Least General Generalizations in Description Logic: Verification and Existence

Jean Christoph Jung,[1] **Carsten Lutz,**[1] **Frank Wolter**[2]
[1] {jeanjung,clu}@uni-bremen.de, University of Bremen, Germany
[2] wolter@liverpool.ac.uk, University of Liverpool, United Kingdom

Abstract

We study two forms of least general generalizations in description logic, the least common subsumer (LCS) and most specific concept (MSC). While the LCS generalizes from examples that take the form of concepts, the MSC generalizes from individuals in data. Our focus is on the complexity of existence and verification, the latter meaning to decide whether a candidate concept is the LCS or MSC. We consider cases with and without a background TBox and a target signature. Our results range from coNP-complete for LCS and MSC verification in the description logic $\mathcal{EL}$ without TBoxes to undecidability of LCS and MSC verification and existence in $\mathcal{ELI}$ with TBoxes. To obtain results in the presence of a TBox, we establish a close link between the problems studied in this paper and concept learning from positive and negative examples. We also give a way to regain decidability in $\mathcal{ELI}$ with TBoxes and study single example MSC as a special case.

1 Introduction

Generalization is a fundamental method in relational learning and inductive logic programming (Plotkin 1970; Muggleton 1991). Given a finite number of positive examples, one seeks a description in a logical language that encompasses all examples and in this sense provides a generalization. To ensure that the description is as informative as possible, one aims at obtaining least general generalizations, that is, generalizations that cannot be made more specific without losing at least one example. Note that computing least general generalizations is a form of supervised learning in which only positive, but no negative examples are given.

In this paper, we study least general generalizations in the context of description logics (DLs), a widely known family of ontology languages that underpin the web ontology language OWL 2 (Baader et al. 2017). In DLs, *concepts* are the building blocks of an ontology and thus a prime target for being learned through generalization. There are in fact several applications in which this is useful, including ontology design by domain experts that are not sufficiently proficient in logical modeling (Baader and Küsters 1998; Baader, Küsters, and Molitor 1999; Baader, Sertkaya, and Turhan 2007; Donini et al. 2009), supporting the improvement and restructuring of an ontology (Cohen, Borgida, and Hirsh 1992; Küsters and Borgida 2001), and creative discovery of novel concepts through conceptual blending (Fauconnier and Turner 2008; Eppe et al. 2018). We focus on the two fundamental DLs $\mathcal{EL}$ and $\mathcal{ELI}$, fragments of first-order Horn logic that can express positive conjunctive existential properties, $\mathcal{ELI}$ extending $\mathcal{EL}$ with inverse roles. Both DLs are natural choices for generalization as their limited expressive power helps to avoid overfitting, that is, we cannot generalize by disjunctively combining descriptions of each single example, but are forced to find a true generalization. In fact, least general generalizations in $\mathcal{EL}$ have received significant attention (Baader, Küsters, and Molitor 1999; Baader 2003; Zarrieß and Turhan 2013) while, somewhat surprisingly, there appears to be no prior work on DLs with inverse roles.

There are two established notions of least general generalization in the DL context. When the examples are given in the form of concepts, the desired generalization is the *least common subsumer (LCS)*, the least general concept that subsumes all examples (Cohen, Borgida, and Hirsh 1992). A natural alternative is to give examples using relational data, which in DLs are represented as an ABox. Traditionally, one uses only a single example, which takes the form of an individual in the data, and then asks for the *most specific concept (MSC)*, that is, the least general concept that the individual is an instance of (Nebel 1990). However, there seems to be no good reason to restrict the MSC to a single example and thus we define it based on multiple examples. In this way, the LCS becomes a special form of MSC in which the data consists of a collection of trees. We remark that $\mathcal{EL}$ and $\mathcal{ELI}$ concepts can be viewed as natural tree query languages for graph databases and knowledge graphs and thus the MSC is useful for data exploration and comprehension, see e.g. (Colucci et al. 2016). It is also related to generating referring expressions (Borgida, Toman, and Weddell 2016).

For both the LCS and the MSC, we study the two decision problems *existence* and *verification*. In fact, both the LCS and the MSC need not exist because there can be an infinite sequence of less and less general generalizations. In verification, one is given a candidate concept and the question is whether the candidate is the LCS or MSC. Verification is relevant, for example, in approaches that try to find the LCS or MSC by refinement operators that move towards less general generalizations in a step-wise fashion (Badea and Nienhuys-Cheng 2000; Lehmann and Hitzler 2010; Lehmann and

Haase 2009) and check after each step whether the least general generalization has already been reached. We consider the case with and without a background TBox and with and without a target signature that the generalization should be formulated in. If the generalization does not exist, one can resort to approximations (Küsters and Molitor 2001; Baader, Sertkaya, and Turhan 2007).

We now summarize our main complexity and undecidability results. They are based on characterizations in terms of simulations between products of universal models, mildly varying characterizations given in (Zarrieß and Turhan 2013; Funk et al. 2019). We start with the case without TBoxes, for which we find LCS and MSC verification in $\mathcal{EL}$ to be coNP-complete. It is well-known that the LCS in $\mathcal{EL}$ always exists (Baader, Küsters, and Molitor 1999), and we complement this by proving that MSC existence in $\mathcal{EL}$ is PSPACE-complete. We then add inverse roles which introduce significant technical challenges. In particular, the structure of the relevant products from the mentioned characterizations is much more complex. As a consequence, the LCS in $\mathcal{ELI}$ is not guaranteed to exist. We prove that LCS and MSC existence and verification are PSPACE-hard and in EXPTIME. The lower bounds require a remarkably intricate construction and show as a by-product that the *product simulation problem* on trees (defined in the paper) is PSPACE-hard.

We then switch to the case with TBoxes, starting with observing a connection to concept learning (Badea and Nienhuys-Cheng 2000; Lehmann and Hitzler 2010; Lehmann and Haase 2009; Lisi 2012; Bühmann et al. 2018; Sarker and Hitzler 2019) and in particular to the concept separability problem (Funk et al. 2019) which asks whether there is a concept that separates given positive examples from given negative examples. It turns out that its complement reduces in polynomial time to MSC existence. Using results from (Funk et al. 2019), this can be used to show that MSC existence is undecidable in $\mathcal{ELI}$ and EXPTIME-complete in $\mathcal{EL}$. The same is true for verification as the two problems are mutually reducible in polynomial time when a TBox can be used. We consider it remarkable that inverse roles have such a dramatic computational effect. We also identify a way around undecidability, namely to consider for the generalization only *symmetry free* $\mathcal{ELI}$ concepts, that is, $\mathcal{ELI}$ concepts that do not admit a subconcept of the form $\exists r.(C \sqcap \exists r^-.D)$. In this case, the complexity drops to EXPTIME again. Up to this point, all mentioned complexity lower bounds and undecidability results hold without a signature restriction on the target concept while all upper bounds apply also with such a restriction. We finally consider the MSC of single examples and show that existence and verification are in PTIME in $\mathcal{EL}$ while they are complete for EXPTIME and 2-EXPTIME in $\mathcal{ELI}$, depending on whether or not we assume the signature to be full. Thus once more, adding inverse roles has a drastic effect.

Note that in the literature, the LCS is sometimes restricted to only constantly many examples. In all of the above results, we do not assume a constant bound on the number of examples. We also make observations regarding that case, though. Without a TBox, the complexity typically drops to PTIME and the same is true for $\mathcal{EL}$ with TBoxes (Zarrieß and Turhan 2013). When both inverse roles and TBoxes are present, however, the complexity tends to not decrease. We remark that in the decidable cases, our constructions yield upper bounds on the role depth of the LCS and MSC, if they exists, which together with the characterizations can be used to actually construct them.

A full version that contains all proof details is available at http://www.informatik.uni-bremen.de/tdki/research/.

2 Preliminaries

We introduce the basics of DLs as required for this paper, for full details see (Baader et al. 2017). Let $\mathsf{N_C}$ be a set of *concept names* and $\mathsf{N_R}$ a set of *role names*, both countably infinite. A *role* is either a role name or an *inverse role* r^-, r a role name. For uniformity, we identify $(r^-)^-$ with r. An $\mathcal{ELI}$ *concept* is formed according to the syntax rule

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

where A ranges over concept names and r over roles. An $\mathcal{EL}$ concept is an $\mathcal{ELI}$ concept that does not use inverse roles. The *depth* of a concept refers to the nesting depth of the operator $\exists r.C$.

For any DL $\mathcal{L}$, an $\mathcal{L}$ *TBox* is a finite set of *concept inclusions* (CIs) $C \sqsubseteq D$, where C and D are $\mathcal{L}$ concepts. Let $\mathsf{N_I}$ be a countably infinite set of *individual names*. An *ABox* $\mathcal{A}$ is a finite set of *concept assertions* $A(a)$ and *role assertions* $r(a, b)$ where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. We often use $r(a, b)$ to denote $r^-(b, a)$ if r is an inverse role. We use $\mathsf{ind}(\mathcal{A})$ to denote the set of all individual names that occur in $\mathcal{A}$. An $\mathcal{L}$ *knowledge base* (KB) $(\mathcal{T}, \mathcal{A})$ consists of an $\mathcal{L}$ TBox $\mathcal{T}$ and an ABox $\mathcal{A}$.

The semantics of DLs is defined in terms of *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set and $\cdot^\mathcal{I}$ maps each concept name $A \in \mathsf{N_C}$ to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$ and each role name $r \in \mathsf{N_R}$ to a binary relation $r^\mathcal{I}$ on $\Delta^\mathcal{I}$. We refer to (Baader et al. 2017) for details on how to extend $\cdot^\mathcal{I}$ to compound concepts. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, a concept assertion $A(a)$ if $a \in A^\mathcal{I}$, and a role assertion $r(a, b)$ if $(a, b) \in r^\mathcal{I}$. $\mathcal{I}$ is a *model* of a TBox, an ABox, or a knowledge base if it satisfies all inclusions and assertions in it. The CI $C \sqsubseteq D$ is a *consequence of the TBox* $\mathcal{T}$, in symbols $\mathcal{T} \models C \sqsubseteq D$, if $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{T}$. For a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, a concept C, and an individual $a \in \mathsf{ind}(\mathcal{A})$, we write $\mathcal{K} \models C(a)$ if $a \in C^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{K}$. For a DL $\mathcal{L}$, $\mathcal{L}$ *instance checking* is the problem to decide, given an $\mathcal{L}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, an $a \in \mathsf{ind}(\mathcal{A})$, and an $\mathcal{L}$ concept C, whether $\mathcal{K} \models C(a)$.

A *signature* Σ is a set of concept and role names. An $\mathcal{L}$ concept is an $\mathcal{L}(\Sigma)$ concept if it uses only concept and role names from Σ, and likewise for other syntactic objects such as TBoxes and ABoxes. The *signature* $\mathsf{sig}(O)$ of a syntactic object O is the set of concept and role names that occur in O. The Σ-*reduct* $\mathcal{I}_{|\Sigma}$ of an interpretation $\mathcal{I}$ is obtained from $\mathcal{I}$ by setting $A^\mathcal{I} = \emptyset$ and $r^\mathcal{I} = \emptyset$ for all concept names A and role names r not in Σ.

Each interpretation $\mathcal{I}$ gives rise to a directed graph $G_\mathcal{I} = (\Delta^\mathcal{I}, \{(d, e) \mid (d, e) \in r^\mathcal{I}\})$ and a corresponding undirected graph $G^u_\mathcal{I}$. We thus apply graph theoretic terminology directly to interpretations, speaking for example about their outdegree. An interpretation is *tree-shaped* (resp. *ditree-shaped*) if $G^u_{\mathcal{I}}$ (resp. $G_{\mathcal{I}}$) is a tree without multiedges, that is, $(d, e) \in r^{\mathcal{I}} \cap s^{\mathcal{I}}$ implies $r = s$ for all roles r, s. Each $\mathcal{ELI}$ (resp. $\mathcal{EL}$) concept C can be viewed as a tree-shaped (resp. ditree-shaped) interpretation and vice versa. All this also applies to ABoxes, which are only a different way to present finite interpretations. We use $\mathcal{A}_C$ to denote the $\mathcal{ELI}$ concept C viewed as a tree-shaped ABox and use ρ_C to denote the root of $\mathcal{A}_C$. For example, $C = A \sqcap \exists r.B \sqcap \exists r^-.\top$ gives $\mathcal{A}_C = \{A(\rho_C), r(\rho_C, b_1), B(b_1), r(b_2, \rho_C)\}$.

Lemma 1 *For all $\mathcal{ELI}$ TBoxes $\mathcal{T}$ and $\mathcal{ELI}$ concepts C, D, $\mathcal{T} \models C \sqsubseteq D$ iff $(\mathcal{T}, \mathcal{A}_C) \models D(\rho_C)$.*

We introduce simulations, universal models, and direct products. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be interpretations. A relation $S \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is an $\mathcal{EL}(\Sigma)$ *simulation* from $\mathcal{I}_1$ to $\mathcal{I}_2$ if for all $d, d' \in \Delta^{\mathcal{I}_1}$ and $e \in \Delta^{\mathcal{I}_2}$:

1. $d \in A^{\mathcal{I}_1}$ and $(d, e) \in S$ imply $e \in A^{\mathcal{I}_2}$, for all $A \in \Sigma$;
2. $(d, d') \in r^{\mathcal{I}_1}$ and $(d, e) \in S$ imply $(d', e') \in S$ and $(e, e') \in r^{\mathcal{I}_2}$ for some $e' \in \Delta^{\mathcal{I}_2}$, for all role names $r \in \Sigma$.

S is an $\mathcal{ELI}(\Sigma)$ *simulation* if Condition 2 also holds for inverse roles r^- with $r \in \Sigma$. Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$ and $(d, e) \in \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$. We write $(\mathcal{I}_1, d) \preceq_{\mathcal{L},\Sigma} (\mathcal{I}_2, e)$ if there exists an $\mathcal{L}(\Sigma)$ simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$ that contains (d, e). We omit Σ if it is the *full signature* $\mathsf{N_C} \cup \mathsf{N_R}$, writing $\preceq_{\mathcal{L}}$ and speaking of $\mathcal{L}$ simulations. It can be checked in polynomial time whether $(\mathcal{I}_1, d) \preceq_{\mathcal{L},\Sigma} (\mathcal{I}_2, e)$. The following lemma shows that $\mathcal{L}(\Sigma)$ simulations characterize preservation of $\mathcal{L}(\Sigma)$ concepts.

Lemma 2 *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, let $\mathcal{I}_1, \mathcal{I}_2$ be interpretations with finite outdegree, and let Σ be a signature. The following are equivalent:*

1. *$(\mathcal{I}_1, d) \preceq_{\mathcal{L},\Sigma} (\mathcal{I}_2, e)$;*
2. *for all $\mathcal{L}(\Sigma)$ concepts C: if $d \in C^{\mathcal{I}_1}$, then $e \in C^{\mathcal{I}_2}$.*

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a KB and $\mathsf{sub}(\mathcal{T})$ be the set of all subconcepts of concepts that occur in $\mathcal{T}$. A *type for* $\mathcal{T}$ is a subset $t \subseteq \mathsf{sub}(\mathcal{T})$ such that $\mathcal{T} \models \bigsqcap t \sqsubseteq D$ implies $D \in t$ for all $D \in \mathsf{sub}(\mathcal{T})$. Denote by T the set of all types for $\mathcal{T}$. When $a \in \mathsf{ind}(\mathcal{A})$, $t, t' \in T$, and r is a role, we write

- $a \leadsto^{\mathcal{K}}_r t$ if $\mathcal{K} \models \exists r.\bigsqcap t(a)$ and t is maximal with this condition, and
- $t \leadsto^{\mathcal{T}}_r t'$ if $\mathcal{T} \models \bigsqcap t \sqsubseteq \exists r.\bigsqcap t'$ and t' is maximal with this condition.

A *path* p for $\mathcal{K}$ is a sequence $ar_0t_1 \cdots r_{n-1}t_n$ such that $a \in \mathsf{ind}(\mathcal{A})$, $r_0, \ldots, r_{n-1}$ are roles, $t_1, \ldots, t_n \in T$, $a \leadsto^{\mathcal{K}}_{r_0} t_1$, and $t_i \leadsto^{\mathcal{T}}_{r_i} t_{i+1}$ for all $i < n$. Let $\mathsf{tail}(p)$ denote the last element of the path p. Define the *universal model* $\mathcal{U}_{\mathcal{K}}$ *of* $\mathcal{K}$ by taking as $\Delta^{\mathcal{U}_{\mathcal{K}}}$ the set of all paths for $\mathcal{K}$ and setting for all concept names A and role names r:

$$
\begin{aligned}
A^{\mathcal{U}_{\mathcal{K}}} &= \{a \in \mathsf{ind}(\mathcal{A}) \mid \mathcal{T}, \mathcal{A} \models A(a)\} \cup \\
&\quad\ \{p \in \Delta^{\mathcal{U}_{\mathcal{K}}} \setminus \mathsf{ind}(\mathcal{A}) \mid A \in \mathsf{tail}(p)\} \\
r^{\mathcal{U}_{\mathcal{K}}} &= \{(a, b) \in \mathsf{ind}(\mathcal{A})^2 \mid r(a, b) \in \mathcal{A}\} \cup \\
&\quad\ \{(p, prt) \mid prt \in \Delta^{\mathcal{U}_{\mathcal{K}}}\} \cup \{(pr^-t, p) \mid pr^-t \in \Delta^{\mathcal{U}_{\mathcal{K}}}\}
\end{aligned}
$$

The *universal model* $\mathcal{U}_{\mathcal{T},C}$ of an $\mathcal{ELI}$ TBox $\mathcal{T}$ and an $\mathcal{ELI}$ concept C is defined as $\mathcal{U}_{\mathcal{K}}$ where $\mathcal{K} = (\mathcal{T}, \mathcal{A}_C)$.

Lemma 3 *For all $\mathcal{ELI}$ KBs $\mathcal{K}$, $\mathcal{ELI}$ concepts C, and $a \in \mathsf{ind}(\mathcal{K})$, $\mathcal{K} \models C(a)$ iff $a \in C^{\mathcal{U}_{\mathcal{K}}}$.*

The *direct product* $\prod_{i=1}^n \mathcal{I}_i$ of interpretations $\mathcal{I}_1, \ldots, \mathcal{I}_n$ is defined by

$$
\begin{aligned}
\Delta^{\prod_{i=1}^n \mathcal{I}_i} &= \Delta^{\mathcal{I}_1} \times \cdots \times \Delta^{\mathcal{I}_n} \\
A^{\prod_{i=1}^n \mathcal{I}_i} &= A^{\mathcal{I}_1} \times \cdots \times A^{\mathcal{I}_n} \\
r^{\prod_{i=1}^n \mathcal{I}_i} &= \{((d_1, \ldots, d_n), (e_1, \ldots, e_n)) \mid \forall i : (d_i, e_i) \in r^{\mathcal{I}_i}\}
\end{aligned}
$$

If $(d_1, \ldots, d_n) \in \Delta^{\prod_{i=1}^n \mathcal{I}_i}$, then we write $\prod_{i=1}^n (\mathcal{I}_i, d_i)$ for the pair $(\prod_{i=1}^n \mathcal{I}_i, (d_1, \ldots, d_n))$.

Lemma 4 *For all $\mathcal{I}_1, \ldots, \mathcal{I}_n$, $(d_1, \ldots, d_n) \in \Delta^{\prod_{i=1}^n \mathcal{I}_i}$, and $\mathcal{ELI}$ concepts C, $(d_1, \ldots, d_n) \in C^{\prod_{i=1}^n \mathcal{I}_i}$ iff $d_i \in C^{\mathcal{I}_i}$ for $1 \leq i \leq n$.*

3 LCS and MSC: Basics

We introduce least common subsumers and most specific concepts, discuss their relationship, and give model-theoretic characterizations for verification and existence. The latter are mild extensions of characterizations established in (Zarrieß and Turhan 2013).

Definition 1 *Let $\mathcal{T}$ be a TBox, $C_1, \ldots, C_n$ concepts called* examples, *$\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, and Σ a signature. An $\mathcal{L}(\Sigma)$ concept D is a* least common $\mathcal{L}(\Sigma)$ subsumer ($\mathcal{L}(\Sigma)$-LCS) of $C_1, \ldots, C_n$ w.r.t. $\mathcal{T}$ *if*

1. *$\mathcal{T} \models C_i \sqsubseteq D$ for all $i = 1, \ldots, n$;*
2. *if $\mathcal{T} \models C_i \sqsubseteq D'$ for all $i = 1, \ldots, n$, D' an $\mathcal{L}(\Sigma)$ concept, then $\mathcal{T} \models D \sqsubseteq D'$.*

If an $\mathcal{L}(\Sigma)$-LCS w.r.t. a TBox $\mathcal{T}$ exists, then it is unique up to equivalence w.r.t. $\mathcal{T}$. We thus speak about *the* $\mathcal{L}(\Sigma)$-LCS. We omit Σ if it contains $\mathsf{sig}(\mathcal{T} \cup \{C_1, \ldots, C_n\})$, speaking of the *$\mathcal{L}$-LCS* w.r.t. $\mathcal{T}$. Clearly, no $\mathcal{L}$-LCS can contain symbols that are not in the TBox or the examples. Thus, all signatures between the finite $\mathsf{sig}(\mathcal{T} \cup \{C_1, \ldots, C_n\})$ and the full signature behave in the same way. We also omit $\mathcal{T}$ if it is empty, speaking of the *$\mathcal{L}(\Sigma)$-LCS*.

Example 1 (1) Let $C_1 = \exists \mathsf{attend}.\mathsf{MLConf}$ and $C_2 = \exists \mathsf{attend}.\mathsf{KRConf}$. Then $\exists \mathsf{attend}.\top$. is the $\mathcal{EL}$ (and $\mathcal{ELI}$) LCS of C_1, C_2. Let $\mathcal{T} = \{\mathsf{MLConf} \sqsubseteq \mathsf{AIConf}, \mathsf{KRConf} \sqsubseteq \mathsf{AIConf}\}$. Then $\exists \mathsf{attend}.\mathsf{AIConf}$ is the $\mathcal{EL}$ (and $\mathcal{ELI}$) LCS of C_1, C_2 w.r.t. $\mathcal{T}$.

(2) The $\mathcal{L}$-LCS, $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, of a single $\mathcal{L}$ concept C w.r.t. an $\mathcal{L}$ TBox $\mathcal{T}$ is just C. For $\Sigma \subsetneq \mathsf{sig}(C)$, however, the $\mathcal{L}(\Sigma)$-LCS of C w.r.t. $\mathcal{T}$ does not always exist. Take, for example, $\mathcal{T} = \{A \sqsubseteq \exists r.A\}$ and $\Sigma = \{r\}$. Then neither the $\mathcal{ELI}(\Sigma)$-LCS nor the $\mathcal{EL}(\Sigma)$-LCS of A w.r.t. $\mathcal{T}$ exists as $\mathcal{T} \models A \sqsubseteq \exists r^n.\top$ for all $n \geq 0$, but there is no $\mathcal{ELI}(\Sigma)$ concept C with $\mathcal{T} \models A \sqsubseteq C$ and $\mathcal{T} \models C \sqsubseteq \exists r^n.\top$ for all n.

Definition 2 *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a KB, $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$ individuals called* examples, *$\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, and Σ a signature. An $\mathcal{L}(\Sigma)$ concept C is a* most specific $\mathcal{L}(\Sigma)$ concept ($\mathcal{L}(\Sigma)$ MSC) of $a_1, \ldots, a_n$ w.r.t. $\mathcal{K}$ *if*

1. *$\mathcal{K} \models C(a_i)$ for all $i = 1, \ldots, n$;*
2. *if $\mathcal{K} \models D(a_i)$ for all $i = 1, \ldots, n$, D an $\mathcal{L}(\Sigma)$ concept, then $\mathcal{T} \models C \sqsubseteq D$.*

Like the LCS, the MSC is unique up to equivalence w.r.t. $\mathcal{T}$ (if it exists) and thus we speak of *the* MSC. We drop Σ if $\Sigma \supseteq \mathsf{sig}(\mathcal{K})$. As for the LCS, a symbol that does not occur in the KB cannot occur in the MSC.

Example 2 (1) In contrast to the $\mathcal{EL}$-LCS, the $\mathcal{EL}$-MSC of a single example does not always exist, even when the TBox is empty, due to cycles in the ABox. For example, for $\mathcal{A} = \{A(a), r(a,a)\}$ the $\mathcal{EL}$-MSC of a w.r.t. $\mathcal{K} = (\emptyset, \mathcal{A})$ does not exist (use that $\mathcal{K} \models \exists r^n.\top(a)$ for all $n \geq 0$). In contrast, the $\mathcal{EL}$-MSC of a w.r.t. $\mathcal{K}' = (\{A \sqsubseteq \exists r.A\}, \mathcal{A})$ is A.

(2) A common proposal to generalize from individuals is to compute the MSC of each individual separately and then generalize by applying the LCS, provided that all MSCs exist (Baader, Küsters, and Molitor 1999). It pays off, however, to directly apply the MSC to multiple individuals. Let, for example, $\mathcal{K} = (\emptyset, \mathcal{A})$, $\mathcal{A} = \{A(a), r(a,a), A(b), s(b,b)\}$. Then the $\mathcal{EL}$-MSC of a alone w.r.t. $\mathcal{K}$ does not exist, and likewise for b. In constrast, the $\mathcal{EL}$-MSC of a, b w.r.t. $\mathcal{K}$ is A.

The following theorem, which is an immediate consequence of Lemma 1, shows that the LCS is a special form of MSC.

Theorem 1 *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{T}$ be an $\mathcal{L}$ TBox, $C_1, \dots, C_n$ $\mathcal{L}$ concepts, and Σ a signature. Then an $\mathcal{L}(\Sigma)$ concept D is the $\mathcal{L}(\Sigma)$-LCS of $C_1, \dots, C_n$ w.r.t. $\mathcal{T}$ iff D is the $\mathcal{L}(\Sigma)$-MSC of $\rho_{C_1}, \dots, \rho_{C_n}$ w.r.t. the KB $(\mathcal{T}, \mathcal{A})$, $\mathcal{A} = \mathcal{A}_{C_1} \cup \dots \cup \mathcal{A}_{C_n}$.*

LCS and MSC give rise to the four decision problems studied in this paper. Let $\mathcal{L}$ be a description logic. *$\mathcal{L}$-LCS existence w.r.t. TBoxes* means to decide, given $\mathcal{L}$ concepts $C_1, \dots, C_n$, an $\mathcal{L}$ TBox $\mathcal{T}$, and a finite signature Σ, whether the $\mathcal{L}(\Sigma)$-LCS of $C_1, \dots, C_n$ w.r.t. $\mathcal{T}$ exists. By the remark made after Definition 1, it is without loss of generality to consider only finite signatures. In particular, we can use $\mathsf{sig}(\mathcal{T} \cup \{C_1, \dots, C_n\})$ instead of the full signature. *$\mathcal{L}$-MSC existence w.r.t. TBoxes* is defined accordingly, the input consisting of a KB $(\mathcal{T}, \mathcal{A})$ with $\mathcal{T}$ an $\mathcal{L}$ TBox, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$, and a finite signature Σ. In *$\mathcal{L}$-LCS* (resp. *$\mathcal{L}$-MSC*) *verification w.r.t. TBoxes*, we are given as an additional input a candidate $\mathcal{L}(\Sigma)$ concept C and the question is whether C is the $\mathcal{L}(\Sigma)$-LCS of $C_1, \dots, C_n$ w.r.t. $\mathcal{T}$ (resp. the $\mathcal{L}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $\mathcal{K}$).

Theorem 1 provides a reduction from $\mathcal{L}$-LCS existence w.r.t. TBoxes to $\mathcal{L}$-MSC existence w.r.t. TBoxes, and likewise for verification. In this reduction, neither the TBox nor the signature nor the number of examples change. We now present a converse reduction which, however, requires to modify the TBox.

Theorem 2 *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. Then $\mathcal{L}$-MSC verification (resp. existence) w.r.t. TBoxes can be reduced in polynomial time to $\mathcal{L}$-LCS verification (resp. existence). This also holds in the full signature case if there are at least two examples.*

Proof. Let $\mathcal{T}$ be an $\mathcal{L}$ TBox, $\mathcal{A}$ an ABox, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$. We may assume w.l.o.g. that $\mathcal{A}$ is the disjoint union of ABoxes $\mathcal{A}_1, \dots, \mathcal{A}_n$ such that $a_i \in \mathsf{ind}(\mathcal{A}_i)$ for $i = 1, \dots, n$. Let X_a be a fresh concept name for every $a \in \mathsf{ind}(\mathcal{A})$ and let $\mathcal{T}'$ be the extension of $\mathcal{T}$ with

$$\begin{array}{lll} X_a & \sqsubseteq A & \text{for all } A(a) \in \mathcal{A}, \\ X_a & \sqsubseteq \exists r.X_{a'} & \text{for all } r(a,a') \in \mathcal{A}. \end{array}$$

(If $\mathcal{L} = \mathcal{ELI}$, then also add $X_a \sqsubseteq \exists r^-.X_{a'}$ if $r(a', a) \in \mathcal{A}$.) Then for every signature Σ that does not contain $\{X_{a_1}, \dots, X_{a_n}\}$ and every $\mathcal{L}(\Sigma)$ concept D, D is the $\mathcal{L}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $(\mathcal{T}, \mathcal{A})$ iff D is the $\mathcal{L}(\Sigma)$-LCS of $X_{a_1}, \dots, X_{a_n}$ w.r.t. $\mathcal{T}'$.

In the case of the full signature, we have to consider the $\mathcal{L}(\Sigma \cup \{X_{a_1}, \dots, X_{a_n}\})$-LCS in place of the $\mathcal{L}(\Sigma)$-LCS. The assumption that there are at least two examples ensures that the concept names X_a cannot occur in the LCS. ❏

We next provide model-theoretic characterizations for MSC verification and existence based on products and simulations. Corresponding characterizations for LCS verification and existence can be obtained in a straightforward way via Theorem 1, see the appendix. Note that Point 1 below can also be viewed as a simulation condition.

Theorem 3 (MSC Verification) *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{L}$ KB, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$, and Σ a signature. An $\mathcal{L}(\Sigma)$ concept C is the $\mathcal{L}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $\mathcal{K}$ iff the following conditions hold:*

1. $(a_1, \dots, a_n) \in C^{\Pi_{i=1}^n \mathcal{U}_\mathcal{K}}$;
2. $\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i) \preceq_{\mathcal{L},\Sigma} \mathcal{U}_{\mathcal{T},C}, \rho_C$.

Proof. By Lemmas 3 and 4, Condition 1 is equivalent to Condition 1 of the definition of MSCs. By Lemmas 2, 3, and 4, Condition 2 is equivalent to Condition 2 of the definition of MSCs. ❏

For an interpretation $\mathcal{I}$ and a $d_0 \in \Delta^\mathcal{I}$, a *$d_0$-path of length k in $\mathcal{I}$* is a sequence $d_0 r_0 \cdots r_{k-1} d_k$ with $(d_i, d_{i+1}) \in r_i^\mathcal{I}$ for all $i < k$, each r_i a (potentially inverse) role. Denote by $\mathsf{tail}(p)$ the last element of p. The *$\mathcal{ELI}, k$-unfolding of $\mathcal{I}$ at d_0*, denoted $(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}, k}$, is the interpretation defined by taking $\Delta^{(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}, k}}$ to be the set of all d_0-paths of length at most k and setting

$$\begin{aligned} A^{(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}, k}} &= \{p \mid \mathsf{tail}(p) \in A^\mathcal{I}\} \\ r^{(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}, k}} &= \{(p, prt) \mid prt \in \Delta^{(\mathcal{I}, a)^{\downarrow \mathcal{ELI}, k}}\} \cup \\ &\qquad \{(pr^-t, p) \mid prt \in \Delta^{(\mathcal{I}, a)^{\downarrow \mathcal{ELI}, k}}\}. \end{aligned}$$

The *$\mathcal{EL}, k$-unfolding of $\mathcal{I}$ at d_0*, denoted $(\mathcal{I}, d_0)^{\downarrow \mathcal{EL}, k}$, is defined accordingly, but only admitting role names in paths. For $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$ and an $\mathcal{L}$ KB $\mathcal{K}$, we use $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, d_i))^{\downarrow \mathcal{L}, k}_{|\Sigma}$ to denote the $\mathcal{L}, k$-unfolding of the Σ-reduct of $\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, d_i)$ at $(d_1, \dots, d_n)$. It can be verified that this interpretation is tree-shaped for $\mathcal{L} = \mathcal{ELI}$ and ditree-shaped for $\mathcal{L} = \mathcal{EL}$ and can thus be viewed as an $\mathcal{L}$ concept C_k.

Theorem 4 (MSC Existence) *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{L}$ KB, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$, and Σ a signature. The following are equivalent, for $C_k = (\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{L}, k}_{|\Sigma}$:*

1. *the $\mathcal{L}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $\mathcal{K}$ exists;*
2. *C_k is the $\mathcal{L}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $\mathcal{K}$, for a $k \geq 0$;*
3. *$\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i) \preceq_{\mathcal{L},\Sigma} (\mathcal{U}_{\mathcal{T},C_k}, \rho_{C_k})$ for some $k \geq 0$.*

Proof. "2 $\Rightarrow$ 1" is trivial. "3 $\Rightarrow$ 2" is an immediate consequence of Theorem 3. For "1 $\Rightarrow$ 3", let the $\mathcal{L}(\Sigma)$-MSC D be of depth k. It then follows from Theorem 3 that $(a_1, \ldots, a_n) \in D^{\prod_{i=1}^n \mathcal{U}_\mathcal{K}}$ which implies $\rho_{C_k} \in D^{\mathcal{U}_{\mathcal{T},C_k}}$. Now Point 3 follows from the definition of the MSC and Lemmas 2, 3, and 4. ❑

Note that Theorems 3 and 4 link MSC-verification and existence, as well as LCS-verification and existence (via Theorem 1) to product simulation problems. For $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, the *$\mathcal{L}$-product simulation problem* is to decide given $(\mathcal{I}_1, d_1), \ldots, (\mathcal{I}_n, d_n), (\mathcal{J}, e)$, whether $\prod_{i=1}^n (\mathcal{I}_i, d_i) \preceq_\mathcal{L} (\mathcal{J}, e)$. These are fundamental problems that have received attention in several areas such as verification and database theory (Harel, Kupferman, and Vardi 2002; Barceló and Romero 2017; ten Cate and Dalmau 2015).

4 Without TBoxes

We start with studying least general generalizations in the case without TBoxes, beginning with verification in $\mathcal{EL}$.

Theorem 5 *In $\mathcal{EL}$, LCS and MSC verification w.r.t. the empty TBox are* CONP*-complete. The lower bounds apply even when the signature is full.*

Proof. (sketch) The upper bound uses Theorem 3, the fact that instance checking in $\mathcal{EL}$ is in PTIME, and the observation that the $\mathcal{EL}$-product simulation problem is in CONP if the interpretation $\mathcal{J}$ is tree-shaped (here, it is even ditree-shaped). In fact, if $(\mathcal{I}, d) \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e)$ with $\mathcal{J}$ tree-shaped, then there is a subinterpretation $\mathcal{I}_0$ of $\mathcal{I}$ of polynomial size such that $(\mathcal{I}_0, d) \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e)$. The lower bound is proved by reducing the satisfiability problem for propositional logic to the complement of $\mathcal{EL}$-LCS verification. It also establishes CONP-hardness of the $\mathcal{EL}$-product simulation problem in the case that $\mathcal{J}$ is tree-shaped. ❑

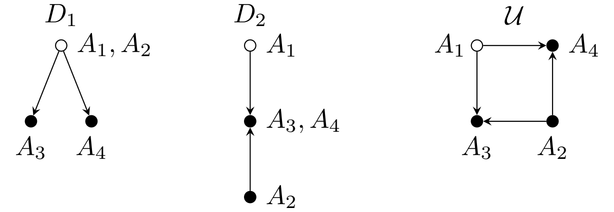
Regarding existence, a first well-known observation is that the $\mathcal{EL}$-LCS always exists, even if the signature is not full. This follows from Theorem 4 and the fact that if $\mathcal{K} = (\emptyset, \mathcal{A}_{C_1} \cup \cdots \cup \mathcal{A}_{C_n})$ then the (reachable part of the) Σ-reduct of $\prod_{i=1}^n (\mathcal{U}_\mathcal{K}, \rho_{C_i})$ is ditree-shaped and coincides with $\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, \rho_{C_i})^{\downarrow \mathcal{EL},k}_{|\Sigma}$, k the maximum depth of $C_1, \ldots, C_n$. In contrast, the $\mathcal{EL}$-MSC does not always exist even with the empty TBox, see Example 2.

Theorem 6 *In $\mathcal{EL}$, MSC existence w.r.t. the empty TBox is* PSPACE*-complete. The lower bound applies even when the signature is full.*

Proof. (sketch) Using Theorem 4, one can show that the $\mathcal{EL}(\Sigma)$-MSC of $a_1, \ldots, a_n$ w.r.t. a KB $\mathcal{K} = (\emptyset, \mathcal{A})$ exists if and only if there is no infinite Σ-path in $\mathcal{A}^n = \prod_{i=1}^n \mathcal{A}$ that starts at $(a_1, \ldots, a_n)$—we view ABoxes as finite interpretations here. We can thus decide existence of the $\mathcal{EL}(\Sigma)$-MSC in polynomial space in the standard way: guess an element a of $\mathcal{A}^n$ and, proceeding step by step, a path through $\mathcal{A}^n$ that starts at $(a_1, \ldots, a_n)$ and follows only role names from Σ. Reject if the element a is seen twice. The lower bound is established by reducing the word problem of deterministic polynomially space-bounded Turing machines. ❑

We next turn to $\mathcal{ELI}$. In contrast to $\mathcal{EL}$, here the LCS does not always exist even when the TBox is empty.

Example 3 Consider the following $\mathcal{ELI}$ concepts D_1, D_2 over concept names $A_1, \ldots, A_4$ and a single role r:



The interpretation $\mathcal{U}$ is the part of $\mathcal{A}_{D_1} \times \mathcal{A}_{D_2}$ that is reachable from its root $\circ$. One can show that the infinite path in $\mathcal{U}$ labeled with $(A_1, r, A_3, r^-, A_2, r, A_4, r^-)^\omega$ is not $\mathcal{ELI}$ simulated by $(\mathcal{U}^{\downarrow \mathcal{ELI},k}, \circ)$, for any $k \geq 0$. Thus, the $\mathcal{ELI}$-LCS of D_1, D_2 does not exist by Theorem 4.

The next theorem summarizes our results regarding $\mathcal{ELI}$.

Theorem 7 *In $\mathcal{ELI}$, LCS and MSC existence and verification w.r.t. the empty TBox are* PSPACE*-hard and in* EXPTIME*. The lower bounds apply when the signature is full.*

Proof. (sketch) The main ingredient to the PSPACE lower bounds is a rather intricate proof that the $\mathcal{ELI}$-product simulation problem is PSPACE-hard already when restricted to tree-shaped interpretations. In fact, this is the case even when interpretations on the left-hand sides are trees of depth two and the interpretation on the right-hand side is fixed (and of depth eleven). It is interesting to contrast this with the fact that the $\mathcal{EL}$-product simulation problem is CONP-complete on tree-shaped interpretations, see the proof of Theorem 5. To obtain a PSPACE lower bound for LCS verification and existence, we then use reductions from $\mathcal{ELI}$-product simulation on tree shaped interpretations.

The upper bound for MSC verification (and thus also for LCS verification) is obtained by recalling that $\mathcal{ELI}$ instance checking is EXPTIME-complete and adapting the EXPTIME upper bound from (Zarrieß and Turhan 2013) for the $\mathcal{EL}$-product simulation problem to $\mathcal{ELI}$.

The EXPTIME upper bound for MSC existence (and thus also for LCS existence) can be proved similarly to the upper bound in Theorem 6. The main difference is that we now work with $\mathcal{ELI}$ simulations rather than $\mathcal{EL}$ simulations and thus need to be more careful about the paths we consider. In fact, we use paths $d_0, r_0, d_1, r_1, d_2, \ldots$ through $\mathcal{A}^n = \prod_{i=1}^n \mathcal{A}$ that start at $d_0 = (a_1, \ldots, a_n)$, follow only Σ-roles, and satisfy the following for all $i \geq 0$: 1. if $r_i = r_{i+1}^-$, then $(\mathcal{A}^n, d_{i+2}) \not\preceq_{\mathcal{ELI},\Sigma} (\mathcal{A}^n, d_i)$; 2. there is no $e \neq d_{i+1}$ such that $r_i(d_i, e) \in \mathcal{A}^n$, $(\mathcal{A}^n, d_{i+1}) \preceq_{\mathcal{ELI},\Sigma} (\mathcal{A}^n, e)$, and $(\mathcal{A}^n, e) \not\preceq_{\mathcal{ELI},\Sigma} (\mathcal{A}^n, d_{i+1})$. ❑

All problems studied in this section are solvable in PTIME if the number of examples is bounded by a constant. This follows from an analysis of the presented upper bound proofs and has in some cases also been established before (Baader, Küsters, and Molitor 1999; Zarrieß and Turhan 2013).

5 With TBoxes

We now add TBoxes to the picture. It turns out that, in this case, we can transfer results from the concept separability problem, which has been considered in concept learning from positive and negative examples (Funk et al. 2019).

Definition 3 *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. An $\mathcal{L}$* learning instance *is a triple $(\mathcal{K}, P, N)$ with $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ an $\mathcal{L}$ KB and $P, N \subseteq \mathsf{ind}(\mathcal{A})$ sets of* positive *and* negative examples. *Let Σ be a signature. An $\mathcal{L}(\Sigma)$* solution *to $(\mathcal{K}, P, N)$ is an $\mathcal{L}(\Sigma)$ concept C such that $\mathcal{K} \models C(a)$ for all $a \in P$ and $\mathcal{K} \not\models C(a)$ for all $a \in N$.*

This definition gives rise to the decision problem of $\mathcal{L}$ *concept separability*: given an $\mathcal{L}$ learning instance $(\mathcal{K}, P, N)$ and a signature Σ, decide whether it admits an $\mathcal{L}(\Sigma)$ solution. As the conjunction of $\mathcal{L}(\Sigma)$ solutions to $(\mathcal{K}, P, \{b\})$, $b \in N$, is an $\mathcal{L}(\Sigma)$ solution to $(\mathcal{K}, P, N)$, it suffices to consider instances with N singleton. Note that in (Funk et al. 2019) only the full signature case is considered.

One can easily derive from (Funk et al. 2019) that $(\mathcal{K}, P, \{b\})$ has an $\mathcal{L}(\Sigma)$ solution iff $\prod_{a\in P}(\mathcal{U}_\mathcal{K}, a) \not\preceq_{\mathcal{L},\Sigma} (\mathcal{U}_\mathcal{K}, b)$. By encoding b as a concept D as in the proof of Theorem 2, we can thus view $\mathcal{L}(\Sigma)$ concept separability as the problem to decide for an $\mathcal{L}$ KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, examples $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$, and an $\mathcal{L}$ concept D whether $\prod_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i) \not\preceq_{\mathcal{L},\Sigma} (\mathcal{U}_{\mathcal{T},D}, \rho_D)$, which is exactly the negation of Condition 2 of the characterization of MSC verification in Theorem 3. This provides the basis for the following.

Theorem 8 *For $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, the complement of $\mathcal{L}$ concept separability can be reduced in polynomial time to $\mathcal{L}$-MSC verification and existence. This also holds for the full signature.*

Proof. (sketch) We consider $\mathcal{EL}$ and the full signature case. Given $\mathcal{K}$, $a_1, \ldots, a_n$, and D, we extend $\mathcal{K}$ by adding assertions $v(\rho_i, a_i), v(\rho_i, b_i), D(b_i)$, where ρ_i and b_i are fresh individuals, v a fresh role name, and $D(b_i)$ stands for $\mathcal{A}_D$ rooted at b_i. Then $\prod_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i) \not\preceq_{\mathcal{EL}} (\mathcal{U}_{\mathcal{T},D}, \rho_D)$ iff $\exists v.D$ is the $\mathcal{EL}$-MSC of $\rho_1, \ldots, \rho_n$ w.r.t. the extended KB (under mild assumptions). For the reduction to MSC existence, we additionally generate infinite r-chains starting at a_i and b_i using CIs $X \sqsubseteq \exists r.X$ and adding $X(a_i)$ and $X(b_i)$ to the ABox, where the concept names X are distinct for distinct a_i but coincide for all b_i. If we assume w.l.o.g. that $n \geq 2$, then $\prod_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i) \not\preceq_{\mathcal{EL}} (\mathcal{U}_{\mathcal{T},D}, \rho_D)$ iff the $\mathcal{EL}$-MSC of $\rho_1, \ldots, \rho_n$ w.r.t. the extended KB exists. ❑

It is shown in (Funk et al. 2019) that $\mathcal{ELI}$ concept separability is undecidable already in the full signature case and even with only two positive examples. We thus obtain the following from Theorems 8 and 2 and the fact that the number of examples remains unchanged under the reductions.

Theorem 9 *In $\mathcal{ELI}$, MSC and LCS verification and existence are undecidable. This is already the case when the signature is full and there are at most two examples.*

It is also shown in (Funk et al. 2019) that $\mathcal{EL}$ concept separability is ExpTime-hard. In this case the number of positive examples is not bounded by a constant.

Theorem 10 *In $\mathcal{EL}$, MSC and LCS verification and existence are* ExpTime*-complete. The lower bounds already apply when the signature is full.*

Proof. (sketch) The lower bounds come from Theorems 8 and 2. ExpTime upper bounds for LCS existence and verification with the full signature are in (Zarrieß and Turhan 2013), the former explicitly and the latter implicitly. They extend to other signatures in a straightforward way. To lift these bounds to the MSC, we use Theorem 2. ❑

When the number of examples is bounded, then all problems in Theorem 10 can be solved in PTime (which was known for LCS existence (Zarrieß and Turhan 2013)).

We close this section with observing that $\mathcal{L}$-MSC verification can be reduced to the complement of concept separability, and thus, by Theorem 8, to $\mathcal{L}$-MSC existence.

Theorem 11 *For $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{L}$-MSC verification can be reduced in polynomial time to the complement of $\mathcal{L}$ concept separability. This also holds for the full signature.*

Proof. (sketch) Recall that Condition 2 of Theorem 3 is the complement of concept separability. By Lemmas 3 and 2, Condition 1 is equivalent to requiring $\mathcal{U}_{\mathcal{T},C}, \rho_C \preceq_\mathcal{L} \mathcal{U}_\mathcal{K}, a_i$, for all i. These simulation checks can be incorporated into Condition 2 by extending the ABox. ❑

6 Symmetry Free $\mathcal{ELI}$

An inspection of the proof of the undecidability results in Theorem 9 reveals that it crucially depends on the MSC and LCS to contain subconcepts of the form $\exists r.(C \sqcap \exists r^-.D)$. Indeed, concept separability is decidable when the TBox is formulated in $\mathcal{ELI}$ while separating concepts are restricted to $\mathcal{EL}$ (Funk et al. 2019). We consider a more general case by restricting the MSC and LCS to *symmetry free $\mathcal{ELI}$* concepts ($\mathcal{ELI}^{\mathsf{sf}}$ *concepts* for short), that is, $\mathcal{ELI}$ concepts that do not contain such subconcepts. With *$\mathcal{ELI}^{\mathsf{sf}}$-LCS and MSC verification and existence w.r.t. $\mathcal{ELI}$ TBoxes*, we mean that the TBox is formulated in $\mathcal{ELI}$ while we seek a least general generalization formulated in $\mathcal{ELI}^{\mathsf{sf}}$. In the case of the LCS, also the examples are formulated in unrestricted $\mathcal{ELI}$.

We start with providing a characterization of $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC existence. To achieve this, we modify the notion of $\mathcal{ELI}, k$-unfolding of an interpretation $\mathcal{I}$ at a $d_0 \in \Delta^\mathcal{I}$ given in Section 3 by restricting the domain of the resulting interpretation to *symmetry free d_0-paths* of length k, that is, to d_0-paths $d_0 r_0 \cdots r_{m-1} d_m$, $m \leq k$, that satisfy $r_i \neq r_{i+1}^-$ for all $i < m$. We speak of the *$\mathcal{ELI}^{\mathsf{sf}}, k$-unfolding of $\mathcal{I}$ at d_0*, denoted $(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}, k}$. We further use $(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}$ to denote the unbounded $\mathcal{ELI}^{\mathsf{sf}}$-unfolding of $\mathcal{I}$ at d_0, that is, the union of all $(\mathcal{I}, d_0)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}, k}$, $k \geq 0$. Now let Σ be a signature. For an $\mathcal{ELI}$ KB $\mathcal{K}$, we use $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, d_i))_{|\Sigma}^{\downarrow \mathcal{ELI}^{\mathsf{sf}}, k}$ to denote the $\mathcal{ELI}^{\mathsf{sf}}, k$-unfolding of the Σ-reduct of $\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, d_i)$ at $(d_1, \ldots, d_n)$. As this interpretation is tree-shaped, it can be viewed as an $\mathcal{ELI}$ concept which is even an $\mathcal{ELI}^{\mathsf{sf}}$ concept.

Theorem 12 **($\mathcal{ELI}^{\mathsf{sf}}$-MSC Existence w.r.t. $\mathcal{ELI}$ TBoxes)** *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELI}$ KB, $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$, and Σ a signature. The following are equivalent, for $C_k = (\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i))_{|\Sigma}^{\downarrow \mathcal{ELI}^{\mathsf{sf}}, k}$:*

1. *the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC of $a_1, \ldots, a_n$ w.r.t. $\mathcal{K}$ exists;*

2. C_k *is the* $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$*-MSC of* $a_1, \ldots, a_n$ *w.r.t.* $\mathcal{K}$*, for a* $k \geq 0$*;*
3. $\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}} \preceq_{\mathcal{ELI},\Sigma} (\mathcal{U}_{\mathcal{T},C_k}, \rho_{C_k})$ *for a* $k \geq 0$.

Since Theorem 1 extends to the case considered in this section, Theorem 12 also yields a characterization for $\mathcal{ELI}^{\mathsf{sf}}$ LCS existence w.r.t. $\mathcal{ELI}$ TBoxes. Theorems 8 and 11 can also be adapted using a version of concept separability where the separating concepts are formulated in $\mathcal{ELI}^{\mathsf{sf}}$. Thus verification reduces to existence in polynomial time and we refrain from giving an explicit characterization.

Theorem 12 provides the basis for proving that symmetry freeness regains decidability.

Theorem 13 $\mathcal{ELI}^{\mathsf{sf}}$*-MSC and LCS existence and verification with respect to* $\mathcal{ELI}$ *TBoxes are* ExpTime*-complete. The lower bounds hold in the full signature case and with only one example.*

The lower bounds are easy to prove by reduction from the subsumption of concept names w.r.t. $\mathcal{ELI}$ TBoxes (Baader, Brandt, and Lutz 2008). For the upper bounds, we use an approach based on automata on infinite trees. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELI}$ KB, $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$, and Σ a signature. Theorem 12 suggests to test emptiness of two tree automata $\mathfrak{A}$ and $\mathfrak{B}$ where $\mathfrak{A}$ accepts precisely the tree-shaped interpretations that admit an $\mathcal{ELI}(\Sigma)$ simulation from $\mathcal{U} := (\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}$ and $\mathfrak{B}$ accepts precisely the tree-shaped interpretations $\mathcal{U}_{\mathcal{T},C_k}, \rho_{C_k}$, $k \geq 0$. In particular, the automaton $\mathfrak{A}$ visits all elements of $\mathcal{U}$ using its states, assigning to each of them a simulating element in the input interpretation. Elements in $\mathcal{U}$ are represented by their type t and the role that led to it—note that these uniquely determine the successors, and that this is not the case without symmetry freeness. We thus have (at least) exponentially many states. To obtain an ExpTime upper bound, we therefore use non-deterministic tree automata (NTA) rather than alternating ones. To avoid having a state for every set of types, we must further make sure that every element in $\mathcal{U}$ is simulated by a different element in the input tree. To have enough room when moving down in the input tree, we slightly refine our characterization.

A simulation S from $\mathcal{I}_1$ to $\mathcal{I}_2$ is *injective* if for all $e \in \Delta^{\mathcal{I}_2}$, there is at most one $d \in \Delta^{\mathcal{I}_1}$ with $(d, e) \in S$. We write $(\mathcal{I}_1, d_1) \preceq^{\mathsf{in}}_{\mathcal{ELI},\Sigma} (\mathcal{I}_2, d_2)$ if there is an injective $\mathcal{ELI}(\Sigma)$-simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$ that contains (d_1, d_2). Let $\mathcal{I}^{\times \ell}$ denote the interpretation that is obtained from a tree-shaped interpretation $\mathcal{I}$ by duplicating every successor in the tree so that it occurs ℓ times.

Lemma 5 *Let* N *be the outdegree of* $\Pi_{i=1}^n \mathcal{U}_\mathcal{K}$*. Then the* $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$*-MSC of* $a_1, \ldots, a_n$ *w.r.t.* $\mathcal{K}$ *exists iff, for some subconcept* D *of* $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}_\Sigma$*, we have:*

$$(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}} \preceq^{in}_{\mathcal{ELI},\Sigma} (\mathcal{U}^{\times N}_{\mathcal{T},D}, \rho_D).$$

Now, $\mathfrak{A}$ accepts the tree-shaped interpretations that admit *injective* $\mathcal{ELI}(\Sigma)$ simulations from $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}$ using exponentially many states. Further, $\mathfrak{B}$ accepts interpretations of the form $\mathcal{U}^{\times N}_{\mathcal{T},D}$ for some D as in the lemma. We first construct an automaton that works over pairs of tree-shaped interpretations and verifies that the first component represents a suitable D and the second component represents $\mathcal{U}_{\mathcal{T},D}$. We then project to the latter and modify the automaton so as to accept all $\mathcal{I}^{\times N}$ with $\mathcal{I}$ accepted before.

7 Single Example MSC

We consider the MSC of a single example, which is the case traditionally studied in the literature. A PTime upper bound for $\mathcal{EL}$ was given in (Zarrieß and Turhan 2013). We show that adding a signature does not affect this result, and that it also holds for verification.

Theorem 14 *In* $\mathcal{EL}$*, single example MSC existence and verification are in* PTime*.*

Proof. (sketch) This is a consequence of the proof of Theorem 13. Applying the constructions from that proof to an $\mathcal{EL}$ TBox instead of an $\mathcal{ELI}$ TBox has two effects: first, all involved automata can be constructed in polynomial time and are of polynomial size; and second Theorem 12 implies that if the $\mathcal{ELI}^{\mathsf{sf}}$-MSC exists, it is actually an $\mathcal{EL}$ concept. ❑

We next show that the $\mathcal{ELI}$ case is dramatically different. In particular, the complexity is much higher and admitting non-full signatures causes an exponential jump in complexity.

Theorem 15 *In* $\mathcal{ELI}$*, single example MSC existence and verification are* 2-ExpTime*-complete in general and* ExpTime*-complete when the signature is full.*

Proof. (sketch) In the full signature case, the lower bound is by reduction from the subsumption of concept names w.r.t. $\mathcal{ELI}$ TBoxes. For unrestricted signatures, we reduce the complement of single example $\mathcal{ELI}$ concept separability, shown 2-ExpTime-hard in (Gutiérrez-Basulto, Jung, and Sabellek 2018), similar to the proof of Theorem 8.

The upper bounds are shown using an automata based approach that is in spirit similar to the approach taken in Section 6. The main difference is that the automaton $\mathfrak{A}$ has to be two-way since it checks for $\mathcal{ELI}$ simulations from $\mathcal{U}_\mathcal{K}, a$. In case of restricted signature, it has to store types in its states, while for the full signature ABox individuals suffice. ❑

8 Discussion

We have analyzed the complexity of LCS and MSC verification and existence in the DLs $\mathcal{EL}$ and $\mathcal{ELI}$, obtaining various complexity results and establishing a close link to concept separability. Topics for future research include tight bounds on the size of the LCS and MSC and studying cases in which the TBoxes is formulated in an expressive DL such as $\mathcal{ALC}$ while the LCS and MSC are formulated in $\mathcal{EL}$ or $\mathcal{ELI}$ (to avoid overfitting). It would also be interesting to study DLs that admit role constraints such as transitive roles and expressive forms of role inclusion. Finally, it would be of interest to study the data complexity, under which the TBox is not regarded as part of the input.

Acknowledgments. Carsten Lutz was supported by the DFG CRC EASE. Frank Wolter was partially supported by EPSRC grant EP/S032207/1.

References

Baader, F., and Küsters, R. 1998. Computing the least common subsumer and the most specific concept in the presence of cyclic aln-concept descriptions. In *Proc. of KI*, 129–140. Springer.

Baader, F.; Brandt, S.; and Lutz, C. 2008. Pushing the $\mathcal{EL}$ envelope further. In *Proc. of OWLED workshop*.

Baader, F.; Küsters, R.; and Molitor, R. 1999. Computing least common subsumers in description logics with existential restrictions. In *Proc. of IJCAI*, 96–103.

Baader, F.; Sertkaya, B.; and Turhan, A. 2007. Computing the least common subsumer w.r.t. a background terminology. *J. Applied Logic* 5(3):392–420.

Baader, F. 2003. Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In *Proc. of IJCAI*, 319–324.

Badea, L., and Nienhuys-Cheng, S. 2000. A refinement operator for description logics. In *Proc. of ILP*, 40–59.

Barceló, P., and Romero, M. 2017. The complexity of reverse engineering problems for conjunctive queries. In *Proc. of ICDT*, 7:1–7:17.

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.

Bühmann, L.; Lehmann, J.; Westphal, P.; and Bin, S. 2018. DL-learner - structured machine learning on semantic web data. In *Proc. of WWW*, 467–471.

Cohen, W. W.; Borgida, A.; and Hirsh, H. 1992. Computing least common subsumers in description logics. In *Proc. of AAAI*, 754–760.

Colucci, S.; Donini, F. M.; Giannini, S.; and Sciascio, E. D. 2016. Defining and computing least common subsumers in RDF. *J. Web Semant.* 39:62–80.

Donini, F. M.; Colucci, S.; Noia, T. D.; and Sciascio, E. D. 2009. A tableaux-based method for computing least common subsumers for expressive description logics. In *Proc. of IJCAI*, 739–745.

Eppe, M.; Maclean, E.; Confalonieri, R.; Kutz, O.; Schorlemmer, M.; Plaza, E.; and Kühnberger, K. 2018. A computational framework for conceptual blending. *Artif. Intell.* 256:105–129.

Fauconnier, G., and Turner, M. 2008. *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Funk, M.; Jung, J. C.; Lutz, C.; Pulcini, H.; and Wolter, F. 2019. Learning description logic concepts: When can positive and negative examples be separated. In *Proc. of IJCAI*.

Gutiérrez-Basulto, V.; Jung, J. C.; and Sabellek, L. 2018. Reverse engineering queries in ontology-enriched systems: The case of expressive Horn description logic ontologies. In *Proc. of IJCAI-ECAI*.

Harel, D.; Kupferman, O.; and Vardi, M. Y. 2002. On the complexity of verifying concurrent transition systems. *Inf. Comput.* 173(2):143–161.

Jung, J.; Lutz, C.; Martel, M.; and Schneider, T. 2017. Query conservative extensions in Horn description logics with inverse roles. In *Proc. of IJCAI-17*.

Küsters, R., and Borgida, A. 2001. What's in an attribute? consequences for the least common subsumer. *J. Artif. Intell. Res.* 14:167–203.

Küsters, R., and Molitor, R. 2001. Approximating most specific concepts in description logics with existential restrictions. In *Proc. of KI*, 33–47.

Lehmann, J., and Haase, C. 2009. Ideal downward refinement in the $\mathcal{EL}$ description logic. In *Proc. of ILP*, 73–87.

Lehmann, J., and Hitzler, P. 2010. Concept learning in description logics using refinement operators. *Machine Learning* 78:203–250.

Lisi, F. A. 2012. A formal characterization of concept learning in description logics. In *Proc. of DL*.

Muggleton, S. 1991. Inductive logic programming. *New Generation Comput.* 8(4):295–318.

Nebel, B. 1990. *Reasoning and Revision in Hybrid Representation Systems*. Springer.

Plotkin, G. 1970. A note on inductive generalizations. *Edinburgh University Press*.

Sarker, M. K., and Hitzler, P. 2019. Efficient concept induction for description logics. In *Proc. of AAAI*, 3036–3043.

ten Cate, B., and Dalmau, V. 2015. The product homomorphism problem and applications. In *Proc. of ICDT*, 161–176.

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

Zarrieß, B., and Turhan, A. 2013. Most specific generalizations w.r.t. general $\mathcal{EL}$-TBoxes. In *Proc. of IJCAI*, 1191–1197.

Notes for Section 3

For the convenience of the reader we formulate the model-theoretic characterizations also for the verification and existence of the LCS. We start with LCS verification. The following characterization follows from Theorems 1 and 3.

Theorem 16 (LCS Verification) *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{T}$ be an $\mathcal{L}$ TBox, $C_1, \ldots, C_n$ $\mathcal{L}$ concepts, and Σ a signature. An $\mathcal{L}(\Sigma)$ concept C is the $\mathcal{L}(\Sigma)$-LCS of $C_1, \ldots, C_n$ w.r.t. $\mathcal{T}$ iff the following conditions hold:*

1. $(\rho_{C_1}, \ldots, \rho_{C_n}) \in C^{\prod_{i=1}^n \mathcal{U}_{\mathcal{T},C_i}}$;
2. $\prod_{i=1}^n (\mathcal{U}_{\mathcal{T},C_i}, \rho_{C_i}) \preceq_{\mathcal{L},\Sigma} \mathcal{U}_{\mathcal{T},C}, \rho_C$.

For LCS existence, the following characterization follows from Theorems 1 and 4.

Theorem 17 (LCS Existence) *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{T}$ be an $\mathcal{L}$ TBox, $C_1, \ldots, C_n$ $\mathcal{L}$ concepts, and Σ a signature. The following are equivalent, for $D_k = (\Pi_{i=1}^n \mathcal{U}_{\mathcal{T},C_i}, \rho_{C_i})^{\downarrow \mathcal{L},k}_{|\Sigma}$:*

1. *the $\mathcal{L}(\Sigma)$-LCS of $C_1, \ldots, C_n$ w.r.t. $\mathcal{T}$ exists;*
2. *D_k is the $\mathcal{L}(\Sigma)$-LCS of $C_1, \ldots, C_n$ w.r.t. $\mathcal{T}$, for some $k \geq 0$;*
3. *$\Pi_{i=1}^n (\mathcal{U}_{\mathcal{T},C_i}, \rho_{C_i}) \preceq_{\mathcal{L},\Sigma} (\mathcal{U}_{\mathcal{T},D_k}, \rho_{D_k})$, for some $k \geq 0$.*

Proofs for Section 4

Theorem 6 *In $\mathcal{EL}$, MSC existence w.r.t. the empty TBox is* PSPACE*-complete. The lower bound applies even when the signature is full.*

Proof. We reduce the word problem for polynomially space bounded Turing machines (TMs), that is, given such a TM M with polynomial space bound $p(n)$, we construct an ABox $\mathcal{A}$ with individuals $a_1, \ldots, a_n$, such that the $\mathcal{EL}$ MSC of $a_1, \ldots, a_n$ w.r.t. $\mathcal{A}$ exists iff M accepts an input w. It is well-known that there is a deterministic polynomially space bounded TM whose halting problem is PSPACE-hard.

For our purposes, a Turing machine $M = (Q, \Gamma, q_0, \delta, F)$ consists of a set of states Q, finite set of tape symbols Γ, an initial state q_0, a set of final states F, and a (partial) transition function $\delta : Q \times \Gamma \to Q \times \Gamma \times \{L, R\}$. There, L and R correspond to the head moving to the left and to the right, respectively. We assume that M halts once it reaches a state $q \in F$, and always continues otherwise.

For the reduction, let $M = (Q, \Gamma, q_0, \delta, F)$ be a $p(n)$-space bounded deterministic TM, and w an input of length n. We construct an ABox $\mathcal{A}$ without any concept assertions. Let us first fix the following individuals:

$$\mathsf{ind}(\mathcal{A}) = \{(q, a, i), (a, i) \mid q \in Q, a \in \Gamma, 1 \leq i \leq p(n)\}.$$

Intuitively, an individual (q, a, i) represents that the content of cell i is a, that the head of the TM is on cell i and that the TM is in state q. Similarly, an individual (a, i) represents that the content of cell i is a (and that the head is not at position i). In the following description the cases $i = 1$ and $i = p(n)$ are not treated in a special way, since we can assume that M does not move its head beyond cell 1 or $p(n)$.

As role names, we use $r_{q,a,i}$, for all $(q, a, i) \in \mathsf{ind}(\mathcal{A})$. Informally, a role assertion $r_{q,a,i}(e, e')$ is included in $\mathcal{A}$ if in state q with the head at cell i and reading tape symbol a, M will change the tape cell represented by e to e'. Note that e and e' may be identical, meaning that the TM transition does not affect the tape cell.

Formally, we include the following role assertions for every $q \in Q$, $a \in \Gamma$, and $i \in \{1, \ldots, p(n)\}$ such that $\delta(q, a) = (q', b, D)$ is defined:

1. Role assertions that affect the direct environment of the head position i:

$$\begin{aligned} r_{q,a,i}((q, a, i), (b, i)) &\qquad \text{if } D = L \\ r_{q,a,i}((a', i-1), (q', a', i-1)) &\qquad \text{if } D = L \\ r_{q,a,i}((a', i+1), (a', i+1)) &\qquad \text{if } D = L \\ r_{q,a,i}((q, a, i), (b, i)) &\qquad \text{if } D = R \\ r_{q,a,i}((a', i-1), (a', i-1)) &\qquad \text{if } D = R \\ r_{q,a,i}((a', i+1), (q', a', i+1)) &\qquad \text{if } D = R \end{aligned}$$

2. Role assertions that do not affect the direct environment of the head position i:

$$r_{q,a,i}((b, j), (b, j)) \quad \text{for all } b \in \Gamma, j \notin \{i-1, i, i+1\}$$

This finishes the construction of $\mathcal{A}$. It remains to specify the individuals $a_1, \ldots, a_{p(n)}$ for the input $w = b_1 \cdots b_n$:

$$\begin{aligned} a_1 &= (q_0, b_1, 1) && \\ a_i &= (b_i, i) && \text{for all } i \in \{2, \ldots, n\} \\ a_i &= (\square, i) && \text{for all } i \in \{n+1, \ldots, p(n)\}, \end{aligned}$$

where $\square$ denotes the blank symbol.

Claim. M accepts w iff the $\mathcal{EL}$-MSC of $a_1, \ldots, a_n$ w.r.t. $\mathcal{A}$ exists.

Proof of the Claim. We provide some insight into the construction of $\mathcal{A}$. For this purpose, let us denote with $\mathcal{A}^{p(n)}$ the $p(n)$-fold product of $\mathcal{A}$. For a configuration α of M, let x_α denote the element of $\mathcal{A}^{p(n)}$ corresponding to this configuration in the natural way. The construction of $\mathcal{A}$ ensures the following:

($*$) if α' is a successor configuration of α then x_α has precisely one successor in $\mathcal{A}^{p(n)}$, namely $x_{\alpha'}$.

Thus, paths in $\mathcal{A}^{p(n)}$ starting in $(a_1, \ldots, a_{p(n)})$ directly correspond to computations of M on input w.

The claim now follows from ($*$) and the fact that the MSC exists iff all paths starting from $(a_1, \ldots, a_{p(n)})$ in $\mathcal{A}^{p(n)}$ are finite. This finishes the proof of the claim, and in fact of the Theorem. ❏

Theorem 5 *In $\mathcal{EL}$, LCS and MSC verification w.r.t. the empty TBox are* CONP*-complete. The lower bounds apply even when the signature is full.*

For the proof of Theorem 5 we require the following lemma. For a tree-shaped interpretation $\mathcal{J}$ and $e \in \Delta^{\mathcal{J}}$ we denote by $\mathcal{J}_e$ the subinterpretation of $\mathcal{J}$ induced by the subtree of $\mathcal{J}$ rooted at e.

Lemma 6 *Let $\mathcal{I}$ and $\mathcal{J}$ be interpretations with $\mathcal{J}$ tree-shaped. If $(\mathcal{I}, d) \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e)$, then there exists a set X with $d \in X \subseteq \Delta^{\mathcal{I}}$ such that $|X| \leq |\Delta^{\mathcal{J}_e}| + 1$ and $(\mathcal{I}_{|X}, d) \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e)$.*

Proof. The proof is by induction on the depth of $\mathcal{J}_e$. Assume first that $\mathcal{J}_e$ has depth 0. If there exists a concept name $A \in \Sigma$ with $d \in A^{\mathcal{I}}$ but $e \notin A^{\mathcal{J}}$, then $X = \{d\}$ is as required. Otherwise there exists a role name $r \in \Sigma$ and d' with $(d, d') \in r^{\mathcal{I}}$. Then $X = \{d, d'\}$ is as required. Now suppose that e has depth $k+1$ and the lemma has been proved for all e' with $\mathcal{J}_{e'}$ of depth $\leq k$. Assume $(\mathcal{I}, d) \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e)$. If there exists a concept name $A \in \Sigma$ with $d \in A^{\mathcal{I}}$ but $e \notin A^{\mathcal{J}}$, then $X = \{d\}$ is as required. Otherwise there exists a role name $r \in \Sigma$ and d' with $(d, d') \in r^{\mathcal{I}}$ such that for all e' with $(e, e') \in r^{\mathcal{J}}$, $(\mathcal{I}, d') \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e')$. Fix d'. By induction hypothesis, we can take for every e' with $(e, e') \in r^{\mathcal{J}}$ a set $X_{e'}$ with $d' \in X_{e'} \subseteq \Delta^{\mathcal{I}}$ such that $|X_{e'}| \leq |\Delta^{\mathcal{J}_{e'}}| + 1$ and $(\mathcal{I}_{|X_{e'}}, d') \not\preceq_{\mathcal{EL},\Sigma} (\mathcal{J}, e')$. Let X be the union of $\{d\}$ and the sets $X_{e'}$, $(e, e') \in r^{\mathcal{J}}$. Then X is as required. ❑

We now give the proof of Theorem 5.

Proof. By Theorem 1, it suffices to give the CONP upper bound for $\mathcal{EL}(\Sigma)$-MSC verification with empty TBox. Assume an ABox $\mathcal{A}$, $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$, a signature Σ, and an $\mathcal{EL}(\Sigma)$ concept C are given. It can be checked in PTIME where $\mathcal{A} \models C(a_i)$ for $i = 1, \ldots, n$. Thus it suffices to show that

$$\prod_{i=1}^{n} (\mathcal{A}, a_i) \not\preceq_{\mathcal{EL},\Sigma} \mathcal{A}_C, \rho_C$$

is in NP, where we regard $\mathcal{A}$, $\prod_{i=1}^{n} \mathcal{A}$, and $\mathcal{A}_C$ as interpretations in the obvious way. But this follows directly since

- by Lemma 6, if $\prod_{i=1}^{n} (\mathcal{A}, a_i) \not\preceq_{\mathcal{EL},\Sigma} \mathcal{A}_C, \rho_C$, then there exists a subset X of $\Delta^{\prod_{i=1}^{n} \mathcal{A}}$ with $|X| \leq |\mathsf{ind}(\mathcal{A}_C)| + 1$ such that $((\prod_{i=1}^{n} \mathcal{A})_{|X}, (a_1, \ldots, a_n)) \not\preceq_{\mathcal{EL},\Sigma} \mathcal{A}_C, \rho_C$;
- the simulation relation can be checked in polynomial time.

For the lower bound, we reduce SAT to LCS verification. Let φ be a formula in CNF that consists of m clauses with n variables $x_1, \ldots, x_n$. We will construct concepts $C_1, \ldots, C_n$ and a concept D such that the following are equivalent:

1. φ is unsatisfiable;
2. $(\prod_{i=1}^{n} \mathcal{A}_{C_i}, \rho_{C_i}) \preceq_{\mathcal{EL}} (\mathcal{A}_D, \rho_D)$;
3. the LCS of $C_1', \ldots, C_n'$ is $\exists s.D$, where $C_i' = \exists s.C_i \sqcap \exists s.D$ for $i = 1, \ldots, n$.

For better readability, we define concepts $C_1, \ldots, C_n$ in terms of interpretations $\mathcal{I}_1, \ldots, \mathcal{I}_n$ as follows:

- $\Delta^{\mathcal{I}_i'} = \{d_i, d_{i0}, d_{i1}\}$;
- $s^{\mathcal{I}_i'} = \{(d_i, d_{i0}), (d_i, d_{i1})\}$;
- $X_j^{\mathcal{I}_i'} = \overline{X}_j^{\mathcal{I}_i'} = \{d_{i0}, d_{i1}\}$, for all $j \neq i$;
- $X_i^{\mathcal{I}_i'} = \{d_{i1}\}$ and $\overline{X}_i^{\mathcal{I}_i'} = \{d_{i0}\}$.

Intuitively, each $\mathcal{I}_i$ has a root d_i with two successors d_{i0}, d_{i1} which "choose" a value for variable x_i. Note that every successor of the root of the product $\prod_{i=1}^{n} \mathcal{I}_i$ corresponds to a variable assignment. We define D via the interpretation $\mathcal{J}$ defined as follows:

- $\Delta^{\mathcal{J}} = \{d, 1, \ldots, m\}$;
- $s^{\mathcal{J}} = \{(d, j) \mid j \in \{1, \ldots, m\}\}$;
- $j \in X_i^{\mathcal{J}}$ iff x_i does not occur positively in clause j, for all $j \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, n\}$;
- $j \in \overline{X}_i^{\mathcal{J}}$ iff x_i does not occur negatively in clause j, for all $j \in \{1, \ldots, m\}$ and all $i \in \{1, \ldots, n\}$.

Note that every element j in $\mathcal{J}$ corresponds to clause j in φ and is labeled with all negated literals from the clause, that is, a successor of the root in $\prod_{i=1}^{n} \mathcal{I}_i$ maps to j iff the corresponding assignment makes the clause false. The concepts $C_1, \ldots, C_n, D$ thus satisfy the equivalence "1 $\Leftrightarrow$ 2" above. For the equivalence "2 $\Leftrightarrow$ 3", we use Theorem 3 (note that it applies to the LCS since the constructed ABoxes are essentially $\mathcal{EL}$ concepts). For "2 $\Rightarrow$ 3", note that

- $\exists s.D$ satisfies Condition 1 of Theorem 3, since it is a conjunct in every C_i';
- $\exists s.D$ satisfies Condition 2 of Theorem 3: first note that every s-successor in the product $\Pi_i \mathcal{I}_{C_i'}$ that involves D is trivially simulated by D; second note that the the s-successor corresponding to $\Pi_i \mathcal{I}$ is simulated by D by Point 2 above.

Conversely, that is, from "3 $\Rightarrow$ 2", Point 2 above follows from Condition 2 of Theorem 3. ❑

Theorem 7 *In $\mathcal{ELI}$, LCS and MSC existence and verification w.r.t. the empty TBox are* PSPACE*-hard and in* EXPTIME. *The lower bounds apply when the signature is full.*

The EXPTIME upper bound for MSC existence in $\mathcal{ELI}$ is established in the following.

Lemma 7 *In $\mathcal{ELI}$, MSC existence w.r.t the empty TBox is decidable in* EXPTIME.

Proof. Assume $\mathcal{K} = (\emptyset, \mathcal{A})$, $a_1, \ldots, a_n \in \mathsf{ind}(\mathcal{A})$, and a signature Σ are given. We show that the $\mathcal{ELI}(\Sigma)$-MSC of $a_1, \ldots, a_n$ w.r.t. $\mathcal{K}$ exists iff there is no infinite path $d_0, r_0, d_1, r_1, d_2, \ldots$ in $\mathcal{I} = \prod_{i=1}^{n} \mathcal{U}_{\mathcal{K}}$ satisfying

(†) $d_0 = (a_1, \ldots, a_n)$ and for all $i \geq 0$: $\mathsf{sig}(r_i) \subseteq \Sigma$ and

1. if $r_i = r_{i+1}^-$, then $(\mathcal{I}, d_{i+2}) \not\preceq_{\mathcal{ELI},\Sigma} (\mathcal{I}, d_i)$;
2. there is no $e \neq d_{i+1}$ such that $(d_i, e) \in r_i^{\mathcal{I}}$, $(\mathcal{I}, d_{i+1}) \preceq_{\mathcal{ELI},\Sigma} (\mathcal{I}, e)$, and $(\mathcal{I}, e) \not\preceq_{\mathcal{ELI},\Sigma} (\mathcal{I}, d_{i+1})$.

To prove this characterization, recall that by Theorem 4 the $\mathcal{ELI}(\Sigma)$-MCS of $a_1, \ldots, a_n$ with respect to $\mathcal{K}$ exists iff there exists $k \geq 0$ such that for $C_k = \Pi_{i=1}^{n} (\mathcal{U}_{\mathcal{K}}, a_i))_{|\Sigma}^{\downarrow \mathcal{ELI},k}$

$$(S_k) \qquad \Pi_{i=1}^{n} (\mathcal{U}_{\mathcal{K}}, a_i) \preceq_{\mathcal{ELI},\Sigma} (\mathcal{U}_{\emptyset,C_k}, \rho_{C_k}).$$

Assume first that there is an infinite path satisfying (†). Then clearly the path cannot be $\mathcal{ELI}(\Sigma)$-simulated by $(\mathcal{U}_{\emptyset,C_k}, \rho_{C_k})$ for any k because in any $\mathcal{U}_{\emptyset,C_k}$ the length of such paths starting at ρ_{C_k} does not exceed k. Conversely, assume there are no infinite paths satisfying (†). Then let k be the length of the longest path satisfying (†). It is readily shown that (S_k) holds, as required. As the universal model $\mathcal{U}_{\mathcal{K}}$ can be constructed in exponential time and the existence of infinite paths in $\prod_{i=1}^{n} \mathcal{U}_{\mathcal{K}}$ satisfying (†) can also be checked in exponential time, the existence of the $\mathcal{ELI}(\Sigma)$-MSC can be decided in EXPTIME. ❑

For the lower bound, we first prove lower bounds for the $\mathcal{ELI}$-product simulation problem for the case of tree-shaped interpretations.

Theorem 18 *The $\mathcal{ELI}$-product simulation problem on tree-shaped interpretations is* PSPACE*-hard.*

We reduce from a tiling problem where the input is a tiling system (T, H, V), an *initial tiling* $\theta = t_1, \ldots, t_n$ with tiles from T, and a *final tile* $t_F \in T$. The goal is to tile a finite rectangle of size $n \times m$, $m \geq 1$ arbitrary, such that the first row is tiled with θ and t_F occurs in the tiling. Formally, a *solution* to a *tiling instance* (T, H, V, θ, t_F), θ of length n, is a mapping $\tau : \{1, \ldots, m\} \times \{1, \ldots, n\} \rightarrow T$, for some $m \geq 2$, such that the following conditions are satisfied:

1. $(\tau(i, j), \tau(i, j + 1)) \in H$ for $1 \leq i \leq m$ and $1 \leq j < n$;
2. $(\tau(i, j), \tau(i + 1, j)) \in V$ for $1 \leq i < m$ and $1 \leq j \leq n$;
3. $(\tau(1, 1), \ldots, \tau(1, n)) = \theta$;
4. t_F is in the range of τ.

Let a tiling instance (T, H, V, θ, t_F) be given. We construct tree interpretations $\mathcal{I}_1, \ldots, \mathcal{I}_{3n}, \mathcal{M}$ such that for suitably chosen d_0 and e_0, $(\prod_{i=1}^{3n} \mathcal{I}_i, d_0) \preceq_{\mathcal{ELI}} (\mathcal{M}, e_0)$ iff there is no solution.

We use the following signature:

1. a single role name r;
2. concept names $T^i_{t,j}$, $t \in T$, $i \in \{1, \ldots, n\}$, $j \in \{1, 2\}$ to express that position i is tiled with t;
3. concept names $M_1, \ldots, M_5$ representing different 'phases' we go through when following a path through the product;
4. concept names M_{ij} with $1 \leq i, j \leq 5$ and $j \in \{i - 1, i + 1\}$, representing transitions between these phases.

Let T_H be the set of triples (t_1, t_2, t_3) with $(t_1, t_2) \in H$ and $(t_2, t_3) \in H$. For every $\tau = (t_1, t_2, t_3) \in T_H$, $i \in \{1, \ldots, n\}$, and $\ell \in \{1, 2\}$, let $S^i_{\tau,\ell}$ be the set of concept names that contains $T^{i-1}_{t_1,\ell}$, $T^i_{t_2,\ell}$, $T^{i+1}_{t_3,\ell}$, and $T^j_{t,\ell}$ for every $j \in \{1, \ldots, i-2, i+2, \ldots, n\}$ and $t \in T$. We admit superscripts -1 and $n + 1$ for uniformity purposes.

The interpretations $\mathcal{I}_i$, $1 \leq i \leq n$ are defined as follows:

- $\mathcal{I}_i$ is a tree of depth two that branches only at the root d^i_0;
- for all $\tau \in T_H$, d^i_0 has an r^--successor e^i_τ which has an r-successor d^i_τ;
- d^i_0 has further r^--successors d^i_{34}, d^i_{43}, d^i_{45},d^i_{54};
- d^i_0 is labeled with M_3, M_4, M_5 and with $T^i_{t,\ell}$ for every $i \in \{1, \ldots, n\}$, $t \in T$, and $\ell \in \{1, 2\}$;
- each d^i_τ is labeled with M_1, M_2, with every concept name from $S^i_{\tau,1}$ and with every concept name $T^i_{t,2}$, $i \in \{1, \ldots, n\}$ and $t \in T$;
- each e^i_τ is labeled with M_{12}, M_{21}, M_{23}, and M_{32};
- each d^i_{jk} is labeled with M_{jk}.

The interpretations $\mathcal{I}_i$, $n + 1 \leq i \leq 2n$ are defined as follows:

- $\mathcal{I}_i$ is a tree of depth two that branches only at the root d^i_0;
- for all $\tau_1, \tau_2 \in T_H$, d^i_0 has an r^--successor $e^i_{\tau_1,\tau_2}$ which has an r-successor $d^i_{\tau_1,\tau_2}$;
- d^i_0 is labeled with M_1, M_5 and with $T^i_{t,\ell}$ for every $i \in \{1, \ldots, n\}$, $t \in T$, and $\ell \in \{1, 2\}$;
- each $d^i_{\tau_1,\tau_2}$ is labeled with M_2, M_3, M_4, with every concept name from $S^i_{\tau_1,1}$ and from $S^i_{\tau_2,2}$;
- each $e^i_{\tau_1,\tau_2}$ is labeled with all concept names M_{jk}.

The interpretations $\mathcal{I}_i$, $2n + 1 \leq i \leq 3n$ are defined as follows:

- $\mathcal{I}_i$ is a tree of depth two that branches only at the root d^i_0;
- for all $\tau \in T_H$, d^i_0 has an r^--successor e^i_τ which has an r-successor d^i_τ;
- d^i_0 has further r^--successors d^i_{12}, d^i_{21}, d^i_{23}, d^i_{32};
- d^i_0 is labeled with M_1, M_2, M_3 and with $T^i_{t,\ell}$ for every $i \in \{1, \ldots, n\}$, $t \in T$, and $\ell \in \{1, 2\}$;
- each d^i_τ is labeled with M_4, M_5, with every concept name from $S^i_{\tau,2}$ and with every concept name $T^i_{t,1}$, $i \in \{1, \ldots, n\}$ and $t \in T$.
- each e^i_τ is labeled with M_{34}, M_{43}, M_{45}, and M_{54};
- each d^i_{jk} is labeled with M_{jk}.

We are mainly interested in paths through $\prod_{i=1}^{3n} \mathcal{I}_i$ that are marked with the following pattern:

$$\begin{aligned} &M_1, r^-, M_{12}, r, \\ &M_2, r^-, M_{23}, r, \\ &M_3, r^-, M_{34}, r, \\ &M_4, r^-, M_{45}, r, \\ &M_5, r^-, M_{54}, r, \ldots \\ &M_4, r^-, M_{43}, r, \\ &M_3, r^-, M_{32}, r, \\ &M_2, r^-, M_{21}, r, \\ &M_1, r^-, M_{12}, r, M_2, \end{aligned} \tag{$*$}$$

We give an informal description of how the mentioned paths are related to rectangle tilings. Note first that, if an element of $\prod_{i=1}^{3n} \mathcal{I}_i$ satisfies some M_i, then this has implications regarding its components. For instance, if an element $(d_1, \ldots, d_{3n})$ satisfies M_3, then $d_1, \ldots, d_n$ are all roots of their respective interpretations and so are $d_{2n+1}, \ldots, d_{3n}$, while $d_{n+1}, \ldots, d_{2n}$ are leaves. We sketch how to obtain a path through $\prod_{i=1}^{3n} \mathcal{I}_i$ that follows pattern $(*)$ and represents any rectangle tiling. Let $\theta_1, \theta_2, \ldots$ be an enumeration of the rows of some the tiling.

- In $\prod_{i=1}^{n} \mathcal{I}_i$, start at those d^i_τ that represent θ_1 by concept names $T^i_{t,1}$. In $\prod_{i=n+1}^{3n} \mathcal{I}_i$, we start at $d^{n+1}_0, \ldots, d^{3n}_0$. This point in the product is labeled M_1.

- Then proceed via an element labeled M_{12} to an element labeled M_2 that represents θ_2 in the $T^i_{t,2}$ and θ_1 in the $T^i_{t,1}$. The choice is in components $\prod_{i=n+1}^{2n} \mathcal{I}_i$ and we remain stationary in $\prod_{i=1}^{n} \mathcal{I}_i$ and in $\prod_{i=2n+11}^{3n} \mathcal{I}_i$.
- Next proceed to the roots in $\prod_{i=1}^{n} \mathcal{I}_i$, remaining stationary in $\prod_{i=n+1}^{3n} \mathcal{I}_i$ (label M_3, via M_{23}). We still represent θ_1 and θ_2 as before. As explained later, this transition serves to verify the vertical matching condition.
- Next proceed to leaves in $\prod_{i=2n+1}^{3n} \mathcal{I}_i$, remaining stationary in $\prod_{i=1}^{2n} \mathcal{I}_i$ (label M_4, via M_{34}). Once more, θ_1 and θ_2 are represented as before. This transition serves no purpose as we move 'upwards' (towards higher indexes) in the $M_1, \ldots, M_5$ sequence. When moving downwards, this transition checks the vertical matching condition while the transition in the previous item serves no purpose.
- Then proceed to the roots in $\prod_{i=n+1}^{2n} \mathcal{I}_i$, remaining stationary in all other components (label M_5, via M_{45}); this preserves the representation of θ_2 via $T^i_{t,2}$, but 'forgets' the representation of θ_1 via $T^i_{t,1}$.
- Now do everything backwards, from M_5 towards M_1; first proceed via an element labeled M_{54} to an element labeled M_4 that represents θ_3 in the $T^i_{t,1}$ and θ_2 in the $T^i_{t,2}$. The choice is in components $\prod_{i=n+1}^{2n} \mathcal{I}_i$ and we remain stationary in all other components; then move to lavel M_3 via M_{43}, and so on.
- After reaching M_1, proceed again in the forward direction, representing θ_4, and so forth.

Of course, there are paths through the product that do not follow this ideal pattern, for different reasons. For instance, the desired sequence of the M_i is not followed, some element does not correspond to a valid row in the tiling, or the vertical matching condition is not met. These undesired paths are captured by 'traps' in the interpretation $\mathcal{M}$ that we construct next.

We assemble $\mathcal{M}$ by starting with a path of length nine, connected by alternating between r^- and r:

$$e_1\, r^-\, e_{12}\, r\, e_2\, r^-\, e_{23}\, r\, e_3\, r^-\, e_{34}\, r\, e_4\, r^-\, e_{45}\, r\, e_5$$

such that each e_i is labeled with M_i and with all concept names $T^i_{t,\ell}$, $1 \leq i \leq n$, $\ell \in \{1,2\}$, and $t \in T \setminus \{t_F\}$; the missing t_F means that any 'proper' path reaching t_F will result in non-simulation. Also, each e_{ij} is labeled with M_{ij} and M_{ji}.

We now add traps to make sure that undesired paths are simulated by $\mathcal{M}$. We start with the case that the desired sequence of the M_i is not followed:

1. To every e_{ij}, we attach an r-successor that is labeled with M_k for every $k \notin \{i,j\}$ and with all concept names $T^{i'}_{t,\ell}$, and that has an r^--successor which has an r-successor that makes true *all* concept names (including all $T^i_{t_F,\ell}$), acting as a well of positivity.
2. To every e_i, we attach an r^--successor that is labeled with M_{jk} whenever $jk \notin \{ii-1, ii+1\}$, and that has an r-successor which is a well of positivity.

Next, we add traps that address defects which concern a single row of the tiling:

3. To each e_{ij}, we attach an r-successor for each $k \in \{1,\ldots,n\}$ and $\ell \in \{1,2\}$. No concept name $T^k_{t,\ell}$ is true there, but all concept names $T^j_{t,m}$ with $(j,m) \neq (k,\ell)$ and $t \in T$, and of course there is a well below it. This has two effects:
   (a) it enforces synchronization of the tiling of each i-th column accross the $\prod_{i=1}^{n} \mathcal{I}_i$ resp. $\prod_{i=n+1}^{2n} \mathcal{I}_i$ resp. $\prod_{i=2n+1}^{3n} \mathcal{I}_i$; so the horizontal tiling condition is satisfied;
   (b) in M_2 and M_4-configurations, it enforces that the tiling of the row that is represented twice, once in $\prod_{i=1}^{n} \mathcal{I}_i$ and once in $\prod_{i=n+1}^{2n} \mathcal{I}_i$, resp. once in $\prod_{i=n+1}^{2n} \mathcal{I}_i$ and once in $\prod_{i=2n+1}^{3n} \mathcal{I}_i$, is identical.

It remains to deal with paths that violate the vertical matching condition:

4. at e_{23}, we attach a trap for each $(t_1,t_2) \notin V$ and each $i \in \{1,\ldots,n\}$, labeled with M_2, with $T^i_{t_1,1}$ and $T^i_{t_2,2}$, and with $T^j_{t,\ell}$ for all $j \neq i$, $\ell \in \{1,2\}$, and $t \in T$;
5. at e_{34}, we attach a trap for each $(t_2,t_1) \notin V$ and each $i \in \{1,\ldots,n\}$, labeled with M_4, with $T^i_{t_2,2}$ and $T^i_{t_1,1}$, and with $T^j_{t,\ell}$ for all $j \neq i$, $\ell \in \{1,2\}$, and $t \in T$.

The initial tiling $\theta = t_1,\ldots,t_n$ gives rise to a sequence of triples $\tau_1,\ldots,\tau_n$ in the obvious way. We are going to use

$$d_0 = (d^1_{\tau_1},\ldots,d^n_{\tau_n},d^{n+1}_0,\ldots,d^{3n}_0).$$

as the starting point for the simulation.

Lemma 8 $(\prod_{i=1}^{3n} \mathcal{I}_i, d_0) \preceq_{\mathcal{ELI}} (\mathcal{M}, e_1)$ *iff there is no solution for* (T,H,V,θ_1,t_F).

Proof. "if". Assume that there is no solution for (T,H,V,θ,t_F). We prove the existence of a simulation from $(\prod_{i=1}^{3n} \mathcal{I}_i, d_0)$ to $(\mathcal{M}, e_1)$.

Let us first introduce some notation. We call a tuple $t_1,\ldots,t_n \in T$ *possible* if there is a mapping τ with $\tau(i,1) = t_1,\ldots,\tau(i,n) = t_n$, for some i and which satisfies Condition 1–3 of a solution (but does not necessarily mention t_F). As there is no solution for (T,H,V,θ,t_F), no tuple that is possible mentions t_F.

Now, we say that d is *k-proper* if d satisfies M_k and

- if $k=1$, then the $T^i_{t,1}$ represent a possible row at d;
- if $k=2$, then the $T^i_{t,1}$ represent a possible row, and the $T^i_{t,2}$ represent some row satisfying the horizontal tiling condition;
- if $k=3$, then the $T^i_{t,\ell}$ represent possible rows, for $\ell \in \{1,2\}$;
- if $k=4$, then the $T^i_{t,2}$ represent a possible row, and the $T^i_{t,1}$ represent some row satisfying the horizontal tiling condition;
- if $k=5$, then the $T^i_{t,2}$ represent a possible row.

We claim that there is a simulation S from $(\prod_{i=1}^{3n} \mathcal{I}_i, d_0)$ to $(\mathcal{M}, e_1)$ which relates all k-proper elements in $\prod_{i=1}^{3n} \mathcal{I}_i$ with e_k. The statement then follows, because the initial tuple d_0 is 1-proper by construction.

We show the arguments only for $k = 2$ because the other cases are similar. Thus, take any d that is 2-proper, and assume $(d, e_2) \in S$. We show how to continue the simulation from (d, e_2). To this end, let d' be an r^--successor of d. We distinguish several cases that can arise by the construction of the $\mathcal{I}_i$:

- if d' does not satisfy one of M_{21} or M_{23}, then d' is simulated by the trap of type 2 at e_2.
- if d' satisfies M_{21}, then we add (d', e_{12}) to the simulation. Now, let d'' be any r-successor of d'. Since d' satisfies M_{21}, we have to be in d_{21}^i for all $i \in \{2n+1, \ldots, 3n\}$ and thus d'' satisfies one of M_1, M_2, M_3 or no M_i at all. We again distinguish cases:
  - if d'' satisfies M_3 or no M_i at all, then d'' is simulated by a trap of type 1 at e_{12};
  - if d'' satisfies M_2, then, by construction of the $\mathcal{I}_i$, d'' is actually d and it is simulated by e_2;
  - if d'' satisfies M_1, the construction of the $\mathcal{I}_i$ implies that the rows represented by $T_{t,1}^i$, at d'' are the same as these rows at represented by $T_{t,1}^i$ at d. Since the latter is possible by assumption, so is the former. Thus, d'' is 1-proper and we know that $(d'', e_1) \in S$.
- if d' satisfies M_{23}, then we add (d', e_{23}) to the simulation and continue as in the previous case. More precisely, let d'' be any r-successor of d'. Since d' satisfies M_{23}, we have to be in d_{23}^i for all $i \in \{2n+1, \ldots, 3n\}$ and thus d'' satisfies one of M_1, M_2, M_3 or no M_i at all. We again distinguish cases:
  - if d'' satisfies M_1 or no M_i at all, then d'' is simulated by a trap of type 1 at e_{23};
  - if d'' satisfies M_2, then, by construction of the $\mathcal{I}_i$, d'' is actually d and it is simulated by e_2;
  - if d'' satisfies M_3, the construction of the $\mathcal{I}_i$ implies that the rows represented by $T_{t,\ell}^i$, $\ell \in \{1, 2\}$ at d'' are the same as these rows at represented by $T_{t,\ell}^i$, $\ell \in \{1, 2\}$ at d. Let $t_1, \ldots t_n$ and $t'_1, \ldots, t'_n$ be the rows represented by $T_{t,1}^i$ and $T_{t,2}^i$, respectively. If $(t_j, t'_j) \notin V$, for some j, then d'' is simulated by a trap of type 4 at e_{23}. Otherwise, the row represented by the $T_{t,2}^i$ at d'' is a valid successor of the row represented by the $T_{t,1}^i$ at d. Overall, d'' is 3-proper and we know that $(d'', e_3) \in S$.

"only if". Assume that there is a solution τ for (T, H, V, θ, t_F), and let t_F occur in the last row. Let $d_0, r^-, d'_0, r, d_1, \ldots, d_n$ be the path which follows the pattern $(*)$ that is contained in the product and which reflects the solution τ. (We can obtain this path by letting τ guide the selection of successors as described above). By construction, the path satisfies the following properties, for all $i \geq 0$:

- if d_i satisfies M_1, then the $T_{t,1}^i$ represent a row of τ;
- if d_i satisfies M_2, then the $T_{t,\ell}^i$ represent rows θ_ℓ, for $\ell \in \{1, 2\}$, respectively, and
  - if d_{i-1} satisfies M_1, θ_2 is a successor row of θ_1 in τ and θ_1 is represented by the $T_{t,1}^i$ at d_{i-1}, and
  - if d_{i-1} satisfies M_3, θ_1 is a successor row of θ_2 in τ and θ_2 is represented by the $T_{t,2}^i$ at d_{i-1};
- if d_i satisfies M_3, then the $T_{t,1}^i$ and the $T_{t,2}^i$ represent rows of τ, and in fact the same rows as the $T_{t,1}^i$ and $T_{t,2}^i$ at d_{i-1};
- if d_i satisfies M_4, then the $T_{t,\ell}^i$ represent rows θ_ℓ, for $\ell \in \{1, 2\}$, respectively, and:
  - if d_{i-1} satisfies M_3, then θ_1 and θ_2 are also represented by the $T_{t,1}^i$ and $T_{t,2}^i$ at d_{i-1};
  - if d_{i-1} satisfies M_5, then θ_1 is a successor row of θ_2 in τ, and θ_2 is represented by the $T_{t,2}^i$ at d_{i-1};
- if d_i satisfies M_5, then the $T_{t,2}^i$ represent a row of τ, and the $T_{t,2}^i$ represent the same row at d_{i-1};
- d'_i satisfies M_{jk} such that d_i satisfies M_j and d_{i+1} satisfies M_k or d_{i+1} satisfies M_j and d_i satisfies M_k.

In order to show that there is no simulation, we show that:

(†) if d_i satisfies M_k it can only be simulated by e_k.

Note that this is a contradiction since d_n satisfies $T_{t_F,\ell}^i$ for some i, ℓ, but none of the e_i does.

We argue inductively. The induction base is given by the fact that d_0 has to be simulated by e_0 in the lemma. In the induction step, we suppose that d_i satisfies some M_k and show that d_{i+1} can only be simulated by e_{k-1} or e_{k+1}, respectively, depending on whether d_{i+1} satisfies M_{k-1} or M_{k+1}.

We show how to argue for $k = 2$, assuming that we are in the downward phase of the construction of the path, that is, d_{i+1} will be labeled with M_3. Based on the invariants given above, it can be verified that $d_{i'}$ and d_{i+1} cannot be simulated by any trap, and thus have to be simulated by e_{23} and e_3, respectively. ❏

To show PSPACE hardness of LCS and MSC verification and existence we first observe that the tiling problem used in the proof of Theorem 18 is still PSPACE hard if one only considers tiling instances (T, H, V, θ, t_F) such that the initial tiling θ only occurs in the first row for any mapping $\tau : \{1, \ldots, m\} \times \{1, \ldots, n\} \to T$, $m \geq 2$, such that the first three conditions for a solution are satisfied:

1. $(\tau(i, j), \tau(i, j+1)) \in H$ for $1 \leq i \leq m$ and $1 \leq j < n$;
2. $(\tau(i, j), \tau(i+1, j)) \in V$ for $1 \leq i < m$ and $1 \leq j \leq n$;
3. $(\tau(1, 1), \ldots, \tau(1, n)) = \theta$.

If this is the case then the pair $(\prod_{i=1}^{3n} \mathcal{I}_i, d_0)$, $(\mathcal{M}, e_1)$ constructed above has the following property which we require in the reduction to LCS and MSC verification and existence. Let $\mathcal{I}$ and $\mathcal{J}$ be interpretations and $d \in \Delta^{\mathcal{I}}$. An $\mathcal{ELI}$ simulation S from $\mathcal{I}$ to $\mathcal{J}$ is called *d-injective* if there exists exactly one $e \in \Delta^{\mathcal{J}}$ with $(d, e) \in S$. We write $(\mathcal{I}, d) \preceq_{\mathcal{ELI}}^{d\text{-inj}} (\mathcal{J}, e)$ if there exists a d-injective $\mathcal{ELI}$ simulation S from $\mathcal{I}$ to $\mathcal{J}$ that contains (d, e). We say that a pair $(\mathcal{I}, d)$, $(\mathcal{J}, e)$ is *oblivious* *to* d*-injectivity* if $(\mathcal{I}, d) \preceq^{d\text{-inj}}_{\mathcal{ELI}} (\mathcal{J}, e)$ iff $(\mathcal{I}, d) \preceq_{\mathcal{ELI}} (\mathcal{J}, e)$. It is easy to show the following.

Lemma 9 *The pair* $(\prod_{i=1}^{3n} \mathcal{I}_i, d_0)$, $(\mathcal{M}, e_1)$ *is oblivious to* d*-injectivity if the input tiling instance* (T, H, V, θ_1, t_F) *is such that the initial tiling* θ_1 *can only occur in the first row for any mapping* τ *satisfying the first three conditions of solutions.*

Now let $C_1, \ldots, C_n$ and D be $\mathcal{ELI}$ concepts and assume that $n \geq 2$. Consider the concepts D_1, D_2 constructed in Example 3 and let

$$D_3 = A_1 \sqcap A_2 \sqcap \exists r.(A_3 \sqcap A_4),$$

where we assume that the signature of D_1, D_2 is disjoint from the signature of $C_1, \ldots, C_n$. Set $C_i' = C_i \sqcap D_1$ if i is even, $C_i' = C_1 \sqcap D_2$ if i is odd, and let $D' = D \sqcap D_3$. Let v be a fresh role name and $E_i = \exists v.C_i' \sqcap \exists v.D'$ for $i = 1, \ldots, n$. PSPACE-hardness of LCS and MSC verification and existence now follow directly from the following reduction.

Lemma 10 *Assume that* $\prod_{i=1}^{n}(\mathcal{U}_{C_i}, \rho_{C_i})$, $(\mathcal{U}_D, \rho_D)$ *is oblivious to* $(\rho_{C_1}, \ldots, \rho_{C_n})$*-injectivity. Then the following conditions are equivalent:*

1. $\prod_{i=1}^{n}(\mathcal{U}_{C_i}, \rho_{C_i}) \preceq_{\mathcal{ELI}} (\mathcal{U}_D, \rho_D)$;
2. $\exists v.D'$ *is the* $\mathcal{ELI}$*-LCS of* $E_1, \ldots, E_n$;
3. *The* $\mathcal{ELI}$*-LCS of* $E_1, \ldots, E_n$ *exists.*

Proof. "$1 \Rightarrow 2$". By Theorem 3, it suffices to show

1. $(\rho_{E_1}, \ldots, \rho_{E_n}) \in (\exists v.D')^{\Pi_{i=1}^{n}\mathcal{U}_{E_i}}$;
2. $\Pi_{i=1}^{n}(\mathcal{U}_{E_i}, \rho_{E_i}) \preceq_{\mathcal{ELI}} (\mathcal{U}_{\exists v.D'}, \rho_{\exists v.D'})$.

Condition 1 follows directly from the construction. For Condition 2, we construct for every v successor $\vec{d} = (d_1, \ldots, d_n)$ of $(\rho_{E_1}, \ldots, \rho_{E_n})$ a $\vec{d}$-injective $\mathcal{ELI}$ simulation $S_{\vec{d}}$ between $\Pi_{i=1}^{n}(\mathcal{U}_{d_i}, d_i)$ and $(\mathcal{U}_{D'}, \rho_{D'})$. Then we are done: let V be the set of v-successors of $(\rho_{E_1}, \ldots, \rho_{E_n})$. Then

$$S = \{((\rho_{E_1}, \ldots, \rho_{E_n}), \rho_{\exists v.D'})\} \cup \bigcup_{\vec{d} \in V} S_{\vec{d}}$$

witnesses $\Pi_{i=1}^{n}(\mathcal{U}_{E_i}, \rho_{E_i}) \preceq_{\mathcal{ELI}} (\mathcal{U}_{\exists v.D'}, \rho_{\exists v.D'})$, as required. Assume $\vec{d} = (d_1, \ldots, d_n)$ is given.

Case 1. $(d_1, \ldots, d_n) = (\rho_{C_1'}, \ldots, \rho_{C_n'})$. To construct a $(\rho_{C_1'}, \ldots, \rho_{C_n'})$-injective $\mathcal{ELI}$ simulation let S_0 be a $(\rho_{C_1}, \ldots, \rho_{C_n})$-injective $\mathcal{ELI}$ simulation between $\Pi_{i=1}^{n}(\mathcal{U}_{C_i}, \rho_{C_i})$ and $(\mathcal{U}_D, \rho_D)$. It exists since $\prod_{i=1}^{n}(\mathcal{U}_{C_i}, \rho_{C_i})$, $(\mathcal{U}_D, \rho_D)$ is oblivious to $(\rho_{C_1}, \ldots, \rho_{C_n})$-injectivity. To define the simulating nodes for the remaining elements of $\Pi_{i=1}^{n}\mathcal{U}_{C_i \sqcap D_i'}$ (where we set $D_i' = D_1$ if i is even and $D_i' = D_2$ if i is odd) simply choose the projection to a fixed component D_i' composed with the obvious $\rho_{D_i'}$-injective $\mathcal{ELI}$ simulation between $(\mathcal{U}_{D_i'}, \rho_{D_i'})$ and $(\mathcal{U}_{D_3}, \rho_{D_3})$. The resulting relation is a $(\rho_{C_1'}, \ldots, \rho_{C_n'})$-injective $\mathcal{ELI}$ simulation.

Case 2. there exists d_i with $d_i = \rho_{D \sqcap D_3}$. Then we take the projection to the component $\mathcal{U}_{D \sqcap D_3}$ and obtain a $(d_1, \ldots, d_n)$-injective $\mathcal{ELI}$ simulation.

"$2 \Rightarrow 3$" is trivial.

"$3 \Rightarrow 1$". We use Theorem 4. There exists $k \geq 0$ such that for $G_k = \Pi_{i=1}^{n}(\mathcal{U}_{E_i}, \rho_{E_i})^{\downarrow \mathcal{ELI}, k}$

$$\Pi_{i=1}^{n}(\mathcal{U}_{E_i}, \rho_{E_i}) \preceq_{\mathcal{ELI}} (\mathcal{U}_{G_k}, \rho_{G_k}).$$

Consider the v successor $(\rho_{C_1'}, \ldots, \rho_{C_n'})$ of $(\rho_{E_1}, \ldots, \rho_{E_n})$ in $\Pi_{i=1}^{n}\mathcal{U}_{E_i}$. There exists a v-successor $(d_1, \ldots, d_n)$ of ρ_{G_k} in $\mathcal{U}_{G_k}$ such that

$$\prod_{i=1}^{n}(\mathcal{U}_{C_i'}, \rho_{C_i'}) \preceq_{\mathcal{ELI}} (\mathcal{U}_{G_k}, (d_1, \ldots, d_n))$$

Note that $(d_1, \ldots, d_n) \neq (\rho_{C_1'}, \ldots, \rho_{C_n'})$ because $n \geq 2$ and the chain constructed in Example 3 shows that there is no $\mathcal{ELI}$ simulation between $\prod_{i=1}^{n}(\mathcal{U}_{D_i'}, \rho_{D_i'})$ and any $(\mathcal{U}_{G_k}, (\rho_{C_1'}, \ldots, \rho_{C_n'}))$, where $D_i' = D_1$ if i is even and $D_i' = D_2$ if i is odd. It follows that there exists $i \leq n$ such that $d_i = \rho_{D \sqcap D_3}$. Then

$$(\mathcal{U}_{G_k}, (d_1, \ldots, d_n)) \preceq_{\mathcal{ELI}} (\mathcal{U}_{D \sqcap D_3}, \rho_{D \sqcap D_3})$$

as the projections onto components are $\mathcal{ELI}$ simulations. By taking the composition of $\mathcal{ELI}$ simulations we obtain

$$\prod_{i=1}^{n}(\mathcal{U}_{C_i'}, \rho_{C_i'}) \preceq_{\mathcal{ELI}} (\mathcal{U}_{D \sqcap D_3}, \rho_{D \sqcap D_3})$$

Condition 1 follows by construction. ❑

Proofs for Section 5

Theorem 8 *For* $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$*, the complement of* $\mathcal{L}$ *concept separability can be reduced in polynomial time to* $\mathcal{L}$*-MSC verification and existence. This also holds for the full signature.*

Proof. Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. Similarly to the proof of Theorem 2, one can show that it suffices to provide a reduction of the following problem: given an $\mathcal{L}$ TBox $\mathcal{T}$, an ABox $\mathcal{A}$ with assertions $A_1(a_1), \ldots, A_n(a_n), B(b)$, where $A_1, \ldots, A_n, B$ are concept names, and a signature Σ containing B, is it the case that

$$\prod_{i=1}^{n}(\mathcal{U}_{\mathcal{K}}, a_i) \preceq_{\mathcal{L}, \Sigma} (\mathcal{U}_{\mathcal{K}}, b),$$

where $\mathcal{K} = (\mathcal{T}, \mathcal{A})$? We start with the reduction for $\mathcal{EL}$.

Assume an $\mathcal{EL}$ TBox $\mathcal{T}$ and an ABox $\mathcal{A}$ with assertions $A_1(a_1), \ldots, A_n(a_n), B(b)$, where $A_1, \ldots, A_n, B$ are concept names, are given. We may assume that $n \geq 2$ and all a_i, $i = 1, \ldots, n$, and b are distinct. Define the *relativisation* $C_{|E}$ of a concept C to a concept name E inductively as follows:

$$\begin{aligned}
\top_{|E} &= E \\
A_{|E} &= E \sqcap A \\
(C \sqcap D)_{|E} &= C_{|E} \sqcap D_{|E} \\
(\exists r.C)_{|E} &= E \sqcap \exists r.C_{|E}
\end{aligned}$$

The *relativisation* $\mathcal{T}_{|E}$ of a TBox $\mathcal{T}$ to a concept name E is defined by setting

$$\mathcal{T}_{|E} = \{C_{|E} \sqsubseteq D_{|E} \mid C \sqsubseteq D \in \mathcal{T}\}$$

Construct a new TBox

$$\mathcal{T}' = \mathcal{T}_{|B_1} \cup \mathcal{T}_{|B_2} \cup \mathcal{T}_{|B_3} \cup \{B_i \sqsubseteq \exists w.B_i \mid i = 1, 2, 3\},$$

where w is a fresh role name and B_1, B_2, B_3 are fresh concept names. Take a fresh role name v and construct an ABox $\mathcal{A}'$ with individuals $\rho_i, a_i, b_i, i = 1, \ldots, n$, and assertions

- $v(\rho_i, a_i), v(\rho_i, b_i), i = 1, \ldots, n$;
- $A_i(a_i), B(b_i), i = 1, \ldots, n$;
- $B_1(a_i)$ if i is even;
- $B_2(a_i)$ if i is odd;
- $B_3(b_i)$ for $i = 1, \ldots, n$.

Let $C_0 = \exists v.(B \sqcap B_3)$, $\mathcal{K}' = (\mathcal{T}', \mathcal{A}')$, and $\Sigma' = \Sigma \cup \{B_1, B_2, B_3, v, w\}$. We show that the following conditions are equivalent:

(a) $\prod_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i) \preceq_{\mathcal{EL},\Sigma} (\mathcal{U}_\mathcal{K}, b)$;

(b) C_0 is the $\mathcal{EL}(\Sigma')$-MSC of $\rho_1, \ldots, \rho_n$ w.r.t. $\mathcal{K}'$;

(c) The $\mathcal{EL}(\Sigma')$-MSC of $\rho_1, \ldots, \rho_n$ w.r.t. $\mathcal{K}'$ exists.

(a) $\Rightarrow$ (b). Assume (a) holds. By Theorem 3, it suffices to show

1. $(\rho_1, \ldots, \rho_n) \in C_0^{\Pi_{i=1}^n \mathcal{U}_{\mathcal{K}'}}$;
2. $\Pi_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, \rho_i) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',C_0}, \rho_{C_0})$.

Condition 1 follows from $((\rho_1, \ldots, \rho_n), (b_1, \ldots, b_n)) \in v^{\Pi_{i=1}^n \mathcal{U}_{\mathcal{K}'}}$ and

$$(b_1, \ldots, b_n) \in (B \sqcap B_3)^{\Pi_{i=1}^n \mathcal{U}_{\mathcal{K}'}}$$

since $C_0 = \exists v.(B \sqcap B_3)$. For Condition 2, it follows from the construction of $\mathcal{K}'$ that is suffices to show that

$$\Pi_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, d_i) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',(B \sqcap B_3)}, \rho_{B \sqcap B_3})$$

for every v-successor $(d_1, \ldots, d_n)$ of $(\rho_1, \ldots, \rho_n)$ in $\Pi_{i=1}^n \mathcal{U}_{\mathcal{K}'}$. If there exists $i \leq n$ with $d_i = b_i$ then this follows from the fact that projections are simulations and the construction of $\mathcal{K}'$. Otherwise $d_i = a_i$ for all $i = 1, \ldots, n$. Denote by

- $\mathcal{U}'_{a_i}$ the tree-shaped subinterpretation rooted at a_i in $\mathcal{U}_{\mathcal{K}'}$;
- $\mathcal{U}_{a_i}$ the tree-shaped subinterpretation rooted at a_i in $\mathcal{U}_\mathcal{K}$;
- $\mathcal{U}'_{B \sqcap B_3}$ the tree-shaped subinterpretation rooted at $\rho_{B \sqcap B_3}$ in $\mathcal{U}_{\mathcal{T}',C_0}$;
- $\mathcal{U}_B$ the tree-shaped subinterpretation rooted at b in $\mathcal{U}_\mathcal{K}$.

Observe that $\mathcal{U}'_{a_i}$ is obtained from $\mathcal{U}_{a_i}$ by adding infinite w-chains starting from each node in $\mathcal{U}_{a_i}$ and making B_1 true in every node if i is even and B_2 in every node if i is odd. Similarly, $\mathcal{U}'_{B \sqcap B_3}$ is obtained from $\mathcal{U}_B$ by adding infinite w-chains starting from each node in $\mathcal{U}_B$ and making B_3 true in every node. By definition, it suffices to show that

$$\Pi_{i=1}^n (\mathcal{U}'_{a_i}, a_i) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}'_{B \sqcap B_3}, \rho_{B \sqcap B_3}).$$

But this follows directly from the fact that

$$\prod_{i=1}^n (\mathcal{U}_{a_i}, a_i) \preceq_{\mathcal{EL},\Sigma} (\mathcal{U}_\mathcal{K}, b)$$

and the assumption that $n \geq 2$ which implies that no $B_i, i = 1, 2$ is satisfied in $\Pi_{i=1}^n (\mathcal{U}'_{a_i}, a_i)$.

(b) $\Rightarrow$ (c) is trivial.

(c) $\Rightarrow$ (a) We use Theorem 4. There exists $k \geq 0$ such that for $G_k = \Pi_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, \rho_i)^{\downarrow \mathcal{EL},k}$

$$\Pi_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, \rho_i) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',G_k}, \rho_{G_k}).$$

Consider $(a_1, \ldots, a_n)$ and the product $\prod_{i=1}^n \mathcal{U}'_{a_i}$ of which it is the root. There exists a v-successor $(d_1, \ldots, d_n)$ of ρ_{G_k} in $\mathcal{U}_{\mathcal{T}',G_k}$ such that

$$\prod_{i=1}^n (\mathcal{U}'_{a_i}, a_i) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',G_k}, (d_1, \ldots, d_n))$$

Note that $(d_1, \ldots, d_n) \neq (a_1, \ldots, a_n)$ because there is an infinite w-chain starting at $(a_1, \ldots, a_n)$ in $\prod_{i=1}^n \mathcal{U}'_{a_i}$ but there is no such w-chain starting at $(a_1, \ldots, a_n)$ in $\mathcal{U}_{\mathcal{T}',G_k}$ because no $B_i, i = 1, 2, 3$, is satisfied in any node of G_k in the subtree generated by $(a_1, \ldots, a_n)$ and w-chains are only generated by the TBox $\mathcal{T}'$ from nodes satisfying at least one B_i. It follows that there exists $i \leq n$ such that $d_i = b_i$. Then

$$(\mathcal{U}_{\mathcal{T}',G_k}, (d_1, \ldots, d_n)) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, b_i)$$

as the projections onto components are $\mathcal{EL}$ simulations. By taking the composition of $\mathcal{EL}$ simulations we obtain

$$\prod_{i=1}^n (\mathcal{U}'_{a_i}, a_i) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, b_i)$$

(a) follows by construction of $\mathcal{K}'$.

We now sketch the proof for $\mathcal{ELI}$. The construction given above for $\mathcal{EL}$ almost works except that in the proof of (a) $\Rightarrow$ (b), for $\mathcal{ELI}$ simulations we have to consider the v-predecessor $(\rho_1, \ldots, \rho_n)$ of $(d_1, \ldots, d_n)$ when proving

$$\Pi_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, d_i) \preceq_{\mathcal{ELI},\Sigma'} (\mathcal{U}_{\mathcal{T}',(B \sqcap B_3)}, \rho_{B \sqcap B_3})$$

to show that Condition 2 for $\mathcal{ELI}$-MSC verification in Theorem 3 holds. In the $\mathcal{EL}$ case, the nodes $\mathcal{EL}$ simulating $(d_1, \ldots, d_n)$ might not have v-predecessors simulating $(\rho_1, \ldots, \rho_n)$. To ensure the existence of appropriate v-predecessors we modify the construction as follows. Given an $\mathcal{ELI}$ TBox $\mathcal{T}$ and an ABox $\mathcal{A}$ with assertions $A_1(a_1), \ldots, A_n(a_n), B(b)$, with $A_1, \ldots, A_n, B$ concept names, and a signature Σ, we construct $\mathcal{T}'$ and $\mathcal{A}'$ as before but add to $\mathcal{T}'$ the CIs $B_3 \sqsubseteq \exists v^-.B_4$ and

$$B_4 \sqsubseteq X \sqcap \bigsqcap_{s \in \mathsf{role}(\mathcal{T}')} \exists s.B_4,$$

where $\mathsf{role}(\mathcal{T}')$ denotes the set of all role names and their inverses used in $\mathcal{T}'$ and $X = \bigsqcap_{A \in \mathsf{sub}(\mathcal{T}')} A$. Now Σ' also contains B_4. Then one can prove that the following conditions are equivalent:

(a) $\prod_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i) \preceq_{\mathcal{ELI},\Sigma} (\mathcal{U}_\mathcal{K}, b)$;

(b) C_0 is the $\mathcal{ELI}(\Sigma')$-MSC of $\rho_1, \ldots, \rho_n$ w.r.t. $\mathcal{K}'$;

(c) The $\mathcal{ELI}(\Sigma')$-MSC of $\rho_1, \ldots, \rho_n$ w.r.t. $\mathcal{K}'$ exists.

The reduction follows directly. ❑

Theorem 11 *For $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $\mathcal{L}$-MSC verification can be reduced in polynomial time to the complement of $\mathcal{L}$ concept separability. This also holds for the full signature.*

Proof. Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{L}$ knowledge base, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$ individuals, Σ a signature, and C an $\mathcal{L}(\Sigma)$ concept. We construct a new ABox $\mathcal{A}'$ as follows:

- start with $\mathcal{A}$ extended with a disjoint copy of $\mathcal{A}$ where every individual $a \in \mathsf{ind}(\mathcal{A})$ is replaced with a';
- take a fresh role name s, and let $\mathcal{A}_{ij}$, with $i, j \in \{1, \dots, n\}$, be (disjoint) copies of $\mathcal{A}_C$ with roots ρ_{ij}. Then add the ABoxes $\mathcal{B}_i$, for every $i \in \{1, \dots, n\}$, $\mathcal{A}_C$ (with root ρ_C), and $\mathcal{B}'$ (also with root ρ_C) defined as follows:

$$\begin{aligned}\mathcal{B}_i = & \bigcup_{j \in \{1,\dots,n\}} \mathcal{A}_{ij} \cup \{s(a_i, \rho_{i1})\} \cup \\ & \{s(\rho_{ij}, \rho_{ij+1}) \mid j \in \{1, \dots, n-1\}\} \\ \mathcal{B}' = & \{s(\rho_C, a'_1)\} \cup \{s(a'_i, a'_{i+1}) \mid i \in \{1, \dots, n-1\}\}\end{aligned}$$

Intuitively, $\mathcal{B}$ adds an s-chain of length n to every a_i in which every element satisfies C, and $\mathcal{B}'$ adds an s-chain to the copies of the individuals a_i.

Let $\mathcal{K}' = (\mathcal{T}, \mathcal{A}')$ and $\Sigma' = \Sigma \cup \{s\}$. Moreover, let $\mathcal{U}$ denote the interpretation that is obtained by taking the union of $\prod_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i)$ and $\mathcal{U}_{\mathcal{T},\mathcal{B}_1}$ (the index is not important), identifying the root a_1 of $\mathcal{U}_{\mathcal{T},\mathcal{B}_1}$ with the root $(a_1, \dots, a_n)$ of the product. Let ρ denote the new root of $\mathcal{U}$. Note that we have:

$$\prod_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, a_i) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}, \rho) \preceq_{\mathcal{L},\Sigma'} \prod_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, a_i). \qquad (1)$$

The second simulation exists since $\mathcal{U}$ is a sub-structure of the product. The first simulation exists because

$$\prod_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, a_i) \preceq_{\mathcal{L},\Sigma} \prod_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i)$$

and because for elements in the product reachable via some s-successor of $(a_1, \dots, a_n)$, any projection is an $\mathcal{L}(\Sigma')$-simulation to $\mathcal{U}_{\mathcal{T},\mathcal{B}_1}$.

Claim. C is the $\mathcal{L}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $\mathcal{K}$ iff $\prod_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, a_i) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, \rho_C)$.

Proof of the Claim. For the "if"-direction, suppose $\prod_{i=1}^n (\mathcal{U}_{\mathcal{K}'}, a_i) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, \rho_C)$. By Equation (1), it follows that

$$(\mathcal{U}, \rho) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, \rho_C).$$

Since s is fresh, we have that $\prod_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{T},\mathcal{A}_C}, \rho_C)$ and $(\mathcal{U}_{\mathcal{T},\mathcal{B}_1}, a_1) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{T},\mathcal{B}'}, \rho_C)$. The former is just Condition 2 of Theorem 3 is satisfied. Moreover, the latter implies that $(\mathcal{U}_{\mathcal{T},\mathcal{A}_{ij}}, \rho_{ij}) \preceq_{\mathcal{L},\Sigma} (\mathcal{U}_\mathcal{K}, a'_i)$, for all $i \in \{1, \dots, n\}$. Thus, we also have $\mathcal{K} \models C(a_i)$, for all i and hence also Condition 1 of Theorem 3 holds.

For "only if", suppose that C satisfies Conditions 1 and 2 of Theorem 3. The former implies that $\mathcal{K} \models C(a_i)$ for all $i \in \{1, \dots, n\}$, and thus $(\mathcal{U}_{\mathcal{T},\mathcal{A}_C}, \rho_C) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_\mathcal{K}, a_i)$, for all i. It can be verified that $(\mathcal{U}_{\mathcal{T},\mathcal{B}_1}, a_1) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{T},\mathcal{B}'}, \rho_C)$. Moreover, Condition 2 reads $\prod_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i) \preceq_{\mathcal{L},\Sigma} (\mathcal{U}_{\mathcal{T},\mathcal{A}_C}, \rho_C)$. Together, we have $(\mathcal{U}, \rho) \preceq_{\mathcal{L},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, \rho_C)$, and the claim follows from Equation (1).

The Claim establishes correctness of the reduction, so it remains to note that the construction of $\mathcal{A}$' can be implemented in polynomial time. ❑

Proofs for Section 6

To show Theorem 12, we first observe the following easily proved relationship between $\mathcal{ELI}$ simulations between $(\mathcal{I}_1, d)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}$ and $(\mathcal{I}_2, e)$ and preservation of $\mathcal{ELI}^{\mathsf{sf}}$ concepts from $(\mathcal{I}_1, d)$ to $(\mathcal{I}_2, e)$.

Lemma 11 *Let $\mathcal{I}_1, \mathcal{I}_2$ have finite outdegree, and let Σ be a signature. The following conditions are equivalent:*

- $(\mathcal{I}_1, d)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}} \preceq_{\mathcal{ELI},\Sigma} (\mathcal{I}_2, e)$*;*
- *for all $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$ concepts C: if $d \in C^{\mathcal{I}_1}$, then $e \in C^{\mathcal{I}_2}$.*

We also state and prove the characterization for MSC verification in $\mathcal{ELI}^{\mathsf{sf}}$.

Theorem 19 *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELI}$ KB, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$, and Σ a signature. An $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$ concept C is the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC of $a_1, \dots, a_n$ with respect to $\mathcal{K}$ if, and only if, the following conditions hold:*

1. $(a_1, \dots, a_n) \in C^{\Pi_{i=1}^n \mathcal{U}_\mathcal{K}}$*;*
2. $(\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}} \preceq_{\mathcal{ELI},\Sigma} \mathcal{U}_{\mathcal{T},C}, \rho_C$*.*

Proof. The proof if similar to the proof of Theorem 3. By Lemmas 3 and 4, Condition 1 is equivalent to Condition 1 of the definition of MSCs. For Condition 2 observe that by Lemmas 3, 4, and 11, Condition 2 is equivalent to Condition 2 of the definition of MSCs. ❑

Theorem 12 **($\mathcal{ELI}^{\mathsf{sf}}$-MSC Existence w.r.t. $\mathcal{ELI}$ TBoxes)** *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELI}$ KB, $a_1, \dots, a_n \in \mathsf{ind}(\mathcal{A})$, and Σ a signature. The following are equivalent, for $C_k = (\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}},k}_{|\Sigma}$:*

1. *the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC of $a_1, \dots, a_n$ w.r.t. $\mathcal{K}$ exists;*

2. *C_k is the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC of $a_1,\ldots,a_n$ w.r.t. $\mathcal{K}$, for a $k \geq 0$;*
3. *$\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i)^{\downarrow\mathcal{ELI}^{\mathsf{sf}}} \preceq_{\mathcal{ELI},\Sigma} (\mathcal{U}_{\mathcal{T},C_k},\rho_{C_k})$ for a $k \geq 0$.*

The proof is similar to the proof of Theorem 4.

Proof. "2 $\Rightarrow$ 1" is trivial. "3 $\Rightarrow$ 2" is an immediate consequence of Theorem 19. For "1 $\Rightarrow$ 3", let the $\mathcal{L}(\Sigma)$-MSC D be of depth k. It follows from Theorem 19 that

$$(a_1,\ldots,a_n) \in D^{\Pi_{i=1}^n \mathcal{U}_{\mathcal{T},\mathcal{K}}}$$

which implies

$$(a_1,\ldots,a_n) \in D^{(\Pi_{i=1}^n \mathcal{U}_{\mathcal{T},\mathcal{K}})^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}}$$

since D is an $\mathcal{ELI}^{\mathsf{sf}}$ concept. As D has depth k and has signature Σ,

$$\rho_{C_k} \in D^{\mathcal{U}_{\mathcal{T},C_k}}.$$

Now, Point 3 follows from the definition of the MSC and Lemmas 3, 4, and 11. ❏

Theorem 13 *$\mathcal{ELI}^{\mathsf{sf}}$-MSC and LCS existence and verification with respect to $\mathcal{ELI}$ TBoxes are* ExpTime*-complete. The lower bounds hold in the full signature case and with only one example.*

We show hardness for LSC verification and existence at the same time, by reducing from concept subsumption relative to general $\mathcal{ELI}$ TBoxes (Baader, Brandt, and Lutz 2008). Hardness for MSC verification and existence then follows from Theorem 1. Let $\mathcal{T}, A, B$ be an input to the subsumption problem. We define a TBox $\mathcal{T}'$ by taking fresh role names r, s and fresh concept names E, F and setting for $C := \exists r.\exists r^-.A$, $D_0 := \exists r.\exists r^-.B$, and $D_1 := \exists r.\exists r^-.E$:

$$\begin{aligned}\mathcal{T}' \;=\; & \mathcal{T} \cup \{C \sqsubseteq \exists s.D_1, D_1 \sqsubseteq \exists s.D_1,\\ & D_0 \sqsubseteq F, F \sqsubseteq \exists s.(\exists r.\top \sqcap F)\}.\end{aligned}$$

Based on Theorems 1, 19, and 12, one can verify that the following conditions are equivalent:

(a) $\mathcal{T} \models A \sqsubseteq B$;

(b) $\exists r.\top \sqcap F$ is the $\mathcal{ELI}^{\mathsf{sf}}$-LCS of C w.r.t $\mathcal{T}'$;

(c) the $\mathcal{ELI}^{\mathsf{sf}}$-LCS of C w.r.t. $\mathcal{T}'$ exists.

This establishes the claimed lower bounds.

Establishing the upper bounds requires more work. We start with Lemma 5.

Lemma 5 *Let N be the outdegree of $\Pi_{i=1}^n \mathcal{U}_\mathcal{K}$. Then the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC of $a_1,\ldots,a_n$ w.r.t. $\mathcal{K}$ exists iff, for some subconcept D of $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}_\Sigma$, we have:*

$$(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}} \preceq^{in}_{\mathcal{ELI},\Sigma} (\mathcal{U}^{\times N}_{\mathcal{T},D},\rho_D).$$

Proof. In the "if"-direction, assume that

$$(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}} \preceq^{\text{in}}_{\mathcal{ELI},\Sigma} \mathcal{U}^{\times N}_{\mathcal{T},D},\rho_D,$$

for some subconcept D of $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}_\Sigma$, and let S be a witnessing injective simulation. Let h be the homomorphism from $\mathcal{U}^{\times N}_{\mathcal{T},D}$ to $\mathcal{U}_{\mathcal{T},D}$ which maps every element to its "original". It should be clear the relation S' defined by

$$S' = \{(d,h(e)) \mid (d,e) \in S\}$$

is an $\mathcal{ELI}(\Sigma)$-simulation between $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}$ and $\mathcal{U}_{\mathcal{T},D},\rho_D$. Since D is a subconcept of $(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}_\Sigma$, it is also a sub-concept of C_k where k is the role depth of D, and thus

$$(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}} \preceq_{\mathcal{ELI},\Sigma} \mathcal{U}_{\mathcal{T},C_k},\rho_{C_k}.$$

By Theorem 12, the MSC exists.

Conversely, suppose the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC exists, and thus, there is a $k \geq 0$ with

$$(\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}} \preceq_{\mathcal{ELI},\Sigma} \mathcal{U}_{\mathcal{T},C_k},\rho_{C_k}.$$

Take $D = C_k$ and let S be the witnessing simulation. It is crucial to observe that, by the definition of $\mathcal{ELI}^{\mathsf{sf}}$ unfolding we have the following property:

($*$) for all $(d,e),(d',e') \in S$: if d' is a successor of d in the tree $\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}$ then e' is an successor of e in the tree $\mathcal{U}_{\mathcal{T},D},\rho_D$.

Intuitively, the simulation always goes "downwards" in the right tree. Based on this insight, we construct an injective simulation S' to $\mathcal{U}^{\times N}_{\mathcal{T},D},\rho_D$ inductively. During the construction, we maintain the invariant that $(d,e) \in S'$ implies $(d,h(e)) \in S$, where h is the homomorphism from $\mathcal{U}^{\times N}_{\mathcal{T},D}$ to $\mathcal{U}_{\mathcal{T},D}$ which maps every element to its "original".

- Start with $S' = \{(a_1,\ldots,a_n),\rho_D\}$;
- For the inductive step, do the following for every $(d,e) \in S'$: let $d_1,\ldots,d_n$ be the all Σ-successors of d in $\Pi_{i=1}^n(\mathcal{U}_\mathcal{K},a_i))^{\downarrow\mathcal{ELI}^{\mathsf{sf}}}$, that is, $n \leq N$. Since $(d,h(e)) \in S$, there are corresponding Σ-successors $e_1,\ldots,e_n$ in $\mathcal{U}_{\mathcal{T},D}$ such that $(d,e_i) \in S$ for all i. Let $e_1',\ldots,e_n'$ be pairwise distinct copies of these nodes in $\mathcal{U}^{\times N}_{\mathcal{T},D}$, and add $(d_i,e_i) \in S'$, for all i.

It can be verified that the invariant is preserved and that S' is an injective simulation because of ($*$). ❏

We give now the automata-based approach to deciding the criterion in Lemma 5. We start with providing the necessary preliminaries. An *n-ary tree* is the set $T = \{1,\ldots,n\}^*$. For a node $ui \in T$, we identify $ui\cdot{-1}$ with u. For an alphabet Θ, a *Θ-labeled tree* is a pair (T,L) with T a tree and $L : T \to \Theta$ a node labeling function. We also recall the notion of *non-deterministic parity tree automata (NTA)*. An NTA over N-ary trees is a tuple $\mathfrak{A} = (Q,\Theta,q_0,\Delta,\Omega)$ where Q is a set of states, Θ is the input alphabet, $q_0 \in Q$ is the initial state, $\Delta \subseteq Q \times \Theta \times Q^N$ is the transition relation, and $\Omega : Q \to \mathbb{N}$ is the priority function. The semantics of NTAs is defined as usual via runs. A *run* of an NTA $\mathfrak{A} = (Q,\Theta,q_0,\Delta,\Omega)$ over an N-ary input (T,L) is a Q-labeled tree (T,r) such that:

- $r(\varepsilon) = q_0$, and
- for all $w \in T$, $(r(w),L(w),r(w1),\ldots,r(wN)) \in \Delta$.

Let $\gamma = i_0 i_1 \cdots$ be an infinite path in (T,r) and denote, for all $j \geq 0$, with q_j the state such that $r(i_j) = (x,q_j)$. The path γ is *accepting* if the largest number m such that $\Omega(q_j) = m$ for infinitely many j is even. A run (T,r) is accepting, if all infinite paths in T_r are accepting. The language accepted by $\mathfrak{A}$, denoted $L(\mathfrak{A})$, is the set of all trees (T, L) for which there is an accepting run.

To encode interpretations we use the alphabet $\Theta = 2^{\Theta_0}$ where

$$\Theta_0 = \mathsf{sub}(\mathcal{T}) \cup \{r, r^- \mid r \text{ occurs in } \mathcal{T}\}.$$

A Θ-labeled tree (T, L) represents the interpretation $\mathcal{I}_L = (T, \cdot^{\mathcal{I}_L})$ given by

$$\begin{aligned} A^{\mathcal{I}_L} &= \{u \mid A \in L(u)\}, \\ r^{\mathcal{I}_L} &= \{(u, u\cdot -1) \mid r^- \in L(u)\} \cup \{(u\cdot -1, u) \mid r \in L(u)\} \end{aligned}$$

for every concept name $A \in \mathsf{sub}(\mathcal{T})$ and role name r that occurs in $\mathcal{T}$. Note that the interpretation $\mathcal{I}_L$ is not necessarily connected; however, we usually identify $\mathcal{I}_L$ with its sub-interpretation induced by all elements reachable from the root. It should be clear that conversely, for every tree-shaped interpretation $\mathcal{I}$ of outdegree $\leq n$, there is an n-ary labeled tree (T, L) such that $\mathcal{I}$ and $\mathcal{I}_L$ are ismorphic.

We also use NTA over the alphabet Θ^2 in which case an input tree (T, L) is treated as two trees (T, L_1), (T, L_2) and thus encodes two interpretations $\mathcal{I}_{L_1}$ and $\mathcal{I}_{L_2}$. Finally, we treat $\mathcal{I}_L$ as a concept if it is finite and has no multiedges.

Lemma 12 *Let N be the outdegree of $\Pi_{i=1}^n \mathcal{U}_\mathcal{K}$.*

1. *There is an NTA $\mathfrak{A}$ such that, for all N^2-ary Θ-labeled trees (T, L), we have:*

$$(T, L) \in L(\mathfrak{A}) \quad \textit{iff} \quad (\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}} \preceq^{\textit{in}}_{\mathcal{ELI},\Sigma} \mathcal{I}_L, \varepsilon.$$

2. *There is an NTA $\mathfrak{B}_0$ over N^2-ary Θ^2-labeled trees such that for every $(T, L) \in L(\mathfrak{B}_0)$, there is a subconcept D of $(\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}_{|\Sigma}$ such that:*
   (a) $\mathcal{I}_{L_1}$ is the concept D, and
   (b) $\mathcal{I}_{L_2}$ is isomorphic to $\mathcal{U}_{\mathcal{T},D}$.
   Conversely, for every subconcept D of $(\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}_{|\Sigma}$ there is some $(T, L) \in L(\mathfrak{B}_0)$ such that $\mathcal{I}_{L_1}$ is D.

Moreover, $\mathfrak{A}$ and $\mathfrak{B}_0$ can be constructed in time exponential in $|\mathcal{K}|$.

In order to prove this lemma, we need a concrete definition of $(\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}$. Let $\mathsf{tp}(\mathcal{T})$ denote the set of all types for $\mathcal{T}$, and $\mathsf{rol}(\mathcal{T})$ be the set of roles that occur in $\mathcal{T}$. For each $r \in \mathsf{rol}(\mathcal{T})$, we define a relation $\rightarrow_r$ on the set $U := D^n \times (\{\varepsilon\} \cup \mathsf{rol}(\mathcal{T}))$ for

$$D = \mathsf{ind}(\mathcal{A}) \cup \mathsf{tp}(\mathcal{T})$$

by taking $(x_1, \ldots, x_n, x) \rightarrow_r (y_1, \ldots, y_n, y)$ iff

- $y = r$ and $x \neq r^-$, and
- for every $i \in \{1, \ldots, n\}$, one of the following is satisfied:
  (i) $x_i, y_i \in \mathsf{ind}(\mathcal{A})$ and $r(x_i, y_i) \in \mathcal{A}$
  (ii) $x_i \in \mathsf{ind}(\mathcal{A})$, $y_i \in \mathsf{tp}(\mathcal{T})$, and $a_i \rightsquigarrow_r^{\mathcal{T},\mathcal{A}} y_i$;
  (iii) $x_i, y_i \in \mathsf{tp}(\mathcal{T})$ and $x_i \rightsquigarrow_r^{\mathcal{T}} y_i$;

Some element $(x_1, \ldots, x_n, x) \in U$ *satisfies* a concept name A if for every $i \in \{1, \ldots, n\}$, either $x_i \in \mathsf{tp}(\mathcal{T})$ and $A \in x_i$, or $x_i \in \mathsf{ind}(\mathcal{A})$ and $\mathcal{U}_\mathcal{K} \models A(x_i)$. A *path* is a sequence $u_0 r_0 u_1 r_1 \cdots r_{n-1} u_n$ such that $u_0 = (a_1, \ldots, a_n, \varepsilon)$, $u_{i-1} \rightarrow_{r_{i-1}} u_i$, for all $i \in \{1, \ldots, n\}$. We denote with PATHS the set of all paths, and with $\mathsf{tail}(p)$ the last element in the sequence p. It can be verified that the interpretation $(\mathcal{U}, (a_1, \ldots, a_n, \varepsilon))$ with $\mathcal{U}$ defined below is isomorphic to $(\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a_i))^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}$:

$$\begin{aligned} \Delta^\mathcal{U} &= \mathsf{PATHS} \\ A^\mathcal{U} &= \{p \mid \mathsf{tail}(p) \text{ satisfies } A\} \\ r^\mathcal{U} &= \{(p, pru') \mid pru' \in \mathsf{PATHS}\} \cup \\ &\quad\ \{(pr^- u, p) \mid pr^- t \in \mathsf{PATHS}\}. \end{aligned}$$

Construction of $\mathfrak{A}$ Informally, the automaton $\mathfrak{A}$ simulates the definition of $\mathcal{U}$ by keeping in its state only the tail of the current path.

More formally, the set Q of states is the smallest set that contains $q_\top$ and $(a_1, \ldots, a_n, \varepsilon)$ and is closed under the relations $\rightarrow_r$ defined above, that is, if $u \in Q$ and $u \rightarrow_r u'$ for some $r \in \mathsf{rol}(\mathcal{T})$, then $u' \in Q$. The initial state is $(a_1, \ldots, a_n, \varepsilon)$, and the transition relation contains $(q_\top, \theta, q_\top^{N^2})$, for all $\theta \in \Theta$, and

$$((x_1, \ldots, x_n, x), \theta, q_1, \ldots, q_{N^2})$$

whenever:

- if $(x_1, \ldots, x_n, x)$ satisfies A then $A \in \theta$, for all $A \in \Sigma$;
- if $x \neq \varepsilon$, then $x \in \theta$;
- each $(y_1, \ldots, y_n, r)$ such that $(x_1, \ldots, x_n, x) \rightarrow_r (y_1, \ldots, y_n, r)$ for some $r \in \mathsf{rol}(\mathcal{T})$ occurs precisely once in $q_1, \ldots, q_{N^2}$; all other q_i are $q_\top$.

It is routine to verify that $\mathfrak{A}$ is as required.

Construction of $\mathfrak{B}_0$ The automaton $\mathfrak{B}_0$ is the intersection of three automata $\mathfrak{A}_0, \mathfrak{A}_1, \mathfrak{A}_2$.

Automaton $\mathfrak{A}_0$ verifies Condition (a) from the Lemma by ensuring that $\mathcal{I}_{L,1}$ is a Σ-concept, that is, finite and without multiedges, and in fact, a subconcept of $\Pi_{i=1}^n (\mathcal{U}_\mathcal{K}, a)^{\downarrow \mathcal{ELI}^{\mathsf{sf}}}_{|\Sigma}$. Realizing this condition as an NTA is relatively straightforward; details are thus omitted.

The automata $\mathfrak{A}_1, \mathfrak{A}_2$ together verify Condition (b) from the Lemma, assuming that $\mathcal{I}_{L_1}$ is some concept C. The first, $\mathfrak{A}_1$, ensures that for all $n \in \Delta^{\mathcal{I}_{L_1}}$ with $L_1(n) \neq \emptyset$ and all $D \in \mathsf{sub}(\mathcal{T})$, we have

$$D(a) \in L_2(n) \quad \text{iff} \quad \mathcal{T}, \mathcal{I}_{L_1} \models D(a). \tag{2}$$

$$r \in L_2(n) \quad \text{iff} \quad r \in L_1(n) \tag{3}$$

Thus, on the elements in $\mathcal{I}_{L_1}$, the interpretation $\mathcal{I}_{L_2}$ is the universal model of $\mathcal{I}_{L_1}$ and $\mathcal{T}$. Based on this, $\mathfrak{A}_2$ just generates (in L_2) the trees induced in the universal model below the elements in $\Delta^{\mathcal{I}_{L_1}}$ by simulating $\rightsquigarrow_r^{\mathcal{T}}$, which is again standard and omitted.

It is rather tedious to specify the automaton $\mathfrak{A}_1$ ensuring (2) directly as an NTA. Instead, we specify $\mathfrak{A}_1$ as a *two-way alternating tree automata (TWAPA)* relying on the fact that every TWAPA can be transformed into an equivalent NTA under an exponential blowup (Vardi 1998).

Two-way Alternating Tree Automata A *two-way alternating parity tree automaton over k-ary trees (TWAPA)* is a tuple $\mathfrak{A} = (Q, \Theta, q_0, \delta, \Omega)$ where Q is a finite set of *states*, Θ is the *input alphabet*, $q_0 \in Q$ is the *initial state*, δ is a *transition function*, and $\Omega : Q \to \mathbb{N}$ is a *priority function* (Vardi 1998). 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 $[k] = \{-1, 0, 1, \ldots, k\}$. The semantics is given in terms of *runs*. More precisely, let (T, L) be a Θ-labeled tree and $\mathfrak{A} = (Q, \Theta, q_0, \delta, \Omega)$ a TWAPA. A *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 $r(y) = (x \cdot i, q')$.

Let $\gamma = i_0 i_1 \cdots$ be an infinite path in T_r and denote, for all $j \geq 0$, with q_j the state such that $r(i_j) = (x, q_j)$. The path γ is *accepting* if the largest number m such that $\Omega(q_j) = m$ for infinitely many j is even. A run (T_r, r) is accepting, if all infinite paths in T_r are accepting. $\mathfrak{A}$ accepts a tree if $\mathfrak{A}$ has an accepting run over it.

Before we can give the TWAPA, we need some preliminary notions, in particular a syntactic characterization of whether $\mathcal{T}, \mathcal{A} \models C(a)$, which can be easily be implemented in a TWAPA. Similar characterizations have been used before, e.g., in (Jung et al. 2017).

Derivation Trees Fix an $\mathcal{ELI}$ knowledge base $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, $a_0 \in \mathsf{ind}(\mathcal{A})$, and $C \in \mathsf{sub}(\mathcal{T})$. A *derivation tree* for an assertion $C_0(a_0)$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$ is a finite $\mathsf{ind}(\mathcal{A}) \times \mathsf{sub}(\mathcal{T})$-labeled tree (T, V) such that:

- $V(\varepsilon) = (a_0, C_0)$;
- if $V(n) = (a, C)$ and neither $C(a) \in \mathcal{A}$ nor $\mathcal{T} \models \top \sqsubseteq C$, one of the following holds:
  - (i) n has successors $n_1, \ldots, n_k$, $k \geq 1$ with $V(n_i) = (a, C_i)$, for $1 \leq i \leq k$ and $\mathcal{T} \models C_1 \sqcap \ldots \sqcap C_k \sqsubseteq C$;
  - (ii) n has a single successors n' with $V(n') = (b, C')$ such that $r(a, b) \in \mathcal{A}$ and $\mathcal{T} \models \exists r.C' \sqsubseteq C$.

Lemma 13 *$\mathcal{T}, \mathcal{A} \models C_0(a_0)$ iff there is a derivation tree for $C_0(a_0)$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$.*

Proof. $(\Leftarrow)$ is clear.

For $(\Rightarrow)$, we construct a sequence of ABoxes $\mathcal{A} = \mathcal{A}_0, \mathcal{A}_1, \ldots$, by obtaining $\mathcal{A}_{i+1}$ from $\mathcal{A}_i$ by applying one of the following two rules:

1. if $C_1(a), \ldots, C_k(a) \in \mathcal{A}$ and $\mathcal{T} \models C_1 \sqcap \ldots \sqcap C_k \sqsubseteq C$ for some $C \in \mathsf{sub}(\mathcal{T})$, then add $C(a)$ to $\mathcal{A}_i$;
2. if $r(a, b), C'(b) \in \mathcal{A}_i$ and $\mathcal{T} \models \exists r.C' \sqsubseteq C$ for some $C \in \mathsf{sub}(\mathcal{T})$, then add $C(a)$.

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

Claim. There is a model $\mathcal{I}$ of $\mathcal{A}^*$ and $\mathcal{T}$ such that $a \in C^{\mathcal{I}}$ implies $C(a) \in \mathcal{A}^*$, for all $a \in \mathsf{ind}(\mathcal{A})$ and $C \in \mathsf{sub}(\mathcal{T})$.

Proof of the Claim. Start with an interpretation $\mathcal{I}_0$ defined by:

$$\Delta^{\mathcal{I}_0} = \mathsf{ind}(\mathcal{A})$$
$$A^{\mathcal{I}_0} = \{a \mid A(a) \in \mathcal{A}^*\}$$
$$r^{\mathcal{I}_0} = \{(a, b) \mid r(a, b) \in \mathcal{A}\}$$

Denote with $\mathcal{A}^*_a$ the set of all $C(a)$ in $\mathcal{A}^*$. Now extend $\mathcal{I}_0$ as follows: For every $a \in \mathsf{ind}(\mathcal{A})$ and every $C \sqsubseteq \exists r.C' \in \mathcal{T}$ such that $\mathcal{T}, \mathcal{A}^*_a \models C(a)$, add the r-successor of a satisfying C' in $\mathcal{U}_{\mathcal{T}, \mathcal{A}^*_a}$ as an r-successor of a to $\mathcal{I}$.

This finishes the construction of $\mathcal{I}$, and it can be verified that it indeed satisfies the requirements of the claim. This finishes the proof of the claim.

Now suppose $\mathcal{T}, \mathcal{A} \models C_0(a_0)$. By the Claim, we have $C_0(a_0) \in \mathcal{A}^*$. Exploiting that the two rules to construct $\mathcal{A}_0, \mathcal{A}_1, \ldots$ are in one-to-one correspondence with Conditions (i) and (ii) from the definition of derivation trees, we can inductively construct a derivation tree for $C_0(a_0)$ in $\mathcal{A}$ w.r.t. $\mathcal{T}$. ❏

We are now in the position to construct the automaton $\mathfrak{A}_1$. It ensures that when a Θ^2-labeled tree (T, L) is accepted, then for all $n \in \Delta^{\mathcal{I}_{L_1}}$ with $L_1(n) \neq \emptyset$, Condition (3) is satisfied, and for all concepts $D \in \mathsf{sub}(\mathcal{T})$:

$(*)$ $D \in L_2(n)$ iff there is a derivation tree for $D(n)$ in $\mathcal{I}_{L_1}$ (viewed as ABox);

By Lemma 13, this condition ensure that Equation (2) above is satisfied. We take $\mathfrak{A}_1 = (Q, \Theta, q_0, \delta, \Omega)$ where

$$Q = \{q_0, q_0'\} \cup \{q_D, \overline{q}_D \mid D \in \mathsf{sub}(\mathcal{T})\} \cup \{q_r, \overline{q}_r, q_{r,D}, \overline{q}_{r,D} \mid r \in \mathsf{rol}(\mathcal{T}), D \in \mathsf{sub}(\mathcal{T})\}$$

and Ω assigns zero to all states, except for states of the form q_D, to which it assigns one.

For Condition $(*)$, we use states q_D for the "$\Rightarrow$" part, and states $\overline{q}_D$ for the "$\Leftarrow$" part. Intuitively, a state q_D assigned to some node n is an obligation to verify the existence of a derivation tree for $D(n)$. Conversely, $\overline{q}_A$ is the obligation that there is *no* such derivation tree. The automaton starts with the following transitions for every $\theta = (\theta_1, \theta_2)$:

$$\delta(q_0, \theta) = \begin{cases} \mathsf{true} & \text{if } \theta_1 = \emptyset \\ q_0' & \text{if } \theta_1 \neq \emptyset \text{ and } \theta_2 \cap \mathsf{rol}(\mathcal{T}) = \theta_1 \cap \mathsf{rol}(\mathcal{T}) \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta(q_0', \theta) = \bigwedge_{i=1}^{N^2} (i, q) \wedge \bigwedge_{D \in \theta_2 \cap \mathsf{sub}(\mathcal{T})} q_D \wedge \bigwedge_{D \in \mathsf{sub}(\mathcal{T}) \setminus \theta_2} \overline{q}_D$$

For states q_D, we implement Conditions (i) and (ii) of derivation trees as transitions. Finiteness of the derivation tree is ensured by the priority assigned to these states. More precisely, we set, for all $\theta \in \Theta$, $\delta(q_D, \theta) = \mathsf{false}$ if $\theta_1 = \emptyset$, $\delta(q_D, \theta) = \mathsf{true}$ if $D \in L_0$ or $\mathcal{T} \models \top \sqsubseteq D$, and otherwise

$$\delta(q_D, \theta) = \bigvee_{\mathcal{T} \models D_1 \sqcap \cdots \sqcap D_k \sqsubseteq D} \big((0, q_{D_1}) \wedge \cdots \wedge (0, q_{D_k})\big) \vee$$

$$\bigvee_{\mathcal{T} \models \exists r.D' \sqsubseteq D} \big(((-1, q_{D'}) \wedge q_r) \vee \bigvee_{i=1}^{N^2} (i, q_{r,D'})\big)$$

Finally, the transitions for q_r and $q_{r,D}$ are as follows:

$$\delta(q_r, \theta) = \begin{cases} \mathsf{true} & \text{if } r \in \theta_1 \\ \mathsf{false} & \text{otherwise} \end{cases}$$

$$\delta(q_{r,D}, \theta) = (0, q_r) \wedge (0, q_D)$$

The transitions for $\overline{q}_D, \overline{q}_r$, and $\overline{q}_{r,D}$ are obtained by dualizing the ones for $q_D, q_r, q_{r,D}$. More precisely, for every such q, we define $\delta(\overline{q}, \theta) = \overline{\delta(q, \theta)}$, where $\overline{\varphi}$ is obtained from φ by exchanging $\wedge$ with $\vee$, true with false, and replacing every state p with $\overline{p}$. This finishes the construction of $\mathfrak{A}_1$. Based on the provided explanations it can be shown that it verifies Condition (2) from above, that is, it computes locally the universal model of C.

This finishes the proof of Lemma 12.

Now, let $\mathfrak{B}_0'$ be the projection of the NTA $\mathfrak{B}_0$ from Lemma 12 to its second component L_2. Then, obtain $\mathfrak{B}_0$ from $\mathfrak{B}_0'$ by modifying $\mathfrak{B}_0'$ such that it accepts

$$\{(T, L)^{\times N} \mid (T, L) \in \mathfrak{B}_0'\}.$$

Note that both modifications can be implemented in polynomial time. Based on Lemmas 5 and 12, it is not hard to verify that the $\mathcal{ELI}^{\mathsf{sf}}(\Sigma)$-MSC of $a_1, \ldots, a_n$ w.r.t. $\mathcal{K}$ exists iff $\mathfrak{A}$ and $\mathfrak{B}$ have non-empty intersection. Since the involved automata can be constructed in exponential time and intersection emptiness for NTAs can be solved in PTIME, the upper bounds in Theorem 13 follow.

Proofs for Section 7

Theorem 15 *In $\mathcal{ELI}$, single example MSC existence and verification are* 2-EXPTIME-*complete in general and* EXPTIME-*complete when the signature is full.*

We start with the lower bounds. For the full signature case, we reduce concept subsumption relative to general $\mathcal{ELI}$-TBoxes which is EXPTIME-hard, already for subsumption between concept names (Baader, Brandt, and Lutz 2008). Let $\mathcal{T}, A, B$ be an input to the subsumption problem. We define a knowledge base $\mathcal{K} = (\mathcal{T}', \mathcal{A})$ by taking

$$\mathcal{T}' = \mathcal{T} \cup \{B \sqsubseteq \exists r.E, E \sqsubseteq \exists r.E \sqcap \exists r^-.E \sqcap \exists r^-.A\}$$
$$\mathcal{A} = \{A(a), r(a, b), r(b, b)\},$$

for fresh names E, r. Then, the following are equivalent.

(a) $\mathcal{T} \models A \sqsubseteq B$;

(b) the $\mathcal{ELI}$-MSC of a w.r.t. $\mathcal{K}$ is A;

(c) the $\mathcal{ELI}$-MSC of a w.r.t. $\mathcal{K}$ exists.

Observe first that $(b) \Rightarrow (c)$ is trivial. For $(a) \Rightarrow (b)$, suppose $\mathcal{T} \models A \sqsubseteq B$. We show via Theorem 3 that A is the $\mathcal{ELI}$-MSC of a w.r.t. $\mathcal{K}$. Condition 1 is satisfied since $A(a) \in \mathcal{A}$. For Condition 2, note first that the universal model $\mathcal{U}_{\mathcal{T}',A}$ has the following structure:

- its root ρ_A satisfies A, B and has an r-successor b satisfying E;
- b is the root of an infinite binary tree in which each node has an r-successor and an r^--successor satisfying E;
- every node in the binary tree has an r^--successor satisfying A;
- below each element satisfying A we find a copy of $\mathcal{U}_{\mathcal{T}',A}$ itself.

It is now straightforward to define an $\mathcal{ELI}$ simulation between $(\mathcal{U}_\mathcal{K}, a)$ and $(\mathcal{U}_{\mathcal{T}',A}, \rho_A)$.

For $(c) \Rightarrow (a)$, suppose C is an $\mathcal{ELI}$-MSC of a w.r.t. $\mathcal{K}$. By Theorem 4, there is some $k \geq 0$ such that $(\mathcal{U}_\mathcal{K}, a) \preceq_{\mathcal{ELI}} (\mathcal{U}_{\mathcal{T}',C_k}, \rho_{C_k})$. Note that there is an infinite r-path starting from a in $\mathcal{U}_\mathcal{K}$. However, if $\mathcal{T} \not\models A \sqsubseteq B$, there is no infinite r-path starting at ρ_{C_k} in any $\mathcal{U}_{\mathcal{T}',C_k}$, a contradiction.

For 2-EXPTIME-hardness in the general case, we reduce the complement of single example $\mathcal{ELI}$ concept separability which has been shown to be 2-EXPTIME-hard in (Gutiérrez-Basulto, Jung, and Sabellek 2018).[1]

Theorem 20 *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. Then the complement of single individual $\mathcal{L}$ concept separability can be reduced in polynomial time to single example $\mathcal{L}$-MSC verification and to single example $\mathcal{L}$-MSC existence.*

Proof. The basic argument is very similar to the proof of Theorem 8. The difference is that the proof of Theorem 8 relies on the fact that at least two positive examples are given in the learning instance which are then used to generate w-chains using distinct concepts names B_1, B_2 which cannot occur in the MSC. Here, instead of using two positive examples we make use of the signature restriction: the concept name generating the w-chains is not in the signature Σ.

Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. We reduce single individual $\mathcal{L}$ concept separability. We can assume without loss of generality that it is formulated as follows: given an $\mathcal{L}$ TBox $\mathcal{T}$, a signature Σ, and an ABox $\mathcal{A} = \{A(a), B(b)\}$, where A, B are concept names with $B \in \Sigma$, decide whether it is the case that $(\mathcal{U}_\mathcal{K}, a) \preceq_{\mathcal{L},\Sigma} (\mathcal{U}_\mathcal{K}, b)$, where $\mathcal{K} = (\mathcal{T}, \mathcal{A})$. We use the notation introduced in the proof of Theorem 8 and start with $\mathcal{EL}$. Construct a new TBox

$$\mathcal{T}' = \mathcal{T}_{|B_1} \cup \mathcal{T}_{|B_2} \cup \{B_i \sqsubseteq \exists w.B_i \mid i = 1, 2\},$$

where w is a fresh role name and B_1, B_2 are fresh concept names. Take a fresh role name v and construct an ABox $\mathcal{A}'$ with individuals ρ, a, b, and assertions

$$v(\rho, a), \quad v(\rho, b), \quad A(a), \quad B(b), \quad B_1(a), \quad B_2(b)$$

[1]In (Gutiérrez-Basulto, Jung, and Sabellek 2018), the authors consider the query by example problem rather than $\mathcal{ELI}$ concept separability. Hence, the semantic characterization is formulated in terms of homomorphisms between universal models rather than $\mathcal{ELI}$ simulations. But as one can work with tree-shaped universal models there is no difference between homomorphisms and $\mathcal{ELI}$ simulations.

Let $C_0 = \exists v.(B \sqcap B_2)$, set $\Sigma' = \Sigma \cup \{w, v, B_2\}$, and let $\mathcal{K}' = (\mathcal{T}', \mathcal{A}')$. We show that the following conditions are equivalent:

(a) $(\mathcal{U}_\mathcal{K}, a) \preceq_{\mathcal{EL},\Sigma} (\mathcal{U}_\mathcal{K}, b)$;

(b) C_0 is the $\mathcal{EL}(\Sigma')$-MSC of ρ w.r.t. $\mathcal{K}'$;

(c) The $\mathcal{EL}(\Sigma')$-MSC of ρ w.r.t. $\mathcal{K}'$ exists.

(a) $\Rightarrow$ (b). Assume (a) holds. By Theorem 3, it suffices to show

1. $\rho \in C_0^{\mathcal{U}_{\mathcal{K}'}}$;
2. $(\mathcal{U}_{\mathcal{K}'}, \rho) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',C_0}, \rho_{C_0})$.

Condition 1 is by construction. For Condition 2, it suffices to show that

$$(\mathcal{U}_{\mathcal{K}'}, d) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',(B\sqcap B_2)}, \rho_{B\sqcap B_2})$$

for $d \in \{a, b\}$. For $d = b$ this is trivial and for $d = a$ this follows directly from (a) and the construction of $\mathcal{K}'$ and Σ'.

(b) $\Rightarrow$ (c) is trivial.

(c) $\Rightarrow$ (a) We use Theorem 4. There exists $k \geq 0$ such that for $G_k = (\mathcal{U}_{\mathcal{K}'}, \rho)^{\downarrow \mathcal{EL},k}_{|\Sigma'}$

$$(\mathcal{U}_{\mathcal{K}'}, \rho) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',G_k}, \rho_{G_k}).$$

Then a, b are in the domain of G_k and

$$(\mathcal{U}_{\mathcal{K}'}, a) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',G_k}, d)$$

for either $d = a$ or $d = b$. But the assumption $d = a$ leads to a contradiction as there is an infinite w-chain starting at a in $\mathcal{U}_{\mathcal{K}'}$ but there is no such w-chain starting at a in $\mathcal{U}_{\mathcal{T}',G_k}$ because no B_i, $i = 1, 2$, is satisfied in any node of G_k in the subtree generated by a (since $B_1 \notin \Sigma'$) and w-chains are only generated by the TBox $\mathcal{T}'$ from nodes satisfying at least one B_i. It follows that $d = b$. Then

$$(\mathcal{U}_{\mathcal{T}',G_k}, a) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{T}',G_k}, b) \preceq_{\mathcal{EL},\Sigma'} (\mathcal{U}_{\mathcal{K}'}, b)$$

and (c) follows by construction of $\mathcal{K}'$.

For $\mathcal{ELI}$ we modify the construction of $\mathcal{T}'$ in exactly the same way as in the proof of Theorem 8. Thus, we construct $\mathcal{T}'$ and $\mathcal{A}'$ as before but add to $\mathcal{T}'$ the CIs $B_2 \sqsubseteq \exists v^-.B_3$ and

$$B_3 \sqsubseteq X \sqcap \bigsqcap_{s \in \mathsf{role}(\mathcal{T}')} \exists s.B_3,$$

where $\mathsf{role}(\mathcal{T}')$ denotes the set of all role names and their inverses used in $\mathcal{T}'$ and $X = \bigsqcap_{A \in \mathsf{sub}(\mathcal{T}')} A$. We also add the concept name B_3 to Σ'. ❏

For the upper bounds, we establish the following Lemma, relying on the encoding of tree-shaped interpretations used in Section "Proofs for Section 6". We concentrate on existence since verification can be reduced, see Theorem 11.

Lemma 14 *Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be an $\mathcal{ELI}$ knowledge base, $a \in \mathsf{ind}(\mathcal{A})$, and Σ a signature, and let N be the outdegree of $\mathcal{U}_\mathcal{K}$. Then, there is an NTA $\mathfrak{A}$ such that for every N^2-ary Θ-labeled tree (T, L) with $\mathcal{I}_L \models \mathcal{T}$, we have:*

$$(T, L) \in L(\mathfrak{A}) \quad \textit{iff} \quad \mathcal{U}_\mathcal{K}, a \preceq_{\mathcal{ELI},\Sigma} \mathcal{I}_L, \varepsilon.$$

Moreover, $\mathfrak{A}$ can be constructed in time double exponential in $|\mathcal{K}|$ in general, and exponential in $|\mathcal{K}|$ if Σ is full.

Proof. We start with the general case. We denote with $\mathsf{rol}_\Sigma(\mathcal{T})$ the set of all Σ-roles that occur in $\mathcal{T}$, and with $\mathsf{tp}(\mathcal{T})$ the set of all types for $\mathcal{T}$. We do not specify the NTA directly, but rather go via two-way alternating automata (TWAPA), see Section "Proofs for Section 6" for precise definitions. The set Q of states of $\mathfrak{A} = (Q, \Theta, q_0, \Delta, \Omega)$ is defined as

$$Q = \{q_x, q_{r,x}, p_{r,x} \mid x \in \mathsf{ind}(\mathcal{A}) \cup \mathsf{tp}(\mathcal{T}), r \in \mathsf{rol}_\Sigma(\mathcal{T})\},$$

and the initial state is q_a. For $b \in \mathsf{ind}(\mathcal{A})$, we denote with t_b the type of b in $\mathcal{U}_\mathcal{K}$, that is,

$$t_b = \{C \in \mathsf{sub}(\mathcal{T}) \mid b \in C^{\mathcal{U}_\mathcal{K}}\}.$$

The transition function δ of $\mathfrak{A}$ assigns $\delta(q_t, \theta) = \mathsf{false}$ and $\delta(q_b, \theta) = \mathsf{false}$ whenever $t \cap \Sigma \not\subseteq \theta$ and $t_b \cap \Sigma \not\subseteq \theta$, respectively. Otherwise, we set:

$$\delta(q_t, \theta) = \bigwedge_{\substack{t \rightsquigarrow_r t', \\ r \in \mathsf{rol}_\Sigma(\mathcal{T})}} \big((0, p_{r,t'}) \vee \bigvee_{1 \leq i \leq N^2} (i, q_{r,t})\big)$$

$$\delta(q_b, \theta) = (0, q_{t_b}) \wedge \bigwedge_{\substack{(b,b') \in r^{\mathcal{U}_\mathcal{K}}, \\ r \in \mathsf{rol}_\Sigma(\mathcal{T})}} \big((0, p_{r,b'}) \vee \bigvee_{1 \leq i \leq N^2} (i, q_{r,b'})\big)$$

The transition function for states of the form $p_{r,x}, q_{r,x}$ is defined by taking, for all $\theta \in \Theta$:

$$\begin{array}{ll} \delta(p_{r,x}, \theta) = \mathsf{false} & \text{if } r^- \notin \theta \\ \delta(p_{r,x}, \theta) = (-1, q_x) & \text{if } r^- \in \theta \\ \delta(q_{r,x}, \theta) = \mathsf{false} & \text{if } r \notin \theta \\ \delta(q_{r,x}, \theta) = (0, q_x) & \text{if } r \in \theta \end{array}$$

It is not hard to verify that $(T, L) \in L(\mathfrak{A})$ iff $\mathcal{U}_{\mathcal{K},a} \preceq_{\mathcal{ELI},\Sigma} \mathcal{I}_L, \varepsilon$, for all (T, L), even without the assumption $\mathcal{I}_L \models \mathcal{T}$. Moreover, the size of $\mathfrak{A}$ is exponential in $|\mathcal{K}|$, and it can be computed in exponential time.

For the full signature case, we obtain $\mathfrak{A}'$ from $\mathfrak{A}$ by dropping the states of the shape $q_t, q_{r,t}, p_{r,t}$, for all types t. The transition function δ' is obtained from δ given above by replacing every atom (i, q) for a dropped state q with true. It is routine to verify that $(T, L) \in L(\mathfrak{A}')$ iff $\mathcal{U}_{\mathcal{K},a} \preceq_{\mathcal{ELI}} \mathcal{I}_L, \varepsilon$, whenever $\mathcal{I}_L \models \mathcal{T}$. Moreover, the size of $\mathfrak{A}'$ is polynomial in $|\mathcal{K}|$, and it can be computed in exponential time.

The required NTAs can now be obtained from $\mathfrak{A}$, $\mathfrak{A}'$ by the standard translation of TWAPAS to NTAs, incurring an exponential blow-up (Vardi 1998). ❏

To finish the upper bound from Theorem 15, let $\mathfrak{B}$ be the projection of $\mathfrak{B}_0$ from Lemma 12 to its second component. It is routine to verify that the $\mathcal{ELI}(\Sigma)$-MSC of a w.r.t. $\mathcal{K}$ exists iff the intersection of $\mathfrak{A}$ and $\mathfrak{B}$ is non-empty.

Since all automata can be computed in double exponential time, and since intersection (non-)emptiness can be verified in PTIME, the 2-EXPTIME upper bound follows. For the full signature case, we obtain an EXPTIME upper bound since in this case $\mathfrak{A}$ and $\mathfrak{B}$ can be computed in exponential time.