Technical Report AC-TR-15-007

April 2015

The Complexity of Equivalence, Entailment, and Minimization in Existential Positive Logic

Simone Bova and Hubie Chen

This is the authors' copy of a paper that appears in J. Comput. Syst. Sci. 81(2): 443-457 (2015).

www.ac.tuwien.ac.at/tr

The Complexity of Equivalence, Entailment, and Minimization in Existential Positive Logic

Simone Bova[*1] **and Hubie Chen**[2,3]

1 **Institut für Informationssysteme**
Technische Universität Wien
A-1040 Wien, Austria
`simone.bova@tuwien.ac.at`

2 **Departamento LSI, Facultad de Informática**
Universidad del País Vasco
E-20018 San Sebastián, Spain
`hubie.chen@ehu.es`

3 **IKERBASQUE, Basque Foundation for Science**
E-48011 Bilbao, Spain

Abstract

The existential positive fragment of first-order logic is the set of formulas built from conjunction, disjunction, and existential quantification. On sentences from this fragment, we study three fundamental computational problems: logical equivalence, entailment, and the problem of deciding, given a sentence and a positive integer k, whether or not the sentence is logically equivalent to a k-variable sentence. We study the complexity of these three problems, and give a description thereof with respect to all relational signatures. In particular, we establish for the first time that, over a signature containing a relation symbol of binary (or higher) arity, all three of these problems are complete for the complexity class Π_2^p of the polynomial hierarchy.

Keywords and phrases Existential Positive Logic, Equivalence, Entailment, Finite-Variable Logics, Computational Complexity

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

1.1 Background and Motivation

The undecidability of the *Entscheidungsproblem*—given a first-order sentence, decide if it is valid—immediately implies the undecidability of the fundamental problems of testing equivalence and testing entailment on input pairs of first-order sentences. Nonetheless, certain fragments of relational first-order logic have been shown to admit equivalence and entailment problems that are decidable. Somewhat recently, these two problems were shown to be decidable for *conjunctive positive logic*, the fragment of formulas built from conjunction ($\wedge$) and both quantifiers ($\forall, \exists$) [6]. These two problems have indeed been long known to be decidable in the more restrictive fragment of *primitive positive logic*, which consists of those formulas built from conjunction ($\wedge$) and existential quantification ($\exists$); indeed, in this fragment these problems admit a relatively tame complexity grading, being both NP-complete [3]. From the decidability of these problems in primitive positive logic, it can be readily verified that these two problems remain decidable in *existential positive logic*, which consists of those

* Supported by ERC (COMPLEX REASON, 239962).

formulas built from conjunction ($\wedge$), disjunction ($\vee$), and existential quantification ($\exists$). In the database and knowledge representation literature, the problem of entailment is often referred to as *query containment* or *subsumption*, and along with equivalence is considered a basic reasoning task.

We here study, in existential positive logic, the complexity of equivalence and entailment, as well as a third basic problem, which we now turn to describe; let us assume, in the rest of this section, that all sentences under discussion are existential positive. A recent study on the complexity of model checking in existential positive logic [5] revealed that the number of variables needed to express a sentence is the crucial parameter determining complexity; specifically, it was shown that on a set $\mathcal{F}$ of bounded-arity sentences, model checking on $\mathcal{F}$ is fixed-parameter tractable if there exists a constant k such that each sentence in $\mathcal{F}$ is logically equivalent to a k-variable sentence (a sentence in which at most k variables are present); otherwise, model checking is not fixed-parameter tractable (under standard complexity-theoretic assumptions). From the perspective of this result, the computational problem of determining exactly how many variables are needed to express a sentence is very well-motivated. Here, we study the following decision version of this problem: given a sentence and a constant k, decide if the sentence is logically equivalent to a k-variable sentence. For the purposes of discussion, let us call this the *minimization* problem. This problem has been studied and shown to be NP-complete in primitive positive logic [7].

1.2 Results

In this paper, we characterize the complexity of minimization, equivalence, and entailment in existential positive logic, over fixed relational signatures. We establish the following results; recall that Π_2^p is a complexity class located at the second level of the *polynomial hierarchy*, and contains both NP and coNP.

We begin by studying the case where a symbol of binary or higher arity is present.

▶ Results 1 (at least binary arity). On a signature containing a relation symbol of at least binary arity, all three of the problems are Π_2^p-complete.

Note that the equivalence problem was shown to be Π_2^p-complete by Sagiv and Yannakakis [11], however, no analysis of the signature was performed in that work. Our result thus strengthens theirs by showing that the same level of hardness can be achieved even (for example) in the case of a signature with a single binary relation symbol.

Hardness for Π_2^p is proved for all three problems via a unified argument that reduces from a quantified version of the classical *graph colorability* problem. For the minimization problem, we in fact show that the problem exhibits this maximal Π_2^p-complete complexity even when k is fixed as any sufficiently large integer.

On a unary signature, by which we mean a signature containing only unary relation symbols, each sentence is logically equivalent to a 1-variable sentence, and hence the minimization problem becomes trivial. We do, though, persist in studying the other two problems on unary signatures, where we demonstrate the following phenomena.

▶ Results 2 (unary arity). The following hold for the equivalence and entailment problems.

- On a unary signature of infinite size, these two problems continue to be Π_2^p-complete.
- On the other hand, on a unary signature of finite size, these two problems are solvable in polynomial time with a constant number of queries to an NP oracle (and thus are not Π_2^p-hard unless the polynomial hierarchy collapses), yet we give a hardness criterion, which, in the case of a signature with two distinct symbols, shows that these problems

are coDP-hard. The class coDP contains a language if it is the union of an NP language and a coNP language; coDP-hardness thus implies both NP-hardness and coNP-hardness.
- Finally, as we discuss in the paper, these two problems are readily verified to be in P on a unary signature with at most one symbol.

To sum up, we obtain a comprehensive complexity profile of the studied problems with respect to all relational signatures, as depicted in the following table, where σ is a relational signature; for precise statements, refer to Section 4, and for more information on the finite unary case, refer to Section 7.2.

	EP_σ-Entail, EP_σ-$\equiv$	EP^k_σ-Expr
σ unary, $\lvert\sigma\rvert \leq 1$	in P	trivial (for all $k \geq 1$)
σ unary finite, $\lvert\sigma\rvert > 1$	coDP-hard, in $\mathrm{P}^{\mathrm{NP[const]}}$	trivial (for all $k \geq 1$)
σ unary infinite	Π^p_2-complete	trivial (for all $k \geq 1$)
$\sigma \ni R$, R at least binary	Π^p_2-complete	Π^p_2-complete (for all $k \geq 6$)

The present article, in part, extends some of the material of an article that appeared in the proceedings of the 17th International Conference on Database Theory (ICDT 2014) and that focused on the problems EP^k_σ-Expr [2].

2 Preliminaries

For an integer $k \geq 0$, we use $\underline{k}$ to denote the set $\{1, \ldots, k\}$, with the convention that $\underline{0} = \emptyset$.

In this paper, we focus on relational first-order logic. A *signature* σ is a set of *relation symbols*, each of which has an associated natural number called its *arity*.

2.1 Structures

A *structure* $\mathbf{A}$ (over signature σ) is specified by a nonempty set A called the *universe* of the structure and denoted by the corresponding non-bold letter, and a relation $R^{\mathbf{A}} \subseteq A^r$ for each arity r relation symbol $R \in \sigma$. A structure is *finite* if its universe is finite.

A collection of structures is said to be *similar* if they share the same signature. Let $\mathbf{A}, \mathbf{B}$ be similar structures on the signature σ. The *union* of $\mathbf{A}$ and $\mathbf{B}$ is the structure $\mathbf{A} \cup \mathbf{B}$ with universe $A \cup B$ and with $R^{\mathbf{A} \cup \mathbf{B}} = R^{\mathbf{A}} \cup R^{\mathbf{B}}$ for each arity r relation symbol $R \in \sigma$. A *homomorphism* from $\mathbf{A}$ to $\mathbf{B}$ is a mapping $h : A \to B$ such that for each symbol $R \in \sigma$, it holds that $h(R^{\mathbf{A}}) \subseteq R^{\mathbf{B}}$, by which is meant that for each tuple $(a_1, \ldots, a_k) \in R^{\mathbf{A}}$, one has $(h(a_1), \ldots, h(a_k)) \in R^{\mathbf{B}}$. We will sometimes simply write $\mathbf{A} \to \mathbf{B}$ to indicate that there exists a homomorphism from $\mathbf{A}$ to $\mathbf{B}$. We say that $\mathbf{A}$ and $\mathbf{B}$ are *homomorphically equivalent* if $\mathbf{A} \to \mathbf{B}$ and $\mathbf{B} \to \mathbf{A}$ both hold.

The structure $\mathbf{B}$ is a *substructure* of the structure $\mathbf{A}$ if $B \subseteq A$ and $R^{\mathbf{B}} \subseteq R^{\mathbf{A}}$ for all relation symbols R. When $\mathbf{B}$ is a substructure of $\mathbf{A}$, there exists a homomorphism h from $\mathbf{A}$ to $\mathbf{B}$, and h fixes each element $b \in B$, the mapping h is said to be a *retraction* from $\mathbf{A}$ to $\mathbf{B}$; when there exists a retraction from $\mathbf{A}$ to $\mathbf{B}$, it is said that $\mathbf{A}$ *retracts* to $\mathbf{B}$. A *core* of the structure $\mathbf{A}$ is a structure $\mathbf{C}$ such that $\mathbf{A}$ retracts to $\mathbf{C}$, but $\mathbf{A}$ does not retract to any proper substructure of $\mathbf{C}$. We will make use of the following well-known facts on cores [9]: (1) each finite structure has a core; (2) all cores of a finite structure are isomorphic. From these facts, it is reasonable to speak of *the* core of a finite structure, which we do, and we use $\mathsf{core}(\mathbf{A})$ to denote a representative from the set of all cores of a finite structure $\mathbf{A}$.

We define the *Gaifman graph* of a structure $\mathbf{B}$ to be the undirected graph $G(\mathbf{B})$ with vertex set B and having an edge $\{b, b'\}$ if and only if b and b' co-occur in a tuple of $\mathbf{B}$.

A *tree decomposition* of an undirected graph G with vertex set B is a pair (T, β) consisting of a tree T and a map $\beta : V^T \to \wp(B)$ defined on the vertex set V^T of T such that, for each vertex $t \in V^T$, it holds that $\beta(t)$ is a non-empty subset of B, called the *bag* of t, and the following conditions hold:

- For each $b \in B$, the vertices $\{t \mid b \in \beta(t)\}$ form a connected subtree of T.
- For each edge $\{b, b'\}$ of G, there exists a vertex $t \in V^T$ such that $\{b, b'\} \subseteq \beta(t)$.

The *width* of a tree decomposition (T, β) is defined as $(\max_{t \in V^T} |\beta(t)|) - 1$. The *treewidth* of an undirected graph G, denoted by $\mathsf{tw}(G)$, is the minimum width over all tree decompositions of G; the treewidth of a structure $\mathbf{B}$, denoted by $\mathsf{tw}(\mathbf{B})$, is defined as $\mathsf{tw}(G(\mathbf{B}))$.

2.2 Formulas

An *atom* (over signature σ) is an equality of variables $(x = y)$ or is a predicate application $R(x_1, \ldots, x_r)$, where $x_1, \ldots, x_r$ are variables, and $R \in \sigma$ is an arity r relation symbol. A *formula* (over signature σ) is built from atoms (over σ), conjunction ($\wedge$), disjunction ($\vee$), universal quantification ($\forall$), and existential quantification ($\exists$). A *sentence* is a formula having no free variables. We let FO denote the set of first-order formulas. For each set L of first-order formulas and each integer $k \geq 1$, we let L^k denote the subset of L containing formulas that use at most k variables, and L_σ denote the subset of L containing formulas over signature σ.

An *existential positive* formula (over signature σ) is a formula built from atoms (over σ) using conjunction, disjunction, and existential quantification; we let EP denote the set of existential positive formulas. A *primitive positive* formula (over signature σ) is a formula built from atoms (over σ) using conjunction and existential quantification; we let PP denote the set of primitive positive formulas.

We use the following standard terminology and notation from logic. For a structure $\mathbf{A}$ and a sentence ϕ over the same signature, we write $\mathbf{A} \models \phi$ if the sentence ϕ is *true* in the structure $\mathbf{A}$. When $\mathbf{A}$ is a structure, f is a mapping from variables to the universe of $\mathbf{A}$, and ψ is a formula over the signature of $\mathbf{A}$, we write $\mathbf{A}, f \models \psi$ to indicate that ψ is satisfied by $\mathbf{A}$ and f. Let ϕ and ψ be sentences over the same signature σ. Then, ϕ *entails* ψ (denoted $\phi \models \psi$) if, for all structures $\mathbf{A}$ over σ, it holds that $\mathbf{A} \models \phi$ implies $\mathbf{A} \models \psi$; also, ϕ and ψ are *logically equivalent* (denoted $\phi \equiv \psi$) if $\phi \models \psi$ and $\psi \models \phi$.

We use the following terminology and notation. Let σ be a signature, let ϕ be a primitive positive formula over σ, and let $\mathbf{A}$ be a finite structure over σ. By the *existential closure* of a formula, we mean the sentence obtained by existentially quantifying the free variables of the formula.

- $\mathbf{C}[\phi]$ denotes the *canonical structure* induced by the existential closure of ϕ, as follows. Let ϕ^c be the existential closure of the prenex form of ϕ. Let $\mathsf{elim}_=(\phi^c)$ be obtained by eliminating equalities from ϕ^c using the following syntactic transformations: for each equality $x = y$ on distinct variables, replace all instances of y with x in the quantifier free part, and remove the quantifier $\exists y$ from the prefix; remove equalities of the form $x = x$. Define $\mathbf{C}[\phi]$ to be the structure having a universe element for each existentially quantified variable in $\mathsf{elim}_=(\phi^c)$, and where, for each $R \in \sigma$, the relation $R^{\mathbf{C}[\phi]}$ contains $(x_1, \ldots, x_k)$ if and only if $R(x_1, \ldots, x_k)$ appears in the quantifier free part of $\mathsf{elim}_=(\phi^c)$.
- $Q[\mathbf{A}]$ denotes the *canonical query* of $\mathbf{A}$, defined as follows. If $A = \{a_1, \ldots, a_n\}$, then

$$Q[\mathbf{A}] = \exists a_1 \ldots \exists a_n \bigwedge_{R \in \sigma} \bigwedge_{(a'_1, \ldots, a'_k) \in R^{\mathbf{A}}} R(a'_1, \ldots, a'_k).$$

Technical Report AC-TR-15-007

We will use the following known fact.

▶ Proposition 1. (Chandra-Merlin [3]) Let ϕ be a sentence and let $\mathbf{A}$ be a finite structure, such that $\phi = Q[\mathbf{A}]$ or $\mathbf{A} = \mathbf{C}[\phi]$. Then, for any structure $\mathbf{B}$, it holds that $\mathbf{A} \to \mathbf{B}$ if and only if $\mathbf{B} \models \phi$.

It is straightforward to verify that the existential closure of any primitive positive formula ϕ is logically equivalent to $Q[\mathbf{C}[\phi]]$, and that every finite structure $\mathbf{A}$ is homomorphically equivalent to $\mathbf{C}[Q[\mathbf{A}]]$.

3 Existential Positive Logic

In this section, we define the problems under study, establish some basic facts on existential positive logic, and place the problems in the complexity class Π_2^p. In related work, Sagiv and Yannakakis [11] showed containment in Π_2^p of the problem EP-Equiv; here, using the formalism of first-order logic, we give a treatment that places all three of the studied problems in Π_2^p.

▶ **Definition 2.** We define the following computational problems:

- EP-Entail: Given a pair (ϕ, ψ) of sentences in EP, decide whether $\phi \models \psi$.
- EP-Equiv: Given a pair (ϕ, ψ) of sentences in EP, decide whether $\phi \equiv \psi$.
- EP-Expr: Given a sentence $\phi \in$ EP and an integer $k \geq 1$, decide whether ϕ is logically equivalent to a sentence in EP^k.

Moreover, for every signature σ and every integer $m \geq 1$, we define the following computational problems as restrictions of the above problems:

- EP_σ-Entail is the restriction of EP-Entail to instances where $\phi, \psi \in \text{EP}_\sigma$.
- EP_σ-$\equiv$ is the restriction of EP-Equiv to instances where $\phi, \psi \in \text{EP}_\sigma$.
- EP_σ^m-Expr is the restriction of EP-Expr to instances where $\phi \in \text{EP}_\sigma$ and $k = m$.

Note that throughout this paper, the only notion of reduction that we use is many-one polynomial-time reduction.

▶ **Definition 3.** A sentence ϕ in EP is in *disjunctive* form if $\phi = \bigvee_{i \in \underline{n}} \phi_i$, where, for all $i \in \underline{n}$, ϕ_i is a sentence in PP; such a disjunctive form is *irredundant* if there do not exist distinct $i, j \in \underline{n}$ such that $\phi_i \models \phi_j$.

We will make use of the following syntactic transformations, which preserve logical equivalence:

$$\exists x(\theta \vee \theta') \equiv \exists x\theta \vee \exists x\theta'; \tag{E1}$$
$$\theta \wedge (\theta' \vee \theta'') \equiv (\theta \wedge \theta') \vee (\theta \wedge \theta''); \tag{E2}$$
$$\exists x\theta \equiv \theta, \qquad \text{if } x \text{ not free in } \theta; \tag{E3}$$
$$\theta \vee \theta' \equiv \theta', \qquad \text{if } \theta \models \theta'. \tag{E4}$$

Given an arbitrary existential positive sentence, an equivalent existential positive sentence in disjunctive form is computable by iterated syntactic replacements exploiting the facts (E1) and (E2) above; also, given an existential positive sentence in disjunctive form, an equivalent existential positive sentence in irredundant disjunctive form is computable by iterated syntactic replacements exploiting the fact (E4) above.

The proof that our computational problems are contained in the complexity class Π_2^p relies on the following lemma.

Technical Report AC-TR-15-007

▶ **Lemma 4.** *Let ϕ and ψ be sentences in* EP*. Let $\bigvee_{i\in\underline{m}}\phi_i$ and $\bigvee_{j\in\underline{n}}\psi_j$ be disjunctive forms in* EP *logically equivalent to ϕ and ψ, respectively. The following hold.*

1. *$\phi \models \psi$ if and only if, for all $i \in \underline{m}$, there exists $j \in \underline{n}$ such that $\phi_i \models \psi_j$.*
2. *If the above disjunctive forms are irredundant and $\phi \equiv \psi$, then $m = n$ and there exists a bijection $\pi\colon \underline{m} \to \underline{m}$ such that for all $i \in \underline{m}$ it holds that $\phi_i \equiv \psi_{\pi(i)}$.*
3. *Let $k \geq 1$ be an integer. Then, ϕ is logically equivalent to a sentence in* EP^k *if and only if, for all $i \in \underline{m}$, there exists $i' \in \underline{m}$ such that $\phi_i \models \phi_{i'}$ and $\phi_{i'}$ is logically equivalent to a sentence in* PP^k*.*

Proof. For (1), the backwards direction is clear. For the forwards direction, let $i \in \underline{m}$. We have $\mathbf{C}[\phi_i] \models \phi$, from which it follows that $\mathbf{C}[\phi_i] \models \psi$. We must then have that there exists $j \in \underline{n}$ such that $\mathbf{C}[\phi_i] \models \psi_j$, from which the result follows from Proposition 1.

For (2), let $i \in \underline{m}$. By (1), there exists $j \in \underline{n}$ such that $\phi_i \models \psi_j$. We claim that $\phi_i \equiv \psi_j$. This is because there exists $i' \in \underline{m}$ such that $\psi_j \models \phi_{i'}$; if $i \neq i'$, then this implies that the disjunctive form for ϕ is not irredundant, a contradiction. Since the disjunctive form for ψ is irredundant, there is a unique $j \in \underline{n}$ satisfying the condition $\phi_i \equiv \psi_j$, and we thus obtain an injection $\pi : \underline{m} \to \underline{n}$, as well as that $m \leq n$. By symmetric reasoning, we obtain that $n \leq m$ and so $m = n$ and the injection π is a bijection.

For (3), first let $\phi \in \mathrm{EP}$. If ϕ is logically equivalent to a sentence in EP^k, say ϕ', then the disjunctive form of ϕ' obtained using the above transformations (E1), (E2) and (E3) is such that each disjunct is a primitive positive sentence in PP^k. This implies that there is an irredundant disjunctive form $\bigvee_{j\in\underline{n}}\psi_j$ logically equivalent to ϕ where each disjunct is in PP^k. By (1), for any $i \in \underline{m}$, there exists $j \in \underline{n}$ such that $\phi_i \models \psi_j$. Since there is a sub-disjunction of $\bigvee_{i\in\underline{m}}\phi_i$ that is irredundant, by (2) there exists $i' \in \underline{m}$ such that $\phi_{i'}$ and ψ_j are logically equivalent. We then have $\phi_i \models \phi_{i'}$, as desired.

Now suppose that $\rho : \underline{m} \to \underline{m}$ is a mapping such that for each $i \in \underline{m}$, it holds that $\phi_i \models \phi_{\rho(i)}$ and each $\phi_{\rho(i)}$ is logically equivalent to a sentence in PP^k. Then ϕ is logically equivalent to $\bigvee_{i\in\underline{m}}\phi_{\rho(i)}$. ◀

We remark that entailment and finite entailment coincide in existential positive logic; this can be seen from the proof of Lemma 4(1).

The conditions in Lemma 4(1) and Lemma 4(3) allow to establish containment in Π_2^p for the problems under consideration.

▶ Proposition 5. The problems EP-Entail, EP-Equiv, and EP-Expr are in the complexity class Π_2^p.

Proof. Let ϕ be a sentence in EP built using variables $x_1, \ldots, x_n$; in polynomial time, it may be transformed to prenex form, so let us assume that ϕ is in prenex form. Let $\mathsf{atoms}(\phi)$ be the set of all atoms occurring in ϕ. For each mapping $f\colon \mathsf{atoms}(\phi) \to \{0,1\}$, let ϕ_f be the primitive positive sentence defined as the existential closure of $\bigwedge_{\alpha\in\mathrm{atoms}(\phi), f(\alpha)=1}\alpha$. Let $\mathsf{eval}(\phi, f)$ denote the result of evaluating the Boolean expression (over $\wedge$ and $\vee$) obtained by replacing, in the quantifier free part of ϕ, every occurrence of α by $f(\alpha)$, for all $\alpha \in \mathrm{atoms}(\phi)$. We observe two facts. First, if $\mathsf{eval}(\phi, f) = 1$, then $\phi_f \models \phi$. Second, let $\mathbf{A}$ be any structure. Let $g\colon \{x_1, \ldots, x_n\} \to A$ be such that the quantifier free part of ϕ is true in $\mathbf{A}$ under g. Let $\{\alpha_1, \ldots, \alpha_k\}$ be the subset of atoms from $\mathrm{atoms}(\phi)$ that are true in in $\mathbf{A}$ under g. Let $f\colon \mathrm{atoms}(\phi) \to \{0,1\}$ be such that $f(\alpha) = 1$ if and only if $\alpha \in \{\alpha_1, \ldots, \alpha_k\}$. Clearly, $\mathbf{A} \models \phi_f$. Moreover, $\mathsf{eval}(\phi, f) = 1$. The two observed facts imply that the disjunctive existential positive sentence $\mathsf{df}(\phi)$ defined by $\bigvee_f \phi_f$, where f ranges over all mappings $f\colon \mathsf{atoms}(\phi) \to \{0,1\}$ such that $\mathsf{eval}(\phi, f) = 1$, is logically equivalent to ϕ.

We prove that EP-Entail is in Π^p_2. Let (ϕ, ψ) be an instance of EP-Entail. We can assume without loss of generality that ϕ and ψ are in prenex form. By the above, $\phi \equiv \mathsf{df}(\phi)$ and $\psi \equiv \mathsf{df}(\psi)$. By Lemma 4(1), $\phi \models \psi$ if and only if the following condition holds: for all disjuncts ϕ_f in $\mathsf{df}(\phi)$, there exists a disjunct ψ_g in $\mathsf{df}(\psi)$ such that $\phi_f \models \psi_g$. To decide this condition, one can check whether for all assignments $f\colon \text{atoms}(\phi) \to \{0, 1\}$, there exist an assignment $g\colon \text{atoms}(\psi) \to \{0, 1\}$ and a map h from $C[\psi_g]$ to $C[\phi_f]$ such that if $\mathsf{eval}(\phi, f) = 1$, then $\mathsf{eval}(\psi, g) = 1$ and h is a homomorphism from $\mathbf{C}[\psi_g]$ to $\mathbf{C}[\phi_f]$. This is justified by the above discussion and Proposition 1. Therefore, EP-Entail is in Π^p_2.

To prove that EP-Equiv is in Π^p_2, note that the problem EP-Entailed $= \{(\psi, \phi) \mid (\phi, \psi) \in \text{EP-Entail}\}$ is in Π^p_2 by the proof just given, and that EP-Equiv = EP-Entail $\cap$ EP-Entailed. The result then follows because Π^p_2 is closed under intersection.

Finally, we prove that EP-Expr is in Π^p_2. Let ϕ and k be an instance of EP-Expr. We have $\phi \equiv \mathsf{df}(\phi)$. Note that, by Lemma 4(3), ϕ is logically equivalent to a sentence in EP^k if and only if, for all $f\colon \mathsf{atoms}(\phi) \to \{0, 1\}$ such that $\mathsf{eval}(\phi, f) = 1$, there exists $g\colon \mathsf{atoms}(\phi) \to \{0, 1\}$ such that $\mathsf{eval}(\phi, g) = 1$, such that $\phi_f \models \phi_g$ and ϕ_g is logically equivalent to a sentence in PP^k. By [7, Theorem 5], ϕ_g is logically equivalent to a sentence in PP^k if and only if $\mathbf{C}[\phi_g]$ has a homomorphically equivalent substructure $\mathbf{S}$ with $\mathsf{tw}(\mathbf{S}) < k$.

So, to check the given instance, one can check whether for all $f\colon \text{atoms}(\phi) \to \{0, 1\}$, there exist $g\colon \text{atoms}(\phi) \to \{0, 1\}$, a substructure $\mathbf{S}$ of $\mathbf{C}[\phi_g]$, a mapping $h : C[\phi_g] \to S$ and a tree decomposition of $\mathbf{S}$ such that: if $\mathsf{eval}(\phi, f) = 1$, then $\mathsf{eval}(\phi, g) = 1$, h is a homomorphism from $\mathbf{C}[\phi_g]$ to $\mathbf{S}$, and $\mathbf{S}$ has a tree decomposition witnessing $\mathsf{tw}(\mathbf{S}) < k$. Note that if there is such a tree decomposition, there is one that has size polynomial in $\mathbf{S}$. ◀

We also note that entailment and equivalence have the same complexity, although in the sequel we find it more transparent to prove complexity results directly for both problems.

▶ Proposition 6. For any signature σ, EP_σ-Entail and EP_σ-$\equiv$ are interreducible.

Proof. Let σ be any signature. Observe that EP_σ-Entail reduces to EP_σ-$\equiv$ via the mapping $(\phi, \psi) \mapsto (\phi, \phi \wedge \psi)$, and EP_σ-$\equiv$ reduces to EP_σ-Entail via the mapping $(\phi, \psi) \mapsto (\phi \vee \psi, \phi \wedge \psi)$. ◀

4 Complexity Results

▶ **Theorem 7.** *Let σ be a signature that contains a relation symbol of at least binary arity. For each $k \geq 6$, the problem EP^k_σ-Expr is Π^p_2-complete.*

Proof. Containment in Π^p_2 follows from Proposition 5. For Π^p_2-hardness, the case where σ contains a binary relation symbol is proved in Theorem 17 in Section 5; the higher-arity case is treated in Section 6. ◀

Note that if σ is a signature that contains only unary relation symbols, then each sentence in EP_σ is logically equivalent to a sentence in EP^1_σ, so the problem EP^k_σ-Expr is trivial for all $k \geq 1$.

▶ **Theorem 8.** *Let σ be a signature that contains a relation symbol of at least binary arity, or contains infinitely many unary relation symbols. The problems EP_σ-Entail and EP_σ-$\equiv$ are Π^p_2-complete.*

Proof. Containment in Π^p_2 follows from Proposition 5. For Π^p_2-hardness, the case where σ contains infinitely many unary relation symbols is proved in Theorem 19; the case where

Technical Report AC-TR-15-007

σ contains a binary relation symbol is proved in Theorem 18; and, the higher-arity case is treated in Section 6. ◀

We refer the reader to Section 7.2 for a discussion of the complexity of EP_σ-Entail and EP_σ-$\equiv$ where the signature σ consists of at least two but finitely many unary relation symbols.

Finally, if σ is empty, then EP_σ-Entail and EP_σ-$\equiv$ are trivial. If $\sigma = \{U\}$ is a signature that consists of one unary relation symbol, then the problems EP_σ-Entail and EP_σ-$\equiv$ are in P. Indeed, note that any sentence ϕ in EP_σ is logically equivalent to either $\exists x(U(x))$ or $\exists x(x = x)$. Moreover, it is possible to decide whether ϕ is logically equivalent to $\exists x(x = x)$ by evaluating the Boolean expression (over $\wedge$ and $\vee$) obtained as follows: first replace in ϕ atoms $U(x)$ by 0 and atoms $x = y$ by 1, and then remove all the quantifiers $\exists x$.

5 The Binary Case

In this section, we prove the hardness results for the case of signatures containing a relation symbol of binary arity. We do this by first presenting the source problem (a Π_2^p-complete problem) from which we will reduce (Section 5.1); then, we present an encoding of labelled digraphs as digraphs which will be used (Section 5.2). Following this, we present the reduction to be used (Section 5.3), and then confirm that the reduction yields the desired hardness result (Section 5.4).

5.1 Source Problem

When $\mathbf{B}$ is a structure, define Π_k-QCSP($\mathbf{B}$) to be the problem of deciding, given a Π_k prenex sentence Φ whose quantifier-free part is a conjunction of atoms without equality, whether or not $\mathbf{B} \models \Phi$; define Σ_k-QCSP($\mathbf{B}$) similarly, with respect to Σ_k sentences. For $q \geq 2$, we define the structure $\mathbf{K}_q$, the clique on q vertices, to be the structure with universe $\underline{q}$ and that interprets the binary relation symbol E by $E^{\mathbf{K}_q} = \{(i, j) \in \underline{q}^2 \mid i \neq j\}$. Our Π_2^p hardness results will be proved by showing reductions from the problems Π_2-QCSP($\mathbf{K}_q$), where $q \geq 3$.

▶ Proposition 9. (follows from [1]) Let $q \geq 3$. For each even $k \geq 2$, the problem Π_k-QCSP($\mathbf{K}_q$) is Π_k^p-complete; and, for each odd $k \geq 3$, the problem Σ_k-QCSP($\mathbf{K}_q$) is Σ_k^p-complete.

Proof. (idea) Let $\mathbf{B}$ be the structure with universe $\{0, 1\}$ and with a single relation, $R^{\mathbf{B}} = \{0, 1\}^3 \setminus \{(0, 0, 0), (1, 1, 1)\}$. Under the bounds on k given in the proposition statement, one has that Π_k-QCSP($\mathbf{B}$) and Σ_k-QCSP($\mathbf{B}$) are Π_k^p-complete and Σ_k^p-complete, respectively; this follows from [4, Theorem 7.2]. We use the construction of [1, Proposition 5.1] to give a reduction from those problems to the present problems. The only modification needed is the following. Each universally quantified variable in an instance of Π_k-QCSP($\mathbf{B}$) or Σ_k-QCSP($\mathbf{B}$) is translated to a universally quantified variable followed by two existentially quantified variables. Such existentially quantified variables can be shifted right without changing the truth-value of the sentence. By the assumed bounds on k, each block of universally quantified variables has a block of existentially quantified variables to its right, so we indeed obtain a reduction that preserves the quantifier prefix (in the sense of being Π_k or Σ_k). ◀

▶ Remark. An inspection of the proof of Proposition 9 yields that the hardness results hold on instances Φ where the quantifier-free part Φ_G has the property that $E^{\mathbf{C}[\Phi_G]}$ is symmetric and irreflexive. In the sequel, we will assume that Φ_G has this property. Indeed, one can always replace $E^{\mathbf{C}[\Phi_G]}$ with its symmetric closure, without affecting the truth-value of Φ on

Technical Report AC-TR-15-007

a structure $\mathbf{K}_q$; and note that any instance where this relation is not irreflexive is false on a structure $\mathbf{K}_q$.

5.2 Auxiliary Structures

We call a structure a *labelled digraph* if it is over a signature that consists of a binary relation symbol E and zero or more unary relation symbols; we call a structure a *digraph* if it is over a signature consisting of just a binary relation symbol E. A digraph or labelled digraph is *symmetric* if it interprets E as a symmetric relation. In previous work [5], a way to encode a given labelled digraph $\mathbf{B}$ as a digraph $\mathbf{B}^*$ was given, and is as follows; we refer to [5, Figure 1] for a pictorial illustration. Let $L_1, \ldots, L_n$ denote the unary symbols of the signature of $\mathbf{B}$. For each $b \in B$, define a gadget digraph $\mathbf{G}_b$ which has universe

$$G_b = \{b^s, b^c, b^d, b^{s1}, b^{t1}, b^{s2}, b^{t2}, \ldots, b^{sn}, b^{tn}, b^t\} \cup \{b^{ui} \mid b \in L_i^{\mathbf{B}}\} \cup \{b^{vi} \mid b \in L_i^{\mathbf{B}}\}$$

and edge relation

$$E^{\mathbf{G}_b} = \{(b^c, b^s), (b^c, b^d), (b^s, b^d), (b^d, b^{s1})\} \cup \{(b^{si}, b^{ti}) \mid i \in \{1, \ldots, n\}\} \cup$$

$$\{(b^{ti}, b^{s(i+1)}) \mid i \in \{1, \ldots, n-1\}\} \cup \{(b^{tn}, b^t)\} \cup \{(b^{ui}, b^{si}), (b^{vi}, b^{ti}), (b^{vi}, b^{ui}) \mid b \in L_i^{\mathbf{B}}\}.$$

For a subset $C \subseteq B$, we define $C^* = \bigcup_{b \in C} G_b$; the digraph $\mathbf{B}^*$ has universe B^* and edge relation

$$E^{\mathbf{B}^*} = (\bigcup_{b \in B} E^{\mathbf{G}_b}) \cup \{(b^t, b'^s) \mid (b, b') \in E^{\mathbf{B}}\}.$$

The key feature of this construction is that it preserves homomorphisms.

▶ **Lemma 10.** *(follows from [5, Lemma 17]) Let $\mathbf{A}, \mathbf{B}$ be labelled digraphs over the same signature. There exists a homomorphism $\mathbf{A} \to \mathbf{B}$ if and only if there exists a homomorphism $h : \mathbf{A}^* \to \mathbf{B}^*$; moreover, when the latter condition holds, the image of h is of the form C^* where $C \subseteq B$.*

Tools for understanding the treewidth of structures of the form $\mathbf{B}^*$ are provided in the following lemmas, which relate the treewidth of such a structure to the treewidth of the structure $\mathbf{B}^+$, defined as follows. When $\mathbf{B}$ is a labelled digraph, the structure $\mathbf{B}^+$ has universe $B^+ = \{b^s, b^t \mid b \in B\}$ and edge relation $E^{\mathbf{B}^+} = \{(b^s, b^t) \mid b \in B\} \cup \{(b^t, b'^s) \mid (b, b') \in E^{\mathbf{B}}\}$.

▶ **Lemma 11.** *([5, Lemma 19]) Let $\mathbf{B}$ be a labelled digraph. It holds that $\mathsf{tw}(\mathbf{B}^*) \leq \max(\mathsf{tw}(\mathbf{B}^+), 5)$.*

▶ **Lemma 12.** *Let $\mathbf{B}$ be a labelled digraph. It holds that $\mathsf{tw}(\mathbf{B}^+) \leq \mathsf{tw}(\mathbf{B}^*)$.*

Proof. Given a tree decomposition (T, β) of $G(\mathbf{B}^*)$, a tree decomposition (T, β') of $G(\mathbf{B}^+)$ having lower or equal width can be obtained by defining $\beta'(t) = f(\beta(t))$, where, for each $b \in B$, the mapping f sends $G_b \setminus \{b^t\}$ to b^s, and sends b^t to b^t. Clearly, it holds for each vertex t of T that $|\beta'(t)| \leq |\beta(t)|$. It is straightforward to verify that (T, β') is a tree decomposition of $G(\mathbf{B}^+)$; note that the connectivity condition is satisfied because $G_b \setminus \{b^t\}$ is connected in $G(\mathbf{B}^*)$, for each $b \in B$. ◀

▶ **Lemma 13.** *Let $\mathbf{B}$ be a symmetric labelled digraph. It holds that $\mathsf{tw}(\mathbf{B}) < \mathsf{tw}(\mathbf{B}^+)$.*

Consider an undirected graph on vertex set V. We say that two subsets C, C' of V *touch* if they have a vertex in common or there is an edge between them. A set of mutually touching connected vertex sets is a *bramble*. We say that a subset S of V *covers* a bramble $\mathcal{M}$ if it non-trivially intersects each set in $\mathcal{M}$. The *order* of a bramble $\mathcal{M}$ is the least number of vertices that covers it. We will use the *tree-width duality theorem* [8], which says that, for $k \geq 0$, a graph has tree-width $\geq k$ if and only if it has a bramble of order $> k$.

Proof. We prove that for each bramble $\mathcal{M}$ of $G(\mathbf{B})$, there exists a bramble $\mathcal{M}^+$ of $G(\mathbf{B}^+)$ of strictly higher order, which suffices by the tree-width duality theorem.

When C is a subset of B, we use C^s to denote the set $\{c^s \mid c \in C\}$, and C^t to denote the set $\{c^t \mid c \in C\}$. Let $\mathcal{M} = \{C_1, \dots, C_n\}$ be a bramble of $G(\mathbf{B})$. Define $\mathcal{M}^+$ to be the set system $\{C_1^s, C_1^t\} \cup \bigcup_{i \geq 2, i \in \underline{n}} \{C_i^s \cup (C_i \setminus C_1)^t, (C_i \setminus C_1)^s \cup C_i^t\}$. We claim that $\mathcal{M}^+$ is a bramble of $G(\mathbf{B}^+)$. We demonstrate this by verifying that each pair of distinct sets in $\mathcal{M}^+$ touch. The following cases are exhaustive, up to symmetry; here, i denotes an element of $\underline{n}$ with $i \geq 2$.

- C_1^s, C_1^t. These touch since for any $c_1 \in C_1$, we have $(c_1^s, c_1^t) \in E^{\mathbf{B}^+}$, and so $\{c_1^s, c_1^t\}$ is an edge in $G(\mathbf{B}^+)$.
- $C_i^s \cup (C_i \setminus C_1)^t, (C_i \setminus C_1)^s \cup C_i^t$. These touch since for any $c_i \in C_i$, we have $(c_i^s, c_i^t) \in E^{\mathbf{B}^+}$, and so $\{c_i^s, c_i^t\}$ is an edge in $G(\mathbf{B}^+)$.
- $C_1^s, C_i^s \cup (C_i \setminus C_1)^t$. If $C_1 \cap C_i$ is non-empty, then so is $C_1^s \cap C_i^s$. Otherwise, there is an edge in $G(\mathbf{B})$ between a vertex $c_1 \in C_1$ and a vertex $c_i \in C_i \setminus C_1$, and so $(c_1^t, c_i^s) \in E^{\mathbf{B}^+}$, implying that the two given sets touch in $G(\mathbf{B}^+)$.
- $C_1^s, (C_i \setminus C_1)^s \cup C_i^t$. If $C_1 \cap C_i$ is non-empty, then let $c \in C_1 \cap C_i$; we have $c^s \in C_1^s$, $c^t \in C_i^t$, and as $(c^s, c^t) \in E^{\mathbf{B}^+}$, the edge $\{c^s, c^t\}$ is present in $G(\mathbf{B}^+)$. Otherwise, there exist vertices $c_1 \in C_1$ and $c_i \in C_i \setminus C_1$ that are adjacent in $G(\mathbf{B})$, and so $(c_i^t, c_1^s) \in E^{\mathbf{B}^+}$, implying that $\{c_1^s, c_i^t\}$ is an edge in $G(\mathbf{B}^+)$.

It remains to show that the order of $\mathcal{M}^+$ is strictly higher than that of $\mathcal{M}$. To show this, we prove that for any cover S^+ of $\mathcal{M}^+$, there exists a cover S of $\mathcal{M}$ with $|S| < |S^+|$. Let S^+ be a cover of $\mathcal{M}^+$, and define S to be the subset of B obtained from removing the s, t superscripts from $S^+ \setminus C_1^t$. Since $C_1^t \in \mathcal{M}^+$, the cover S^+ must contain an element of C_1^t, from which it follows that $|S| < |S^+|$. We now verify that S covers $\mathcal{M}$. We have that S covers C_1, since S^+ covers C_1^s. When $i \geq 2$, we have that S covers C_i, since S^+ covers $C_i^s \cup (C_i \setminus C_1)^t$, which implies that S covers $C_i \cup (C_i \setminus C_1) = C_i$. ◀

5.3 Reduction

In this technical section, we prepare the elements for the proof of the main hardness result (Theorem 17 in Section 5). More specifically, Lemma 15 implements a polynomial-time mapping of the source problem to the target problem, whereas Lemma 14 and Lemma 16 assist in proving its correctness.

Let $\forall y_1 \dots \forall y_m \exists x_1 \dots \exists x_n \phi_G$ be an instance of Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$. Relative to this instance, we define the following objects.

- Let τ be the signature $\{E\} \cup \{U_{y_1}, \dots, U_{y_m}\} \cup \{U_1, \dots, U_q\}$, where the U_{y_i} and the U_j are unary relation symbols.
- We define the following formulas of signature τ.

$$\phi_K = (\bigwedge_{i \in \underline{q}} U_i(i)) \wedge (\bigwedge_{i,j \in \underline{q}, i \neq j} E(i,j))$$

For each $i \in \underline{m}$, $j \in \underline{q}$,

$$\lambda_{y_i \to j} = U_{y_i}(j)$$
$$\phi_{y_i \to j} = \lambda_{y_i \to j} \wedge \bigwedge_{k \in \underline{q}, k \neq j} (E(y_i, k) \wedge E(k, y_i))$$

For each $f : \{y_1, \ldots, y_m\} \to \underline{q}$,

$$\lambda_f = \bigwedge_{i \in \underline{m}} \lambda_{y_i \to f(y_i)}$$
$$\phi_f = \bigwedge_{i \in \underline{m}} \phi_{y_i \to f(y_i)}$$

Observe that, for each mapping $f : \{y_1, \ldots, y_m\} \to \underline{q}$,

$$\phi_f = \lambda_f \wedge \bigwedge_{i \in \underline{m}, k \in \underline{q}, f(y_i) \neq k} (E(y_i, k) \wedge E(k, y_i)), \tag{†}$$

up to a permutation of the conjuncts. In the sequel, we formally view ϕ_G as a formula of signature τ, so that, for instance, $\phi_G \wedge \phi_K \wedge \phi_f$ is a formula of signature τ, and $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$ is a structure of signature τ.

▶ **Lemma 14.** *Let $\forall y_1 \ldots \forall y_m \exists x_1 \ldots \exists x_n \phi_G$ be an instance of Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$. If a mapping $f : \{y_1, \ldots, y_m\} \to \underline{q}$ has an extension $f' : \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to \underline{q}$ such that $\mathbf{K}_q, f' \models \phi_G$, then the following hold.*
1. $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*$ maps homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_f]^*$.
2. If $q \geq 5$, then $\mathsf{tw}(\mathbf{C}[\phi_K \wedge \lambda_f]^*) \leq q$.

Proof. For the first part, it is sufficient to prove that $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$ maps homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_f]$; the statement then follows by Lemma 10. Note that the universes of the structures are $C[\phi_K \wedge \lambda_f] = \underline{q}$ and $C[\phi_G \wedge \phi_K \wedge \phi_f] = \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \cup \underline{q}$.

Let $f' : \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to \underline{q}$ be an extension of $f : \{y_1, \ldots, y_m\} \to \underline{q}$ such that $\mathbf{K}_q, f' \models \phi_G$. Let $h \colon \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \cup \underline{q} \to \underline{q}$ be the extension of f' defined by $h(j) = j$ for all $j \in \underline{q}$. We claim that h is a homomorphism from $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$ to $\mathbf{C}[\phi_K \wedge \lambda_f]$. By hypothesis, h maps homomorphically $\mathbf{C}[\phi_G]$ into $\mathbf{C}[\phi_K]$. Clearly, h maps homomorphically $\mathbf{C}[\phi_K \wedge \lambda_f]$ into $\mathbf{C}[\phi_K \wedge \lambda_f]$. By (†), it suffices to show that h is a homomorphism from $\mathbf{C}[\bigwedge_{i \in \underline{m}, k \in \underline{q}, f(y_i) \neq k}(E(y_i, k) \wedge E(k, y_i))]$ to $\mathbf{C}[\phi_K]$. Suppose that the tuples $(y_i, k), (k, y_i)$ occur in the first structure. Then, by definition, $f(y_i) \neq k$. Therefore $(h(y_i), h(k)) = (f(y_i), k) \in E^{\mathbf{C}[\phi_K]}$, and $(h(k), h(y_i)) = (k, f(y_i)) \in E^{\mathbf{C}[\phi_K]}$.

For the second part, assume $q \geq 5$. Then, by Lemma 11, it is sufficient to prove that $\mathsf{tw}(\mathbf{C}[\phi_K \wedge \lambda_f]^+) \leq q$. We establish $\mathsf{tw}(\mathbf{C}[\phi_K \wedge \lambda_f]^+) \leq q$ by providing a tree decomposition of width q of $\mathbf{C}[\phi_K \wedge \lambda_f]^+$. It is straightforward to check that a path of q vertices $v_1, \ldots, v_q$, where the bag on v_j is $\{k^s \mid k \in \underline{q}\} \cup \{j^t\}$ for all $j \in \underline{q}$, gives the required tree decomposition. ◀

▶ **Lemma 15.** *There exists a polynomial-time algorithm that, given an instance*

$$\phi = \forall y_1 \ldots \forall y_m \exists x_1 \ldots \exists x_n \phi_G$$

of Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$, computes two sentences $\phi', \phi'' \in \mathrm{EP}_{\{E\}}$, where E is a binary relation symbol, such that ϕ' is logically equivalent to the disjunctive form

$$\bigvee_{f : \{y_1, \ldots, y_m\} \to \underline{q}} Q[\mathbf{C}[\phi_K \wedge \lambda_f]^*], \tag{F1}$$

ϕ'' is logically equivalent to the disjunctive form

$$\bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*], \tag{F2}$$

and the following hold:

1. *The disjunctive forms (F1) and (F2) are irredundant.*
2. *For all $f : \{y_1, \ldots, y_m\} \to \underline{q}$, $\mathbf{C}[\phi_K \wedge \lambda_f]^*$ maps homomorphically to $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*$; and consequently, $\phi'' \models \phi'$.*
3. *If $\mathbf{K}_q \models \phi$, then ϕ' and ϕ'' are logically equivalent.*

Proof. Let $\phi = \forall y_1 \ldots \forall y_m \exists x_1 \ldots \exists x_n \phi_G$ be an instance of Π_2-QCSP($\mathbf{K}_q$). The algorithm, given ϕ, constructs in polynomial-time the existential positive sentences

$$\phi' = Q[\mathbf{C}[\phi_K]^*] \wedge \bigwedge_{i\in\underline{m}} \bigvee_{j\in\underline{q}} Q[\mathbf{C}[\lambda_{y_i\to j}]^*];$$

$$\phi'' = Q[\mathbf{C}[\phi_G]^*] \wedge Q[\mathbf{C}[\phi_K]^*] \wedge \bigwedge_{i\in\underline{m}} \bigvee_{j\in\underline{q}} Q[\mathbf{C}[\phi_{y_i\to j}]^*].$$

We claim that:

$$\phi' \equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\phi_K \wedge \lambda_f]^*]]; \tag{G1}$$

$$\phi'' \equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*]. \tag{G2}$$

It is sufficient to observe the following logical equivalences. For (G1),

$$\begin{aligned}
\phi' &= Q[\mathbf{C}[\phi_K]^*] \wedge \bigwedge_{i\in\underline{m}} \bigvee_{j\in\underline{q}} Q[\mathbf{C}[\lambda_{y_i\to j}]^*] \\
&\equiv Q[\mathbf{C}[\phi_K]^*] \wedge \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\lambda_f]^*] \\
&\equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} (Q[\mathbf{C}[\phi_K]^*] \wedge Q[\mathbf{C}[\lambda_f]^*]) \\
&\equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[(\mathbf{C}[\phi_K] \cup \mathbf{C}[\lambda_f])^*] \\
&\equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\phi_K \wedge \lambda_f]^*].
\end{aligned}$$

For (G2), we similarly have

$$\begin{aligned}
\phi'' &= Q[\mathbf{C}[\phi_G]^*] \wedge Q[\mathbf{C}[\phi_K]^*] \wedge \bigwedge_{i\in\underline{m}} \bigvee_{j\in\underline{q}} Q[\mathbf{C}[\phi_{y_i\to j}]^*] \\
&\equiv Q[\mathbf{C}[\phi_G]^*] \wedge Q[\mathbf{C}[\phi_K]^*] \wedge \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\phi_f]^*] \\
&\equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} (Q[\mathbf{C}[\phi_G]^*] \wedge Q[\mathbf{C}[\phi_K]^*] \wedge Q[\mathbf{C}[\phi_f]^*]) \\
&\equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[(\mathbf{C}[\phi_G] \cup \mathbf{C}[\phi_K] \cup \mathbf{C}[\phi_f])^*] \\
&\equiv \bigvee_{f:\{y_1,\ldots,y_m\}\to\underline{q}} Q[\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*].
\end{aligned}$$

To prove the stated properties, we observe preliminarily that ϕ_f contains all conjuncts of λ_f by (†), thus $\mathbf{C}[\phi_K \wedge \lambda_f]$ is a substructure of $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$.

We prove the first property. By Proposition 1, it is sufficient to check that if f and g are distinct mappings from $\{y_1, \dots, y_m\}$ to $\underline{q}$, then $\mathbf{C}[\phi_K \wedge \lambda_f]$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_g]$. By Lemma 10, this implies that $\mathbf{C}[\phi_K \wedge \lambda_f]^*$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_g]^*$, which settles irredundancy of (F1); in turn it follows, by the substructure observation, that $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*$ does not map homomorphically to $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_g]^*$, which settles irredundancy of (F2). Assume for a contradiction that h maps $\mathbf{C}[\phi_K \wedge \lambda_f]$ homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_g]$. By definition of ϕ_K, it holds that $U_i^{\mathbf{C}[\phi_K]} = \{i\}$ for all $i \in \underline{q}$, therefore h acts identically on $\underline{q}$. Let $j \in \underline{m}$ be such that $f(y_j) = k \neq k' = g(y_j)$. By definition of λ_f and λ_g, it holds that $U_{y_j}^{\mathbf{C}[\phi_f]} = \{k\}$ and $U_{y_j}^{\mathbf{C}[\phi_g]} = \{k'\}$. Therefore, $h(k) = k'$, a contradiction.

We prove the second property. It suffices to prove the first part; that $\phi'' \models \phi'$ is then a consequence by appeal to Proposition 1. Let f be any mapping from $\{y_1, \dots, y_m\}$ to $\underline{q}$. By the observation that $\mathbf{C}[\phi_K \wedge \lambda_f]$ is a substructure of $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$, we have that $\mathbf{C}[\phi_K \wedge \lambda_f]$ maps homomorphically to $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$; the statement then follows by Lemma 10.

We prove the third property. Assume $\mathbf{K}_q \models \phi$. Let f be any mapping of $\{y_1, \dots, y_m\}$ to $\underline{q}$. Then, there exists an extension $f' \colon \{y_1, \dots, y_m, x_1, \dots, x_n\} \to \underline{q}$ of f such that $\mathbf{K}_q, f' \models \phi_G$. Then, by Lemma 14(1), $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*$ maps homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_f]^*$, which implies that $Q[\mathbf{C}[\phi_K \wedge \lambda_f]^*] \models Q[\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*]$. Therefore, by Lemma 4(1), $\phi' \models \phi''$. Then, by the second property proved above, $\phi' \equiv \phi''$. ◀

▶ **Lemma 16.** *Let $\forall y_1 \dots \forall y_m \exists x_1 \dots \exists x_n \phi_G$ be an instance of Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$. Let $f : \{y_1, \dots, y_m\} \to \underline{q}$ be a mapping, and suppose that $\mathsf{tw}(\mathsf{core}(\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*)) \leq q$. Then, f has an extension $f' : \{y_1, \dots, y_m, x_1, \dots, x_n\} \to \underline{q}$ such that $\mathbf{K}_q, f' \models \phi_G$.*

In the proof, we will use the following notation: when $\mathbf{B}$ is a structure on signature σ, and $\sigma' \subseteq \sigma$, use $\mathsf{red}_{\sigma'}(\mathbf{B})$ to denote the *reduct* of $\mathbf{B}$ on σ', that is, the structure on σ' naturally obtained from $\mathbf{B}$ by forgetting the interpretations of the symbols not in σ'.

Proof. Set $\mathbf{A} = \mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$. Since each core of $\mathbf{A}^*$ is the image of an endomorphism of $\mathbf{A}^*$, then by Lemma 10, each core of $\mathbf{A}^*$ has universe of the form S^* where $S \subseteq A$. Let $S \subseteq A$ be a subset with this property, and let $\mathbf{S}$ be the substructure of $\mathbf{A}$ induced on S. By assumption, $\mathsf{tw}(\mathbf{S}^*) \leq q$. By Lemma 12, $\mathsf{tw}(\mathbf{S}^+) \leq q$. By Lemma 13, $\mathsf{tw}(\mathbf{S}) < q$. It follows that $\mathsf{red}_E(\mathbf{S})$ has a homomorphism to $\mathbf{K}_q$ because, by the remark in Section 5.1 and the construction, $\mathsf{red}_E(\mathbf{S})$ is irreflexive. Since $\mathbf{A}$ has a homomorphism to $\mathbf{S}$, we have that $\mathsf{red}_E(\mathbf{A})$ has a homomorphism to $\mathsf{red}_E(\mathbf{S})$, and by transitivity of the homomorphism relation, we have that $\mathsf{red}_E(\mathbf{A})$ has a homomorphism h to $\mathbf{K}_q$. Observe that $\mathsf{red}_E(\mathbf{A}) = \mathbf{C}[\phi_G \wedge (\bigwedge_{i,j \in \underline{q}, i \neq j} E(i,j)) \wedge \bigwedge_{i \in \underline{m}, k \in \underline{q}, f(y_i) \neq k} (E(y_i,k) \wedge E(k,y_i))]$.

By relabelling the elements of $\mathbf{K}_q$ if necessary, it can be assumed that h is the identity map on $\underline{q}$. Therefore, since $\mathbf{K}_q, h \models \bigwedge_{i \in \underline{m}, k \in \underline{q}, f(y_i) \neq k} (E(y_i,k) \wedge E(k,y_i))$, we have that h is an extension of f. Indeed, for all $i \in \underline{m}$ and $k \in \underline{q}$ such that $f(y_i) \neq k$, we have $(h(y_i), h(k)) = (h(y_i), k) \in E^{\mathbf{K}_q}$, which implies $h(y_i) \neq k$.

Since $\mathbf{K}_q, h \models \phi_G$, we obtain the result. ◀

5.4 Hardness Results

▶ **Theorem 17.** *Let σ be a signature that contains a relation symbol E of binary arity. For each $k \geq 6$, the problem EP^k_σ-$\textsc{Expr}$ is Π^p_2-hard.*

Proof. Assume $q \geq 5$. We show that there is a reduction from Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$ to EP_σ^{q+1}-Expr, where $\sigma = \{E\}$ and E is a binary relation symbol; this suffices by Proposition 9.

Let $\phi = \forall y_1 \ldots \forall y_m \exists x_1 \ldots \exists x_n \phi_G$ be an instance of Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$. The reduction uses the algorithm in Lemma 15 to compute in polynomial-time the sentence $\phi'' \in \mathrm{EP}_\sigma$ defined there. We prove that $\mathbf{K}_q \models \phi$ if and only if ϕ'' is logically equivalent to a sentence in EP_σ^{q+1}.

Assume that $\mathbf{K}_q \models \phi$. By Lemma 15(3), we have that ϕ'' is logically equivalent to ϕ'. Now look at the formula shown to be logically equivalent to ϕ' in that lemma (Lemma 15). For each $f : \{y_1, \ldots, y_m\} \to \underline{q}$, by Lemma 14(2) and the assumption that $q \geq 5$, we have $\mathsf{tw}(\mathbf{C}[\phi_K \wedge \lambda_f]^*) \leq q$, and therefore by [7, Theorem 5], $Q[\mathbf{C}[\phi_K \wedge \lambda_f]^*]$ is logically equivalent to a primitive positive sentence in PP^{q+1}. Therefore, ϕ'' is logically equivalent to a sentence in EP_σ^{q+1}.

Assume that $\mathbf{K}_q \not\models \phi$. Let $f \colon \{y_1, \ldots, y_m\} \to \underline{q}$ be a mapping such that for all mappings $f' \colon \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to \underline{q}$ extending f it holds that $\mathbf{K}_q, f' \not\models \phi_G$. Then, by Lemma 16, $\mathsf{tw}(\mathsf{core}(\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*)) > q$. Therefore, by [7, Theorem 5], $Q[\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*]$ is not logically equivalent to a primitive positive sentence in PP^{q+1}. Since, by Lemma 15(1), the disjunctive form in (F2) is irredundant, so by Lemma 4(3), ϕ'' is a "No" instance of EP_σ^{q+1}-Expr. ◀

▶ **Theorem 18.** *Let σ be a signature that contains a relation symbol E of binary arity. The problems EP_σ-Entail and EP_σ-$\equiv$ are Π_2^p-hard.*

Proof. By Proposition 9, it is sufficient to show that there are reductions from Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$ to EP_σ-Entail and from Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$ to EP_σ-$\equiv$, where $\sigma = \{E\}$ and E is a binary relation symbol.

Let $\phi = \forall y_1 \ldots \forall y_m \exists x_1 \ldots \exists x_n \phi_G$ be an instance of Π_2-$\mathsf{QCSP}(\mathbf{K}_q)$. We give one reduction that works for both problems. The reduction uses the algorithm in Lemma 15 to compute in polynomial time the pair of existential positive sentences (ϕ', ϕ'') specified there. We prove that the reduction is correct for both EP_σ-Entail and EP_σ-$\equiv$.

Assume that $\mathbf{K}_q \models \phi$. By Lemma 15(3), $\phi' \equiv \phi''$. Now assume that $\mathbf{K}_q \not\models \phi$. Let $f \colon \{y_1, \ldots, y_m\} \to \underline{q}$ be such that for all extensions $f' \colon \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to \underline{q}$ of f it holds that $\mathbf{K}_q, f' \not\models \phi_G$. Then $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_f]$, and by Lemma 10, $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_f]^*$. Moreover, for all $g \colon \{y_1, \ldots, y_m\} \to \underline{q}$ distinct from f, $\mathbf{C}[\phi_K \wedge \lambda_f]$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_g]$ by Lemma 15(1), therefore $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_g]$, so that by Lemma 10 again, $\mathbf{C}[\phi_G \wedge \phi_K \wedge \phi_f]^*$ does not map homomorphically to $\mathbf{C}[\phi_K \wedge \lambda_g]^*$. Therefore, by Lemma 4(1), $\phi' \not\models \phi''$. ◀

6 The Higher-Arity Case

In this section, E is a binary relation symbol and σ is a signature that contains a relation symbol of at least binary arity. We fix a relational symbol $R \in \sigma$ of arity $r \geq 2$, and relative to this choice, we define the following objects.

- For every sentence $\phi \in \mathrm{EP}_{\{E\}}$, let $\phi_{E \rightsquigarrow R}$ denote the sentence in EP_σ obtained by replacing in ϕ subformulas of the form $E(x, y)$ by $R(x, y, \ldots, y)$.
- For every sentence $\phi \in \mathrm{EP}_\sigma$, let $\phi_{R \rightsquigarrow E}$ denote the sentence in $\mathrm{EP}_{\{E\}}$ obtained by replacing in ϕ subformulas of the form $R(x, y, \ldots, y)$ by $E(x, y)$.
- For every structure $\mathbf{A}$ over $\{E\}$, let $\mathbf{A}_{E \rightsquigarrow R}$ denote the structure over σ, with universe A, obtained by interpreting R over $\{(a, a', a', \ldots, a') \mid (a, a') \in E^{\mathbf{A}}\}$ and every $S \in \sigma \setminus \{R\}$ over the empty set.

- For every structure $\mathbf{A}$ over σ, let $\mathbf{A}_{R \rightsquigarrow E}$ denote the structure over $\{E\}$, with universe A, obtained by interpreting E over $\{(a, a') \mid (a, a', a', \ldots, a') \in R^{\mathbf{A}}\}$.

Proof of Higher-Arity Case of Theorem 7

Let $k \geq 6$. We prove that the problem EP^k_σ-Expr is Π^p_2-hard. By Theorem 17, it is sufficient to give a reduction from $\text{EP}^k_{\{E\}}$-Expr, where E is a binary relation symbol, to EP^k_σ-Expr.

The reduction, given an instance ϕ of $\text{EP}^k_{\{E\}}$-Expr, returns the instance $\phi_{E \rightsquigarrow R}$ of EP^k_σ-Expr. We check that ϕ is logically equivalent to an existential positive sentence in EP^k if and only if $\phi_{E \rightsquigarrow R}$ is logically equivalent to an existential positive sentence in EP^k.

If ϕ is logically equivalent to an existential positive sentence in EP^k, then let $\bigvee_{i \in \underline{n}} \phi_i$ be an existential positive sentence in disjunctive form logically equivalent to ϕ such that $\phi \in \text{PP}^k$ for all $i \in \underline{n}$; such sentence exists by Lemma 4(3). Then, for every structure $\mathbf{A}$ over the signature σ,

$$\begin{aligned}
\mathbf{A} \models \phi_{E \rightsquigarrow R} & \text{ if and only if } \mathbf{A}_{R \rightsquigarrow E} \models \phi \\
& \text{ if and only if } \mathbf{A}_{R \rightsquigarrow E} \models \bigvee_{i \in \underline{n}} \phi_i \\
& \text{ if and only if } (\mathbf{A}_{R \rightsquigarrow E})_{E \rightsquigarrow R} \models \bigvee_{i \in \underline{n}} (\phi_i)_{E \rightsquigarrow R} \\
& \text{ if and only if } \mathbf{A} \models \bigvee_{i \in \underline{n}} (\phi_i)_{E \rightsquigarrow R}.
\end{aligned}$$

Therefore, $\phi_{E \rightsquigarrow R}$ is logically equivalent to a sentence in EP^k.

Conversely, assume that $\phi_{E \rightsquigarrow R}$ is logically equivalent to an existential positive sentence in EP^k. We claim that $\phi_{E \rightsquigarrow R}$ is logically equivalent to an existential positive sentence of the form

$$\bigvee_{i \in \underline{n}} (\chi_i)_{E \rightsquigarrow R},$$

where, χ_i in $\text{PP}^k_{\{E\}}$ for all $i \in \underline{n}$. Now, let $\mathbf{A}$ be any structure over the signature $\{E\}$. Then,

$$\begin{aligned}
\mathbf{A} \models \phi & \text{ if and only if } \mathbf{A}_{E \rightsquigarrow R} \models \phi_{E \rightsquigarrow R} \\
& \text{ if and only if } \mathbf{A}_{E \rightsquigarrow R} \models \bigvee_{i \in \underline{n}} (\chi_i)_{E \rightsquigarrow R} \\
& \text{ if and only if } (\mathbf{A}_{E \rightsquigarrow R})_{R \rightsquigarrow E} \models \bigvee_{i \in \underline{n}} \chi_i \\
& \text{ if and only if } \mathbf{A} \models \bigvee_{i \in \underline{n}} \chi_i.
\end{aligned}$$

Therefore, ϕ is logically equivalent to a sentence in EP^k.

We prove the claim. Let $\phi_{E \rightsquigarrow R}$ be given. Using (E1)-(E4), we compute an irredundant disjunctive form logically equivalent to $\phi_{E \rightsquigarrow R}$; by construction, such sentence has the form $\bigvee_{i \in \underline{n}} (\phi_i)_{E \rightsquigarrow R}$, where ϕ_i is a sentence in $\text{PP}_{\{E\}}$ for all $i \in \underline{n}$. Let $i \in \underline{n}$. By Lemma 4(3), $(\phi_i)_{E \rightsquigarrow R}$ is logically equivalent to a sentence in PP^k_σ. Then, by [7, Theorem 5], $\mathsf{tw}(\mathsf{core}(\mathbf{C}[(\phi_i)_{E \rightsquigarrow R}])) < k$. Note that the canonical query of $\mathsf{core}(\mathbf{C}[(\phi_i)_{E \rightsquigarrow R}])$ has the form $(\psi_i)_{E \rightsquigarrow R}$ for some sentence ψ_i in $\text{PP}_{\{E\}}$. Then, by [7, Theorem 7], $(\psi_i)_{E \rightsquigarrow R}$ is logically equivalent to a sentence in PP^k_σ of the form $(\chi_i)_{E \rightsquigarrow R}$ for some $\chi_i \in \text{PP}^k_{\{E\}}$; this is because the syntactic replacements used to derive $(\chi_i)_{E \rightsquigarrow R}$ from $(\psi_i)_{E \rightsquigarrow R}$ are: $\exists x(\theta \wedge \theta') \equiv \exists x \theta \wedge \theta'$, if x is not free in θ'; $\exists x\theta \equiv \exists y\theta[x/y]$, if y is not free in θ and x is free for y in θ, where, $\theta[x/y]$ is obtained by replacing free occurrences of x in θ by y.

Proof of Higher-Arity Case of Theorem 8

We prove that the problems EP_σ-Entail and EP_σ-$\equiv$ are Π_2^p-hard. By Theorem 18, it is sufficient to give a reduction from $\mathrm{EP}_{\{E\}}$-Entail to EP_σ-Entail, and from $\mathrm{EP}_{\{E\}}$-$\equiv$ to EP_σ-Entail.

We give one reduction that works for both problems. The reduction, given an instance (ϕ, ψ) of $\mathrm{EP}_{\{E\}}$-Entail (respectively, $\mathrm{EP}_{\{E\}}$-$\equiv$), returns the instance $(\phi_{E\rightsquigarrow R}, \psi_{E\rightsquigarrow R})$ of EP_σ-Entail (respectively, EP_σ-$\equiv$).

We check correctness. It is sufficient to prove that $\phi \models \psi$ if and only if $\phi_{E\rightsquigarrow R} \models \psi_{E\rightsquigarrow R}$. Let $\mathbf{A}$ be any structure over $\{E\}$. Then,

$$\begin{aligned}\mathbf{A} \models \phi \text{ implies } & \mathbf{A}_{E\rightsquigarrow R} \models \phi_{E\rightsquigarrow R}\\ \text{implies } & \mathbf{A}_{E\rightsquigarrow R} \models \psi_{E\rightsquigarrow R}\\ \text{implies } & \mathbf{A} \models \psi.\end{aligned}$$

Conversely, let $\mathbf{A}$ be any structure over σ. Then,

$$\begin{aligned}\mathbf{A} \models \phi_{E\rightsquigarrow R} \text{ implies } & \mathbf{A}_{R\rightsquigarrow E} \models (\phi_{E\rightsquigarrow R})_{R\rightsquigarrow E}\\ \text{implies } & \mathbf{A}_{R\rightsquigarrow E} \models \phi\\ \text{implies } & \mathbf{A}_{R\rightsquigarrow E} \models \psi\\ \text{implies } & \mathbf{A} \models \psi_{E\rightsquigarrow R},\end{aligned}$$

which settles the theorem.

7 The Unary Case

7.1 Infinite Unary Signature

▶ **Theorem 19.** *Let σ be a signature that contains infinitely many unary relation symbols. The problems EP_σ-Entail and EP_σ-$\equiv$ are Π_2^p-hard.*

Proof. We give a single reduction to both problems from the Π_2^p-complete problem Π_2^p-QBF, which is the problem of deciding, given a sentence in propositional logic consisting of a Π_2^p quantifier prefix followed by a quantifier-free formula, whether or not the sentence is true.

Let $\Phi = \forall y_1 \dots \forall y_m \exists x_1 \dots \exists x_n \phi$ be an instance of Π_2^p-QBF. Let $S^0, S^1, S_1^0, S_1^1, \dots, S_m^0, S_m^1$ be pairwise distinct unary relation symbols in σ. Define

$$\alpha = \bigvee_{i \in \underline{m}} (\exists z S_i^0(z) \wedge \exists z S_i^1(z))$$

and

$$\beta = (\exists z S^0(z)) \wedge (\exists z S^1(z)) \wedge \bigwedge_{i \in \underline{m}} (\exists z S_i^0(z) \vee \exists z S_i^1(z)).$$

Let ϕ^* be the quantifier-free existential positive formula on σ obtained from ϕ by pushing negations to the variable level, and mapping the resulting propositional literals according to

$$y_i \mapsto S_i^1(y_i), \neg y_i \mapsto S_i^0(y_i) \text{ and } x_j \mapsto S^1(x_j), \neg x_j \mapsto S^0(x_j)$$

for all $i \in \underline{m}$ and $j \in \underline{n}$. The reduction outputs the pair (ψ, ψ') whose components are defined by

$$\psi = \alpha \vee \beta \text{ and } \psi' = \alpha \vee (\beta \wedge \exists y_1 \ldots \exists y_m \exists x_1 \ldots \exists x_n \phi^*).$$

It holds that $\psi' \models \psi$. We claim that $\psi \models \psi'$ if and only if Φ is true, which suffices to give the theorem.

($\Rightarrow$) Let $f : \{y_1, \ldots, y_m\} \to \{0,1\}$ be any assignment. Let $\mathbf{B}_f$ be a structure such that $(S^0)^{\mathbf{B}_f} = \{0\}$, $(S^1)^{\mathbf{B}_f} = \{1\}$, and such that $S_i^j = \{j\}$ if $f(y_i) = j$, and $S_i^j = \emptyset$ otherwise (for all $i \in \underline{m}, j \in \{0,1\}$). We have $\mathbf{B}_f \models \neg\alpha \wedge \beta$; by the assumption that $\psi \models \psi'$, it follows that $\mathbf{B}_f \models \exists y_1 \ldots \exists y_m \exists x_1 \ldots \exists x_n \phi^*$. By the definitions of ϕ^* and $\mathbf{B}_f$, we have that $\mathbf{B}_f, f \models \exists x_1 \ldots \exists x_n \phi^*$. This implies that there is an extension $f' : \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to \{0,1\}$ of f such that $\mathbf{B}_f, f' \models \phi^*$, and it follows from the definitions of ϕ^* and $\mathbf{B}_f$ that f' satisfies ϕ.

($\Leftarrow$) Let $\mathbf{B}$ be an arbitrary structure on σ. If there exists $i \in \underline{m}$ such that both $(S_i^0)^{\mathbf{B}}$ and $(S_i^1)^{\mathbf{B}}$ are non-empty, then $\mathbf{B} \models \alpha$ and thus $\mathbf{B} \models \psi$ and $\mathbf{B} \models \psi'$. Otherwise, it holds that $\mathbf{B} \not\models \alpha$, which we assume for the rest of the proof.

If there exists $i \in \underline{m}$ such that $(S_i^0)^{\mathbf{B}} = (S_i^1)^{\mathbf{B}} = \emptyset$, then $\mathbf{B} \not\models \beta$ and both $\mathbf{B} \not\models \psi$ and $\mathbf{B} \not\models \psi'$ hold. Otherwise, for all $i \in \underline{m}$, exactly one of $(S_i^0)^{\mathbf{B}}$ and $(S_i^1)^{\mathbf{B}}$ is non-empty; define $f : \{y_1, \ldots, y_m\} \to \{0,1\}$ so that $f(y_i) = j$ if and only if $(S_i^j)^{\mathbf{B}}$ is non-empty. Assume that $\mathbf{B} \models \psi$. Then $\mathbf{B} \models \beta$, implying that $(S^0)^{\mathbf{B}}$ and $(S^1)^{\mathbf{B}}$ are both non-empty. Since by assumption Φ is true, the assignment f has an extension $f' : \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to \{0,1\}$ satisfying ϕ. Define an assignment $f^* : \{y_1, \ldots, y_m, x_1, \ldots, x_n\} \to B$ where $f'(x_i) = j$ implies $f^*(x_i) \in (S^j)^{\mathbf{B}}$ (for all $i \in \underline{n}$), and where $f'(y_i) = j$ implies $f^*(y_i) \in (S_i^j)^{\mathbf{B}}$ (for all $i \in \underline{m}$); the latter property can be enforced due to the definition of f and the fact that f' extends f. We have $\mathbf{B}, f^* \models \phi^*$ and thus $\mathbf{B} \models (\beta \wedge \exists y_1 \ldots \exists y_m \exists x_1 \ldots \exists x_n \phi^*)$, implying that $\mathbf{B} \models \psi'$. ◀

7.2 Finite Unary Signature

Let us suppose throughout this section that σ is a signature that consists of finitely many unary relation symbols.

We first describe complexity upper bounds, focusing on the entailment problem (this is justified by appeal to Proposition 6). Relative to a signature σ, let us define a *profile* as a subset of $\wp(\sigma)$ (the power set of σ). When $\mathbf{B}$ is a structure on signature σ and $b \in B$, define $\sigma_{\mathbf{B},b}$ as the set $\{U \in \sigma \mid b \in U^{\mathbf{B}}\}$. Let us define the *profile of* $\mathbf{B}$ as the set $\{\sigma_{\mathbf{B},b} \mid b \in B\}$.

We show that deciding whether or not a structure $\mathbf{B}$ satisfies an EP formula depends only on the profile of $\mathbf{B}$. Say that a profile P' dominates a profile P if for each element $S \in P$, there exists a superset S' of S that is contained in P'. Observe that the profile of $\mathbf{B}'$ *dominates* that of $\mathbf{B}$ if there exists a function $f : B \to B'$ such that, for all $b \in B$, it holds that $\sigma_{\mathbf{B},b} \subseteq \sigma_{\mathbf{B}',f(b)}$.

▶ **Lemma 20.** *Suppose that σ consists of finitely many unary relation symbols. Let $\mathbf{B}$ and $\mathbf{B}'$ be structures over σ. If the profile of $\mathbf{B}'$ dominates that of $\mathbf{B}$ via $f : B \to B'$, then for each existential positive formula ϕ and any assignment $h : V \to B$ defined on the variables over which formulas are written,*

$$\mathbf{B}, h \models \phi \text{ implies } \mathbf{B}', f(h) \models \phi$$

Proof. This is proved straightforwardly by induction on the structure of the formula ϕ. ◀

In particular, observe that, if two structures have the same profile, then for any existential positive sentence ϕ, one of the structures satisfies ϕ if and only if the other one does. For a fixed signature σ there are finitely many profiles, and each is realized by a structure of size at most $\wp(\sigma)$. We thus obtain the following.

▶ Proposition 21. Suppose that σ consists of finitely many unary relation symbols. The problem EP_σ-ENTAIL is in P^{NP} via a polynomial-time algorithm that makes a constant number of queries (in parallel) to an NP oracle.

Indeed, we obtain that the number of queries needed is bounded by two times the number of profiles (as, for a particular structure $\mathbf{B}$, checking that $\mathbf{B} \models \phi$ implies $\mathbf{B} \models \psi$ requires two NP queries). One can, however, exhibit tighter bounds in terms of the number of queries required. For instance, let P be a profile and let P' be the subset of P that contains a subset $S \in P$ if and only if no strict superset of S appears in P. The profiles P and P' dominate each other, and thus by the lemma above, checking if a structure with profile P satisfies a sentence is equivalent to checking if a structure with profile P' satisfies the sentence. The profile P' is an *antichain* in the sense that no two of its elements are comparable (in the $\subseteq$ ordering). We have thus argued that, on an instance (ϕ, ψ) of EP_σ-ENTAIL, one needs only to check that, for each profile P that is an antichain, $\mathbf{B}_P \models \phi$ implies $\mathbf{B}_P \models \psi$, where $\mathbf{B}_P$ is a structure whose profile is P; this can be carried out in polynomial time with $2 \cdot A$ queries to an NP oracle, where A denotes the number of profiles that are antichains. Clearly, for every non-empty unary signature, the number of profiles that are antichains is strictly less than the number of profiles; for instance, $\{\{U_1\}, \{U_1, U_2\}\}$ and $\{\{U_1\}\}$ are profiles of $\sigma = \{U_1, U_2\}$, but the first is not an antichain as $\{U_1\} \subset \{U_1, U_2\}$.

We now turn to discuss complexity lower bounds. As usual, let SAT denote the problem of deciding, given a propositional formula ϕ, whether or not ϕ is satisfiable. Recall that DP is the complexity class that contains a language if it is equal to the intersection of an NP language and a coNP language. The problem SAT-UNSAT, defined as $\{(\phi, \phi') \mid \phi \in \mathrm{SAT} \text{ and } \phi' \notin \mathrm{SAT}\}$, is known to be DP-complete [10, Theorem17.1]. Let SIS (short for "SAT implies SAT") be defined as the problem $\{(\phi, \phi') \mid \phi \in \mathrm{SAT} \text{ implies } \phi' \in \mathrm{SAT}\}$. We have that SIS is the complement of SAT-UNSAT, in the sense that, by definition, a pair (ϕ, ϕ') of formulas is in SIS if and only if it is not in SAT-UNSAT; hence, the problem SIS is coDP-complete. For $M \geq 2$, let M-SIS denote the problem $\{(x_1, \ldots, x_m) \mid x_1, \ldots, x_m \in \mathrm{SIS}\}$. We will exhibit reductions from these problems.

Let us say that a set $T \subseteq \wp(\sigma)$ containing antichains is a *system of profiles* if it does not contain two distinct profiles P, P' such that P dominates P'; and, for each $P \in T$, it holds that $|P| > 1$.

▶ **Example 22.** Suppose that $|\sigma| \geq 2$; for concreteness, suppose that $\sigma \supseteq \{U_1, U_2\}$. Then, the one-element set $\{\{\{U_1\}, \{U_2\}\}\}$ is a system of profiles. Moreover, in the case that $\sigma = \{U_1, U_2\}$, it can be verified that this is the only non-empty system of profiles, since $\{\{U_1\}, \{U_2\}\}$ is the only antichain in $\wp(\sigma)$ of size strictly greater than 1. Now suppose that $|\sigma| \geq 3$; for concreteness, suppose that $\sigma \supseteq \{U_1, U_2, U_3\}$. Then, the three-element set $\{\{\{U_1\}, \{U_2, U_3\}\}, \{\{U_2\}, \{U_1, U_3\}\}, \{\{U_3\}, \{U_1, U_2\}\}\}$ is a system of profiles.

For a signature σ, let M_σ denote the maximum size over all systems of profiles. We have the following result.

▶ **Theorem 23.** *Suppose that σ consists of finitely many unary relation symbols. The problem M_σ-SIS reduces to* EP_σ-ENTAIL.

Proof. Fix T to be an antichain of profiles of size M_σ. For each subset $S \subseteq \sigma$, let θ_S denote the formula $(\exists x \bigwedge_{U \in S} U(x))$, and for each profile P on σ, let θ_P denote the formula $\bigwedge_{S \in P} \theta_S$. Observe that a structure $\mathbf{B}$ satisfies θ_P if and only if its profile dominates the profile P.

Set $M = M_\sigma$, and let $((\phi_1, \psi_1), \ldots, (\phi_M, \psi_M))$ be an instance of M-SIS. Let $P_1, \ldots, P_m$ be a listing of the profiles in T. For each $i \in \underline{m}$, fix $V_i^0, V_i^1 \in P_i$ to be distinct elements of P_i. For each $i \in \underline{m}$, define ϕ_i^* to be the EP sentence obtained from ϕ_i by propagating all negations to the atomic level; substituting each positive literal x with $\theta_{V_i^1}(x)$ and each negative literal $\neg x$ with $\theta_{V_i^0}(x)$; and, existentially quantifying all variables in front of the resulting formula. For each $i \in \underline{m}$, define ψ_i^* to be the EP sentence obtained from ψ_i analogously. Observe that when, for an index $i \in \underline{m}$, a structure $\mathbf{B}$ has profile P_i, it holds that $\mathbf{B} \models \phi_i^*$ if and only if ϕ_i is satisfiable, and likewise $\mathbf{B} \models \psi_i^*$ if and only if ψ_i is satisfiable.

The reduction, given the named instance of M-SIS, outputs the pair (ϕ, ψ) defined by $\phi = (\bigvee_Q \theta_Q) \vee (\bigvee_{i \in \underline{m}} (\theta_{P_i} \wedge \phi_i^*))$ and $\psi = (\bigvee_Q \theta_Q) \vee (\bigvee_{i \in \underline{m}} (\theta_{P_i} \wedge \psi_i^*))$, where in each sentence the first disjunction is over the profiles Q that strictly dominate T; we say that a profile Q strictly dominates T if there is a $P \in T$ such that Q dominates P and $Q \notin T$. We claim that $\phi \models \psi$ if and only if the instance of M-SIS is a *yes* instance.

($\Rightarrow$): Let $i \in \underline{m}$ be arbitrary; we show that $(\phi_i, \psi_i) \in$ SIS. Let $\mathbf{B}_i$ be a structure with profile P_i. To satisfy a disjunct in ϕ, a structure must satisfy either a sentence θ_Q, for a profile Q that strictly dominates T, or a sentence θ_{P_i}. Of these sentences, $\mathbf{B}_i$ satisfies only θ_{P_i}, and so by definition of ϕ, we have that $\mathbf{B}_i \models \phi$ if and only if $\mathbf{B}_i \models \phi_i^*$. Likewise, we have that $\mathbf{B}_i \models \psi$ if and only if $\mathbf{B}_i \models \psi_i^*$. Under the assumption, $\mathbf{B}_i \models \phi$ implies $\mathbf{B}_i \models \psi$, and hence the satisfiability of ϕ_i implies the satisfiability of ψ_i.

($\Leftarrow$): Let $\mathbf{B}$ be an arbitrary structure with profile Q. We consider three cases. If the profile Q does not dominate any profile P_i in T, then $\mathbf{B} \not\models \phi$. If the profile Q is equal to a profile $P_i \in T$, then, by the discussion in the previous paragraph, the conditions $\mathbf{B} \models \phi$ and $\mathbf{B} \models \phi_i^*$ are equivalent; likewise, the conditions $\mathbf{B} \models \psi$ and $\mathbf{B} \models \psi_i^*$ are equivalent. As observed, the first pair of conditions occurs exactly when ϕ_i is satisfiable; this implies (by hypothesis) that ψ_i is satisfiable, which occurs exactly when the second pair of conditions occurs. Finally, if the profile Q strictly dominates T, then θ_Q appears as a disjunct of both ϕ and ψ, and both $\mathbf{B} \models \phi$ and $\mathbf{B} \models \psi$ hold. ◀

Note that, in particular, if $|\sigma| \geq 2$, we obtain that SIS reduces to EP$_\sigma$-Entail, as such a σ has a system of profiles of size 1 (see Example 22).

References

1. F. Börner, A. A. Bulatov, H. Chen, P. Jeavons, and A. A. Krokhin. The Complexity of Constraint Satisfaction Games and QCSP. *Inform. Comput.*, 207(9):923–944, 2009.
2. S. Bova and H. Chen. The Complexity of Width Minimization for Existential Positive Queries. In *Proceedings of ICDT'14*, pp. 235–244, 2014.
3. A. K. Chandra and P. M. Merlin. Optimal Implementation of Conjunctive Queries in Relational Data Bases. In *Proceedings of STOC'77*, pp. 77–90, 1977.
4. H. Chen. A Rendezvous of Logic, Complexity, and Algebra. *ACM Computing Surveys*, 42(1), 2009.
5. H. Chen. On the Complexity of Existential Positive Queries. *ACM Trans. Comput. Log.*, 15(1):9, 2014.
6. H. Chen, F. Madelaine, and B. Martin. Quantified Constraints and Containment Problems. In *Proceedings of LICS'08*, 2008.
7. V. Dalmau, P. G. Kolaitis, and M. Y. Vardi. Constraint Satisfaction, Bounded Treewidth, and Finite-Variable Logics. In *Proceedings of CP'02*, 2002.
8. R. Diestel. *Graph Theory, 4th Edition*, Springer, 2012.
9. P. Hell and J. Nesetril. The Core of a Graph. *Discrete Math.*, 109:117–126, 1992.
10. C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1995.

11 Y. Sagiv and M. Yannakakis. Equivalences among relational expressions with the union and difference operators. *Journal of the ACM*, 27(4):633–655, 1980.