \geq 0$. The MSO Ehrenfeucht-Fraïssé game $EF^n_{MSO}((\mathfrak{M}, \bar{A}, \bar{a}), (\mathfrak{N}, \bar{B}, \bar{b}))$ of length n on expanded structures $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ is defined as for the first-order case, except that each time she chooses a structure, Spoiler can choose either an element or an admissible subset of its domain. For a given $A_{r+1} \in \mathbb{A}_{\mathfrak{M}}$ chosen by Spoiler, $(\mathfrak{M}, \bar{A}, \bar{a})$ is expanded to $(\mathfrak{M}, \bar{A}, A_{r+1}, \bar{a})$. Duplicator then responds by choosing $B_{r+1} \in \mathbb{A}_{\mathfrak{N}}$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ is expanded to $(\mathfrak{N}, \bar{B}, B_{r+1}, \bar{b})$. The game goes on with the so expanded structures. The winning conditions are as follows: if at some point of the game $\bar{a} \mapsto \bar{b}$ is not a finite partial isomorphism from $(\mathfrak{M}, \bar{A}, A_{r+1})$ to $(\mathfrak{N}, \bar{B}, B_{r+1})$, then Spoiler wins immediately. If after each round the relation is a finite partial isomorphism, then the game is won by Duplicator.

Theorem 9 (Adequacy). *Assume a finite relational MSO language. Given $\mathfrak{M}$ and $\mathfrak{N}$, $\bar{A} \in \mathbb{A}^r_{\mathfrak{M}}$, $\bar{B} \in \mathbb{A}^r_{\mathfrak{N}}$, $\bar{a} \in dom(\mathfrak{M})^s$, $\bar{b} \in dom(\mathfrak{N})^s$ and $r \geq 0$, $s \geq 0$, $n \geq 0$, Duplicator has a winning strategy in the game $EF^n_{MSO}((\mathfrak{M}, \bar{A}, \bar{a}), (\mathfrak{N}, \bar{B}, \bar{b}))$ iff $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ satisfy the same MSO formulas of quantifier depth n. In particular, Duplicator has a winning strategy in all EF_{MSO}-games of finite length between $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ if and only if $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ satisfy the same MSO formulas.*

Proof. $\Rightarrow$ From the existence of a winning strategy for Duplicator in $EF^n_{MSO}((\mathfrak{M}, \bar{A}, \bar{a}), (\mathfrak{N}, \bar{B}, \bar{b}))$ to the fact that $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ satisfy the same MSO formulas of quantifier depth n.

By induction on n.

Base step: With 0 round the initial match between the distinguished objects must have been a partial isomorphism for Duplicator to win. Thus $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ agree on all atomic formulas and on their Boolean combinations (which are precisely the formulas of quantifiers depth 0).

Inductive step: The inductive hypothesis says that, for any two expanded MSO structures, if Duplicator can win their comparison game over n rounds, then they agree on all MSO formulas up to quantifier depth n. Now assume that for some $(\mathfrak{M}, \bar{A}, \bar{a})$, $(\mathfrak{N}, \bar{B}, \bar{b})$ Duplicator has a winning strategy for the game over $n+1$ rounds. Consider any MSO formula ϕ of quantifier depth $n+1$. Any such sentence should be equivalent to a Boolean combination of atoms and formulas of the form $\exists x_i\ \chi(x_i)$ and $\exists X_i\ \psi(X_i)$, with $\chi(x_i)$, $\psi(X_i)$ of quantifier depth at most n. Thus it suffices to show that $(\mathfrak{M}, \bar{A}, \bar{a})$, $(\mathfrak{N}, \bar{B}, \bar{b})$ agree on the latter forms. They do so on atoms, as Duplicator can certainly win over 0 rounds. So let suppose $(\mathfrak{M}, \bar{A}, \bar{a}) \models \exists X_i\ \psi(X_i)$ (the case $(\mathfrak{M}, \bar{A}, \bar{a}) \models \exists x_i\ \chi(x_i)$ is symmetric). Then for some $A_i \in \mathbb{A}_{\mathfrak{M}}$, $(\mathfrak{M}, \bar{A}, A_i, \bar{a}) \models \psi(Xi)$. Now, Duplicator's given winning strategy has a response for whatever Spoiler might do in the $n+1$ round game. In particular, let Spoiler select A_i in $\mathbb{A}_{\mathfrak{M}}$. Then Duplicator has a response B_i in $\mathbb{A}_{\mathfrak{N}}$ such that her remaining strategy still gives her a win in the n-round game played on $(\mathfrak{M}, \bar{A}, A_i, \bar{a})$ and $(\mathfrak{N}, \bar{B}, B_i, \bar{b})$. By the inductive hypothesis, these expanded structures agree on all formulas up to quantifier depth n and hence also on $\psi(X_i)$. Therefore $(\mathfrak{N}, \bar{B}, B_i, \bar{b}) \models \psi(X_i)$ and hence $(\mathfrak{N}, \bar{B}, \bar{b}) \models \exists X_i\ \psi(X_i)$.

$\Leftarrow$ From the fact that $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ satisfy the same MSO formulas of quantifier depth n to the existence of a winning strategy for Duplicator in $EF^n_{MSO}(\mathfrak{M}, \mathfrak{N})$.

Base step: Doing nothing is a strategy for Duplicator.

Inductive step: The inductive hypothesis says that, for any two expanded MSO structures, if they agree on all MSO formulas up to quantifier depth n, then Duplicator has a winning strategy in the n-round corresponding game. Now, assume that some structures $(\mathfrak{M}, \bar{A}, \bar{a})$, $(\mathfrak{N}, \bar{B}, \bar{b})$ agree on all MSO formulas of quantifier depth $n+1$. We can infer that Duplicator has a winning strategy in the $n+1$-round game. The first move in her strategy is as follows. Let Spoiler choose $A_i \in \mathbb{A}_{\mathfrak{M}}$ (the case where she rather chooses a_i in $dom(\mathfrak{M})$ is symmetric). Now, Duplicator looks at the set of MSO formulas of quantifier depth $n+1$ that hold of A_i in $(\mathfrak{M}, \bar{A}, \bar{a})$. By the finiteness lemma, this set is finite modulo logical equivalence, and hence, one existential formula $\exists X_i\ \psi(X_i)$ true in the structure summarizes all this information. As $(\mathfrak{M}, \bar{A}, \bar{a})$, $(\mathfrak{N}, \bar{B}, \bar{b})$ agree on all MSO formulas of quantifier depth $n+1$, and $\exists X_i\ \psi(X_i)$ is such a sentence, it also holds in $(\mathfrak{N}, \bar{B}, \bar{b})$. So, Duplicator can choose a witness B_i. Then, the so expanded structures $(\mathfrak{M}, \bar{A}, A_i, \bar{a})$, $(\mathfrak{N}, \bar{B}, B_i, \bar{b})$ agree on all MSO sentences up to quantifier depth n, and by the inductive hypothesis, Duplicator has a winning strategy in the remaining n-round game between them. Her initial response plus the latter gives her over-all strategy over $n+1$ rounds.

□

Note that this proof holds for MSO with any semantics (e.g. standard, Henkin...).

We are interested in "choice of an element" versus "quantification", but neither the exact domain of quantification does never play any role in our reasoning.

Corollary 3. For structures $\mathfrak{M}$, $\mathfrak{N}$ and $n \geq 0$, Duplicator has a winning strategy in $EF^n_{MSO}(\mathfrak{M}, \mathfrak{N})$ if and only if $\mathfrak{M} \equiv^n_{MSO} \mathfrak{N}$. In particular, Duplicator has a winning strategy in all EF_{MSO}-games of finite length between $\mathfrak{M}$ and $\mathfrak{N}$ if and only if $\mathfrak{M} \equiv_{MSO} \mathfrak{N}$.

C.2 Ehrenfeucht-Fraïssé Game for $\mathsf{FO(TC^1)}$

The game that we will be introducing in this section had been already mentioned in passing by Grädel in [9] as an alternative to the game he used. We will show that it is adequate on Henkin-structures.

Definition 25 ($\mathsf{FO(TC^1)}$ Ehrenfeucht-Fraïssé game)**.** In $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$ there are two types of moves, $\exists$ (or point) moves and $\mathsf{FO(TC^1)}$ moves. Each point move extends an assignment $\{\bar{a} \mapsto \bar{b}\}$ with elements $a_k \in dom(\mathfrak{M}), b_k \in dom(\mathfrak{N})$. Each $\mathsf{FO(TC^1)}$ move extends an assignment $\{\bar{a} \mapsto \bar{b}\}$ with elements $a_k, a_{k+1} \in dom(\mathfrak{M}), b_k, b_{k+1} \in dom(\mathfrak{N})$. After each move, Spoiler chooses the kind of move to be played. We assume that the assignment $\{\bar{a} \mapsto \bar{b}\}$ has to be extended. The $\exists$ move is defined as in the FO case. The $\mathsf{FO(TC^1)}$ move is as follows:

Spoiler considers two pebbles (a_i, b_i) and (a_j, b_j) on the board and depending on the structure that he chooses to consider, he plays:

- either $A \in \mathbb{A}_{\mathfrak{M}}$ with $a_i \in A$ and $a_j \notin A$. Duplicator then answers with $B \in \mathbb{A}_{\mathfrak{N}}$ such that $b_i \in B$ and $b_j \notin B$. Spoiler now picks $b_k \in B, b_{k+1} \notin B$ and Duplicator answers with $a_k \in A, a_{k+1} \notin A$.
- either $B \in \mathbb{A}_{\mathfrak{N}}$ with $b_i \in B$ and $b_j \notin B$. Duplicator then answers with $A \in \mathbb{A}_{\mathfrak{M}}$ such that $a_i \in A$ and $a_j \notin A$. Spoiler now picks $a_k \in A, a_{k+1} \notin A$ and Duplicator answers with $b_k \in B, b_{k+1} \notin B$.

In each $\mathsf{FO(TC^1)}$ move, the assignment is extended with $a_k \mapsto b_k, a_{k+1} \mapsto b_{k+1}$. After n moves, Duplicator has won if the constructed assignment $\bar{a} \mapsto \bar{b}$ is a partial isomorphism (i.e. the game continues with the two new pebbles in each structure, but the sets A and B are forgotten).

Theorem 10 (Adequacy)**.** *Assume a finite relational $\mathsf{FO(TC^1)}$ language. Given $\mathfrak{M}$ and $\mathfrak{N}$, $\bar{a} \in M^s$, $\bar{b} \in N^s$ and $r \geq 0$, $s \geq 0$, $n \geq 0$, Spoiler has a winning strategy in the game $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$ iff there is a $\mathsf{FO(TC^1)}$ formula of quantifier depth n distinguishing $(\mathfrak{M}, \bar{a})$ and $(\mathfrak{N}, \bar{b})$.*

Proof.

⇒ From the existence of a winning strategy for Spoiler in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$ to the existence of a $\mathsf{FO(TC^1)}$ formula of quantifier depth n distinguishing $(\mathfrak{M}, \bar{a})$ and $(\mathfrak{N}, \bar{b})$.

By induction on n.

Base step: With 0 round the initial match between distinguished objects must have failed to be a partial isomorphism for Spoiler to win. This implies that $(\mathfrak{M}, \bar{a})$ and $(\mathfrak{N}, \bar{b})$ disagree on some atomic formula.

Inductive step: The inductive hypothesis says that for any two structures, if Spoiler can win their comparison game over n rounds, then the structures disagree on some $\mathsf{FO}(\mathsf{TC}^1)$ formula of quantifier depth n. Now assume that for some structures $(\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b})$, Spoiler has a winning strategy for the game over $n+1$ rounds. Let us reason on Spoiler's first move in the game. It can either be a $\mathsf{FO}(\mathsf{TC}^1)$ or an $\exists$ move.

If it is an $\exists$ move, then it means that Spoiler picks an element a in one of the two structures, so that no matter what element b Duplicator picks in the other, Spoiler has an n-round winning strategy. But then we can use the induction hypothesis, and find for each such b a formula $\phi_b(x)$ that distinguishes $(\mathfrak{M}, \bar{a}, a)$ from $(\mathfrak{N}, \bar{b}, b)$. In fact we can assume that in each case the respective formula is true of $(\mathfrak{M}, \bar{a}, a)$ and false of $(\mathfrak{N}, \bar{b}, b)$ (by negating the formula if needed). Now take the big conjunction $\phi(x)$ of all these formulas (which is equivalent to a finite formula according to the finiteness lemma) and prefix it with an existential quantifier. Then the resulting formula is true in $(\mathfrak{M}, \bar{a})$ but false in $(\mathfrak{N}, \bar{b})$. It is true in $(\mathfrak{M}, \bar{a})$ if we pick a for the existentially quantified variable. And no matter which element we pick in $(\mathfrak{N}, \bar{b})$, it will always falsify one of the conjuncts in the formula, by construction. So, the new formula is false in $(\mathfrak{N}, \bar{b})$. I.e., $\exists x \phi(x)$ of quantifier depth $n+1$ distinguishes $(\mathfrak{M}, \bar{a})$ and $(\mathfrak{N}, \bar{b})$.

If Spoiler's first move is a $\mathsf{FO}(\mathsf{TC}^1)$ move, then it means that Spoiler picks a subset in one structure, let say $A \in \mathbb{A}_{\mathfrak{M}}$ (with $a_i \in A$ and $a_j \notin A$), so that no matter which $B \in \mathbb{A}_{\mathfrak{N}}$ (with $b_i \in B$ and $b_j \notin B$) Duplicator picks in the other structure, Spoiler can pick $b_k \in B$, $b_{k+1} \notin B$ such that no matter which $a_k \in A$, $a_{k+1} \notin A$ Duplicator picks, Spoiler has an n-round winning strategy. For each B that might be chosen by Duplicator, Spoiler's given strategy gives a fixed couple b_k, b_{k+1}. For each response a_k, a_{k+1} of Duplicator, we thus obtain by inductive hypothesis a discriminating formula $\phi_{B,a_k,a_{k+1}}(x, y)$ that we can assume to be true in $(\mathfrak{N}, \bar{b})$ for b_k, b_{k+1} and false in $(\mathfrak{M}, \bar{a})$ for a_k, a_{k+1}. Now for each B, let us take the big conjunction $\Phi_B(x, y)$ of all these formulas (which is finite, by the finiteness lemma). We can then construct the big disjunction $\Phi(x, y)$ (again finite, by the same lemma) of all the formulas $\Phi_B(x, y)$.

Considering the first round in the game together with the inductive hypothesis, note that it holds in $(\mathfrak{M}, \bar{a})$ that $\exists X(a_i \in X \wedge a_j \notin X \wedge \forall xy((x \in X \wedge y \notin X) \rightarrow \neg\Phi(x, y)))$. Indeed, by induction hypothesis, any couple $a_k \in A, a_{k+1} \notin A$ that Duplicator might choose in $dom(\mathfrak{M})$ will always falsify at least one of the conjuncts of each $\Phi_B(x, y)$. Finally, the formula $\Phi(x, y)$ being constructed as the disjunction of all the formulas $\Phi_B(x, y)$, any such couple a_k, a_{k+1} will also falsify $\Phi(x, y)$. Now $\exists X(a_i \in X \wedge a_j \notin X \wedge \forall xy((x \in X \wedge y \notin X) \rightarrow \neg\Phi(x, y)))$ is equivalent[5] to $\exists X(a_i \in X \wedge a_j \notin X \wedge \neg\exists xy(x \in X \wedge \Phi(x, y) \wedge y \notin X))$, which means that $(\mathfrak{M}, \bar{a}) \not\models [TC_{xy}\Phi(x, y)](a_i, a_j)$.

On the other hand for the same reasons, note that it holds in $(\mathfrak{N}, \bar{b})$ that $\forall X((b_i \in X \wedge b_j \notin X) \rightarrow \exists xy(x \in X \wedge y \notin X \wedge \Phi(x, y)))$. Indeed, by induction hypothesis, for each B that Duplicator might choose in $\mathbb{A}_{\mathfrak{N}}$ Spoiler will always be able to find a couple $b_k \in B, b_{k+1} \notin B$ satisfying all the conjuncts of the corresponding formulas $\Phi_B(x, y)$. Finally, the formula $\Phi(x, y)$ being constructed as the disjunction of all the formulas $\Phi_B(x, y)$, such a couple a_k, a_{k+1} will also satisfy $\Phi(x, y)$. Now $\forall X((b_i \in$

[5] As $\neg(p \rightarrow q) \equiv p \wedge \neg q$.

$X \wedge b_j \notin X) \rightarrow \exists xy(x \in X \wedge y \notin X \wedge \Phi(x,y)))$ is equivalent[6] to $\forall X(b_i \notin X \vee b_j \in X \vee \exists xy(x \in X \wedge y \notin X \wedge \Phi(x,y)))$, which means that $(\mathfrak{N}, \bar{b}) \models [TC_{xy}\Phi(x,y)](b_i, b_j)$.

Let u be a name for the parameters a_i, b_i and v for b_i, b_j. $[TC_{xy}\Phi(x,y)](u,v)$ of quantifier depth $n+1$ distinguishes $(\mathfrak{N}, \bar{a})$ and $(\mathfrak{M}, \bar{b})$.

$\Leftarrow$ From the existence of a $\mathsf{FO(TC^1)}$ formula of quantifier depth n distinguishing $(\mathfrak{M}, \bar{a})$ and $(\mathfrak{N}, \bar{b})$ to the existence of a winning strategy for Spoiler in $EF^n_{FO+TC}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$.

Base step: Doing nothing is a strategy for Spoiler.

Inductive step: The inductive hypothesis says that, for any two structures, if they disagree on some $\mathsf{FO(TC^1)}$ formula of quantifier depth n, then Duplicator has a winning strategy in the n-round game. Now, assume that some expanded structures $(\mathfrak{M}, \bar{a})$, $(\mathfrak{N}, \bar{b})$ disagree on some $\mathsf{FO(TC^1)}$ formula χ of quantifier depth $n+1$. Any such formula must be equivalent to a Boolean combination of formulas of the form $\exists x\psi(x)$ and $[TC_{xy}\phi(x,y)](u,v)$ with ψ, ϕ of quantifier depth at most n. If χ distinguishes the two structures, then there is at least one component of this Boolean combination which suffices distinguishing them.

Let us first suppose that it is of the form $\exists x\psi(x)$ and such that $(\mathfrak{M}, \bar{a}) \models \exists x\psi(x)$ whereas $(\mathfrak{N}, \bar{b}) \not\models \exists x\psi(x)$. Then it means that there exists an object $a \in \mathit{dom}(\mathfrak{M})$ such that $(\mathfrak{M}, \bar{a}) \models \psi(a)$ whereas for every object $b \in \mathit{dom}(\mathfrak{N})$, $(\mathfrak{N}, \bar{b}) \not\models \psi(b)$. But then we can use our induction hypothesis and find for each such b a winning strategy for Spoiler in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}, a), (\mathfrak{N}, \bar{b}, b))$. We can infer that Spoiler has a winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$. His first move consists in picking the object a in M and for each response b in N of Duplicator, the remaining of his winning strategy is the same as in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}, a), (\mathfrak{N}, \bar{b}, b))$.

Let us now suppose that $[TC_{xy}\phi(x,y)](u,v)$ of quantifier depth $n+1$ distinguishes the two structures such that $(\mathfrak{M}, \bar{a}) \models [TC_{xy}\phi(x,y)](u,v)$ i.e. it holds in $(\mathfrak{M}, \bar{a})$ that $\forall X((a_i \in X \wedge a_j \notin X) \rightarrow \exists xy(x \in X \wedge y \notin X \wedge \phi(x,y)))$, whereas $(\mathfrak{N}, \bar{b}) \not\models [TC_{xy}\phi(x,y)](u,v)$ i.e. it holds in $(\mathfrak{N}, \bar{b})$ that $\exists X(b_i \in X \wedge b_j \notin X \wedge \neg\exists xy(x \in X \wedge \phi(x,y) \wedge y \notin X))$. We want to show that Spoiler has a winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$. Let us describe her first move. She first chooses $(\mathfrak{N}, \bar{b})$ and $B \in \mathbb{A}_{\mathfrak{N}}$ such that $b_i \in B \wedge b_j \notin B \wedge \neg\exists xy(x \in B \wedge \phi(x,y) \wedge y \notin B)$. By definition of TC, such a set exists. Duplicator has to respond by picking a set A in $\mathbb{A}_{\mathfrak{M}}$. Spoiler then picks $a_k \in A$ and $a_{k+1} \notin A$ such that $(\mathfrak{M}, \bar{a}) \models \phi(a_k, a_{k+1})$. This is possible because by definition of TC, for any possible choice A of Duplicator we have $\exists xy(x \in A \wedge y \notin A \wedge \phi(x,y))$. But that means that Duplicator is now stuck and has to pick $b_k \in B$ and $b_{k+1} \notin B$ such that $(\mathfrak{N}, \bar{b}) \not\models \phi(b_k, b_{k+1})$. Consequently, we have $(\mathfrak{N}, \bar{b}, b_k, b_{k+1}) \not\models \phi(x,y)$, whereas $(\mathfrak{M}, \bar{a}, a_k, a_{k+1}) \models \phi(x,y)$. As $\phi(x,y)$ is of quantifier depth n, by induction hypothesis, Spoiler has a winning strategy in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}, a_k, a_{k+1}), (\mathfrak{N}, \bar{b}, b_k, b_{k+1}))$. The remaining of Spoiler's winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}), (\mathfrak{N}, \bar{b}))$ (i.e. after her first move, that we already accounted for) is consequently as in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}, \bar{a}, a_k, a_{k+1}), (\mathfrak{N}, \bar{b}, b_k, b_{k+1}))$.

$\square$

[6] As $p \rightarrow q \equiv \neg p \vee q$.

Corollary 4. For structures $\mathfrak{M}$, $\mathfrak{N}$ and $n \geq 0$, Duplicator has a winning strategy in $EF^n_{\mathsf{FO(TC^1)}}(\mathfrak{M}, \mathfrak{N})$ if and only if $\mathfrak{M} \equiv^n_{\mathsf{FO(TC^1)}} \mathfrak{N}$. In particular, Duplicator has a winning strategy in all $EF_{\mathsf{FO(TC^1)}}$-games of finite length between $\mathfrak{M}$ and $\mathfrak{N}$ if and only if $\mathfrak{M} \equiv_{\mathsf{FO(TC^1)}} \mathfrak{N}$.

C.3 Ehrenfeucht-Fraïssé Game for $\mathsf{FO(LFP^1)}$

There are two classical equivalent syntactic ways to define the syntax of $\mathsf{FO(LFP^1)}$: the one we used in Section 1.2 and an other one, dispensing with restrictions to positive formulas, but allowing negations only in front of atomic formulas and introducing a greatest fixpoint operator as the dual of the least fixpoint operator (also $\forall$ cannot be defined using $\exists$ and has to be introduced separately, similarly for the Boolean connectives). This second way to define $\mathsf{FO(LFP^1)}$ turns out to be more convenient to define an adequate Ehrenfeucht-Fraïssé game. The game is suitable to use on Henkin structures because the semantics on which it relies is merely a syntactical variant of the one given in Section 3. Now the $\mathsf{FO(LFP^1)}$ formulas $[LFP_{x,X}\phi(x, X)]y$ and $[GFP_{x,X}\phi(x, X)]y$, stating that a point belongs to the least fixpoint, or respectively, to the greatest fixpoint induced by the formula ϕ satisfy the following equations:

$$[LFP_{x,X}\phi(x, X)]y \leftrightarrow \forall X(\neg Xy \rightarrow \exists x(\neg Xx \wedge \phi(x, X)))$$
$$[GFP_{x,X}\phi(x, X)]y \leftrightarrow \exists X(Xy \wedge \forall x(Xx \rightarrow \phi(x, X)))$$

This is the key idea behind an Ehrenfeucht-Fraïssé game defined by Uwe Bosse in [2] for least fixpoint logic LFP (i.e. where fixpoints are not only considered for monadic operators, but for any n-ary operator). $\mathsf{FO(LFP^1)}$ being simply the monadic fragment of LFP, the game for LFP can be adapted to $\mathsf{FO(LFP^1)}$ in a straightforward way:

Definition 26 ($\mathsf{FO(LFP^1)}$ Ehrenfeucht-Fraïssé game)**.** Let $n \geq 0$, $r \geq 0$, $s \geq 0$. In the game $EF^n_{\mathsf{FO(LFP^1)}}((\mathfrak{M}, \bar{A}, \bar{a}), (\mathfrak{N}, \bar{B}, \bar{b}))$, there are two types of moves, point and fixpoint moves. Each move extends an assignment $\bar{a} \mapsto \bar{b}, \bar{A} \mapsto \bar{B}$ with elements $a_s \in dom(\mathfrak{M}), b_s \in dom(\mathfrak{N})$, and possibly (in the case of fixpoint moves) with sets $A_r \in \mathbb{A}_{\mathfrak{M}}, B_r \in \mathbb{A}_{\mathfrak{N}}$. After each move, Spoiler chooses the kind of move to be played. We assume that the assignment $\bar{a} \mapsto \bar{b}, \bar{A} \mapsto \bar{B}$ has to be extended. Now the following moves are possible:

- $\exists$ move: Spoiler chooses $a_{s+1} \in dom(\mathfrak{M})$ and Duplicator $b_{s+1} \in dom(\mathfrak{N})$.
- $\forall$ move: Spoiler chooses $b_{s+1} \in dom(\mathfrak{N})$ and Duplicator $a_{s+1} \in dom(\mathfrak{M})$.

In each point move, the assignment is extended by $a_{s+1} \mapsto b_{s+1}$.

- LFP move: Spoiler chooses $B_{r+1} \in \mathbb{A}_{\mathfrak{N}} \setminus \{dom(\mathfrak{N})\}$ with some pebble $b_i \notin B_{r+1}$ and Duplicator responds with $A_{r+1} \in \mathbb{A}_{\mathfrak{M}} \setminus \{dom(\mathfrak{M})\}$.

  Now Spoiler chooses in M a new element $a_{s+1} \notin A_{r+1}$ and Duplicator answers in N with $b_{s+1} \notin B_{r+1}$.

- GFP move: Spoiler chooses $A_{r+1} \in \mathbb{A}_{\mathfrak{M}} \setminus \{dom(\mathfrak{M})\}$ with some pebble $a_i \in A_{r+1}$ and Duplicator responds with $B_{r+1} \in \mathbb{A}_{\mathfrak{N}} \setminus \{dom(\mathfrak{N})\}$ such that $B_{r+1} \neq \emptyset$.

  Now Spoiler chooses in $dom(\mathfrak{N})$ a new element $b_{s+1} \in B_{r+1}$ and Duplicator answers in $dom(\mathfrak{M})$ with $a_{s+1} \in A_{r+1}$.

In each fixpoint move the assignment is extended by $A_{r+1} \mapsto B_{r+1}, a_{s+1} \mapsto b_{s+1}$.

After n moves, Duplicator has won if the constructed element assignment $\bar{a} \mapsto \bar{b}$ is a partial isomorphism and for the subset assignment $\bar{A} \mapsto \bar{B}$, for any $1 \leq j \leq r, i \leq s$:

$$a_i \in A_j \text{ implies } b_i \in B_j$$

We call an assignment with these properties a *posimorphism*.

Theorem 11 (Adequacy). *Assume a finite relational* $\mathsf{FO(LFP^1)}$ *language. Given* $\mathfrak{M}$ *and* $\mathfrak{N}$*,* $\bar{A} \in \mathbb{A}^r_{\mathfrak{M}}$*,* $\bar{B} \in \mathbb{B}^r_{\mathfrak{N}}$*,* $\bar{a} \in dom(\mathfrak{M})^s$*,* $\bar{b} \in dom(\mathfrak{N})^s$ *and* $r \geq 0$*,* $r \geq 0$*,* $n \geq 0$*, Duplicator has a winning strategy in the game* $EF^n_{\mathsf{FO(LFP^1)}}((\mathfrak{M}, \bar{A}, \bar{a}), (\mathfrak{N}, \bar{B}, \bar{b}))$ *iff* $(\mathfrak{M}, \bar{A}, \bar{a})$ *and* $(\mathfrak{N}, \bar{B}, \bar{b})$ *satisfy the same* $\mathsf{FO(LFP^1)}$ *formula of quantifier depth* n*.*

Proof. We refer the reader to Uwe Bosse [2]. □

D Fusion Lemmas on Henkin-Structures

Let $\Lambda \in \{\mathsf{MSO}, \mathsf{FO(TC^1)}, \mathsf{FO(LFP^1)}\}$. In this Appendix, we show our analogues of Feferman-Vaught theorem for fusions of Λ-Henkin-structures. We refer to them as Λ-*fusion lemmas* in the main part of the paper, even though they will be formally stated as theorems or corollaries below. What we show is, more precisely, that fusion of Λ-Henkin-structures preserve Λ-equivalence.

In order to give inductive proofs for MSO and $\mathsf{FO(LFP^1)}$, it will be more convenient to consider parametrized Λ-Henkin-structures where the set of set parameters is closed under union, this notion being defined below. This is safe because whenever two parametrized structures $(\mathfrak{M}, \bar{A}, \bar{a})$ and $(\mathfrak{N}, \bar{B}, \bar{b})$ are n-Λ-equivalent, it follows trivially that $\mathfrak{M}$ and $\mathfrak{N}$ considered together with a subset of this set of parameters are also n-Λ-equivalent.

Definition 27. Let $A_1, \ldots, A_k$ be a finite sequence of set parameters. We define $(A_1, \ldots, A_k)^{\cup}$ as the finite sequence of set parameters obtained by closing the set $\{A_1, \ldots, A_k\}$ under union in such a way that $(A_1, \ldots, A_k)^{\cup} = \{\bigcup_{i \in I} A_i | I \subseteq \{1, \ldots, k\}\}$. (We additionally assume that this set is ordered in a fixed canonical way, depending on the index sets I.)

D.1 Fusion Lemma for MSO

Theorem 12. *Whenever* $(\mathfrak{M}_i, \bar{A}_i, \bar{a}_i) \equiv^n_{MSO} (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i)$ *for all* $1 \leq i \leq k$ *(with* $\bar{a}_i$ *a sequence of first-order parameters of the form* $a_{i_1}, \ldots, a_{i_m}$ *with* $m \in \mathbb{N}$ *and* $\bar{A}_i$ *a sequence of set parameters of the form* $A_{i_1}, \ldots, A_{i_{m'}}$ *with* $m' \in \mathbb{N}$*, similarly for the* $\bar{b}_i$ *and* $\bar{B}_i$*), then also* $\bigoplus^f_{1 \leq i \leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k \equiv^n_{MSO} \bigoplus^f_{1 \leq i \leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k$*.*

Proof. We define a winning strategy for Duplicator in the game $EF^n_{MSO}((\bigoplus^f_{1 \leq i \leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k), (\bigoplus^f_{1 \leq i \leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k))$ out of her winning strategies in the games $EF^n_{MSO}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$ by induction on n.

Base step: $n = 0$, doing nothing is a strategy for Duplicator. We need to show that $(\bigoplus^f_{1 \leq i \leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k)$ and $(\bigoplus^f_{1 \leq i \leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k)$ agree on all atomic formulas. Now in the fusion structures, each atomic formula is defined by f in terms of a σ^*-quantifier free formula that is evaluated in the corresponding disjoint union structure. So it is enough to show that the disjoint union structures agree on all atomic σ^*-formulas and on their Boolean combinations. The initial match between the distinguished objects in $(\mathfrak{M}_i, \bar{A}_i, \bar{a_i})$ and $(\mathfrak{N}_i, \bar{B}_i, \bar{b_i})$ is a partial isomorphism for every $1 \leq i \leq k$, so it is also one for $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}$ and $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}$ i.e. the two disjoint union structures extended with FO parameters agree on all σ^*-atomic formulas. We still need to show that it is also one for $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, \bar{a_1}, \ldots, \bar{a_k}$ and $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, \bar{b_1}, \ldots, \bar{b_k}$ i.e. the two disjoint union structures extended with FO parameters and the closure under union of set parameters agree on all σ^*-atomic formulas. It is enough to point that for any parameter a_{i_j}, for any $I \subseteq \{i_1, \ldots, i_{m'}, \ldots, k_1, k_{m'}\}$ by construction of $\bigcup_{i\in I} A_i$ in $(\bar{A_1}, \ldots, \bar{A_k})^{\cup}$, the following are equivalent:

- $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, \bar{a_1}, \ldots, \bar{a_k} \models \bigcup_{i\in I} A_i a_{i_j}$,
- $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, A_{i_l}, \bar{a_1}, \ldots, \bar{a_k} \models A_{i_l} a_{i_j}$ for some i_l in I.

Similarly for any parameter b_{i_j}, by construction of $\bigcup_{i\in I} B_i$ in $(\bar{B_1}, \ldots, \bar{B_k})^{\cup}$, the following are equivalent:

- $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, \bar{b_1}, \ldots, \bar{b_k} \models \bigcup_{i\in I} B_i b_{i_j}$,
- $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, B_{i_l}, \bar{b_1}, \ldots, \bar{b_k} \models B_{i_l} b_{i_j}$ for some i_l in I.

But by Duplicator's winning strategy in the small structure games, we know that the following are equivalent:

- $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, A_{i_l}, \bar{a_1}, \ldots, \bar{a_k} \models A_{i_l} a_{i_j}$ for some i_l in I.
- $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, B_{i_l}, \bar{b_1}, \ldots, \bar{b_k} \models B_{i_l} b_{i_j}$ for some i_l in I.

So the following are also equivalent:

- $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, \bar{a_1}, \ldots, \bar{a_k} \models \bigcup_{i\in I} A_i a_{i_j}$,
- $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, \bar{b_1}, \ldots, \bar{b_k} \models \bigcup_{i\in I} B_i b_{i_j}$,

So the two extended disjoint union structures agree on all σ^*-atomic formulas. Now relying on the semantics of Boolean connectives, it can be shown by induction on the complexity of quantifier free sentences that they also agree on all Boolean combinations of atomic σ^*-sentences.

Inductive step: the inductive hypothesis says that whenever Duplicator has a winning strategy in $EF^n_{MSO}((\mathfrak{M}_i, \bar{A}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{B}_i, \bar{b_i}))$ for all $1 \leq i \leq k$, he also has one in $EF^n_{MSO}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, \bar{a_1}, \ldots, \bar{a_k}), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, \bar{b_1}, \ldots, \bar{b_k}))$. We want to show that this also holds when the length of the games is $n + 1$. Suppose Duplicator has a winning strategy in $EF^{n+1}_{MSO}((\mathfrak{M}_i, \bar{A}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{B}_i, \bar{b_i}))$ for all $1 \leq i \leq k$. We describe Duplicator's answer to Spoiler's first move in $EF^{n+1}_{MSO}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{A_1}, \ldots, \bar{A_k}, \bar{a_1}, \ldots, \bar{a_k}), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{B_1}, \ldots, \bar{B_k}, \bar{b_1}, \ldots, \bar{b_k}))$. It then follows by induction hypothesis, that he has a winning strategy in the remaining n-length game.

- Spoiler's first move is a point move. Suppose Spoiler picks a in $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$. Then $a \in dom(\mathfrak{M}_i)$ for some $1 \leq i \leq k$. So Duplicator uses his winning strategy in $EF^{n+1}_{MSO}((\mathfrak{M}_i, \bar{A}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{B}_i, \bar{b_i}))$ to pick $b \in dom(\mathfrak{N}_i)$, so that he still has a winning strategy in $EF^n_{MSO}((\mathfrak{M}_i, \bar{A}_i, \bar{a_i}, a), (\mathfrak{N}_i, \bar{B}_i, \bar{b_i}, b))$. By induction hypothesis he also has one in the remaining n-length game $EF^n_{MSO}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A_1}, \ldots, \bar{A_k})^{\cup}, \bar{a_1}, \ldots, \bar{a_k}, a), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B_1}, \ldots, \bar{B_k})^{\cup}, \bar{b_1}, \ldots, \bar{b_k}, b))$.

- Spoiler's first move is a set move. Suppose Spoiler chooses a set A in the set of admissible subsets of $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$. Then A is necessarily of the form $A_1 \cup \ldots \cup A_k$, with A_i an admissible subset of $\mathfrak{M}_i$. We now define locally his response $B = B_1 \cup \ldots \cup B_k$, using his winning strategies in the small structures, so that he still has a winning strategy in $EF^n_{MSO}((\mathfrak{M}_i, \bar{A}_i, A_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, B_i, \bar{b}_i))$ for all $1 \leq i \leq k$. By induction hypothesis, he also has one in the remaining n length game $EF^n_{MSO}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, A_1, \ldots, \bar{A}_k, A_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, B_1, \ldots, \bar{B}_k, B_k)^{\cup}, \bar{b}_1, \ldots, b$ (Note that this is enough, because $A \in (\bar{A}_1, A_1, \ldots, \bar{A}_k, A_k)^{\cup}$.)

□

Now an analogue of this result for disjoint unions can easily be derived as a corollary of Theorem 12. For the convenience of the reader, we provide here the detailed argument:

Corollary 5. Whenever $(\mathfrak{M}_i, \bar{A}_i, \bar{a}_i) \equiv^n_{MSO} (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i)$ for all $1 \leq i \leq k$ (with $\bar{a}_i$ a sequence of first-order parameters of the form $a_{i_1}, \ldots, a_{i_m}$ with $m \in \mathbb{N}$ and $\bar{A}_i$ a sequence of set parameters of the form $A_{i_1}, \ldots, A_{i_{m'}}$ with $m' \in \mathbb{N}$, similarly for the $\bar{b}_i$ and $\bar{B}_i$), then also $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k \equiv^n_{MSO} \biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k$.

Proof. Let $(\mathfrak{M}_i, \bar{A}_i, \bar{a}_i) \equiv^n_{MSO} (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i)$ for all $1 \leq i \leq k$ (with $\bar{a}_i$ a sequence of first-order parameters of the form $a_{i_1}, \ldots, a_{i_m}$ with $m \in \mathbb{N}$ and $\bar{A}_i$ a sequence of set parameters of the form $A_{i_1}, \ldots, A_{i_{m'}}$ with $m' \in \mathbb{N}$, similarly for the $\bar{b}_i$ and $\bar{B}_i$).

Now consider the following expansions $\mathfrak{M}'_i$ and $\mathfrak{N}'_i$ of the σ structures $\mathfrak{M}_i$ and $\mathfrak{N}_i$ to $\sigma^* = \sigma \cup \{Q_1, \ldots, Q_k\}$: the interpretation of Q_j is empty in $\mathfrak{M}'_i$ (respectively $\mathfrak{N}'_i$) whenever $i \neq j$ and it is the domain of $\mathfrak{M}'_i$ (respectively $\mathfrak{N}'_i$) whenever $i = j$.

Clearly $(\mathfrak{M}'_i, \bar{A}_i, \bar{a}_i) \equiv^n_{MSO} (\mathfrak{N}'_i, \bar{B}_i, \bar{b}_i)$ for all $1 \leq i \leq k$.

Now consider a mapping f such that for every n-ary predicate $P \in \sigma^*$, $f(P) = Px_1 \ldots x_n$. By Theorem 12 we have that $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}'_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k \equiv^n_{MSO} \bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k$.

Corollary 5 follows because $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}'_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k$ and $\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k$ are isomorphic to $\biguplus_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k$ and $\biguplus_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k$ respectively. □

An other important corollary of Theorem 12 is the fact that fusions of MSO-Henkin structures are also MSO-Henkin structures. Let us the stress the importance of this fact, which is needed for the correcteness of our main completeness argument.

Corollary 6. $\mathbb{A}_{\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i}$ is closed under MSO parametric definability and so $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$ is a MSO-Henkin structure.

Proof. First note that the following are equivalent:

- A is MSO parametrically definable in $\mathfrak{M}$,
- there is a finite sequence of parameters $\bar{a}, \bar{A}$ such that A is defined by a MSO formula ϕ of quantifier depth n using $\bar{a}, \bar{A}$,
- for any two points a and a' in $dom(\mathfrak{M})$, if they are MSO n-indistinguishable using $\bar{a}, \bar{A}$, then $a \in A$ iff $a' \in A$.

Now suppose there is $A \subseteq dom(\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i)$ MSO parametrically definable in $\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i$ using $\bar{a'}, \bar{A'}$, but $A \notin \mathbb{A}_{\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i}$. So it means that for some $1 \leq i \leq k$, $A_i = A \cap dom(\mathfrak{M}_i)$ is not MSO parametrically definable in $\mathfrak{M}_i$ i.e. there are two MSO parametrically indistinguishable points $a \in A$, $a' \notin A$. So for all n, for all sequence of parameters $\bar{a}, \bar{A}$ in $\mathfrak{M}_i$, $(\mathfrak{M}_i, \bar{a}, \bar{A}, a) \equiv^n_{\mathsf{FO(TC^1)}} (\mathfrak{M}_i, \bar{a}, \bar{A}, a')$ and by the fusion lemma,[7] $\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i, \bar{a}, \bar{A}, \bar{a'}, \bar{A'}, a \equiv^n_{\mathsf{FO(TC^1)}} \bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i, \bar{a}, \bar{A}, \bar{a'}, \bar{A'}, a'$. But this entails that A is not MSO parametrically definable using $\bar{a'}, \bar{A'}$ in $\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i$. □

Corollary 7. $\mathbb{A}_{\biguplus_{1\leq i\leq k}\mathfrak{M}_i}$ is closed under MSO parametric definability and so $\biguplus_{1\leq i\leq k}\mathfrak{M}_i$ is a MSO-Henkin structure.

Proof. Analoguous to the proof of Corollary 6 (because $\mathbb{A}_{\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i} = \mathbb{A}_{\biguplus_{1\leq i\leq k}\mathfrak{M}_i}$). □

D.2 Fusion Lemma for FO(TC^1)

As TC moves can only be played when there are already two pebbles on the board, it is more convenient to show first a version of our FO(TC^1) fusion lemma in which each small structure comes with at least two parameters. This allows us to define Duplicator's answer to a TC move played in a big structure, by means of his winning strategies in the corresponding small structures. We then derive as a corollary the fusion lemma for non parametrized structures.

Theorem 13. *Whenever $(\mathfrak{M}_i, \bar{a_i}) \equiv^n_{FO(TC^1)} (\mathfrak{N}_i, \bar{b_i})$ for all $1 \leq i \leq k$ (with $\bar{a_i}$ a sequence of distinct parameters of the form $a_{i_1}, \ldots, a_{i_m}$ with $m \in \mathbb{N}$ and $m \geq 2$, similarly for the $\bar{b_i}$), then also $\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k} \equiv^n_{FO(TC^1)} \bigoplus^f_{1\leq i\leq k}\mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}$. As a special case, in the case of single point structures (structures which domain contains only one point), we allow the parameters to be non distinct objects.*

Proof. We define a winning strategy for Duplicator in the game $EF^n_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}), (\bigoplus^f_{1\leq i\leq k}\mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}))$ out of her winning strategies in the games $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ by induction on n.

Base step: $n = 0$, doing nothing is a strategy for Duplicator. We need to show that the $\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}$ and $\bigoplus^f_{1\leq i\leq k}\mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}$ agree on all atomic formulas. Now in the fusion structures, each atomic formula is defined by f in terms of a σ^*-quantifier free formula that is evaluated in the corresponding disjoint union structure. So it is enough to show that the disjoint union structures agree on all atomic σ^*-formulas and on their Boolean combinations. The initial match between the distinguished objects in $(\mathfrak{M}_i, \bar{a_i})$ and $(\mathfrak{N}_i, \bar{b_i})$ is a partial isomorphism for every $1 \leq i \leq k$, so it is also one for $\biguplus_{1\leq i\leq k}\mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}$ and $\biguplus_{1\leq i\leq k}\mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}$ i.e. the two disjoint union structures agree on all σ^*-atomic formulas. Now relying on the semantics of Boolean connectives, it can be shown by induction on the complexity of quantifier free sentences that they also agree on all Boolean combinations of atomic σ^*-sentences.

Inductive step: the inductive hypothesis says that whenever Duplicator has a winning strategy in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ for some $(\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i})$ satisfying the required conditions on parameters and $1 \leq i \leq k$, he also has one in $EF^n_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k}\mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}), (\bigoplus^f_{1\leq i\leq k}\mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}))$.

[7] There is no need to consider the case where $\bar{a'}, \bar{A'}$ is empty, because if a set is parametrically definable using no parameter, it is also definable using parameters.

We want to show that this also holds when the length of the game is $n+1$. Suppose Duplicator has a winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ for all $1 \leq i \leq k$. We describe Duplicator's answer to Spoiler's first move in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}))$. It then follows by induction hypothesis, that he has a winning strategy in the remaining n-length game.

- Spoiler's first move is an $\exists$ move. Suppose Spoiler chooses a point $a \in dom(\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i)$, then $a \in dom(\mathfrak{M}_i)$ for some $1 \leq i \leq k$. So Duplicator can use his winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ and pick a corresponding point b in the other structure. Now he still has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a), (\mathfrak{N}_i, \bar{b_i}, b))$. So by induction hypothesis he also has one in the remaining n length game $EF^{n}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}, a), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}, b))$.

- Spoiler's first move is a TC move. Suppose Spoiler chooses a set A in the set of admissible subsets of $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$. Then A is necessarily of the form $A_1 \cup \ldots \cup A_k$, with A_i an admissible subset (possibly empty) of $\mathfrak{M}_i$. Her response $B = B_1 \cup \ldots \cup B_k$ can now be defined locally for each B_i using her winning strategies in the small structures. So let Spoiler choose $A = A_1 \cup \ldots \cup A_k$. Keeping in mind that each non single point small structure comes with at least two distinct parameters, there are four cases:

  a) in $dom(\mathfrak{M}_i)$, there is a distinguished object inside, but also outside A_i, so Duplicator considers A_i together with these two parameters and constructs B_i by using his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$.
  b) in $dom(\mathfrak{M}_i)$, there are only distinguished objects inside A_i[8], so Duplicator considers any one of these distinguished objects, let say a_j and looks at $A_i \backslash \{a_j\}$ together with some parameter inside A_i, so that he can use his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ to construct an answer that we call B'_i. Now $B_i = B'_i \cup \{b_j\}$;
  c) in $dom(\mathfrak{M}_i)$, there are only distinguished objects outside A_i,[9] so Duplicator similarly considers some distinguished object a_j and looks at $A_i \cup \{a_j\}$ together with some other parameter outside A_i, so that he can use his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ to construct an answer that we call B'_i. Now $B_i = B'_i \backslash \{b_j\}$;
  d) $\mathfrak{M}_i$ is a single point structure, then $B_i = \emptyset$ if $A_i = \emptyset$ and $B_i = dom(\mathfrak{M}_i)$ if $A_i = dom(\mathfrak{N}_i)$.

  Once $B = B_1 \cup \ldots \cup B_k$ has been constructed, Spoiler picks two points $b \in B$ and $b' \notin B$. There are two cases:

  1. b and b' belong to the domain of one and the same small structure $\mathfrak{N}_i$; now $dom(\mathfrak{M}_i)$ is as previously described in $a), b), c)$ (but not d), because two distinct points cannot belong to one and the same single point structure) and in each case Duplicator does the following:

[8] Note that as a special case we may have $A_i = dom(\mathfrak{M}_i)$.
[9] Note that as a special case we may have $A_i = \emptyset$.

a) Duplicator answers with a, a' according to his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$, so that he still has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a, a'), (\mathfrak{N}_i, \bar{b_i}, b, b'))$. By induction hypothesis he also has one in the remaining n length game $EF^{n}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}, a, a'), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}, b, b'))$;

b) suppose first that $b' \neq b_j$, so Duplicator considers $A_i \backslash \{a_j\}$ together with a_j and with some other parameter inside this set and uses his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ to pick corresponding a, a' in $\mathfrak{M}_i$, so that he still has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a, a'), (\mathfrak{N}_i, \bar{b_i}, b, b'))$. By induction hypothesis he also has one in the remaining n length game $EF^{n}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}, a, a'), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}, b, b'))$; otherwise $b = b_j$, then $a = a_j$ because the parameter a_j already matches b i.e. Duplicator has a winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a), (\mathfrak{N}_i, \bar{b_i}, b))$, so Duplicator uses his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a), (\mathfrak{N}_i, \bar{b_i}, b))$ to pick a', answering as if it was a point move (i.e a' has to be n-equivalent to b'), so that he still has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a, a'), (\mathfrak{N}_i, \bar{b_i}, b, b'))$. By induction hypothesis he also has one in the remaining n length game $EF^{n}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}, a, a'), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}, b, b'))$.
Now there is some additional condition $a' \notin A_i$ that Duplicator shall also maintain in order to respect the rules of the game. But there has to be an n-equivalent point to b' which is outside A_i. Indeed, instead of b, Spoiler could have picked any other point $b^* \in B_i$ together with $b' \notin B_i$ and Duplicator's winning strategy would have provided a correct answer $a^* \in A_i$, $a' \notin A_i$, which means that Duplicator would have found some a' point which is at least n-equivalent to b' and outside A_i (because if Duplicator has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a^*, a'), (\mathfrak{N}_i, \bar{b_i}, b^*, b'))$ then he also has one in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a'), (\mathfrak{N}_i, \bar{b_i}, b'))$ and hence in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a, a'), (\mathfrak{N}_i, \bar{b_i}, b, b'))$).

c) suppose first that $b \neq b_j$, so Duplicator considers $A_i \cup \{a_j\}$ together with a_j and with some other parameter outside this set and uses his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$, so that he still has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a, a'), (\mathfrak{N}_i, \bar{b_i}, b, b'))$. By induction hypothesis he also has one in the remaining n length game $EF^{n}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}, a, a'), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}, b, b'))$; otherwise $b' = b_j$, then $a' = a_j$ because the parameter a_j already matches b' i.e. Duplicator has a winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a'), (\mathfrak{N}_i, \bar{b_i}, b'))$, so we can show by a similar argument as the one used in the above item, that he can use his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a'), (\mathfrak{N}_i, \bar{b_i}, b'))$ to pick $a \in A_i$, so that he still has a winning strategy in $EF^{n}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a, a'), (\mathfrak{N}_i, \bar{b_i}, b, b'))$. By induction hypothesis he also has one in the remaining n length game $EF^{n}_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bar{a_1}, \ldots, \bar{a_k}, a, a'), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, \bar{b_1}, \ldots, \bar{b_k}, b, b'))$.

2. otherwise $b \in dom(\mathfrak{N}_i, \bar{b_i})$ and $b' \in dom(\mathfrak{N}_j, \bar{b_j})$ with $i \neq j$; we can again use a similar argument to show that Duplicator can use his winning strategy in $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}), (\mathfrak{N}_i, \bar{b_i}))$ and $EF^{n+1}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_j, \bar{a_j}), (\mathfrak{N}_j, \bar{b_j}))$ to pick

a, a' in the right side of the structure (i.e. inside or outside A_i), so that he still has a winning strategy in $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, \bar{a_i}, a), (\mathfrak{N}_i, \bar{b_i}, b))$ and $EF^n_{\mathsf{FO(TC^1)}}((\mathfrak{M}_j, \bar{a_j}, a'), (\mathfrak{N}_j, \bar{b_j}, b'))$ (in the special case where for instance, $\mathfrak{M}_j$ is a single point structure, Duplicator picks the only available point in the other structure). By induction hypothesis he also has one in the remaining n length game $EF^n_{\mathsf{FO(TC^1)}}((\bigoplus^f_{1\le i\le k} \mathfrak{M}_i, \bar{a_1}, \dots, \bar{a_k}, a, a'), (\bigoplus^f_{1\le i\le k} \mathfrak{N}_i, \bar{b_1}, \dots, \bar{b_k}, b, b'))$.

□

We now show a corollary of the preceding lemma, in which the small structure do not come with any distinguished objects:

Corollary 8. Whenever $\mathfrak{M}_i \equiv^n_{\mathsf{FO(TC^1)}} \mathfrak{N}_i$ for all $1 \le i \le k$, then also $\bigoplus^f_{1\le i\le k} \mathfrak{M}_i \equiv^n_{\mathsf{FO(TC^1)}} \bigoplus^f_{1\le i\le k} \mathfrak{N}_i$.

Proof. We know that Spoiler's first two moves in $EF^{n+1}_{\mathsf{FO(TC^1)}}(\bigoplus^f_{1\le i\le k} \mathfrak{M}_i, \bigoplus^f_{1\le i\le k} \mathfrak{N}_i)$ must be quantifier moves, because the TC move can only be played once there are two pebbles on the board. Let us look at the first move. Suppose Spoiler plays a point $a \in dom(\bigoplus^f_{1\le i\le k} \mathfrak{M}_i)$. So $a \in dom(\mathfrak{M}_i)$ for some $1 \le i \le k$. By Duplicator's winning strategy in $EF^n_{\mathsf{FO(TC^1)}}(\mathfrak{M}_i, \mathfrak{N}_i)$, he has an answer $b \in dom(\mathfrak{N}_i)$ such that $(\mathfrak{M}_i, a) \equiv^n_{\mathsf{FO(TC^1)}} (\mathfrak{N}_i, b)$. Let us rename a with a_{i_1} and b with b_{i_1}. Similarly, for every $j \neq i$ such that $1 \le j \le k$, fix some random point a_{j_1} coming from the domain of $\mathfrak{M}_j$, Spoiler could have played this point and so Duplicator would have had an adequate answer b_{j_1} such that $(\mathfrak{M}_j, a_{j_1}) \equiv^n_{\mathsf{FO(TC^1)}} (\mathfrak{N}_j, b_{j_1})$. Now for the second round in the game, some point $a' = a_{l_2}$ or $b' = b_{l_2}$ coming from the domain of respectively $\mathfrak{M}_l$ or $\mathfrak{N}_l$ will be played by Spoiler and Duplicator will be able to answer so that $(\mathfrak{M}_l, a_{l_1}, a_{l_2}) \equiv^{n-2}_{\mathsf{FO(TC^1)}} (\mathfrak{N}_l, b_{l_1}, b_{l_2})$. Similarly, for each $\mathfrak{M}_j$ such that $j \neq l$, we can find points such that $(\mathfrak{M}_j, a_{j_1}, a_{j_2}) \equiv^{n-2}_{\mathsf{FO(TC^1)}} (\mathfrak{N}_i, b_{j_1}, b_{j_2})$. Now as for all $1 \le i \le k$, Duplicator has a winning strategy in $EF^{n-2}_{\mathsf{FO(TC^1)}}((\mathfrak{M}_i, a_{i_1}, a_{i_2}), (\mathfrak{N}_i, b_{i_1}, b_{i_2}))$, by the previous lemma, he has one in $EF^{n-2}_{\mathsf{FO(TC^1)}}(\bigoplus^f_{1\le i\le k} \mathfrak{M}_i, a_{1_1}, a_{1_2}, \dots, a_{k_1}, a_{k_2}), (\bigoplus^f_{1\le i\le k} \mathfrak{N}_i, b_{1_1}, b_{1_2}, \dots, b_{k_1}, b_{k_2}))$, so he also has one in $EF^{n-2}_{\mathsf{FO(TC^1)}}(\bigoplus^f_{1\le i\le k} \mathfrak{M}_i, a, a'), (\bigoplus^f_{1\le i\le k} \mathfrak{N}_i, b, b'))$. □

Corollary 9. Whenever $\mathfrak{M}_i \equiv^n_{\mathsf{FO(TC^1)}} \mathfrak{N}_i$ for all $1 \le i \le k$, then also $\biguplus_{1\le i\le k} \mathfrak{M}_i \equiv^n_{\mathsf{FO(TC^1)}} \biguplus_{1\le i\le k} \mathfrak{N}_i$.

Proof. Analoguous to the proof of Corollary 5. □

Corollary 10. $\mathbb{A}_{\bigoplus^f_{1\le i\le k} \mathfrak{M}_i}$ is closed under $\mathsf{FO(TC^1)}$ parametric definability and so $\bigoplus^f_{1\le i\le k} \mathfrak{M}_i$ is a $\mathsf{FO(TC^1)}$-Henkin structure.

Proof. Analoguous to the proof of Corollary 6. □

Corollary 11. $\mathbb{A}_{\biguplus_{1\le i\le k} \mathfrak{M}_i}$ is closed under $\mathsf{FO(TC^1)}$ parametric definability and so $\biguplus_{1\le i\le k} \mathfrak{M}_i$ is a $\mathsf{FO(TC^1)}$-Henkin structure.

Proof. Analoguous to the proof of Corollary 7. □

D.3 Fusion Lemma for FO(LFP^1)

The situation parallels the FO(TC^1) case. As LFP moves can only be played when there is already one pebble on the board, it is more convenient to show first a version of our FO(LFP^1) fusion lemma in which each small structure comes with at least one FO parameter. This allows us to define Duplicator's answer to a LFP move played in the big structure, by means of his winning strategies in the small structures. We then derive as a corollary the fusion lemma for non parametrized structures.

Theorem 14. *Whenever* $(\mathfrak{M}_i, \bar{A}_i, \bar{a}_i) \equiv^n_{\mathit{FO(LFP^1)}} (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i)$ *for all* $1 \leq i \leq k$ *(with* $\bar{a}_i$ *a non empty sequence of parameters of the form* $a_{i_1}, \ldots, a_{i_m}$ *with* $m \geq 0$*, similarly for the* $\bar{b}_i$*), then also* $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^\cup, \bar{a}_1, \ldots, \bar{a}_k \equiv^n_{\mathit{FO(LFP^1)}} \bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^\cup, \bar{b}_1, \ldots, \bar{b}_k$*.*

Proof. We define a winning strategy for Duplicator in the game $EF^n_{\mathsf{FO(LFP^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^\cup, \bar{a}_1, \ldots, \bar{a}_k), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^\cup, \bar{b}_1, \ldots, \bar{b}_k))$ out of her winning strategies in the games $EF^n_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$ by induction on n.

Base step: $n = 0$, doing nothing is a strategy for Duplicator (this can be justified by a similar argument as in the MSO case).

Inductive step: the inductive hypothesis says that whenever Duplicator has a winning strategy in $EF^n_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$ for some $(\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i)$ satisfying the required conditions on parameters and $1 \leq i \leq k$, he also has one in $EF^n_{\mathsf{FO(LFP^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^\cup, \bar{a}_1, \ldots, \bar{a}_k), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^\cup, \bar{b}_1, \ldots, \bar{b}_k))$.

We want to show that this also holds when the length of the games is $n + 1$. Suppose Duplicator has a winning strategy in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$ for all $1 \leq i \leq k$. We describe Duplicator's answer to Spoiler's first move in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, \ldots, \bar{A}_k)^\cup, \bar{a}_1, \ldots, \bar{a}_k), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, \ldots, \bar{B}_k)^\cup, \bar{b}_1, \ldots, \bar{b}_k))$. It then follows by induction hypothesis, that he has a winning strategy in the remaining n-length game.

- Spoiler's first move is an $\exists$ move.

  Same argument as for MSO and FO(TC^1).

- Spoiler's first move is a $\forall$ move.

  Symmetric.

- Spoiler's first move is a GFP move.

  Suppose Spoiler chooses a set A in the set of admissible subsets of $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$ with some pebble $a_{i_j} \in A$. Then A is necessarily of the form $A_1 \cup \ldots \cup A_k$, with A_i an admissible subset of $\mathfrak{M}_i$. Her response $B = B_1 \cup \ldots \cup B_k$ can now be defined locally for each B_i using her winning strategies in the small structures. So let Spoiler choose $A = A_1 \cup \ldots \cup A_k$. Keeping in mind that each small structure comes with at least one parameter, there are four cases:

  1) in $dom(\mathfrak{M}_i)$, there is a distinguished object inside A_i and $A_i \neq dom(\mathfrak{M}_i)$, so Duplicator considers A_i together with this parameter and constructs B_i by using his winning strategy in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$.

2) in $dom(\mathfrak{M}_i)$, there are only distinguished objects outside A_i and $A_i \neq \emptyset$, so Duplicator considers any one of these distinguished objects, let say a_j and looks at $A_i \cup \{a_j\}$, so that he can use his winning strategy in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$ to construct an answer that we call B'_i. Now $B_i = B'_i \backslash \{b_j\}$. This is a correct answer, because the (posimorphism) condition to be maintained is that for every pebble a_l on the board at the end of the game, $a_l \in A_i \Rightarrow b_l \in B_i$. But by Duplicator's winning strategy in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, A_i \cup \{a_j\}, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, B'_i, \bar{b}_i))$, we know already that for every such pebble, $a_l \in A_i \cup \{a_j\} \Rightarrow b_l \in B'_i$, so also $a_l \in A_i \Rightarrow b_l \in B'_i \backslash \{b_j\}$.

3) $B_i = dom(\mathfrak{M}_i)$. So $A_i = dom(\mathfrak{N}_i)$. As pebbles are only chosen using Duplicator's winning strategies in the small structures, the posimorphism condition will be maintained.

4) $B_i = \emptyset$. So $A_i = \emptyset$. As no pebble can belong to this set, the posimorphism condition will be maintained.

Now that $B = B_1 \cup \ldots \cup B_k$ has been constructed, Spoiler picks a new element $b \in B$ which belongs to the domain of one particular small structure $\mathfrak{N}_i$ (so $b \in B_i$) and $dom(\mathfrak{M}_i)$ is as previously described either in 1), 2) or 3) (but not 4), because b cannot belong to the empty set) and in each case Duplicator does the following:

1) Duplicator answers with a according to his winning strategy in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, \bar{b}_i))$;

2) Duplicator again considers $A_i \cup \{a_j\}$ and answers according to his winning strategy in $EF^{n+1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, A_i \cup \{a_j\}, \bar{a}_i), (\mathfrak{N}_i, \bar{B}_i, B'_i, \bar{b}_i))$. This is safe, because the pebble to be chosen has to be fresh, so it won't be a_j;

3) Duplicator picks some random pebble a_j in $dom(\mathfrak{M}_i)$ and considers $dom(\mathfrak{M}_i) \backslash \{a_j\}$. His winning strategy provides him with a correct answer.

So in any case (either 1), 2) or 3)), Duplicator has a winning strategy in $EF^{n}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, \bar{A}_i, A_i, \bar{a}_i, a), (\mathfrak{N}_i, \bar{B}_i, B_i, \bar{b}_i, b))$. Now for all $j \neq i$, $1 \leq j \leq k$, he also has one in $EF^{n}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_j, \bar{A}_j, A_j, \bar{a}_j), (\mathfrak{N}_j, \bar{B}_j, B_j, \bar{b}_j))$. So by induction hypothesis, he has one in the remaining n length game $EF^{n}_{\mathsf{FO(LFP^1)}}((\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, (\bar{A}_1, A_1, \ldots, \bar{A}_k, A_k)^{\cup}, \bar{a}_1, \ldots, \bar{a}_k, a),$
$(\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, (\bar{B}_1, B_1, \ldots, \bar{B}_k, B_k)^{\cup}, \bar{b}_1, \ldots, \bar{b}_k, b))$.

- Spoiler's first move is a LFP move.

  Symmetric.

□

Corollary 12. Whenever $\mathfrak{M}_i \equiv^n_{\mathsf{FO(LFP^1)}} \mathfrak{N}_i$ for all $1 \leq i \leq k$, then also $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i \equiv^n_{\mathsf{FO(LFP^1)}} \bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i$.

Proof. We know that Spoiler's first move in $EF^{n+1}_{\mathsf{FO(LFP^1)}}(\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, \bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i)$ must be a FO quantifier move, because the LFP move can only be played once there is a pebble on the board. Let us look at the first move. Suppose Spoiler plays a point $a \in dom(\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i)$. So $a \in dom(\mathfrak{M}_i)$ for some $1 \leq i \leq k$. By Duplicator's winning strategy in $EF^{n}_{\mathsf{FO(LFP^1)}}(\mathfrak{M}_i, \mathfrak{N}_i)$, he has an answer $b \in dom(\mathfrak{N}_i)$ such that

$(\mathfrak{M}_i, a) \equiv^n_{\mathsf{FO(LFP^1)}} (\mathfrak{N}_i, b)$. Let us rename a with a_i and b with b_i. Similarly, for every $j \neq i$ such that $1 \leq j \leq k$, fix some random point a_j coming from the domain of $\mathfrak{M}_j$, Spoiler could have played this point and so Duplicator would have had an adequate answer b_j such that $(\mathfrak{M}_j, a_j) \equiv^n_{\mathsf{FO(LFP^1)}} (\mathfrak{N}_j, b_j)$. Now as for all $1 \leq i \leq k$, Duplicator has a winning strategy in $EF^{n-1}_{\mathsf{FO(LFP^1)}}((\mathfrak{M}_i, a_i), (\mathfrak{N}_i, b_i))$, by the previous lemma, he has one in $EF^{n-1}_{\mathsf{FO(LFP^1)}}(\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, a_1, \ldots, a_k), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, b_1, \ldots, b_k))$, so he also has one in $EF^{n-1}_{\mathsf{FO(LFP^1)}}(\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i, a), (\bigoplus^f_{1\leq i\leq k} \mathfrak{N}_i, b))$. □

Corollary 13. Whenever $\mathfrak{M}_i \equiv^n_{\mathsf{FO(LFP^1)}} \mathfrak{N}_i$ for all $1 \leq i \leq k$, then also $\biguplus_{1\leq i\leq k} \mathfrak{M}_i \equiv^n_{\mathsf{FO(LFP^1)}} \biguplus_{1\leq i\leq k} \mathfrak{N}_i$.

Proof. Analoguous to the proof of Corollary 5. □

Corollary 14. $\mathbb{A}_{\bigoplus^f_{1\leq i\leq k}}$ is closed under $\mathsf{FO(LFP^1)}$ parametric definability and so $\bigoplus^f_{1\leq i\leq k} \mathfrak{M}_i$ is a $\mathsf{FO(LFP^1)}$-Henkin structure.

Proof. Analoguous to the proof of Corollary 6. □

Corollary 15. $\mathbb{A}_{\biguplus_{1\leq i\leq k} \mathfrak{M}_i}$ is closed under $\mathsf{FO(LFP^1)}$ parametric definability and so $\biguplus_{1\leq i\leq k} \mathfrak{M}_i$ is a $\mathsf{FO(LFP^1)}$-Henkin structure.

Proof. Analoguous to the proof of Corollary 7. □