

Enriching $\mathcal{EL}$ -Concepts with Greatest Fixpoints

Carsten Lutz¹ and Robert Piro² and Frank Wolter²

Abstract. We investigate the expressive power and computational complexity of $\mathcal{EL}^\nu$, the extension of the lightweight description logic $\mathcal{EL}$ with concept constructors for greatest fixpoints. It is shown that $\mathcal{EL}^\nu$ has the same expressive power as $\mathcal{EL}$ extended with simulation quantifiers and that it can be characterized as a largest fragment of monadic second-order logic that is preserved under simulations and has finite minimal models. As in basic $\mathcal{EL}$, all standard reasoning problems for general TBoxes can be solved in polynomial time. $\mathcal{EL}^\nu$ has a range of very desirable properties that $\mathcal{EL}$ itself is lacking. Firstly, least common subsumers w.r.t. general TBoxes as well as most specific concepts always exist and can be computed in polynomial time. Secondly, $\mathcal{EL}^\nu$ shares with $\mathcal{EL}$ the Craig interpolation property and the Beth definability property, but in contrast to $\mathcal{EL}$ allows the computation of interpolants and explicit concept definitions in polynomial time.

1 INTRODUCTION

The well-known description logic (DL) $\mathcal{ALC}$ is usually regarded as the basic DL that comprises all Boolean concept constructors and from which more expressive DLs are derived by adding further expressive means. This fundamental role of $\mathcal{ALC}$ is largely due to its well-behavedness regarding logical, model-theoretic, and computational properties which can, in turn, be explained nicely by the fact that $\mathcal{ALC}$ -concepts can be characterized as the bisimulation invariant fragment of first-order logic (FO): an FO formula is invariant under bisimulation if, and only if, it is equivalent to an $\mathcal{ALC}$ -concept [23, 12, 16]. For example, invariance under bisimulation can explain the tree-model property of $\mathcal{ALC}$ and its favorable computational properties [25]. In the above characterization, the condition that $\mathcal{ALC}$ is a fragment of FO is much less important than its bisimulation invariance. In fact, $\mathcal{ALC}\mu$, the extension of $\mathcal{ALC}$ with fixpoint operators, is not a fragment of FO, but inherits almost all important properties of $\mathcal{ALC}$ [7, 11, 20]. Similar to $\mathcal{ALC}$, $\mathcal{ALC}\mu$'s fundamental role (in particular in its formulation as the modal μ -calculus) can be explained by the fact that $\mathcal{ALC}\mu$ -concepts comprise exactly the bisimulation invariant fragment of monadic second-order logic (MSO) [14, 7]. Indeed, from a purely theoretical viewpoint it is hard to explain why $\mathcal{ALC}$ rather than $\mathcal{ALC}\mu$ forms the logical underpinning of current ontology language standards; the facts that μ -calculus concepts can be hard to grasp and that, despite the same theoretical complexity, efficient reasoning in $\mathcal{ALC}\mu$ is more challenging than in $\mathcal{ALC}$ are probably the main reasons.

In recent years, the development of very large ontologies and the use of ontologies to access instance data has led to a revival of inter-

est in *tractable* DLs. The main examples are $\mathcal{EL}$ [4] and DL-Lite [8], the logical underpinnings of the OWL profiles OWL2 EL and OWL2 QL, respectively. In contrast to $\mathcal{ALC}$, a satisfactory characterization of the expressivity of such DLs is still missing, and a first aim of this paper is to fill this gap for $\mathcal{EL}$. To this end, we characterize $\mathcal{EL}$ as a maximal fragment of FO that is preserved under *simulations* and has *finite minimal models*. Note that preservation under simulations alone would characterize $\mathcal{EL}$ with disjunctions, and the existence of minimal models reflects the “Horn-logic character” of $\mathcal{EL}$.

The second and main aim of this paper, however, is to introduce and investigate two equi-expressive extensions of $\mathcal{EL}$ with greatest fixpoints, $\mathcal{EL}^\nu$ and $\mathcal{EL}^{\nu+}$, and to prove that they stand in a similar relationship to $\mathcal{EL}$ as $\mathcal{ALC}\mu$ to $\mathcal{ALC}$. To this end, we prove that $\mathcal{EL}^\nu$ (and therefore also $\mathcal{EL}^{\nu+}$, which admits mutual fixpoints and is exponentially more succinct than $\mathcal{EL}^\nu$) can be characterized as a maximal fragment of MSO that is preserved under *simulations* and has *finite minimal models*. Similar to $\mathcal{ALC}\mu$, $\mathcal{EL}^\nu$ and $\mathcal{EL}^{\nu+}$ inherit many good properties of $\mathcal{EL}$ such as its Horn-logic character and the crucial fact that *reasoning with general concept inclusions (GCIs) is still tractable*. In contrast to $\mathcal{ALC}\mu$, the development of practical decision procedures is thus no obstacle to using $\mathcal{EL}^{\nu+}$. Moreover, $\mathcal{EL}^{\nu+}$ has a number of very useful properties that $\mathcal{EL}$ and most of its extensions are lacking. To begin with, we show that in $\mathcal{EL}^{\nu+}$ *least common subsumers (lcs)* w.r.t. general TBoxes always exist and can be computed in polynomial time (for a bounded number of concepts). This result can be regarded as an extension of similar results for least common subsumers w.r.t. *classical* TBoxes in $\mathcal{EL}$ with greatest fixpoint semantics in [2]. Similarly, in $\mathcal{EL}^{\nu+}$ *most specific concepts* always exist and can be computed in linear time; a result which also generalizes [2]. Secondly, we show that $\mathcal{EL}^{\nu+}$ has the *Beth definability property* with explicit definitions being computable in polytime and of polynomial size. It has been convincingly argued in [22, 21] that this property is of great interest for structuring TBoxes and for ontology based data access. Another application of $\mathcal{EL}^{\nu+}$ is demonstrated in [15], where the succinct representations of definitions in $\mathcal{EL}^{\nu+}$ are used to develop polytime algorithms for decomposing certain general $\mathcal{EL}$ -TBoxes.

To prove these result and provide a better understanding of the modeling capabilities of $\mathcal{EL}^{\nu+}$ we show that it has the same expressive power as extensions of $\mathcal{EL}$ by means of *simulation quantifiers*, a variant of second-order quantifiers that quantifies “modulo a simulation of the model”; in fact, the relationship between simulation quantifiers and $\mathcal{EL}^{\nu+}$ is somewhat similar to the relationship between $\mathcal{ALC}\mu$ and bisimulation quantifiers [10].

Proofs are omitted for brevity. The reader is referred to [17] for a version of this paper containing all proofs.

¹ Department of Computer Science, University of Bremen, Germany, email: clu@informatik.uni-bremen.de

² Department of Computer Science, University of Liverpool, UK, email: {Robert.Piro,wolter}@liverpool.ac.uk

2 PRELIMINARIES

Let N_C and N_R be countably infinite and mutually disjoint sets of concept and role names. $\mathcal{EL}$ -concepts are built according to the rule

$$C := A \mid \top \mid \perp \mid C \sqcap D \mid \exists r.C,$$

where $A \in N_C$, $r \in N_R$, and C, D range over $\mathcal{EL}$ -concepts³. An $\mathcal{EL}$ -concept inclusion takes the form $C \sqsubseteq D$, where C, D are $\mathcal{EL}$ -concepts. As usual, we use $C \equiv D$ to abbreviate $C \sqsubseteq D, D \sqsubseteq C$. A general $\mathcal{EL}$ -TBox $\mathcal{T}$ is a finite set of $\mathcal{EL}$ -concept inclusions. An ABox assertion is an expression of the form $A(a)$ or $r(a, b)$, where a, b are from a countably infinite set of individual names N_I , $A \in N_C$, and $r \in N_R$. An ABox is a finite set of ABox assertions. By $\text{Ind}(\mathcal{A})$ we denote the set of individual names in $\mathcal{A}$. An $\mathcal{EL}$ -knowledge base (KB) is a pair $(\mathcal{T}, \mathcal{A})$ that consists of an $\mathcal{EL}$ -TBox $\mathcal{T}$ and an ABox $\mathcal{A}$.

The semantics of $\mathcal{EL}$ is based on interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where the domain $\Delta^{\mathcal{I}}$ is a non-empty set, and $\cdot^{\mathcal{I}}$ is a function mapping each concept name A to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$, each role name r to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$, and each individual name a to an element $a^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$. The interpretation $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ of $\mathcal{EL}$ -concepts C in an interpretation $\mathcal{I}$ is defined in the standard way [5], and so are models of TBoxes, ABoxes, and KBs. We will often make use of the fact that $\mathcal{EL}$ -concepts can be regarded as formulas in FO (and, therefore, MSO) with unary predicates from N_C , binary predicates from N_R , and exactly one free variable [5]. We will often not distinguish between $\mathcal{EL}$ -concepts and their translations into FO/MSO.

We now introduce $\mathcal{EL}^{\nu}$, the extension of $\mathcal{EL}$ with greatest fixpoints and the main language studied in this paper. $\mathcal{EL}^{\nu}$ -concepts are defined like $\mathcal{EL}$ -concepts, but additionally allow the greatest fixpoint constructor $\nu X.C$, where X is from a countably infinite set of (concept) variables N_V and C an $\mathcal{EL}^{\nu}$ -concept. A variable is *free* in a concept C if it occurs in C at least once outside the scope of any ν -constructor that binds it. An $\mathcal{EL}^{\nu}$ -concept is *closed* if it does not contain any free variables. An $\mathcal{EL}^{\nu}$ -concept inclusion takes the form $C \sqsubseteq D$, where C, D are closed $\mathcal{EL}^{\nu}$ -concepts. The semantics of the greatest fixpoint constructor is as follows, where $\mathcal{V}$ is an *assignment* that maps variables to subsets of $\Delta^{\mathcal{I}}$ and $\mathcal{V}[X \mapsto W]$ denotes $\mathcal{V}$ modified by setting $\mathcal{V}(X) = W$:

$$(\nu X.C)^{\mathcal{I}, \mathcal{V}} = \bigcup \{W \subseteq \Delta^{\mathcal{I}} \mid W \subseteq C^{\mathcal{I}, \mathcal{V}[X \mapsto W]}\}$$

Example 1 For the concept $C = \nu X.(\exists \text{has_parent}.X)$, we have $e \in C^{\mathcal{I}}$ if, and only if, there is an infinite has_parent-chain starting at e in $\mathcal{I}$, i.e., there exist $e_0, e_1, e_2, \dots$ such that $e = e_0$ and $(e_i, e_{i+1}) \in \text{has_parent}^{\mathcal{I}}$ for all $i \geq 0$.

We can now form the TBox $\mathcal{T} = \{\text{Human_being} \sqsubseteq C\}$ stating that every human being has an infinite chain of parents.

We will also consider an extended version of the ν -constructor that allows to capture mutual recursion. It has been considered e.g. in [9, 24] and used in a DL context in [20]; it can be seen as a variation of the fixpoint equations considered in [7]. The constructor has the form $\nu_i X_1 \dots X_n. C_1, \dots, C_n$ where $1 \leq i \leq n$. The semantics is defined by setting $(\nu_i X_1 \dots X_n. C_1, \dots, C_n)^{\mathcal{I}, \mathcal{V}}$ to

$$\bigcup \{W_i \mid \exists W_1, \dots, W_{i-1}, W_{i+1}, \dots, W_n \text{ s.t. for } 1 \leq j \leq n: \\ W_j \subseteq C_j^{\mathcal{I}, \mathcal{V}[X_1 \mapsto W_1, \dots, X_n \mapsto W_n]}\}$$

We use $\mathcal{EL}^{\nu+}$ to denote $\mathcal{EL}$ extended with this mutual greatest fixpoint constructor. Clearly, $\nu X.C \equiv \nu_1 X.C$, thus every $\mathcal{EL}^{\nu}$ -concept

³ In the literature, $\mathcal{EL}$ is typically defined without $\perp$. The sole purpose of including $\perp$ here is to simplify the formulation of some results.

is equivalent to an $\mathcal{EL}^{\nu+}$ -concept. We now consider the converse direction. Firstly, the following proposition follows immediately from well known results on mutual fixpoint constructors [7].

Proposition 2 *For every $\mathcal{EL}^{\nu+}$ -concept one can construct an equivalent $\mathcal{EL}^{\nu}$ -concept.*

In this paper, we define the *length* of a concept C as the number of occurrences of symbols in it. Then the translation in Proposition 2 yields an exponential blow-up and one can show that indeed there is a sequence of $\mathcal{EL}^{\nu+}$ -concepts $C_0, C_1, \dots$ such that C_i is of length $p(i)$, p a polynomial, whereas the shortest $\mathcal{EL}^{\nu}$ -concept equivalent to C_i is of length at least 2^i [17].

By extending the translation of $\mathcal{EL}$ -concepts into FO in the obvious way, one can translate closed $\mathcal{EL}^{\nu+}$ -concepts into MSO formulas with one free first-order variable. We will often not distinguish between $\mathcal{EL}^{\nu+}$ -concepts and their translation into MSO.

3 CHARACTERIZING $\mathcal{EL}$ USING SIMULATIONS

The purpose of this section is to provide a model-theoretic characterization of $\mathcal{EL}$ as a fragment of FO that is similar in spirit to the well-known characterization of $\mathcal{ALC}$ as the bisimulation-invariant fragment of FO. To this end, we first characterize $\mathcal{EL}^{\sqcup}$, the extension of $\mathcal{EL}$ with the disjunction constructor $\sqcup$, as the fragment of FO that is preserved under simulation. Then we characterize the fragment $\mathcal{EL}$ of $\mathcal{EL}^{\sqcup}$ using, in addition, the existence of minimal models. A *pointed interpretation* is a pair $(\mathcal{I}, d)$ consisting of an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$. A *signature* Σ is a set of concept and role names.

Definition 3 (Simulations) Let $(\mathcal{I}_1, d_1)$ and $(\mathcal{I}_2, d_2)$ be pointed interpretations and Σ a signature. A relation $S \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is a Σ -simulation between $(\mathcal{I}_1, d_1)$ and $(\mathcal{I}_2, d_2)$, in symbols $S : (\mathcal{I}_1, d_1) \leq_{\Sigma} (\mathcal{I}_2, d_2)$, if $(d_1, d_2) \in S$ and the following conditions hold:

1. for all concept names $A \in \Sigma$ and all $(e_1, e_2) \in S$, if $e_1 \in A^{\mathcal{I}_1}$ then $e_2 \in A^{\mathcal{I}_2}$;
2. for all role names $r \in \Sigma$, all $(e_1, e_2) \in S$, and all $e'_1 \in \Delta^{\mathcal{I}_1}$ with $(e_1, e'_1) \in r^{\mathcal{I}_1}$, there exists $e'_2 \in \Delta^{\mathcal{I}_2}$ such that $(e_2, e'_2) \in r^{\mathcal{I}_2}$ and $(e'_1, e'_2) \in S$.

If such an S exists, then we also say that $(\mathcal{I}_2, d_2)$ Σ -simulates $(\mathcal{I}_1, d_1)$ and write $(\mathcal{I}_1, d_1) \leq_{\Sigma} (\mathcal{I}_2, d_2)$.

If $\Sigma = N_C \cup N_R$, then we omit Σ and use the term *simulation* to denote Σ -simulations and $(\mathcal{I}_1, d_1) \leq (\mathcal{I}_2, d_2)$ stands for $(\mathcal{I}_1, d_1) \leq_{\Sigma} (\mathcal{I}_2, d_2)$. It is well-known that the description logic $\mathcal{EL}$ is intimately related to the notion of a simulation, see for example [3, 18]. In particular, $\mathcal{EL}$ -concepts are preserved under simulations in the sense that if $d_1 \in C^{\mathcal{I}_1}$ for an $\mathcal{EL}$ -concept C and $(\mathcal{I}_1, d_1) \leq (\mathcal{I}_2, d_2)$, then $d_2 \in C^{\mathcal{I}_2}$. This observation, which clearly generalizes to $\mathcal{EL}^{\sqcup}$, illustrates the (limitations of the) modeling capabilities of $\mathcal{EL}/\mathcal{EL}^{\sqcup}$. We now strengthen it to an exact characterization of the expressive power of these logics relative to FO.

Let $\varphi(x)$ be an FO-formula (or, later, MSO-formula) with one free variable x . We say that $\varphi(x)$ is *preserved under simulations* if, and only if, for all $(\mathcal{I}_1, d_1)$ and $(\mathcal{I}_2, d_2)$, $\mathcal{I}_1 \models \varphi[d_1]$ and $(\mathcal{I}_1, d_1) \leq (\mathcal{I}_2, d_2)$ implies $\mathcal{I}_2 \models \varphi[d_2]$.

Theorem 4 *An FO-formula $\varphi(x)$ is preserved under simulations if, and only if, it is equivalent to an $\mathcal{EL}^{\sqcup}$ -concept.*

To characterize $\mathcal{EL}$, we add a central property of Horn-logics on top of preservation under simulations. Let $\mathcal{L}$ be a set of FO (or, later, MSO) formulas, each with one free variable. We say that $\mathcal{L}$ *has (finite) minimal models* if, and only if, for every $\varphi(x) \in \mathcal{L}$ there exists a (finite) pointed interpretation $(\mathcal{I}, d)$ such that for all $\psi(x) \in \mathcal{L}$, we have $\mathcal{I} \models \psi[d]$ if, and only if, $\forall x.(\varphi(x) \rightarrow \psi(x))$ is a tautology.

Theorem 5 *The set of $\mathcal{EL}$ -concepts is a maximal set of FO-formulas that is preserved under simulations and has minimal models (equivalently: has finite minimal models): if $\mathcal{L}$ is a set of FO-formulas that properly contains all $\mathcal{EL}$ -concepts, then either it contains a formula not preserved under simulations or it does not have (finite) minimal models.*

We note that de Rijke and Kurtonina have given similar characterizations of various non-Boolean fragments of $\mathcal{ALC}$. In particular, Theorem 4 is rather closely related to results proved in [16] and would certainly have been included in the extensive list of characterizations given there had $\mathcal{EL}$ already been as popular as it is today. In contrast, the novelty of Theorem 5 is that it makes the Horn character of $\mathcal{EL}$ explicit through minimal models while the characterizations of disjunction-free languages in [16] are based on simulations that take sets (rather than domain-elements) as arguments.

4 SIMULATION QUANTIFIERS AND $\mathcal{EL}^\nu$

To understand and characterize the expressive power and modeling capabilities of $\mathcal{EL}^\nu$, we introduce three distinct types of simulation quantifiers and show that, in each case, the resulting language has the same expressive power as $\mathcal{EL}^\nu$.

Simulating interpretations. The first language $\mathcal{EL}^{si}$ extends $\mathcal{EL}$ by the concept constructor $\exists^{sim}(\mathcal{I}, d)$, where $(\mathcal{I}, d)$ is a finite pointed interpretation in which only finitely many $\sigma \in \mathbf{N}_C \cup \mathbf{N}_R$ have a non-empty interpretation $\sigma^\mathcal{I} \subseteq \Delta^\mathcal{I}$. The semantics of $\exists^{sim}(\mathcal{I}, d)$ is defined by setting for all interpretations $\mathcal{J}$ and $e \in \Delta^\mathcal{J}$,

$$e \in (\exists^{sim}(\mathcal{I}, d))^\mathcal{J} \text{ iff } (\mathcal{I}, d) \leq (\mathcal{J}, e).$$

Example 6 Let $\mathcal{I}$ be an interpretation such that $\Delta^\mathcal{I} = \{d\}$, $(d, d) \in \text{has.parent}^\mathcal{I}$, and $\sigma^\mathcal{I} = \emptyset$ for all remaining role and concept names σ . Then $\exists^{sim}(\mathcal{I}, d)$ is equivalent to the concept $\nu X.(\exists \text{has.parent}.X)$ from Example 1.

To attain a better understanding of the constructor $\exists^{sim}$, it is interesting to observe that every $\mathcal{EL}^{si}$ -concept is equivalent to a concept of the form $\exists^{sim}(\mathcal{I}, d)$.

Lemma 7 *For every $\mathcal{EL}^{si}$ -concept C one can construct, in linear time, an equivalent concept of the form $\exists^{sim}(\mathcal{I}, d)$.*

Proof By induction on the construction of C . If $C = A$ for a concept name A , then let $\mathcal{I} = (\{d\}, \cdot^\mathcal{I})$, where $A^\mathcal{I} = \{d\}$ and $\sigma^\mathcal{I} = \emptyset$ for all symbols distinct from A . Clearly, A and $\exists^{sim}(\mathcal{I}, d)$ are equivalent. For $C_1 = \exists^{sim}(\mathcal{I}_1, d_1)$ and $C_2 = \exists^{sim}(\mathcal{I}_2, d_2)$ assume that $\Delta^{\mathcal{I}_1} \cap \Delta^{\mathcal{I}_2} = \{d_1\} = \{d_2\}$. Then $\exists^{sim}(\mathcal{I}_1 \cup \mathcal{I}_2, d_1)$ is equivalent to $C_1 \sqcap C_2$, where $\Delta^{\mathcal{I}_1 \cup \mathcal{I}_2} = \Delta^{\mathcal{I}_1} \cup \Delta^{\mathcal{I}_2}$, and $\sigma^{\mathcal{I}_1 \cup \mathcal{I}_2} = \sigma^{\mathcal{I}_1} \cup \sigma^{\mathcal{I}_2}$ for all $\sigma \in \mathbf{N}_C \cup \mathbf{N}_R$. For $C = \exists r. \exists^{sim}(\mathcal{I}, d)$ construct a new interpretation $\mathcal{I}'$ by adding a new node e to $\Delta^\mathcal{I}$ and setting $(e, d) \in r^\mathcal{I}'$. Then $\exists^{sim}(\mathcal{I}', e)$ and C are equivalent.

We will show that there are polynomial translations between $\mathcal{EL}^{si}$ and $\mathcal{EL}^{\nu+}$. When using $\mathcal{EL}^{\nu+}$ in applications and to provide a translation from $\mathcal{EL}^{\nu+}$ to $\mathcal{EL}^{si}$, it is convenient to have available a “syn-tactic” simulation operator.

Simulating models of TBoxes. The second language $\mathcal{EL}^{st}$ extends $\mathcal{EL}$ by the concept constructor $\exists^{sim}\Sigma.(\mathcal{T}, C)$, where Σ is a finite signature, $\mathcal{T}$ a general TBox, and C a concept. To admit nestings of $\exists^{sim}$, the concepts of $\mathcal{EL}^{st}$ are defined by simultaneous induction; namely, $\mathcal{EL}^{st}$ -concepts, concept inclusions, and general TBoxes are defined as follows:

- every $\mathcal{EL}$ -concept, concept inclusion, and general TBox is an $\mathcal{EL}^{st}$ -concept, concept inclusion, and general TBox, respectively;
- if $\mathcal{T}$ is a general $\mathcal{EL}^{st}$ -TBox, C an $\mathcal{EL}^{st}$ -concept, and Σ a finite signature, then $\exists^{sim}\Sigma.(\mathcal{T}, C)$ is an $\mathcal{EL}^{st}$ -concept;
- if C, D are $\mathcal{EL}^{st}$ -concepts, then $C \sqsubseteq D$ is a $\mathcal{EL}^{st}$ -concept inclusion;
- a general $\mathcal{EL}^{st}$ -TBox is a finite set of $\mathcal{EL}^{st}$ -concept inclusions.

The semantics of $\exists^{sim}\Sigma.(\mathcal{T}, C)$ is as follows:

$$d \in (\exists^{sim}\Sigma.(\mathcal{T}, C))^\mathcal{I} \text{ iff there exists } (\mathcal{J}, e) \text{ such that } \mathcal{J} \text{ is a model of } \mathcal{T}, e \in C^\mathcal{J} \text{ and } (\mathcal{J}, e) \leq_\Gamma (\mathcal{I}, d), \text{ where } \Gamma = (\mathbf{N}_C \cup \mathbf{N}_R) \setminus \Sigma.$$

Example 8 Let $\mathcal{T} = \{A \sqsubseteq \exists \text{has.parent}.A\}$ and $\Sigma = \{A\}$. Then $\exists^{sim}\Sigma.(\mathcal{T}, A)$ is equivalent to the concept $\exists^{sim}(\mathcal{I}, d)$ defined in Example 6.

We will later exploit the fact that $\exists^{sim}\Sigma.(\mathcal{T}, C)$ is equivalent to $\exists^{sim}\Sigma \cup \{A\}.(\mathcal{T}', A)$, where A is a fresh concept name and $\mathcal{T}' = \mathcal{T} \cup \{A \sqsubseteq C\}$. Another interesting (but subsequently unexploited) observation is that we can w.l.o.g. restrict Σ to singleton sets since

$$\begin{aligned} \exists^{sim}(\{\sigma\} \cup \Sigma).(\mathcal{T}, C) &\equiv \exists^{sim}\{\sigma\}.(\emptyset, \exists^{sim}\Sigma.(\mathcal{T}, C)) \\ \exists^{sim}\emptyset.(\mathcal{T}, C) &\equiv \exists^{sim}\{B\}.(\mathcal{T}, C) \end{aligned}$$

where B is a concept name that does not occur in $\mathcal{T}$ and C .

Simulating models of KBs. The third language $\mathcal{EL}^{sa}$ extends $\mathcal{EL}$ by the concept constructor $\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a)$, where a is an individual name in the ABox $\mathcal{A}$, $\mathcal{T}$ is a TBox, and Σ a finite signature. More precisely, we define $\mathcal{EL}^{sa}$ -concepts, concept inclusions, general TBoxes, and KBs, by simultaneous induction as follows:

- every $\mathcal{EL}$ -concept, concept inclusion, general TBox, and KB is an $\mathcal{EL}^{sa}$ -concept, concept inclusion, general TBox, and KB, respectively;
- if $(\mathcal{T}, \mathcal{A})$ is a general $\mathcal{EL}^{sa}$ -KB, a an individual name in $\mathcal{A}$, and Σ a finite signature, then $\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a)$ is an $\mathcal{EL}^{sa}$ -concept;
- if C, D are $\mathcal{EL}^{sa}$ -concepts, then $C \sqsubseteq D$ is an $\mathcal{EL}^{sa}$ -concept inclusion;
- a general $\mathcal{EL}^{sa}$ -TBox is a finite set of $\mathcal{EL}^{sa}$ -concept inclusions;
- an $\mathcal{EL}^{sa}$ -KB is a pair $(\mathcal{T}, \mathcal{A})$ consisting of a general $\mathcal{EL}^{sa}$ -TBox and an ABox.

The semantics of $\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a)$ is as follows:

$$d \in (\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a))^\mathcal{I} \text{ iff there exists } d \in (\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a))^\mathcal{I} \text{ iff there exists a model } \mathcal{J} \text{ of } (\mathcal{T}, \mathcal{A}) \text{ such that } (\mathcal{J}, a^\mathcal{J}) \leq_\Gamma (\mathcal{I}, d), \text{ where } \Gamma = (\mathbf{N}_C \cup \mathbf{N}_R) \setminus \Sigma.$$

Example 9 Let $\mathcal{T} = \emptyset$, $\mathcal{A} = \{\text{has.parent}(a, a)\}$, and $\Sigma = \emptyset$. Then $\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a)$ is equivalent to the concept $\exists^{sim}(\mathcal{I}, d)$ defined in Example 6.

Let $\mathcal{L}_1, \mathcal{L}_2$ be sets of concepts. We say that $\mathcal{L}_2$ is *polynomially at least as expressive as* $\mathcal{L}_1$, in symbols $\mathcal{L}_1 \leq_p \mathcal{L}_2$, if for every $C_1 \in \mathcal{L}_1$ one can construct in polynomial time a $C_2 \in \mathcal{L}_2$ such that C_1 and C_2 are equivalent. We say that $\mathcal{L}_1, \mathcal{L}_2$ are *polynomially equivalent*, in symbols $\mathcal{L}_1 \equiv_p \mathcal{L}_2$, if $\mathcal{L}_1 \leq_p \mathcal{L}_2$ and $\mathcal{L}_2 \leq_p \mathcal{L}_1$.

Theorem 10 *The languages $\mathcal{EL}^{\nu+}, \mathcal{EL}^{si}, \mathcal{EL}^{st}$, and $\mathcal{EL}^{sa}$ are polynomially equivalent.*

We provide sketches of proofs of $\mathcal{EL}^{si} \leq_p \mathcal{EL}^{\nu+}, \mathcal{EL}^{\nu+} \leq_p \mathcal{EL}^{st}, \mathcal{EL}^{st} \leq_p \mathcal{EL}^{sa}$, and $\mathcal{EL}^{sa} \leq_p \mathcal{EL}^{si}$.

$\mathcal{EL}^{si} \leq_p \mathcal{EL}^{\nu+}$. By Lemma 7, considering $\mathcal{EL}^{si}$ -concepts of the form $\exists^{sim}(\mathcal{T}, d)$ is sufficient. Each such concept is equivalent to the $\mathcal{EL}^{\nu+}$ -concept $\nu_\ell d_1 \dots d_n. C_1, \dots, C_n$, where the domain $\Delta^\mathcal{I} = \{d_1, \dots, d_n\}$ is regarded as a set of concept variables, $d = d_\ell$, and

$$C_i = \bigcap \{A \mid d_i \in A^\mathcal{I}\} \cap \bigcap \{\exists r. d_j \mid (d_i, d_j) \in r^\mathcal{I}\}.$$

$\mathcal{EL}^{\nu+} \leq_p \mathcal{EL}^{st}$. Let C be a closed $\mathcal{EL}^{\nu+}$ -concept. An equivalent $\mathcal{EL}^{st}$ -concept is constructed by replacing each subconcept of C of the form $\nu_\ell X_1, \dots, X_n. C_1, \dots, C_n$ with an $\mathcal{EL}^{st}$ -concept, proceeding from the inside out. We assume that for every variable X that occurs in the original $\mathcal{EL}^{\nu+}$ -concept C , there is a concept name A_X that does not occur in C . Now $\nu_\ell X_1, \dots, X_n. C_1, \dots, C_n$ (which potentially contains free variables) is replaced with the $\mathcal{EL}^{st}$ -concept

$$\exists^{sim}\{A_{X_1}, \dots, A_{X_n}\}. (\{A_{X_i} \sqsubseteq C_i^\downarrow \mid 1 \leq i \leq n\}, A_{X_\ell})$$

where $C_i^\downarrow$ is obtained from C_i by replacing every variable X with the concept name A_X .

$\mathcal{EL}^{st} \leq_p \mathcal{EL}^{sa}$. Let C be an $\mathcal{EL}^{st}$ -concept. As already observed, we may assume that D is a concept name in all subconcepts $\exists^{sim}\Sigma.(\mathcal{T}, D)$ of C . Now replace each $\exists^{sim}\Sigma.(\mathcal{T}, A)$ in C , proceeding from the inside out, by $\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a)$, where $\mathcal{A} = \{A(a)\}$. The resulting concept is equivalent to C .

$\mathcal{EL}^{sa} \leq_p \mathcal{EL}^{si}$. To prove this inclusion, we make use of *canonical models* for $\mathcal{EL}^{sa}$ -KBs, and extension of the canonical models used for $\mathcal{EL}$ in [4]. In particular, canonical models for $\mathcal{EL}^{sa}$ can be constructed by an extension of the algorithm given in [4], see [17] for details.

Theorem 11 (Canonical model) *For every consistent $\mathcal{EL}^{sa}$ -KB $(\mathcal{T}, \mathcal{A})$, one can construct in polynomial time a model $\mathcal{I}_{\mathcal{T}, \mathcal{A}}$ of $(\mathcal{T}, \mathcal{A})$ with $|\Delta^{\mathcal{I}_{\mathcal{T}, \mathcal{A}}}|$ bounded by twice the size of $(\mathcal{T}, \mathcal{A})$ and such that for every model $\mathcal{J}$ of $(\mathcal{T}, \mathcal{A})$, we have $(\mathcal{I}_{\mathcal{T}, \mathcal{A}}, a^{\mathcal{I}_{\mathcal{T}, \mathcal{A}}}) \leq (\mathcal{J}, a^{\mathcal{J}})$ for all $a \in \text{Ind}(\mathcal{A})$.*

To prove $\mathcal{EL}^{sa} \leq_p \mathcal{EL}^{si}$, it suffices to show that any outermost occurrence of a concept of the form $\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a)$ in an $\mathcal{EL}^{sa}$ -concept C can be replaced with the equivalent $\mathcal{EL}^{si}$ -concept $\exists^{sim}(\mathcal{I}_{\mathcal{T}, \mathcal{A}}^\Sigma, a)$, where $\mathcal{I}_{\mathcal{T}, \mathcal{A}}^\Sigma$ denotes $\mathcal{I}_{\mathcal{T}, \mathcal{A}}$ except that all $\sigma \in \Sigma$ are interpreted as empty sets. First let $d \in (\exists^{sim}\Sigma.(\mathcal{T}, \mathcal{A}, a))^{\mathcal{J}}$. Then there is a model $\mathcal{I}'$ of $(\mathcal{T}, \mathcal{A})$ such that $(\mathcal{I}', a^{\mathcal{I}'}) \leq_\Sigma (\mathcal{J}, d)$. By Theorem 11, $(\mathcal{I}_{\mathcal{T}, \mathcal{A}}, a^{\mathcal{I}_{\mathcal{T}, \mathcal{A}}}) \leq (\mathcal{I}', a^{\mathcal{I}'})$. Thus, by closure of simulations under composition, $(\mathcal{I}_{\mathcal{T}, \mathcal{A}}^\Sigma, a) \leq_\Sigma (\mathcal{J}, d)$ as required. The converse direction follows from the condition that $\mathcal{I}_{\mathcal{T}, \mathcal{A}}$ is a model of $(\mathcal{T}, \mathcal{A})$. This finishes our proof sketch for Theorem 10.

It is interesting to note that, as a consequence of the proofs of Theorem 10, for every $\mathcal{EL}^{\nu+}$ -concept there is an equivalent $\mathcal{EL}^{\nu+}$ -concept of polynomial size in which the greatest fixpoint constructor is not nested, and similarly for $\mathcal{EL}^{st}, \mathcal{EL}^{sa}$. An important consequence of the existence of canonical models, as granted by Theorem 11, is that reasoning in our family of extensions of $\mathcal{EL}$ is

tractable. Recall that *KB consistency* is the problem of deciding whether a given KB has a model; *subsumption w.r.t. general TBoxes* is the problem of deciding whether a subsumption $C \sqsubseteq D$ follows from a general TBox $\mathcal{T}$ (in symbols, $\mathcal{T} \models C \sqsubseteq D$); and the *instance problem* is the problem of deciding whether an assertion $C(a)$ follows from a KB $(\mathcal{T}, \mathcal{A})$ (in symbols, $(\mathcal{T}, \mathcal{A}) \models C(a)$).

Theorem 12 (Tractable reasoning) *Let $\mathcal{L}$ be any of the languages $\mathcal{EL}^\nu, \mathcal{EL}^{\nu+}, \mathcal{EL}^{si}, \mathcal{EL}^{st}$, or $\mathcal{EL}^{sa}$. Then KB consistency, subsumption w.r.t. TBoxes, and the instance problem can be decided in PTIME.*

Proof (sketch) By Theorem 10, it suffices to concentrate on $\mathcal{L} = \mathcal{EL}^{sa}$. The PTIME decidability of KB consistency is proved in [17] as part of the algorithm that constructs the canonical model. Subsumption w.r.t. general TBoxes can be polynomially reduced in the standard way to the instance problem. Finally, by Theorem 11, we can decide the instance problem as follows: to decide whether $(\mathcal{T}, \mathcal{A}) \models C(a)$, where we can w.l.o.g. assume that $C = A$ for a concept name A , we check whether $(\mathcal{T}, \mathcal{A})$ is inconsistent or $a^{\mathcal{I}_{\mathcal{T}, \mathcal{A}}} \in A^{\mathcal{I}_{\mathcal{T}, \mathcal{A}}}$. Both can be done in PTIME.

Besides of the canonical model of a KB from Theorem 11, we also require the canonical model $\mathcal{I}_{\mathcal{T}, C}$ of a general $\mathcal{EL}^{\nu+}$ -TBox $\mathcal{T}$ and concept C which is defined by taking the reduct not interpreting A of the canonical model $\mathcal{I}_{\mathcal{T}', \mathcal{A}}$ for $\mathcal{T}' = \mathcal{T} \cup \{A \sqsubseteq C\}$ and $\mathcal{A} = \{A(a)\}$ (A a fresh concept). We set $d_C = a^{\mathcal{I}_{\mathcal{T}, C}}$. $\mathcal{I}_{\mathcal{T}, C}$ is a model of $\mathcal{T}$ with $d_C \in C^{\mathcal{I}_{\mathcal{T}, C}}$ such that $(\mathcal{I}_{\mathcal{T}, C}, d_C) \leq (\mathcal{J}, e)$ for all models $\mathcal{J}$ of $\mathcal{T}$ with $e \in C^{\mathcal{J}}$.

5 CHARACTERIZING $\mathcal{EL}^\nu$ USING SIMULATIONS

When characterizing $\mathcal{EL}$ as a fragment of first-order logic in Theorem 5, our starting point was the observation that $\mathcal{EL}$ -concepts are preserved under simulations and that $\mathcal{EL}$ is a Horn logic, thus having finite minimal models. The same is true for $\mathcal{EL}^\nu$: first, $\mathcal{EL}^\nu$ -concepts are preserved under simulations, as $\mathcal{EL}^{si}$ is obviously preserved under simulations and, by Theorem 10, every $\mathcal{EL}^\nu$ -concept is equivalent to an $\mathcal{EL}^{si}$ -concept. And second, a finite minimal model of an $\mathcal{EL}^\nu$ -concept C is given by the canonical model $(\mathcal{I}_{\mathcal{T}, C}, d_C)$ defined above for $\mathcal{T} = \emptyset$. However, $\mathcal{EL}^\nu$ is clearly not a fragment of FO. Instead, it relates to MSO in exactly the way that $\mathcal{EL}$ related to FO.

Theorem 13 *The set of $\mathcal{EL}^\nu$ -concepts is a maximal set of MSO-formulas that is preserved under simulations and has finite minimal models: if $\mathcal{L}$ is a set of MSO-formulas that properly contains all $\mathcal{EL}^\nu$ -concepts, then either it contains a formula not preserved under simulations or it does not have finite minimal models.*

Proof Assume that $\mathcal{L} \supsetneq \mathcal{EL}^\nu$ is preserved under simulations and has finite minimal models. Let $\varphi(x) \in \mathcal{L}$. We have to show that $\varphi(x)$ is equivalent to an $\mathcal{EL}^\nu$ -concept. To this end, take a finite minimal model of φ , i.e., an interpretation $\mathcal{I}$ and a $d \in \Delta^\mathcal{I}$ such that for all $\psi(x) \in \mathcal{L}$ we have that $\forall x. (\varphi(x) \rightarrow \psi(x))$ is valid iff $\mathcal{I} \models \psi[d]$. We will show that φ is equivalent to (the MSO translation of) $\exists^{sim}(\mathcal{I}, d)$. We may assume that $\exists^{sim}(\mathcal{I}, d) \in \mathcal{L}$. Since $d \in (\exists^{sim}(\mathcal{I}, d))^\mathcal{I}$, we thus have that $\forall x. (\varphi(x) \rightarrow \exists^{sim}(\mathcal{I}, d)(x))$ is valid. Conversely, assume that $d' \in (\exists^{sim}(\mathcal{I}, d))^\mathcal{J}$ for some interpretation $\mathcal{J}$. Then $(\mathcal{I}, d) \leq (\mathcal{J}, d')$. We have $(\mathcal{I}, d) \models \varphi[d]$. Thus, by preservation of $\varphi(x)$ under simulations, $\mathcal{J} \models \varphi[d']$. Thus $\forall x. (\exists^{sim}(\mathcal{I}, d)(x) \rightarrow \varphi(x))$ is also valid. This finishes the proof.

A number of closely related characterizations remain open. For example, we conjecture that an extension of Theorem 4 holds for $\mathcal{EL}^{\nu, \sqcup}$ and MSO (instead of $\mathcal{EL}$ and FO). Also, it is open whether Theorem 13 still holds if finite minimal models are replaced by arbitrary minimal models.

6 APPLICATIONS

The μ -calculus is considered to be extremely well-behaved regarding its expressive power and logical properties. The aim of this section is to take a brief look at the expressive power of its $\mathcal{EL}$ -analogues $\mathcal{EL}^{\nu}$ and $\mathcal{EL}^{\nu+}$. In particular, we show that $\mathcal{EL}^{\nu+}$ is more well-behaved than $\mathcal{EL}$ in a number of respects. Throughout this section, we will not distinguish between the languages previously proved polynomially equivalent.

To begin with, we construct the *least common subsumer* (LCS) of two concepts w.r.t. a general $\mathcal{EL}^{\nu+}$ -TBox (the generalization to more than two concepts is straightforward). Given a general $\mathcal{EL}^{\nu+}$ -TBox $\mathcal{T}$ and concepts C_1, C_2 , a concept C is called a *LCS* of C_1, C_2 w.r.t. $\mathcal{T}$ in $\mathcal{EL}^{\nu+}$ if $\mathcal{T} \models C_i \sqsubseteq C$ for $i = 1, 2$; if $\mathcal{T} \models C_i \sqsubseteq D$ for $i = 1, 2$ and D a $\mathcal{EL}^{\nu+}$ -concept, then $\mathcal{T} \models C \sqsubseteq D$. It is known [2] that in $\mathcal{EL}$ the LCS does not always exist.

Example 14 In $\mathcal{EL}$, the LCS of A, B w.r.t.

$$\mathcal{T} = \{A \sqsubseteq \exists \text{has_parent}.A, B \sqsubseteq \exists \text{has_parent}.B\}$$

does not exist. In $\mathcal{EL}^{\nu}$, however, the LCS of A, B w.r.t. $\mathcal{T}$ is given by $\nu X. \exists \text{has_parent}.X$ (see Example 1).

To construct the LCS in $\mathcal{EL}^{\nu+}$, we adopt the product construction used in [2] for the case of classical TBoxes with a fixpoint semantics. For interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$, the *product* $\mathcal{I}_1 \times \mathcal{I}_2$ is defined by setting $\Delta^{\mathcal{I}_1 \times \mathcal{I}_2} = \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$, $(d_1, d_2) \in A^{\mathcal{I}_1 \times \mathcal{I}_2}$ iff $d_i \in A^{\mathcal{I}_i}$ for $i = 1, 2$, and $((d_1, d_2), (d'_1, d'_2)) \in r^{\mathcal{I}_1 \times \mathcal{I}_2}$ iff $(d_i, d'_i) \in r^{\mathcal{I}_i}$ for $i = 1, 2$.

Theorem 15 (LCS) Let $\mathcal{T}$ be a general $\mathcal{EL}^{\nu+}$ -TBox and C_1 and C_2 be $\mathcal{EL}^{\nu+}$ -concepts. Then $\exists^{sim}(\mathcal{I}_{\mathcal{T}, C_1} \times \mathcal{I}_{\mathcal{T}, C_2}, (d_{C_1}, d_{C_2}))$ is the LCS of C_1, C_2 w.r.t. $\mathcal{T}$ in $\mathcal{EL}^{\nu+}$.

The same product construction has been used in [2] for the case of classical TBoxes with a fixpoint semantics, which, however, additionally require a notion of conservative extension (see Section 7).

Our second result concerns the most specific concept, which plays an important role in the bottom-up construction of knowledge bases and has received considerable attention in the context of $\mathcal{EL}$ [2, 6]. Formally, a concept C is the *most specific concept* (MSC) for an individual a in a knowledge base $(\mathcal{T}, \mathcal{A})$ in $\mathcal{EL}^{\nu+}$ if $(\mathcal{T}, \mathcal{A}) \models C(a)$ and for every $\mathcal{EL}^{\nu+}$ -concept D with $(\mathcal{T}, \mathcal{A}) \models D(a)$, we have $\mathcal{T} \models C \sqsubseteq D$. In $\mathcal{EL}$, the MSC need not exist, as is witnessed by the knowledge base $(\emptyset, \{\text{has_parent}(a, a)\})$, where the MSC for a is non-existent.

Theorem 16 (MSC) In $\mathcal{EL}^{\nu+}$, the MSC always exists for any a in any KB $(\mathcal{T}, \mathcal{A})$ and is given as $\exists^{sim}\emptyset.(\mathcal{T}, \mathcal{A}, a)$.

In [2], the MSC in $\mathcal{EL}$ -KBs based on classical TBoxes with a fixpoint semantics is defined. The relationship between $\mathcal{EL}^{\nu+}$ and fixpoint TBoxes is discussed in more detail in Section 7.

We now turn our attention to issues of definability and interpolation. From now on, we use $\text{sig}(C)$ to denote the set of concept and

role names used in the concept C . A concept C is a Σ -concept if $\text{sig}(C) \subseteq \Sigma$. Let $\mathcal{T}$ be a general $\mathcal{EL}^{\nu+}$ -TBox, C an $\mathcal{EL}^{\nu+}$ -concept and Γ a finite signature.

We start with considering the fundamental notion of a Γ -definition. The question addressed here is whether a given concept can be expressed in an equivalent way by referring only to the symbols in a given signature Γ [22, 21]. Formally, a Γ -concept D is an *explicit Γ -definition* of a concept C w.r.t. a TBox $\mathcal{T}$ if, and only if, $\mathcal{T} \models C \equiv D$ (i.e., C and D are equivalent w.r.t. $\mathcal{T}$). Clearly, explicit Γ -definitions do not always exist in any of the logics studied in this paper: for example, there is no explicit $\{A\}$ -definition of B w.r.t. the TBox $\{A \sqsubseteq B\}$. However, it is not hard to show the following using the fact that $\exists^{sim}\Sigma.(\mathcal{T}, C)$ is the most specific Γ -concept that subsumes C w.r.t. $\mathcal{T}$.

Proposition 17 Let C be an $\mathcal{EL}^{\nu+}$ -concept, $\mathcal{T}$ a general $\mathcal{EL}^{\nu+}$ -TBox and Γ a signature. There exists an explicit Γ -definition of C w.r.t. $\mathcal{T}$ iff $\exists^{sim}\Sigma.(\mathcal{T}, C)$ is such a definition ($\Sigma = \text{sig}(\mathcal{T}, C) \setminus \Gamma$).

It is interesting to note that if $\mathcal{T}$ happens to be a general $\mathcal{EL}$ -TBox and C an $\mathcal{EL}$ -concept and there exists an explicit Γ -definition of C w.r.t. $\mathcal{T}$, then the concept $\exists^{sim}\Sigma.(\mathcal{T}, C)$ from Proposition 17 is equivalent w.r.t. $\mathcal{T}$ to an $\mathcal{EL}$ -concept over Γ . This follows from the fact that $\mathcal{EL}$ has the Beth definability property (see below for a definition) which follows immediately from interpolation results proved for $\mathcal{EL}$ in [15]. The advantage of giving explicit Γ -definitions in $\mathcal{EL}^{\nu+}$ even when $\mathcal{T}$ and C are formulated in $\mathcal{EL}$ is that Γ -definitions in $\mathcal{EL}^{\nu+}$ are of polynomial size while the following example shows that they may be exponentially large in $\mathcal{EL}$.

Example 18 Let $\mathcal{T}$ consist of $A_i \equiv \exists r_i.A_{i+1} \sqcap \exists s_i.A_{i+1}$ for $0 \leq i < n$, and $A_n \equiv \top$. Let $\Gamma = \{r_0, \dots, r_{n-1}, s_0, \dots, s_{n-1}\}$. Then A_0 has an explicit Γ -definition w.r.t. $\mathcal{T}$ in $\mathcal{EL}$, namely C_0 , where $C_i \equiv \exists r_i.C_{i+1} \sqcap \exists s_i.C_{i+1}$ and $C_n \equiv \top$. This definition is of exponential size and it is easy to see that there is no shorter Γ -definition of A_0 w.r.t. $\mathcal{T}$ in $\mathcal{EL}$.

Say that a concept C is *implicitly Γ -defined* w.r.t. $\mathcal{T}$ iff $\mathcal{T} \cup \mathcal{T}_{\Gamma} \models C \equiv C_{\Gamma}$, where $\mathcal{T}_{\Gamma}$ and C_{Γ} are obtained from $\mathcal{T}$ and C , respectively, by replacing each $\sigma \notin \Gamma$ by a fresh symbol σ' . The Beth definability property, which was studied in a DL context in [22, 21], ensures that concepts that are implicitly Γ -defined have an explicit Γ -definition.

Theorem 19 (Beth Property) $\mathcal{EL}^{\nu+}$ has the polynomial Beth definability property: for every general $\mathcal{EL}^{\nu+}$ -TBox $\mathcal{T}$, concept C , and signature Γ such that C is implicitly Γ -defined w.r.t. $\mathcal{T}$, there is an explicit Γ -definition w.r.t. $\mathcal{T}$, namely $\exists^{sim}(\text{sig}(\mathcal{T}, C) \setminus \Gamma).(\mathcal{T}, C)$.

The proof of Theorem 19 relies on $\mathcal{EL}^{\nu}$ having a certain interpolation property. Say that two general TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ are Δ -inseparable w.r.t. $\mathcal{EL}^{\nu}$ if $\mathcal{T}_1 \models C \sqsubseteq D$ iff $\mathcal{T}_2 \models C \sqsubseteq D$ for all $\mathcal{EL}^{\nu}$ -inclusions $C \sqsubseteq D$.

Theorem 20 (Interpolation) Let $\mathcal{T}_1 \cup \mathcal{T}_2 \models C \sqsubseteq D$ and assume that $\mathcal{T}_1$ and $\mathcal{T}_2$ are Δ -inseparable w.r.t. $\mathcal{EL}^{\nu}$ for $\Delta = \text{sig}(\mathcal{T}_1, C) \cap \text{sig}(\mathcal{T}_2, D)$. Then the Δ -concept $F = \exists^{sim}\Delta.(\mathcal{T}_1, C)$, $\Sigma = \text{sig}(\mathcal{T}_1, C) \setminus \Delta$, is an interpolant of C, D w.r.t. $\mathcal{T}_1, \mathcal{T}_2$; i.e. $\mathcal{T}_1 \models C \sqsubseteq F$ and $\mathcal{T}_2 \models F \sqsubseteq D$.

We show how Theorem 19 follows from Theorem 20. Assume that $\mathcal{T} \cup \mathcal{T}_{\Gamma} \models C \equiv C_{\Gamma}$, where $\mathcal{T}, \mathcal{T}_{\Gamma}, C, C_{\Gamma}$ satisfy the conditions of Theorem 19. Then $\mathcal{T}$ and $\mathcal{T}_{\Gamma}$ are Γ -inseparable and $\Gamma \supseteq \text{sig}(\mathcal{T}, C) \cap$

$\text{sig}(\mathcal{T}_\Gamma, C_\Gamma)$. Thus, by Theorem 20, $\mathcal{T} \models \exists^{sim} \Sigma.(\mathcal{T}_\Gamma, C_\Gamma) \subseteq C$ for $\Sigma = \text{sig}(\mathcal{T}_\Gamma, C_\Gamma) \setminus \Gamma$. Now Theorem 19 follows from the fact that $\exists^{sim} \Sigma.(\mathcal{T}_\Gamma, C_\Gamma)$ is equivalent to $\exists^{sim} \Sigma'.(\mathcal{T}, C)$ for $\Sigma' = \text{sig}(\mathcal{T}, C) \setminus \Gamma$.

In [15], it is shown that $\mathcal{EL}$ also has this interpolation property. However, the advantage of using $\mathcal{EL}^{\nu+}$ is that interpolants are of polynomial size. The decomposition algorithm for $\mathcal{EL}$ given in [15] crucially depends on this property of $\mathcal{EL}^{\nu+}$.

7 RELATION TO TBOXES WITH FIXPOINT SEMANTICS

There is a tradition of considering DLs that introduce fixpoints at the TBox level instead of at the concept level [19, 20, 1]. In [3], Baader proposes and analyzes such a DL based on $\mathcal{EL}$ and greatest fixpoints. This DL, which we call $\mathcal{EL}^{\text{gfp}}$ here, differs from $\mathcal{EL}^\nu$ in that (i) TBoxes are classical TBoxes rather than sets of GCIs $C \sqsubseteq D$, i.e., sets of expressions $A \equiv C$ with $A \in \mathbf{N}_C$ and C a concept (cycles are allowed) and (ii) the ν -concept constructor is not present; instead, a greatest fixpoint semantics is adopted for TBoxes.

On the concept level, $\mathcal{EL}^\nu$ is clearly strictly more expressive than $\mathcal{EL}^{\text{gfp}}$: since fixpoints are introduced at the TBox level, concepts of $\mathcal{EL}^{\text{gfp}}$ coincide with $\mathcal{EL}$ -concepts, and thus there is no $\mathcal{EL}^{\text{gfp}}$ -concept equivalent to the $\mathcal{EL}^\nu$ -concept $\nu X. \exists r. X$. In the following, we show that $\mathcal{EL}^\nu$ is also more expressive than $\mathcal{EL}^{\text{gfp}}$ on the TBox level, even if we restrict $\mathcal{EL}^\nu$ -TBoxes to (possibly cyclic) concept definitions, as in $\mathcal{EL}^{\text{gfp}}$. We use the standard notion of logical equivalence, i.e., two TBoxes $\mathcal{T}$ and $\mathcal{T}'$ are *equivalent* iff $\mathcal{T}$ and $\mathcal{T}'$ have precisely the same models. As observed by Schild in the context of $\mathcal{ALC}$ [20], every $\mathcal{EL}^{\text{gfp}}$ -TBox $\mathcal{T} = \{A_1 \equiv C_1, \dots, A_n \equiv C_n\}$ is equivalent in this sense to the $\mathcal{EL}^{\nu+}$ -TBox $\{A_i \equiv \nu_i X_1, \dots, X_n. C'_1, \dots, C'_n \mid 1 \leq i \leq n\}$, where each C'_i is obtained from C_i by replacing each A_j with X_j , $1 \leq j \leq n$. Note that since we are translating to mutual fixpoints, the size of the resulting TBox is polynomial in the size of the original one. In the converse direction, there is no equivalence-preserving translation.

Lemma 21 *For each $\mathcal{EL}^{\text{gfp}}$ -TBox, there is an equivalent $\mathcal{EL}^{\nu+}$ -TBox of polynomial size, but no $\mathcal{EL}^{\text{gfp}}$ -TBox is equivalent to the $\mathcal{EL}^\nu$ -TBox $\mathcal{T}_0 = \{A \equiv P \sqcap \nu X. \exists r. X\}$.*

Proof It is not difficult to show that for every $\mathcal{EL}^{\text{gfp}}$ -TBox $\mathcal{T}$, defined concept name A in $\mathcal{T}$, and role name r , at least one of the following holds:

- there is an $m \geq 0$ such that $\mathcal{T} \models A \sqsubseteq \exists r^m. \top$ implies $n \leq m$ or
- $\mathcal{T} \models A \sqsubseteq \exists r^n. B$ for some $n > 0$ and defined concept name B .

Since neither of these is true for $\mathcal{T}_0$, $\mathcal{T}$ is not equivalent to $\mathcal{T}_0$.

Restricted to classical TBoxes, $\mathcal{EL}^{\text{gfp}}$ and $\mathcal{EL}^\nu$ become equi-expressive if the strict notion of equivalence used above is replaced with one based on conservative extensions, thus allowing the introduction of new concept names that are suppressed from logical equivalence.

8 Conclusion

We have introduced and investigated the extensions $\mathcal{EL}^\nu$ and $\mathcal{EL}^{\nu+}$ of $\mathcal{EL}$ with greatest fixpoint operators. The main result of this paper is that $\mathcal{EL}^{\nu+}$ can be regarded as a completion of $\mathcal{EL}$ regarding its expressive power in which reasoning is still tractable, but where many

previously non-existent concepts (such as the LCS and MCS) exist and/or can be expressed more succinctly (such as interpolants and explicit concept definitions). Interestingly, the alternative extension of $\mathcal{EL}$ by smallest rather than greatest fixpoints is much less well-behaved. For example, even the addition of transitive closure to $\mathcal{EL}$ leads to non-tractable reasoning problems [13].

REFERENCES

- [1] F. Baader. Using automata theory for characterizing the semantics of terminological cycles. *Annals of Mathematics and Artificial Intelligence*, 18(2–4):175–219, 1996.
- [2] F. Baader. Least common subsumers and most specific concepts in description logics with existential restrictions and terminological cycles. In *Proc. of IJCAI03*, pages 319–324. Morgan Kaufmann, 2003.
- [3] F. Baader. Terminological cycles in a description logic with existential restrictions. In *Proc. of IJCAI03*, pages 325–330. Morgan Kaufmann, 2003.
- [4] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI05*. Morgan Kaufmann, 2005.
- [5] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, and P. Patel-Schneider. *The Description Logic Handbook: Theory, implementation and applications*. Cambridge Univ. Press, 2003.
- [6] F. Baader and F. Distel. A finite basis for the set of $\mathcal{EL}$ -implications holding in a finite model. In *Proc. of ICFCA8*, volume 4933 of *LNAI*, pages 46–61. Springer, 2008.
- [7] J. Bradfield and C. Stirling. Modal mu-calculus. In *Handbook of Modal Logic*. Elsevier, 2006.
- [8] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati. Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. *J. of Automated Reasoning*, 39(3):385–429, 2007.
- [9] J. W. de Bakker. *Mathematical Theory of Program Correctness*. Prentice-Hall, 1980.
- [10] T. French. *Bisimulation quantifiers for modal logics*. PhD thesis, University of Western Australia, 2006.
- [11] G. De Giacomo and M. Lenzerini. Boosting the Correspondence between Description Logics and Propositional Dynamic Logics. In *Proc. of AAAI94*, pages 205–212. AAAI Press, 1994.
- [12] V. Goranko and M. Otto. Model theory of modal logic. In *Handbook of Modal Logic*. Elsevier, 2007.
- [13] C. Haase and C. Lutz. Complexity of Subsumption in the $\mathcal{EL}$ Family of Description Logics: Acyclic and Cyclic TBoxes. In *Proc. of ECAI08*, pages 25–29, 2008.
- [14] D. Janin and I. Walukiewicz. On the expressive completeness of the propositional mu-calculus with respect to monadic second order logic. In *Proc. of CONCUR96*, volume 1119 of *PNCS*, pages 263–277. Springer, 1996.
- [15] B. Konev, C. Lutz, D. Ponomaryov, and F. Wolter. Decomposing description logic ontologies. In *Proc. of KR10*, 2010.
- [16] N. Kurtonina and M. de Rijke. Expressiveness of concept expressions in first-order description logics. *Artificial Intelligence*, 107, 1999.
- [17] C. Lutz, R. Piro and F. Wolter. Enriching EL-concepts with Greatest Fixpoints. *Technical Report, University of Liverpool*, 2010.
- [18] C. Lutz and F. Wolter. Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. *J. of Symbolic Computation*, 45(2):194–228, 2010.
- [19] B. Nebel. Terminological cycles: Semantics and computational properties. In *Principles of Semantic Networks*. Morgan Kaufmann, 1991.
- [20] K. Schild. Terminological cycles and the propositional μ -calculus. In *Proc. of KR94*, pages 509–520. Morgan Kaufmann, 1994.
- [21] I. Seylan, E. Franconi, and J. de Bruijn. Effective query rewriting with ontologies over DBBoxes. In *Proc. of IJCAI09*, pages 923–925, 2009.
- [22] B. ten Cate, W. Conradie, M. Marx, and Y. Venema. Definitorially complete description logics. In *Proc. of KR06*, pages 79–89. AAAI Press, 2006.
- [23] J. van Benthem. *Modal Correspondence Theory*. Mathematical Institute, University of Amsterdam, 1976.
- [24] M. Vardi and P. Wolper. Automata theoretic techniques for modal logics of programs. In *Proc. of STOC84*, pages 446–456. ACM Press, 1984.
- [25] M. Y. Vardi. Why is modal logic so robustly decidable? In *Descriptive Complexity and Finite Models*, pages 149–184, 1996.