Inclusion Logic and Fixed Point Logic

Pietro Galliani[1] **and Lauri Hella**[2]

1 **University of Helsinki**
pgallian@gmail.com
2 **University of Tampere**
lauri.hella@uta.fi

Abstract

We investigate the properties of Inclusion Logic, that is, First Order Logic with Team Semantics extended with inclusion dependencies. We prove that Inclusion Logic is equivalent to Greatest Fixed Point Logic, and we prove that all union-closed first-order definable properties of relations are definable in it. We also provide an Ehrenfeucht-Fraïssé game for Inclusion Logic, and give an example illustrating its use.

1998 ACM Subject Classification F.4.1 Mathematical Logic

Keywords and phrases Dependence Logic, Team Semantics, Fixpoint Logic, Inclusion

Digital Object Identifier 10.4230/LIPIcs.CSL.2013.281

1 Introduction

Inclusion Logic [10], FO($\subseteq$), is a novel logical formalism designed for expressing inclusion dependencies between variables. It is closely related to Dependence Logic [24], FO(D), which is the extension of First Order Logic by functional dependencies between variables. Dependence Logic initially arose as a variant of *Branching Quantifier Logic* [13] and of *Independence-Friendly Logic* [14, 22], and its study has sparked the development of a whole family of logics obtained by adding various dependency conditions to First Order Logic.

All these logics are based on Team Semantics [16, 24] which is a generalization of Tarski Semantics. In Team Semantics, formulas are satisfied or not satisfied by *sets* of assignments, called *teams*, rather than by single assignments. This semantics was introduced in [16] for the purpose of defining a compositional equivalent for the Game Theoretic Semantics of Independence-Friendly Logic [14, 22], but it was soon found out to be of independent interest. See [9] for a, mostly up-to-date, account of the research on Team Semantics.

Like Branching Quantifier Logic and Independence-Friendly Logic, Dependence Logic has the same expressive power as Existential Second Order Logic Σ^1_1: every FO(D)-sentence is equivalent to some Σ^1_1-sentence, and vice versa [24]. The semantics of Dependence Logic is downwards closed in the sense that if a team X satisfies a formula ϕ in a model M, then all subteams $Y \subseteq X$ also satisfy ϕ in M. The equivalence between FO(D) and Σ^1_1 was extended to formulas in [19], where it was proved that FO(D) captures exactly the downwards closed Σ^1_1-definable properties of teams.

Other variants of Dependence Logic that have been studied are Conditional Independence Logic FO($\perp_c$) [12], Independence Logic FO($\perp$) [12, 25], Exclusion Logic FO($\mid$) [10] and Inclusion/Exclusion Logic FO($\subseteq, \mid$) [10]. All the logics in this family arise from dependency notions that have been studied in Database Theory. In particular, FO(D) is based on *functional dependencies* introduced by Armstrong [1], FO($\subseteq$) is based on *inclusion dependencies* [8, 3], FO($\mid$) is based on *exclusion dependencies* [4], and FO($\perp$) is based on *independence conditions* [11].

Computer Science Logic 2013 (CSL'13).
Editor: Simona Ronchi Della Rocca; pp. 281–295
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The expressive power of all these logics, with the exception of FO($\subseteq$), is well understood. It is known that, with respect to sentences, they are all equivalent with Σ_1^1. With respect to formulas, FO($\mid$) is equivalent with FO(D) [10]; and FO($\subseteq, \mid$), FO($\perp_c$) and FO($\perp$) are all equivalent to each other [10, 25]. Moreover, FO($\perp_c$) (and hence also FO($\subseteq, \mid$) and FO($\perp$)) captures all Σ_1^1-definable properties of teams [10].

On the other hand, relatively little is known about the expressive power of Inclusion Logic, and the main purpose of the present work is precisely to remedy this. What little is known about this formalism can be found in [10], and amounts to the following: With respect to formulas, FO($\subseteq$) is strictly weaker than $\Sigma_1^1 \equiv$ FO($\perp_c$) and incomparable with FO(D) $\equiv$ FO($\mid$). This is simply because the semantics of FO($\subseteq$) is not downwards closed, but is closed under unions: if both teams X and Y satisfy a formula ϕ in a model M, then $X \cup Y$ also satisfies ϕ in M. Moreover, FO($\subseteq$) is stronger than First Order Logic over sentences, and it is contained in Σ_1^1; but it was unknown whether it is equivalent to Σ_1^1, or whether FO($\subseteq$)-formulas could define all union closed Σ_1^1-definable properties of teams.

In this paper we show that the answer to both of these problems is negative. In fact, we give a complete characterization for the expressive power of FO($\subseteq$) in terms of Positive Greatest Fixed Point Logic GFP$^+$: We prove that every FO($\subseteq$)-sentence is equivalent to some GFP$^+$-sentence, and vice versa (Corollary 17).

Fixed point logics have a central role in the area of Descriptive Complexity Theory. By the famous result of Immerman [17] and Vardi [26], Least Fixed Point Logic LFP captures PTIME on the class of ordered finite models. Furthermore, it is well known that on finite models, LFP is equivalent to GFP$^+$. Thus, we obtain a novel characterization for PTIME: a class of ordered finite models is in PTIME if and only if it is definable by a sentence of FO($\subseteq$).

In addition to the equivalence with GFP$^+$, we prove that all union-closed first-order definable properties of teams are definable in Inclusion Logic (Corollary 26). Thus, it is not possible to increase the expressive power of FO($\subseteq$) by adding first-order definable union-closed dependencies. On the other hand, it is an interesting open problem, whether FO($\subseteq$) can be extended by some natural set $\mathbf{D}$ of union-closed dependencies such that the extension FO($\subseteq, \mathbf{D}$) captures all union-closed Σ_1^1-definable properties of teams.

We also introduce a new Ehrenfeucht-Fraïssé game that characterizes the expressive power of Inclusion Logic (Theorem 29). Our game is a modification of the EF game for Dependence Logic defined in [24]. Although the EF game has a clear second order flavour, it is still more manageable than the usual EF game for Σ_1^1; we illustrate this by describing a concrete winning strategy for Duplicator in the case of models with empty signature (Proposition 30). Due to the equivalence between FO($\subseteq$) and GFP$^+$ we see that the EF game for Inclusion Logic is also a novel EF game for GFP$^+$; it is quite different in structure from the one introduced in [2]. It may be hoped that this new game and its variants could be of some use for studying the expressive power of fixed point logics.

2 Preliminaries

2.1 Team Semantics

In this section, we will recall the definition of the Team Semantics for First Order Logic. For simplicity reasons, we will assume that all our expressions are in negation normal form.

▶ **Definition 1.** Let M be a first order model and let V be a set of variables. A *team* X over M with *domain* $\mathtt{Dom}(X) = V$ is a set of assignments $s : V \to \mathtt{Dom}(M)$. Given a tuple

$\vec{t} = (t_1, \ldots, t_n)$ of terms with variables in V and an assignment $s \in X$, we write $\vec{t}\langle s\rangle$ for the tuple $(t_1\langle s\rangle, \ldots, t_n\langle s\rangle)$, where $t\langle s\rangle$ denotes the value of the term t with respect to s in the model M. Furthermore, we write $X(\vec{t})$ for the relation $\{\vec{t}\langle s\rangle : s \in X\}$.

A (non-deterministic) *choice function* for a team X over a set A is a function $H : X \to \mathcal{P}(A) \setminus \{\emptyset\}$. The set of all choice functions for X over A is denoted by $\mathcal{C}(X, A)$.

▶ **Definition 2** (Team Semantics for First Order Logic[1])**.** Let M be a first order model and let X be a team over it. Then, for all first-order literals α, variables v, and formulas ϕ and ψ over the signature of M and with free variables in $\texttt{Dom}(X)$,
TS-lit: $M \models_X \alpha$ iff for all $s \in X$, $M \models_s \alpha$ in the usual Tarski Semantics sense;
TS-$\vee$**:** $M \models_X \phi \vee \psi$ iff $X = Y \cup Z$ for some Y and Z such that $M \models_Y \phi$ and $M \models_Z \psi$;
TS-$\wedge$**:** $M \models_X \phi \wedge \psi$ iff $M \models_X \phi$ and $M \models_X \psi$;
TS-$\exists$**:** $M \models_X \exists v\phi$ iff there exists a function $H \in \mathcal{C}(X, \texttt{Dom}(M))$ such that $M \models_{X[H/v]} \psi$, where $X[H/v] = \{s[m/v] : s \in X, m \in H(s)\}$;
TS-$\forall$**:** $M \models_X \forall v\phi$ iff $M \models_{X[M/v]} \phi$, where $X[M/v] = \{s[m/v] : s \in X, m \in \texttt{Dom}(M)\}$.

The next theorem can be proved by structural induction on ϕ:

▶ **Theorem 3** (Team Semantics and Tarski Semantics)**.** *For all first order formulas $\phi(\vec{v})$, all models M and all teams X, $M \models_X \phi$ if and only if for all $s \in X$, $M \models_s \phi$ with respect to Tarski Semantics.*

Thus, in the case of First Order Logic it is possible to reduce Team Semantics to Tarski Semantics. What is then the point of working with the technically more complicated Team Semantics? As we will see in the next subsection, the answer is that Team Semantics allows us to extend First Order Logic in novel and interesting ways.

Note that on every model M, there are two teams with empty domain: the empty team $\emptyset$, and the team $\{\emptyset\}$ containing the empty assignment $\emptyset$. All the logics that we consider in this paper have the *empty team property*: $M \models_\emptyset \phi$ for every formula ϕ and model M. Thus, we say that a *sentence ϕ is true* in a model M if $M \models_{\{\emptyset\}} \phi$. If this is the case, we drop the subscript $\{\emptyset\}$, and write just $M \models \phi$.

2.2 Dependencies in Team Semantics

As we saw, in Team Semantics formulas are satisfied or not satisfied by sets of assignments, called *teams*; and a team corresponds in a natural way to a relation over the domain of the model. Therefore, any property of relations can be made to correspond to some property of teams, which we can then add to our language as a new atomic formula. In particular, we can do so for database-theoretic dependency notions, thus obtaining the following *generalized atoms*:[2]

▶ **Definition 4** (Dependence Atoms)**.** Let $\vec{t}_1$, $\vec{t}_2$, $\vec{t}_3$ be tuples of terms over some vocabulary. Then, for all models M and all teams X over M whose domain contains the variables of $\vec{t}_1\vec{t}_2\vec{t}_3$,
TS-fdep: $M \models_X =\!(\vec{t}_1, \vec{t}_2)$ if and only if, for all $s, s' \in X$, $\vec{t}_1\langle s\rangle = \vec{t}_1\langle s'\rangle \Rightarrow \vec{t}_2\langle s\rangle = \vec{t}_2\langle s'\rangle$;
TS-exc: For $|\vec{t}_1| = |\vec{t}_2|$, $M \models_X \vec{t}_1 \mid \vec{t}_2$ if and only if $X(\vec{t}_1) \cap X(\vec{t}_2) = \emptyset$;

[1] What we present here is the so-called *lax* version of Team Semantics. There also exists a *strict* version, with somewhat different rules for disjunction and existential quantification. As discussed in [10], the lax semantics has more convenient properties for the case of Inclusion Logic.
[2] The notion of "generalized atom" is defined formally in [20].

TS-inc: For $|\vec{t}_1| = |\vec{t}_2|$, $M \models_X \vec{t}_1 \subseteq \vec{t}_2$ if and only if $X(\vec{t}_1) \subseteq X(\vec{t}_2)$;
TS-ind: $M \models_X \vec{t}_1 \bot \vec{t}_2$ if and only if for all $s, s' \in X$ there exists an $s'' \in X$ with $\vec{t}_1\langle s''\rangle = \vec{t}_1\langle s\rangle$ and $\vec{t}_2\langle s''\rangle = \vec{t}_2\langle s'\rangle$;
TS-cond-ind: $M \models_X \vec{t}_2 \bot_{\vec{t}_1} \vec{t}_3$ if and only if for all $s, s' \in X$ with $\vec{t}_1\langle s\rangle = \vec{t}_1\langle s'\rangle$ there exists an $s'' \in X$ with $(\vec{t}_1\vec{t}_2)\langle s''\rangle = (\vec{t}_1\vec{t}_2)\langle s\rangle$ and $(\vec{t}_1\vec{t}_3)\langle s''\rangle = (\vec{t}_1\vec{t}_3)\langle s'\rangle$.

These atoms correspond respectively to *functional dependencies* [1], to *exclusion dependencies* [4], to *inclusion dependencies* [8, 3], to *independence conditions* [11], and to *conditional independence conditions*[3]; and by adding them to the language of First Order Logic we can obtain various logics, whose principal known properties we will now briefly recall.

Dependence Logic FO(D) is obtained by adding functional dependence atoms to the language of First Order Logic. It is the oldest and the most studied among the logics that we will discuss in this work, having been introduced in the seminal book [24] as an alternative approach to the study of *Branching* [13] and *Independence-Friendly* [14, 22] Quantification. It is *downwards closed*, in the sense that, for all models M, Dependence Logic formulas ϕ and teams X, if $M \models_X \phi$ then $M \models_Y \phi$ for all subsets Y of X.

On the level of sentences, Dependence Logic has the same expressive power as Existential Second Order Logic Σ^1_1.

▶ **Theorem 5** ([27, 6, 24])**.** *Every* FO(D)*-sentence is equivalent to some* Σ^1_1*-sentence, and vice versa. In particular,* FO(D) *captures* NP *on finite models.*

The equivalence between FO(D) and Σ^1_1 was extended to formulas by Kontinen and Väänänen, who proved the following characterization:

▶ **Theorem 6** ([19])**.** *Let ϕ be a* FO(D)*-formula with free variables in $\vec{v}$. Then there exists a Σ^1_1-sentence $\Phi(R)$, where R is a $|\vec{v}|$-ary relation symbol which occurs only negatively in Φ, such that*

$$M \models_X \phi \iff (M, X(\vec{v})) \models \Phi(R) \textit{ for all models } M \textit{ and teams } X \neq \emptyset.$$

Conversely, for any such $\Phi(R)$ there exists an FO(D)*-formula ϕ such that the above holds.*

Thus, FO(D) is the strongest logic that can be obtained by adding Σ^1_1-definable downwards-closed dependence conditions to First-Order Logic. Indeed, any such condition will be expressible as $\exists S(X(\vec{v}) \subseteq S \wedge \Phi(S))$ for some Φ in Σ^1_1, and therefore it will be equivalent to some FO(D)-formula.

Exclusion Logic FO(|), on the other hand, is the logic obtained by adding exclusion atoms to First-Order Logic. It was introduced in [10], where it was shown to be equivalent to Dependence Logic with respect to formulas.

Conditional Independence Logic FO($\bot_c$), which was introduced in [12], adds conditional independence atoms $\vec{t}_2 \bot_{\vec{t}_1} \vec{t}_3$ to the language of First Order Logic. Like FO(D), FO($\bot_c$) is equivalent to Σ^1_1 with respect to sentences, and also with respect to formulas:

▶ **Theorem 7** ([12])**.** *Every* FO($\bot_c$)*-sentence is equivalent to some* Σ^1_1*-sentence, and vice versa.*

▶ **Theorem 8** ([10])**.** *A class of relations is definable in Conditional Independence Logic if and only if it contains the empty relation and it is* Σ^1_1*-definable.*

---

[3] As observed in [7], conditional independence atoms also correspond to *embedded multivalued dependencies.*

Therefore, Conditional Independence Logic is the strongest logic that can be obtained by adding Σ^1_1-definable dependencies which are true of the empty relation to First Order Logic. In particular, this implies that every FO(D) formula (and, therefore, every FO(|) formula) is equivalent to some FO($\perp_c$) formula.[4] However, the converse is not true, since FO($\perp_c$) formulas are not, in general, downwards closed.

Furthermore, **Inclusion/Exclusion Logic** FO($\subseteq$, |) – that is, the logic obtained by adding inclusion *and* exclusion dependencies to First Order Logic – was proved in [10] to be equivalent with FO($\perp_c$) with respect to formulas.

Finally, **Independence Logic** FO($\perp$) is the logic obtained by adding only non-conditional dependence atoms $\vec{t}_1 \perp \vec{t}_2$ to First Order Logic. As proved in [25], Independence Logic and Conditional Independence Logic are also equivalent with respect to formulas.

Inclusion Logic FO($\subseteq$) is obtained by adding inclusion atoms to First Order Logic. It is not downwards closed, but it is *closed under unions* in the following sense: if ϕ is an FO($\subseteq$)-formula, M is a model, and X_i, $i \in I$, are teams on M such that $M \models_{X_i} \phi$ for all $i \in I$, then $M \models_X \phi$, where $X = \bigcup_{i \in I} X_i$. (For a proof, see [10]).

Relatively little is known about the expressive power of FO($\subseteq$), and the main purpose of the present work is precisely to remedy this. Here we recall the following results from [10]:

1. On the level of formulas, FO($\subseteq$) is strictly weaker than FO($\perp_c$) $\equiv$ FO($\perp$) $\equiv \Sigma^1_1$, and incomparable with FO(D) $\equiv$ FO(|).
2. The complement of the transitive closure of any first-order formula $\phi(\vec{x}, \vec{y})$ is definable in FO($\subseteq$); hence, FO($\subseteq$) is strictly stronger than First Order Logic on sentences.
3. On the level of sentences, FO($\subseteq$) is contained in Σ^1_1.

We give next a couple of further examples of the expressive power of FO($\subseteq$).

► **Example 9.** (a) Consider the sentence $\phi := \exists x \exists y (y \subseteq x \wedge Exy)$. Let $M = (\texttt{Dom}(M), E^M)$ be a finite model. Then $M \models \phi$ if and only if E^M contains a cycle, i.e., there are $a_0, \ldots, a_{n-1} \in \texttt{Dom}(M)$ such that $(a_i, a_{i+1}) \in E^M$ for all $i < n-1$, and $(a_{n-1}, a_0) \in E^M$.

The idea here is the following: the first existential quantifier gives a set C of values for x, and the formula $\exists y (y \subseteq x \wedge Exy)$ then says that for every $a \in C$ there is a $b \in C$ such that $(a, b) \in E^M$.

(b) Let ψ be the FO($\subseteq$)-sentence $\exists w (\exists u (Pu \wedge u \subseteq w) \wedge \forall u (Ewu \to \exists v (Euv \wedge v \subseteq w)))$. Then $M \models \psi$ if and only if player I has a winning strategy in the following game $G(M)$: Player I starts by choosing some element $a_0 \in P^M$. In each odd round $i+1$, player II chooses an element a_{i+1} such that $(a_i, a_{i+1}) \in E^M$. In each even round $i+1$, player I chooses an element a_{i+1} such that $(a_i, a_{i+1}) \in E^M$. The first player unable to move according to the rules, loses the game. Player I wins all infinite plays of the game.

The class K of all finite models M such that player II has a winning strategy in $G(M)$ is an equivalent to Immerman's *alternating graph accessibility problem*, AGAP. It is well known that AGAP is a complete problem for PTIME with respect to quantifier free reductions ([18]).

2.3 Greatest Fixed Point Logic

Let $\psi(R, \vec{x})$ be a first-order formula such that the arity of R, $\texttt{ar}(R)$, is equal to the length $k = |\vec{x}|$ of the tuple $\vec{x}$. If M is a model, then ψ defines an operation $\Gamma = \Gamma_{M,\psi}$ on the set

[4] This was already shown in [12], in which it was shown that any dependence atom $=\!(\vec{t}_1, \vec{t}_2)$ is equivalent to the conditional independence atom $\vec{t}_2 \perp_{\vec{t}_1} \vec{t}_2$.

$\mathcal{P}(\mathtt{Dom}(M)^k)$ of k-ary relations on $\mathtt{Dom}(M)$ as follows:

$$\Gamma(P) := \{\vec{a} : (M, P) \models_{s[\vec{a}/\vec{x}]} \psi(R, \vec{x})\} \text{ for each } P \in \mathcal{P}(\mathtt{Dom}(M)^k).$$

A relation P is a *fixed point* of the operation $\Gamma_{M,\psi}$ on M if $\Gamma(P) = P$. Furthermore, P is the *greatest fixed point* (*least fixed point*) of $\Gamma_{M,\psi}$ if $Q \subseteq P$ ($P \subseteq Q$, respectively) for all fixed points Q of $\Gamma_{M,\psi}$.

It is well known that if R occurs only positively in ψ, then for every model M, $\Gamma_{M,\psi}$ has a greatest fixed point (and a least fixed point). Moreover, the greatest fixed point P of $\Gamma_{M,\psi}$ has the following characterization: $P = \bigcup\{Q \subseteq \mathtt{Dom}(M)^k : Q \subseteq \Gamma_{M,\psi}(Q)\}$ (see, e.g. [21]).

▶ **Definition 10.** *Greatest Fixpoint Logic*, GFP, is obtained by adding to First Order Logic the *greatest fixed point operator* $[\mathsf{gfp}_{R,\vec{x}}\psi(R,\vec{x})]\vec{t}$, where R is a relation variable with $\mathtt{ar}(R) = |\vec{x}|$, $\psi(R,\vec{x})$ is a formula in which R occurs only positively, and $\vec{t}$ is a tuple of terms with $|\vec{t}| = |\vec{x}|$. The semantics of the operator gfp is defined by the clause:

- $M \models_s [\mathsf{gfp}_{R,\vec{x}}\psi(R,\vec{x})]\vec{t}$ if and only if $\vec{t}\langle s\rangle$ is in the greatest fixed point of $\Gamma_{M,\psi}$.

Positive Greatest Fixed Point Logic, GFP$^+$, is the fragment of Greatest Fixed Point Logic in which fixed point operators occur only positively.

Least Fixpoint Logic, LFP, similarly, introduces an operator $[\mathsf{lfp}_{R,\vec{x}}\psi(R,\vec{x})]\vec{t}$, again for R occurring only positively in ψ, such that $M \models [\mathsf{lfp}_{R,\vec{x}}\psi(R,\vec{x})]\vec{t}$ if and only if $\vec{t}\langle s\rangle$ is in the least fixed point of $\Gamma_{M,\psi}$.

Fixed point logics have been the object of a vast amount of research, especially because of their applications in Finite Model Theory and Descriptive Complexity Theory. In particular, Least Fixed Point Logic captures the complexity class PTIME that consists of all problems that are solvable in polynomial time:

▶ **Theorem 11** ([17, 26]). *A class of linearly ordered finite models is definable in* LFP *if and only if it can be recognized in* PTIME*.*

Another important result is that on finite models, Greatest Fixed Point Logic has the same expressive power as Least Fixed Point Logic.

▶ **Theorem 12** ([17]). *Over finite models,* GFP$^+$ *(as well as* GFP*) is equivalent to* LFP*.*

We will also make use of the following normal form result for Positive Greatest Fixed Point Logic:

▶ **Theorem 13** ([23, 17]). *Every* GFP$^+$*-sentence ϕ is equivalent to a* GFP$^+$*-sentence of the form $\exists\vec{z}\,[\mathsf{gfp}_{R,\vec{x}}\,\psi(R,\vec{x})]\vec{z}$, where ψ is a first-order formula.*

3 Inclusion Logic captures GFP$^+$

We will now prove that Inclusion Logic has exactly the same expressive power as Positive Greatest Fixed Point Logic. Since the semantics of GFP$^+$ is defined in terms of single assignments instead of teams, the equivalence of FO($\subseteq$) and GFP$^+$ on formulas has to be formulated in a bit indirect way; see Theorems 15 and 16 below.

We start with a lemma that connects teams and the greatest fixed point operator:

▶ **Lemma 14.** *Let $\psi(S,\vec{x})$ be a* GFP$^+$*-formula with free variables in $\vec{x} = (x_1,\dots,x_n)$ such that S is n-ary and occurs only positively in ψ, let M be a model, and let Y a team on M.*

(a) *If $(M, Y(\vec{x})) \models_s \psi(S,\vec{x})$ for all $s \in Y$, then $M \models_s [\mathsf{gfp}_{S,\vec{x}}\,\psi(S,\vec{x})]\vec{x}$ for all $s \in Y$.*

(b) *If Y is a maximal team such that $M \models_s [\mathsf{gfp}_{S,\vec{x}}\, \psi(S,\vec{x})]\vec{x}$ for all $s \in Y$, then $(M, Y(\vec{x})) \models_s \psi(S,\vec{x})$ for all $s \in Y$.*

Proof. Note that $(M, Y(\vec{x})) \models_s \psi(S,\vec{x})$ for all $s \in Y$ if and only if $Y(\vec{x}) \subseteq \Gamma_{M,\psi}(Y(\vec{x}))$. Thus, claim (a) follows from the fact that the greatest fixed point of $\Gamma_{M,\psi}$ is the union of all relations Q such that $Q \subseteq \Gamma_{M,\psi}(Q)$. Claim (b) follows from the observation that if Y is a maximal team such that $M \models_s [\mathsf{gfp}_{S,\vec{x}}\, \psi(S,\vec{x})]\vec{x}$ for all $s \in Y$, then $Y(\vec{x})$ is the greatest fixed point of $\Gamma_{M,\psi}$. ◀

We will next prove that every FO($\subseteq$)-formula can be expressed in GFP$^+$.

▶ **Theorem 15.** *For every* FO($\subseteq$)*-formula $\phi(\vec{x})$ with free variables in $\vec{x} = (x_1, \ldots, x_n)$ there is a* GFP$^+$*-formula $\phi^* = \phi^*(R,\vec{x})$ such that $\mathtt{ar}(R) = |\vec{x}|$, R occurs only positively in ϕ^*, and*

$$M \models_X \phi(\vec{x}) \iff (M, X(\vec{x})) \models_s \phi^*(R,\vec{x}) \text{ for all } s \in X$$

holds for all models M and teams X with $\mathtt{Dom}(X) = \{x_1, \ldots, x_n\}$.

Proof. The proof is by structural induction on ϕ.

1. If $\phi(\vec{x})$ is a first-order literal, let $\phi^*(R,\vec{x})$ be just $\phi(\vec{x})$. Then we have

$$\begin{aligned} M \models_X \phi(\vec{x}) &\iff M \models_s \phi(\vec{x}) \text{ for all } s \in X \\ &\iff (M, X(\vec{x})) \models_s \phi(\vec{x}) \text{ for all } s \in X. \end{aligned}$$

2. If $\phi(\vec{x})$ is an inclusion atom $\vec{t_1} \subseteq \vec{t_2}$, let $\phi^*(R,\vec{x})$ be $\exists \vec{z}(R\vec{z} \wedge \vec{t_1}(\vec{x}) = \vec{t_2}(\vec{z}))$, where $\vec{z}$ is a tuple of new variables. Note that $(M, X(\vec{x})) \models_h \vec{t_1}(\vec{x}) = \vec{t_2}(\vec{z})$ for an assignment h defined on $\vec{x}\vec{z}$ if and only if there are two assignments s, s' defined on $\vec{x}$ such that $\vec{t_1}\langle s \rangle = \vec{t_2}\langle s' \rangle$ and $h = s \cup (s' \circ f)$, where f is the function $f(z_i) = x_i$. Thus, we see that $(M, X(\vec{x})) \models_s \phi^*(R,\vec{x})$ for all $s \in X$ if and only if for every $s \in X$ there is an $s' \in X$ such that $\vec{t_1}\langle s \rangle = \vec{t_2}\langle s' \rangle$, as desired.
3. Assume next that $\phi(\vec{x})$ is of the form $\psi(\vec{x}) \vee \theta(\vec{x})$. Then we define

$$\phi^*(R,\vec{x}) := [\mathsf{gfp}_{S,\vec{x}}\, (R\vec{x} \wedge \psi^*(S,\vec{x}))]\vec{x} \vee [\mathsf{gfp}_{T,\vec{x}}\, (R\vec{x} \wedge \theta^*(T,\vec{x}))]\vec{x}.$$

If $M \models_X \phi(\vec{x})$, then there exist teams Y and Z such that $X = Y \cup Z$, $M \models_Y \psi(\vec{x})$ and $M \models_Z \theta(\vec{x})$. By induction hypothesis, $(M, Y(\vec{x})) \models_s \psi^*(S,\vec{x})$, and consequently $(M, X(\vec{x}), Y(\vec{x})) \models_s R\vec{x} \wedge \psi^*(S,\vec{x})$, holds for all $s \in Y$. Hence, by Lemma 14, $(M, X(\vec{x})) \models_s [\mathsf{gfp}_{S,\vec{x}}\, (R\vec{x} \wedge \psi^*(S,\vec{x}))]\vec{x}$ holds for all $s \in Y$.
In the same way we see that $(M, X(\vec{x})) \models_s [\mathsf{gfp}_{T,\vec{x}}\, (R\vec{x} \wedge \theta^*(T,\vec{x}))]\vec{x}$ holds for all $s \in Z$. Thus, we conclude that $(M, X(\vec{x})) \models_s \phi^*(R,\vec{x})$ for all $s \in X$.
To prove the converse, assume that $(M, X(\vec{x})) \models_s \phi^*(R,\vec{x})$ for all $s \in X$. Let Y be the set of all assignments $s \in X$ that satisfy the first disjunct of $\phi^*(R,\vec{x})$, and let Z be the set of assignments $s \in X$ that satisfy the second disjunct. Then Y is the maximal team such that, for all $s \in Y$, $(M, X(\vec{x})) \models_s [\mathsf{gfp}_{S,\vec{x}}\, (R\vec{x} \wedge \psi^*(S,\vec{x}))]\vec{x}$. It follows from Lemma 14 that $(M, X(\vec{x}), Y(\vec{x})) \models_s R\vec{x} \wedge \psi^*(S,\vec{x})$ for all $s \in Y$. Thus, $(M, Y(\vec{x})) \models_s \psi^*(S,\vec{x})$ for all $s \in Y$, and by induction hypothesis, $M \models_Y \psi(\vec{x})$. In the same way we see that $M \models_Z \theta(\vec{x})$. Finally, since $X = Y \cup Z$, we conclude that $M \models_X \phi(\vec{x})$.
4. If $\phi(\vec{x}) = \psi(\vec{x}) \wedge \theta(\vec{x})$, we define simply $\phi^*(R,\vec{x}) := \psi^*(R,\vec{x}) \wedge \theta^*(R,\vec{x})$. The claim follows then directly from the induction hypothesis.
5. If $\phi(\vec{x})$ is of the form $\exists v\, \psi(\vec{x}v)$, let $\phi^*(R,\vec{x})$ be $\exists v [\mathsf{gfp}_{S,\vec{x}v}\, (R\vec{x} \wedge \psi^*(S,\vec{x}v))]\vec{x}v$. Then $M \models_X \phi(\vec{x})$ if and only if there is a function $H \in \mathcal{C}(X, \mathtt{Dom}(M))$ such that $M \models_Y \psi(\vec{x}v)$,

where $Y = X[H/v]$. By the induction hypothesis, this is equivalent to $(M, Y(\vec{x}v)) \models_h \psi^*(S, \vec{x}v)$ being true for all $h \in Y$. This, in turn, is equivalent with the condition

$$(M, X(\vec{x}), Y(\vec{x}v)) \models_h R\vec{x} \wedge \psi^*(S, \vec{x}v) \text{ for all } h \in Y. \quad (1)$$

If condition (1) holds, then by Lemma 14, $(M, X(\vec{x})) \models_h [\mathsf{gfp}_{S,\vec{x}v}(R\vec{x} \wedge \psi^*(S, \vec{x}v))]\vec{x}v$ holds for all $h \in Y$. Since every $s \in X$ has an extension $h \in Y$, it follows that $(M, X(\vec{x})) \models_s \phi^*(R, \vec{x})$ for all $s \in X$.
On the other hand, if $(M, X(\vec{x})) \models_s \phi^*(R, \vec{x})$ for all $s \in X$, we define $H \in \mathcal{C}(X, \mathtt{Dom}(M))$ to be the function such that

$$H(s) := \{a \in \mathtt{Dom}(M) : (M, X(\vec{x})) \models_{s[a/v]} [\mathsf{gfp}_{S,\vec{x}v}(R\vec{x} \wedge \psi^*(S, \vec{x}v))]\vec{x}v\},$$

and let $Y = X[H/v]$. Then Y is the maximal team such that

$$(M, X(\vec{x})) \models_h [\mathsf{gfp}_{S,\vec{x}v}(R\vec{x} \wedge \psi^*(S, \vec{x}v))]\vec{x}v$$

for all $h \in Y$, whence condition (1) follows from Lemma 14.

6. If $\phi(\vec{x})$ is of the form $\forall v\, \psi(\vec{x}v)$, let $\phi^*(R, \vec{x})$ be $\forall v[\mathsf{gfp}_{S,\vec{x}v}(R\vec{x} \wedge \psi^*(S, \vec{x}v))](\vec{x}v)$. The proof of the claim is similar to the case of existential quantification.

◀

In proving that GFP$^+$-sentences can be expressed in FO($\subseteq$) we will use the normal form given in Theorem 13. Thus, it suffices to find translations for first-order formulas, and formulas obtained by a single application of the gfp-operator to first-order formulas.

▶ **Theorem 16.** *Let $\eta(R, \vec{x}, \vec{y})$ be a first-order formula such that R occurs only positively in η, $\mathtt{ar}(R) = |\vec{x}| = n$, and the free variables of η are in $\vec{x}\vec{y}$.*

(a) *There exists an FO($\subseteq$)-formula $\eta^+(\vec{x}, \vec{y})$ such that for all models M and teams X on M*

$$M \models_X \eta^+(\vec{x}, \vec{y}) \iff (M, X(\vec{x})) \models_s \eta(R, \vec{x}, \vec{y}) \text{ for every } s \in X$$

(b) *If $\vec{y}$ is empty, and $\vec{z}$ is an n-tuple of variables not occurring in η, then there exists an FO($\subseteq$)-formula $\tilde{\eta}(\vec{z})$ such that for all models M and teams X on M*

$$M \models_X \tilde{\eta}(\vec{z}) \iff M \models_s [\mathsf{gfp}_{R,\vec{x}}\, \eta(R, \vec{x})]\vec{z} \text{ for every } s \in X$$

Proof. (a) We prove the claim by structural induction on η.

1. If $\eta(R, \vec{x}, \vec{y})$ is a first-order literal not containing the relation symbol R, we define $\eta^+ := \eta$. Then $M \models_X \eta^+$ if and only if $M \models_s \eta$ for every $s \in X$. Since R does not occur in η, this is equivalent with $(M, X(\vec{x})) \models_s \eta$ for all $s \in X$, as required.
2. If η is of the form $R\vec{t}$, we define $\eta^+(\vec{x}, \vec{y}) := \vec{t} \subseteq \vec{x}$. Then we have

$$\begin{aligned} M \models_X \eta^+(\vec{x}, \vec{y}) &\iff \forall s \in X\, \exists s' \in X : \vec{t}\langle s\rangle = \vec{x}\langle s'\rangle \\ &\iff \forall s \in X : \vec{t}\langle s\rangle \in X(\vec{x}) \\ &\iff \forall s \in X : (M, X(\vec{x})) \models_s R\vec{t}. \end{aligned}$$

3. If η is of the form $\alpha(R, \vec{x}, \vec{y}) \vee \beta(R, \vec{x}, \vec{y})$, let $\vec{u} = (u_1, \ldots, u_n)$ be a tuple of new variables and let $\eta^+(\vec{x}, \vec{y})$ be the formula $\exists \vec{u}\big((\vec{u} \subseteq \vec{x}) \wedge (\alpha^+(\vec{u}, \vec{x}\vec{y}) \vee \beta^+(\vec{u}, \vec{x}\vec{y}))\big)$. Here we assume as induction hypothesis that $M \models_Y \alpha^+(\vec{u}, \vec{x}\vec{y})$ if and only if $(M, Y(\vec{u})) \models_h \alpha(R, \vec{x}, \vec{y})$ for all $h \in Y$, and similarly for $\beta^+(\vec{u}, \vec{x}\vec{y})$ and $\beta(R, \vec{x}, \vec{y})$.
Suppose first that $M \models_X \eta^+(\vec{x}, \vec{y})$. Then there is a function $H \in \mathcal{C}(X, \mathtt{Dom}(M)^n)$ such that $X[H/\vec{u}](\vec{u}) \subseteq X(\vec{x})$, and furthermore, $X[H/\vec{u}]$ can be split into two subteams Y

and Z such that $M \models_Y \alpha^+(\vec{u}, \vec{x}\vec{y})$ and $M \models_Z \beta^+(\vec{u}, \vec{x}\vec{y})$. Now take any $s \in X$ and let $h \in X[H/\vec{u}]$ be an extension of s. If $h \in Y$ then $(M, Y(\vec{u})) \models_h \alpha(R, \vec{x}, \vec{y})$. Since $Y(\vec{u}) \subseteq X[H/\vec{u}](\vec{u}) \subseteq X(\vec{x})$, $\vec{x}\vec{y}\langle h\rangle = \vec{x}\vec{y}\langle s\rangle$ and R occurs only positively in α, we have $(M, X(\vec{x})) \models_s \alpha(R, \vec{x}, \vec{y})$. Similarly, if $h \in Z$ then $(M, X(\vec{x})) \models_s \beta(R, \vec{x}, \vec{y})$. Thus, $(M, X(\vec{x})) \models_s \alpha(R, \vec{x}, \vec{y}) \vee \beta(R, \vec{x}, \vec{y})$ for all $s \in X$, as required.

Conversely, suppose that for any $s \in X$, $(M, X(\vec{x})) \models_s \alpha(R, \vec{x}, \vec{y}) \vee \beta(R, \vec{x}, \vec{y})$. Now let $H \in \mathcal{C}(X, \texttt{Dom}(M)^n)$ be the function such that $H(s) = X(\vec{x})$ for all $s \in X$. Note first that clearly $M \models_{X[H/\vec{u}]} \vec{u} \subseteq \vec{x}$. Let $Y = \{h \in X[H/\vec{u}] : (M, X(\vec{x})) \models_h \alpha(R, \vec{x}, \vec{y})\}$ and $Z = \{h \in X[H/\vec{u}] : (M, X(\vec{x})) \models_h \beta(R, \vec{x}, \vec{y})\}$. By hypothesis, $X[H/\vec{u}] = Y \cup Z$.

If $Y \neq \emptyset$, then $Y(\vec{u}) = X[H/\vec{u}](\vec{u}) = X(\vec{x})$: indeed, if $(M, X(\vec{x})) \models_h \alpha(R, \vec{x}, \vec{y})$ then the same holds for all h' which differ from h only with respect to $\vec{u}$, since $\vec{u}$ is not free in α. Therefore $(M, Y(\vec{u})) \models_h \alpha(R, \vec{x}, \vec{y})$ for all $h \in Y$, and thus $M \models_Y \alpha^+(\vec{u}, \vec{x}\vec{y})$. If instead $Y = \emptyset$, then $M \models_Y \alpha^+(\vec{u}, \vec{x}\vec{y})$ trivially. Similarly, $M \models_Z \beta^+(\vec{u}, \vec{x}\vec{y})$, and therefore $M \models_{X[H/\vec{u}]} \alpha^+(\vec{u}, \vec{x}\vec{y}) \vee \beta^+(\vec{u}, \vec{x}\vec{y})$, whence the function H witnesses that $M \models_X \eta^+$.

4. If η is $\alpha(R, \vec{x}, \vec{y}) \wedge \beta(R, \vec{x}, \vec{y})$, let $\eta^+(\vec{x}, \vec{y})$ be $\alpha^+(\vec{x}, \vec{y}) \wedge \beta^+(\vec{x}, \vec{y})$. Then the claim follows directly from the induction hypothesis.
5. If $\eta(R, \vec{x}, \vec{y})$ is $\exists v\, \alpha(R, \vec{x}, \vec{y}v)$, let $\eta^+(\vec{x}, \vec{y})$ be $\exists v\, \alpha^+(\vec{x}, \vec{y}v)$; here we assume w.l.o.g. that v is not among the variables in $\vec{x}\vec{y}$. Then $M \models_X \eta^+(\vec{x}, \vec{y})$ if and only if there is a function $H \in \mathcal{C}(X, \texttt{Dom}(M))$ such that $M \models_{X[H/v]} \alpha^+(\vec{x}, \vec{y}v)$. Since $X[H/v](\vec{x}) = X(\vec{x})$, by induction hypothesis this is equivalent with the condition

$$(M, X(\vec{x})) \models_h \alpha(R, \vec{x}, \vec{y}v) \text{ holds for all } h \in X[H/v]. \tag{2}$$

If condition (2) is true, then clearly $(M, X(\vec{x})) \models_s \eta(R, \vec{x}, \vec{y})$ for all $s \in X$. Conversely, if $(M, X(\vec{x})) \models_s \eta(R, \vec{x}, \vec{y})$ holds for all $s \in X$, then (2) is true for the function H such that $H(s) = \{a \in \texttt{Dom}(M) : (M, X(\vec{x})) \models_{s[a/v]} \alpha(R, \vec{x}, \vec{y}v)\}$.
6. If $\eta(\vec{R}, \vec{x}, \vec{y})$ is $\forall v\, \alpha(\vec{R}, \vec{x}, \vec{y}v)$, let $\eta^+(\vec{x}, \vec{y})$ be $\forall v\, \alpha^+(\vec{x}, \vec{y}v)$. The proof of the claim is similar as in the previous case.

(b) Let $\vec{z}$ be an n-tuple of variables not occurring in η. We define $\tilde{\eta}(\vec{z})$ to be the formula $\exists \vec{x}(\vec{z} \subseteq \vec{x} \wedge \eta^+(\vec{x}))$, where η^+ is the FO($\subseteq$)-formula corresponding to $\eta(R, \vec{x})$, as given in claim (a). Suppose first that $M \models_X \tilde{\eta}(\vec{z})$. Then there is a function $H \in \mathcal{C}(X, \texttt{Dom}(M)^n)$ such that $M \models_Y \eta^+(\vec{x})$, and $\vec{z}\langle h\rangle \in Y(\vec{x})$ for all $h \in Y$, where $Y = X[H/\vec{x}]$. Thus, by claim (a), $(M, Y(\vec{x})) \models_h \eta(R, \vec{x})$ holds for all $h \in Y$. It follows now from Lemma 14 that $M \models_h [\mathsf{gfp}_{R,\vec{x}}\, \eta(R, \vec{x})]\vec{x}$ for all $h \in Y$. Since every $s \in X$ has an extension $h \in Y$, and $\vec{z}\langle s\rangle = \vec{z}\langle h\rangle \in Y(\vec{x})$, we conclude that $M \models_s [\mathsf{gfp}_{R,\vec{x}}\, \eta(R, \vec{x})]\vec{z}$ for all $s \in X$.

To prove the converse, assume that $M \models_s [\mathsf{gfp}_{R,\vec{x}}\, \eta(R, \vec{x})]\vec{z}$ for all $s \in X$. Let P be the greatest fixed point of the formula $\eta(R, \vec{x})$ (with respect to R and $\vec{x}$) on the model M, and let $H \in \mathcal{C}(X, \texttt{Dom}(M)^n)$ be the function such that $H(s) = P$ for every $s \in X$. Let $Y = X[H/\vec{x}]$. Then $(M, Y(\vec{x})) \models_h \eta(R, \vec{x})$ for all $h \in Y$, whence by claim (a), we have $M \models_Y \eta^+(\vec{x})$. Moreover, $\vec{z}\langle h\rangle \in Y(\vec{x}) = P$ for all $h \in H$, whence $M \models_Y \vec{z} \subseteq \vec{x}$. Thus, the function H witnesses that $M \models_X \exists \vec{x}(\vec{z} \subseteq \vec{x} \wedge \eta^+(\vec{x}))$. ◀

Note that in the case of disjunction above, it was necessary to "store" the possible values of $\vec{x}$ into the values of a new tuple $\vec{u}$ of variables: otherwise, by splitting the team X into two subteams we could have lost information about $X(\vec{x})$.

The equivalence of FO($\subseteq$) and GFP$^+$ follows now from the two theorems above:

▶ **Corollary 17.** *For any* FO($\subseteq$)*-sentence* ϕ *there exists an equivalent* GFP$^+$*-sentence* θ*, and vice versa.*

Proof. If ϕ is an FO($\subseteq$)-sentence, then by Theorem 15, there is a formula $\phi^*(R,x)$ such that for all models M and teams X, $M \models_X \phi$ if and only if $(M, X(x)) \models_s \phi^*(R,x)$ for all $s \in X$. Thus, $M \models \phi$ if and only if $M \models \forall x\, [\mathsf{gfp}_{R,x}\, \phi^*(R,x)]x$.

On the other hand, if ψ is a GFP$^+$-sentence, then by Theorem 13, we can assume that it is of the form $\exists \vec{z}\, [\mathsf{gfp}_{R,\vec{x}}\, \eta(R,\vec{x})]\vec{z}$, where η is a first-order formula. It follows now from Theorem 16(b) that ψ is equivalent to the FO($\subseteq$)-sentence $\exists \vec{z}\, \tilde{\eta}(\vec{z})$. ◀

▶ **Corollary 18.** *A class of linearly ordered finite models is definable in* FO($\subseteq$) *if and only if it can be recognized in* PTIME.

This connection between Inclusion Logic, Fixed Point Logic and descriptive complexity may be of great value for the further development of the area. In particular, it implies that fragments and extensions of FO($\subseteq$) can be made to correspond to various fragments and extensions of PTIME. Hence, results concerning their relationships may lead to insights which may be valuable in complexity theory, and vice versa.

4 First-Order Union Closed Properties

From Corollary 17 it follows immediately that Inclusion Logic is strictly weaker than Σ^1_1. As an immediate consequence, not all Σ^1_1-definable union-closed properties of relations can be expressed in Inclusion Logic. For example, consider the atom

TS-$\mathcal{R}$: $M \models \mathcal{R}(xyzw)$ if and only if there exist two functions $f, g : \mathtt{Dom}(M) \to \mathtt{Dom}(M)$ such that, for all $a, b \in \mathtt{Dom}(M)$, $(a, f(a), b, g(b)) \in X(xyzw)$.

It is easy to see that the atom $\mathcal{R}$ is union-closed. On the other hand, it can be seen that that the sentence $\forall x \exists y \forall z \exists w (\mathcal{R}(xyzw) \wedge (x = z \leftrightarrow y = w) \wedge (y = z \to x = w) \wedge x \neq y)$ holds in a finite model if and only if it contains an even number of elements.[5] Since even cardinality is not definable in GFP, it follows that $\mathcal{R}$ is not definable in FO($\subseteq$).

But what about first order definable union-closed properties? As we will now see, all such properties are indeed definable in Inclusion Logic; and, therefore, it is not possible to increase the expressive power of Inclusion Logic by adding any first order definable union-closed dependency.

▶ **Definition 19.** A sentence $\phi(R)$ is *myopic* if it is of the form $\forall \vec{x}(R\vec{x} \to \theta(R,\vec{x}))$ for some first-order formula θ in which R occurs only positively.

It follows at once from Theorem 16 that myopic sentences correspond to Inclusion Logic-definable properties:

▶ **Proposition 20.** *Let $\phi(R) = \forall \vec{x}(R\vec{x} \to \theta(R,\vec{x}))$ be a myopic sentence. Then there exists an* FO($\subseteq$)*-formula $\phi^+(\vec{x})$ such that, for all models M and teams X,*

$$M \models_X \phi^+(\vec{x}) \text{ if and only if } (M, X(\vec{x})) \models \phi(R).$$

Proof. Consider $\theta(R,\vec{x})$: by Theorem 16(a), there exists an FO($\subseteq$)-formula $\theta^+(\vec{x})$ such that for all models M and teams X,

$$\begin{aligned} M \models_X \theta^+(\vec{x}) &\iff \forall s \in X : (M, X(\vec{x})) \models_s \theta(R,\vec{x}) \\ &\iff (M, X(\vec{x})) \models \forall \vec{x}(R\vec{x} \to \theta(R,\vec{x})), \end{aligned}$$

as required. ◀

[5] The proof of this fact mirrors that of the example in [24], §4.1. In brief, the sentence asserts that the function f mapping x to y is the same as the function g mapping z to w, that this function is an involution, and that this function has no fixed points.

It is also easy to see that all myopic properties are union-closed. We will now prove the converse implication: if $\phi(R)$ is a first order sentence that defines a union-closed property of relations, then it is equivalent to some myopic sentence. From this preservation theorem it will follow at once that all union-closed first-order properties of relations are definable in Inclusion Logic.

First, let us recall some model-theoretic machinery:

▶ **Definition 21** (ω-big models). A model A of signature Σ is ω-big if for all finite tuples $\vec{a}$ of elements of it and for all models $(B, \vec{b}, S)$ such that $(A, \vec{a}) \equiv (B, \vec{b})$ there exists a relation P over A such that $(A, \vec{a}, P) \equiv (B, \vec{b}, S)$.

▶ **Definition 22** (ω-saturated models). A model A is ω-saturated if for every finite set C of elements of A, all complete 1-types over C with respect to A are realized in A.

The proofs of the following model-theoretic results can be found in [15].

▶ **Theorem 23** ([15], Theorem 8.2.1). *Let A be a model. Then A has an ω-big elementary extension.*

▶ **Theorem 24** ([15], Lemma 8.3.4). *Let A and B be ω-saturated structures over a finite signature and such that, for all sentences $\chi(R)$ in which R occurs only positively,*

$$A \models \chi(R) \implies B \models \chi(R).$$

Then there are elementary substructures C and D of A and B and a bijective homomorphism $f : C \to D$ which fixes all relation symbols except R.

▶ **Theorem 25** (Essentially [15], Theorem 8.1.2). *Suppose that A is ω-big and $\vec{a}$ is a finite tuple of elements. Then $(A, \vec{a})$ is ω-saturated.*

Using these results, we can prove our representation theorem:

▶ **Theorem 26.** *Let $\phi(R)$ be a first order sentence that defines a union-closed property of R. Then ϕ is equivalent to some myopic sentence. Consequently, every first-order definable union-closed property of relations is definable in* FO($\subseteq$).

Proof. Let $T = \{\phi'(R) : \phi'(R) \text{ is myopic}, \phi(R) \models \phi'(R)\}$. If we can show that $T \models \phi(R)$, we are done: indeed, by compactness this implies that ϕ is equivalent to a finite conjunction $\forall \vec{x}(R\vec{x} \to \theta_1(R, \vec{x})) \wedge \ldots \wedge \forall \vec{x}(R\vec{x} \to \theta_n(R, \vec{x}))$ of myopic sentences, which of course is equivalent to $\forall \vec{x}(R\vec{x} \to (\theta_1(R, \vec{x}) \wedge \ldots \wedge \theta_n(R, \vec{x})))$.

So, let B' be a model satisfying T, and let B be an ω-big elementary extension of B'. We need to show that $B \models \phi(R)$ (and, therefore, $B' \models \phi(R)$).

Now choose an arbitrary tuple $\vec{b}$ of elements such that $B \models R\vec{b}$, and let Γ be the theory

$$\Gamma = \{R\vec{a}, \phi(R)\} \cup \{\psi(R, \vec{a}) : R \text{ only negative in } \psi, B \models \psi(R, \vec{b})\}.$$

Γ is satisfiable: indeed, if it were not then by compactness there would be formulas $\psi_1(R, \vec{x}), \ldots, \psi_n(R, \vec{x})$ in which R occurs only negatively such that

$$\phi(R) \models \forall \vec{x}\Big(R\vec{x} \to \bigvee_{1 \leq i \leq n} \neg\psi_i(R, \vec{x})\Big).$$

But this is a myopic formula, and therefore it would have to hold in B, which is a contradiction since $B \models \psi_i(R, \vec{b})$ for all $1 \leq i \leq n$.

Now let $(A, \vec{a})$ be an ω-saturated model of Γ. If R occurs only positively in $\chi(R, \vec{x})$ and $A \models \chi(R, \vec{a})$, then $B \models \chi(R, \vec{b})$; otherwise $\neg\chi(R, \vec{a})$ would be in Γ. Furthermore, since B is ω-big, $(B, \vec{b})$ is ω-saturated. Thus, there are elementary substructures $(C, \vec{a})$ and $(D, \vec{b})$ of $(A, \vec{a})$ and $(B, \vec{b})$ and a bijective homomorphism $f : C \to D$ that fixes all relations except R.

Let $S = f(R^C)$. Then $S \subseteq R^D$, since f is an homomorphism; and f is actually an isomorphism between $(C, \vec{a})$ and $(D[S/R], \vec{b})$, since f fixes even R between these two models. Now, $C \models R\vec{a} \wedge \phi(R)$, whence $D \models S\vec{b} \wedge \phi(S)$. Furthermore, since $S \subseteq R$ we have that $D \models \forall\vec{x}(S\vec{x} \to R\vec{x})$.

Now, $(D, \vec{b})$ is an elementary substructure of $(B, \vec{b})$ and B is a ω-big model: therefore, there exists a relation P over B such that $(D, \vec{b}, S) \equiv (B, \vec{b}, P)$. In particular, this implies that $B \models P\vec{b} \wedge \phi(P) \wedge P \subseteq R$: there is a subset of R^B which contains $\vec{b}$ and satisfies ϕ.

But we chose $\vec{b}$ as an arbitrary tuple in R^B. So we have that R^B is the union of a family of relations $P_{\vec{b}}$, where $\vec{b}$ ranges over R^B; and $B \models \phi(P_{\vec{b}})$ for all such $\vec{b}$. Since $\phi(R)$ is closed under unions, this implies that $B \models \phi(R)$, as required. ◀

5 An EF Game for Inclusion Logic

We will now define an Ehrenfeucht-Fraïssé game for Inclusion Logic. This game is an obvious variant of the one defined in [24] for Dependence Logic:

▶ **Definition 27.** Let A and B be two models over the same signature, let $n \in \mathbb{N}$, and let X and Y be two teams with the same domain over A and B, respectively. Then the two-player game $G_n(A, X, B, Y)$ is defined as follows:

1. The initial position p_0 is (X, Y);
2. For each $i \in \{1, \ldots, n\}$, let p_{i-1} be (X_{i-1}, Y_{i-1}). Then Spoiler makes a move of one of the following types:
   Splitting: Spoiler chooses two teams X', X'' such that $X_{i-1} = X' \cup X''$. Then Duplicator chooses two teams Y', Y'' such that $Y_{i-1} = Y' \cup Y''$. Then Spoiler chooses whether the next position p_i is (X', Y') or (X'', Y'').
   Supplementing: Spoiler chooses a variable v and a function $H : X_{i-1} \to \mathcal{P}(\mathtt{Dom}(A))\backslash\{\emptyset\}$. Then Duplicator chooses a function $K : Y_{i-1} \to \mathcal{P}(\mathtt{Dom}(B))\backslash\{\emptyset\}$, and the new position p_i is $(X_{i-1}[H/v], Y_{i-1}[K/v])$.
   Duplication: Spoiler chooses a variable v. The next position p_i is $(X_{i-1}[A/v], Y_{i-1}[B/v])$.
3. The final position $p_n = (X_n, Y_n)$ is *winning for Spoiler* if and only if there exists a formula α which is either a first-order literal, or an inclusion atom, such that $A \models_{X_n} \alpha$, but $B \not\models_{Y_n} \alpha$. Otherwise, the final position is winning for Duplicator.

The rank of an Inclusion Logic formula is also defined much in the same way as the rank of a Dependence Logic formula:

▶ **Definition 28.** Let ϕ be an FO($\subseteq$)-formula. Then we define its *rank* $\mathtt{rk}(\phi) \in \mathbb{N}$ by structural induction on ϕ, as follows:

1. If ϕ is a first-order literal or an inclusion atom, $\mathtt{rk}(\phi) = 0$;
2. $\mathtt{rk}(\psi \wedge \theta) = \max(\mathtt{rk}(\psi), \mathtt{rk}(\theta))$;
3. $\mathtt{rk}(\psi \vee \theta) = \max(\mathtt{rk}(\psi), \mathtt{rk}(\theta)) + 1$;
4. $\mathtt{rk}(\exists v\psi) = \mathtt{rk}(\forall v\psi) = \mathtt{rk}(\psi) + 1$.

The next theorem shows that our games behave as required with respect to our notion of rank. Its proof is practically the same as for the EF game for FO(D) in [24].

▶ **Theorem 29.** *Let A and B be models and X and Y teams on A and B. Then Duplicator has a winning strategy in $G_n(A, X, B, Y)$ if and only if*

$$A \models_X \phi \implies B \models_Y \phi$$

holds for all FO(⊆)*-formulas ϕ with $\mathtt{rk}(\phi) \leq n$.*

Due to the equivalence between FO(⊆) and GFP$^+$ we can conclude at once that the EF game for Inclusion Logic is also a novel EF game for GFP$^+$, rather different in structure from the one introduced in [2]. It may be hoped that this new game and its variants could be of some use for studying the expressive power of fixed point logics.

Although the EF game for Inclusion Logic has a clear second order flavour, it is still manageable: we will next show that Duplicator has a concrete winning strategy, when the models are simple enough.

▶ **Proposition 30.** *Let $A = \{1, \ldots, n\}$ and $B = \{1, \ldots, n+1\}$ be two finite models over the empty signature. Then for all* FO(⊆)*-sentences ϕ of rank $\leq n$,*

$$A \models \phi \implies B \models \phi.$$

Proof. It suffices to specify a winning strategy for Duplicator in the game $G_n(A, \{\emptyset\}, B, \{\emptyset\})$. Our aim for such a strategy is to preserve the following property for n turns:

- If the current position is (X, Y) then

$$Y = \bigcup \{\pi[X] : \pi \in I(A, B)\}, \tag{3}$$

where $I(A, B)$ is the set of all 1-1 functions $A \to B$, $\pi[X] = \{\pi(s) : s \in X\}$ and $\pi(s)$ denotes the assignment $\pi \circ s$.

The property (3) is trivially true for $(\{\emptyset\}, \{\emptyset\})$. Furthermore, as long as (3) holds, Spoiler does not win. Indeed, if α is a first-order literal such that $A \models_s \alpha$ for all $s \in X$, then, since all $s' \in Y$ are of the form $\pi(s)$ for some $s \in X$ and the signature is empty, we have $B \models_{s'} \alpha$ for all $s' \in Y$. Similarly, suppose that $A \models_X \vec{u} \subseteq \vec{w}$, and let $s' \in Y$. Then $s' = \pi(s)$ for some $s \in X$ and some $\pi \in I(A, B)$, and there exists a $h \in X$ such that $\vec{u}\langle s\rangle = \vec{w}\langle h\rangle$. But then $\pi(h) \in Y$, and $\vec{w}\langle\pi(h)\rangle = \vec{u}\langle\pi(s)\rangle = \vec{u}\langle s'\rangle$, as required.

Thus, we only need to verify that Duplicator can maintain property (3) for n rounds. Suppose that at round $i < n$ the current position (X, Y) has property (3), and let us consider the possible moves of Spoiler:

Splitting: Suppose that Spoiler splits X into X_1 and X_2. Then let Duplicator reply by splitting Y into $Y_j = \{s' \in Y : \exists \pi \in I(A, B) \exists s \in X_j \text{ such that } \pi(s) = s'\}$ for $j \in \{1, 2\}$. Then $Y = Y_1 \cup Y_2$, and it is straightforward to check that both possible successors (X_1, Y_1) and (X_2, Y_2) have property (3).

Supplementing: Suppose that Spoiler chooses a function $H \in \mathcal{C}(X, A)$. Then let Duplicator reply with the function $K \in \mathcal{C}(Y, B)$ defined as

$$K(s') = \{\pi(a) : \exists \pi \in I(A, B) \exists s \in X \text{ such that } \pi(s) = s' \text{ and } a \in H(s)\}$$

for each $s' \in Y$. We leave it to the reader to verify that the next position $(X[H/v], Y[K/v])$ has property (3).

Duplication: If Spoiler chooses a duplication move, the next position is $(X[M/v], Y[M/v])$. We check that this new position satisfies property (3).

Let $s[a/v] \in X[A/v]$ and let $\pi \in I(A, B)$. Since $s \in X$, we have that $\pi(s) \in Y$, and therefore $\pi(s)[\pi(a)/v] = \pi(s[a/v]) \in Y[B/v]$.

Conversely, let $s' \in Y$ and let b be any element of B. We need to show that $s'[b/v] = \pi(s[a/v])$ for some $\pi \in I(A, B)$, $s \in X$ and $a \in \mathtt{Dom}(A)$.

By induction hypothesis, there exists $\pi \in I(A, B)$ and $s \in X$ such that $\pi(s) = s'$. If b is in the range of π, then $s'[b/v] = \pi(s[a/v])$, where $a = \pi^{-1}(b)$. On the other hand, if b is not in the range of π, then since $i < n$, there is an element $a \in A$ which is not in the range of s. Now $s[a/v] \in X[A/v]$, and $s'[b/v] = \pi'(s[a/v])$, where $\pi' \in I(A, B)$ is a function such that $\pi'(a) = b$ and $\pi'(c) = \pi(c)$ for all c in the range of s. ◀

From Proposition 30 it immediately follows that *even cardinality* (and other similar cardinality properties) of finite models is not definable in Inclusion Logic. This, of course, follows already from the equivalence of FO($\subseteq$) and GFP$^+$, as it is well-known that non-trivial cardinality properties are not definable in fixed point logics.

6 Conclusions and Further Work

In this work, we proved a number of results concerning the expressive power of inclusion Logic. We showed that this logic is strictly weaker than Σ_1^1, and corresponds in fact to Positive Greatest Fixed Point Logic. Furthermore, we showed that all union-closed first-order properties of relations correspond to the satisfaction conditions of Inclusion Logic formulas, and we also defined a new Ehrenfeucht-Fraïssé game for it.

Due to the connection between Inclusion Logic and fixed point logics, the study of this formalism may have interesting applications in descriptive complexity theory. In [5], Durand and Kontinen established some correspondences between fragments of Dependence Logic and fragments of NP; in the same way, one may hope to find correspondences between fragments of Inclusion Logic and fragments of PTIME.

Furthermore, we may inquire about extensions of Inclusion Logic. For example, is there any natural union-closed dependency notion $\mathbf{D}$ such that FO($\subseteq, \mathbf{D}$) defines all Σ_1^1 union-closed properties of relations? By the results in Section 4, we know that if this is the case, then $\mathbf{D}$ is not first-order.

Acknowledgements. Pietro Galliani gratefully acknowledges the support of grant 264917 of the Academy of Finland. We thank Erich Grädel, Miika Hannula, Juha Kontinen and Jouko Väänänen for a number of highly useful suggestions and comments. We especially thank Miika Hannula for pointing out an error in a previous version of the paper. Finally, we thank the referees for a number of useful suggestions and comments.

References

1 William W. Armstrong. Dependency Structures of Data Base Relationships. In *Proc. of IFIP World Computer Congress*, pages 580–583, 1974.

2 Uwe Bosse. An "Ehrenfeucht-Fraïssé game" for fixpoint logic and stratified fixpoint logic. In *Computer science logic*, pages 100–114. Springer, 1993.

3 Marco A. Casanova, Ronald Fagin, and Christos H. Papadimitriou. Inclusion dependencies and their interaction with functional dependencies. In *Proceedings of the 1st ACM SIGACT-SIGMOD symposium on Principles of database systems*, PODS '82, pages 171–176, New York, NY, USA, 1982. ACM.

4 Marco A. Casanova and Vânia M. P. Vidal. Towards a sound view integration methodology. In *Proceedings of the 2nd ACM SIGACT-SIGMOD symposium on Principles of database systems*, PODS '83, pages 36–47, New York, NY, USA, 1983. ACM.

5 Arnaud Durand and Juha Kontinen. Hierarchies in dependence logic. *ACM Trans. Comput. Log.*, 13(4), 2012, 31 pages.

6 Herbert B. Enderton. Finite partially-ordered quantifiers. *Mathematical Logic Quarterly*, 16(8):393–397, 1970.

7 Fredrik Engström. Generalized quantifiers in dependence logic. *Journal of Logic, Language and Information*, 21(3):299–324, 2012.

8 Ronald Fagin. A normal form for relational databases that is based on domains and keys. *ACM Transactions on Database Systems*, 6:387–415, September 1981.

9 Pietro Galliani. *The Dynamics of Imperfect Information.* PhD thesis, University of Amsterdam, September 2012.

10 Pietro Galliani. Inclusion and exclusion dependencies in team semantics: On some logics of imperfect information. *Annals of Pure and Applied Logic*, 163(1):68 – 84, 2012.

11 Dan Geiger, Azaria Paz, and Judea Pearl. Axioms and algorithms for inferences involving probabilistic independence. *Information and Computation*, 91(1):128–141, 1991.

12 Erich Grädel and Jouko Väänänen. Dependence and Independence. *Studia Logica*, 101(2):399–410, 2013.

13 Leon Henkin. Some Remarks on Infinitely Long Formulas. In *Infinitistic Methods. Proc. Symposium on Foundations of Mathematics*, pages 167–183. Pergamon Press, 1961.

14 Jaakko Hintikka and Gabriel Sandu. Informational independence as a semantic phenomenon. In J.E Fenstad, I.T Frolov, and R. Hilpinen, editors, *Logic, methodology and philosophy of science*, pages 571–589. Elsevier, 1989.

15 Wilfrid Hodges. *A Shorter Model Theory.* Cambridge University Press, 1997.

16 Wilfrid Hodges. Compositional Semantics for a Language of Imperfect Information. *Journal of the Interest Group in Pure and Applied Logics*, 5 (4):539–563, 1997.

17 Neil Immerman. Relational queries computable in polynomial time. *Information and control*, 68(1):86–104, 1986.

18 Neil Immerman. Languages That Capture Complexity Classes. *SIAM Journal of Computing*, 16:760–778, 1987.

19 Juha Kontinen and Jouko Väänänen. On definability in dependence logic. *Journal of Logic, Language and Information*, 3(18):317–332, 2009.

20 Antti Kuusisto. Defining a double team semantics for generalized quantifiers. URN:ISBN:978-951-44-8882-5, 2012.

21 Leonid Libkin. *Elements of Finite Model Theory.* Springer, 2004.

22 Allen L. Mann, Gabriel Sandu, and Merlijn Sevenster. *Independence-Friendly Logic: A Game-Theoretic Approach.* Cambridge University Press, 2011.

23 Yannis Moschovakis. *Elementary Induction on Abstract Structures.* North Holland, 1974.

24 Jouko Väänänen. *Dependence Logic.* Cambridge University Press, 2007.

25 Jouko Väänänen and Pietro Galliani. On dependence logic. ArXiv:1305.5948, 2013.

26 Moshe Y. Vardi. The complexity of relational query languages. In *Proceedings of the fourteenth annual ACM symposium on Theory of computing*, pages 137–146. ACM, 1982.

27 Wilbur John Walkoe. Finite partially-ordered quantification. *The Journal of Symbolic Logic*, 35(4):pp. 535–555, 1970.