

Interpolants and Explicit Definitions in Extensions of the Description Logic \mathcal{EL}

Marie Fortin , Boris Konev , Frank Wolter

University of Liverpool

{mfortin, konev, wolter}@liverpool.ac.uk

Abstract

We show that the vast majority of extensions of the description logic \mathcal{EL} do not enjoy the Craig interpolation nor the projective Beth definability property. This is the case, for example, for \mathcal{EL} with nominals, \mathcal{EL} with the universal role, \mathcal{EL} with a role inclusion of the form $r \circ s \sqsubseteq s$, and for \mathcal{ELI} . It follows in particular that the existence of an explicit definition of a concept or individual name cannot be reduced to subsumption checking via implicit definability. We show that nevertheless the existence of interpolants and explicit definitions can be decided in polynomial time for standard tractable extensions of \mathcal{EL} (such as \mathcal{EL}^{++}) and in EXPTIME for \mathcal{ELI} and various extensions. It follows that these existence problems are not harder than subsumption which is in sharp contrast to the situation for expressive DLs. We also obtain tight bounds for the size of interpolants and explicit definitions and the complexity of computing them: single exponential for tractable standard extensions of \mathcal{EL} and double exponential for \mathcal{ELI} and extensions. We close with a discussion of Horn-DLs such as Horn- \mathcal{ALCI} .

1 Introduction

The *projective Beth definability property* (PBDP) of a description logic (DL) \mathcal{L} states that a concept or individual name is explicitly definable under an \mathcal{L} -ontology \mathcal{O} by an \mathcal{L} -concept using symbols from a signature Σ of concept, role, and individual names if, and only if, it is implicitly definable using Σ under \mathcal{O} . The importance of the PBDP for DL research stems from the fact that it provides a polynomial time reduction of the problem to decide the existence of an explicit definition to the well understood problem of subsumption checking. The existence of explicit definitions is important for numerous knowledge engineering tasks and applications of description logic ontologies, for example, the extraction of equivalent acyclic TBoxes from ontologies (ten Cate et al. 2006; ten Cate, Franconi, and Seylan 2013), the computation of referring expressions (or definite descriptions) for individuals (Artale et al. 2021b), the equivalent rewriting of ontology-mediated queries into concepts (Seylan, Franconi, and de Bruijn 2009; Lutz, Seylan, and Wolter 2019; Toman and Weddell 2021), the construction of alignments between ontologies (Geleta, Payne, and Tamma 2016), and the decomposition of ontologies (Konev et al. 2010).

The PBDP is often investigated in tandem with the *Craig interpolation property* (CIP) which states that if an \mathcal{L} -concept is subsumed by another \mathcal{L} -concept under some \mathcal{L} -ontology then one finds an interpolating \mathcal{L} -concept using the shared symbols of the two input concepts only. In fact, the CIP implies the PBDP and the interpolants obtained using the CIP can serve as explicit definitions.

Many standard Boolean DLs such as \mathcal{ALC} , \mathcal{ALCI} , and \mathcal{ALCQI} enjoy the CIP and PBDP and sophisticated algorithms for computing interpolants and explicit definitions have been developed (ten Cate, Franconi, and Seylan 2013). Important exceptions are the extensions of any of the above DLs with nominals and/or role hierarchies. In fact, it has recently been shown that the problem of deciding the existence of an interpolant/explicit definition becomes 2EXPTIME-complete for \mathcal{ALCO} (\mathcal{ALC} with nominals) and for \mathcal{ALCH} (\mathcal{ALC} with role hierarchies). This result is in sharp contrast to the EXPTIME-completeness of the same problem for \mathcal{ALC} itself inherited from the EXPTIME-completeness of subsumption under \mathcal{ALC} -ontologies (Artale et al. 2021a).

Our aim in this article is threefold: (1) determine which members of the \mathcal{EL} -family of DLs enjoy the CIP/PBDP; (2) investigate the complexity of deciding the existence of interpolants/explicit definitions for those that do not enjoy it; and (3) establish tight bounds on the size of interpolants/explicit definitions and the complexity of computing them.

In what follows we discuss our main results. It has been shown in (Konev et al. 2010; Lutz, Seylan, and Wolter 2019) already that \mathcal{EL} and \mathcal{EL} with role hierarchies enjoy the CIP and PBDP. Rather surprisingly, it turns out that none of the remaining standard DLs in the \mathcal{EL} -family enjoy the CIP nor the PBDP.

Theorem 1. *The following DLs do not enjoy the CIP nor PBDP:*

1. \mathcal{EL} with the universal role,
2. \mathcal{EL} with nominals,
3. \mathcal{EL} with a single role inclusion $r \circ s \sqsubseteq s$,
4. \mathcal{EL} with role hierarchies and a transitive role,
5. the extension \mathcal{ELI} of \mathcal{EL} with inverse roles.

In Points 2 to 5, the CIP/PBDP also fails if the universal role can occur in interpolants/explicit definitions.

Theorem 1 also has interesting consequences that are not explicitly stated. For instance, it follows that neither the DL \mathcal{EL}^{++} introduced in (Baader, Brandt, and Lutz 2005) nor the extension of \mathcal{ELI} with any combination of nominals, role hierarchies, or transitive roles enjoy the CIP/PBDP. With the exception of the failure of the CIP/PBDP for \mathcal{EL} with nominals (without the universal role in interpolants/explicit definitions) (Artale et al. 2021b), our results are new.

It follows from Theorem 1 that the behaviour of extensions of \mathcal{EL} is fundamentally different from extensions of \mathcal{ALC} : adding role hierarchies to \mathcal{ALC} does not preserve the CIP/PBDP (Konev et al. 2009) but it does for \mathcal{EL} ; on the other hand, adding the universal role or inverse roles to \mathcal{ALC} preserves the CIP/PBDP (ten Cate, Franconi, and Seylan 2013) but it does not for \mathcal{EL} .

Theorem 1 leaves open the behaviour of a few natural DLs between \mathcal{EL} and its extension with arbitrary role inclusions. For instance, what happens if one only adds transitive roles or, more generally, role inclusions using a single role name only? To cover these cases we show a general result that implies that these DLs enjoy the CIP and PBDP. In particular, it follows that in Point 4 of Theorem 1 the combination of role hierarchies with a transitive role is necessary for failure of the CIP/PBDP.

We next discuss our main result about tractable extensions of \mathcal{EL} .

Theorem 2. *For \mathcal{EL} and any extension with any combination of nominals, role inclusions, the universal role, or \perp , the existence of interpolants and explicit definitions is in PTIME. If an interpolant/explicit definition exists, then there exists one of at most exponential size that can be computed in exponential time. This bound is optimal.*

It follows that for tractable extensions of \mathcal{EL} the complexity of deciding the existence of interpolants and explicit definitions does not depend on the CIP/PBDP, in sharp contrast to the behaviour of \mathcal{ALCO} and \mathcal{ALCH} . Moreover, the proof shows how interpolants and explicit definitions can be computed from the canonical models introduced in (Baader, Brandt, and Lutz 2005), if they exist. It applies derivation trees (first introduced in (Bienvenu, Lutz, and Wolter 2013) for DLs without nominals and role hierarchies) to estimate the size of interpolants and provide an exponential time algorithm for computing them.

Theorem 3. *For \mathcal{ELI} and any extension with any combination of nominals, the universal role, or \perp , the existence of interpolants and explicit definitions is EXPTIME-complete. If an interpolant/explicit definition exists, then there exists one of at most double exponential size that can be computed in double exponential time. This bound is optimal.*

The proof of Theorem 3 shows how an interpolant or explicit definition can be extracted from a (potentially infinite) tree-shaped canonical model. The EXPTIME complexity bound is proved using an encoding as an emptiness problem for tree automata that also uses derivation trees. It does not seem possible to obtain tight bounds on the size of interpolants using derivation trees; instead we generalize transfer sequences for this purpose (also first introduced in (Bienvenu, Lutz, and Wolter 2013)).

In the final section, we consider expressive Horn-DLs such as Horn- \mathcal{ALCI} . We first observe that Theorem 3 also holds for Horn- \mathcal{ALCI} and extensions with nominals and the universal role, provided one asks for interpolants and explicit definitions in \mathcal{ELI} (and extensions with nominals and the universal role, respectively). If one admits expressive Horn-concepts as interpolants or explicit definitions, then sometimes interpolants and explicit definitions exist that previously did not exist. We show that nevertheless the CIP/PBDP also fail in this case for DLs including Horn- \mathcal{ALC} , \mathcal{ELI} , and Horn- \mathcal{ALCI} .

Detailed proofs are given in the arxiv version of this article.

2 Related Work

The CIP and PBDP have been investigated extensively in databases, with applications to query rewriting under views and query compilation (Toman and Weddell 2011; Benedikt et al. 2016). The computation of explicit definitions under Horn ontologies can be seen as an instance of query reformulation under constraints (Deutsch, Popa, and Tannen 2006) which has been a major research topic for many years. The Chase and Backchase approach that is central to this research closely resembles our use of canonical models. We do not assume, however, that the chase terminates. In (Benedikt et al. 2016; 2017), it is shown that the reformulation of CQs into CQs under tgds can be reduced to entailment using Lyndon interpolation of first-order logic. By linking reformulation into CQs and definability using concepts, this approach can potentially be used to obtain alternative proofs of complexity upper bounds for interpolant and explicit definition existence in our languages. Also relevant is the investigation of interpolation in basic modal logic (Maksimova and Gabbay 2005) and hybrid modal logic (Areces, Blackburn, and Marx 2001; ten Cate 2005).

The main aim of this article is to investigate explicit definability of concept and individual names under ontologies. We have therefore chosen a definition of the CIP and interpolants that generalizes the projective Beth definability property and explicit definability in a natural and useful way, following (ten Cate, Franconi, and Seylan 2013). There are, however, other notions of Craig interpolation that are of interest. Of particular importance for modularity and various other purposes is the following version: if \mathcal{O} is an ontology and $C \sqsubseteq D$ an inclusion such that $\mathcal{O} \models C \sqsubseteq D$, then there exists an ontology \mathcal{O}' in the shared signature of \mathcal{O} and $C \sqsubseteq D$ such that $\mathcal{O}' \models C \sqsubseteq D$. This property has been considered for \mathcal{EL} and various extensions in (Sofronie-Stokkermans 2008; Konev et al. 2010). Currently, it is unknown whether there exists any interesting relationship between this version of the CIP and the version we investigate in this article.

Craig interpolants should not be confused with uniform interpolants (or forgetting) (Lutz, Seylan, and Wolter 2012; Lutz and Wolter 2011; Nikitina and Rudolph 2014; Koopmann and Schmidt 2015). Uniform interpolants generalize Craig interpolants in the sense that a uniform interpolant is

an interpolant for a fixed antecedent and any formula implied by the antecedent and sharing with it a fixed set of symbols.

Interpolant and explicit definition existence have only recently been investigated for logics that do not enjoy the CIP or PBDP. Extending work on Boolean DLs we discussed already, it is shown that they become harder than validity also in the guarded and two-variable fragment (Jung and Wolter 2021). The interpolant existence problem for linear temporal logic LTL (equivalently, first-order logic) is considered in (Place and Zeitoun 2016). In the context of referring expressions, explicit definition existence is investigated in (Artale et al. 2021b), see also (Borgida, Toman, and Weddell 2016).

3 Preliminaries

Let \mathbb{N}_C , \mathbb{N}_R , and \mathbb{N}_I be disjoint and countably infinite sets of *concept*, *role*, and *individual names*. A *role* is a role name r or an *inverse role* r^- , with r a role name. *Nominals* take the form $\{a\}$, where a is an individual name. The *universal role* is denoted by u . \mathcal{ELIO}_u -concepts C are defined by the following syntax rule:

$$C, C' ::= \top \mid A \mid \{a\} \mid C \sqcap C' \mid \exists r.C$$

where A ranges over concept names, a over individual names, and r over roles (including the universal role). Fragments of \mathcal{ELIO}_u are defined as usual. For example, \mathcal{ELI} -concepts are \mathcal{ELIO}_u -concepts without nominals and the universal role, and \mathcal{EL} -concepts are \mathcal{ELI} -concepts without inverse roles. Given any of the DLs \mathcal{L} introduced above, an \mathcal{L} -concept inclusion (\mathcal{L} -CI) takes the form $C \sqsubseteq D$ with C, D \mathcal{L} -concepts. An \mathcal{L} -ontology \mathcal{O} is a finite set of \mathcal{L} -CIs.

We also consider ontologies with *role inclusions* (RIs), expressions of the form $r_1 \circ \dots \circ r_n \sqsubseteq r$ with r_1, \dots, r_n, r role names. An \mathcal{ELRO}_u -ontology with RIs is called an \mathcal{ELRO}_u -ontology. A set of RIs is a *role hierarchy* if all its RIs are of the form $r \sqsubseteq s$ with r, s role names.

A *signature* Σ is a set of concept, role, and individual names, uniformly referred to as (*non-logical*) *symbols*. We follow common practice and do not regard the universal role u as a non-logical symbol as its interpretation is fixed. We use $\text{sig}(X)$ to denote the set of symbols used in any syntactic object X such as a concept or an ontology. If \mathcal{L} is a DL and Σ a signature, then an $\mathcal{L}(\Sigma)$ -concept C is an \mathcal{L} -concept with $\text{sig}(C) \subseteq \Sigma$. The *size* $\|X\|$ of a syntactic object X is the number of symbols needed to write it down.

The semantics of DLs is given in terms of *interpretations* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set (the *domain*) and $\cdot^{\mathcal{I}}$ is the *interpretation function*, assigning to each $A \in \mathbb{N}_C$ a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, to each $r \in \mathbb{N}_R$ a relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and to each $a \in \mathbb{N}_I$ an element $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$. The interpretation $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ of a concept C in \mathcal{I} is defined as usual, see (Baader et al. 2017). An interpretation \mathcal{I} satisfies a CI $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ and an RI $r_1 \circ \dots \circ r_n \sqsubseteq r$ if $r_1^{\mathcal{I}} \circ \dots \circ r_n^{\mathcal{I}} \subseteq r^{\mathcal{I}}$. We say that \mathcal{I} is a *model* of an ontology \mathcal{O} if it satisfies all inclusions in it. If α is a CI or RI, we write $\mathcal{O} \models \alpha$ if all models of \mathcal{O} satisfy α . We write $\mathcal{O} \models C \equiv D$ if $\mathcal{O} \models C \sqsubseteq D$ and $\mathcal{O} \models D \sqsubseteq C$.

An ontology is in *normal form* if its CIs are of the form

$$\top \sqsubseteq A, \quad A_1 \sqcap A_2 \sqsubseteq B, \quad A \sqsubseteq \{a\}, \quad \{a\} \sqsubseteq A,$$

and

$$A \sqsubseteq \exists r.B, \quad \exists r.B \sqsubseteq A$$

where A, A_1, A_2, B are concept names, r is a role or the universal role, and a is an individual name. It is well known that for any \mathcal{ELIO}_u -ontology \mathcal{O} with or without RIs one can construct in polynomial time a conservative extension \mathcal{O}' using the same constructors as \mathcal{O} that is in normal form.

$\mathcal{L}(\Sigma)$ -concepts can be characterized using $\mathcal{L}(\Sigma)$ -simulations which we define next. Let \mathcal{I} and \mathcal{J} be interpretations. A relation $S \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{J}}$ is called an $\mathcal{EL}(\Sigma)$ -simulation between \mathcal{I} and \mathcal{J} if the following conditions hold:

1. if $d \in A^{\mathcal{I}}$ and $(d, e) \in S$, then $e \in A^{\mathcal{J}}$, for all $A \in \mathbb{N}_C \cap \Sigma$;
2. if $d = a^{\mathcal{I}}$ and $(d, e) \in S$, then $e = a^{\mathcal{J}}$, for all $a \in \mathbb{N}_I \cap \Sigma$;
3. if $(d, d') \in r^{\mathcal{I}}$ and $(d, e) \in S$, then there exists e' with $(e, e') \in r^{\mathcal{J}}$ and $(d', e') \in S$, for all $r \in \mathbb{N}_R \cap \Sigma$.

S is called an $\mathcal{EL}(\Sigma)$ -simulation if $\Delta^{\mathcal{I}}$ is the domain of S and an $\mathcal{ELIO}(\Sigma)$ -simulation if Condition 3 also holds for inverse roles from Σ . Condition 2 is dropped if \mathcal{L} does not use nominals. We write $(\mathcal{I}, d) \preceq_{\mathcal{L}, \Sigma} (\mathcal{J}, e)$ if there exists an $\mathcal{L}(\Sigma)$ -simulation S between \mathcal{I} and \mathcal{J} with $(d, e) \in S$. We write $(\mathcal{I}, d) \preceq_{\mathcal{L}, \Sigma} (\mathcal{J}, e)$ if $d \in C^{\mathcal{I}}$ implies $e \in C^{\mathcal{J}}$ for all $\mathcal{L}(\Sigma)$ -concepts C . The following characterization is well known (Lutz and Wolter 2010; Lutz, Piro, and Wolter 2011).

Lemma 1. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_u, \mathcal{ELO}, \mathcal{ELO}_u, \mathcal{ELI}, \mathcal{ELI}_u\}$. Then $(\mathcal{I}, d) \preceq_{\mathcal{L}, \Sigma} (\mathcal{J}, e)$ implies $(\mathcal{I}, d) \preceq_{\mathcal{L}, \Sigma} (\mathcal{J}, e)$. The converse direction holds if \mathcal{J} is finite.*

4 Craig Interpolation Property and Projective Beth Definability Property

We introduce the Craig interpolation property (CIP) as defined in (ten Cate, Franconi, and Seylan 2013) and the projective Beth definability property (PBDP) and prove Theorem 1 from the introduction to this article. We observe that the CIP implies the PBDP, but lack a proof of the converse direction. Nevertheless, all DLs considered in this paper enjoying the PBDP also enjoy the CIP.

Set $\text{sig}(\mathcal{O}, C) = \text{sig}(\mathcal{O}) \cup \text{sig}(C)$, for any ontology \mathcal{O} and concept C . Let $\mathcal{O}_1, \mathcal{O}_2$ be \mathcal{L} -ontologies and let C_1, C_2 be \mathcal{L} -concepts. Then an \mathcal{L} -concept D is called an \mathcal{L} -interpolant¹ for $C_1 \sqsubseteq C_2$ under $\mathcal{O}_1, \mathcal{O}_2$ if

- $\text{sig}(D) \subseteq \text{sig}(\mathcal{O}_1, C_1) \cap \text{sig}(\mathcal{O}_2, C_2)$;
- $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqsubseteq D$;
- $\mathcal{O}_1 \cup \mathcal{O}_2 \models D \sqsubseteq C_2$.

¹Important variations of this definition are to drop \mathcal{O}_2 in Point 2 and \mathcal{O}_1 in Point 3, respectively, or to consider only one ontology $\mathcal{O} = \mathcal{O}_1 = \mathcal{O}_2$ and regard the signature Σ of the interpolant as an input given independently from \mathcal{O}, C_1, C_2 . This has an effect on the CIP, but our results on interpolant computation and existence are not affected.

Definition 1. A DL \mathcal{L} has the *Craig interpolation property (CIP)* if for any \mathcal{L} -ontologies $\mathcal{O}_1, \mathcal{O}_2$ and \mathcal{L} -concepts C_1, C_2 such that $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqsubseteq C_2$ there exists an \mathcal{L} -interpolant for $C_1 \sqsubseteq C_2$ under $\mathcal{O}_1, \mathcal{O}_2$.

We next define the relevant definability notions. Let \mathcal{O} be an ontology and A a concept name. Let $\Sigma \subseteq \text{sig}(\mathcal{O})$ be a signature. An $\mathcal{L}(\Sigma)$ -concept C is an *explicit $\mathcal{L}(\Sigma)$ -definition of A under \mathcal{O}* if $\mathcal{O} \models A \equiv C$. We call A *explicitly definable in $\mathcal{L}(\Sigma)$ under \mathcal{O}* if there is an explicit $\mathcal{L}(\Sigma)$ -definition of A under \mathcal{O} . The Σ -reduct $\mathcal{I}_{|\Sigma}$ of an interpretation \mathcal{I} coincides with \mathcal{I} except that no symbol that is not in Σ is interpreted in $\mathcal{I}_{|\Sigma}$. A concept A is called *implicitly definable using Σ under \mathcal{O}* if the Σ -reduct of any model \mathcal{I} of \mathcal{O} determines the set $A^{\mathcal{I}}$; in other words, if \mathcal{I} and \mathcal{J} are both models of \mathcal{O} such that $\mathcal{I}_{|\Sigma} = \mathcal{J}_{|\Sigma}$, then $A^{\mathcal{I}} = A^{\mathcal{J}}$. It is easy to see that implicit definability can be reformulated as a standard reasoning problem as follows: a concept name $A \notin \Sigma$ is implicitly definable using Σ under \mathcal{O} iff $\mathcal{O} \cup \mathcal{O}_\Sigma \models A \equiv A'$, where \mathcal{O}_Σ is obtained from \mathcal{O} by replacing every symbol X not in Σ (including A) uniformly by a fresh symbol X' .

Definition 2. A DL \mathcal{L} has the *projective Beth definable property (PBDP)* if for any \mathcal{L} -ontology \mathcal{O} , concept name A , and signature $\Sigma \subseteq \text{sig}(\mathcal{O})$ the following holds: if A is implicitly definable using Σ under \mathcal{O} , then A is explicitly $\mathcal{L}(\Sigma)$ -definable under \mathcal{O} .

Remark 1. The CIP implies the PBDP. To see this, assume that an \mathcal{L} -ontology \mathcal{O} , concept name A and a signature Σ are given, and that A is implicitly definable from Σ under \mathcal{O} . Then $\mathcal{O} \cup \mathcal{O}_\Sigma \models A \equiv A'$, with \mathcal{O}_Σ defined above. Take an \mathcal{L} -interpolant C for $A \sqsubseteq A'$ under $\mathcal{O}, \mathcal{O}_\Sigma$. Then C is an explicit $\mathcal{L}(\Sigma)$ -definition of A under \mathcal{O} .

Remark 2. The PBDP implies that implicitly definable nominals are explicitly definable and that, more generally, every implicitly definable concept C is explicitly definable. This can be shown by adding $A \equiv C$ to the ontology for a fresh concept name A and asking for an explicit definition of A in the extended ontology.

Remark 3. The CIP and PBDP are invariant under adding \perp (interpreted as the empty set) to the languages introduced above. The straightforward proof is given in the appendix of the full version.

We next prove that the majority of tractable extensions of \mathcal{EL} does not enjoy the CIP nor PBDP.

Theorem 1. *The following DLs do not enjoy the CIP nor PBDP:*

1. \mathcal{EL} with the universal role,
2. \mathcal{EL} with nominals,
3. \mathcal{EL} with a single role inclusion $r \circ s \sqsubseteq s$,
4. \mathcal{EL} with role hierarchies and a transitive role,
5. \mathcal{EL} with inverse roles.

In Points 2 to 5, the CIP/PBDP also fails if the universal role can occur in interpolants/explicit definitions.

Proof. We first show that \mathcal{EL}_u does not enjoy the PBDP. Point 1 then follows using Remark 1. We define an \mathcal{EL}_u -ontology \mathcal{O}_u , signature Σ , and concept name A such that A

is implicitly definable using Σ under \mathcal{O}_u but not $\mathcal{EL}_u(\Sigma)$ -explicitly definable under \mathcal{O}_u . Define \mathcal{O}_u as the following set of CIs:

$$\begin{aligned} A \sqsubseteq B, \quad D \sqcap \exists u.A \sqsubseteq E, \quad B \sqsubseteq \exists r.C \\ C \sqsubseteq D, \quad B \sqcap \exists r.(C \sqcap E) \sqsubseteq A, \end{aligned}$$

and let $\Sigma = \{B, D, E, r\}$. We have $\mathcal{O}_u \models A \equiv B \sqcap \forall r.(D \rightarrow E)$,² so A is implicitly definable using Σ under \mathcal{O}_u . The interpretations \mathcal{I} and \mathcal{I}' given in Figure 1 show that A is not explicitly $\mathcal{EL}_u(\Sigma)$ -definable under \mathcal{O}_u . Indeed, \mathcal{I}

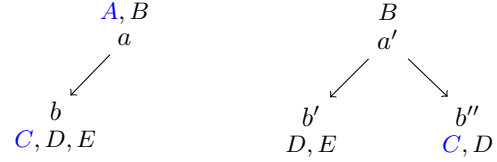


Figure 1: Interpretations \mathcal{I} (left) and \mathcal{I}' (right) used for \mathcal{O}_u .

and \mathcal{I}' are both models of \mathcal{O}_u , $a \in A^{\mathcal{I}}$, $a' \notin A^{\mathcal{I}'}$, and the relation $\{(a, a'), (b, b')\}$ is a $\mathcal{EL}_u(\Sigma)$ -simulation between \mathcal{I} and \mathcal{I}' . As $\mathcal{EL}_u(\Sigma)$ -concepts are preserved under $\mathcal{EL}_u(\Sigma)$ -simulations (Lemma 1), if $\mathcal{O}_u \models A \equiv F$ for some $\mathcal{EL}_u(\Sigma)$ -concept F , then from $a \in A^{\mathcal{I}}$ we obtain $a \in F^{\mathcal{I}}$. This implies $a' \in F^{\mathcal{I}'}$, and so $a' \in A^{\mathcal{I}'}$. As $a' \notin A^{\mathcal{I}'}$, we obtain a contradiction.

We next prove Point 2. An example from (Artale et al. 2021b) shows that $\mathcal{EL}\mathcal{O}$ does not enjoy the CIP/PBDP. Here we show that $\mathcal{EL}\mathcal{O}$ does not enjoy the CIP/PBDP, even if interpolants/explicit definitions are from $\mathcal{EL}\mathcal{O}_u$. Let \mathcal{O}_n contain the following CIs:

$$\begin{aligned} A \sqsubseteq \exists r.(E \sqcap \{c\}), \quad \top \sqsubseteq \exists s.(Q_2 \sqcap \exists s.\{c\}) \\ \exists s.(Q_1 \sqcap Q_2 \sqcap \exists s.\{c\}) \sqsubseteq A, \quad \exists s.E \sqsubseteq Q_1 \end{aligned}$$

and let $\Sigma = \{c, s, Q_1\}$. Observe that A is implicitly definable using Σ under \mathcal{O}_n as $\mathcal{O}_n \models A \equiv \forall s.(\exists s.\{c\} \rightarrow Q_1)$. The relation $\{(a, a'), (b, b'), (c, c')\}$ is an $\mathcal{EL}\mathcal{O}_u(\Sigma)$ -simulation between the interpretations \mathcal{I} and \mathcal{I}' defined in Figure 2. Now we can apply the same argument as in Point 1 to show that A is not explicitly $\mathcal{EL}\mathcal{O}_u(\Sigma)$ -definable under \mathcal{O}_n .

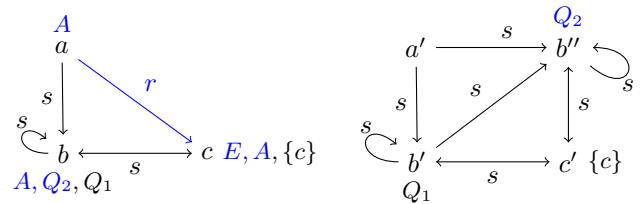


Figure 2: Interpretations \mathcal{I} (left) and \mathcal{I}' (right) used for \mathcal{O}_n .

For Point 3, let \mathcal{O}_r contain

$$A \sqsubseteq \exists r.E, \quad E \sqsubseteq \exists s.B, \quad \exists s.B \sqsubseteq A, \quad r \circ s \sqsubseteq s,$$

²Here and in what follows we use standard \mathcal{ALC} syntax and semantics and set $C \rightarrow D := \neg C \sqcup D$ (Baader et al. 2017).

and let $\Sigma = \{s, E\}$. Then A is implicitly definable using Σ under \mathcal{O}_r since

$$\mathcal{O}_r \models \forall x(A(x) \leftrightarrow \exists y(E(y) \wedge \forall z(s(y, z) \rightarrow s(x, z))).$$

We show that there does not exist any $\mathcal{EL}_u(\Sigma)$ -explicit definition of A under \mathcal{O}_r . The interpretations \mathcal{I} and \mathcal{I}' given

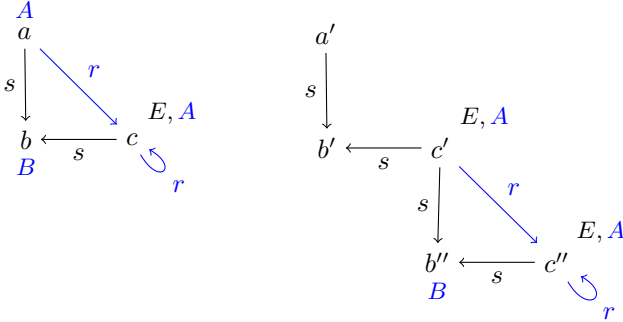


Figure 3: Interpretations \mathcal{I} (left) and \mathcal{I}' (right) used for \mathcal{O}_r .

in Figure 3 are both models of \mathcal{O}_r , $a \in A^{\mathcal{I}}$, $a' \notin A^{\mathcal{I}'}$, and the relation $\{(a, a'), (b, b'), (c, c')\}$ is an $\mathcal{EL}_u(\Sigma)$ -simulation between \mathcal{I} and \mathcal{I}' . One can now show in the same way as in Point 1 that no $\mathcal{EL}_u(\Sigma)$ -definition of A under \mathcal{O}_r exists.

Point 4 is shown in the appendix of the full version using a modification of the ontology used for Point 3.

To prove Point 5, obtain an \mathcal{ELI} -ontology \mathcal{O}_i from \mathcal{O}_u defined above by replacing the second CI of \mathcal{O}_u by $D \sqcap \exists r^-.A \sqsubseteq E$. Let, as before, $\Sigma = \{B, D, E, r\}$. Then A is implicitly definable from Σ under \mathcal{O}_i (the same explicit definition works), but A is not explicitly $\mathcal{EL}_u(\Sigma)$ -definable under \mathcal{O}_i (the same interpretations \mathcal{I} and \mathcal{I}' work). \square

We next discuss a general positive result on interpolation and explicit definition existence that shows that Theorem 1 is essentially optimal. A set \mathcal{R} of RIs is *safe* for a signature Σ if for each RI $r_1 \circ \dots \circ r_n \sqsubseteq r \in \mathcal{R}$, $n \geq 1$, if $\{r_1, \dots, r_n, r\} \cap \Sigma \neq \emptyset$ then $\{r_1, \dots, r_n, r\} \subseteq \Sigma$.

Theorem 4. *Let $\mathcal{O}_1, \mathcal{O}_2$ be \mathcal{EL} -ontologies with RIs, C_1, C_2 be \mathcal{EL} -concepts, and set $\Sigma = \text{sig}(\mathcal{O}_1, C_1) \cap \text{sig}(\mathcal{O}_2, C_2)$. Assume that the set of RIs in $\mathcal{O}_1 \cup \mathcal{O}_2$ is safe for Σ and $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqsubseteq C_2$. Then an \mathcal{EL} -interpolant for $C_1 \sqsubseteq C_2$ under $\mathcal{O}_1, \mathcal{O}_2$ exists.*

The proof technique is based on simulations and similar to (Konev et al. 2010; Lutz, Seylan, and Wolter 2019). Theorem 4 has a few interesting consequences. For instance, \mathcal{EL} with transitive roles enjoys both the CIP and PBPD since transitivity is expressed by the role inclusion $r \circ r \sqsubseteq r$ which is safe for any signature (as it only uses a single role name).

5 Interpolant and Explicit Definition Existence

We introduce interpolant and explicit definition existence as decision problems and establish a polynomial time reduction of the latter to the former. We then show that it suffices to consider ontologies in normal form and that the addition of \perp does not affect the complexity of the decision problems.

Definition 3. *Let \mathcal{L} be a DL. Then \mathcal{L} -interpolant existence is the problem to decide for any \mathcal{L} -ontologies $\mathcal{O}_1, \mathcal{O}_2$ and \mathcal{L} -concepts C_1, C_2 whether there exists an \mathcal{L} -interpolant for $C_1 \sqsubseteq C_2$ under $\mathcal{O}_1, \mathcal{O}_2$.*

Observe that interpolant existence reduces to checking $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqsubseteq C_2$ for logics with the CIP but that this is not the case for logics without the CIP.

Definition 4. *Let \mathcal{L} be a DL. Then \mathcal{L} -explicit definition existence is the problem to decide for any \mathcal{L} -ontology \mathcal{O} , signature Σ , and concept name A whether A is explicitly definable in $\mathcal{L}(\Sigma)$ under \mathcal{O} .*

Remark 4. There is a polynomial time reduction of \mathcal{L} -explicit definition existence to \mathcal{L} -interpolant existence. Moreover, any algorithm computing \mathcal{L} -interpolants also computes \mathcal{L} -explicit definitions and any bound on the size of \mathcal{L} -interpolants provides a bound on the size of \mathcal{L} -explicit definitions. The proof is similar to the proof of Remark 1.

We next observe that replacing the original ontologies by a conservative extension preserves interpolants and explicit definitions. Thus, it suffices to consider ontologies in normal form and interpolants for inclusions between concept names.

Lemma 2. *Let $\mathcal{O}_1, \mathcal{O}_2$ be ontologies and C_1, C_2 concepts in any DL \mathcal{L} considered in this paper. Then one can compute in polynomial time \mathcal{L} -ontologies $\mathcal{O}'_1, \mathcal{O}'_2$ in normal form and with fresh concept names A, B such that an \mathcal{L} -concept C is an interpolant for $C_1 \sqsubseteq C_2$ under $\mathcal{O}_1, \mathcal{O}_2$ iff it is an interpolant for $A \sqsubseteq B$ under $\mathcal{O}'_1, \mathcal{O}'_2$.*

Proof. Let \mathcal{O}'_1 and \mathcal{O}'_2 be normal form conservative extensions of $\mathcal{O}_1 \cup \{A \equiv C\}$ and, respectively, $\mathcal{O}_2 \cup \{B \equiv C\}$, computed in polynomial time. One can show that \mathcal{O}'_1 and \mathcal{O}'_2 are as required. \square

Remark 5. Assume that \mathcal{L} is any of the DLs introduced above and let \mathcal{L}_\perp denote its extension with \perp . Then \mathcal{L} -interpolant existence and \mathcal{L} -explicit definition existence can be reduced in polynomial time to \mathcal{L}_\perp -interpolant existence and \mathcal{L}_\perp -explicit definition existence, respectively. The converse direction also holds modulo an oracle deciding whether $\mathcal{O} \models C \sqsubseteq \perp$.

6 Interpolant and Explicit Definition Existence in Tractable \mathcal{EL} Extensions

The aim of this section is to analyse interpolants and explicit definitions for extensions of \mathcal{EL} with any combination of nominals, role inclusions, or the universal role. We show the following result from the introduction.

Theorem 2. *For \mathcal{EL} and any extension with any combination of nominals, role inclusions, the universal role, or \perp , the existence of interpolants and explicit definitions is in PTIME. If an interpolant/explicit definition exists, then there exists one of at most exponential size that can be computed in exponential time. This bound is optimal.*

Before we start with a sketch of the proof we give instructive examples showing that the exponential bound on the size of explicit definitions is optimal.

Example 1. Variants of the following example have already been used for various succinctness arguments in DL. Let

$$\begin{aligned} \mathcal{O}_b = & \{A \sqsubseteq M \sqcap \exists r_1.B_1 \sqcap \exists r_2.B_1\} \cup \\ & \{B_i \sqsubseteq \exists r_1.B_{i+1} \sqcap \exists r_2.B_{i+1} \mid 1 \leq i < n\} \cup \\ & \{B_n \sqsubseteq B, \exists r_1.B \sqcap \exists r_2.B \sqsubseteq B, B \sqcap M \sqsubseteq A\} \end{aligned}$$

and $\Sigma_0 = \{r_1, r_2, B_n, M\}$. A triggers a marker M and a binary tree of depth n whose leaves are decorated with B_n . Conversely, if B_n is true at all leaves of a binary tree of depth n , then B is true at all nodes of the tree and B together with M entail A at its root. Let, inductively, $C_0 := B_n$ and $C_{i+1} = \exists r_1.C_i \sqcap \exists r_2.C_i$, for $0 < i < n$, and $C = M \sqcap C_n$. Then C is the smallest explicit $\mathcal{EL}(\Sigma_0)$ -definition of A under \mathcal{O}_b . Next let

$$\begin{aligned} \mathcal{O}_p = & \{r_i \circ r_i \sqsubseteq r_{i+1} \mid 0 \leq i < n\} \cup \\ & \{A \sqsubseteq \exists r_0.B, B \sqsubseteq \exists r_0.B, \exists r_n.B \sqsubseteq A\} \end{aligned}$$

and $\Sigma_1 = \{r_0, B\}$. Then $\exists r_0^{2^n}.B$ is the smallest explicit $\mathcal{EL}(\Sigma_1)$ -definition of A under \mathcal{O}_p .

Observe that using \mathcal{O}_b one enforces explicit definitions of exponential size by generating a binary tree of linear depth whereas using \mathcal{O}_p this is achieved by generating a path of exponential length. The latter can only happen if role inclusions are used in the ontology. One insight provided by the exponential upper bound on the size of explicit definitions in Theorem 2 is that the two examples cannot be combined to enforce a binary tree of exponential depth.

To continue with the proof we introduce ABoxes as a technical tool that allows us to move from interpretations to (potentially incomplete) sets of facts and concepts. An *ABox* \mathcal{A} is a (possibly infinite) set of assertions of the form $A(x)$, $r(x, y)$, $\{a\}(x)$, and $\top(x)$ with $A \in \mathbf{N}_C$, $r \in \mathbf{N}_R$, $a \in \mathbf{N}_I$, and x, y individual variables (we call individuals used in ABoxes variables to distinguish them from individual names used in nominals). We denote by $\text{ind}(\mathcal{A})$ the set of individual variables in \mathcal{A} . A Σ -ABox is an ABox using symbols from Σ only. Models of ABoxes are defined as usual. We do not make the unique name assumption.

Every interpretation \mathcal{I} defines an ABox $\mathcal{A}_{\mathcal{I}}$ by identifying every $d \in \Delta^{\mathcal{I}}$ with a variable x_d and taking $A(x_d)$ if $d \in A^{\mathcal{I}}$, $r(x_c, x_d)$ if $(c, d) \in r^{\mathcal{I}}$, $\{a\}(x_d)$ if $a^{\mathcal{I}} = d$. Conversely, ABoxes \mathcal{A} define interpretations in the obvious way (by identifying variables x, y if $\{a\}(x), \{a\}(y) \in \mathcal{A}$). We associate with every ABox \mathcal{A} a directed graph $G_{\mathcal{A}} = (\text{ind}(\mathcal{A}), \bigcup_{r \in \mathbf{N}_R} \{(x, y) \mid r(x, y) \in \mathcal{A}\})$. Let Γ be a set of individual names. Then \mathcal{A} is *ditree-shaped modulo* Γ if after dropping some facts of the form $r(x, y)$ with $\{a\}(y) \in \mathcal{A}$ for some $a \in \Gamma$, it is *ditree-shaped* in the sense that $G_{\mathcal{A}}$ is acyclic and $r(x, y) \in \mathcal{A}$ and $s(x, y) \in \mathcal{A}$ imply $r = s$. A *pointed ABox* is a pair \mathcal{A}, x with $x \in \text{ind}(\mathcal{A})$. Then $\mathcal{EL}\mathcal{O}_u(\Sigma)$ -concepts correspond to pointed Σ -ABoxes \mathcal{A}, x such that \mathcal{A} is ditree-shaped modulo $\mathbf{N}_I \cap \Sigma$ and $\mathcal{EL}\mathcal{O}(\Sigma)$ -concepts correspond to *rooted* pointed Σ -ABoxes \mathcal{A}, x such that \mathcal{A} is ditree-shaped modulo $\mathbf{N}_I \cap \Sigma$, where \mathcal{A}, x is called rooted if for every $y \in \text{ind}(\mathcal{A})$ there is a path from x to y in $G_{\mathcal{A}}$. We write $\mathcal{O}, \mathcal{A} \models C(x)$ if $x^{\mathcal{I}} \in C^{\mathcal{I}}$ for every model \mathcal{I} of \mathcal{O} and \mathcal{A} .

Given an $\mathcal{EL}\mathcal{R}\mathcal{O}_u$ -ontology \mathcal{O} in normal form and a concept name A , one can construct in polynomial time the *canonical model* $\mathcal{I}_{\mathcal{O}, A}$ of \mathcal{O} and A using the approach introduced in (Baader, Brandt, and Lutz 2005). More generally, the canonical model $\mathcal{I}_{\mathcal{O}, A}$ for an ABox \mathcal{A} and ontology \mathcal{O} can be constructed in polynomial time and is a model of both \mathcal{O} and \mathcal{A} such that for any $\mathcal{EL}\mathcal{O}_u$ -concept C using symbols from \mathcal{O} only and any $x \in \text{ind}(\mathcal{A})$,

$$(\dagger) \quad \mathcal{O}, \mathcal{A} \models C(x) \text{ iff } x \in C^{\mathcal{I}_{\mathcal{O}, \mathcal{A}}},$$

details are given in the appendix of the full version. We let $\mathcal{I}_{\mathcal{O}, A} = \mathcal{I}_{\mathcal{O}, \mathcal{A}}$ with $\mathcal{A} = \{A(\rho_A)\}$. Note that in (Baader, Brandt, and Lutz 2005) the condition (\dagger) is only stated for subconcepts C of the ontology \mathcal{O} , thus (\dagger) requires a proof.

Example 2. The interpretations \mathcal{I} defined in the proof of Theorem 1 define canonical models $\mathcal{I}_{\mathcal{O}, A}$ with $\rho_A = a$ for the ontologies $\mathcal{O} \in \{\mathcal{O}_u, \mathcal{O}_n, \mathcal{O}_r, \mathcal{O}_i\}$. The interpretations \mathcal{I}' define canonical models $\mathcal{I}_{\mathcal{O}, \mathcal{A}^\Sigma}$ with \mathcal{A}^Σ the Σ -reduct of $\mathcal{I}_{\mathcal{O}, A}$ regarded as an ABox and $\rho_A = a'$.

The *directed unfolding* of a pointed Σ -ABox \mathcal{A}, x into a pointed Σ -ABox \mathcal{A}^u, x that is ditree-shaped modulo $\Sigma \cap \mathbf{N}_I$ is defined in the standard way. In the *rooted directed unfolding*, nodes that cannot be reached from x via role names are dropped.

Assume now that \mathcal{O} is in normal form and A a concept name. Let $\mathcal{A}_{\mathcal{O}, A}^\Sigma$ be the Σ -reduct of the canonical model $\mathcal{I}_{\mathcal{O}, A}$, regarded as an ABox. Denote by $\mathcal{A}_{\mathcal{O}, A}^{\Sigma, u}, \rho_A$ the directed unfolding of $\mathcal{A}_{\mathcal{O}, A}^\Sigma, \rho_A$, by $\mathcal{A}_{\mathcal{O}, A}^{\downarrow \Sigma}, \rho_A$ the sub-ABox of $\mathcal{A}_{\mathcal{O}, A}^\Sigma$ rooted in ρ_A , and by $\mathcal{A}_{\mathcal{O}, A}^{\downarrow \Sigma, u}, \rho_A$ its rooted directed unfolding. Theorem 2 is a direct consequence of the following characterization of interpolants.

Theorem 5. *There exists a polynomial p such that the following conditions are equivalent for all $\mathcal{EL}\mathcal{R}\mathcal{O}_u$ -ontologies $\mathcal{O}_1, \mathcal{O}_2$ in normal form, concept names A, B , and $\Sigma = \text{sig}(\mathcal{O}_1, A) \cap \text{sig}(\mathcal{O}_2, B)$:*

1. *An $\mathcal{EL}\mathcal{O}_u$ -interpolant for $A \sqsubseteq B$ under $\mathcal{O}_1, \mathcal{O}_2$ exists;*
2. *$\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma \models B(\rho_A)$;*
3. *there exists a finite subset \mathcal{A} of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, u}$ with $|\text{ind}(\mathcal{A})| \leq 2^{p(\|\mathcal{O}_1 \cup \mathcal{O}_2\|)}$ such that the $\mathcal{EL}\mathcal{O}_u$ -concept corresponding to \mathcal{A}, ρ_A is an $\mathcal{EL}\mathcal{O}_u$ -interpolant for $A \sqsubseteq B$ under $\mathcal{O}_1, \mathcal{O}_2$.*

The same equivalences hold if in Points 1 to 3, $\mathcal{EL}\mathcal{O}_u$ is replaced by $\mathcal{EL}\mathcal{O}$, $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma$ by $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\downarrow \Sigma}$, and $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, u}$ by $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\downarrow \Sigma, u}$.

In Point 3, \mathcal{A} can be computed in exponential time, if it exists.

Note that the polynomial time decidability of interpolant existence follows from Point 2 of Theorem 5 (and the tractability of $\mathcal{EL}\mathcal{R}\mathcal{O}_u$ (Baader, Brandt, and Lutz 2005)).

Example 3. Our proof of Theorem 2 can be regarded as an application of Theorem 5: by Example 2, the interpretations \mathcal{I} and \mathcal{I}' coincide with the canonical models $\mathcal{I}_{\mathcal{O}, A}$ and $\mathcal{I}_{\mathcal{O}, \mathcal{A}^\Sigma}$ and so $\rho_A = a' \notin A^{\mathcal{I}_{\mathcal{O}, \mathcal{A}^\Sigma}}$ is equivalent to $\mathcal{O}, \mathcal{A}_{\mathcal{O}, A}^\Sigma \not\models A(\rho_A)$ (Point 2 in Theorem 5).

The following example illustrates the difference between the existence of explicit definitions in $\mathcal{EL}\mathcal{O}$ and $\mathcal{EL}\mathcal{O}_u$ and thus the need for moving to the ABoxes $\mathcal{A}_{\mathcal{O},A}^{\downarrow\Sigma}$, and $\mathcal{A}_{\mathcal{O},A}^{\downarrow\Sigma,u}$ if one does not admit the universal role in explicit definitions.

Example 4. Let $\mathcal{O} = \{A \sqsubseteq \{b\}, A \sqsubseteq \exists r.B, B \sqsubseteq \exists s.A\}$ and let $\Sigma = \{b, B\}$. Then A is explicitly $\mathcal{EL}\mathcal{O}_u(\Sigma)$ -definable under \mathcal{O} since $\mathcal{O} \models A \equiv \{b\} \sqcap \exists u.B$ but A is not explicitly $\mathcal{EL}\mathcal{O}(\Sigma)$ -definable. Note that in this case $\mathcal{A}_{\mathcal{O},A}^{\Sigma} = \{\{b\}(\rho_A), B(y)\}$ but $\mathcal{A}_{\mathcal{O},A}^{\downarrow\Sigma} = \{\{b\}(\rho_A)\}$.

We next sketch the proof idea for Theorem 5 for the case with universal role in interpolants. We show “1. \Rightarrow 2.”, observe that “3. \Rightarrow 1.” is trivial, and then sketch the proof of “2. \Rightarrow 3.” and the exponential time algorithm computing interpolants, details are provided in the appendix of the full version. For “1. \Rightarrow 2.” assume that C is an $\mathcal{EL}\mathcal{O}_u(\Sigma)$ -concept with (i) $\mathcal{O}_1 \cup \mathcal{O}_2 \models A \sqsubseteq C$ and (ii) $\mathcal{O}_1 \cup \mathcal{O}_2 \models C \sqsubseteq B$. By (†) and (i), $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma} \models C(\rho_A)$. But then by (ii) $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma} \models B(\rho_A)$, as required.

If one does not impose a bound on the size of \mathcal{A} in Point 3, then one can prove “2. \Rightarrow 3.” using compactness and a generalization of *unraveling tolerance* according to which $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma}$ and $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma,u}$ entail the same $C(\rho_A)$ (Lutz and Wolter 2017; Hernich et al. 2020). As we are interested in an exponential bound on the size of \mathcal{A} (and a deterministic exponential time algorithm computing it) we require a more syntactic approach. Our proof of “2. \Rightarrow 3.” is based on *derivation trees* which represent a derivation of a fact $C(a)$ from an ontology \mathcal{O} and ABox \mathcal{A} using a labeled tree. Our derivation trees generalize those introduced in (Bienvenu, Lutz, and Wolter 2013; Baader et al. 2016) to languages with nominals and role inclusions. Reflecting the use of individual names and concept names in the construction of the domain of the canonical model (Baader, Brandt, and Lutz 2005), we assume $a \in \Delta := \text{ind}(\mathcal{A}) \cup ((\mathbf{N}_C \cup \mathbf{N}_I) \cap \text{sig}(\mathcal{O}))$ and $C \in \Theta := \{\top\} \cup (\mathbf{N}_C \cap \text{sig}(\mathcal{O})) \cup \{\{a\} \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})\}$. Then a derivation tree (T, V) for $(a, C) \in \Delta \times \Theta$ is a tree T with a labeling function $V : T \rightarrow \Delta \times \Theta$ such that $V(\varepsilon) = (a, C)$ and (V, T) satisfies rules stating under which conditions the label of n is derived in one step from the labels of the successors of n . To illustrate, the existence of successors n_1, n_2 of n with $V(n_1) = (a, C_1)$ and $V(n_2) = (a, C_2)$ justifies $V(n) = (a, C)$ if $\mathcal{O} \models C_1 \sqcap C_2 \sqsubseteq C$. The rules are given in the appendix of the full version, we only discuss the rule used to capture derivations using RIs: $V(n) = (a_1, C)$ is justified if there are role names r_2, \dots, r_{2k-2}, r such that (a_{2k}, C') is a label of a successor of n , $\mathcal{O} \models \exists r.C' \sqsubseteq C$, $\mathcal{O} \models r_2 \circ \dots \circ r_{2k-2} \sqsubseteq r$, and the situation depicted in Figure 4 holds, where the “dotted lines” stand for ‘either $a_i = a_{i+1}$ or some $(a_i, \{c\}), (a_{i+1}, \{c\})$ with $c \in \mathbf{N}_I$ are labels of successors of n ’, and \hat{r}_i stands for ‘either $r(a_i, a_{i+1}) \in \mathcal{A}$ or some (a_i, C_i) is a label of a successor of n and $\mathcal{O} \models C_i \sqsubseteq \exists r_i.\{a_{i+1}\}$ if $a_{i+1} \in \mathbf{N}_I$ and $\mathcal{O} \models C_i \sqsubseteq \exists r_i.a_{i+1}$ if $a_{i+1} \in \mathbf{N}_C$ ’. Moreover, for all $a_i \neq a_1, 1 \leq i \leq 2k$, there exists a successor of n with label (a_i, D) for some D . The soundness of this rule should be clear, completeness can be shown similarly to the analysis

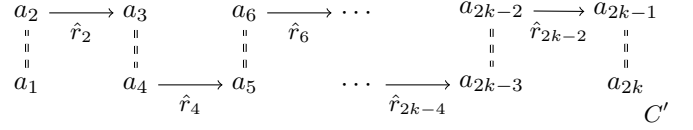


Figure 4: Rule for Role Inclusions.

of canonical models.

The length of the sequence a_1, \dots, a_{2k} can be exponential (for instance, in Example 1 for the fact (ρ_A, A) in $\mathcal{O}_p, \mathcal{A}_{\mathcal{O}_p, A}^{\Sigma_1}$). One can show, however, that its length can be bounded without affecting completeness by $2^{q(\|\mathcal{O}\| + \|\mathcal{A}\|)}$ with q a polynomial. The following lemma summarizes the main properties of derivation trees.

Lemma 3. *Let \mathcal{O} be an $\mathcal{EL}\mathcal{RO}_u$ -ontology in normal form and \mathcal{A} a finite $\text{sig}(\mathcal{O})$ -ABox. Then*

1. $\mathcal{O}, \mathcal{A} \models A(x)$ if and only if there is a derivation tree for $A(x)$ in \mathcal{O}, \mathcal{A} . Moreover, if a derivation tree exists, then there exists one of depth and outdegree bounded by $(\|\mathcal{A}\| + \|\mathcal{O}\|) \times \|\mathcal{O}\|$ which can be constructed in exponential time in $\|\mathcal{O}\| + \|\mathcal{A}\|$.
2. If (T, V) is a derivation tree for $A(x)$ in \mathcal{O}, \mathcal{A} of at most exponential size, then one can construct in exponential time (in $\|\mathcal{A}\| + \|\mathcal{O}\|$) a derivation tree (T', V') for $A(x)$ in $\mathcal{O}, \mathcal{A}^u$ with \mathcal{A}^u the directed unfolding of \mathcal{A} modulo $\Sigma = \text{sig}(\mathcal{A}) \cap \mathbf{N}_I$ and T' of the same depth as T and such that the outdegree of T' does not exceed $\max\{3, 3n\}$ with n the length of the longest chain $a_1 \dots a_n$ used in the rule for RIs in the derivation tree (T, V) .

Proof. We sketch the idea. For Point 1, the bound on the depth of derivation trees can be proved by observing that one can assume (using a standard pumping argument) that the labels of distinct nodes on a single path are distinct and the bound on the outdegree can be proved by observing that one can trivially assume that all successor nodes of a node have distinct labels. For the construction of derivation trees, let F_n denote the set of facts in $\Delta \times \Theta$ for which there is a derivation tree of depth at most n . Then one can construct in exponential time derivation trees for all facts in any F_n , $n \leq (\|\mathcal{A}\| + \|\mathcal{O}\|) \times \|\mathcal{O}\|$ by starting with derivation trees of depth 0 for members of F_0 , and then constructing derivation trees of depth $i + 1$ for members of F_{i+1} using the trees for members of F_0, \dots, F_i . For Point 2, the transformation of (T, V) into (T', V') is by induction over rule application, the only interesting step being the rule for RIs. Using the ontology \mathcal{O}_p of Example 1 one can see that the exponential blow-up of the outdegree is unavoidable. \square

We are now in the position to complete the sketch of the proof of “2. \Rightarrow 3.” Assume that Point 2 holds. Then $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma} \models B(\rho_A)$. By Point 1 of Lemma 3 we can construct a derivation tree (T, V) for (ρ_A, B) in $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma}$ of polynomial depth and outdegree in exponential time. By Point 2 of Lemma 3 we can transform (T, V) into a derivation tree (T', V') for (ρ_A, B) in $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma,u}$ in exponential time. Now let \mathcal{A} be the

restriction of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, u}$ to all $x \in \text{ind}(\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, u})$ which occur in a label of V' . Then (T', V') is also a derivation tree for (ρ_A, B) in $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}$ and so $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A} \models B(\rho_A)$. It follows that the $\mathcal{EL}\mathcal{O}_u(\Sigma)$ -concept corresponding to \mathcal{A} is an interpolant for $A \sqsubseteq B$ under $\mathcal{O}_1 \cup \mathcal{O}_2$. Its size is at most exponential in $\|\mathcal{O}_1 \cup \mathcal{O}_2\|$ since (T', V') is at most exponential in $\|\mathcal{O}_1 \cup \mathcal{O}_2\| + \|\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma}\|$, and so also in $\|\mathcal{O}_1 \cup \mathcal{O}_2\|$.

7 Interpolant and Explicit Definition Existence in $\mathcal{EL}\mathcal{I}$ and Extensions

We analyze interpolants and explicit definitions for $\mathcal{EL}\mathcal{I}$ and its extensions with nominals and universal roles, and show the following result from the introduction.

Theorem 3. *For $\mathcal{EL}\mathcal{I}$ and any extension with any combination of nominals, the universal role, or \perp , the existence of interpolants and explicit definitions is EXPTIME-complete. If an interpolant/explicit definition exists, then there exists one of at most double exponential size that can be computed in double exponential time. This bound is optimal.*

The double exponential lower bound on the size of explicit definitions and interpolants is shown in the appendix of the full version. The proof is inspired by similar lower bounds for the size of FO-rewritings and uniform interpolants (Lutz and Wolter 2010; Nikitina and Rudolph 2014). To prove the remaining claims of Theorem 3, we lift Theorem 5 to $\mathcal{EL}\mathcal{I}$. The main differences are that (1) we now associate undirected graphs with ABoxes and also unfold along inverse roles; (2) that canonical models become potentially infinite but tree-shaped; (3) that therefore deciding the new variant of Point 2 of Theorem 5 is not an instance of standard entailment checking in $\mathcal{EL}\mathcal{I}$, instead we give a reduction to emptiness checking for tree automata; and (4) that to bound the size of \mathcal{A} in Point 3, we employ transfer sequences (and not derivation trees) to represent how facts are derived.

In more detail, associate with every ABox \mathcal{A} the undirected graph $G_{\mathcal{A}}^u = (\text{ind}(\mathcal{A}), \bigcup_{r \in \text{NR}} \{\{x, y\} \mid r(x, y) \in \mathcal{A}\})$. We say that \mathcal{A} is *tree-shaped* if $G_{\mathcal{A}}^u$ is acyclic, $r(x, y) \in \mathcal{A}$ and $s(x, y) \in \mathcal{A}$ imply $r = s$, and $r(x, y) \in \mathcal{A}$ implies $s(y, x) \notin \mathcal{A}$ for any s . \mathcal{A} is *tree-shaped modulo a set Γ of individual names* if after dropping some facts $r(x, y)$ with $\{a\}(x)$ or $\{a\}(y) \in \mathcal{A}$ for some $a \in \Gamma$ it is tree-shaped. We observe that $\mathcal{EL}\mathcal{I}\mathcal{O}_u(\Sigma)$ -concepts correspond to pointed Σ -ABoxes \mathcal{A}, x such that \mathcal{A} is tree-shaped modulo $\text{N}_1 \cap \Sigma$. $\mathcal{EL}\mathcal{I}\mathcal{O}(\Sigma)$ -concepts correspond to *weakly rooted* pointed Σ -ABoxes \mathcal{A}, x such that \mathcal{A} is tree-shaped modulo $\text{N}_1 \cap \Sigma$, where \mathcal{A}, x is called weakly rooted if for every $y \in \text{ind}(\mathcal{A})$ there is a path from x to y in $G_{\mathcal{A}}^u$.

For every $\mathcal{EL}\mathcal{I}\mathcal{O}_u$ -ontology \mathcal{O} and concept A there exists a (potentially infinite) pointed canonical model $\mathcal{I}_{\mathcal{O}, A}, \rho_A$ such that the ABox $\mathcal{A}_{\mathcal{O}, A}$ corresponding to $\mathcal{I}_{\mathcal{O}, A}$ is tree-shaped modulo $\text{N}_1 \cap \text{sig}(\mathcal{O})$. The property (\dagger) used in the context of canonical models for tractable extensions of \mathcal{EL} holds here as well. We also require the *undirected unfolding* of a pointed Σ -ABox \mathcal{A}, x into a pointed Σ -ABox \mathcal{A}^*, x which is tree-shaped modulo $\Sigma \cap \text{N}_1$. In the *rooted undirected unfolding*, nodes that cannot be reached from x via

roles are dropped.

Assume now that \mathcal{O} is in normal form and A a concept name. Let $\mathcal{A}_{\mathcal{O}, A}^{\Sigma}$ be the Σ -reduct of the canonical model $\mathcal{I}_{\mathcal{O}, A}$, regarded as an ABox. Denote by $\mathcal{A}_{\mathcal{O}, A}^{\Sigma, *}, \rho_A$ the undirected unfolding of $\mathcal{A}_{\mathcal{O}, A}^{\Sigma}, \rho_A$, by $\mathcal{A}_{\mathcal{O}, A}^{\downarrow w \Sigma}, \rho_A$ the sub-ABox of $\mathcal{A}_{\mathcal{O}, A}^{\Sigma}$ weakly rooted in ρ_A , and by $\mathcal{A}_{\mathcal{O}, A}^{\downarrow w \Sigma, *}, \rho_A$ its rooted undirected unfolding. Then we lift Theorem 5 as follows.

Theorem 6. *There exists a polynomial p such that the following conditions are equivalent for all $\mathcal{EL}\mathcal{I}\mathcal{O}_u$ -ontologies $\mathcal{O}_1, \mathcal{O}_2$ in normal form, concept names A, B , and $\Sigma = \text{sig}(\mathcal{O}_1, A) \cap \text{sig}(\mathcal{O}_2, B)$:*

1. *An $\mathcal{EL}\mathcal{I}\mathcal{O}_u$ -interpolant for $A \sqsubseteq B$ under $\mathcal{O}_1, \mathcal{O}_2$ exists;*
2. *$\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma} \models B(\rho_A)$;*
3. *there exists a finite subset \mathcal{A} of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, *}$ with $|\text{ind}(\mathcal{A})| \leq 2^{2^{p(\|\mathcal{O}_1 \cup \mathcal{O}_2\|)}}$ such that the $\mathcal{EL}\mathcal{I}\mathcal{O}_u$ -concept corresponding to \mathcal{A}, ρ_A is an $\mathcal{EL}\mathcal{I}\mathcal{O}_u$ -interpolant for $A \sqsubseteq B$ under $\mathcal{O}_1, \mathcal{O}_2$.*

The same equivalences hold if in Points 1 to 3, $\mathcal{EL}\mathcal{I}\mathcal{O}_u$ is replaced by $\mathcal{EL}\mathcal{I}\mathcal{O}$, $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma}$ by $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\downarrow w \Sigma}$, and $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, *}$ by $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\downarrow w \Sigma, *}$.

In Point 3, \mathcal{A} can be computed in double exponential time, if it exists.

We first sketch how tree automata are used to show that Point 2 entails an exponential time upper bound for deciding the existence of an interpolant. To this end we represent finite prefix-closed subsets \mathcal{A} of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma}$ as trees and design

- a non-deterministic tree automaton over finite trees (NTA), \mathfrak{A}_1 , that accepts exactly those trees that represent prefix-closed finite subsets of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma}$;
- a two-way alternating tree automaton over finite trees (2ATA), \mathfrak{A}_2 , that accepts exactly those trees that represent a pointed ABox \mathcal{A}, ρ with $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A} \models B(\rho)$.

Similar tree automata techniques have been used e.g. in (Jung et al. 2020). \mathfrak{A}_1 is constructed using the definition of canonical models; its states are essentially types occurring in the canonical model and it can be constructed in exponential time. The 2ATA \mathfrak{A}_2 tries to construct a derivation tree for $B(\rho)$ in $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}$, given as input a tree representing \mathcal{A}, ρ . It has polynomially many states, and can thus be turned into an equivalent NTA with exponentially many states (Vardi 1998). By taking the intersection with \mathfrak{A}_1 , one can then check in exponential time whether $L(\mathfrak{A}_1) \cap L(\mathfrak{A}_2) \neq \emptyset$, that is, whether $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma} \models B(\rho_A)$.

We return to the proof of Theorem 6. The interesting implication is “2. \Rightarrow 3.” and the double exponential computation of interpolants. In this case we use transfer sequences to obtain a bound on the size of the subset \mathcal{A} of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, *}$ needed to derive $B(\rho_A)$ (we note that for $\mathcal{EL}\mathcal{I}$ without nominals one can also use the automata encoding above). Transfer sequences describe how facts are derived in a tree-shaped ABox and allow to determine when individuals a and b behave sufficiently similar so that the subtree rooted at a can

be replaced by the subtree rooted at b (Bienvenu, Lutz, and Wolter 2013) without affecting a derivation. This technique can be used to show that one can always choose a prefix closed subset \mathcal{A} of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^{\Sigma, *}$ of at most exponential depth. This also implies that \mathcal{A} can be obtained in double exponential time by constructing the canonical model up to depth $2^{q(\|\mathcal{O}_1 \cup \mathcal{O}_2\|)}$ with q a polynomial.

8 Expressive Horn Description Logics

We address two questions regarding expressive Horn-DLs. (1) Can our results for \mathcal{ELI} and extensions be lifted to more expressive Horn-DLs? (2) In the examples provided in the proof of Theorem 1 we sometimes (for example, for \mathcal{EL}_u and \mathcal{ELI}) construct explicit Horn-DL definitions to show implicit definability of concept names. Are Horn-DL concepts always sufficient to obtain an explicit definition if an implicit definition exists? We provide a positive answer to (1) if one only admits $\mathcal{ELI}\mathcal{O}_u$ -concepts (or fragments) as interpolants/explicit definitions and a negative answer to (2) in the sense that \mathcal{ELI} and various other Horn-DLs do not enjoy the CIP/PBDP even if one admits Horn-DL concepts as interpolants/explicit definitions.

We introduce expressive Horn DLs (Hustadt, Motik, and Sattler 2005), presented here in the form proposed in (Lutz and Wolter 2012). *Horn- $\mathcal{ALCI}\mathcal{O}_u$ -concepts* R and *Horn- $\mathcal{ALCI}\mathcal{O}_u$ -CIs* $L \sqsubseteq R$ are defined by the syntax rules

$$R, R' ::= \top \mid \perp \mid A \mid \neg A \mid \{a\} \mid \neg\{a\} \mid R \sqcap R' \mid L \rightarrow R \mid \exists r.R \mid \forall r.R$$

$$L, L' ::= \top \mid \perp \mid A \mid L \sqcap L' \mid L \sqcup L' \mid \exists r.L$$

with A ranging over concept names, a over individual names, and r over roles (including the universal role). As usual, the fragment of Horn- $\mathcal{ALCI}\mathcal{O}_u$ without nominals and the universal role is denoted by Horn- \mathcal{ALCI} and Horn- \mathcal{ALC} denotes the fragment of Horn- \mathcal{ALCI} without inverse roles.

Theorem 7. *Let $(\mathcal{L}, \mathcal{L}')$ be the pair (Horn- \mathcal{ALCI} , \mathcal{ELI}) or the pair (Horn- $\mathcal{ALCI}\mathcal{O}_u$, $\mathcal{ELI}\mathcal{O}_u$). Then*

- *deciding the existence of an \mathcal{L}' -interpolant for an \mathcal{L}' -CI $C \sqsubseteq D$ under \mathcal{L} -ontologies $\mathcal{O}_1, \mathcal{O}_2$ is EXPTIME-complete;*
- *deciding the existence of an explicit $\mathcal{L}'(\Sigma)$ -definition of a concept name A under an \mathcal{L} -ontology \mathcal{O} is EXPTIME-complete.*

Moreover, if an \mathcal{L} -interpolant/explicit definition exists, then there exists one of at most double exponential size that can be computed in double exponential time.

Theorem 7 follows from Theorem 3 and the fact that for any \mathcal{L} -ontology one can construct in polynomial time an \mathcal{L}' -ontology in normal form that is a conservative extension of \mathcal{L} (see (Bienvenu et al. 2016) for a similar result). We next show that despite the fact that Horn- \mathcal{ALCI} -concepts sometimes provide explicit definitions if none exist in \mathcal{ELI} (proof of Theorem 1), they are not sufficient to prove the CIP/PBDP.

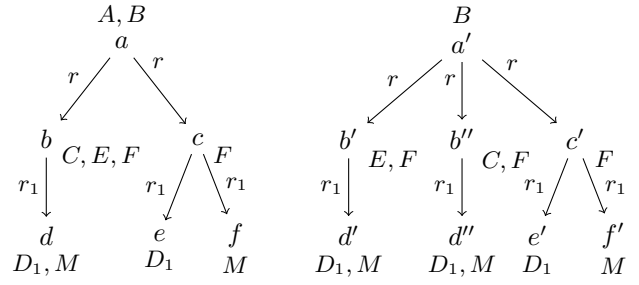


Figure 5: Interpretations \mathcal{I} (left) and \mathcal{I}' (right).

Theorem 8. *There exists an ontology \mathcal{O} in Horn- \mathcal{ALC} (and in \mathcal{ELI}), a signature Σ , and a concept name A such that A is implicitly definable using Σ under \mathcal{O} but does not have an explicit Horn- $\mathcal{ALCI}\mathcal{O}_u(\Sigma)$ -definition.*

Proof. We modify the ontology used in the proof of Point 1 of Theorem 1. Let $\Sigma = \{B, D_1, E, r, r_1\}$ and let \mathcal{O} contain $B \sqcap \exists r.(C \sqcap E) \sqsubseteq A$ and the following CIs:

$$\begin{aligned} A &\sqsubseteq B, & B &\sqsubseteq \forall r.F, & B &\sqsubseteq \exists r.C, & C &\sqsubseteq F \sqcap \forall r_1.D_1, \\ F &\sqsubseteq \exists r_1.D_1 \sqcap \exists r_1.M, \\ A &\sqsubseteq \forall r.((F \sqcap \exists r_1.(D_1 \sqcap M)) \rightarrow E). \end{aligned}$$

Intuitively, the final two CIs should be read as

$$\begin{aligned} F &\sqsubseteq \exists r_1.D_1 \\ A &\sqsubseteq \forall r.((F \sqcap \forall r_1.D_1) \rightarrow E) \end{aligned}$$

and the concept name M is introduced to achieve this in a projective way as the latter CI is not in Horn- \mathcal{ALCI} .

A is implicitly definable using Σ under \mathcal{O} since

$$\mathcal{O} \models A \equiv B \sqcap \forall r.(\forall r_1.D_1 \rightarrow E).$$

To show that A is not explicitly Horn- $\mathcal{ALCI}\mathcal{O}_u(\Sigma)$ -definable under \mathcal{O} consider the interpretations \mathcal{I} and \mathcal{I}' in Figure 5. The claim follows from the facts that \mathcal{I} and \mathcal{I}' are models of \mathcal{O} , $a \in A^{\mathcal{I}}$, $a' \notin A^{\mathcal{I}'}$, but $a \in F^{\mathcal{I}}$ implies $a' \in F^{\mathcal{I}'}$ holds for every Horn- $\mathcal{ALCI}\mathcal{O}_u(\Sigma)$ -concept F . The latter can be proved by observing that there exists a Horn- $\mathcal{ALCI}\mathcal{O}_u(\Sigma)$ -simulation between \mathcal{I} and \mathcal{I}' (Jung et al. 2019) containing $(\{a\}, a)$, we refer the reader to the appendix of the full version. To obtain an example in \mathcal{ELI} , it suffices to take a conservative extension of \mathcal{O} in \mathcal{ELI} . \square

9 Discussion

For a few important extensions of $\mathcal{EL}/\mathcal{ELI}$ the complexity of interpolant and explicit definition existence remains to be investigated. Examples include extensions of \mathcal{ELI} with role inclusions, and extensions of \mathcal{EL} or \mathcal{ELI} with functional roles or more general number restrictions. It would also be of interest to investigate interpolant existence if Horn-concepts are admitted as interpolants (using, for example, the games introduced in (Jung et al. 2019)). Finally, the question arises whether there exists at all a decidable Horn language extending, say, Horn- \mathcal{ALCI} , with the CIP/PBDP. We note that Horn-FO enjoys the CIP (Exercise

6.2.6 in (Chang and Keisler 1998)) but is undecidable and that we show in the appendix of the full version that the Horn fragment of the guarded fragment does not enjoy the CIP/PBDP.

Acknowledgments

This research was supported by the EPSRC UK grant EP/S032207/1.

References

- Areces, C.; Blackburn, P.; and Marx, M. 2001. Hybrid logics: Characterization, interpolation and complexity. *J. Symb. Log.* 66(3):977–1010.
- Artale, A.; Jung, J. C.; Mazzullo, A.; Ozaki, A.; and Wolter, F. 2021a. Living without Beth and Craig: Explicit definitions and interpolants in description logics with nominals and role hierarchies. In *Proc. of AAAI*.
- Artale, A.; Mazzullo, A.; Ozaki, A.; and Wolter, F. 2021b. On free description logics with definite descriptions. In *Proc. of KR*.
- Baader, F.; Bienvenu, M.; Lutz, C.; and Wolter, F. 2016. Query and predicate emptiness in ontology-based data access. *J. Artif. Intell. Res.* 56:1–59.
- Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.
- Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the \mathcal{EL} envelope. In *Proc. of IJCAI*, 364–369.
- Benedikt, M.; Leblay, J.; ten Cate, B.; and Tsamoura, E. 2016. *Generating Plans from Proofs: The Interpolation-based Approach to Query Reformulation*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers.
- Benedikt, M.; Kostylev, E. V.; Mogavero, F.; and Tsamoura, E. 2017. Reformulating queries: Theory and practice. In *IJCAI*, 837–843.
- Bienvenu, M.; Hansen, P.; Lutz, C.; and Wolter, F. 2016. First order-rewritability and containment of conjunctive queries in horn description logics. In *IJCAI*, 965–971.
- Bienvenu, M.; Lutz, C.; and Wolter, F. 2013. First-order rewritability of atomic queries in Horn description logics. In *Proc. of IJCAI*.
- Borgida, A.; Toman, D.; and Weddell, G. E. 2016. On referring expressions in query answering over first order knowledge bases. In *Proc. of KR*, 319–328.
- Chang, C., and Keisler, H. J. 1998. *Model Theory*. Elsevier.
- Deutsch, A.; Popa, L.; and Tannen, V. 2006. Query reformulation with constraints. *SIGMOD Rec.* 35(1):65–73.
- Geleta, D.; Payne, T. R.; and Tamma, V. A. M. 2016. An investigation of definability in ontology alignment. In Blomqvist, E.; Ciancarini, P.; Poggi, F.; and Vitali, F., eds., *Proc. of EKAW*, 255–271.
- Hernich, A.; Lutz, C.; Papacchini, F.; and Wolter, F. 2020. Dichotomies in ontology-mediated querying with the guarded fragment. *ACM Trans. Comput. Log.* 21(3):20:1–20:47.
- Hustadt, U.; Motik, B.; and Sattler, U. 2005. Data complexity of reasoning in very expressive description logics. In *IJCAI*, 466–471.
- Jung, J. C., and Wolter, F. 2021. Living without Beth and Craig: Definitions and interpolants in the guarded and two-variable fragments. In *Proc. of LICS*.
- Jung, J. C.; Papacchini, F.; Wolter, F.; and Zakharyashev, M. 2019. Model comparison games for horn description logics. In *Proc. of LICS*, 1–14. IEEE.
- Jung, J. C.; Lutz, C.; Martel, M.; and Schneider, T. 2020. Conservative extensions in horn description logics with inverse roles. *J. Artif. Intell. Res.* 68:365–411.
- Konev, B.; Lutz, C.; Walther, D.; and Wolter, F. 2009. Formal properties of modularisation. In *Modular Ontologies*, volume 5445 of *Lecture Notes in Computer Science*. Springer. 25–66.
- Konev, B.; Lutz, C.; Ponomaryov, D. K.; and Wolter, F. 2010. Decomposing description logic ontologies. In *Proc. of KR*. AAAI Press.
- Koopmann, P., and Schmidt, R. A. 2015. Uniform interpolation and forgetting for ALC ontologies with aboxes. In *Proc. of AAAI*, 175–181. AAAI Press.
- Lutz, C., and Wolter, F. 2010. Deciding inseparability and conservative extensions in the description logic EL. *J. Symb. Comput.* 45(2):194–228.
- Lutz, C., and Wolter, F. 2011. Foundations for uniform interpolation and forgetting in expressive description logics. In *Proc. of IJCAI*, 989–995. IJCAI/AAAI.
- Lutz, C., and Wolter, F. 2012. Non-uniform data complexity of query answering in description logics. In Brewka, G.; Eiter, T.; and McIlraith, S. A., eds., *Proc. of KR*.
- Lutz, C., and Wolter, F. 2017. The data complexity of description logic ontologies. *Logical Methods in Computer Science* 13(4).
- Lutz, C.; Piro, R.; and Wolter, F. 2011. Description logic tboxes: Model-theoretic characterizations and rewritability. In Walsh, T., ed., *Proc. of IJCAI*, 983–988. IJCAI/AAAI.
- Lutz, C.; Seylan, I.; and Wolter, F. 2012. An automata-theoretic approach to uniform interpolation and approximation in the description logic EL. In *Proc. of KR*. AAAI Press.
- Lutz, C.; Seylan, I.; and Wolter, F. 2019. The data complexity of ontology-mediated queries with closed predicates. *Logical Methods in Computer Science* 15(3).
- Maksimova, L., and Gabbay, D. 2005. *Interpolation and Definability in Modal and Intuitionistic Logics*. Clarendon Press.
- Nikitina, N., and Rudolph, S. 2014. (Non-)succinctness of uniform interpolants of general terminologies in the description logic EL. *Artif. Intell.* 215:120–140.
- Place, T., and Zeitoun, M. 2016. Separating regular languages with first-order logic. *Log. Methods Comput. Sci.* 12(1).
- Seylan, I.; Franconi, E.; and de Bruijn, J. 2009. Effective

query rewriting with ontologies over dboxes. In *Proc. of IJCAI*, 923–925.

Sofronie-Stokkermans, V. 2008. Interpolation in local theory extensions. *Log. Methods Comput. Sci.* 4(4).

ten Cate, B.; Conradie, W.; Marx, M.; and Venema, Y. 2006. Definitorially complete description logics. In *Proc. of KR*, 79–89. AAAI Press.

ten Cate, B.; Franconi, E.; and Seylan, İ. 2013. Beth definability in expressive description logics. *J. Artif. Intell. Res.* 48:347–414.

ten Cate, B. 2005. Interpolation for extended modal languages. *J. Symb. Log.* 70(1):223–234.

Toman, D., and Weddell, G. E. 2011. *Fundamentals of Physical Design and Query Compilation*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers.

Toman, D., and Weddell, G. E. 2021. FO rewritability for OMQ using beth definability and interpolation. In Homola, M.; Ryzhikov, V.; and Schmidt, R. A., eds., *Proc. of DL*. CEUR-WS.org.

Vardi, M. Y. 1998. Reasoning about the past with two-way automata. In *Proc. of ICALP'98*, 628–641.

A Further Preliminaries

We call an ontology \mathcal{O}' a *conservative extension* of an ontology \mathcal{O} if $\mathcal{O}' \models \alpha$ for all $\alpha \in \mathcal{O}$ and every model \mathcal{I} of \mathcal{O} can be expanded to a model \mathcal{J} of \mathcal{O}' by modifying the interpretation of symbols in $\text{sig}(\mathcal{O}') \setminus \text{sig}(\mathcal{O})$. In other words, the $\text{sig}(\mathcal{O})$ -reducts of \mathcal{I} and \mathcal{J} coincide. The following result is folklore (Baader, Brandt, and Lutz 2005).

Lemma 4. *Let \mathcal{L} be any DL from $\mathcal{EL}, \mathcal{ELI}, \mathcal{EL\mathcal{O}}, \mathcal{ELRO}, \mathcal{ELIO}$ or an extension with the universal role, and let \mathcal{O} be an \mathcal{L} -ontology. Then one can construct in polynomial time an \mathcal{L} -ontology \mathcal{O}' in normal form such that \mathcal{O}' is a conservative extension of \mathcal{O} .*

We next give a more detailed introduction to ABoxes and how they relate to concepts. Recall that an ABox \mathcal{A} is a (possibly infinite) set of assertions of the form $A(x)$, $r(x, y)$, $\{a\}(x)$, and $\top(x)$ with $A \in \mathbf{N}_C$, $r \in \mathbf{N}_R$, $a \in \mathbf{N}_I$, and x, y individual variables. An ABox is *factorized* if $\{a\}(x), \{a\}(y) \in \mathcal{A}$ imply $x = y$.

ABox assertions are interpreted in an interpretation \mathcal{I} using a *variable assignment* v that maps individual variables to elements of $\Delta^{\mathcal{I}}$. Then \mathcal{I}, v satisfies an assertion $A(x)$ if $v(x) \in A^{\mathcal{I}}$, $r(x, y)$ if $(v(x), v(y)) \in r^{\mathcal{I}}$, $\{a\}(x)$ if $a^{\mathcal{I}} = v(x)$, and $\top(x)$ is always satisfied. \mathcal{I}, v satisfies an ABox if it satisfies all assertions in it. We write $\mathcal{I} \models \mathcal{A}[x \mapsto d]$ if there exists an assignment v with $v(x) = d$ such that \mathcal{I}, v satisfies \mathcal{A} . We say that an assertion $A_0(x_0)$ is *entailed by an ontology \mathcal{O} and ABox \mathcal{A}* , in symbols $\mathcal{O}, \mathcal{A} \models A_0(x_0)$, if $v(x) \in A_0^{\mathcal{I}}$ for all models \mathcal{I} of \mathcal{O} and assignments v such that \mathcal{I}, v satisfy \mathcal{A} . This is the standard notion of entailment from a knowledge base consisting of an ontology and an ABox. Deciding entailment is in PTIME for the DLs between \mathcal{EL} and \mathcal{EL}_u^{++} (Baader, Brandt, and Lutz 2005) and EXPTIME-complete for the DLs between \mathcal{ELI} and \mathcal{ELIO}_u (Baader et al. 2017).

Every interpretation \mathcal{I} defines a factorized ABox $\mathcal{A}_{\mathcal{I}}$ by identifying every $d \in \Delta^{\mathcal{I}}$ with a variable x_d and taking $A(x_d)$ if $d \in A^{\mathcal{I}}$, $r(x_c, x_d)$ if $(c, d) \in r^{\mathcal{I}}$, $\{a\}(x_d)$ if $a^{\mathcal{I}} = d$. Conversely, factorized ABoxes define interpretations in the obvious way.

The following lemma provides a formal description of the relationship between ABoxes that are ditree-shaped modulo some set of individual names and $\mathcal{EL\mathcal{O}}$ -concepts.

Lemma 5. *For any $\mathcal{EL\mathcal{O}}_u(\Sigma)$ -concept C one can construct in polynomial time a pointed Σ -ABox \mathcal{A}, x such that \mathcal{A} is ditree-shaped modulo $\mathbf{N}_I \cap \Sigma$ and $d \in C^{\mathcal{I}}$ iff $\mathcal{I} \models \mathcal{A}[x \mapsto d]$, for all interpretations \mathcal{I} and $d \in \Delta^{\mathcal{I}}$.*

Conversely, for any pointed Σ -ABox \mathcal{A}, x such that \mathcal{A} is a ditree-shaped ABox modulo Γ , one can construct in polynomial time an $\mathcal{EL\mathcal{O}}_u(\Sigma)$ -concept C such that $\Gamma = \mathbf{N}_I \cap \Sigma$ and $d \in C^{\mathcal{I}}$ iff $\mathcal{I} \models \mathcal{A}_C[x \mapsto d]$, for all interpretations \mathcal{I} and $d \in \Delta^{\mathcal{I}}$.

The above also holds if one replaces $\mathcal{EL\mathcal{O}}_u(\Sigma)$ -concepts by $\mathcal{EL\mathcal{O}}(\Sigma)$ -concepts and requires the pointed ABoxes to be rooted.

We define a *canonical model* $\mathcal{I}_{\mathcal{O}, A_0}$ for an \mathcal{ELRO}_u -ontology \mathcal{O} in normal form and a concept name A_0 . This has been done in (Baader, Brandt, and Lutz 2005), but as

we do not use canonical models for subsumption or instance checking we give a succinct model-theoretic construction.

Assume \mathcal{O} and A_0 are given and \mathcal{O} is in normal form. Define an equivalence relation \sim on the set of individual names a in $\text{sig}(\mathcal{O})$ by setting $a \sim b$ if $\mathcal{O} \models \exists u. A_0 \sqcap \{a\} \sqsubseteq \{b\}$. Let $[a] = \{b \in \text{sig}(\mathcal{O}) \mid a \sim b\}$ and set $\Delta_I = \{\{a\} \mid a \in \text{sig}(\mathcal{O})\}$. Say that a concept name A is *absorbed* by an individual name a if $\mathcal{O} \models \exists u. A_0 \sqcap A \sqsubseteq \{a\}$ and let Δ_C denote the set of concept names A in \mathcal{O} such that $\mathcal{O} \models A_0 \sqsubseteq \exists u. A$ and A is not absorbed by any individual name.

Now let $\Delta^{\mathcal{I}_{\mathcal{O}, A_0}} = \Delta_I \cup \Delta_C$ and let

$$\begin{aligned} A^{\mathcal{I}_{\mathcal{O}, A_0}} &= \{[a] \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \mid \mathcal{O} \models \exists u. A_0 \sqcap \{a\} \sqsubseteq A\} \cup \\ &\quad \{B \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \mid \mathcal{O} \models \exists u. A_0 \sqcap B \sqsubseteq A\} \\ a^{\mathcal{I}_{\mathcal{O}, A_0}} &= [a] \\ r^{\mathcal{I}_{\mathcal{O}, A_0}} &= \{([a], [b]) \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \times \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \mid \\ &\quad \mathcal{O} \models \exists u. A_0 \sqcap \{a\} \sqsubseteq \exists r. \{b\}\} \cup \\ &\quad \{([a], B) \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \times \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \mid \\ &\quad \mathcal{O} \models \exists u. A_0 \sqcap \{a\} \sqsubseteq \exists r. B\} \cup \\ &\quad \{(B, [a]) \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \times \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \mid \\ &\quad \mathcal{O} \models \exists u. A_0 \sqcap B \sqsubseteq \exists r. \{a\}\} \cup \\ &\quad \{(A, B) \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \times \Delta^{\mathcal{I}_{\mathcal{O}, A_0}} \mid \\ &\quad \mathcal{O} \models \exists u. A_0 \sqcap A \sqsubseteq \exists r. B\} \end{aligned}$$

for every concept name $A \in \text{N}_C$, $a \in \text{sig}(\mathcal{O}) \cap \text{N}_I$, and $r \in \text{N}_R$. We often denote the nodes $[a]$ and A by $\rho_{[a]}$ or, for simplicity, ρ_a and, respectively, ρ_A . If A_0 is absorbed by an individual a we still often denote $\rho_{[a]}$ by ρ_{A_0} .

Lemma 6. *The canonical model $\mathcal{I}_{\mathcal{O}, A_0}$ is a model of \mathcal{O} and for every model \mathcal{J} of \mathcal{O} and any $d \in \Delta^{\mathcal{J}}$ with $d \in A_0^{\mathcal{J}}$, $(\mathcal{I}_{\mathcal{O}, A_0}, \rho_{A_0}) \preceq_{\mathcal{E}\mathcal{L}\mathcal{O}_u, \Sigma} (\mathcal{J}, d)$, where Σ is any signature.*

Proof. We first show that $\mathcal{I}_{\mathcal{O}, A_0}$ is a model of \mathcal{O} . It is straightforward to show that $\mathcal{I}_{\mathcal{O}, A_0}$ satisfies the CIs of the form $\top \sqsubseteq A$, $A_1 \sqcap A_2 \sqsubseteq A$, $A \sqsubseteq \{a\}$, $\{a\} \sqsubseteq A$.

Assume now that $A \sqsubseteq \exists r. B \in \mathcal{O}$ and $\rho_C \in A^{\mathcal{I}_{\mathcal{O}, A_0}}$ with C of the form a or A . We have $\mathcal{O} \models \exists u. A_0 \sqsubseteq \exists u. C$, $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq A$. Thus $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq \exists r. B$. But then $(\rho_C, \rho_B) \in r^{\mathcal{I}_{\mathcal{O}, A_0}}$ and $\rho_B \in B^{\mathcal{I}_{\mathcal{O}, A_0}}$. Thus $\rho_C \in (\exists r. B)^{\mathcal{I}_{\mathcal{O}, A_0}}$, as required.

Assume now that $\exists r. A \sqsubseteq B \in \mathcal{O}$ and $\rho_C \in (\exists r. A)^{\mathcal{I}_{\mathcal{O}, A_0}}$. Then there exists ρ_D such that $(\rho_C, \rho_D) \in r^{\mathcal{I}_{\mathcal{O}, A_0}}$ and $\rho_D \in A^{\mathcal{I}_{\mathcal{O}, A_0}}$. Hence $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq \exists r. D$ and $\mathcal{O} \models \exists u. A_0 \sqcap D \sqsubseteq A$. Thus, $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq \exists r. A$. Hence since $\exists r. A \sqsubseteq B \in \mathcal{O}$, $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq B$. But then $\rho_C \in B^{\mathcal{I}_{\mathcal{O}, A_0}}$, as required.

Finally, assume that $r_1 \circ \dots \circ r_n \sqsubseteq r \in \mathcal{O}$ and $(\rho_C, \rho_D) \in r_1^{\mathcal{I}_{\mathcal{O}, A_0}} \circ \dots \circ r_n^{\mathcal{I}_{\mathcal{O}, A_0}}$. Then there are $\rho_{C_0}, \dots, \rho_{C_n}$ with $(\rho_{C_i}, \rho_{C_{i+1}}) \in r_{i+1}^{\mathcal{I}_{\mathcal{O}, A_0}}$ for all $i < n$, where $C_0 = C$ and $C_n = D$. We obtain $\mathcal{O} \models \exists u. A_0 \sqcap C_i \sqsubseteq \exists r_{i+1}. C_{i+1}$ for all $i < n$. Thus $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq \exists r_1 \dots \exists r_n. D$. Hence $\mathcal{O} \models \exists u. A_0 \sqcap C \sqsubseteq \exists r. D$. Hence $(\rho_C, \rho_D) \in r^{\mathcal{I}_{\mathcal{O}, A_0}}$, as required.

Let \mathcal{J} be a model of \mathcal{O} with $A_0^{\mathcal{J}} \neq \emptyset$. Define a relation between $\Delta^{\mathcal{I}_{\mathcal{O}, A_0}}$ and $\Delta^{\mathcal{J}}$ as follows: for any $\rho_C \in \Delta^{\mathcal{I}_{\mathcal{O}, A_0}}$

and $d \in \Delta^{\mathcal{J}}$, let $(\rho_C, d) \in S$ if $d \in C^{\mathcal{J}}$. One can now show that this is well-defined and that for any ρ_C there exists a $d \in \Delta^{\mathcal{J}}$ with $(\rho_C, d) \in S$. It is straightforward to show that S is a $\mathcal{E}\mathcal{L}\mathcal{O}_u(\Sigma)$ -simulation, as required. \square

The following observation is a consequence of Lemma 1 and Lemma 6.

Lemma 7. *Let \mathcal{O} be an $\mathcal{E}\mathcal{L}\mathcal{R}\mathcal{O}_u$ -ontology in normal form, A_0 a concept name, and C an $\mathcal{E}\mathcal{L}\mathcal{O}_u$ -concept. Then the following conditions are equivalent:*

1. $\rho_{A_0} \in C^{\mathcal{I}_{\mathcal{O}, A_0}}$;
2. $\mathcal{O} \models A_0 \sqsubseteq C$.

Next assume that \mathcal{O} and an ABox \mathcal{A} are given. Assume \mathcal{O} is in normal form. Then one can construct in polynomial time a canonical model $\mathcal{I}_{\mathcal{O}, \mathcal{A}}$ of \mathcal{O} that satisfies \mathcal{A} via an assignment $v_{\mathcal{O}, \mathcal{A}}$. The details are straightforward, and we only give the main properties of $\mathcal{I}_{\mathcal{O}, \mathcal{A}}$.

Lemma 8. *Given an $\mathcal{E}\mathcal{L}\mathcal{R}\mathcal{O}_u$ -ontology \mathcal{O} in normal form and an ABox \mathcal{A} one can construct in polynomial time a model $\mathcal{I}_{\mathcal{O}, \mathcal{A}}$ of \mathcal{O} and an assignment $v_{\mathcal{O}, \mathcal{A}}$ such that for all $x \in \text{ind}(\mathcal{A})$ and all $\mathcal{E}\mathcal{L}\mathcal{O}_u$ -concepts C the following conditions are equivalent:*

1. $v_{\mathcal{O}, \mathcal{A}}(x) \in C^{\mathcal{I}_{\mathcal{O}, \mathcal{A}}}$;
2. $\mathcal{O}, \mathcal{A} \models C(x)$.

The following lemma provides a formal description of the relationship between ABoxes that are tree-shaped modulo some set of individual names and $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}$ -concepts.

Lemma 9. *For any $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u(\Sigma)$ -concept C one can construct in polynomial time a pointed Σ -ABox \mathcal{A}, x such that \mathcal{A} is tree-shaped modulo $\text{N}_I \cap \Sigma$ and $d \in C^{\mathcal{I}}$ iff $\mathcal{I} \models \mathcal{A}[x \mapsto d]$, for all interpretations \mathcal{I} and $d \in \Delta^{\mathcal{I}}$.*

Conversely, for any pointed Σ -ABox \mathcal{A}, x such that \mathcal{A} is a tree-shaped ABox modulo Γ , one can construct in polynomial time an $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u(\Sigma)$ -concept C such that $\Gamma = \text{N}_I \cap \Sigma$ and $d \in C^{\mathcal{I}}$ iff $\mathcal{I} \models \mathcal{A}_C[x \mapsto d]$, for all interpretations \mathcal{I} and $d \in \Delta^{\mathcal{I}}$.

The above also holds if one replaces $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u(\Sigma)$ -concepts by $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}(\Sigma)$ -concepts and requires the pointed ABoxes to be weakly rooted.

B Proof for Section 4

We start by proving Remark 3.

Proof of Remark 3. We have to show that the CIP and PBDP are invariant under adding \perp (interpreted as the empty set) to the languages introduced in this paper. Assume that \mathcal{L} is any such language and let \mathcal{L}_{\perp} denote its extension with \perp . We claim that \mathcal{L} enjoys the CIP/PBDP iff \mathcal{L}_{\perp} does. We show this for the CIP, the proof for the PBDP is similar. Assume first that $C \sqsubseteq D$ and $\mathcal{O}_1, \mathcal{O}_2$ are a counterexample to the CIP of \mathcal{L} . Then they are also a counterexample to the CIP of \mathcal{L}_{\perp} . Conversely, assume that $C \sqsubseteq D$ and $\mathcal{O}_1, \mathcal{O}_2$ are a counterexample to the CIP of \mathcal{L}_{\perp} . We may assume that no CI in $\mathcal{O}_1 \cup \mathcal{O}_2$ uses \perp in the concept on its left hand side (if it does, the CI is redundant). Let B be a fresh concept name and replace \perp by B in \mathcal{O}_1 and \mathcal{O}_2 . Also add to \mathcal{O}_i the CIs

$$\exists r. B \sqsubseteq B, \quad B \sqsubseteq A \sqcap \exists r. B$$

for all role names r in $\text{sig}(\mathcal{O}_i)$ and $A \in \text{sig}(\mathcal{O}_i)$. We also let r range over inverse roles in $\text{sig}(\mathcal{O}_i)$ if \mathcal{L} admits inverse roles, the universal role if \mathcal{L} admits the universal role, and A over nominals in $\text{sig}(\mathcal{O}_i)$ if \mathcal{L} admits nominals. Let \mathcal{O}'_i denote the resulting ontology. Then it is easy to see that $C \sqsubseteq D$ and $\mathcal{O}'_1, \mathcal{O}'_2$ are a counterexample to the CIP of \mathcal{L} .

We continue with a few comments and missing proofs for Theorem 1.

Theorem 1. *The following DLs do not enjoy the CIP nor PBDDP:*

1. \mathcal{EL} with the universal role,
2. \mathcal{EL} with nominals,
3. \mathcal{EL} with a single role inclusion $r \circ s \sqsubseteq s$,
4. \mathcal{EL} with role hierarchies and a transitive role,
5. \mathcal{EL} with inverse roles.

In Points 2 to 5, the CIP/PBDDP also fails if the universal role can occur in interpolants/explicit definitions.

Proof. We first supply a proof for Point 4. Let \mathcal{O}_{rs} contain

$$A \sqsubseteq \exists s.E, \quad E \sqsubseteq \exists s_1.B, \quad \exists s_2.B \sqsubseteq A,$$

$$s_1 \sqsubseteq s, \quad s \sqsubseteq s_2, \quad s \circ s \sqsubseteq s,$$

and let $\Sigma = \{s_1, s_2, E\}$. Then A is implicitly definable using Σ under \mathcal{O}_{rs} since

$$\mathcal{O}_{rs} \models \forall x(A(x) \leftrightarrow \exists y(E(y) \wedge \forall z(s_1(y, z) \rightarrow s_2(x, z))).$$

In the same way as above, the interpretations \mathcal{I} and \mathcal{I}' given in Figure 6 show that A has no $\mathcal{EL}_u(\Sigma)$ -definition under \mathcal{O}_{rs} .

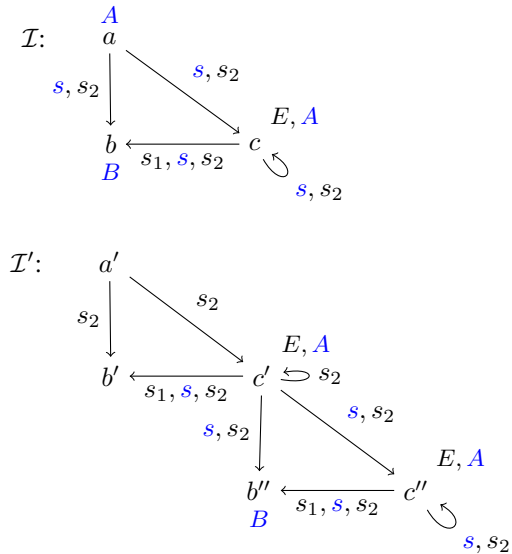


Figure 6: Interpretations \mathcal{I} and \mathcal{I}' used for \mathcal{O}_{rs} .

We next observe that Point 5 can easily be strengthened. The concept name A does not only have no explicit $\mathcal{EL}_u(\Sigma)$ -definition, but no such definition exists in



Figure 7: Interpretations \mathcal{I} (left) and \mathcal{I}' (right) for \mathcal{O}_i .

the positive fragment of \mathcal{ALCC}_u . To see this, consider the interpretations given in Figure 7. Observe that the interpretations $\mathcal{I}, \mathcal{I}'$ show that A is not definable under \mathcal{O}_i using any concept constructed from Σ using $\sqcap, \sqcup, \exists, \forall$ since for any such concept F we have for $(x, x') \in \{(a, a'), (b, b'), (c, c'), (c, c'')\}$ that $x \in F^{\mathcal{I}}$ implies $x' \in F^{\mathcal{I}'}$. Of course, the interpretations \mathcal{I} and \mathcal{I}' given in Figure 7. also demonstrate that concepts with implicit definitions in \mathcal{EL}_u may not have explicit definitions in positive \mathcal{ALCC}_u . The interpretations depicted in Figure 7 differ from the interpretations constructed previously in that they are not the canonical models. The nodes c and c' are not enforced by the ontology but are needed to ensure $\forall r.E$ does not distinguish a and a' . \square

We defer the proof of Theorem 4 to the end of Section D as we need the canonical model and ABox unfolding machinery developed in that section.

C Proofs for Section 5

We give a proof for Remark 5.

Proof of Remark 5. Assume that \mathcal{L} is any DL introduced in this paper and let \mathcal{L}_\perp denote its extension with \perp . The polynomial time reductions of \mathcal{L} -interpolant existence and \mathcal{L} -explicit definition existence to \mathcal{L}_\perp -interpolant existence and \mathcal{L}_\perp -explicit definition existence, respectively, are trivial. For the converse direction, we consider the CIP, the reduction for the PBDDP is similar. The idea is the same as in Remark 3. Assume that $C \sqsubseteq D$ and $\mathcal{O}_1, \mathcal{O}_2$ are in \mathcal{L}_\perp . If $\mathcal{O}_1 \cup \mathcal{O}_2 \models C \sqsubseteq \perp$, then an interpolant exists and we are done. Assume $\mathcal{O}_1 \cup \mathcal{O}_2 \not\models C \sqsubseteq \perp$. We may assume that no CI in $\mathcal{O}_1 \cup \mathcal{O}_2$ uses \perp in the concept on its left hand side (if it does, the CI is redundant). Now let B be a fresh concept name and replace \perp by B in C, D, \mathcal{O}_1 , and \mathcal{O}_2 . Also add to \mathcal{O}_i the CIs

$$\exists r.B \sqsubseteq B, \quad B \sqsubseteq A \sqcap \exists r.B$$

for all role names r in $\text{sig}(\mathcal{O}_i)$ and $A \in \text{sig}(\mathcal{O}_i)$. We also let r range over inverse roles in $\text{sig}(\mathcal{O}_i)$ if \mathcal{L} admits inverse roles, the universal role if \mathcal{L} admits the universal role, and A over nominals in $\text{sig}(\mathcal{O}_i)$ if \mathcal{L} admits nominals. Let \mathcal{O}'_i denote the resulting ontology. Then there exists an \mathcal{L}_\perp -interpolant for $C \sqsubseteq D$ under $\mathcal{O}_1, \mathcal{O}_2$ iff there exists an \mathcal{L} -interpolant for $C \sqsubseteq D$ under $\mathcal{O}'_1, \mathcal{O}'_2$.

D Proofs for Section 6

We first give a proof of the polynomial time decidability of interpolant existence that has not been discussed in the main

paper. Then we provide the missing proofs from the main paper.

The following complexity upper bound proof does not provide an upper bound on the size of interpolants/explicit definitions, but is more elementary than the one we sketched in the main paper.

We start by proving a characterization for the existence of interpolants using canonical models and simulations.

Lemma 10. *Let $\mathcal{O}_1, \mathcal{O}_2$ be \mathcal{ELRO}_u -ontologies in normal form, A, B concept names, and $\mathcal{L} \in \{\mathcal{ELO}, \mathcal{ELO}_u\}$. Let $\Sigma = \text{sig}(\mathcal{O}_1, A) \cap \text{sig}(\mathcal{O}_2, B)$. Then there does not exist an \mathcal{L} -interpolant for $A \sqsubseteq B$ under $\mathcal{O}_1, \mathcal{O}_2$ iff there exists a model \mathcal{J} of $\mathcal{O}_1 \cup \mathcal{O}_2$ and $d \in \Delta^{\mathcal{J}}$ such that*

1. $d \notin B^{\mathcal{J}}$;
2. $(\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}, \rho_A) \preceq_{\mathcal{L}, \Sigma} (\mathcal{J}, d)$.

Proof. Assume an \mathcal{L} -interpolant F exists, but there exists a model \mathcal{J} of $\mathcal{O}_1 \cup \mathcal{O}_2$ and $d \in \Delta^{\mathcal{J}}$ satisfying the conditions of the lemma. As $\mathcal{O}_1 \cup \mathcal{O}_2 \models A \sqsubseteq F$, by Lemma 6, we obtain $\rho_A \in F^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$. By Lemma 1, $d \in F^{\mathcal{J}}$. We have derived a contradiction to the condition that $d \notin B^{\mathcal{J}}$, \mathcal{J} is a model of $\mathcal{O}_1 \cup \mathcal{O}_2$, and $\mathcal{O}_1 \cup \mathcal{O}_2 \models F \sqsubseteq B$.

Assume no \mathcal{L} -interpolant exists. Let

$$\Gamma = \{C \in \mathcal{L}(\Sigma) \mid \rho_A \in C^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}\}$$

By Lemma 7 and compactness, there exists a model \mathcal{J} of $\mathcal{O}_1 \cup \mathcal{O}_2$ and $d \in \Delta^{\mathcal{J}}$ such that $d \in C^{\mathcal{J}}$ for all $C \in \Gamma$ but $d \notin B^{\mathcal{J}}$. We may assume that \mathcal{J} is ω -saturated.³ Thus, by a straightforward generalization of Lemma 1 from finite to ω -saturated interpretations, $(\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}, \rho_A) \preceq_{\mathcal{L}, \Sigma} (\mathcal{J}, d)$, and \mathcal{J} satisfies the conditions of the lemma. \square

The characterization provided in Lemma 10 can be checked in polynomial time. Consider a fresh concept name X_d for each $d \in \Delta^{\mathcal{I}}$ for $\mathcal{I} = \mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}$. We define the $\mathcal{ELO}_u(\Sigma)$ diagram $\mathcal{D}(\mathcal{I})$ of \mathcal{I} as the ontology consisting of the following CIs:

- $X_d \sqsubseteq A$, for every $A \in \Sigma$ and $d \in \Delta^{\mathcal{I}}$;
- $X_{bx} \sqsubseteq \{b\}$, for every $b \in \Sigma$;
- $X_d \sqsubseteq \exists r.X_{d'}$, for every $r \in \Sigma$ and $(d, d') \in r^{\mathcal{I}}$;
- $X_d \sqsubseteq \exists u.X_{d'}$, for every $d, d' \in \Delta^{\mathcal{I}}$.

Denote by \mathcal{I}_{Σ} the Σ -reduct of the interpretation \mathcal{I} . Now it is straightforward to show that there exists a model \mathcal{J} of $\mathcal{O}_1 \cup \mathcal{O}_2$ and $d \in \Delta^{\mathcal{J}}$ such that the conditions of Lemma 10 hold for $\mathcal{L} = \mathcal{ELO}_u$ iff $\mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{D}((\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A})_{|\Sigma}) \not\models X_{\rho_A} \sqsubseteq B$. The latter condition can be checked in polynomial time. If we aim at interpolants without the universal role we simply remove the CIs of the final item from the definition of $\mathcal{D}(\mathcal{I})$, denote the resulting set of inclusions by $\mathcal{D}'(\mathcal{I})$ and have that there exists a model \mathcal{J} of $\mathcal{O}_1 \cup \mathcal{O}_2$ and $d \in \Delta^{\mathcal{J}}$ such that the conditions of Lemma 10 hold for $\mathcal{L} = \mathcal{ELO}$ iff $\mathcal{O}_1 \cup \mathcal{O}_2 \cup \mathcal{D}'((\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A})_{|\Sigma}) \not\models X_{\rho_A} \sqsubseteq B$.

³See (Chang and Keisler 1998) for an introduction to ω -saturated interpretations and their properties.

Directed Unfolding of ABox. We give a precise definition of the directed unfolding of an ABox. Let \mathcal{A} be a factorized Σ -ABox and $\Gamma = \mathbb{N}_1 \cap \Sigma$. The *directed unfolding* of \mathcal{A} into a ditree-shaped ABox \mathcal{A}^u modulo Γ is defined as follows. The individuals of \mathcal{A}^u are the words $w = x_0 r_1 \cdots r_n x_n$ with r_1, \dots, r_n role names and $x_0, \dots, x_n \in \text{ind}(\mathcal{A})$ such that $\{a\}(x_i) \notin \mathcal{A}$ for any $i \neq 0$ and $a \in \Gamma$ and $r_{i+1}(x_i, x_{i+1}) \in \mathcal{A}$ for all $i < n$. We set $\text{tail}(w) = x_n$ and define

- $A(w) \in \mathcal{A}^u$ if $A(\text{tail}(w)) \in \mathcal{A}$, for $A \in \mathbb{N}_C$;
- $r(w, wrx) \in \mathcal{A}^u$ if $r(\text{tail}(w), x) \in \mathcal{A}$ and $r(w, x) \in \mathcal{A}^u$ if $\{a\}(x) \in \mathcal{A}$ for some $a \in \Gamma$ and $r(\text{tail}(w), x) \in \mathcal{A}$, for $r \in \mathbb{N}_R$;
- $\{a\}(x) \in \mathcal{A}^u$ if $\{a\}(x) \in \mathcal{A}$, for $a \in \Gamma$ and $x \in \text{ind}(\mathcal{A})$.

Derivation Trees. Fix an \mathcal{ELRO}_u -ontology \mathcal{O} in normal form, a $\text{sig}(\mathcal{O})$ -ABox \mathcal{A} , and recall the definition of Δ and Θ . Let $(a, C) \in \Delta \times \Theta$. A *derivation tree* for the assertion (a, C) in \mathcal{O}, \mathcal{A} is a finite $\Delta \times \Theta$ -labeled tree (T, V) , where T is a set of nodes and $V : T \rightarrow \Delta \times \Theta$ the labeling function, such that

- $V(\varepsilon) = (a, C)$;
- if $V(n) = (a, C)$, then (i) $a \in \text{ind}(\mathcal{A})$ and $C = \top$ or (ii) $C(a) \in \mathcal{A}$ or (iii) $a \in \mathbb{N}_1$ and $C = \{a\}$ or
 1. $a = C = A$ for a concept name A and n has a successor n' with $V(n') = (b, A)$; or
 2. $a = C = A$ for a concept name A and n has a successor n' such that $V(n') = (b, C')$ and $\mathcal{O} \models C' \sqsubseteq \exists u.A$; or
 3. n has successors n_1, n_2 with $V(n_i) = (a, C_i)$ for $i = 1, 2$ and $\mathcal{O} \models C_1 \sqcap C_2 \sqsubseteq C$; or
 4. n has successors n_1, n_2, n_3 with $V(n_1) = (b, C)$, $V(n_2) = (a, \{c\})$, and $V(n_3) = (b, \{c\})$; or
 5. the conditions of the rule for RIs discussed in the main paper hold: there are role names r_2, \dots, r_{2k-2}, r and members $a = a_1, \dots, a_{2k}$ of Δ such that (a_{2k}, C') is a label of a successor of n , $\mathcal{O} \models \exists r.C' \sqsubseteq C$, $\mathcal{O} \models r_2 \circ \cdots \circ r_{2k-2} \sqsubseteq r$, and the situation depicted in Figure 4 holds, where the ‘‘dotted lines’’ stand for ‘either $a_i = a_{i+1}$ or some $(a_i, \{c\}), (a_{i+1}, \{c\})$ with $c \in \mathbb{N}_1$ are labels of successors of n , and \hat{r}_i stands for ‘either $r(a_i, a_{i+1}) \in \mathcal{A}$ or some (a_i, C_i) is a label of a successor of n and $\mathcal{O} \models C_i \sqsubseteq \exists r_i.\{a_{i+1}\}$ if $a_{i+1} \in \mathbb{N}_1$ and $\mathcal{O} \models C_i \sqsubseteq \exists r_i.a_{i+1}$ if $a_{i+1} \in \mathbb{N}_C$ ’. Moreover, for all $a_i \neq a$, $1 < i \leq 2k$, there exists a successor of n with label (a_i, D) for some D ; or
 6. n has a successor n' with $V(n') = (b, C')$ and $\mathcal{O} \models \exists u.C' \sqsubseteq C$.

The purpose of Conditions 1 and 2 is to establish that it follows from \mathcal{O} and \mathcal{A} that A is not empty. In this case (A, A) is derived. The purpose of the remaining rules should be clear.

Example 5. We use the ontology from Example 1. Recall that

$$\mathcal{O}_p = \{r_i \circ r_i \sqsubseteq r_{i+1} \mid 0 \leq i < n\} \cup \{A \sqsubseteq \exists r_0.B, B \sqsubseteq \exists r_0.B, \exists r_n.B \sqsubseteq A\}$$

Then $\mathcal{I}_{\mathcal{O}_p, A}$ is defined by setting

$$\begin{aligned}\Delta^{\mathcal{I}_{\mathcal{O}_p, A}} &= \{\rho_A, y\} \\ A^{\mathcal{I}_{\mathcal{O}_p, A}} &= \{\rho_A\} \\ B^{\mathcal{I}_{\mathcal{O}_p, A}} &= \{y\} \\ r_i^{\mathcal{I}_{\mathcal{O}_p, A}} &= \{(\rho_A, y), (y, y)\}, \text{ for } 0 \leq i \leq n.\end{aligned}$$

Recall that $\Sigma = \{r_0, B\}$ and that $\exists r_0^{2^n}.B$ is an explicit definition of A using Σ under \mathcal{O}_p . Consider the ABox $\mathcal{A}_{|\Sigma}$ corresponding to the Σ -reduct of $\mathcal{I}_{\mathcal{O}_p, A}$. Then a derivation tree (T, V) for (ρ_A, A) in $\mathcal{O}_p, \mathcal{A}_{|\Sigma}$ is defined by setting $V(\varepsilon) = (\rho_A, A)$ and taking a single successor n of ε with $V(n) = (y, B)$. In the notation of Rule 5, we have $a_1 = a_2 = \rho_A$ and $a_3 = \dots = a_{2^n} = y$. We use that $\mathcal{O}_p \models r_0^{2^n} \sqsubseteq r_n$ and $\mathcal{O}_p \models \exists r_n.B \sqsubseteq A$.

We next show Part 1 of Lemma 3.

Proof of Part 1 of Lemma 3. Let \mathcal{O} be an \mathcal{ELRO}_u -ontology in normal form and \mathcal{A} a finite $\text{sig}(\mathcal{O})$ -ABox. Assume (x, A) with $x \in \text{ind}(\mathcal{A})$ and $A \in \Theta$ is given. It is straightforward to show by induction that if there is a derivation tree for (x, A) in \mathcal{O}, \mathcal{A} , then $\mathcal{O}, \mathcal{A} \models A(x)$. We construct a sequence of ABoxes $\mathcal{A}_0, \mathcal{A}_1, \dots$ as follows. Define \mathcal{A}_0 as the union of \mathcal{A} and all assertions $\{a\}(a)$ with a an individual name in \mathcal{O} and $\top(x)$ with $x \in \text{ind}(\mathcal{A})$. Let \mathcal{A}_{i+1} be obtained from \mathcal{A}_i by applying one of the following rules:

1. if $A(b) \in \mathcal{A}_i$, then add $A(A)$ to \mathcal{A}_i ;
2. if $C'(b) \in \mathcal{A}_i$ and $\mathcal{O} \models C' \sqsubseteq \exists u.A$, then add $A(A)$ to \mathcal{A}_i ;
3. if $C_1(a), C_2(a) \in \mathcal{A}_i$ and $\mathcal{O} \models C_1 \sqcap C_2 \sqsubseteq C$, then add $C(a)$ to \mathcal{A}_i ;
4. if $C(b), \{c\}(a), \{c\}(b) \in \mathcal{A}_i$, then add $C(a)$ to \mathcal{A}_i ;
5. if there is a sequence a_1, \dots, a_{2^k} of elements of Δ and a sequence $r_2, r_4, \dots, r_{2^{k-2}}$ of role names such that $a = a_1$ and for every a_{2^j+1} either $a_{2^j+1} = a_{2^j+2}$ or there is c with $\{c\}(a_{2^j+1}), \{c\}(a_{2^j+2}) \in \mathcal{A}_i$ such that for every a_{2^j} :
 - $r_{2^j}(a_{2^j}, a_{2^j+1}) \in \mathcal{A}$; or
 - $a_{2^j+1} \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$ and there exists $C_{2^j} \in (\mathbf{N}_C \cup \mathbf{N}_I) \cap \text{sig}(\mathcal{O})$ such that $C_{2^j}(a_{2^j}) \in \mathcal{A}_i$ and $\mathcal{O} \models C_{2^j} \sqsubseteq \exists r_{2^j}.\{a_{2^j+1}\}$; or
 - $a_{2^j+1} \in \mathbf{N}_C \cap \text{sig}(\mathcal{O})$ and there exists $C_{2^j} \in (\mathbf{N}_C \cup \mathbf{N}_I) \cap \text{sig}(\mathcal{O})$ such that $C_{2^j}(a_{2^j}) \in \mathcal{A}_i$ and $\mathcal{O} \models C_{2^j} \sqsubseteq \exists r_{2^j}.a_{2^j+1}$
and there exist $C' \in (\mathbf{N}_C \cup \mathbf{N}_I) \cap \text{sig}(\mathcal{O})$ and a role name r such that $C'(a_{2^k}) \in \mathcal{A}_i$, $\mathcal{O} \models \exists r.C' \sqsubseteq A$, and $\mathcal{O} \models r_2 \circ r_4 \circ \dots \circ r_{2^{k-2}} \sqsubseteq r$, then add $A(a)$ to \mathcal{A}_i .
6. if $C'(b) \in \mathcal{A}_i$ and $\mathcal{O} \models \exists u.C' \sqsubseteq C$, then add $C(a)$ to \mathcal{A}_i .

Note that the sequence is finite, and denote by \mathcal{A}^* the final ABox.

Claim. There is a model \mathcal{I}, v of \mathcal{A}^* and \mathcal{O} such that for all $x \in \text{ind}(\mathcal{A})$ and $A \in \mathbf{N}_C, v(x)^{\mathcal{I}} \in A^{\mathcal{I}}$ implies $A(x) \in \mathcal{A}^*$.

Proof of the Claim. For all $a, b \in \text{ind}(\mathcal{A}^*)$, we write $a \sim b$ if $a = b$ or $\{c\}(a), \{c\}(b) \in \mathcal{A}^*$ for some c . Notice that

due to Rule 4, $a \sim b$ implies $C(a) \in \mathcal{A}^*$ if and only if $C(b) \in \mathcal{A}^*$. It follows that \sim is an equivalence relation. We let $[a]$ denote the equivalence class of a . Start with an interpretation \mathcal{I}_0 defined by:

$$\begin{aligned}\Delta^{\mathcal{I}_0} &= \text{ind}(\mathcal{A}^*)/\sim \\ A^{\mathcal{I}_0} &= \{[a] \mid A(a) \in \mathcal{A}^*\} \\ a^{\mathcal{I}_0} &= \{[a]\} \\ r^{\mathcal{I}_0} &= \{([a], [b]) \mid \exists a' \in [a], b' \in [b]. r(a', b') \in \mathcal{A}^*\}.\end{aligned}$$

By definition, \mathcal{I}_0 satisfies all CIs in \mathcal{O} that do not involve role names or the universal role. We next extend \mathcal{I}_0 by adding pairs of the form $([a], [b])$ with $b \in \mathbf{N}_C \cup \mathbf{N}_I$ to the interpretation of role names. In detail, if $[a] \in \Delta^{\mathcal{I}_0}$ and there exist $C \in \mathbf{N}_C \cup \mathbf{N}_I$ with $a \in C^{\mathcal{I}_0}$ and $c \in \mathbf{N}_I$ with $c \in [b]$ such that $\mathcal{O} \models C \sqsubseteq \exists r.\{c\}$, then add $([a], [b])$ to $r^{\mathcal{I}_0}$. Also, if $[a] \in \Delta^{\mathcal{I}_0}$ and there exist $C \in \mathbf{N}_C \cup \mathbf{N}_I$ with $a \in C^{\mathcal{I}_0}$ and $A \in \mathbf{N}_C$ with $A \in [b]$ such that $\mathcal{O} \models C \sqsubseteq \exists r.A$, then add $([a], [b])$ to $r^{\mathcal{I}_0}$. Finally, add any pair $([a], [b])$ to $r^{\mathcal{I}_0}$ if there exists an RI $r_1 \circ \dots \circ r_n \sqsubseteq r$ that follows from \mathcal{O} such that $([a], [b])$ is in relation $r_1 \circ \dots \circ r_n$ under the updated interpretations of r_1, \dots, r_n . This defines an interpretation \mathcal{I} . By Rule 2 all CIs of the form $A \sqsubseteq \exists r.B$ are satisfied in \mathcal{I} . By definition, all RIs in \mathcal{O} are satisfied in \mathcal{I} . By Rules 5 and 6, all CIs of the form $\exists r.B \sqsubseteq A$ are satisfied as well. This finishes the proof of the claim.

Now suppose $\mathcal{O}, \mathcal{A} \models A_0(x_0)$. By the Claim, we have $A_0(x_0) \in \mathcal{A}^*$. Since the six rules to construct $\mathcal{A}_0, \mathcal{A}_1, \dots$ are in one-to-one correspondence with Conditions (1)–(6) from the definition of derivation trees, we can inductively construct a derivation tree for $A_0(x_0)$ in \mathcal{A} w.r.t. \mathcal{O} .

The remaining claims made in Part 1 of Lemma 3 have been shown in the main paper already. \square

We next come to Part 2 of Lemma 3. The following example illustrates how one can construct from a derivation tree of $A(x)$ in \mathcal{O}, \mathcal{A} a derivation tree in $\mathcal{O}, \mathcal{A}^u$ with \mathcal{A}^u the directed unfolding of \mathcal{A} . The derivation tree has the same depth but the outdegree might be exponential.

Example 6. Recall the ontology \mathcal{O}_p and concept name A from Example 5. We consider the Σ -reduct $\mathcal{A}_{|\Sigma}$ of the ABox \mathcal{A} corresponding to the canonical model $\mathcal{I}_{\mathcal{O}_p, A}$. It is defined by $\mathcal{A}_{|\Sigma} = \{r_0(\rho_A, y), r_0(y, y), B(y)\}$. The directed unfolding $\mathcal{A}_{|\Sigma}^u$ has individuals

$$\rho_A, \rho_A r_0 y, \rho_A r_0 y r_0 y, \dots$$

and the assertions

$$B(\rho_A r_0 y), B(\rho_A x r_0 y r_0 y), \dots$$

$$r_0(\rho_A, \rho_A r_0 y), r_0(\rho_A r_0 y, \rho_A r_0 y r_0 y), \dots$$

In a derivation tree (T', V') for $A(\rho_0)$ in $\mathcal{O}_p, \mathcal{A}_{|\Sigma}$ we require that ε has 2^n successors labeled with:

$$(\rho_A r_0 y, B), (\rho_A r_0 y r_0 y, B), \dots, (\rho_A (r_0 y)^{2^n}, B).$$

We now give the general construction of the derivation tree in the directed unfolding from a derivation tree in the original ABox.

Proof of Part 2 of Lemma 3. Assume that (T, V) is a derivation tree for $A(x)$ in \mathcal{O}, \mathcal{A} of at most exponential size. We obtain a very similar derivation tree (T', V') for $A(x)$ in $\mathcal{O}, \mathcal{A}^u$ with \mathcal{A}^u the directed unfolding of \mathcal{A} modulo $\Sigma = \text{sig}(\mathcal{A}) \cap \mathbb{N}_1$. In fact, with the exception of Condition 5, the construction is identical. For Condition 5, one potentially has to introduce "copies" of the nodes in T which correspond to the fresh individuals introduced in the unfolded ABox.

In the following construction of (T', V') the following holds: if the label of n in (T, V) is (a, C) , then the label of copies n' of n in (T', V') takes the form (w, C) with $\text{tail}(w) = a$. Moreover, if $\{b\}(a) \in \mathcal{A}$ for some $b \in \Sigma \cap \mathbb{N}_1$ or $a \in \mathbb{N}_1 \cup \mathbb{N}_C$, then the label of n' is identical to the label of n . Note V' is a mapping from T' to $\Delta' \times \Theta$ with

$$\Delta' = \text{ind}(\mathcal{A}^u) \cup ((\mathbb{N}_C \cup \mathbb{N}_1) \cap \text{sig}(\mathcal{O}))$$

In detail, we define (T', V') as follows from (T, V) , starting with the root by setting $V'(\varepsilon) := V(\varepsilon) = (x, A)$.

Assume inductively that m is a copy of n , $V(n) = (a, C)$, and $V'(m) = (w, C)$. To define the successors of m and their labelings we consider the possible derivation steps for (a, C) in \mathcal{O}, \mathcal{A} : (i) if $a \in \text{ind}(\mathcal{A})$ and $C = \top$, then $w \in \text{ind}(\mathcal{A}^u)$ and $C = \top$; (ii) if $C(a) \in \mathcal{A}$, then $C(w) \in \mathcal{A}^u$; (iii) if $a \in \mathbb{N}_1$ and $C = \{a\}$, then $V'(m) = (a, \{a\})$. We next consider the cases 1 to 6:

1. $a = C = A$ for a concept name A and n has a successor n' with $V(n') = (b, A)$: take a copy m' of n' as the only successor of m and set $V'(m') = (b, A)$.
2. $a = C = A$ for a concept name A and n has a successor n' such that $V(n') = (b, C')$ and $\mathcal{O} \models C' \sqsubseteq \exists u.A$: take a copy m' of n' as the only successor of m and set $V'(m') = (b, C)$.
3. n has successors n_1, n_2 with $V(n_i) = (a, C_i)$ and $\mathcal{O} \models C_1 \sqcap C_2 \sqsubseteq C$: take copies m_1, m_2 of n_1, n_2 as the successors of m and set $V'(m_i) = (w, C_i)$.
4. n has successors n_1, n_2, n_3 with $V(n_1) = (b, C)$, $V(n_2) = (a, \{c\})$, and $V(n_3) = (b, \{c\})$: take copies m_1, m_2, m_3 of n_1, n_2, n_3 as successors of m and set $V(m_1) = (b, C)$, $V(m_2) = (w, \{c\})$, and $V(m_3) = (b, \{c\})$.
5. Suppose that n has successors such that the conditions of Point 5 for derivation trees hold for r_2, \dots, r_{2k-2}, r and members $a = a_1, \dots, a_{2k}$ of Δ . We define the new members b_1, \dots, b_{2k} of Δ' and relevant successors of m with labeling by induction. We set $b_1 = w$. Assume that b_{2i+1} has been defined and (b_{2i+1}, D) is the label of a copy of a successor of n with label (a_{2i+1}, D) .

Case 1. $a_{2i+1} = a_{2i+2}$. Then we set $b_{2i+2} := b_{2i+1}$.

Case 2. There exists $c \in \mathbb{N}_1$ and successors n_1, n_2 of n with $V(n_1) = (a_{2i+1}, \{c\})$ and $V(n_2) = (a_{2i+2}, \{c\})$. Then we let $b_{2i+2} := a_{2i+2}$ and we introduce copies m_1, m_2 of n_1, n_2 with $V'(m_1) = (b_{2i+1}, \{c\})$ and $V'(m_2) = (a_{2i+2}, \{c\})$.

Now assume that b_{2i} has been defined and (b_{2i}, D) is the label of a copy of a successor of n with label (a_{2i}, D) .

Case 1. $r_{2i}(a_{2i}, a_{2i+1}) \in \mathcal{A}$ and $V(n') = (a_{2i+1}, D')$ for some successor n' of n . If $\{b\}(a_{2i+1}) \in \mathcal{A}$ for some $b \in \mathbb{N}_1$, then we set $b_{2i+1} = a_{2i+1}$ and introduce a copy m' of n and set $V'(m') = (a_{2i+1}, D')$. Observe that $r_{2i}(b_{2i}, a_{2i+1}) \in \mathcal{A}^u$. Otherwise (if no b with $\{b\}(a_{2i+1}) \in \mathcal{A}$ exists), we set $b_{2i+1} = b_{2i}r_{2i}a_{2i+1}$ and introduce a copy m' of n' and set $V'(m') = (b_{2i+1}, D')$.

Case 2. $a_{2i+1} \in (\mathbb{N}_1 \cup \mathbb{N}_C) \cap \text{sig}(\mathcal{O})$, $V(n_1) = (a_{2i+1}, D')$ for some successor n_1 of n , and $V(n_2) = (a_{2i}, F)$ for a successor n_2 of n and $\mathcal{O} \models F \sqsubseteq \exists r_{2i}.a_{2i+1}$ (if $a_{2i+1} \in \mathbb{N}_C$) or $\mathcal{O} \models F \sqsubseteq \exists r_{2i}.\{a_{2i+1}\}$ (if $a_{2i+1} \in \mathbb{N}_1$), respectively. Then we introduce copies m_1, m_2 of n_1, n_2 and set $b_{2i+1} = a_{2i+1}$, $V'(m_1) = (b_{2i+1}, D')$, and $V'(m_2) = (b_{2i}, F)$.

6. n has a successor n' with $V(n') = (b, C')$ and $\mathcal{O} \models \exists u.C' \sqsubseteq C$: then introduce a copy m' of n' and set $V'(m') = (b, C)$.

Then (T', V') is a derivation tree for $A(x)$ in $\mathcal{O}, \mathcal{A}^u$ satisfying the conditions of the lemma. \square

The proof of "2. \Rightarrow 3." of Theorem 5 is now as sketched in the main paper. Note also that we can construct \mathcal{A} in exponential time since we can construct the derivation tree for B in $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^\Sigma$ in exponential time, then lift it to a derivation tree in its unfolding in exponential time, and from that derivation tree obtain the individuals in the ABox \mathcal{A} in exponential time.

A proof of the statement of Theorem 5 for interpolants without the universal role is obtained from the proof above in a straightforward way.

We conclude this section with a deferred proof of Theorem 4.

Theorem 4. Let $\mathcal{O}_1, \mathcal{O}_2$ be \mathcal{EL} -ontologies with RIs, C_1, C_2 be \mathcal{EL} -concepts, and set $\Sigma = \text{sig}(\mathcal{O}_1, C_1) \cap \text{sig}(\mathcal{O}_2, C_2)$. Assume that the set of RIs in $\mathcal{O}_1 \cup \mathcal{O}_2$ is safe for Σ and $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqsubseteq C_2$. Then an \mathcal{EL} -interpolant for $C_1 \sqsubseteq C_2$ under $\mathcal{O}_1, \mathcal{O}_2$ exists.

Proof. For convenience of notation, we assume w.l.o.g., by Lemma 2, that \mathcal{O}_1 and \mathcal{O}_2 are in normal form, $A \in \text{sig}(\mathcal{O}_1)$, $B \in \text{sig}(\mathcal{O}_2)$ and $\{A, B\} \cap \Sigma = \emptyset$. Suppose for a proof by contradiction that $\mathcal{O}_1 \cup \mathcal{O}_2 \models A \sqsubseteq B$ but there exists no \mathcal{EL} -interpolant for $A \sqsubseteq B$. Then $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^{\downarrow \Sigma} \not\models B(\rho_A)$. Moreover, since the language under consideration contains neither nominals nor the universal role, this strengthens to $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^\Sigma \not\models B(\rho_A)$.

Let \mathcal{J}_0 be the canonical model of $\mathcal{O}_1 \cup \mathcal{O}_2$ and $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^\Sigma$. In what follows, we identify the domain of $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^\Sigma$ and individuals of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^\Sigma$, and consider both to be subsets of the domain of \mathcal{J}_0 . By the properties of the canonical model, we then have $\rho_A \notin B^{\mathcal{J}_0}$. Furthermore, as $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}$ is a model for both $\mathcal{O}_1 \cup \mathcal{O}_2$ and $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}^\Sigma$, there exists a $\text{sig}(\mathcal{O}_1 \cup \mathcal{O}_2)$ -simulation S between \mathcal{J}_0 and $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}$ such that $(x, x) \in S$ for all $x \in \Delta^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}}}$.

$$\begin{aligned}
\Delta^{\mathcal{J}_{i+1}} &= \Delta^{\mathcal{J}_i}, \\
A^{\mathcal{J}_{i+1}} &= A^{\mathcal{J}_i}, \text{ for all } A \in \mathbf{N}_C, \\
r^{\mathcal{J}_{i+1}} &= r^{\mathcal{J}_i} \cup \left\{ (d_1, d_{n+1}) \mid \begin{array}{l} r_1 \circ \dots \circ r_n \sqsubseteq r \in \mathcal{O}_1 \cup \mathcal{O}_2 \\ \{d_1, \dots, d_{n+1}\} \subseteq \Delta^{\mathcal{J}_i}, (d_1, d_{n+1}) \notin r^{\mathcal{J}_i} \\ (d_k, d_{k+1}) \in r_k^{\mathcal{J}_i} \text{ for all } 1 \leq k \leq n \end{array} \right\}
\end{aligned}$$

Figure 8: Definition of \mathcal{J}_{i+1} .

Consider an interpretation \mathcal{J}_1 defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{J}_1} &= \Delta^{\mathcal{J}_0}, \\
P^{\mathcal{J}_1} &= P^{\mathcal{J}_0} \cup P^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}, \text{ for all } P \in (\text{sig}(\mathcal{O}_1) \setminus \Sigma), \\
P^{\mathcal{J}_1} &= P^{\mathcal{J}_0}, \text{ for all } P \notin (\text{sig}(\mathcal{O}_1) \setminus \Sigma),
\end{aligned}$$

where P is a concept or role name. If $\mathcal{J}_1 \models \mathcal{O}_1 \cup \mathcal{O}_2$ we immediately derive a contradiction as we then have $\rho_A \in A^{\mathcal{J}_1}$ and $\rho_A \notin B^{\mathcal{J}_1}$, contradicting $\mathcal{O}_1 \cup \mathcal{O}_2 \models A \sqsubseteq B$.

- If $\mathcal{O}_1 \cup \mathcal{O}_2$ does not contain RIs, as \mathcal{J}_0 and \mathcal{J}_1 are identical on all elements except $\Delta^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$, for all $x \in \Delta^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$ the relation S is a $\text{sig}(\mathcal{O}_1 \cup \mathcal{O}_2)$ -simulation between \mathcal{J}_1 and $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}$. Conversely, the embedding of $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}$ into \mathcal{J}_1 generates a simulation, that is $(\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}, x) \preceq_{\mathcal{E}\mathcal{L}, \text{sig}(\mathcal{O}_1)} (\mathcal{J}_1, x)$ for all $x \in \Delta^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$. By Lemma 1, for any $\text{sig}(\mathcal{O}_1)$ - $\mathcal{E}\mathcal{L}$ -concept C and for all $x \in \Delta^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$ we have $x \in C^{\mathcal{J}_1}$ if, and only if $x \in C^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$. Thus, \mathcal{J}_1 is a model of CIs in \mathcal{O}_1 . By construction $\mathcal{J}_1 \models \mathcal{O}_2$.
- Suppose that $\mathcal{O}_1 \cup \mathcal{O}_2$ contains RIs. Since the interpretation \mathcal{J}_1 may not satisfy some RIs, we consider a sequence of interpretations \mathcal{J}_i obtained by extending the interpretations of roles in \mathcal{J}_1 to satisfy RIs. We give the construction of \mathcal{J}_{i+1} , for $i \geq 1$, in Figure 8.

A simple inductive argument shows that by the safety condition and the fact that $(d_1, d_{n+1}) \notin r^{\mathcal{J}_i}$ we have that $\{r_1, \dots, r_n, r\} \subseteq \text{sig}(\mathcal{O}_1)$.

Furthermore, we prove by induction that the relation S is a $\text{sig}(\mathcal{O}_1 \cup \mathcal{O}_2)$ -simulation between \mathcal{J}_{i+1} and $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}$. For $i = 1$ this has been established above. For the induction step it suffices to consider r -successors of d_1 in \mathcal{J}_{i+1} , where r is from the definition of \mathcal{J}_{i+1} above. By the induction hypothesis, S is a $\text{sig}(\mathcal{O}_1 \cup \mathcal{O}_2)$ -simulation between \mathcal{J}_i and $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}$. Then there exist $\{v_2, \dots, v_{n+1}\} \subseteq \Delta^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$ with $(d_{j+1}, v_{j+1}) \in S$ and $(v_1, v_{n+1}) \in r_i^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$ for $j \in \{1, \dots, n\}$. As $\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}$ is a model of \mathcal{O}_1 , we have $(v_1, v_{n+1}) \in r^{\mathcal{I}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}}$ and $(d_{n+1}, v_{n+1}) \in S$ as required.

As $\mathcal{E}\mathcal{L}$ canonical models defined in this paper are finite, there exists $N > 0$ such that for all $i > N$, $\mathcal{J}_i = \mathcal{J}_N$. It can be seen that \mathcal{J}_N satisfies all RIs in $\mathcal{O}_1 \cup \mathcal{O}_2$ and the satisfaction of CIs is proved similarly to the case above. Then \mathcal{J}_N is a model of $\mathcal{O}_1 \cup \mathcal{O}_2$ with $\rho_A \in A^{\mathcal{J}_N}$ and $\rho_A \notin B^{\mathcal{J}_N}$, contradicting $\mathcal{O}_1 \cup \mathcal{O}_2 \models A \sqsubseteq B$.

□

E Proofs for Section 7

The section is organized as follows. We first introduce canonical models and derivation trees for $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u$. We then give the automata based proof of the EXPTIME upper bound for interpolant existence. We then show the double exponential lower bound on the size of explicit definitions, the implication “2. \Rightarrow 3.” of Theorem 6, and that interpolants can be computed in double exponential time.

Canonical Models. Assume \mathcal{O} is an $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u$ ontology in normal form and A a concept name with $A \in \text{sig}(\mathcal{O})$. We introduce the canonical model $\mathcal{I}_{\mathcal{O}, A}$. Let $\text{sub}(\mathcal{O})$ denote the set of subconcepts of concepts in \mathcal{O} , and denote by $\text{sub}^{\exists}(\mathcal{O})$ the set of $\exists r.\{a\}$ with r or r^- a role name in $\text{sig}(\mathcal{O})$ and $a \in \text{sig}(\mathcal{O})$. We may assume that $\exists u.A \in \text{sub}(\mathcal{O})$. An \mathcal{O} -type is a subset τ of $\text{sub}(\mathcal{O}) \cup \text{sub}^{\exists}(\mathcal{O})$ such that $\mathcal{O} \models \prod_{C \in \tau} C \sqsubseteq C'$ implies $C' \in \tau$ for all concepts $C' \in \text{sub}(\mathcal{O}) \cup \text{sub}^{\exists}(\mathcal{O})$. We sometimes identify τ and $\prod_{C \in \tau} C$. For a role r , we write $\tau_1 \rightsquigarrow_r \tau_2$ if τ_2 is a maximal (w.r.t. inclusion) \mathcal{O} -type such that $\mathcal{O} \models \tau_1 \sqsubseteq \exists r.\tau_2$. Note that the set of all \mathcal{O} -types and relation \rightsquigarrow_r can be computed in exponential time.

For any concept name B , τ_B denotes the minimal \mathcal{O} -type containing B and $\exists u.A$. Similarly, for any individual a , τ_a denotes the minimal \mathcal{O} -type containing $\{a\}$ and $\exists u.A$. Let $S = S_C \cup S_N$ with $S_C = \{\tau_B \mid \mathcal{O} \models A \sqsubseteq \exists u.B\}$ and $S_N = \{\tau_a \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})\}$. The *canonical model* $\mathcal{I}_{\mathcal{O}, A}$ of \mathcal{O} and A is defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_{\mathcal{O}, A}} &= \{\tau_0 r_1 \tau_1 \dots r_n \tau_n \mid \tau_0 \in S, \tau_1, \dots, \tau_n \notin S_N, \\
&\quad r_1, \dots, r_n \in \mathbf{N}_R \cup \mathbf{N}_R^-, \tau_i \rightsquigarrow_{r_{i+1}} \tau_{i+1}\} \\
a^{\mathcal{I}_{\mathcal{O}, A}} &= \tau_a \\
B^{\mathcal{I}_{\mathcal{O}, A}} &= \{w \mid w \in \Delta^{\mathcal{I}_{\mathcal{O}, A}}, B \in \text{tail}(w)\} \\
r^{\mathcal{I}_{\mathcal{O}, A}} &= \{(w, wr\tau) \mid w, wr\tau \in \Delta^{\mathcal{I}_{\mathcal{O}, A}}\} \cup \\
&\quad \{(wr^- \tau, w) \mid w, wr^- \tau \in \Delta^{\mathcal{I}_{\mathcal{O}, A}}\} \cup \\
&\quad \{r(w, \tau_a) \mid \exists r.\{a\} \in \text{tail}(w)\} \cup \\
&\quad \{r(\tau_a, w) \mid \exists r^-. \{a\} \in \text{tail}(w)\}
\end{aligned}$$

We also use ρ_A to denote τ_A . The following properties of canonical models can be proved in a standard way.

Lemma 11. *For all $\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u$ -ontologies \mathcal{O} in normal form and concept names $A \in \text{sig}(\mathcal{O})$:*

1. $\mathcal{I}_{\mathcal{O}, A}$ is a model of \mathcal{O} ;
2. for every model \mathcal{J} of \mathcal{O} and any $d \in \Delta^{\mathcal{J}}$ with $d \in A^{\mathcal{J}}$, $(\mathcal{I}_{\mathcal{O}, A}, \rho_A) \preceq_{\mathcal{E}\mathcal{L}\mathcal{I}\mathcal{O}_u, \Sigma} (\mathcal{J}, d)$;

3. for every $\mathcal{ELIO}_u(\text{sig}(\mathcal{O}))$ -concept C , $\mathcal{O} \models A \sqsubseteq C$ iff and only if $\rho_A \in C^{\mathcal{I}_{\mathcal{O},A}}$.

We use $\mathcal{A}_{\mathcal{O},A}$ to denote the ABox associated with the canonical model $\mathcal{I}_{\mathcal{O},A}$, and $\mathcal{A}_{\mathcal{O},A}^\Sigma$ its Σ -reduct. We denote the individuals x_{τ_a} and x_{τ_B} by x_a and x_B , respectively and observe that $x_a = x_b$ iff $\mathcal{O} \models \{a\} \sqcap \exists u.A \sqsubseteq \{b\}$ and $x_B = x_a$ iff $\mathcal{O} \models B \sqcap \exists u.A \sqsubseteq \{a\}$.

Undirected Unfolding of an ABox. We give a precise definition of the undirected unfolding of an ABox. Let \mathcal{A} be a Σ -ABox and $\Gamma = \mathbf{N}_I \cap \Sigma$. The *undirected unfolding* of \mathcal{A} into a tree-shaped ABox \mathcal{A}^* modulo Γ is defined as follows. The individuals of \mathcal{A}^* are the set of words $w = x_0 r_1 \dots r_n x_n$ with r_1, \dots, r_n roles and $x_0, \dots, x_n \in \text{ind}(\mathcal{A})$ such that $\{a\}(x_i) \notin \mathcal{A}$ for any $i \neq 0$ and $a \in \Gamma$, and $r_{i+1}(x_i, x_{i+1}) \in \mathcal{A}$ if r_{i+1} is a role name and $r_{i+1}^-(x_{i+1}, x_i) \in \mathcal{A}$ if r_{i+1} is an inverse role, for all $i < n$. We set $\text{tail}(w) = x_n$ and let

- $A(w) \in \mathcal{A}^*$ if $A(\text{tail}(w)) \in \mathcal{A}$, for $A \in \mathbf{N}_C$;
- $r(w, wrx) \in \mathcal{A}^*$ if $r(\text{tail}(w), x) \in \mathcal{A}$ and $r(w, x) \in \mathcal{A}^*$ if $\{a\}(x) \in \mathcal{A}$ for some $a \in \Gamma$ and $r(\text{tail}(w), x) \in \mathcal{A}$, for $r \in \mathbf{N}_R$;
- $r(wr^-x, w) \in \mathcal{A}^*$ if $r(x, \text{tail}(w)) \in \mathcal{A}$ and $r(x, w) \in \mathcal{A}^*$ if $\{a\}(x) \in \mathcal{A}$ for some $a \in \Gamma$ and $r(x, \text{tail}(w)) \in \mathcal{A}$, for $r \in \mathbf{N}_R$;
- $\{a\}(x) \in \mathcal{A}^*$ if $\{a\}(x) \in \mathcal{A}$, for $a \in \Gamma$ and $x \in \text{ind}(\mathcal{A})$.

Derivation Trees. Fix an \mathcal{ELIO}_u -ontology \mathcal{O} in normal form and an ABox \mathcal{A} , $x_0 \in \text{ind}(\mathcal{A})$ and $A_0 \in \mathbf{N}_C$. Let $\Theta_1 = \text{ind}(\mathcal{A}) \cup (\mathbf{N}_I \cap \text{sig}(\mathcal{O}))$, and $\Theta_2 = \mathbf{N}_C \cap \text{sig}(\mathcal{O}) \cup \{\{a\} \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})\} \cup \{\exists u.A \mid A \in \mathbf{N}_C \cap \text{sig}(\mathcal{O})\}$. A *derivation tree* for the assertion $A_0(x_0)$ in \mathcal{O}, \mathcal{A} is a finite $\Theta_1 \times \Theta_2$ -labeled tree (T, V) , where T is a set of nodes and $V : T \rightarrow \Theta_1 \times \Theta_2$ the labeling function, such that:

- $V(\varepsilon) = (x_0, A_0)$;
- If $V(n) = (x, C)$ with $x \in \text{ind}(\mathcal{A})$, then $C(x) \in \mathcal{A}$ or $\mathcal{O} \models \top \sqsubseteq C$ or
 1. n has successors n_1, \dots, n_k , $k \geq 1$ with $V(n_i) = (a_i, C_i)$, such that $a_i = x$ or $a_i \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$ for all i , and defining $C'_i = C_i$ if $a_i = x$, and $C'_i = \exists u.(\{a_i\} \sqcap C_i)$ otherwise, we have $\mathcal{O} \models C'_1 \sqcap \dots \sqcap C'_k \sqsubseteq C$; or
 2. $C = \exists u.A$ and n has a single successor n' with $V(n') = (y, \exists u.A)$; or
 3. n has a single successor n' with $V(n') = (y, A)$ such that $r(x, y) \in \mathcal{A}$ and $\mathcal{O} \models \exists r.A \sqsubseteq C$ (where r is a role name or an inverse role).
- If $V(n) = (a, C)$ with $a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$, then $C = \{a\}$ or:
 4. There exists $x \in \text{ind}(\mathcal{A})$ such that n has successors n_1, \dots, n_k , $k \geq 1$ with $V(n_i) = (a_i, C_i)$ and $a_i = x$ or $a_i \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$ for all i , and, defining $C'_i = C_i$ if $a_i = x$, and $C'_i = \exists u.(\{a_i\} \sqcap C_i)$ otherwise, we have $\mathcal{O} \models C'_1 \sqcap \dots \sqcap C'_k \sqsubseteq \exists u.(\{a\} \sqcap C)$.

Note that a special case of rule 1 is when n has two successors labeled $(x, \{a\})$ and (a, C) , and a special case of rule 4 is when n has two successors labeled $(x, \{a\})$ and (x, C) .

We now prove the analogue of Lemma 3 for \mathcal{ELIO}_u , except not considering the size of derivation trees.

Lemma 12. *Let \mathcal{O} be an \mathcal{ELIO}_u -ontology in normal form and \mathcal{A} a finite $\text{sig}(\mathcal{O})$ -ABox. Then*

1. $\mathcal{O}, \mathcal{A} \models A_0(x_0)$ iff and only if there is a derivation tree for $A_0(x_0)$ in \mathcal{O}, \mathcal{A} .
2. If (T, V) is a derivation tree for $A_0(x_0)$ in \mathcal{O}, \mathcal{A} , then one can construct a derivation tree (T', V') for $A_0(x_0)$ in $\mathcal{O}, \mathcal{A}^*$, with \mathcal{A}^* the undirected unfolding of \mathcal{A} , and such that $T = T'$.

Proof. We start with the proof of Part 1. (\Leftarrow) is straightforward. For (\Rightarrow), we construct a sequence of ABoxes $\mathcal{A}_0, \mathcal{A}_1, \dots$ generalized with assertions of the form $(\exists u.A)(x)$. Take $\mathcal{A}_0 = \mathcal{A} \cup \{\{a\}(x_a) \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})\}$ where the x_a 's are fresh individual variables. Let \mathcal{A}_{i+1} be obtained from \mathcal{A}_i by applying one of the following rule, where C is a concept of the form $C \in \mathbf{N}_C$ or $C = \{a\}$ or $C = \exists u.A$, and $x, y \in \text{ind}(\mathcal{A}_i)$:

1. if $C_1(x_1), \dots, C_k(x_k) \in \mathcal{A}_i$, with $x_i = x$ or $x_i = x_{a_i}$ for some $a_i \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$, and $\mathcal{O} \models C'_1 \sqcap \dots \sqcap C'_k \sqsubseteq C$, where $C'_i = C_i$ if $x_i = x$ and $C'_i = \exists u.(\{a_i\} \sqcap C_i)$ if $x_i = x_{a_i}$, then add $C(x)$;
2. if $(\exists u.A)(y) \in \mathcal{A}_i$ then add $(\exists u.A)(x)$;
3. if $r(x, y), A(y) \in \mathcal{A}_i$ and $\mathcal{O} \models \exists r.A \sqsubseteq C$, then add $C(x)$;
4. if $C_1(x_1), \dots, C_k(x_k) \in \mathcal{A}_i$, with $x_i = x$ or $x_i = x_{a_i}$ for some $a_i \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$, and $\mathcal{O} \models C'_1 \sqcap \dots \sqcap C'_k \sqsubseteq \exists u.(\{a\} \sqcap C)$, where $C'_i = C_i$ if $x_i = x$ and $C'_i = \exists u.(\{a_i\} \sqcap C_i)$ if $x_i = x_{a_i}$, then add $C(x_a)$.

Note that the sequence is finite, and denote by \mathcal{A}^* the final ABox.

Claim. There is a model \mathcal{I}, v of \mathcal{A}^* and \mathcal{O} such that for all $x \in \text{ind}(\mathcal{A})$ and $A \in \mathbf{N}_C$, $v(x)^{\mathcal{I}} \in A^{\mathcal{I}}$ implies $A(x) \in \mathcal{A}^*$.

Proof of the Claim. For all $x, y \in \text{ind}(\mathcal{A}^*)$, we write $x \sim y$ if $\{a\}(x), \{a\}(y) \in \mathcal{A}^*$ for some $a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O})$. Notice that if $\{a\}(x), \{a\}(y), C(x) \in \mathcal{A}^*$, then $C(x_a) \in \mathcal{A}^*$ by rule 4, and $C(y) \in \mathcal{A}^*$ by rule 1. Therefore, $x \sim y$ implies $C(x) \in \mathcal{A}^*$ if and only if $C(y) \in \mathcal{A}^*$. In particular, \sim is an equivalence relation. We let $[x]$ denote the equivalence class of x . Start with an interpretation \mathcal{I}_0 defined by:

$$\begin{aligned} \Delta^{\mathcal{I}_0} &= \text{ind}(\mathcal{A}^*) / \sim \\ A^{\mathcal{I}_0} &= \{[x] \mid A(x) \in \mathcal{A}^*\} \\ a^{\mathcal{I}_0} &= [x_a] \\ r^{\mathcal{I}_0} &= \{([x], [y]) \mid r(x, y) \in \mathcal{A}^*\}. \end{aligned}$$

Let C_x denote the conjunction of all concepts of the form $C \in \mathbf{N}_C$, $C = \{a\}$, $C = \exists u.A$, or $C = \exists u.(\{a\} \sqcap A)$ such that $\mathcal{A}^* \models C(x)$. Let \mathcal{I}_x denote the canonical model for \mathcal{O} and C_x rooted at $[x]$. Due to rule 1 and the universality of \mathcal{I}_x , for every concept name or nominal C , we have $[x] \in C^{\mathcal{I}_0}$

if and only if $[x] \in C^{\mathcal{I}_x}$. Similarly, because of rule 4, for every $a \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})$, $a^{\mathcal{I}_x} \in C^{\mathcal{I}_x}$ if and only if $a^{\mathcal{I}_0} \in C^{\mathcal{I}_0}$.

We can now define \mathcal{I} as follows: $\Delta^{\mathcal{I}}$ is the disjoint union of $\Delta^{\mathcal{I}_0}$ and all elements in domains $\Delta^{\mathcal{I}_x} \setminus (\{[x]\} \cup \{a^{\mathcal{I}_x} \mid a \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})\})$. Interpretations of concept names and nominals are inherited from the \mathcal{I}_0 or \mathcal{I}_x each element comes from. Finally, $r^{\mathcal{I}}$ is obtained by taking the union of $r^{\mathcal{I}_0}$ and all $r^{\mathcal{I}_x}$ after replacing edges to/from $a^{\mathcal{I}_x}$ with edges to/from $a^{\mathcal{I}_0}$. It is clear that for the variable assignment $v(x) = [x]$, \mathcal{I}_0, v satisfies \mathcal{A}^* , and thus so does \mathcal{I}, v .

By rule 1, all concept inclusions of \mathcal{O} of the form $\top \sqsubseteq A$, $A_1 \sqcap A_2 \sqsubseteq B$, $A \sqsubseteq \{a\}$ and $\{a\} \sqsubseteq A$ are satisfied by \mathcal{I}_0 . They are also satisfied by every \mathcal{I}_x (since \mathcal{I}_x is a model of \mathcal{O}), and thus by \mathcal{I} . Now consider a concept inclusion $A \sqsubseteq \exists r.B \in \mathcal{O}$, where r is a role name or an inverse role. Recall that for every a and x , $a^{\mathcal{I}} \in B^{\mathcal{I}}$ if and only if $a^{\mathcal{I}_x} \in B^{\mathcal{I}_x}$. Therefore, for all $d \in \Delta^{\mathcal{I}_x}$, $d \in (\exists r.B)^{\mathcal{I}_x}$ implies $d \in (\exists r.B)^{\mathcal{I}}$. The case $A \sqsubseteq \exists u.B$ is similar. Since every \mathcal{I}_x satisfies $A \sqsubseteq \exists r.B$, so does \mathcal{I} . Similarly, every concept inclusion $\exists r.B \sqsubseteq A \in \mathcal{O}$ is satisfied in \mathcal{I} : if the witness pair for $\exists r.B$ is part of \mathcal{I}_0 , this follows from rule 3, and if not, then it is part of some \mathcal{I}_x , which is by definition a model of \mathcal{O} . For concept inclusions of the form $\exists u.B \sqsubseteq A \in \mathcal{O}$, we can observe that if there exists some $d' \in \Delta^{\mathcal{I}}$ such that $d' \in B^{\mathcal{I}}$, then $(\exists u.B)$ is in C_x for some x , i.e., by rule 2, for all x .

Finally, for all $x \in \text{ind}(\mathcal{A})$ and $A \in \mathbf{N}_c$, $[x]^{\mathcal{I}} \in A^{\mathcal{I}}$ implies $[x]^{\mathcal{I}_0} \in A^{\mathcal{I}_0}$, i.e., $A(x) \in \mathcal{A}^*$. This concludes the proof of the claim.

Now suppose $\mathcal{O}, \mathcal{A} \models A_0(x_0)$. By the Claim, we have $A_0(x_0) \in \mathcal{A}^*$. Since the four rules to construct $\mathcal{A}_0, \mathcal{A}_1, \dots$ are in one-to-one correspondence with Conditions (1)–(4) from the definition of derivation trees, we can inductively construct a derivation tree for $A_0(x_0)$ in \mathcal{A} w.r.t. \mathcal{O} . This concludes the proof of Part 1.

The proof of Part 2 is similar to that of Lemma 3. We define (T, V') as follows from (T, V) , starting with the root by setting $V'(\varepsilon) = V(\varepsilon) = (x_0, A_0)$. At each step, if $V(n) = (a, C)$ then $V'(n) = (w, C)$ for some w such that $\text{tail}(w) = a$. To define the labelings of the successors of n , we consider the possible derivation steps for (a, C) in \mathcal{A} .

1. $a = x \in \text{ind}(\mathcal{A})$, and n has successors n_1, \dots, n_k , $k \geq 1$ with $V(n_i) = (a_i, C_i)$, such that $a_i = x$ or $a_i \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})$ for all i , and defining $C'_i = C_i$ if $a_i = x$, and $C'_i = \exists u.(\{a_i\} \sqcap C_i)$ otherwise, we have $\mathcal{O} \models C'_1 \sqcap \dots \sqcap C'_k \sqsubseteq C$. Take $V'(n_i) = (w, C_i)$ if $x_i = x$, and $V'(n_i) = (a_i, C_i)$ if $a_i \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})$.
2. $C = \exists u.A$ and n has a single successor n' with $V(n') = (y, \exists u.A)$. Take $V'(n') = (y, \exists u.A)$.
3. n has a single successor n' with $V(n') = (y, A)$ such that $r(a, y) \in \mathcal{A}$ and $\mathcal{O} \models \exists r.A \sqsubseteq C$ (where r is a role name or an inverse role). Take $V'(n') = (wry, A)$.
4. $a \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})$ and there exists $x \in \text{ind}(\mathcal{A})$ such that n has successors n_1, \dots, n_k , $k \geq 1$ with $V(n_i) = (a_i, C_i)$ and $a_i = x$ or $a_i \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})$ for all i , and, defining $C'_i = C_i$ if $a_i = x$, and $C'_i = \exists u.(\{a_i\} \sqcap C_i)$ otherwise,

we have $\mathcal{O} \models C'_1 \sqcap \dots \sqcap C'_k \sqsubseteq \exists u.(\{a\} \sqcap C)$. Take $V'(n_i) = (x, C_i)$ if $x_i = x$, and $V'(n_i) = (a_i, C_i)$ if $a_i \in \mathbf{N}_1 \cap \text{sig}(\mathcal{O})$.

Then (T, V') is a derivation tree for $A_0(x_0)$ in \mathcal{A}^* w.r.t. \mathcal{O} . \square

Tree Automata. A *tree* is a non-empty set $T \subseteq (\mathbb{N} \setminus \{0\})^*$ closed under prefixes and such that $n \cdot (i + 1) \in T$ implies $n \cdot i \in T$. It is *k-ary* if $T \subseteq \{1, \dots, k\}^*$. The node ε is the *root* of T . As a convention, we take $n \cdot 0 = n$ and $(n \cdot i) \cdot -1 = n$. Note that $\varepsilon \cdot -1$ is undefined. Given an alphabet Θ , a Θ -*labeled tree* is a pair (T, L) consisting of a tree T and a node-labeling function $L : T \rightarrow \Theta$.

A *non-deterministic tree automaton (NTA)* over finite *k-ary trees* is a tuple $\mathfrak{A} = (Q, \Theta, I, \Delta)$, where Q is a set of states, Θ is the input alphabet, $I \subseteq Q$ is the set of initial states, and $\Delta \subseteq Q \times \Theta \times \bigcup_{0 \leq \ell \leq k} Q^\ell$ is the transition relation. A *run* of an NTA $\mathfrak{A} = (Q, \Theta, I, \Delta)$ over a *k-ary input* (T, L) is a Q -labeled tree (T, r) such that for all $x \in T$ with children y_1, \dots, y_ℓ , $(r(x), L(x), r(y_1), \dots, r(y_\ell)) \in \Delta$. It is *accepting* if $r(\varepsilon) \in I$. The language accepted by \mathfrak{A} , denoted $L(\mathfrak{A})$, is the set of all finite *k-ary* Θ -labeled trees over which \mathfrak{A} has an accepting run.

A *two-way alternating tree automaton over finite k-ary trees (2ATA)* is a tuple $\mathfrak{A} = (Q, \Theta, q_0, \delta)$ where Q is a finite set of *states*, Θ is the *input alphabet*, $q_0 \in Q$ is the *initial state*, and δ is a *transition function*. The transition function δ maps every state q and input letter $\theta \in \Theta$ to a positive Boolean formula $\delta(q, \theta)$ over the truth constants *true* and *false* and *transition atoms* of the form $(i, q) \in [k] \times Q$, where $[k] = \{-1, 0, 1, \dots, k\}$. The semantics is given in terms of *runs*. More precisely, let (T, L) be a finite *k-ary* Θ -labeled tree and $\mathfrak{A} = (Q, \Theta, q_0, \delta)$ a 2ATA. An *accepting run of* \mathfrak{A} *over* (T, L) is a $(T \times Q)$ -labeled tree (T_r, r) such that:

1. $r(\varepsilon) = (\varepsilon, q_0)$, and
2. for all $y \in T_r$ with $r(y) = (x, q)$, there is a subset $S \subseteq [k] \times Q$ such that $S \models \delta(q, L(x))$ and for every $(i, q') \in S$, there is some successor y' of y in T_r with $r(y') = (x \cdot i, q')$.

The language accepted by \mathfrak{A} , denoted $L(\mathfrak{A})$, is the set of all finite *k-ary* Θ -labeled trees (T, L) for which there is an accepting run.

From a 2ATA \mathfrak{A} , one can compute in exponential time an NTA \mathfrak{A}' whose number of states is exponential in the number of states of \mathfrak{A} and such that $L(\mathfrak{A}) = L(\mathfrak{A}')$ (Vardi 1998).

Interpolant Existence. We now give the proof that Point 2 in Theorem 6 entails an exponential time upper bound for deciding the existence of an interpolant. We focus on the case of \mathcal{ELIO}_u . Let $\mathcal{O}_1, \mathcal{O}_2$ be \mathcal{ELIO}_u -ontologies in normal form, $A, B \in \mathbf{N}_c$, and $\Sigma = \text{sig}(\mathcal{O}_1, A) \cap \text{sig}(\mathcal{O}_2, B)$. We can assume that $A \in \text{sig}(\mathcal{O}_1)$ and $B \in \text{sig}(\mathcal{O}_2)$.

As our proof relies on tree automata, let us first explain how we represent ABoxes that are tree-shaped modulo $\mathbf{N}_1 \cap$

Σ as trees over the alphabet 2^Λ , where

$$\begin{aligned}\Lambda = & \mathbf{N}_C \cap \Sigma \cup \\ & \{\{a\} \mid a \in \mathbf{N}_I \cap \Sigma\} \cup \\ & \{r, r^- \mid r \in \mathbf{N}_R \cap \Sigma\} \cup \\ & \{\exists r. \{a\} \mid r \in \mathbf{N}_R \cap \Sigma \wedge a \in \mathbf{N}_I \cap \Sigma\} \cup \\ & \{\exists r^-. \{a\} \mid r \in \mathbf{N}_R \cap \Sigma \wedge a \in \mathbf{N}_I \cap \Sigma\}.\end{aligned}$$

Intuitively, the nodes of the tree correspond to the individual variables of the ABox; labels $C \in \mathbf{N}_C$, $\{a\}$, $\exists r. \{a\}$, $\exists r^-. \{a\}$ indicate concepts that hold at the current node, while labels r or r^- are used to indicate which roles (if any) connect a node to its parent. Note that there need not be such a label r or r^- , so connected nodes in the tree representation are not necessarily connected in the ABox.

More precisely, we associate with every 2^Λ -labeled tree (T, L) the following ABox, where x_a are fresh individual variables:

$$\begin{aligned}\mathcal{A}_{(T,L)} = & \{\top(x) \mid x \in T\} \cup \\ & \{\{a\}(x_a) \mid \exists x \in T : \{a\} \in L(x)\} \cup \\ & \{\{a\}(x) \mid x \in T \wedge \{a\} \in L(x)\} \cup \\ & \{B(x) \mid x \in T \wedge B \in L(x)\} \cup \\ & \{r(x, x \cdot i) \mid x \cdot i \in T \wedge r \in \mathbf{N}_R \wedge r \in L(x \cdot i)\} \cup \\ & \{r(x \cdot i, x) \mid x \cdot i \in T \wedge r \in \mathbf{N}_R \wedge r^- \in L(x \cdot i)\} \cup \\ & \{r(x, x_a) \mid x \in T \wedge \exists r. \{a\} \in L(x)\} \cup \\ & \{r(x_a, x) \mid x \in T \wedge \exists r^-. \{a\} \in L(x)\}.\end{aligned}$$

Notice that $\mathcal{A}_{(T,L)}$ is tree-shaped modulo $\mathbf{N}_I \cap \Sigma$. Conversely, for every ABox \mathcal{A} that is tree-shaped modulo $\mathbf{N}_I \cap \Sigma$, there exists a (not necessarily unique) tree (T, L) such that $\mathcal{A} = \mathcal{A}_{(T,L)}$. In addition, if the degree of G_A^u is less than k , then there exists a k -ary tree (T, L) such that $\mathcal{A} = \mathcal{A}_{(T,L)}$. For instance, $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma$ can be represented by a k -ary tree for any k larger than the number of concept inclusions in $\mathcal{O}_1 \cup \mathcal{O}_2$.

We also denote by $\mathcal{A}_{(T,L)}^\Sigma$ the Σ -reduct of $\mathcal{A}_{(T,L)}$.

We describe below an NTA \mathfrak{A}_1 with exponentially many states accepting trees that represent prefix-closed finite subsets of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma$, and a 2-ATA \mathfrak{A}_2 with polynomially many states accepting trees (T, L) such that $\mathcal{A}_{(T,L)} \models B(\varepsilon)$. The existence of an interpolant then reduces to the non-emptiness of $L(\mathfrak{A}_1) \cap L(\mathfrak{A}_2)$.

Definition of \mathfrak{A}_1 . We represent the canonical model for $\mathcal{O}_1 \cup \mathcal{O}_2$ and A by a tree with τ_A at the root of the tree, other $\tau \in S$ inserted at arbitrary positions in the tree, and $\tau_0 r_1 \tau_1 \cdots r_n \tau_n$ below $\tau_0 r_1 \tau_1 \cdots r_{n-1} \tau_{n-1}$ if $n > 0$. We want \mathfrak{A}_1 to accept finite subsets of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma$ obtained by keeping a prefix-closed finite subset of nodes, and possibly removing some concepts and relations from the labels (including all concepts and relations not in Σ). To do so, the automaton will simply guess in its state the type of each node, and check that all guesses are locally consistent by allowing only transitions that match the definition of canonical

models. Concretely, the states of the automaton consist of a pair of \mathcal{O} -types, where state (τ, τ') should be interpreted as the parent node having type τ and the current node type τ' .

To keep the definition simple, the automaton also accepts trees where, compared to the canonical model, some nodes are duplicated (that is, we do not require that the node corresponding to some $\tau \in \Delta^{\mathcal{I}_{\mathcal{O}, A}} \cap S$ is unique). This does not change the set of concepts entailed at the root.

We take $\mathfrak{A}_1 = (Q_1, 2^\Lambda, I_1, \Delta_1)$, where

- $Q_1 = (S \cup \{\perp\}) \times S$, where S is the set of \mathcal{O} -types introduced in the definition of $\mathcal{I}_{\mathcal{O}, A}$;
- $I_1 = \{(\perp, \tau_A)\}$;
- For states $q = (\tau, \tau')$, $q_1 = (\tau_1, \tau'_1), \dots, q_\ell = (\tau_\ell, \tau'_\ell) \in Q_1$ and input letter $\alpha \subseteq \Lambda$, $(q, \alpha, q_1, \dots, q_\ell) \in \Delta_1$ if the following conditions are satisfied, for all $1 \leq i \leq \ell$:
 - the current state and label are consistent with the definition of the canonical model: for all $r \in \alpha$, $\tau \rightsquigarrow_r \tau'$;
 - the set of concepts associated with α is a subset of the \mathcal{O} -type τ' : $\alpha \cap (\text{sub}(\mathcal{O}) \cup \text{sub}^\exists(\mathcal{O})) \subseteq \tau'$;
 - the current type τ' is stored in the state of all child nodes: for all $1 \leq i \leq \ell$, $\tau_i = \tau'$.

Note that \mathfrak{A}_1 can be computed in exponential time.

Lemma 13. $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma \models B(\rho_A)$ if and only if there exists $(T, L) \in L(\mathfrak{A}_1)$ such that $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{(T,L)} \models B(\varepsilon)$, where ε is the root of (T, L) .

Proof. The run of \mathfrak{A}_1 on some $(T, L) \in L(\mathfrak{A}_1)$ can be used to define a homomorphism from $\mathcal{A}_{(T,L)}, \varepsilon$ to $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma, \rho_A$. Therefore, if $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{(T,L)} \models B(\varepsilon)$ then $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma \models B(\rho_A)$. Conversely, if $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma \models B(\rho_A)$ then there exists a finite subset \mathcal{A} of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma$ such that $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A} \models B(\rho_A)$. Take as (T, L) any finite prefix of an encoding of $\mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma$ that contains all nodes corresponding to individuals in \mathcal{A} . Then the labeling of (T, L) with the full types from the canonical model defines an accepting run of \mathfrak{A}_1 on (T, L) , and $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{(T,L)} \models B(\varepsilon)$. \square

Definition of \mathfrak{A}_2 . The construction of $\mathfrak{A}_2 = (Q_2, 2^\Lambda, q_B, \delta_2)$ relies on derivation trees. Intuitively, runs of \mathfrak{A}_2 on some (T, L) correspond to derivation trees

for $B(\varepsilon)$ in $\mathcal{O}_1 \cup \mathcal{O}_2$, $\mathcal{A}_{(T,L)}$. The states of \mathfrak{A}_2 are

$$\begin{aligned} Q_2 = & \{q_{A'} \mid A' \in \mathbf{N}_C \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\} \cup \\ & \{q_{\{a\}} \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\} \cup \\ & \{q_{\exists r.A'}, q_{\exists r^-.A'} \mid r \in \mathbf{N}_R \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2), \\ & \quad A' \in \mathbf{N}_C \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\} \cup \\ & \{q_{\exists u.A'} \mid A' \in \mathbf{N}_C \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\} \cup \\ & \{q_{\exists u.(\{a\} \sqcap A')} \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2), \\ & \quad A' \in \mathbf{N}_C \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\} \cup \\ & \{q_{\exists u.(\{a\} \sqcap \{b\})} \mid a, b \in \mathbf{N}_I \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\} \cup \\ & \{q_r, q_{r^-} \mid r \in \mathbf{N}_R \cap \Sigma\} \cup \\ & \{q_{\exists r.\{a\}}, q_{\exists r^-. \{a\}} \mid a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2), \\ & \quad r \in \mathbf{N}_R \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)\}. \end{aligned}$$

Intuitively, state q_C is used to check that C is entailed at the current node. States q_r and $q_{\exists r.\{a\}}$ are used to check the label of the current node. The initial state is q_B , as we are trying to construct a derivation tree for B at the root.

Let us now define the transition relation. From a state q_r or $q_{\exists r.\{a\}}$, where $r \in \mathbf{N}_R \cup \mathbf{N}_R^-$ and $a \in \mathbf{N}_I$, the automaton simply checks the current label:

$$\begin{aligned} \delta_2(q_r, \alpha) &= \begin{cases} \text{true} & \text{if } r \in \alpha \\ \text{false} & \text{if } r \notin \alpha \end{cases} \\ \delta_2(q_{\exists r.\{a\}}, \alpha) &= \begin{cases} \text{true} & \text{if } \exists r.\{a\} \in \alpha \\ \text{false} & \text{if } \exists r.\{a\} \notin \alpha. \end{cases} \end{aligned}$$

From a state $q_{\exists r.A'}$, with $r \in \mathbf{N}_R \cup \mathbf{N}_R^-$ and $A' \in \mathbf{N}_C$, the automaton checks that the current node has an r -successor from which there exists a run starting in $q_{A'}$. This r -successor can be (i) the parent of the current node, i.e. there is a run from q_{r^-} from the current node and a run from $q_{A'}$ from the parent node, (ii) some i -th child of the current node, i.e. there is a run from q_r and one from $q_{A'}$ from the i -th child, or (iii) an individual a , i.e. there is a run from $q_{\exists r.\{a\}}$ and from $q_{\exists u.(\{a\} \sqcap A')}$ from the current node:

$$\begin{aligned} \delta_2(q_{\exists r.A'}, \alpha) &= (0, q_{r^-}) \wedge (-1, q_{A'}) \vee \\ & \quad \bigvee_{1 \leq i \leq k} (i, q_{A'}) \wedge (i, q_r) \vee \\ & \quad \bigvee_{a \in \mathbf{N}_I \cap \Gamma} (0, q_{\exists r.\{a\}}) \wedge (0, q_{\exists u.(\{a\} \sqcap A')}). \end{aligned}$$

From a state $q_{\exists u.A'}$, the automaton checks if (i) condition 1 from derivation trees can be applied, that is, there exist concepts C_1, \dots, C_n of the form $B', \{a\}, \exists u.(\{a\} \sqcap B')$ or $\exists u.(\{a\} \sqcap \{b\})$ such that $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqcap \dots \sqcap C_n \models \exists u.A'$ and there exists a run from each q_{C_i} from the current node, or (ii) condition 2 from derivation trees can be applied, which can be checked by propagating the search for a run from $q_{\exists u.A'}$ to all neighbouring nodes, or (iii) condition 3 from derivation trees can be applied, that is, there exists r, B' such that $\mathcal{O}_1 \cup \mathcal{O}_2 \models \exists r.B' \sqsubseteq \exists u.A'$ and the automaton has

a run from $q_{\exists r.B'}$ starting from the current node:

$$\begin{aligned} \delta_2(q_{\exists u.A'}, \alpha) &= \bigvee_{\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqcap \dots \sqcap C_n \models \exists u.A'} \bigwedge_{1 \leq i \leq n} (0, q_{C_i}) \vee \\ & \quad \bigvee_{i \in \{-1, 1, \dots, k\}} (i, q_{\exists u.A'}) \vee \\ & \quad \bigvee_{\mathcal{O}_1 \cup \mathcal{O}_2 \models \exists r.B' \sqsubseteq \exists u.A'} (0, q_{\exists r.B'}). \end{aligned}$$

From a state $q_{\exists u.C}$ where $C = \{a\} \sqcap A'$ or $C = \{a\} \sqcap \{b\}$ with $b \neq a$, the automaton checks if condition 4 from derivation trees can be applied either (i) taking the current node as x , that is, there exist concepts C_1, \dots, C_n of the form $B', \{b\}, \exists u.(\{b\} \sqcap B')$ or $\exists u.(\{b\} \sqcap \{c\})$ such that $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqcap \dots \sqcap C_n \models \exists u.C$ and there exists a run from each q_{C_i} from the current node, or (ii) taking some other node as x , which can be checked by propagating the search for a run from $q_{\exists u.C}$ to all neighbouring nodes:

$$\begin{aligned} \delta_2(q_{\exists u.C}, \alpha) &= \bigvee_{\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqcap \dots \sqcap C_n \models \exists u.C} \bigwedge_{1 \leq i \leq n} (0, q_{C_i}) \vee \\ & \quad \bigvee_{i \in \{-1, 1, \dots, k\}} (i, q_{\exists u.C}). \end{aligned}$$

We also set

$$\delta_2(q_{\exists u.(\{a\} \sqcap \{a\})}, \alpha) = \text{true}.$$

For $C = \{a\}$ or $C \in \mathbf{N}_C$, $\delta(q_C, \alpha) = \text{true}$ if $C \in \alpha$ or $\mathcal{O}_1 \cup \mathcal{O}_2 \models \top \sqsubseteq C$, and otherwise, the automaton checks if conditions 1 or 3 from derivation trees can be applied:

$$\begin{aligned} \delta_2(q_C, \alpha) &= \bigvee_{\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqcap \dots \sqcap C_n \models C} \bigwedge_{1 \leq i \leq n} (0, q_{C_i}) \vee \\ & \quad \bigvee_{\mathcal{O}_1 \cup \mathcal{O}_2 \models \exists r.B' \sqsubseteq C} (0, q_{\exists r.B'}). \end{aligned}$$

Lemma 14. *For all finite k -ary 2^Λ -labeled trees (T, L) , we have $(T, L) \in L(\mathfrak{A}_2)$ if and only if $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{(T,L)} \models B(\varepsilon)$.*

Proof. We observe that for all (T, L) ,

- For all $\text{sig}(\mathcal{O}_1, \mathcal{O}_2)$ -concept C of the form $C = A'$, $C = \{a\}$ or $C = \exists u.A'$ with $A' \in \mathbf{N}_C$ and $a \in \mathbf{N}_I$, \mathfrak{A} has a run starting from state q_C on (T, L) if and only if there exists a derivation tree for (ε, C) in $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{(T,L)}$.
- For all $a \in \mathbf{N}_I \cap \text{sig}(\mathcal{O}_1, \mathcal{O}_2)$, for all $\text{sig}(\mathcal{O}_1, \mathcal{O}_2)$ -concept $C = \{b\}$ or $C = A' \in \mathbf{N}_C$, \mathfrak{A} has a run starting from state $q_{\exists u.(\{a\} \sqcap C)}$ if and only if there exists a derivation tree for (a, C) in $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{(T,L)}$. \square

From \mathfrak{A}_2 , one can construct an equivalent NTA \mathfrak{A}'_2 with exponentially many states (Vardi 1998). By Lemmas 13 and 14, we have $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}'_{\mathcal{O}_1 \cup \mathcal{O}_2} \models B(\rho_A)$ if and only if $L(\mathfrak{A}_1) \cap L(\mathfrak{A}'_2) = \emptyset$, which can be checked in exponential time.

Lower Bound for Explicit Definitions. We construct an \mathcal{ELI} -ontology \mathcal{O} , signature Σ , and concept name A such that the smallest explicit $\mathcal{ELI}(\Sigma)$ -definition of A under \mathcal{O} is of double exponential size in $\|\mathcal{O}\|$. \mathcal{O} is a variant of ontologies constructed in (Lutz and Wolter 2010; Nikitina and Rudolph 2014) and defined as follows. It contains $\top \sqsubseteq \exists r. \top \sqcap \exists s. \top$,

$$\begin{aligned} A &\sqsubseteq M \sqcap \overline{X_0} \sqcap \dots \sqcap \overline{X_n} \\ \exists \sigma^-. (\overline{X_i} \sqcap X_0 \sqcap \dots \sqcap X_{i-1}) &\sqsubseteq X_i & \sigma \in \{r, s\}, i \leq n \\ \exists \sigma^-. (X_i \sqcap X_0 \sqcap \dots \sqcap X_{i-1}) &\sqsubseteq \overline{X_i} & \sigma \in \{r, s\}, i \leq n \\ \exists \sigma^-. (\overline{X_i} \sqcap \overline{X_j}) &\sqsubseteq \overline{X_i} & \sigma \in \{r, s\}, j < i \leq n \\ \exists \sigma^-. (X_i \sqcap \overline{X_j}) &\sqsubseteq X_i & \sigma \in \{r, s\}, j < i \leq n \\ X_0 \sqcap \dots \sqcap X_n &\sqsubseteq L \end{aligned}$$

and

$$L \sqsubseteq B, \quad \exists r. B \sqcap \exists s. B \sqsubseteq B, \quad B \sqcap M \sqsubseteq A.$$

Let $\Sigma = \{M, r, s, L\}$. Note that A triggers a marker M and a binary tree of depth 2^n using counter concept names X_0, \dots, X_n and $\overline{X_0}, \dots, \overline{X_n}$. A concept name L is made true at the leaves. Conversely, if L is true at the leaves of a binary tree of depth 2^n then B is true at all nodes of the tree and A is entailed by M and B at its root. Define inductively

$$C_0 = L, \quad C_{k+1} = \exists r. C_k \sqcap \exists s. C_k, \quad C = C_{2^n} \sqcap M.$$

Then C is the smallest explicit $\mathcal{ELI}(\Sigma)$ -definition of A under \mathcal{O} .

Transfer Sequences. For the proof of “2. \Rightarrow 3.” of Theorem 6 and the proof that interpolants can be computed in double exponential time we require an extension of the notion of transfer sequences first introduced in (Bienvenu, Lutz, and Wolter 2013) to logics with nominals.

Assume that Condition 2 of Theorem 6 holds. So we have $\mathcal{ELI}\mathcal{O}_u$ -ontologies $\mathcal{O}_1, \mathcal{O}_2$ in normal form, concept names A, B , and $\Sigma = \text{sig}(\mathcal{O}_1, A) \cap \text{sig}(\mathcal{O}_2, B)$ such that $\mathcal{O}_1 \cup \mathcal{O}_2, \mathcal{A}_{\mathcal{O}_1 \cup \mathcal{O}_2, A}^\Sigma \models \overline{B}(\rho_A)$. Set $\mathcal{O} = \mathcal{O}_1 \cup \mathcal{O}_2$. We use $\mathcal{A}_{\mathcal{O}, A}$ to denote the ABox associated with the canonical model $\mathcal{I}_{\mathcal{O}, A}$. We require some notation for the individuals that occur in $\mathcal{A}_{\mathcal{O}, A}$. We set $a \sim b$ if $\mathcal{O} \models \{a\} \sqcap \exists u. A \sqsubseteq \{b\}$ and set $[a] = \{b \in \text{sig}(\mathcal{O}) \mid a \sim b\}$. We say that concept name E is *absorbed* by a if $\mathcal{O} \models E \sqcap \exists u. A \sqsubseteq \{a\}$. We denote the individual x_{τ_a} of $\mathcal{A}_{\mathcal{O}, A}$ by x_a and the individuals x_{τ_E} of $\mathcal{A}_{\mathcal{O}, A}$ by x_E . Note that $x_a = x_b$ if $a \sim b$ and $x_a = x_A$ if A is absorbed by a .

Given $w \in \text{ind}(\mathcal{A}_{\mathcal{O}, A})$, we call the individuals of the form $ww' \in \text{ind}(\mathcal{A}_{\mathcal{O}, A})$ the *subtree* of $\mathcal{A}_{\mathcal{O}, A}$ rooted at w .

By compactness we have a finite subset \mathcal{A} of $\mathcal{A}_{\mathcal{O}, A}$ containing x_A such that $\mathcal{O}, \mathcal{A}_{|\Sigma} \models B(x_A)$. We may assume that \mathcal{A} is prefix closed and that $\mathcal{A}_{|\Sigma}$ contains

- $\{a\}(x_a)$ and $A(x_A)$ for all $a, A \in \Sigma$;
- $\top(x_A)$ and $\top(x_a)$ for all $a, A \in \text{sig}(\mathcal{O}) \setminus \Sigma$;

We obtain the ABox \mathcal{A}_Σ from $\mathcal{A}_{|\Sigma}$ by adding the assertions

- $\{a\}(x_{a, \text{new}})$ and $\top(x_{A, \text{new}})$, for all $a, A \in \text{sig}(\mathcal{O}) \setminus \Sigma$, where $x_{a, \text{new}}$ and $x_{A, \text{new}}$ are fresh individuals.

Let I denote the set of individuals x_a, x_A with $a, A \in \text{sig}(\mathcal{O})$ and let I_{new} denote the set of individuals $x_{a, \text{new}}, x_{A, \text{new}}$ with $a, A \in \text{sig}(\mathcal{O}) \setminus \Sigma$. Observe that $\mathcal{O}, \mathcal{A}_\Sigma$ and $\mathcal{O}, \mathcal{A}_{|\Sigma}$ entail the same assertions $C(a)$ for $a \in \text{ind}(\mathcal{A}_{|\Sigma})$, so the additional individuals do not influence what is entailed. In fact, we introduce the individuals I_{new} only to enable explicit bookkeeping about when in a transfer sequence (defined below) an assertion of the form $C(a)$ or $\exists u. A$ is derived.

We aim to define a small subset \mathcal{A}' of \mathcal{A}_Σ such that $\mathcal{O} \models A \sqsubseteq C$ for the concept C corresponding to \mathcal{A}' and such that still $\mathcal{O}, \mathcal{A}' \models B(x_A)$. If \mathcal{A}' has at most exponential depth in the size of \mathcal{O} then we are done, as then \mathcal{A}' is of at most double exponential size in the size of \mathcal{O} . We obtain \mathcal{A}' from \mathcal{A}_Σ by determining w and $ww' \in \text{ind}(\mathcal{A}) \setminus (I \cup I_{\text{new}})$ which behave ‘sufficiently similar’ such that if we obtain \mathcal{A}' from \mathcal{A} by replacing the subtree rooted at w in \mathcal{A}_Σ by the subtree rooted at ww' , then we still have $\mathcal{O}, \mathcal{A}' \models B(x_A)$ and $\mathcal{O} \models A \sqsubseteq C$ for the concept C defined by \mathcal{A}' . The replacement of subtrees is then performed exhaustively.

For w and ww' to be sufficiently similar, we firstly require that $\text{tail}(w) = \text{tail}(ww')$ (with $\text{tail}(w)$ the final type in w for any w). This ensures that $\mathcal{O} \models A \sqsubseteq C$ for the concept C corresponding to \mathcal{A}' . This also has the consequence that \mathcal{A}' is (isomorphic) to a prefix closed subABox of \mathcal{A}_Σ . For the second condition for being sufficiently similar, we apply the notion of transfer sequences (Bienvenu, Lutz, and Wolter 2013). To define transfer sequences, we consider derivations using \mathcal{O} and *intermediate* ABoxes \mathcal{B} such that

$$I \cup I_{\text{new}} \subseteq \text{ind}(\mathcal{B}) \subseteq \text{ind}(\mathcal{A}_\Sigma)$$

We admit \mathcal{B} to contain equations $x_e = x_{e'}$ for $x_e, x_{e'} \in I \cup I_{\text{new}}$, with the obvious semantics. Consider such an intermediate \mathcal{B} and $w \in \text{ind}(\mathcal{B}) \setminus (I \cup I_{\text{new}})$. Then the set $D_{\mathcal{B}}(w)$ is defined as the set of assertions α with $\mathcal{O}, \mathcal{B}' \models \alpha$ and α of the form

- $A(c)$ or $\{a\}(c)$ with $A, a \in \text{sig}(\mathcal{O})$ and $c \in \{w\} \cup I \cup I_{\text{new}}$; or
- $r(w, c)$ with $r \in \mathbb{N}_R \cup \mathbb{N}_R^-$ and $c \in I \cup I_{\text{new}}$;
- $r(c, d)$ with $r \in \mathbb{N}_R \cup \mathbb{N}_R^-$ and $c, d \in I \cup I_{\text{new}}$;
- $c = d$ with $c, d \in I \cup I_{\text{new}}$.

and $\mathcal{B}' = \mathcal{B} \cup \{A(x_{A, \text{new}}) \mid \mathcal{O}, \mathcal{B} \models \exists u. A\}$. For $w \in \text{ind}(\mathcal{B}) \setminus (I \cup I_{\text{new}})$, let

- \mathcal{B}_w^\downarrow denote the restriction of \mathcal{B} to the individuals in the subtree of \mathcal{B} rooted at w and $I \cup I_{\text{new}}$; and let
- \mathcal{B}_w^\uparrow be the ABox obtained from \mathcal{B} by dropping \mathcal{B}_w^\downarrow from \mathcal{B} except for w itself and $I \cup I_{\text{new}}$.

Define the *transfer sequence* $\mathcal{X}_0, \mathcal{X}_1, \dots$ of (\mathcal{A}_Σ, w) w.r.t. \mathcal{O} as follows:

$$\begin{aligned} \mathcal{X}_0 &= D_{(\mathcal{A}_\Sigma)_w^\downarrow}(w) \\ \mathcal{X}_1 &= D_{(\mathcal{A}_\Sigma)_w^\uparrow \cup \mathcal{X}_0}(w) \\ \mathcal{X}_2 &= D_{(\mathcal{A}_\Sigma)_w^\downarrow \cup \mathcal{X}_1}(w) \\ \mathcal{X}_3 &= \dots \end{aligned}$$

Intuitively, we first consider the set \mathcal{X}_0 of assertions that are entailed by \mathcal{O} and \mathcal{A}_Σ at $\{w\} \cup I \cup I_{\text{new}}$ if we only use assertions in $\mathcal{A}_{\Sigma_w^\downarrow}$. We update \mathcal{A}_Σ by those assertions. Next we consider the set \mathcal{X}_1 of assertions that are entailed by \mathcal{O} and the updated \mathcal{A}_Σ at $\{w\} \cup I \cup I_{\text{new}}$ if we only use assertions in the updated $\mathcal{A}_{\Sigma_w^\downarrow}$. We update \mathcal{A}_Σ again, and so on. It is not difficult to see that if $w, ww' \in \text{ind}(\mathcal{A}_\Sigma) \setminus (I \cup I_{\text{new}})$ and

- the restrictions of \mathcal{A}_Σ to $\{w\} \cup (I \cup I_{\text{new}})$ and $\{ww'\} \cup (I \cup I_{\text{new}})$ coincide (modulo renaming w to ww') and
- the transfer sequences of (\mathcal{A}_Σ, w) w.r.t. \mathcal{O} coincides with the transfer sequence of $(\mathcal{A}_\Sigma, ww')$ w.r.t. \mathcal{O} (modulo renaming w to ww')

then one can replace $\mathcal{A}_{\Sigma_w^\downarrow}$ by $\mathcal{A}_{\Sigma_{ww'}^\downarrow}$ in \mathcal{A}_Σ and it still holds that $\mathcal{O}, \mathcal{A}' \models B(x_A)$ for the resulting ABox \mathcal{A}' . If in addition we require that $\text{tail}(w) = \text{tail}(ww')$, then the resulting ABox is (isomorphic to) a prefix closed sub ABox of \mathcal{A}_Σ and so the concept corresponding to the ABox \mathcal{A}' is still entailed by A w.r.t. \mathcal{O} .

By performing the above replacement exhaustively, we obtain a prefix closed subset \mathcal{A} of \mathcal{A}_Σ that is of depth $\leq 2^q(\|\mathcal{O}\|)$ with q a polynomial and therefore has the properties required for Point 3 of Theorem 6. Such an \mathcal{A} can be constructed in at most double exponential time since one can construct the canonical model $\mathcal{I}_{\mathcal{O}, \mathcal{A}}$ up to nodes of depth $\leq 2^q(\|\mathcal{O}\|)$ in double exponential time.

The claims stated in Theorem 6 for interpolants without the universal role are shown by modifying the proof above in a straightforward way.

F Proofs for Sections 8 and 9

We first complete the proof of Theorem 8 by showing that there is a Horn- \mathcal{ALCI} -simulation between the interpretations \mathcal{I} and \mathcal{I}' defined in Figure 5. The definition of Horn-simulations is as follows. For any two sets X and Y and a binary relation R , we set

- $XR^\uparrow Y$ if for all $x \in X$ there exists $y \in Y$ with $(x, y) \in R$;
- $XR^\downarrow Y$ if for all $y \in Y$ there exists $x \in X$ with $(x, y) \in R$.

A relation $Z \subseteq \mathcal{P}(\Delta^{\mathcal{I}}) \times \Delta^{\mathcal{I}'}$ is a *Horn- $\mathcal{ALCI}(\Sigma)$ -simulation* between \mathcal{I} and \mathcal{I}' if $(X, b) \in Z$ implies $X \neq \emptyset$ and the following hold:

- for any $A \in \Sigma$, if $(X, b) \in Z$ and $X \subseteq A^{\mathcal{I}}$, then $b \in A^{\mathcal{I}'}$;
- for any role r in Σ , if $(X, b) \in Z$ and $XR^{\uparrow} Y$, then there exist $Y' \subseteq Y$ and $b' \in \Delta^{\mathcal{I}'}$ with $(b, b') \in r^{\mathcal{I}'}$ and $(Y', b') \in Z$;
- for any role r in Σ , if $(X, b) \in Z$ and $(b, b') \in r^{\mathcal{I}'}$, then there is $Y \subseteq \Delta^{\mathcal{I}}$ with $XR^{\downarrow} Y$ and $(Y, b') \in Z$;
- if $(X, b) \in Z$, then $\mathcal{I}', b \preceq_{\mathcal{ELI}, \Sigma} \mathcal{I}, a$ for every $a \in X$ (where $\preceq_{\mathcal{ELI}, \Sigma}$ indicates that we have a simulation that does not only respect role names in Σ but also the inverse of role names in Σ).

We write $\mathcal{I}, X \preceq_{\text{Horn}, \Sigma} \mathcal{I}', b$ if there exists a Horn- $\mathcal{ALCI}(\Sigma)$ -simulation Z between \mathcal{I} and \mathcal{I}' such that $(X, b) \in Z$. It is shown in (Jung et al. 2019) that if $\mathcal{I}, X \preceq_{\text{Horn}, \Sigma} \mathcal{I}', b$, then all Horn- $\mathcal{ALCI}(\Sigma)$ -concepts true in all nodes in X are also true in b .

Now observe that the relation Z between $2^{\Delta^{\mathcal{I}'}}$ and $\Delta^{\mathcal{I}'}$ containing all pairs $(\{x\}, x')$, $(\{b, c\}, b'')$, and $(\{d, e\}, d'')$ is a Horn- $\mathcal{ALCI}(\Sigma)$ -simulation between the interpretations \mathcal{I} and \mathcal{I}' defined in Figure 5, as required.

We next observe that moving to the Horn fragment Horn-GF of the guarded fragment is not sufficient to obtain a logic in which interpolants/explicit definitions always exist. To this end we modify the ontology given in the proof of Theorem 8. In detail, let \mathcal{O}' contain the following CIs:

$$\begin{aligned} A &\sqsubseteq B \\ B &\sqsubseteq \forall r.F \\ F &\sqsubseteq \exists r_1.D_1 \sqcap \exists r_2.D_2 \sqcap \exists r_1.M \sqcap \exists r_2.M \\ A &\sqsubseteq \forall r.((F \sqcap \exists r_1.(D_1 \sqcap M) \sqcap \exists r_2.(D_2 \sqcap M)) \rightarrow E) \\ B &\sqsubseteq \exists r.C \\ C &\sqsubseteq F \sqcap \forall r_1.D_1 \sqcap \forall r_2.D_2 \end{aligned}$$

and also $B \sqcap \exists r.(C \sqcap E) \sqsubseteq A$. Define the signature Σ by setting $\Sigma = \{B, D_1, D_2, E, r, r_1, r_2\}$. We note that, intuitively, the third and fourth CI should be read as

$$\begin{aligned} F &\sqsubseteq \exists r_1.D_1 \sqcap \exists r_2.D_2 \\ A &\sqsubseteq \forall r.((F \sqcap \forall r_1.D_1 \sqcap \forall r_2.D_2) \rightarrow E) \end{aligned}$$

and the concept name M is introduced to achieve this in a projective way as the latter CI is not in Horn- \mathcal{ALCI} .

We first observe that A is implicitly definable from Σ under \mathcal{O}' since

$$\mathcal{O}' \models A \equiv B \sqcap \forall r.(\forall r_1.D_1 \sqcap \forall r_2.D_2 \rightarrow E).$$

We next sketch the proof that A is not explicitly Horn-GF(Σ)-definable under \mathcal{O}' . For a definition of Horn-GF and Horn-GF simulations we refer the reader to (Jung et al. 2019). Now consider the interpretations \mathcal{I} and \mathcal{I}' defined in Figure 9. Both \mathcal{I} and \mathcal{I}' are models of \mathcal{O}' , $a \in A^{\mathcal{I}}$, $a' \notin A^{\mathcal{I}'}$, but $a \in F^{\mathcal{I}}$ implies $a' \in F^{\mathcal{I}'}$ holds for every Horn-GF(Σ)-formula F , and the claim follows. The latter can be proved by observing that there exists a Horn-GF(Σ)-simulation between \mathcal{I} and \mathcal{I}' (Jung et al. 2019) containing $(\{a\}, a)$. In fact, one can show that the relation Z containing all pairs $(\{x\}, x')$, $(\{b, c\}, b'')$, and $(\{d, e\}, d'')$ is a Horn-GF(Σ)-simulation.

We finally make a few observations regarding the Horn fragment of first-order logic. Recall that *Horn-FO* is defined as the closure of formulas of the form $R(\vec{t})$,

$$R_1(\vec{t}_1) \wedge \dots \wedge R_n(\vec{t}_n) \rightarrow R(\vec{t}), \quad R_1(\vec{t}_1) \wedge \dots \wedge R_n(\vec{t}_n) \rightarrow \perp$$

under conjunction, universal quantification, and existential quantification, where $\vec{t}_1, \dots, \vec{t}_n, \vec{t}$ are sequences of individual variables and individual names (Chang and Keisler 1998). According to Exercise 6.2.6 in (Chang and Keisler 1998) Horn-FO has the following property.

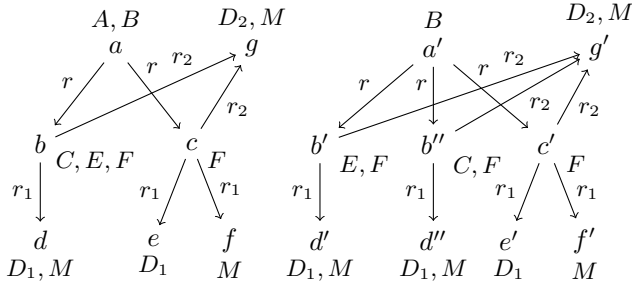


Figure 9: Interpretations \mathcal{I} (left) and \mathcal{I}' (right) used for \mathcal{O}' .

Theorem 9. Let φ, ψ be sentences in Horn-FO such that $\varphi \wedge \psi$ is not satisfiable. Then there exists a sentence χ in Horn-FO such that $\text{sig}(\chi) \subseteq \text{sig}(\varphi) \cap \text{sig}(\psi)$, $\varphi \models \chi$, and $\chi \wedge \psi$ is not satisfiable.

We directly obtain the following interpolation result.

Theorem 10. Let $\mathcal{O}_1, \mathcal{O}_2$ be Horn- \mathcal{ALCTO}_u -ontologies and let C_1, C_2 be Horn- \mathcal{ALCTO}_u -concepts such that $\mathcal{O}_1 \cup \mathcal{O}_2 \models C_1 \sqsubseteq C_2$. Then there exists a formula $\chi(x)$ in Horn-FO such that

- $\text{sig}(\chi) \subseteq \text{sig}(\mathcal{O}_1, C_1) \cap \text{sig}(\mathcal{O}_2, C_2)$;
- $\mathcal{O}_1 \models \forall x(C_1(x) \rightarrow \chi(x))$;
- $\mathcal{O}_2 \models \forall x(\chi(x) \rightarrow C_2(x))$.

Proof. Take a fresh unary relation symbol $A(x)$ and a fresh individual name c . Let φ be the conjunction of all sentences in $\mathcal{O}_1 \cup \{C_1(c)\}$ and let ψ be the conjunction of all sentences in $\mathcal{O}_2 \cup \{\forall x(C_2(x) \leftrightarrow A(x)), \neg A(c)\}$. Then φ and ψ are both equivalent to sentences in Horn-FO. By definition $\varphi \wedge \psi$ is not satisfiable. Thus there exists a Horn-FO sentence χ using only c and symbols in $\text{sig}(\mathcal{O}_1, C_1) \cap \text{sig}(\mathcal{O}_2, C_2)$ such that $\varphi \models \chi$ and $\chi \wedge \psi$ is not satisfiable. Thus:

- $\mathcal{O}_1 \models C_1(c) \rightarrow \chi$;
- $\mathcal{O}_2 \cup \{\forall x(C_2(x) \leftrightarrow A(x))\} \models \chi \rightarrow A(c)$.

Replace c by x in $\chi, C_1(c)$, and $A(c)$. Then

- $\mathcal{O}_1 \models \forall x(C_1(x) \rightarrow \chi(x))$;
- $\mathcal{O}_2 \models \forall x(\chi(x) \rightarrow C_2(x))$,

as required. \square

Applied to Horn- \mathcal{ALCT} ontologies and concepts we thus always obtain an interpolant in Horn-FO and an interpolant in \mathcal{ALCT} (since \mathcal{ALCT} enjoys the CIP (ten Cate, Franconi, and Seylan 2013)).

It would be interesting to find out whether there exists an interpolant in the intersection of Horn-FO and \mathcal{ALCT} and whether it is possible to give an informative syntactic description of that intersection.