Deciding Inseparability and Conservative Extensions in the Description Logic $\mathcal{EL}$

Carsten Lutz

Fachbereich Mathematik und Informatik, Universität Bremen, Germany

Frank Wolter

Department of Computer Science, University of Liverpool, UK

Abstract

We study the problem of deciding whether two ontologies are inseparable w.r.t. a signature Σ, i.e., whether they have the same consequences in the signature Σ. A special case is to decide whether an extension of an ontology is conservative. By varying the language in which ontologies are formulated and the query language that is used to describe consequences, we obtain different versions of the problem. We focus on the lightweight description logic $\mathcal{EL}$ as an ontology language, and consider query languages based on (i) subsumption queries, (ii) instance queries over ABoxes, (iii) conjunctive queries over ABoxes, and (iv) second-order logic. For query languages (i) to (iii), we establish ExpTime-completeness of both inseparability and conservative extensions. Case (iv) is equivalent to a model-theoretic version of inseparability and conservative extensions, and we prove it to be undecidable. We also establish a number of robustness properties for inseparability.

Key words: Description logic, ontologies, conservative extension, modularity.

1. Introduction

The main use of ontologies in computer science is to provide a reference vocabulary for some domain of interest. In logic-based ontology languages such as description logics (DLs), this vocabulary is represented as predicate symbols whose meaning is formalized using (a finite axiomatization of) a logical theory (2) formulated in these symbols.

⋆ This research was partly supported by EPSRC grant EP/E065279/1.

Email addresses: `clu@informatik.uni-bremen.de` (Carsten Lutz), `frank@csc.liv.ac.uk` (Frank Wolter).

URLs: `http://informatik.uni-bremen.de/tdki/` (Carsten Lutz), `http://www.csc.liv.ac.uk/~frank/` (Frank Wolter).

Recent applications of ontologies, such as in health care and the bio-sciences, have led to the development of very large ontologies that capture an extensive vocabulary. Notable examples include the Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT), which comprises almost 0.5 million vocabulary items (22); and the thesaurus of the US national cancer institute (NCI), which comprises more than 40.000 such items (20). The design, maintenance, and customization of ontologies of this size are highly non-trivial tasks that are supported by various tool suites, many of which are based on DL reasoning systems.

Currently, the main service provided by such systems is to compute subsumption, a basic reasoning service that helps to make explicit the structure of the vocabulary. While being very useful, subsumption alone does not suffice to support the complex engineering patterns used in the design and customization of large-scale ontologies. In particular, subsumption provides only limited support for complex operations such as the import, merging, combination, re-use, refinement, and extension of ontologies. The consequences of these operations are hard to analyze and easily introduce unintended changes to the logical theory that describes the vocabulary. Therefore, additional tool support is required to identify theory changes. We give two concrete examples:

Ontology refinement. Suppose an ontology designer wants to extend an ontology with new axioms that refine the description of a particular part Σ of the vocabulary. In this case, he usually intends to preserve the theory (and thus the meaning) of most or all of the non-Σ-symbols. For example, when a medical ontology is extended to refine the axiomatization of the vocabulary for X-ray diagnostics, the theory that describes the vocabulary of anatomy and drugs are not expected to change. Thus, an appropriate reasoning service is to check for such unexpected theory change, and to report it to the designer.

Ontology import. Suppose an ontology designer wants to import an existing ontology into the one he is currently designing. For example, a medical ontology might be imported into an ontology about the health-care regulations of a particular country. It is then typically intended to use the vocabulary Σ of the imported ontology with its original meaning. However, if the symbols from Σ are used to define new symbols in the importing ontology, it may happen that new consequences about Σ become derivable and thus the Σ-theory changes. As in the previous example, reasoning support should identify such theory changes and report them to the user.

In this paper, we propose *Σ-inseparability of two ontologies* as a fundamental notion for addressing problems of this kind. In short, two ontologies are Σ-inseparable if they have the same logical consequences formulated in the signature (vocabulary) Σ. For the operations on ontologies mentioned above, checking for Σ-inseparability is a central reasoning service. Additionally, Σ-inseparability plays a fundamental role in defining notions of a module inside an ontology. While we do not directly address modularity in this paper, we note that understanding Σ-inseparability is crucial for any approach to modularity: an ontology module should be independent from its host ontology, and thus Σ-inseparable from the overall ontology regarding its own vocabulary Σ (10; 12; 16). We also note that conservative extensions are the special case of Σ-inseparability where one ontology is included in the other. Like Σ-inseparability, conservative extensions have been proposed as a useful reasoning service for ontologies and were used to formalize modularity (1; 12; 10; 14; 18).

Above, we have defined Σ-inseparability of two ontologies in terms of their logical consequences, but we have not made explicit the logical language that is used to formulate these consequences. From now on, we call this language the *query language* and say that two ontologies are Σ-inseparable w.r.t. a query language $\mathcal{QL}$ iff they have as consequences the same $\mathcal{QL}$ queries that use only symbols from Σ. When studying conservative extensions between logical theories in mathematical logic, the query language typically coincides with the language in which the theories are formulated. In DLs, ontologies are formulated as TBoxes, which are sets of concept inclusions. In analogy with mathematical logic, one can thus define Σ-inseparability of two DL TBoxes based on the query language that consists of all concept inclusions. Indeed, this is useful for applications in which the user is mainly interested in subsumption between concepts, and it is one of the choices that we consider in this paper.

In other applications, concept inclusions are not appropriate as a query language for Σ-inseparability. An important example is the use of an ontology to access instance data stored in an ABox using as a query mechanism either instance retrieval or conjunctive query answering. In this case, the query language on which Σ-inseparability is based should ensure that two ontologies are Σ-inseparable iff they give the same answers to any (instance or conjunctive) query over any possible ABox. We will show that the query language based on concept inclusions is too weak for this purpose, and introduce two additional query language that can be used to define appropriate notions of Σ-inseparability: one based on instance retrieval and one based on conjunctive query answering. Finally, we also consider full second-order logic as a query language. The resulting notion of Σ-inseparability is equivalent to a model-theoretic version in which two ontologies are Σ-inseparable iff the classes of Σ-reducts of their models coincide. This notion has been extensively investigated in the context of modular software design (13; 19).

We study the following three aspects of Σ-inseparability:

(i) The computational complexity of deciding Σ-inseparability of two ontologies.

(ii) The relation between different versions of Σ-inseparability, which are obtained from the different query languages discussed above.

(iii) Robustness properties, which guarantee that Σ-inseparability is preserved under natural modifications of the ontologies and signatures involved.

The notions of Σ-inseparability defined in this paper can be used with ontologies formulated in any standard DL. However, the concrete results obtained for Points (i)-(iii) above depend on the choice of the ontology language. In this paper, we concentrate on ontologies formulated in the lightweight description logic $\mathcal{EL}$ (7; 4). This decision is motivated by the fact that many large-scale ontologies, such as those originating in the life sciences, are formulated in $\mathcal{EL}$ or mild extensions thereof. Concrete examples include Snomed CT and (early versions of) the NCI ontology.

The central result of this paper is that deciding Σ-inseparability and conservative extensions is ExpTime-complete for the three versions of Σ-inseparability derived from DL query languages (concept inclusions, instance retrieval, conjunctive queries). For inseparability based on second-order logic (equivalently, model-theoretic inseparability), we prove undecidability. We also show that (a) inseparability based on concept inclusions coincides with inseparability based on instance retrieval, and (b) inseparability based on conjunctive queries coincides with inseparability based on concept inclusions that are formulated in an extension of $\mathcal{EL}$ with the universal role. Finally, we postulate two robustness properties and show that all versions of Σ-inseparability considered in this paper enjoy these properties.

This paper is organized as follows. In Section 2, we introduce inseparability and query languages, state the relationships between different versions of inseparability, and introduce and analyze robustness properties. Section 3 introduces some technical tools that we use extensively in the remainder of the paper, namely simulations, canonical models, and local entailment. An ExpTime upper bound for Σ-entailment based on concept inclusions is established in Section 4. In Section 5, we prove the relationships between different notions of inseparability as stated in Section 2 and use them to prove the ExpTime upper bound for inseparability based on instance retrieval and conjunctive query answering. A matching lower bound, which applies already to the case of conservative extensions, is established in Section 6. Undecidability of inseparability based in second-order logic is proved in Section 7. Finally, we discuss some open questions in Section 8. To improve readability, many proof details are deferred to the appendix.

2. Preliminaries

We introduce the description logic $\mathcal{EL}$ as well as (different versions of) inseparability and the related notions of entailment and conservative extensions. We also describe the relationship between the different versions of inseparability and introduce and investigate two robustness properties.

2.1. The Description Logic $\mathcal{EL}$

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

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

where A ranges over $\mathsf{N_C}$, r ranges over $\mathsf{N_R}$, and C, D range over $\mathcal{EL}$-concepts. The semantics of $\mathcal{EL}$ is defined by means of *interpretations* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where the interpretation *domain* $\Delta^\mathcal{I}$ is a non-empty set, and $\cdot^\mathcal{I}$ is a function mapping each concept name A to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$, each role name r to a binary relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$, and each individual name a to an element $a^\mathcal{I} \in \Delta^\mathcal{I}$. The function $\cdot^\mathcal{I}$ is inductively extended to arbitrary concepts by setting

$$\begin{aligned} \top^\mathcal{I} &:= \Delta^\mathcal{I} \\ (C \sqcap D)^\mathcal{I} &:= C^\mathcal{I} \cap D^\mathcal{I} \\ (\exists r.C)^\mathcal{I} &:= \{d \in \Delta^\mathcal{I} \mid \exists e \in C^\mathcal{I} : (d,e) \in r^\mathcal{I}\}. \end{aligned}$$

An *$\mathcal{EL}$-TBox* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C and D are $\mathcal{EL}$-concepts. We write $C \doteq D$ as abbreviation for the two CIs $C \sqsubseteq D$ and $D \sqsubseteq C$. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$, written $\mathcal{I} \models C \sqsubseteq D$, if $C^\mathcal{I} \subseteq D^\mathcal{I}$. $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$ if it satisfies all CIs in $\mathcal{T}$. As an example, here is a simple $\mathcal{EL}$-TBox $\mathcal{T}_1$:

$$\begin{aligned}
\mathsf{Mother} &\doteq \mathsf{Female} \sqcap \exists \mathsf{has\text{-}child}.\mathsf{Human} \\
\mathsf{Father} &\doteq \mathsf{Male} \sqcap \exists \mathsf{has\text{-}child}.\mathsf{Human} \\
\mathsf{Male} &\sqsubseteq \mathsf{Human} \\
\mathsf{Female} &\sqsubseteq \mathsf{Human}
\end{aligned}$$

When introducing entailment, inseparability, and conservative extensions in the subsequent section, it is important to be precise about the concept and role names that occur in a concept or TBox. We use the notion of a *signature*, which is a finite subset of $\mathsf{N_C} \cup \mathsf{N_R}$. The signature $\mathsf{sig}(C)$ of a concept C is the set of concept and role names that occur in C, and likewise for the signature $\mathsf{sig}(\mathcal{T})$ of a TBox $\mathcal{T}$. If $\mathsf{sig}(C) \subseteq \Sigma$, we call C an $\mathcal{EL}_\Sigma$*-concept.*

In description logic, an important way to query a TBox is subsumption (2). For two $\mathcal{EL}$-concepts C, D and a TBox $\mathcal{T}$, we say that C is *subsumed* by D w.r.t. $\mathcal{T}$ (written $\mathcal{T} \models C \sqsubseteq D$) iff all models of $\mathcal{T}$ satisfy the CI $C \sqsubseteq D$. Thus, a *subsumption query* is a concept implication $C \sqsubseteq D$. *Subsumption query answering* means to decide whether $\mathcal{T} \models C \sqsubseteq D$, given the query $C \sqsubseteq D$ and the TBox $\mathcal{T}$. For example, reconsider the above TBox $\mathcal{T}_1$. It is easy to see that $\mathcal{T}_1 \models \mathsf{Mother} \sqsubseteq \mathsf{Human}$.

2.2. Entailment, Inseparability, Conservative Extensions

We introduce the three main notions studied in this paper: entailment between TBoxes, which is the most basic notion; inseparability, which is defined in terms of entailment; and conservative extensions, which are a special case of inseparability. All these notions depend on the query language that is used to query a TBox. Subsumption queries are one possible choice, but we shall also consider other options. To treat such query languages in a uniform way, we adopt a rather general view on them: in what follows, a *query language* is a set of sentences of second-order logic with second-order variables for unary and binary relations, and in the signature consisting of the set $\mathsf{N_C}$ of unary predicates and the set $\mathsf{N_R}$ of binary predicates.

Just like queries, $\mathcal{EL}$-TBoxes can also be viewed in the framework of second-order logic. The following well-known inductive translation (2) transforms $\mathcal{EL}$-concepts C into an equivalent first-order formula with one free variable x:

$$\begin{aligned}
A^\sharp &= A(x) \\
(C \sqcap D)^\sharp &= C^\sharp \sqcap D^\sharp \\
(\exists r.C)^\sharp &= \exists y.(r(x,y) \wedge C^\sharp(y/x))
\end{aligned}$$

A concept inclusion $C \sqsubseteq D$ thus corresponds to a first-order sentence $\forall x.(C^\sharp \Rightarrow D^\sharp)$, and a TBox to a conjunction of such sentences. From now on, we will not distinguish between $\mathcal{EL}$-concepts and their translation into first-order logic, and likewise for concept inclusions and TBoxes. Thus, it makes sense to write $\mathcal{T} \models \varphi$ for an $\mathcal{EL}$-TBox $\mathcal{T}$ and a second-order sentence φ to denote second-order entailment. As usual, the signature $\mathsf{sig}(\varphi)$ of a second-order sentence φ is defined as the set of predicates used in it.

Definition 1 (Entailment, inseparability, conservative extension). Let $\mathcal{QL}$ be a query language, Σ a signature, and $\mathcal{T}_1$, $\mathcal{T}_2$ TBoxes. Then

- $\mathcal{T}_1$ Σ-*entails* $\mathcal{T}_2$ w.r.t. $\mathcal{QL}$, written $\mathcal{T}_1 \sqsubseteq_{\Sigma}^{\mathcal{QL}} \mathcal{T}_2$, if $\mathcal{T}_2 \models \varphi$ implies $\mathcal{T}_1 \models \varphi$ for all sentences $\varphi \in \mathcal{QL}$ with $\mathsf{sig}(\varphi) \subseteq \Sigma$;
- $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-*inseparable* w.r.t. $\mathcal{QL}$ if $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ and $\mathcal{T}_2$ Σ-entails $\mathcal{T}_1$;
- $\mathcal{T}_2$ is a Σ-*conservative extension* of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}$ if $\mathcal{T}_2 \supseteq \mathcal{T}_1$ and $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-inseparable w.r.t. $\mathcal{QL}$;
- $\mathcal{T}_2$ is a *conservative extension* of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}$ if $\mathcal{T}_2$ is a Σ-conservative extension of $\mathcal{T}_2$ w.r.t. $\mathcal{QL}$, with $\Sigma = \mathsf{sig}(\mathcal{T}_1)$.

A $\mathcal{QL}$-sentence φ is a *witness* for the non-entailment $\mathcal{T}_1 \not\sqsubseteq_{\Sigma}^{\mathcal{QL}} \mathcal{T}_2$ if $\mathsf{sig}(\varphi) \subseteq \Sigma$, $\mathcal{T}_1 \not\models \varphi$, and $\mathcal{T}_2 \models \varphi$.

The notions of Σ-inseparability, Σ-conservative extensions, and conservative extensions are all defined in terms of Σ-entailment. When developing algorithms, we may thus concentrate on Σ-entailment. Only when giving counterexamples and complexity lower bounds, we will consider conservative extensions as the most special case.

We now give three examples of query languages, all based on subsumption. First, the simple language $\mathcal{QL}_{\mathsf{CN}}$ consists of all concept inclusions $A \sqsubseteq B$, with A and B concept *names* or the top concept $\top$. This query language is useful if we are only interested in the *classification* of TBoxes, i.e., the partial order on the concept names in the TBox induced by the subsumption relation. Indeed, two TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ over a signature Σ have the same classification if and only if they are Σ-inseparable w.r.t. $\mathcal{QL}_{\mathsf{CN}}$. Similarly, if $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}_{\mathsf{CN}}$, then $\mathcal{T}_2$ only extends the existing classification of $\mathcal{T}_1$ with new classes, but does not change it in any other way. Reconsider the example TBox $\mathcal{T}_1$ from Section 2.1, and let $\mathcal{T}_2$ be $\mathcal{T}_1$ extended with the following:

$$\exists \mathsf{has\text{-}child}.\mathsf{Human} \sqsubseteq \mathsf{Parent}$$
$$\mathsf{Parent} \sqsubseteq \mathsf{Human}.$$

Then $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}_{\mathsf{CN}}$ because the only new inclusion $A \sqsubseteq B$, where A, B are concept names, derivable from $\mathcal{T}_2$ is $\mathsf{Parent} \sqsubseteq \mathsf{Human}$ but Parent is not in the signature of $\mathcal{T}_1$. It is easy to decide Σ-entailment w.r.t. $\mathcal{QL}_{\mathsf{CN}}$ by computing all subsumptions between the (finitely many) concept names from Σ.

Second and more interesting, the language $\mathcal{QL}_{\mathcal{EL}} \supsetneq \mathcal{QL}_{\mathsf{CN}}$ consists of all concept inclusions $C \sqsubseteq D$ between (possibly composite) $\mathcal{EL}$-concepts C and D. Intuitively, $\mathcal{QL}_{\mathcal{EL}}$ is appropriate if we are interested not only in the classification of a TBox, but in *all* consequences of the TBox that can be expressed in terms of concept inclusions. It is easy to see that Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ implies Σ-entailment w.r.t. $\mathcal{QL}_{\mathsf{CN}}$. The converse is not true: take the example TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ from above. Then $\mathcal{T}_2$ is *not* a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$, a witness being

$$\exists \mathsf{has\text{-}child}.\mathsf{Human} \sqsubseteq \mathsf{Human}.$$

Deciding Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ is much less trivial than w.r.t. $\mathcal{QL}_{\mathsf{CN}}$, and we will study this problem in detail in the main part of this paper. For brevity, we write $\mathcal{T}_1 \sqsubseteq_{\Sigma} \mathcal{T}_2$ if $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$.

We can define other query languages $\mathcal{QL}_{\mathcal{L}}$ by replacing the $\mathcal{EL}$-concepts in $\mathcal{QL}_{\mathcal{EL}}$ with concepts formulated in another description logic $\mathcal{L}$, i.e., $\mathcal{QL}_{\mathcal{L}}$ consists of all concept implications $C \sqsubseteq D$ with C and D $\mathcal{L}$-concepts. In general, different choices of $\mathcal{L}$ give rise to distinct notions of Σ-entailment w.r.t. $\mathcal{QL}_\mathcal{L}$. As our third example, we consider the case $\mathcal{L} = \mathcal{ALC}$, where $\mathcal{ALC}$ is the extension of $\mathcal{EL}$ with a negation constructor $\neg C$ which has the obvious semantics $(\neg C)^\mathcal{I} = \Delta^\mathcal{I} \setminus C^\mathcal{I}$. We use $\forall r.C$ to abbreviate $\neg\exists r.\neg C$. Consider the TBoxes

$$\begin{aligned} \mathcal{T}_1 : \qquad \mathsf{Human} &\sqsubseteq \exists \mathsf{eats}.\top \\ \mathsf{Plant} &\sqsubseteq \exists \mathsf{grows\text{-}in}.\mathsf{Area} \\ \mathsf{Vegetarian} &\sqsubseteq \mathsf{Healthy} \end{aligned}$$

$$\begin{aligned} \mathcal{T}_2 : \qquad \mathsf{Human} &\sqsubseteq \exists \mathsf{eats}.\mathsf{Food} \\ \mathsf{Food} \sqcap \mathsf{Plant} &\sqsubseteq \mathsf{Vegetarian} \end{aligned}$$

where $\mathcal{T}_2$ additionally contains all the CIs of $\mathcal{T}_1$. Then, $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}_\mathcal{EL}$, as can be checked using the semantic criteria introduced below. However, $\mathcal{T}_2$ is not a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}_\mathcal{ALC}$, as witnessed by

$$\mathsf{Human} \sqcap \forall \mathsf{eats}.\mathsf{Plant} \sqsubseteq \exists \mathsf{eats}.\mathsf{Vegetarian}.$$

For deciding Σ-conservative extensions (of $\mathcal{EL}$-TBoxes) w.r.t. $\mathcal{QL}_\mathcal{ALC}$, we can use the algorithm for deciding conservative extensions in $\mathcal{ALC}$, given in (14). As the above example shows, this algorithm cannot be used to decide Σ-conservative extensions w.r.t. $\mathcal{QL}_\mathcal{EL}$.

2.3. ABoxes and Conjunctive Queries

In some applications, queries are asked against knowledge bases rather than TBoxes. Such a knowledge base enriches a TBox with instance data, stored in an ABox.

Let $\mathsf{N_I}$ be a countably infinite set of *individual names*. An *$\mathcal{EL}$-ABox* is a finite set of assertions of the form $C(a)$ and $r(a, b)$, where C is an $\mathcal{EL}$-concept, r a role, and $a, b \in \mathsf{N_I}$. An *$\mathcal{EL}$-knowledge base (KB)* is a pair $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ consisting of an $\mathcal{EL}$-TBox and an $\mathcal{EL}$-ABox. To interpret ABoxes, we consider interpretations $\mathcal{I}$ which additionally assign to each $a \in \mathsf{N_I}$ an element $a^\mathcal{I} \in \Delta^\mathcal{I}$. An interpretation $\mathcal{I}$ *satisfies* an assertion $C(a)$ if $a^\mathcal{I} \in C^\mathcal{I}$ and an assertion $r(a, b)$ if $(a^\mathcal{I}, b^\mathcal{I}) \in r^\mathcal{I}$. If α is an ABox assertion and $\mathcal{I}$ satisfies α, we write $\mathcal{I} \models \alpha$. $\mathcal{I}$ is a model of an ABox $\mathcal{A}$ if it satisfies all assertions in $\mathcal{A}$. It is a model of a KB $(\mathcal{T}, \mathcal{A})$ if it is a model of both $\mathcal{T}$ and $\mathcal{A}$. The *signature* of an ABox $\mathcal{A}$ is defined as the set of concept and role names occurring in $\mathcal{A}$. Observe that individual names are not part of the signature.

When working with knowledge bases, there are several options for querying. In this paper, we consider the two most important ones: instance retrieval and conjunctive query answering. For an $\mathcal{EL}$-concept C, a KB $\mathcal{K} = (\mathcal{T}, \mathcal{A})$, and an individual name a, we say that a is an *instance* of C w.r.t. $\mathcal{K}$ (written $\mathcal{K} \models C(a)$) if all models of $\mathcal{K}$ satisfy the assertion $C(a)$. Now, an *instance query* is a concept C and *instance query answering* means, given the query C and a knowledge base $\mathcal{K}$, to produce all *answers* to C w.r.t. $\mathcal{K}$, i.e., all $a \in \mathsf{N_I}$ such that $\mathcal{K} \models C(a)$.

A *conjunctive query* is an expression of the form $q = \exists \mathbf{y}.\psi(\mathbf{x}, \mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ are sequences of variables taken from a fixed and countably infinite set of variables $\mathsf{N_V}$, and ψ is a conjunction of atoms $C(v)$ and $r(u, v)$ with C an $\mathcal{EL}$-concept, r a role name, and $u, v \in \mathbf{x} \cup \mathbf{y}$. The variables in $\mathbf{x}$ are called *answer variables*, and those in $\mathbf{y}$ *bound variables*. To make the answer variables in q explicit, we write $q(\mathbf{x})$. The *signature* of a conjunctive query is defined as the set of concept and role names occurring in it.

Let $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ be a knowledge base, $q = \exists y_1, \dots, y_m.\psi$ a conjunctive query with answer variables $\mathbf{x} = x_1, \dots, x_n$, and $\mathbf{a} = a_1, \dots, a_n$ a sequence of individual names. Then $\mathbf{a}$ is an *answer* to q w.r.t. $\mathcal{K}$, written $\mathcal{K} \models q(\mathbf{a})$, if for every model $\mathcal{I}$ of $\mathcal{K}$, there exists a mapping $\tau : \mathsf{N_V} \rightarrow \Delta^{\mathcal{I}}$ such that

- $\tau(x_i) = a_i$, for $1 \leq i \leq n$;
- for every atom $C(v) \in q$, $\tau(v) \in C^{\mathcal{I}}$;
- for every atom $r(u, v) \in q$, $(\tau(u), \tau(v)) \in r^{\mathcal{I}}$.

Conjunctive query answering means, given C, $\mathcal{K}$, and q, to produce all answers to q w.r.t. $\mathcal{K}$.

In most applications, the instance data in an ABox has a different status than the conceptual knowledge in the TBox. Often, the TBox is developed while the ABox is not yet known. Moreover, even if an initial ABox is known, the ABox usually changes frequently over the lifespan of an application. Therefore, to analyze the consequences of changes to TBoxes, we quantify over *all* possible ABoxes that could possibly be used together with the TBox.

We now define the corresponding notions of Σ-entailment.

Definition 2. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be $\mathcal{EL}$-TBoxes and Σ a signature. Then

- $\mathcal{T}_1$ Σ-*entails* $\mathcal{T}_2$ *w.r.t.* $\mathcal{QL}^i_{\mathcal{EL}}$ iff the following holds for all Σ-ABoxes $\mathcal{A}$, Σ-concepts C, and $a \in \mathsf{N_I}$:
$$(\mathcal{T}_2, \mathcal{A}) \models C(a) \Rightarrow (\mathcal{T}_1, \mathcal{A}) \models C(a).$$
- $\mathcal{T}_1$ Σ-*entails* $\mathcal{T}_2$ *w.r.t.* $\mathcal{QL}^q_{\mathcal{EL}}$ iff the following holds for all Σ-ABoxes $\mathcal{A}$, conjunctive Σ-queries q with k free variables, and k-tuples $\mathbf{a}$ of individual names in $\mathsf{N_I}$:
$$(\mathcal{T}_2, \mathcal{A}) \models q(\mathbf{a}) \Rightarrow (\mathcal{T}_1, \mathcal{A}) \models q(\mathbf{a}).$$

In this definition, the terms "Σ-entails w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$" and "$\Sigma$-entails w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$" are defined without specifying the query languages $\mathcal{QL}^i_{\mathcal{EL}}$ and $\mathcal{QL}^q_{\mathcal{EL}}$. It is not hard, however, to define first-order query languages $\mathcal{QL}^i_{\mathcal{EL}}$ and $\mathcal{QL}^q_{\mathcal{EL}}$ that are compatible with Definition 2 and fit into the schema of query languages used in Definition 1. We only consider the case of instance retrieval and leave the conjunctive query case to the reader. For every individual name $a \in \mathsf{N_I}$, fix a variable x_a. Then an ABox $\mathcal{A}$ can be translated to a first-order formula

$$\mathcal{A}^\sharp := \bigwedge_{C(a) \in \mathcal{A}} C^\sharp(x_a) \wedge \bigwedge_{r(a,b) \in \mathcal{A}} r(x_a, x_b),$$

and an assertion $C(a)$ into a first-order formula $C^\sharp(x_a)$. The query language $\mathcal{QL}^i_{\mathcal{EL}}$ is now defined as the set of all first-order sentences $\forall \mathbf{x}, x_a.(\mathcal{A}^\sharp \rightarrow C^\sharp(x_a))$ with $\mathbf{x}$ the set of all variables in $\mathcal{A}^\sharp$.

In applications in which the ABox does not change frequently, it can also make sense to consider entailment and inseparability between knowledge bases instead of between TBoxes. In this case, the ABox is part of the two theories that are compared, and not universally quantified as in Definition 2. This problem turns out to be computationally much simpler. In fact, tractability of inseparability of knowledge bases will be a corollary of our investigation of inseparability for TBoxes, see Definition 25 and Lemma 29 below.

2.4. Relating Query Languages

We discuss the relationship between the query languages $\mathcal{QL}_{\mathcal{EL}}$, $\mathcal{QL}^i_{\mathcal{EL}}$, and $\mathcal{QL}^q_{\mathcal{EL}}$. It is not hard to see that

(1) Σ-entailment w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$ implies Σ-entailment w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$, and

(2) Σ-entailment w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$ implies Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$.

Indeed, (1) holds since every instance query C can be seen as a conjunctive query $C(v)$, and (2) follows from the fact that $\mathcal{T} \models C \sqsubseteq D$ iff $\mathcal{K} \models D(a)$ with $\mathcal{K} = (\mathcal{T}, \{C(a)\})$.

Now for the converses of (1) and (2). Somewhat surprisingly, the converse of (2) is true, and we will prove this in Section 5.1. In contrast, the converse of (1) is false. To see this, consider the TBox $\mathcal{T}_1$ from Section 2.1, and let $\mathcal{T}_2$ be $\mathcal{T}_1$ extended with the old-fashioned statement

$$\mathsf{Father} \sqsubseteq \exists \mathsf{spouse}.\mathsf{Female}.$$

Then, $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$, but not w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$, as witnessed by the ABox $\{\mathsf{Father}(a)\}$ and the query $\exists v.\mathsf{Female}(v)$.

Interestingly, there is a moderate extension of $\mathcal{QL}_{\mathcal{EL}}$ that is still based on subsumption queries, and for which Σ-entailment coincides with Σ-entailment in $\mathcal{QL}^q_{\mathcal{EL}}$. Let u be a fresh role name not in $\mathsf{N_R}$, and call it the *universal role*. The set of $\mathcal{EL}^u$*-concepts* consists of all $\mathcal{EL}$-concepts C and all concepts of the form $\exists u.C$, where C is an $\mathcal{EL}$-concept. Note that we do not allow nesting of the $\exists u.C$ constructor inside any constructor. Interpretations $\mathcal{I}$ are required to interpret the universal role as $u^{\mathcal{I}} = \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The role name u is not part of the signature of any concept, hence $\mathsf{sig}(C) = \mathsf{sig}(\exists u.C)$ for any $\mathcal{EL}$-concept C, and similarly for concept inclusions. Observe that the signature of the first-order translation $(\exists u.C)^\sharp = \exists x.C^\sharp(x)$ of $\exists u.C$ coincides with the signature of $\exists u.C$. $\mathcal{EL}^u$-concepts C with $\mathsf{sig}(C) \subseteq \Sigma$ are called $\mathcal{EL}^u_\Sigma$-concepts.

The query language $\mathcal{QL}^u_{\mathcal{EL}}$ consists of all concept inclusions $C \sqsubseteq D$ such that C is an $\mathcal{EL}$-concept and D an $\mathcal{EL}^u$-concept. Clearly, Σ-entailment w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$ implies Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$. To show that the converse does not hold, we can reuse the example from above showing that the converse of (2) fails. In fact, the subsumption $\mathsf{Father} \sqsubseteq \exists u.\mathsf{Female}$ is a witness for the fact that $\mathcal{T}_2$ is not a conservative extension of $\mathcal{T}_1$ w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$. Although maybe not too interesting in its own right, the query language $\mathcal{QL}^u_{\mathcal{EL}}$ is one of the central query languages studied in this paper. This is due to the fact that, in Section 5.2, we show that Σ-entailment in $\mathcal{QL}^u_{\mathcal{EL}}$ coincides with Σ-entailment in $\mathcal{QL}^q_{\mathcal{EL}}$. In what follows, we write $\mathcal{T}_1 \sqsubseteq^u_\Sigma \mathcal{T}_2$ to state that $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$.

The following theorem sums up the main results stated in this section. It will be proved in Section 5.

Theorem 3. *The following equivalences hold for any two TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and any signature Σ:*

- $\mathcal{T}_1 \sqsubseteq^i_\Sigma \mathcal{T}_2$ *iff* $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$.
- $\mathcal{T}_1 \sqsubseteq^q_\Sigma \mathcal{T}_2$ *iff* $\mathcal{T}_1 \sqsubseteq^u_\Sigma \mathcal{T}_2$.

Thus, it suffices to study Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ and $\mathcal{QL}^u_{\mathcal{EL}}$. This is what we will do in the following sections.

2.5. Properties of Σ-Inseparability

To use Σ-inseparability in applications such as the ones mentioned in the introduction, it is important to properly understand its behavior. In this section, we postulate two useful robustness properties for Σ-inseparability, and show that they are enjoyed by the notions of Σ-inseparability studied in this paper. The first property is concerned with extensions of the signature Σ by additional symbols.

Definition 4. Let $\mathcal{L}$ be a description logic and $\mathcal{QL}$ a query language. We say that the pair $(\mathcal{L}, \mathcal{QL})$ is *robust for signature extensions* if for all $\mathcal{L}$-TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$, we have the following: if $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ w.r.t. $\mathcal{QL}$, then $\mathcal{T}_1$ Σ'-entails $\mathcal{T}_2$ w.r.t. $\mathcal{QL}$ for every Σ' with $\mathsf{sig}(\mathcal{T}_2) \cap \Sigma' \subseteq \Sigma$.

Robustness under signature extensions is of particular interest for the query languages $\mathcal{QL}^i_{\mathcal{EL}}$ and $\mathcal{QL}^q_{\mathcal{EL}}$. We consider $\mathcal{QL}^q_{\mathcal{EL}}$, the argument for $\mathcal{QL}^i_{\mathcal{EL}}$ is similar. Assume that $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-inseparable w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$. Then the answers to conjunctive Σ-queries q of the KB $(\mathcal{T}_1, \mathcal{A})$ coincide with the answers to q of the KB $(\mathcal{T}_2, \mathcal{A})$, for every Σ-ABox $\mathcal{A}$. Robustness under signature extensions implies that even if the ABox and the query contain additional symbols not occurring in $(\mathsf{sig}(\mathcal{T}_1) \cup \mathsf{sig}(\mathcal{T}_2)) \setminus \Sigma$, the answers still coincide. This property is critical for applications in which it is not possible to restrict ABoxes and conjunctive queries to a fixed signature Σ.

Robustness under signature extensions is closely related to Craig interpolation, a property that is studied in mathematical logic and applied, for example, in the area of modular software specification (13; 24; 19). In this paper, we use Craig interpolation of $\mathcal{EL}$ as established in (21) to prove robustness under vocabulary extensions. We first state the interpolation property of $\mathcal{EL}$.

Theorem 5. *$\mathcal{EL}$ has Craig interpolation: for every TBox $\mathcal{T}$ and $\varphi \in \mathcal{QL}_{\mathcal{EL}}$ with $\mathcal{T} \models \varphi$, there exists a TBox $I(\mathcal{T}, \varphi)$ (called an interpolant of $\mathcal{T}$ and φ) such that $\mathsf{sig}(I(\mathcal{T}, \varphi)) \subseteq \mathsf{sig}(\mathcal{T}) \cap \mathsf{sig}(\varphi)$, $\mathcal{T} \models I(\mathcal{T}, \varphi)$, and $I(\mathcal{T}, \varphi) \models \varphi$.*

Corollary 6. *$(\mathcal{EL}, \mathcal{QL})$ is robust for signature extensions, for all $\mathcal{QL}$ among $\mathcal{QL}_{\mathcal{EL}}$, $\mathcal{QL}^u_{\mathcal{EL}}$, $\mathcal{QL}^i_{\mathcal{EL}}$, and $\mathcal{QL}^q_{\mathcal{EL}}$.*

Proof. By Theorem 3, it is sufficient to prove this result for $\mathcal{QL}_{\mathcal{EL}}$ and $\mathcal{QL}^u_{\mathcal{EL}}$. For $\mathcal{QL}_{\mathcal{EL}}$ the proof is by a straightforward application of the Craig interpolation property of $\mathcal{EL}$. Assume $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$ and $\Sigma \subseteq \Sigma'$ with $\mathsf{sig}(\mathcal{T}_2) \cap \Sigma' \subseteq \Sigma$. Let $\mathcal{T}_2 \models C \sqsubseteq D$, where C, D are $\mathcal{EL}_{\Sigma'}$-concepts. Take an interpolant $I(\mathcal{T}_2, C \sqsubseteq D)$. Then $\mathsf{sig}(I(\mathcal{T}_2, C \sqsubseteq D)) \subseteq \Sigma$. Hence $\mathcal{T}_1 \models I(\mathcal{T}_2, C \sqsubseteq D)$ and this yields $\mathcal{T}_1 \models C \sqsubseteq D$, as required. The proof for $\mathcal{QL}^u_{\mathcal{EL}}$ also uses the Craig interpolation property of $\mathcal{EL}$ and is given in Section B of the appendix. □

It follows from Corollary 6 that when deciding whether $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$, we can w.l.o.g. assume that $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$ because Σ'-entailment follows for all signatures Σ' with $\Sigma' \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$.

We now introduce the second robustness property.

Definition 7. We say that a pair $(\mathcal{L}, \mathcal{QL})$ has the *join-modularity property* if for all TBoxes $\mathcal{T}_1$, $\mathcal{T}_2$, the following holds: if $\mathcal{T}_1$ and $\mathcal{T}_2$ are Σ-inseparable w.r.t. $\mathcal{QL}$ and $\mathsf{sig}(\mathcal{T}_1) \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$, then $\mathcal{T}_1 \cup \mathcal{T}_2$ and $\mathcal{T}_i$ are Σ-inseparable w.r.t. $\mathcal{QL}$, for $i = 1, 2$.

Join-modularity is of interest for collaborative ontology development. For example, assume that two ontology developers extend a given ontology $\mathcal{T}_0$ independently of each other, obtaining extended ontologies $\mathcal{T}_1$ and $\mathcal{T}_2$ with $\mathcal{T}_0 \subseteq \mathcal{T}_i$ and such that $\mathcal{T}_i$ is a conservative extension of $\mathcal{T}_0$, for $i \in \{1, 2\}$. If the two developers worked on different parts of the ontology, it is safe to assume that $\mathsf{sig}(\mathcal{T}_1) \cap \mathsf{sig}(\mathcal{T}_2) = \mathsf{sig}(\mathcal{T}_0)$. Now, the join-modularity property implies that the joint extension $\mathcal{T}_1 \cup \mathcal{T}_2$ is also a conservative extension of $\mathcal{T}_0$: since $\mathcal{T}_1$ and $\mathcal{T}_2$ are conservative extensions of $\mathcal{T}_0$, $\mathcal{T}_1$ and $\mathcal{T}_2$ are $\mathsf{sig}(\mathcal{T}_0)$-inseparable; it follows by join-modularity that they are $\mathsf{sig}(\mathcal{T}_0)$-inseparable from $\mathcal{T}_1 \cup \mathcal{T}_2$.

The join-modularity property is closely related to the Robinson consistency property studied in mathematical logic and applied, similarly to the interpolation property, in modular software specification (13). If a logic satisfies certain criteria, Robinson consistency property and Craig interpolation are known to be equivalent. Unfortunately, to the best our knowledge, the criteria considered in the literature do not apply to $\mathcal{EL}$.

Theorem 8. *$(\mathcal{EL}, \mathcal{QL})$ has the join-modularity property, for all $\mathcal{QL}$ among $\mathcal{QL}_{\mathcal{EL}}$, $\mathcal{QL}^u_{\mathcal{EL}}$, $\mathcal{QL}^i_{\mathcal{EL}}$, and $\mathcal{QL}^q_{\mathcal{EL}}$.*

By Theorem 3, it is sufficient to prove this result for $(\mathcal{EL}, \mathcal{QL}_{\mathcal{EL}})$ and $(\mathcal{EL}, \mathcal{QL}^u_{\mathcal{EL}})$. We provide the proof in Section B of the appendix.

We close this section with the observation that $(\mathcal{EL}, \mathcal{QL}_{\mathsf{CN}})$ does not have the join-modularity property. The TBoxes $\mathcal{T}_1 = \{A_0 \sqsubseteq \exists r.B\}$ and $\mathcal{T}_2 = \{\exists r.B \sqsubseteq A_1\}$ are Σ-inseparable w.r.t. $\mathcal{QL}_{\mathsf{CN}}$, for $\Sigma = \{A_0, A_1, r, B\}$, but $\mathcal{T}_1 \cup \mathcal{T}_2 \models A_0 \sqsubseteq A_1$ and $\mathcal{T}_i \not\models A_0 \sqsubseteq A_1$, for $i = 1, 2$.

3. Simulations, Canonical Models, and Local Entailment

The purpose of this section is to establish some notions that are crucial to our algorithms for deciding Σ-entailment and their correctness proofs: we recall the tight connection between $\mathcal{EL}$ and simulations on graphs, introduce a certain canonical model construction for $\mathcal{EL}$-concepts and TBoxes, and define a local version of entailment between TBoxes.

Definition 9 (Simulation)**.** Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be interpretations and Σ a signature. A relation $S \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is a Σ-*simulation* from $\mathcal{I}_1$ to $\mathcal{I}_2$ if the following holds:

- for all concept names $A \in \Sigma$ and all $(d_1, d_2) \in S$ with $d_1 \in A^{\mathcal{I}_1}$ we have $d_2 \in A^{\mathcal{I}_2}$;
- for all role names $r \in \Sigma$, all $(d_1, d_2) \in S$, and all $e_1 \in \Delta^{\mathcal{I}_1}$ with $(d_1, e_1) \in r^{\mathcal{I}_1}$, there exists $e_2 \in \Delta^{\mathcal{I}_2}$ such that $(d_2, e_2) \in r^{\mathcal{I}_2}$ and $(e_1, e_2) \in S$.

The Σ-simulation S is called *full* if the domain $\mathsf{dom}(S)$ of S coincides with $\Delta^{\mathcal{I}_1}$. For $d_1 \in \Delta^{\mathcal{I}_1}$, $d_2 \in \Delta^{\mathcal{I}_2}$, we write

- $(\mathcal{I}_1, d_1) \leq_\Sigma (\mathcal{I}_2, d_2)$ if there is a Σ-simulation S with $(d_1, d_2) \in S$ and
- $(\mathcal{I}_1, d_1) \leq^{\mathsf{full}}_\Sigma (\mathcal{I}_2, d_2)$ if there is a full Σ-simulation S with $(d_1, d_2) \in S$.

If $\Sigma = \mathsf{N_C} \cup \mathsf{N_R}$, we simply speak of a simulation and write $\leq$ instead of $\leq_\Sigma$.

Let $\mathcal{I}$ be an interpretation, Σ a signature, and $d \in \Delta^{\mathcal{I}}$. Then we define the abbreviation $d^{\Sigma,\mathcal{I}} := \{C \in \mathcal{EL}_\Sigma \mid d \in C^{\mathcal{I}}\}$ and $d^{\Sigma,\mathcal{I},u} := \{C \in \mathcal{EL}^u_\Sigma \mid d \in C^{\mathcal{I}}\}$. The following theorem establishes a fundamental connection between simulations and $\mathcal{EL}$-concepts. The proof is standard, and therefore omitted, see e.g. (9).

Theorem 10.

(i) *If* $(\mathcal{I}_1, d_1) \leq_\Sigma (\mathcal{I}_2, d_2)$ *, then* $d_1^{\Sigma,\mathcal{I}_1} \subseteq d_2^{\Sigma,\mathcal{I}_2}$*. Conversely, if* $\mathcal{I}_1$ *and* $\mathcal{I}_2$ *are finite and* $d_1^{\Sigma,\mathcal{I}_1} \subseteq d_2^{\Sigma,\mathcal{I}_2}$*, then* $(\mathcal{I}_1, d_1) \leq_\Sigma (\mathcal{I}_2, d_2)$*.*

(ii) *If* $(\mathcal{I}_1, d_1) \leq_\Sigma^{\mathsf{full}} (\mathcal{I}_2, d_2)$ *, then* $d_1^{\Sigma,\mathcal{I}_1,u} \subseteq d_2^{\Sigma,\mathcal{I}_2,u}$*. Conversely, if* $\mathcal{I}_1$ *and* $\mathcal{I}_2$ *are finite and* $d_1^{\Sigma,\mathcal{I}_1,u} \subseteq d_2^{\Sigma,\mathcal{I}_2,u}$*, then* $(\mathcal{I}_1, d_1) \leq_\Sigma^{\mathsf{full}} (\mathcal{I}_2, d_2)$*.*

The following example illustrates the difference between simulations and full simulations. Let $\Sigma = \{A\}$ and assume that $\mathcal{I}_1$ has domain $\Delta^{\mathcal{I}_1} = \{d, d'\}$ and that $A^{\mathcal{I}_1} = \{d'\}$, $r^{\mathcal{I}_1} = \{(d, d')\}$. Further assume that $\mathcal{I}_2$ has domain $\Delta^{\mathcal{I}_2} = \{d\}$ and that $A^{\mathcal{I}_2} = r^{\mathcal{I}_2} = \emptyset$. Then $S = \{(d, d)\}$ is a Σ-simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$, but there does not exist a full Σ-simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$ containing (d, d). This is reflected by the fact that $d \in (\exists u.A)^{\mathcal{I}_1}$ but $d \notin (\exists u.A)^{\mathcal{I}_2}$.

We use $\mathsf{sub}(C)$ to denote the set of subconcepts of a concept C, including C itself. For a TBox $\mathcal{T}$, we use $\mathsf{sub}(\mathcal{T})$ to denote the set of all subconcepts of concepts which occur in $\mathcal{T}$.

Definition 11 (Canonical model)**.** Let C be an $\mathcal{EL}$-concept and $\mathcal{T}$ a TBox. The *canonical model* $\mathcal{I}_{C,\mathcal{T}} = (\Delta^{C,\mathcal{T}}, \cdot^{C,\mathcal{T}})$ of C and $\mathcal{T}$ is defined as follows:

- $\Delta^{C,\mathcal{T}} = \{C\} \cup \{C' \mid \exists r.C' \in \mathsf{sub}(C) \cup \mathsf{sub}(\mathcal{T}), \mathcal{T} \models C \sqsubseteq \exists u.C'\}$;
- $D \in A^{\mathcal{I}_{C,\mathcal{T}}}$ iff $\mathcal{T} \models D \sqsubseteq A$, for all $A \in \mathsf{N_C}$;
- $(D, D') \in r^{\mathcal{I}_{C,\mathcal{T}}}$ iff $\mathcal{T} \models D \sqsubseteq \exists r.D'$ and $\exists r.D' \in \mathsf{sub}(\mathcal{T})$ or $\exists r.D'$ is a conjunct in D, for all $r \in \mathsf{N_R}$.

In the last item, the phrase "$\exists r.D'$ is a conjunct in D" also includes the case that $D = \exists r.D'$. Clearly, the size of $\mathcal{I}_{C,\mathcal{T}}$ is polynomial in the size of C and $\mathcal{T}$. Since subsumption in $\mathcal{EL}$ w.r.t. TBoxes is decidable in polynomial time (7) and the proof is easily extended to $\mathcal{EL}^u$, $\mathcal{I}_{C,\mathcal{T}}$ can also be constructed in time polynomial in the size of C and $\mathcal{T}$. We note that the model $\mathcal{I}_{C,\mathcal{T}}$ as defined here is a refinement of the model defined in (3) to prove correctness of the algorithm in (4). We now establish some basic properties of canonical models. The proof of this and all following results of this section can be found in Appendix A.

Lemma 12. *Let* C *be an* $\mathcal{EL}$*-concept and* $\mathcal{T}$ *a TBox. Then*

(1) for all $E \in \Delta^{\mathcal{I}_{C,\mathcal{T}}}$*, we have* $E \in E^{\mathcal{I}_{C,\mathcal{T}}}$*;*

(2) $\mathcal{I}_{C,\mathcal{T}} \models \mathcal{T}$*.*

(3) $(\mathcal{I}_{C,\mathcal{T}}, D) \leq (\mathcal{I}_{C',\mathcal{T}}, D)$*, for all* $\mathcal{EL}$*-concepts* C' *and all* $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}}} \cap \Delta^{\mathcal{I}_{C',\mathcal{T}}}$*.*

Clearly, Points (1) and (2) of Lemma 12 imply that $\mathcal{I}_{C,\mathcal{T}}$ of $\mathcal{T}$ satisfying the concept C. Point (3) states that the behavior of points in a canonical model $\mathcal{I}_{C,\mathcal{T}}$ depends only on $\mathcal{T}$, but not on C. In the remainder of this paper, we will use Points (1) to (3) of Lemma 12 without explicit reference to this lemma. The next lemma relates canonical models $\mathcal{I}_{C,\mathcal{T}}$ to other models of C and $\mathcal{T}$ (Point (1)), and to subsumption w.r.t. $\mathcal{T}$ (Points (1) and (2)). Similar lemmas for the case of $\mathcal{EL}$ without TBoxes have been established in (5).

Lemma 13. *Let* C *and* D *be* $\mathcal{EL}$*-concepts and* $\mathcal{T}$ *a TBox. Then the following holds:*

(1) For all models $\mathcal{I}$ *of* $\mathcal{T}$ *and all* $d \in \Delta^{\mathcal{I}}$*, the following conditions are equivalent:*

(a) $d \in C^{\mathcal{I}}$*;*

(b) $(\mathcal{I}_{C,\mathcal{T}}, C) \leq (\mathcal{I}, d)$;
(c) $(\mathcal{I}_{C,\mathcal{T}}, C) \leq^{\mathsf{full}} (\mathcal{I}, d)$.

(2) *The following conditions are equivalent:*
(a) $\mathcal{T} \models C \sqsubseteq D$;
(b) $C \in D^{\mathcal{I}_{C,\mathcal{T}}}$;
(c) $(\mathcal{I}_{D,\mathcal{T}}, D) \leq (\mathcal{I}_{C,\mathcal{T}}, C)$.

(3) *The following conditions are equivalent:*
(a) $\mathcal{T} \models C \sqsubseteq \exists u.D$;
(b) $C \in (\exists u.D)^{\mathcal{I}_{C,\mathcal{T}}}$.

We now provide a local version of entailment between TBoxes. More precisely, we consider pairs $(\mathcal{T}, C)$ of a TBox and a concept, and are interested in the consequences that C has in models of $\mathcal{T}$. The term "local" refers to the intuition that concepts are interpreted locally in an interpretation, whereas TBoxes are interpreted globally.

Definition 14 (Local Entailment)**.** Let C_1 and C_2 be $\mathcal{EL}$-concepts, $\mathcal{T}_1$ and $\mathcal{T}_2$ TBoxes, and Σ a signature. Then

- $(\mathcal{T}_1, C_1)$ locally Σ-entails $(\mathcal{T}_2, C_2)$, w.r.t. $\mathcal{EL}$, in symbols $(\mathcal{T}_1, C_1) \sqsubseteq_\Sigma (\mathcal{T}_2, C_2)$, iff for all $\mathcal{EL}_\Sigma$-concepts E,
$$\mathcal{T}_2 \models C_2 \sqsubseteq E \Rightarrow \mathcal{T}_1 \models C_1 \sqsubseteq E.$$
- $(\mathcal{T}_1, C_1)$ locally Σ-entails $(\mathcal{T}_2, C_2)$ w.r.t. $\mathcal{EL}^u$, in symbols $(\mathcal{T}_1, C_1) \sqsubseteq_\Sigma^u (\mathcal{T}_2, C_2)$, iff for all $\mathcal{EL}^u_\Sigma$-concepts E,
$$\mathcal{T}_2 \models C_2 \sqsubseteq E \Rightarrow \mathcal{T}_1 \models C_1 \sqsubseteq E.$$

The following lemma characterizes local Σ-entailment in terms of simulations. Since the largest Σ-simulation between two finite graphs can be computed in polynomial time (9), the lemma implies that local Σ-entailment w.r.t. $\mathcal{EL}$ and $\mathcal{EL}^u$ can be decided in polynomial time.

Lemma 15. *Let $\mathcal{T}_1, \mathcal{T}_2$ be TBoxes and C_1, C_2 $\mathcal{EL}$-concepts and Σ a signature. Then*

- $(\mathcal{T}_1, C_1) \sqsubseteq_\Sigma (\mathcal{T}_2, C_2)$ *iff* $(\mathcal{I}_{C_2,\mathcal{T}_2}, C_2) \leq_\Sigma (\mathcal{I}_{C_1,\mathcal{T}_1}, C_1)$;
- $(\mathcal{T}_1, C_1) \sqsubseteq_\Sigma^u (\mathcal{T}_2, C_2)$ *iff* $(\mathcal{I}_{C_2,\mathcal{T}_2}, C_2) \leq_\Sigma^{\mathsf{full}} (\mathcal{I}_{C_1,\mathcal{T}_1}, C_1)$.

Our algorithm deciding whether $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$ will systematically search for witnesses $C \sqsubseteq D$ for $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$ (and similarly for $\mathcal{T}_1 \sqsubseteq_\Sigma^u \mathcal{T}_2$). Clearly, $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$ iff there exists an $\mathcal{EL}_\Sigma$-concept C such that $(\mathcal{T}_1, C) \not\sqsubseteq_\Sigma (\mathcal{T}_2, C)$. Since Lemma 15 implies that local Σ-entailment can be decided in polynomial time, it thus provides some first evidence that, when searching for witnesses $C \sqsubseteq D$ for $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$, the difficult part is to identify a suitable concept C.

4. Deciding Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$

An initial observation about deciding Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ is that minimal witness sentences for non-Σ-entailment may be quite large. Let $\mathcal{T}_1$ be the empty TBox and $\Sigma = \{A, B, r, s\}$. For each $n \geq 0$, we define a TBox $\mathcal{T}_n'$. It has additional concept names

$X_0, \ldots, X_{n-1}$ and $\overline{X}_0, \ldots, \overline{X}_{n-1}$ that are used to represent a binary counter X: if X_i is true, then the i-th bit is positive and if $\overline{X}_i$ is true, then it is negative. Define $\mathcal{T}'_n$ as

$$A \sqsubseteq \overline{X}_0 \sqcap \cdots \sqcap \overline{X}_{n-1} \quad (1)$$

$$\textstyle\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X}_i \sqcap X_0 \sqcap \cdots \sqcap X_{i-1}) \sqsubseteq X_i \quad \text{for all } i < n \quad (2)$$

$$\textstyle\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap X_0 \sqcap \cdots \sqcap X_{i-1}) \sqsubseteq \overline{X}_i \quad \text{for all } i < n \quad (3)$$

$$\textstyle\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap \overline{X}_j) \sqsubseteq \overline{X_i} \quad \text{for all } j < i < n \quad (4)$$

$$\textstyle\bigsqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap \overline{X}_j) \sqsubseteq X_i \quad \text{for all } j < i < n \quad (5)$$

$$X_0 \sqcap \cdots \sqcap X_{n-1} \sqsubseteq B \quad (6)$$

Observe that Lines 2-5 implement incrementation of the counter X. Then the smallest consequence of $\mathcal{T}_1 \cup \mathcal{T}'_n$ in the signature Σ which is not a consequence of $\mathcal{T}_1$ is $C_{2^n} \sqsubseteq B$, where:

$$C_0 = A$$

$$C_i = \exists r.C_{i-1} \sqcap \exists s.C_{i-1}$$

Clearly, C_{2^n} is doubly exponentially large in the size of $\mathcal{T}_1$ and $\mathcal{T}'_n$. If we use structure sharing (i.e., define the size of C_{2^n} as the number of its distinct subconcepts), it is still exponentially large.

We now design a ExpTime algorithm deciding Σ-entailment. At the end of Section 3, we have seen that when searching for a witness for $\mathcal{T}_1 \not\models_\Sigma \mathcal{T}_2$ is sufficient to search for a C such that $\mathcal{I}_{C,\mathcal{T}_2} \not\leq_\Sigma \mathcal{I}_{C,\mathcal{T}_1}$. Using Lemma 15, we now derive a characterization of non-Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ which can be implemented almost directly. We start with a technical lemma.

Lemma 16. *Suppose $\mathcal{T} \models C \sqsubseteq \exists r.D$, where C, D are $\mathcal{EL}$-concepts. Then one of the following holds:*
- *there is a conjunct $\exists r.C'$ of C such that $\mathcal{T} \models C' \sqsubseteq D$;*
- *there is a $\exists r.C' \in \mathsf{sub}(\mathcal{T})$ such that $\mathcal{T} \models C \sqsubseteq \exists r.C'$ and $\mathcal{T} \models C' \sqsubseteq D$.*

Proof. Let $\mathcal{T} \models C \sqsubseteq \exists r.D$. By Point 2 of Lemma 13, $C \in (\exists r.D)^{\mathcal{I}_{C,\mathcal{T}}}$. Thus, there is a $C' \in D^{\mathcal{I}_{C,\mathcal{T}}}$ such that $(C, C') \in r^{\mathcal{I}_{C,\mathcal{T}}}$. By definition of $\mathcal{I}_{C,\mathcal{T}}$, (i) $\exists r.C'$ is a conjunct of C or (ii) $\exists r.C' \in \mathsf{sub}(\mathcal{T})$ and $\mathcal{T} \models C \sqsubseteq \exists r.C'$. In both cases it follows from Point 2 of Lemma 13 that $\mathcal{T} \models C' \sqsubseteq D$. ☐

The *outdegree* of a concept C is the maximum cardinality of any set P of pairs of the form (r, C'), with r a role name and C' a concept, such that $\bigsqcap_{(r,C') \in P} \exists r.C' \in \mathsf{sub}(C)$. We use $|C|$ and $|\mathcal{T}|$ to denote the *length* of a C and a TBox $\mathcal{T}$, i.e., the number of occurrences of symbols needed to write it.

Proposition 17. *Assume $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$. $\mathcal{T}_1$ does not Σ-entail $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ iff there exists an $\mathcal{EL}_\Sigma$-concept C and a concept $D \in \mathsf{sub}(\mathcal{T}_2)$ such that*
(a) $\mathcal{T}_2 \models C \sqsubseteq D$;
(b) $(\mathcal{T}_1, C) \not\leq_\Sigma (\mathcal{T}_2, D)$;
(c) the outdegree of C is bounded by $|\mathcal{T}_2|$.

Proof. We first show that if there exist an $\mathcal{EL}_\Sigma$-concept C and $D \in \mathsf{sub}(\mathcal{T}_2)$ with (a) and (b), then $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$. Assume that (a) and (b) are satisfied for C and D. By (b), there is an $\mathcal{EL}_\Sigma$-concept E with $\mathcal{T}_2 \models D \sqsubseteq E$ and $\mathcal{T}_1 \not\models C \sqsubseteq E$. From the former and (a), we get $\mathcal{T}_2 \models C \sqsubseteq E$, which implies that $\mathcal{T}_1$ does not Σ-entail $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$.

Now we show that from $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$ follows the existence of C and D satisfying (a) and (b). If there exists $C \sqsubseteq D$ which follows from $\mathcal{T}_2$ but not from $\mathcal{T}_1$ with $\mathsf{sig}(C) \subseteq \Sigma$ and D a Σ-concept in $\mathsf{sub}(\mathcal{T}_2)$, then we are done: we have $\mathcal{T}_2 \models D \sqsubseteq D$ and $\mathcal{T}_1 \not\models C \sqsubseteq D$, therefore $(\mathcal{T}_1, C) \not\sqsubseteq_\Sigma (\mathcal{T}_2, D)$. Assume that no such inclusion separating the two TBoxes exists.

Let $C \sqsubseteq D$ be a witness for $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$ such that no witness $C' \sqsubseteq D'$ with D' shorter than D exists. Then D is of the form $\exists r.D'$:

- If $D = \top$, then $\mathcal{T}_1 \models C \sqsubseteq D$, contradicting the fact that $C \sqsubseteq D$ separates the two TBoxes.
- If D is an atomic concept, then $D \in \mathsf{sub}(\mathcal{T}_2)$, which we have assumed not to be the case.
- If D is a conjunction $D_1 \sqcap D_2$, then $\mathcal{T}_2 \models C \sqsubseteq D_i$ for all $i \in \{1, 2\}$ and $\mathcal{T}_1 \not\models C \sqsubseteq D_i$ for some $i \in \{1, 2\}$. Thus, one of $C \sqsubseteq D_1$ and $C \sqsubseteq D_2$ separates the two TBoxes, contradicting the minimality of D.

By Lemma 16, $\mathcal{T}_2 \models C \sqsubseteq \exists r.D'$ implies that one of the following holds:

(1) there exists a conjunct $\exists r.C'$ of C such that $\mathcal{T}_2 \models C' \sqsubseteq D'$;

(2) there exists $\exists r.C' \in \mathsf{sub}(\mathcal{T}_2)$ such that $\mathcal{T}_2 \models C \sqsubseteq \exists r.C'$ and $\mathcal{T}_2 \models C' \sqsubseteq D'$.

We first show that (1) cannot be true. Assume it is. Then we have $\mathcal{T}_1 \models C' \sqsubseteq D'$ because otherwise $C' \sqsubseteq D'$ is a witness, contradicting the minimality of D. It follows that $\mathcal{T}_1 \models \exists r.C' \sqsubseteq \exists r.D'$. Since $\exists r.C'$ is a conjunct of C, $\mathcal{T}_1 \models C \sqsubseteq \exists r.C'$. Together with $\mathcal{T}_1 \models \exists r.C' \sqsubseteq \exists r.D'$, we obtain $\mathcal{T}_1 \models C \sqsubseteq \exists r.D' = D$. It follows that $\mathcal{T}_1 \models C \sqsubseteq D$, contradicting the fact that $C \sqsubseteq D$ is a witness.

Thus, (2) applies. We show that the concepts C and $\exists r.C'$ (substituted for D) satisfy Conditions (a) and (b). First, $\mathcal{T}_2 \models C \sqsubseteq \exists r.C'$ establishes Condition (a). For Condition (b), observe that $\mathcal{T}_1 \not\models C \sqsubseteq \exists r.D'$ and $\mathcal{T}_2 \models \exists r.C' \sqsubseteq \exists r.D'$. This means $(\mathcal{T}_1, C) \not\sqsubseteq_\Sigma (\mathcal{T}_2, \exists r.C')$.

We have shown that $\mathcal{T}_1$ does not Σ-entail $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ iff there exist C and D such that (a) and (b) hold. It thus remains to show that one can find such C and D satisfying constraint (c) as well, whenever $\mathcal{T}_1$ does not Σ-entail $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$. This is done in Section C of the appendix. □

The main benefit of this characterization is that when searching for a subsumption $\mathcal{T}_2 \models C \sqsubseteq D$ with $\mathsf{sig}(C \sqsubseteq D) \subseteq \Sigma$ which does not follow from $\mathcal{T}_1$, it allows us to concentrate on concepts D of a very simple form, namely subconcepts of $\mathcal{T}_2$. This is achieved by considering $\mathsf{sig}(\mathcal{T}_2)$-concepts instead of $\mathcal{EL}_\Sigma$-concepts as in the definition of Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$.

We now devise an algorithm for deciding whether $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$. To check whether $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$, the algorithm searches for an $\mathcal{EL}_\Sigma$-concept C such that for some $D \in \mathsf{sub}(\mathcal{T}_2)$, Points (a)–(c) of Proposition 17 are satisfied. Intuitively, it proceeds in rounds. In the first round, the algorithm considers the case where C is a conjunction of concept names in Σ. For every such C and all $D \in \mathsf{sub}(\mathcal{T}_2)$, it checks whether Points (a) and (b) are satisfied. By Lemma 15, this can be done in polynomial time. If all tests fail, the second round is started in which the algorithm considers concepts C of the form $F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, where F_0 is a conjunction of concept names and P is a set of pairs (r, E) with r a role name and E a candidate for C from the first round (i.e., E is also a conjunction of concept names). Because of Point (c), it will be sufficient to consider sets P of cardinality bounded by $|\mathcal{T}_2|$. To check if such a concept C satisfies Points (a) and (b), we exploit the information that we have gained about the concepts E in the previous round. If again no suitable C is found, then in the third round we use the Cs from the second round as the Es in $F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, and so on.

For the algorithm to terminate and run in exponential time, we have to introduce a condition that indicates when enough candidates C have been inspected in order to know that there is no witness $C \sqsubseteq D$. To obtain such a termination condition and to avoid having to deal with double exponentially large concepts, our algorithm will not construct the candidate concepts C directly, but rather use a certain data structure to represent relevant information about C. The relevant information about C is suggested by Proposition 17: for each C, we take the quadruple

$$C^\sharp = (F, K_{\mathcal{T}_1}(C), K_{\mathcal{T}_2}(C), K_{\mathcal{T}_1,\mathcal{T}_2}(C)),$$

where F is the conjunction of all concept names occurring in the top-level conjunction of C (if there are none, then $F = \top$) and

- $K_{\mathcal{T}}(C) = \{D \in \mathsf{sub}(\mathcal{T}) \mid \mathcal{T} \models C \sqsubseteq D\}$;
- $K_{\mathcal{T}_1,\mathcal{T}_2}(C) = \{D \in \mathsf{sub}(\mathcal{T}_2) \mid (\mathcal{T}_1, C) \sqsubseteq_\Sigma (\mathcal{T}_2, D)\}$.

We call this the *quadruple determined by* C. By Proposition 17, the quadruple $C^\sharp$ determined by a concept C gives us enough information to decide whether C is the left hand side of a witness. In addition, it contains enough information to enable the recursive search described above. In what follows F, F_0, etc. range over conjunctions of concept names and the concept $\top$, and when writing $C = F \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$ we assume that P is a finite set of pairs (r, E) in which r is a role and E an $\mathcal{EL}$-concept.

Now the following lemma (proved in the appendix) states how $K_{\mathcal{T}}(C)$ is computed recursively during the search described above.

Lemma 18. *Let $\mathcal{T}$ be a TBox and $C = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$. Then*

$$K_{\mathcal{T}}(C) = K_{\mathcal{T}}(F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.(\bigsqcap_{D\in K_{\mathcal{T}}(E)} D)).$$

The algorithm deciding non-Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ is shown in Figure 1. Observe that the Condition $\mathcal{Q}_2 \setminus \mathcal{Q}_3 \neq \emptyset$ corresponds to satisfaction of Points (a) and (b) in Proposition 17. Also observe that, in Point (b) of the definition of $\mathcal{F}_3$, we refer to the canonical model $\mathcal{I}_{D,\mathcal{T}_i}$ for the relevant concepts D. These models are constructed in polynomial time when needed. To show that this algorithm really implements the initial description given at the beginning of this section, we make explicit the concepts that we describe by means of the quadruples constructed in Step 3 of Figure 1. This is done by the following lemma, which will also be a central ingredient to our correctness proof.

Lemma 19. *Let $(F_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)$ be the quadruple obtained from F_0 and Q in Step (3) of Figure 1. Let, for each $(r, q) \in Q$, $C_{r,q}$ be some concept such that $C^\sharp_{r,q} = q$. Then $C^\sharp = (F_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)$, where C is defined as $C = F_0 \sqcap \bigsqcap_{(r,q)\in Q} \exists r.C_{r,q}$.*

Input: TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and signature $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$.

(1) Compute the set $\mathcal{N}_0$ of quadruples determined by conjunctions of concept names from Σ.

(2) if $\mathcal{N}_0$ contains a quadruple $(F, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3)$ such that $\mathcal{Q}_2 \setminus \mathcal{Q}_3 \neq \emptyset$, then output "$\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$".

(3) Generate the sequence $\mathcal{N}_1, \mathcal{N}_2, \ldots$ of sets of quadruples with $\mathcal{N}_{i+1} = \mathcal{N}_i \cup \mathcal{N}_i'$, where $\mathcal{N}_i'$ is the set of quadruples $(F_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)$ which can be obtained from a conjunction F_0 of concept names from Σ and a set $Q \subseteq (\mathsf{N_R} \cap \Sigma) \times \mathcal{N}_i$ of cardinality not exceeding $|\mathcal{T}_2|$ in the following way:

- $\mathcal{F}_1 = K_{\mathcal{T}_1}(F_0 \sqcap \bigsqcap_{(r,(F,\mathcal{Q}_1,\mathcal{Q}_2,\mathcal{Q}_3)) \in Q} \exists r.(\bigsqcap_{D \in \mathcal{Q}_1} D))$;
- $\mathcal{F}_2 = K_{\mathcal{T}_2}(F_0 \sqcap \bigsqcap_{(r,(F,\mathcal{Q}_1,\mathcal{Q}_2,\mathcal{Q}_3)) \in Q} \exists r.(\bigsqcap_{D \in \mathcal{Q}_2} D))$;
- $\mathcal{F}_3 = \{D \mid D \in \mathsf{sub}(\mathcal{T}_2)$ and
  - (a) for all $A \in \Sigma$, $A \in K_{\mathcal{T}_2}(D)$ implies $A \in \mathcal{F}_1$;
  - (b) if $(D, D') \in r^{\mathcal{I}_{D,\mathcal{T}_2}}$ with $r \in \Sigma$, then
    - (i) there is a tuple $(r, (F, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3)) \in Q$ such that $D' \in \mathcal{Q}_3$
    - or (ii) there is $\exists r.C' \in \mathcal{F}_1$ with $(\mathcal{T}_1, C') \sqsubseteq_\Sigma (\mathcal{T}_2, D')$

  $\}$

This is done until $\mathcal{N}_i$ contains a quadruple $(F, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3)$ such that $\mathcal{Q}_2 \setminus \mathcal{Q}_3 \neq \emptyset$, or $\mathcal{N}_{i+1} = \mathcal{N}_i$. Output "$\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$" if the first condition applies. Otherwise, output "$\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$".

Fig. 1. Algorithm deciding non-Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$

Proof. Let $(F_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)$ and C be as in the lemma. It is trivial that F_0 is as required. By Lemma 18, $\mathcal{F}_1$ and $\mathcal{F}_2$ are as required. It remains to consider $\mathcal{F}_3$. Fix $D \in \mathsf{sub}(\mathcal{T}_2)$. By Lemma 15, $(\mathcal{T}_1, C) \sqsubseteq_\Sigma (\mathcal{T}_2, D)$ iff $(\mathcal{I}_{D,\mathcal{T}_2}, D) \leq_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, C)$. By definition of simulations, we therefore have $D \in K_{\mathcal{T}_1,\mathcal{T}_2}(C)$ iff the following holds:

(1) for all concept names $A \in \Sigma$, $A \in K_{\mathcal{T}_2}(D)$ implies $A \in K_{\mathcal{T}_1}(C)$;

(2) for all $r \in \Sigma$ and D' with $(D, D') \in r^{\mathcal{I}_{D,\mathcal{T}_2}}$ there exists C' with $(C, C') \in r^{\mathcal{I}_{C,\mathcal{T}_1}}$ and $(\mathcal{I}_{D,\mathcal{T}_2}, D') \leq_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, C')$.

Point 1 is checked under Point (a) in the definition of $\mathcal{F}_3$ of the algorithm in Figure 1 since, as we have seen already, $K_{\mathcal{T}_1}(C) = \mathcal{F}_1$. For Point 2, $(C, C') \in r^{\mathcal{I}_{C,\mathcal{T}_1}}$ is equivalent to (i) $\exists r.C'$ is a conjunct of C or (ii) $\exists r.C' \in K_{\mathcal{T}_1}(C)$. In Case (i), $C' = C_{r,q}$ for some $(r, q) \in Q$ and $(\mathcal{T}_1, C') \sqsubseteq_\Sigma (\mathcal{T}_2, D')$ iff D' is an element of the fourth component of q. By Lemma 15, this is what is checked in (b.i) in the definition of $\mathcal{F}_3$ of the algorithm. In Case (ii), $\exists r.C' \in \mathcal{F}_1$ and, by Lemma 15, $(\mathcal{I}_{D,\mathcal{T}_2}, D') \leq_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, C')$ iff $(\mathcal{T}_1, C') \sqsubseteq_\Sigma (\mathcal{T}_2, D')$. This condition is exactly what is checked in (b.ii) in the definition of $\mathcal{F}_3$ of the algorithm. $\square$

Proposition 20. *The algorithm in Figure 1 is sound, complete, and runs in exponential time.*

Proof. Soundness follows from Proposition 17 and Lemma 19. For completeness, assume that $\mathcal{T}_1$ does not Σ-entail $\mathcal{T}_2$ w.r.t. $\mathcal{QL}_{\mathcal{EL}}$. By Proposition 17, there exists $\mathcal{EL}_\Sigma$-concept C of outdegree not exceeding $|\mathcal{T}_2|$ and $D \in \mathsf{sub}(\mathcal{T}_2)$ such that $\mathcal{T}_2 \models C \sqsubseteq D$ and $(\mathcal{T}_1, C) \not\sqsubseteq_\Sigma (\mathcal{T}_2, D)$. If C is a conjunction of concept names, then the algorithm outputs "$\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$" in Step 2. Now suppose C has quantifier depth $n \geq 1$. Using Lemma 19, one can easily show by induction on i that for all $i \geq 0$, the set $\mathcal{N}_i$ contains all quadruples determined by subconcepts C' of C of quantifier depth smaller than i. Hence, the algorithm outputs "$\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$" after computing some $\mathcal{N}_i$ with $i \leq n$.

For termination and complexity, observe that, by Lemma 15, the quadruple determined by a conjunction of concept names from Σ can be computed in polynomial time. Hence Steps 1 and 2 run in exponential time. For Step 3 observe that the number of tuples $(F, \mathcal{Q}_1, \mathcal{Q}_2, \mathcal{Q}_3)$ with F a conjunction of concept names from Σ and $\mathcal{Q}_i \subseteq \mathsf{sub}(\mathcal{T}_1 \cup \mathcal{T}_2)$ is bounded by $2^{4|\mathcal{T}_1 \cup \mathcal{T}_2|}$. It follows that $\mathcal{N}_i = \mathcal{N}_{i+1}$ for some $i \leq 2^{4|\mathcal{T}_1 \cup \mathcal{T}_2|}$. Hence, the algorithm terminates and to show that it runs in exponential time it remains to check that $\mathcal{N}_{i+1}$ can be computed in exponential time from $\mathcal{N}_i$. This follows from the following: first, the number of pairs (F_0, Q), with F_0 a conjunction of concept names from Σ and $Q \subseteq (\mathsf{N_R} \cap \Sigma) \times \mathcal{N}_i$ of cardinality not exceeding $|\mathcal{T}_2|$, is still only exponential in $|\mathcal{T}_1 \cup \mathcal{T}_2|$; and second, the computation of $(F_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3)$ from F_0 and Q in Figure 1 can be done in time polynomial in $|\mathcal{T}_1 \cup \mathcal{T}_2|$. ☐

In Figure 1, we assume that $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$. But, as observed above already, $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$ iff $\mathcal{T}_1 \sqsubseteq_{\Sigma \cap \mathsf{sig}(\mathcal{T}_2)} \mathcal{T}_2$ because of robustness under vocabulary extensions. Thus, by applying the algorithm to $\Sigma \cap \mathsf{sig}(\mathcal{T}_2)$, we obtain a general decision procedure for Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ and have proved the following result.

Theorem 21. *Σ-entailment of $\mathcal{EL}$-TBoxes w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ is in* ExpTime.

5. Σ-entailment w.r.t. Other Query Languages

In this section, we first prove the equivalences stated in Theorem 3 and then provide an extension of the decision procedure for Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ to Σ-entailment w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$.

5.1. Equivalence of Σ-entailment w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$ and $\mathcal{QL}_{\mathcal{EL}}$

To prove that Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ implies Σ-entailment w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$, we first show that answering an instance query $(\mathcal{T}, \mathcal{A}) \models D(a)$ can be decomposed into two parts that separate reasoning with the TBox $\mathcal{T}$ from reasoning with the ABox $\mathcal{A}$.

Lemma 22. *Let $\mathcal{T}$ be a TBox, Σ a signature, and $\mathcal{A}$ an $\mathcal{EL}_\Sigma$-ABox.*

(1) For every $\mathcal{EL}_\Sigma$-concept D and $a \in \mathsf{N_I}$, $(\mathcal{T}, \mathcal{A}) \models D(a)$ iff there exists an $\mathcal{EL}_\Sigma$-concept C such that $\mathcal{T} \models C \sqsubseteq D$ and $\mathcal{A} \models C(a)$.

(2) For every $\mathcal{EL}^u_\Sigma$-concept D and $a \in \mathsf{N_I}$, $(\mathcal{T}, \mathcal{A}) \models D(a)$ iff there exists an $\mathcal{EL}^u_\Sigma$-concept C such that $\mathcal{T} \models C \sqsubseteq D$ and $\mathcal{A} \models C(a)$.

The first equivalence of Theorem 3 is now an easy consequence of Lemma 22.

Proposition 23. *For all TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and any signature Σ: $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$ iff $\mathcal{T}_1 \sqsubseteq^i_\Sigma \mathcal{T}_2$.*

Proof. Suppose $\mathcal{T}_1 \not\sqsupseteq_\Sigma \mathcal{T}_2$. Take $\mathcal{EL}_\Sigma$-concepts C and D such that $\mathcal{T}_2 \models C \sqsubseteq D$, but $\mathcal{T}_1 \not\models C \sqsubseteq D$. Let $\mathcal{A} = \{C(a)\}$. Then $(\mathcal{T}_2, \mathcal{A}) \models D(a)$ but $(\mathcal{T}_1, \mathcal{A}) \not\models D(a)$. Hence $\mathcal{T}_1 \not\sqsupseteq^i_\Sigma \mathcal{T}_2$. Conversely, assume $\mathcal{T}_1 \not\sqsupseteq^i_\Sigma \mathcal{T}_2$. Take a Σ-ABox $\mathcal{A}$, an $\mathcal{EL}_\Sigma$-concept D and $a \in \mathsf{N_I}$ such that $(\mathcal{T}_2, \mathcal{A}) \models D(a)$ but $(\mathcal{T}_1, \mathcal{A}) \not\models D(a)$. Then, by Lemma 22, Point 1, there exists an $\mathcal{EL}_\Sigma$-concept C such that $\mathcal{T}_2 \models C \sqsubseteq D$ and $\mathcal{A} \models C(a)$. Again by Lemma 22, Point 1, $\mathcal{T}_1 \not\models C \sqsubseteq D$. Hence $\mathcal{T}_1 \not\sqsupseteq_\Sigma \mathcal{T}_2$. ☐

Proposition 23 and Theorem 21 yield the following result.

Theorem 24. *Σ-entailment of $\mathcal{EL}$-TBoxes w.r.t. $\mathcal{QL}^i_{\mathcal{EL}}$ is in* ExpTime.

5.2. Equivalence of Σ-entailment w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$ and $\mathcal{QL}^u_{\mathcal{EL}}$

We first provide a notion of Σ-entailment between knowledge bases. Instead of inclusions between concepts we now consider answers to conjunctive queries two KBs give.

Definition 25. Let $(\mathcal{T}_1, \mathcal{A}_1)$ and $(\mathcal{T}_2, \mathcal{A}_2)$ be KBs and Σ a signature. $(\mathcal{T}_1, \mathcal{A}_1)$ *Σ-query entails* $(\mathcal{T}_2, \mathcal{A}_2)$, in symbols $(\mathcal{T}_1, \mathcal{A}_1) \sqsupseteq^q_\Sigma (\mathcal{T}_2, \mathcal{A}_2)$, if for all conjunctive Σ-queries q with k free variables and k-tuples $\mathbf{a}$ of individual names in $\mathsf{N_I}$:

$$(\mathcal{T}_2, \mathcal{A}_2) \models q(\mathbf{a}) \Rightarrow (\mathcal{T}_1, \mathcal{A}_1) \models q(\mathbf{a}).$$

The difference to Definition 2 is that here we do not define entailment between TBoxes by considering answers to queries over arbitrary ABoxes, but we fix two KBs each consisting of a TBox and an ABox and then consider the answers to queries these KBs give. It turns out that this entailment relation is much easier to characterize semantically than the former. We now give such a semantic characterization of Σ-query entailment between KBs (Lemma 29) and then use this characterization to prove that Σ-entailment between TBoxes w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$ is equivalent to Σ-entailment between TBoxes w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$ (Proposition 30).

To start with, we extend the notion of canonical models discussed above to canonical models for KBs $(\mathcal{T}, \mathcal{A})$. Denote by $\mathsf{obj}(\mathcal{A})$ the set of individual names occurring in an ABox $\mathcal{A}$. For any TBox $\mathcal{T}$, ABox $\mathcal{A}$ and finite set Ob of individual names with $\mathsf{obj}(\mathcal{A}) \subseteq \mathsf{Ob}$, the canonical model $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}$ is defined as follows: fix some $b_{\mathsf{aux}} \notin \mathsf{Ob}$ and set

- $\Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}} = \mathsf{Ob} \cup \{b_{\mathsf{aux}}\} \cup \{C \mid \exists r.C \in \mathsf{sub}(\mathcal{T} \cup \mathcal{A}), (\mathcal{T}, \mathcal{A}) \models \exists u.C(a)\}$;
- $a^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}} = a$, for all $a \in \mathsf{Ob}$;
- $a^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}} = b_{\mathsf{aux}}$, for all $a \in \mathsf{N_I} \setminus \mathsf{Ob}$.
- $d \in A^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}}$ iff $d = a^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}}$ for some $a \in \mathsf{N_I}$ and $(\mathcal{T}, \mathcal{A}) \models A(a)$ or $d = C \in \mathsf{N_C} \cap \Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}}$ and $\mathcal{T} \models C \sqsubseteq A$, for all $A \in \mathsf{N_C}$;
- $(d_1, d_2) \in r^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}}$ iff one of the following three conditions holds:
  - · $d_1, d_2 \in \mathsf{N_I}$ and $r(d_1, d_2) \in \mathcal{A}$ or
  - · $d_1 = a \in \mathsf{N_I}$ and $d_2 = C \in \mathsf{N_C}$ and $(\mathcal{T}, \mathcal{A}) \models \exists r.C(a)$ or
  - · $d_1 = C_1 \in \mathsf{N_C}$ and $d_2 = C_2 \in \mathsf{N_C}$ and $\mathcal{T} \models C_1 \sqsubseteq \exists r.C_2$, for all $r \in \mathsf{N_R}$.

We set $\mathcal{I}_{\mathcal{A},\mathcal{T}} = \mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{obj}(\mathcal{A})}$. To describe basic properties of canonical model for KBs, we extend the notion of simulations to simulations preserving individuals.

Definition 26. Let $\mathsf{Ob} \subseteq \mathsf{N_I}$. A Σ-simulation S between two models $\mathcal{I}_1$ and $\mathcal{I}_2$ *preserves* Ob if $(a^{\mathcal{I}_1}, a^{\mathcal{I}_2}) \in S$, for all $a \in \mathsf{Ob}$. We write $\mathcal{I}_1 \leq^{\mathsf{Ob}}_\Sigma \mathcal{I}_2$ if there exists a Σ-simulation between $\mathcal{I}_1$ and $\mathcal{I}_2$ preserving Ob and we write $\mathcal{I}_1 \leq^{\mathsf{Ob},\mathsf{full}}_\Sigma \mathcal{I}_2$ if there exists a full such Σ-simulation. A *Σ-homomorphism preserving* Ob is a full Σ-simulation which is a function.

The following lemma establishes the main properties of canonical models for KBs.

Lemma 27. *Let $\mathcal{T}$ be a TBox, $\mathcal{A}$ an ABox, and $\mathsf{Ob} \supseteq \mathsf{obj}(\mathcal{A})$ a finite set of individual names. Then $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}$ is a model of $(\mathcal{T},\mathcal{A})$ and the following holds:*

(1) For all finite sets $\mathsf{Ob}', \mathsf{Ob}'' \supseteq \mathsf{obj}(\mathcal{A})$: $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}'} \leq^{\mathsf{Ob},\mathsf{full}} \mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}''}$.

(2) For all models $\mathcal{I}$ of $\mathcal{T}$ the following are equivalent:

- *(a) $\mathcal{I} \models \mathcal{A}$;*
- *(b) $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}} \leq^{\mathsf{Ob},\mathsf{full}} \mathcal{I}$.*

(3) For all assertions α of the form $C(a)$ and $r(a,b)$, where C is a $\mathcal{EL}^u$-concept, $r \in \mathsf{N_R}$, and $a,b \in \mathsf{N_I}$, the following conditions are equivalent:

- *(a) $(\mathcal{T},\mathcal{A}) \models \alpha$;*
- *(b) $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}} \models \alpha$.*

Proof. With the exception of Point 1, the proof is similar to the proof of Lemma 13 and left to the reader. For Point 1, observe that $\mathcal{C} = \mathsf{N_C} \cap \Delta^{\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}}}$ does not depend on Ob and that

$$S = \{(D,D) \mid D \in \mathcal{C}\} \cup \{(b^{\mathcal{I},\mathcal{A},\mathsf{Ob}'}, c^{\mathcal{I},\mathcal{A},\mathsf{Ob}''}) \mid b,c \in \mathsf{N_I} \setminus \mathsf{obj}(\mathcal{A})\} \cup \{(a,a) \mid a \in \mathsf{obj}(\mathcal{A})\}$$

is a full Σ-simulation preserving $\mathsf{N_I}$ between $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}'}$ and $\mathcal{I}_{\mathcal{A},\mathcal{T},\mathsf{Ob}''}$. □

Let $\mathcal{I}$ be a model and Ob a non-empty set of individual names. Any model $\mathcal{I}^{\mathsf{Ob},*}$ with the following properties is called an *unravelling* of $\mathcal{I}$ w.r.t. Ob (where $W = \{a^{\mathcal{I}} \mid a \in \mathsf{Ob}\}$):

- $\Delta^{\mathcal{I}^{\mathsf{Ob},*}}$ is the set of all words $d_0 r_1 d_1 r_2 \cdots r_n d_n$, $n \geq 0$, such that $d_0 \in W$ and $(d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}}$ for $i < n$;
- $r^{\mathcal{I}^{\mathsf{Ob},*}} = \{(wd, wdrd') \mid wd, wdrd' \in \Delta^{\mathcal{I}^{\mathsf{Ob},*}}\} \cup \{(d,d') \in W^2 \mid (d,d') \in r^{\mathcal{I}}\}$, for $r \in \mathsf{N_R}$;
- $A^{\mathcal{I}^{\mathsf{ob},*}} = \{wd \mid d \in A^{\mathcal{I}}\}$, for $A \in \mathsf{N_C}$;
- $a^{\mathcal{I}^{\mathsf{ob},*}} = a^{\mathcal{I}}$, for $a \in \mathsf{Ob}$.

Observe that the relation $S \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}^{\mathsf{Ob},*}}$ consisting of all pairs (d, wd) with $d \in \Delta^{\mathcal{I}}$ and $wd \in \Delta^{\mathcal{I}^{\mathsf{Ob},*}}$ is a bisimulation (i.e., a simulation in both directions) between $\mathcal{I}$ and $\mathcal{I}^{\mathsf{Ob},*}$. It follows that if $\mathcal{I} \models (\mathcal{T},\mathcal{A})$ and $\mathsf{Ob} \supseteq \mathsf{obj}(\mathcal{A})$, then $\mathcal{I}^{\mathsf{Ob},*} \models (\mathcal{T},\mathcal{A})$.

Lemma 28. *Let Ob be a non-empty finite set of individual names, Σ a signature, and $\mathcal{I}, \mathcal{J}$ models such that $a^{\mathcal{I}} \neq b^{\mathcal{I}}$ for distinct $a,b \in \mathsf{Ob}$. The following conditions are equivalent:*

- *$\mathcal{I} \leq_{\Sigma}^{\mathsf{Ob},\mathsf{full}} \mathcal{J}$;*
- *There exists a Σ-homomorphism from $\mathcal{I}^{\mathsf{Ob},*}$ to $\mathcal{J}$ preserving Ob.*

Proof. Straightforward and left to the reader. □

We are now in a position to characterize Σ-query entailment between KBs. Observe that it follows from Point 3 of the characterization below that Σ-query entailment between KBs is decidable in polynomial time.

Lemma 29. *Let $(\mathcal{T}_1,\mathcal{A}_1)$ and $(\mathcal{T}_2,\mathcal{A}_2)$ be KBs, Σ a signature, and $b \notin \mathsf{obj}(\mathcal{A}_1)$. Then the following conditions are equivalent:*

- $(\mathcal{T}_1,\mathcal{A}_1) \sqsubseteq_{\Sigma}^{q} (\mathcal{T}_2,\mathcal{A}_2)$.

- *For all assertions α of the form $C(a)$ and $r(a,b)$, where C is a $\mathcal{EL}^u_\Sigma$-concept, $r \in \Sigma \cap \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$:*
$$(\mathcal{T}_2, \mathcal{A}_2) \models \alpha \Rightarrow (\mathcal{T}_1, \mathcal{A}_1) \models \alpha.$$
- $\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2} \leq_\Sigma^{\mathsf{obj}(\mathcal{A}_2)\cup\{b\},\mathsf{full}} \mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}$.

Proof. The implication from Point 1 to Point 2 is trivial.

Point 2 implies Point 3. Assume that Point 3 does not hold. By the definition of canonical models, simulations, and Theorem 10, at least one of the following conditions holds:

(a) there exist $b_1, b_2 \in \mathsf{obj}(\mathcal{A}_2) \cup \{b\}$ and $r \in \Sigma$ such that $(b_1^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}, b_2^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}) \in r^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}$ and $(b_1^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}, b_2^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}) \notin r^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}$;

(b) there exists an $a \in \mathsf{obj}(\mathcal{A}_2) \cup \{b\}$ and a $\mathcal{EL}^u_\Sigma$-concept C such that $a^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}} \in C^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}$ and $a^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}} \notin C^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}$;

(c) there exists an $\mathcal{EL}^u_\Sigma$-concept C such that $C^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}} \neq \emptyset$ and $C^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}} = \emptyset$.

For suppose that none of the conditions (a)-(c) holds. As (c) does not hold, by Theorem 10, for every $d \in \Delta^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}$ there exists a $d' \in \Delta^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}$ and a Σ-simulation S_d with $(d,d') \in S_d$. Moreover, as (b) does not hold, we may assume that $d' = a^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}$ whenever $d = a^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}$ and $a \in \mathsf{obj}(\mathcal{A}_2) \cup \{b\}$. Using the assumption that (a) does not hold, it follows immediately that $\bigcup_{d \in \Delta^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}} S_d$ is a full Σ-simulation preserving $\mathsf{obj}(\mathcal{A}_2) \cup \{b\}$. We have derived a contradiction.

We now show that each of the conditions (a)-(c) implies that Point 2 does not hold.

Suppose (a) does not hold. By Lemma 27 (3), we have $(\mathcal{T}_2, \mathcal{A}_2) \models r(b_1, b_2)$ and $(\mathcal{T}_1, \mathcal{A}_1) \not\models r(b_1, b_2)$. Thus Point 2 does not hold.

Suppose (b) does not hold. By Lemma 27 (3), we have $(\mathcal{T}_2, \mathcal{A}_2) \models C(a)$ and $(\mathcal{T}_1, \mathcal{A}_1) \not\models C(a)$. Again, Point 2 does not hold.

Suppose (c) does not hold. Take any individual name a. By Lemma 27 (3), we have $(\mathcal{T}_2, \mathcal{A}_2) \models \exists u.C(a)$ and $(\mathcal{T}_1, \mathcal{A}_1) \not\models \exists u.C(a)$. Again, Point 3 does not hold.

Point 3 implies Point 1. Assume Point 3 holds and let $(\mathcal{T}_2, \mathcal{A}_2) \models q(\mathbf{a})$. Take a model $\mathcal{J}$ of $(\mathcal{T}_1, \mathcal{A}_1)$. We show that $\mathcal{J} \models q(\mathbf{a})$. Let Ob be the union of $\mathsf{obj}(\mathcal{A}_1 \cup \mathcal{A}_2)$ and the set of individual names occurring in $\mathbf{a}$. Then $\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2} \leq_\Sigma^{\mathsf{obj}(\mathcal{A}_2)\cup\{b\},\mathsf{full}} \mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}$ implies
$$\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2,\mathsf{Ob}} \leq_\Sigma^{\mathsf{Ob},\mathsf{full}} \mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}$$
because for the largest full Σ-simulation between $\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}$ and $\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}$ preserving $\mathsf{obj}(\mathcal{A}_2) \cup \{b\}$ we have $(b^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}}, d) \in S$ for all $d \in \Delta^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}$ so that we obtain the required full Σ-simulation by adding (c,d) to S for all $c \in \mathsf{Ob} \setminus \mathsf{obj}(\mathcal{A}_2)$ and $d \in \Delta^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}$. Observe that by Lemma 27 (1),
$$\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1} \leq^{\mathsf{Ob},\mathsf{full}} \mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1,\mathsf{Ob}}.$$
Moreover, since $\mathcal{J}$ is a model of $(\mathcal{T}_1, \mathcal{A}_1)$, we obtain from Lemma 27 (2),
$$\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1,\mathsf{Ob}} \leq^{\mathsf{Ob},\mathsf{full}} \mathcal{J}.$$
Because of transitivity of the relation $\leq_\Sigma^{\mathsf{Ob},\mathsf{full}}$, we obtain from the three Σ-simulations above that
$$\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2,\mathsf{Ob}} \leq_\Sigma^{\mathsf{Ob},\mathsf{full}} \mathcal{J}.$$

By Lemma 28, there is a Σ-homomorphism from $\mathcal{I}^{\mathsf{Ob},*}_{\mathcal{A}_2,\mathcal{T}_2,\mathsf{Ob}}$ to $\mathcal{J}$ preserving Ob and, therefore, from $\mathcal{I}^{\mathsf{Ob},*}_{\mathcal{A}_2,\mathcal{T}_2,\mathsf{Ob}} \models q(\mathbf{a})$ we obtain $\mathcal{J} \models q(\mathbf{a})$, as required. □

Note that it is not sufficient to have $\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2} \leq_\Sigma^{\mathsf{obj}(\mathcal{A}_2),\mathsf{full}} \mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}$ in Point 3 of Lemma 29: consider the TBoxes $\mathcal{T}_1 = \emptyset$ and $\mathcal{T}_2 = \{A \equiv \top\}$, the ABox $\mathcal{A}_1 = \mathcal{A}_2 = \{A(a)\}$, and the signature $\Sigma = \{A\}$. Then $(\mathcal{T}_1, \mathcal{A}_1) \not\sqsupseteq_\Sigma^q (\mathcal{T}_2, \mathcal{A}_2)$ because $(\mathcal{T}_2, \mathcal{A}_2) \models A(b)$ and $(\mathcal{T}_1, \mathcal{A}_1) \not\models A(b)$ for $b \neq a$. But we have $\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2} \leq_\Sigma^{\mathsf{obj}(\mathcal{A}_2),\mathsf{full}} \mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}$ because $\Delta^{\mathcal{I}_{\mathcal{A}_2,\mathcal{T}_2}} \times \{a^{\mathcal{I}_{\mathcal{A}_1,\mathcal{T}_1}}\}$ is a full Σ-simulation preserving $\{a\}$.

Proposition 30. *For all TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and any signature Σ: $\mathcal{T}_1 \sqsupseteq_\Sigma^u \mathcal{T}_2$ iff $\mathcal{T}_1 \sqsupseteq_\Sigma^q \mathcal{T}_2$.*

Proof. Suppose $\mathcal{T}_1 \not\sqsupseteq_\Sigma^u \mathcal{T}_2$. Take an $\mathcal{EL}_\Sigma$-concept C and $\mathcal{EL}_\Sigma^u$-concept D such that $\mathcal{T}_2 \models C \sqsubseteq D$ but $\mathcal{T}_1 \not\models C \sqsubseteq D$. Let $\mathcal{A} = \{C(a)\}$. If D is an $\mathcal{EL}_\Sigma$-concept, then $(\mathcal{T}_1, \mathcal{A}) \not\models D(a)$ but $(\mathcal{T}_2, \mathcal{A}) \models D(a)$. If $D = \exists u.D'$, then $(\mathcal{T}_1, \mathcal{A}) \not\models \exists x D(x)$, but $(\mathcal{T}_2, \mathcal{A}) \models \exists x D(x)$. It follows that $\mathcal{T}_1 \not\sqsupseteq_\Sigma^q \mathcal{T}_2$.

Conversely, let $\mathcal{T}_1 \not\sqsupseteq_\Sigma^q \mathcal{T}_2$. By the equivalence of Points 1 and 2 in Lemma 29 (and the fact that $(\mathcal{T}_1, \mathcal{A}) \models r(a, b)$ iff $(\mathcal{T}_2, \mathcal{A}) \models r(a, b)$, for any assertion $r(a, b)$), $(\mathcal{T}_1, \mathcal{A}) \not\models D(a)$ but $(\mathcal{T}_2, \mathcal{A}) \models D(a)$, for some Σ-ABox $\mathcal{A}$ and $\mathcal{EL}_\Sigma^u$-concept D. Then, by Lemma 22, Point 2, there exists a $\mathcal{EL}_\Sigma^u$-concept C such that $\mathcal{T}_2 \models C \sqsubseteq D$ and $\mathcal{A} \models C(a)$. Again by Lemma 22, Point 2, $\mathcal{T}_1 \not\models C \sqsubseteq D$. If D is an $\mathcal{EL}$-concept, then we can assume that C is an $\mathcal{EL}$-concept. (To see this observe that $\mathcal{T} \models \exists u.C_0 \sqsubseteq D_0$ implies $\mathcal{T} \models \top \sqsubseteq D_0$, for all $\mathcal{EL}$-concepts C_0, D_0 and TBoxes $\mathcal{T}$.) Thus, we even have $\mathcal{T}_1 \not\sqsupseteq_\Sigma \mathcal{T}_2$. If D is an $\mathcal{EL}^u$-concept, let $C' = C_0$ if $C = \exists u.C_0$ and $C' = C$, otherwise. Then we still have $\mathcal{T}_2 \models C' \sqsubseteq D$ and $\mathcal{T}_1 \not\models C' \sqsubseteq D$. Hence $\mathcal{T}_1 \not\sqsupseteq_\Sigma^u \mathcal{T}_2$. □

5.3. The Algorithm for $\mathcal{QL}_{\mathcal{EL}}^u$

The aim of this section is extend the algorithm from Figure 1 to an algorithm deciding non-Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}^u$. Before we go into this, we establish an illustrative lemma which shows that the difference between Σ-entailment w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ and $\mathcal{QL}_{\mathcal{EL}}^u$ is due to non-Σ roles in the TBox $\mathcal{T}_2$.

Lemma 31. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes, and Σ a signature that contains all role names occurring in $\mathcal{T}_2$. Then $\mathcal{T}_1 \sqsupseteq_\Sigma \mathcal{T}_2$ iff $\mathcal{T}_1 \sqsupseteq_\Sigma^u \mathcal{T}_2$.*

Proof. The "if" direction is clear. For the "only if" direction, let C, D be $\mathcal{EL}_\Sigma$-concepts such that $\mathcal{T}_1 \not\models C \sqsubseteq \exists u.D$, and $\mathcal{T}_2 \models C \sqsubseteq \exists u.D$. The latter implies that, in the canonical model $\mathcal{I}_{C,\mathcal{T}_2}$, there is a $d \in \Delta^{\mathcal{I}_{C,\mathcal{T}_2}}$ with $d \in D^{\mathcal{I}_{C,\mathcal{T}_2}}$. This implies that there is a sequence $r_1 \cdots r_k$ of role names from $\mathsf{sig}(C) \cup \mathsf{sig}(\mathcal{T}_2)$ such that d is reachable from C in $\mathcal{I}_{C,\mathcal{T}_2}$ along $d_1, \ldots, d_{k-1} \in \Delta^{\mathcal{I}_{C,\mathcal{T}_2}}$:

$$(C, d_1) \in r_1^{\mathcal{I}_{C,\mathcal{T}_2}}, \quad (d_1, d_2) \in r_2^{\mathcal{I}_{C,\mathcal{T}_2}}, \ldots, \quad (d_{k-1}, d) \in r_k^{\mathcal{I}_{C,\mathcal{T}_2}}.$$

By Point (2) of Lemma 13, this implies $\mathcal{T}_2 \models C \sqsubseteq \exists r_1. \cdots \exists r_k.D$. Since $\mathsf{sig}(\mathcal{T}_2) \cap \mathsf{N_R} \subseteq \Sigma$, $\exists r_1. \cdots \exists r_k.D$ is a Σ-concept. Moreover, $\mathcal{T}_1 \not\models C \sqsubseteq \exists u.D$ and $\emptyset \models \exists r_1. \cdots \exists r_k.D \sqsubseteq \exists u.D$ implies $\mathcal{T}_1 \not\models C \sqsubseteq \exists r_1. \cdots \exists r_k.D$. Thus, $\mathcal{T}_1 \not\sqsupseteq_\Sigma \mathcal{T}_2$. □

Now for the extension of the algorithm in Figure 1. To take into account consequences of the form $C \sqsubseteq \exists u.D$ we work, in addition to the sets $K_{\mathcal{T}}(C)$, with the set

$$K^u_{\mathcal{T}}(C) = \{D \mid \exists r.D \in \mathsf{sub}(\mathcal{T}), \mathcal{T} \models C \sqsubseteq \exists u.D\}.$$

We extend Proposition 17 as follows.

Proposition 32. *Assume $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$. $\mathcal{T}_1$ does not Σ-entail $\mathcal{T}_2$ w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$ iff*

(1) there exist $\mathcal{EL}_\Sigma$-concepts C and D satisfying the conditions of Proposition 17 or

(2) there exists an $\mathcal{EL}_\Sigma$-concept C and $D \in K^u_{\mathcal{T}_2}(C)$ such that

(a) there does not exist $D' \in \Delta^{\mathcal{I}_{C,\mathcal{T}_1}}$ with $(\mathcal{T}_1, D') \sqsubseteq_\Sigma (\mathcal{T}_2, D)$;

(b) the outdegree of C is bounded by $|\mathcal{T}_2|$.

Proof. Assume $\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$. By Proposition 17, $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$ iff Point 1 is satisfied. So it remains to consider the case $\mathcal{T}_1 \sqsubseteq_\Sigma \mathcal{T}_2$. Thus, by Lemma 15, there exists an $\mathcal{EL}_\Sigma$-concept C such that $(\mathcal{I}_{C,\mathcal{T}_2}, C) \leq_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, C)$ but $(\mathcal{I}_{C,\mathcal{T}_2}, C) \not\leq^{\mathsf{full}}_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, C)$. This means that there exists $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}_2}}$ such that

- there is no path from C to D following the relation $\bigcup_{r\in\Sigma} r^{\mathcal{I}_{C,\mathcal{T}_2}}$;
- there does not exist a $D' \in \Delta^{\mathcal{I}_{C,\mathcal{T}_1}}$ with $(\mathcal{I}_{C,\mathcal{T}_2}, D) \leq_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, D')$.

Take such a D. It follows that $D \in K^u_{\mathcal{T}_2}(C)$ because for all $D_0 \in \Delta^{\mathcal{I}_{C,\mathcal{T}_2}} \setminus K^u_{\mathcal{T}_2}(C)$ there exists a path violating Point 1. Hence, by Lemma 15, C and D are as required for (a). It remains to show that one can obtain C and D satisfying, in addition, (b). This is shown in Lemma 50 in the appendix.

Conversely, suppose Point 1 or Point 2 holds. If Point 1 holds, then $\mathcal{T}_1 \not\sqsubseteq_\Sigma \mathcal{T}_2$, and so $\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$. Now suppose that Point 2 holds. Take C and $D \in K^u_{\mathcal{T}_2}(C)$ such that (a) holds. Then $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}_2}}$ but there does not exist $D' \in \Delta^{\mathcal{I}_{C,\mathcal{T}_1}}$ with $(\mathcal{I}_{C,\mathcal{T}_2}, D) \leq_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, D')$ (Lemma 15). This implies $(\mathcal{I}_{C,\mathcal{T}_2}, C) \not\leq^{\mathsf{full}}_\Sigma (\mathcal{I}_{C,\mathcal{T}_1}, C)$ and so $\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$. ☐

We now give the algorithm deciding $\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$ by extending the algorithm from Figure 1. The additional code implements directly Point 2 of Proposition 32. Instead of quadruples representing concepts, we now work with 7-tuples where the additional three entries store the information relevant for dealing with the universal role. Namely, we set for $C = F \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$,

$$C^\sharp = (F, K_{\mathcal{T}_1}(C), K_{\mathcal{T}_2}(C), K_{\mathcal{T}_1,\mathcal{T}_2}(C), K^u_{\mathcal{T}_1}(C), K^u_{\mathcal{T}_2}(C), \bigcup_{C'\in\Delta^{\mathcal{I}_{C,\mathcal{T}_1}}\setminus\{C\}} K_{\mathcal{T}_1,\mathcal{T}_2}(C')),$$

Due to the following lemma and Lemma 18, 7-tuples can be computed recursively similar to the 4-tuples used before.

Lemma 33. *Let $\mathcal{T}$ be a TBox, $C = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$ and $D = \bigsqcap_{E\in K_{\mathcal{T}}(C)} E$. Then*

$$K^u_{\mathcal{T}}(C) = K^u_{\mathcal{T}}(D) \cup \bigcup_{(r,E)\in P} K^u_{\mathcal{T}}(E).$$

The algorithm is now given in Figure 2. Observe that, compared to Figure 1, we have only added one more sufficient condition (the second condition in Steps 2 and 3) under which the algorithm outputs $\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$ and the computation of the three new

Input: TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$ and signature $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$.

(1) Compute the set $\mathcal{N}_0$ of 7-tuples determined by conjunctions of concept names in Σ.

(2) if $\mathcal{N}_0$ contains a 7-tuple $(F, \mathcal{Q}_1, \ldots, \mathcal{Q}_6)$ such that $\mathcal{Q}_2 \setminus \mathcal{Q}_3 \neq \emptyset$ or $\mathcal{Q}_5 \setminus (\mathcal{Q}_3 \cup \mathcal{Q}_6) \neq \emptyset$, then output "$\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$".

(3) Generate the sequence $\mathcal{N}_1, \mathcal{N}_2, \ldots$ of sets of 7-tuples with $\mathcal{N}_{i+1} = \mathcal{N}_i \cup \mathcal{N}_i'$, where $\mathcal{N}_i'$ is the set of 7-tuples $(F_0, \mathcal{F}_1, \ldots, \mathcal{F}_6)$ which can be obtained from a conjunction F_0 of concept names from Σ and a set $Q \subseteq (\mathsf{N_R} \cap \Sigma) \times \mathcal{N}_i$ of cardinality not exceeding $|\mathcal{T}_2|$ in the following way:

- $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ are computed from the components $\mathcal{Q}_1$, $\mathcal{Q}_2$ and $\mathcal{Q}_3$ of the 7-tuples in Q as in Figure 1.
- $\mathcal{F}_4 = K^u_{\mathcal{T}_1}(\bigsqcap_{D \in \mathcal{F}_1} D) \cup \bigcup_{(r,(F,\mathcal{Q}_1,\ldots,\mathcal{Q}_6)) \in Q} \mathcal{Q}_4$.
- $\mathcal{F}_5 = K^u_{\mathcal{T}_2}(\bigsqcap_{D \in \mathcal{F}_2} D) \sqcap \bigcup_{(r,(F,\mathcal{Q}_1,\ldots,\mathcal{Q}_6)) \in Q} \mathcal{Q}_5$.
- $\mathcal{F}_6 = \bigcup_{(r,(F,\mathcal{Q}_1,\ldots,\mathcal{Q}_6)) \in Q} \mathcal{Q}_6 \cup \bigcup_{C' \in \mathcal{F}_4} K_{\mathcal{T}_1,\mathcal{T}_2}(C')$

This is done until $\mathcal{N}_i$ contains a 7-tuple $(F, \mathcal{Q}_1, \ldots, \mathcal{Q}_6)$ such that $\mathcal{Q}_2 \setminus \mathcal{Q}_3 \neq \emptyset$ or $\mathcal{Q}_5 \setminus (\mathcal{Q}_3 \cup \mathcal{Q}_6) \neq \emptyset$ or $\mathcal{N}_{i+1} = \mathcal{N}_i$. Output "$\mathcal{T}_1 \not\sqsubseteq^u_\Sigma \mathcal{T}_2$" if one of the two first conditions applies. Otherwise, output "$\mathcal{T}_1 \sqsubseteq^u_\Sigma \mathcal{T}_2$".

Fig. 2. Algorithm deciding Σ-entailment w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$.

components $\mathcal{F}_4$, $\mathcal{F}_5$, and $\mathcal{F}_6$ of the new 7-tuples generated in Step 3. The new condition, $\mathcal{Q}_5 \setminus (\mathcal{Q}_3 \cup \mathcal{Q}_6) \neq \emptyset$, corresponds exactly to Point 2 of Proposition 32: there exists $D \in \mathcal{Q}_5 \setminus (\mathcal{Q}_3 \cup \mathcal{Q}_6)$ iff there exists $D \in K^u_{\mathcal{T}_2}(C)$ (meaning $D \in \mathcal{Q}_5$) such that there does not exist $D' \in \Delta^{\mathcal{I}_{C,\mathcal{T}_1}} = \{C\} \cup \{C' \mid \exists r.C' \in \mathsf{sub}(C)\} \cup K^u_{\mathcal{T}_1}(C)$ with $(\mathcal{T}_1, D') \sqsubseteq_\Sigma (\mathcal{T}_2, D)$. To prove completeness and soundness it is, therefore, sufficient to prove that the computations of $\mathcal{F}_4$, $\mathcal{F}_5$ and $\mathcal{F}_6$ are correct. For $\mathcal{F}_4$ and $\mathcal{F}_5$ this follows from Lemma 33, and for $\mathcal{F}_6$ this is trivial. Termination after at most exponentially many steps can be proved similarly to the proof for the algorithm in Figure 1 and is left to the reader. With Lemma 30, we thus obtain the following result.

Theorem 34. *Σ-entailment of $\mathcal{EL}$-TBoxes w.r.t. $\mathcal{QL}^q_{\mathcal{EL}}$ is in* ExpTime.

6. ExpTime-hardness

We prove that the ExpTime upper bounds stated in Theorem 21, 24, and 34 are tight by establishing matching lower bounds. The lower bounds apply already to conservative extensions, i.e., the special case of Σ-inseparability where $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\Sigma = \mathsf{sig}(\mathcal{T}_1)$. By the equivalences established in the preceding section, it suffices to consider the query languages $\mathcal{QL}_{\mathcal{EL}}$ and $\mathcal{QL}^u_{\mathcal{EL}}$. We start with the former.

The proof is by reduction of the problem of determining whether Player 1 has a winning strategy in version G_5 of the two-player game Peek which was introduced and proved to be ExpTime-complete in (23). An instance of Peek is a tuple $(\Gamma_1, \Gamma_2, \Gamma_I, \varphi)$ where:

- Γ_1 and Γ_2 are disjoint, finite sets of Boolean variables, with variables in Γ_1 under the control of Player 1, and variables in Γ_2 under the control of Player 2;
- $\Gamma_I \subseteq (\Gamma_1 \cup \Gamma_2)$ are the variables that are true in the initial state of the game;
- φ is a propositional logic formula over the variables $\Gamma_1 \cup \Gamma_2$ which represents the winning condition.

The game is played in a series of rounds. Each round produces an assignment for the variables in $\Gamma_1 \cup \Gamma_2$, and the game starts with the initial assignment Γ_I. The players alternate, with Player 1 moving first. In each turn of Player $i \in \{1, 2\}$, he selects a variable from Γ_i whose truth value is flipped to reach the next assignment. All other variables retain their truth value. A player may also make a skip move, i.e., not change any of his variables. Player 1 wins if the formula φ ever becomes true. Player 2 wins if he can forever prevent φ from becoming true.

Formally, a *configuration* of Peek is a pair (t, p) where t is a truth assignment for the variables in $\Gamma_1 \cup \Gamma_2$ and $p \in \{1, 2\}$ indicates the player that has to move next. A *winning strategy for Player 1* is a finite node-labeled tree (V, E, ℓ) where ℓ is a node labeling function that assigns to each node a configuration of G such that

(1) the root is labeled with $(\Gamma_I, 1)$;
(2) if an inner node is labeled with $(t, 1)$, then it has a single successor labeled $(t', 2)$, where t' is obtained from t by switching the truth value of at most one variable from Γ_1;
(3) if an inner node is labeled with $(t, 2)$, then it has ℓ successors labeled $(t_0, 1), \ldots, (t_\ell, 1)$, where $t_0, \ldots, t_\ell$ are the configurations of G that can be obtained from t by switching the truth value of at most one variable from Γ_2;
(4) if a leaf is labeled (t, i), then t satisfies φ.

Given a game instance $G = (\Gamma_1, \Gamma_2, \Gamma_I, \varphi)$, we define TBoxes $\mathcal{T}_G$ and $\mathcal{T}'_G$ such that $\mathcal{T}_G \cup \mathcal{T}'_G$ is not a conservative extension of $\mathcal{T}_G$ iff Player 1 has a winning strategy in G. Intuitively, witnesses $C \sqsubseteq D$ against conservativity are such that C describes a winning strategy for Player 1 in G and, conversely, every winning strategy can be converted into a witness against conservativity. For convenience, we assume that the set of variables $\Gamma_1 \cup \Gamma_2$ is of the form $\{0, \ldots, n-1\}$ for some $n \geq 1$. To describe winning strategies as concepts, we use the following symbols:

- $V_0, \ldots, V_{n-1}$ and $\overline{V}_0, \ldots, \overline{V}_{n-1}$ to describe the truth values of the variables;
- $F_0, \ldots, F_n$ to denote the variable that is flipped to reach the current configuration, with F_n indicating a skip move;
- P_1, P_2 to denote the player which moves next;
- a single role name r.

Since $\mathcal{EL}$-concepts correspond to trees in an obvious way (every existential restriction $\exists r.C$ gives rise to an edge), it is not hard to see how winning strategies can be represented as a concept formulated in the above signature.

In $\mathcal{T}_G$, we additionally use a concept name B that will occur on the right-hand side of witnesses against conservativity, and a concept name M that serves as a marker. The construction of $\mathcal{T}_G$ starts with saying that the players alternate:

$$\exists r.P_1 \sqsubseteq P_2$$
$$\exists r.P_2 \sqsubseteq P_1$$

Then, we say that P_1 and P_2 should be disjoint. The idea is as follows: every concept C which implies that $P_1 \sqcap P_2$ is true somewhere in the model is subsumed by the concept name B already w.r.t. $\mathcal{T}_G$, and thus cannot occur on the left-hand side of a witness $C \sqsubseteq B$. Here we use the concept name M:

$$\begin{aligned} P_1 \sqcap P_2 &\sqsubseteq M \\ \exists r.M &\sqsubseteq M \\ M &\sqsubseteq B \end{aligned}$$

We also need disjointness conditions for truth values and flipping markers:

$$\begin{aligned} V_i \sqcap \overline{V}_i &\sqsubseteq M \quad \text{for all } i < n \\ F_i \sqcap F_j &\sqsubseteq M \quad \text{for all } i, j \leq n \text{ with } i \neq j \end{aligned}$$

Next, we say that if the marker F_i is set in a configuration, then the variable V_i flips:

$$\begin{aligned} \exists r.(F_i \sqcap V_i) &\sqsubseteq \overline{V}_i \quad \text{for all } i < n \\ \exists r.(F_i \sqcap \overline{V}_i) &\sqsubseteq V_i \quad \text{for all } i < n \end{aligned}$$

If a marker F_j for a different variable V_j is set, then V_i does not flip:

$$\begin{aligned} \exists r.(F_i \sqcap V_j) &\sqsubseteq V_j \quad \text{for all } i \leq n \text{ and } j < n \text{ with } i \neq j \\ \exists r.(F_i \sqcap \overline{V}_j) &\sqsubseteq \overline{V}_j \quad \text{for all } i \leq n \text{ and } j < n \text{ with } i \neq j \end{aligned}$$

Additionally, we should ensure that at least one of the F_i markers is true in every configuration. This cannot be done in a straightforward way in $\mathcal{T}_G$, and we will use the TBox $\mathcal{T}'_G$.

To define $\mathcal{T}'_G$, we start with translating the formula φ into a set of CIs. W.l.o.g., we assume that φ is in NNF. For each $\psi \in \mathsf{sub}(\varphi)$, we introduce a concept name X_ψ. For each $\psi \in \mathsf{sub}(\varphi)$, we use $\sigma(\psi)$ to denote

- the concept name X_ψ if ψ is a non-literal and
- the concept name from $V_0, \dots, V_{n-1}, \overline{V}_0, \dots, \overline{V}_{n-1}$ corresponding to ψ if ψ is a literal.

For each non-literal $\psi \in \mathsf{sub}(\varphi)$, $\mathcal{T}'_G$ contains the following CI:

- if $\psi = \vartheta \wedge \chi$, then the CI is $\sigma(\vartheta) \sqcap \sigma(\chi) \sqsubseteq X_\psi$;
- if $\psi = \vartheta \vee \chi$, then the CIs are $\sigma(\vartheta) \sqsubseteq X_\psi$ and $\sigma(\chi) \sqsubseteq X_\psi$.

To continue, let $\Gamma_1 = \{0, \dots, k-1\}$ and $\Gamma_2 = \{k, \dots, n\}$, and introduce concept names $N, N', N'', N_0, \dots, N_{n-1}$ to be used as markers. The markers will help to ensure that (i) each variable has a truth value in every configuration, (ii) a least one of the flipping indicators $F_0, \dots, F_n$ is set in every configuration, and (iii) the flipping indicator denotes a variable controlled by the player who moved to reach the current configuration. The markers are set as follows:

$$\begin{aligned} V_i &\sqsubseteq N_i \quad \text{for all } i < n \\ \overline{V}_i &\sqsubseteq N_i \quad \text{for all } i < n \\ F_i &\sqsubseteq N' \quad \text{for all } i \in \{0, \dots, k-1, n\} \\ F_i &\sqsubseteq N'' \quad \text{for all } i \in \{k, \dots, n\} \end{aligned}$$

Next, we set the marker N if the encoded truth assignment satisfies φ and (i)-(iii) are satisfied:

$$X_\varphi \sqcap P_1 \sqcap N'' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1} \sqsubseteq N$$

$$X_\varphi \sqcap P_2 \sqcap N' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1} \sqsubseteq N$$

Then, the marker N is pulled up inductively ensuring that if Player 1 is to move, there is the required single successor, and if Player 2 is to move, there are the required $k+1$ successors:

$$P_1 \sqcap N'' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1} \sqcap \exists r.N \sqsubseteq N$$

$$P_2 \sqcap N' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1} \sqcap \bigsqcap_{i \in \{0,\dots,k-1,n\}} \exists r.(N \sqcap F_i) \sqsubseteq N$$

We require that P_1 moves first and that the initial configuration is labeled as described by Γ_I. Only if this is satisfied, the concept name B from $\mathcal{T}_G$ is implied:

$$P_1 \sqcap N \sqcap \bigsqcap_{i \in \Gamma_I} V_i \sqcap \bigsqcap_{i \notin \Gamma_I} \overline{V}_i \sqsubseteq B$$

Finally, we also deal with the case where already Γ_I satisfies φ:

$$P_1 \sqcap X_\varphi \sqcap \bigsqcap_{i \in \Gamma_I} V_i \sqcap \bigsqcap_{i \notin \Gamma_I} \overline{V}_i \sqsubseteq B$$

The following lemma is proved in Appendix E.

Lemma 35. *Player 1 has a winning strategy in G iff $\mathcal{T}_G \cup \mathcal{T}'_G$ is not a conservative extension of $\mathcal{T}_G$.*

We have thus shown the following result.

Theorem 36. *Deciding conservative extensions w.r.t. $\mathcal{QL}_{\mathcal{EL}}$ is* ExpTime*-hard and thus* ExpTime*-complete.*

Together with Lemmas 23, 30, and 31 and since the only role name in $\mathcal{T}'_G$ is from $\Sigma = \mathsf{sig}(\mathcal{T}_G)$, we obtain the following corollary.

Corollary 37. *For $\mathcal{QL} \in \{\mathcal{QL}_{\mathcal{EL}}, \mathcal{QL}^i_{\mathcal{EL}}, \mathcal{QL}^q_{\mathcal{EL}}\}$, deciding conservative extensions w.r.t. $\mathcal{QL}$ is* ExpTime*-hard and thus* ExpTime*-complete.*

7. Model Conservativity

We consider Σ-entailment w.r.t. second-order logic. Denote by SO the set of second-order sentences with second-order variables for sets and binary relations in the signature with unary predicates from $\mathsf{N_C}$ and binary predicates from $\mathsf{N_R}$. Clearly, Σ-entailment between two TBoxes w.r.t. SO implies Σ-entailment w.r.t. any other query language introduced so far.

We start with observing that Σ-entailment w.r.t. SO can be easily characterized model-theoretically without using any query language. Say that two interpretations $\mathcal{I}$ and $\mathcal{J}$ *coincide on a signature* Σ, in symbols $\mathcal{I}|_\Sigma = \mathcal{J}|_\Sigma$, if $\Delta^\mathcal{I} = \Delta^\mathcal{J}$ and $X^\mathcal{I} = X^\mathcal{J}$ for all $X \in \Sigma$.

Definition 38 (Semantic Σ-consequence, model conservative extension)**.** Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes and Σ a signature. Then

- $\mathcal{T}_2$ is a *semantic Σ-consequence* of $\mathcal{T}_1$ if for every model $\mathcal{I}$ of $\mathcal{T}_1$ there exists a model $\mathcal{J}$ of $\mathcal{T}_2$ that coincides with $\mathcal{I}$ on Σ.
- $\mathcal{T}_2$ is a *model conservative extension* of $\mathcal{T}_1$ if $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{T}_2$ is a semantic Σ-consequence of $\mathcal{T}_1$ for $\Sigma = \mathsf{sig}(\mathcal{T}_1)$.

Model conservative extensions are a well-known notion in mathematical logic and modular software verification (13). The relation between deduction-based notions of conservativity and model-conservativity in modular software design is discussed in (25; 26; 8). The following lemma relates Σ-entailment w.r.t. SO and semantic Σ-consequence.

Lemma 39. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes and Σ a signature. Then*

- *$\mathcal{T}_2$ is a semantic Σ-consequence of $\mathcal{T}_1$ iff $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ w.r.t.* SO*.*
- *$\mathcal{T}_2$ is a model-conservative extension of $\mathcal{T}_1$ iff $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ w.r.t.* SO*.*

Proof. Point 2 follows from Point 1, so we concentrate on Point 1. The implication from left to right follows from the fact that no second-order formula using only symbols from Σ can distinguish two models whose Σ-reducts are isomorphic. For the other direction observe that $\mathcal{T}_2 \models \exists S_1 \cdots \exists S_n. \bigwedge \mathcal{T}_2$ with $\{S_1, \ldots, S_n\} = \mathsf{sig}(\mathcal{T}_2) \setminus \Sigma$. Thus, if $\mathcal{T}_1$ Σ-entails $\mathcal{T}_2$ w.r.t. SO, then $\mathcal{T}_1 \models \exists S_1 \cdots \exists S_n. \bigwedge \mathcal{T}_2$ which means that for every interpretation $\mathcal{I}$ satisfying $\mathcal{T}_1$ there exists an interpretation $\mathcal{J}$ of $\mathcal{T}_2$ which coincides with $\mathcal{I}$ on Σ, as required. $\Box$

The proof also shows that if a TBox $\mathcal{T}_2$ is not Σ-entailed by a TBox $\mathcal{T}_1$ w.r.t. SO, then there is a witness of the form $\exists S_1 \cdots \exists S_n. \bigwedge \mathcal{T}_2$.

For TBoxes formulated in the description logic $\mathcal{ALC}$, model conservativity has been proved Π^1_1-hard in (18). In this section, we show that model conservative extensions, and therefore also semantic Σ-consequence, are undecidable even in $\mathcal{EL}$ (we leave Π^1_1-hardness as an open question). The proof is by reduction of the halting problem for deterministic Turing machines on the empty tape. We assume w.l.o.g. that the Turing machines are such that (i) the initial state is not reachable from itself, (ii) the halting state does not allow any further transitions, and (iii) all transitions move the head either right or left. Let $M = (Q, \Gamma, \Delta, q_0, q_h)$ be such a deterministic Turing machine, where Q is a set of states, Γ an alphabet, Δ a partial transition function, $q_0 \in Q$ the starting state, and $q_h \in Q$ the halting state. We construct TBoxes $\mathcal{T}_M$ and $\mathcal{T}'_M$ such that $\mathcal{T}_M \cup \mathcal{T}'_M$ is not a model conservative extension of $\mathcal{T}_M$ iff M reaches q_h from q_0 on the empty tape. We use the following concept and role names for describing computations of M:

- the elements of Q and Γ as concept names;
- concept names head, before, and after to represent the relation of a tape cell to the head position;
- role names n (for *next tape cell*) and s (for *successor configuration*).

Our construction is such that models $\mathcal{I}$ of $\mathcal{T}_M$ for which there does not exist a model $\mathcal{J}$ of $\mathcal{T}'_M$ which coincides with $\mathcal{I}$ on $\mathsf{sig}(\mathcal{T}_M)$ describe halting computations of M on the empty tape. Essentially, such models have the form of a grid, with the vertical edges labeled s and the horizontal ones labeled n. Thus, each row represents a configuration. We will enforce the roles n and s to be functional, except at row 0 and column 0 (because this does not seem possible). Therefore, the actual grid representing the computation of M starts at row 1 and column 1.

We start with the definition of $\mathcal{T}_M$. For now, it is easiest to simply assume n and s to be functional and confluent (which will be enforced later by $\mathcal{T}'_M$). We first set before and after correctly, exploiting the assumed functionality of n:

$$\exists n.\mathsf{before} \sqsubseteq \mathsf{before} \quad \exists n.\mathsf{head} \sqsubseteq \mathsf{before} \quad \mathsf{head} \sqsubseteq \exists n.\mathsf{after} \quad \mathsf{after} \sqsubseteq \exists n.\mathsf{after}.$$

Then we say that states are uniform over the tape: for all $q \in Q$,

$$q \sqsubseteq \exists n.q \quad \exists n.q \sqsubseteq q.$$

Exploiting that q_0 cannot reach itself and the above uniformity, we say that the tape is initially blank (where $b \in \Gamma$ is the blank symbol):

$$q_0 \sqsubseteq b.$$

For each transition $\delta(q,a) = (q',a',L)$, exploiting confluence of n and s, we set

$$\exists n.(q \sqcap \mathsf{head} \sqcap a) \sqsubseteq \exists s.(q' \sqcap \mathsf{head} \sqcap \exists n.a'),$$

and for each transition $\delta(q,a) = (q',a',R)$,

$$(q \sqcap \mathsf{head} \sqcap a) \sqsubseteq \exists s.(a' \sqcap q' \sqcap \exists n.\mathsf{head}).$$

We also say that symbols not under the head do not change: for all $a \in \Gamma$, put

$$a \sqcap \mathsf{before} \sqsubseteq \exists s.a, \;\; a \sqcap \mathsf{after} \sqsubseteq \exists s.a.$$

We would like to say that certain concept names such as before and head are disjoint. Since disjointness cannot be expressed in $\mathcal{EL}$, we revert to a trick that will become clear when $\mathcal{T}'_M$ is defined. For now, we introduce a concept name D that serves as a marker for problems with disjointness: for all $q, q' \in Q$ with $q \neq q'$ and all $a, a' \in \Gamma$ with $a \neq a'$, put

$$q \sqcap q' \sqsubseteq D \quad a \sqcap a' \sqsubseteq D \quad \mathsf{before} \sqcap \mathsf{head} \sqsubseteq D \quad \mathsf{head} \sqcap \mathsf{after} \sqsubseteq D \quad \mathsf{before} \sqcap \mathsf{after} \sqsubseteq D.$$

Up to now, we simply have assumed the described grid structure, but we did not enforce it. In $\mathcal{T}_M$, we cannot do much more than saying that every point has the required successors:

$$\top \sqsubseteq \exists n.\top \sqcap \exists s.\top.$$

We now define $\mathcal{T}'_M$, introducing new concept names N, A, B and a new role u_0. The concept name N serves as a marker. It is enforced to be true at the origin of the relevant part of the grid (point (1,1)) if the described computation reaches the halting state:

$$q_h \sqsubseteq N \quad \exists n.N \sqsubseteq N \quad \exists s.N \sqsubseteq N$$

It remains to ensure that for a model $\mathcal{I}$ of $\mathcal{T}_M$ there does not exist a model $\mathcal{J}$ of $\mathcal{T}'_M$ which coincides with $\mathcal{I}$ on $\mathsf{sig}(\mathcal{T}_M)$ iff (i) r and s are functional, (ii) r and s are confluent, (iii) $D^{\mathcal{I}} = \emptyset$ (because then there are no problems with disjointness), (iv) the described computation starts in the starting state with the head on the left-most cell and reaches the halting state. Surprisingly, all this can be achieved with two simple CIs:

$$\exists n.\exists s.(N \sqcap q_0 \sqcap \mathsf{head}) \sqsubseteq \exists u_0.(\exists n.\exists s.A \sqcap \exists s.\exists n.B)$$

$$A \sqcap B \sqsubseteq \exists u_0.D$$

The following lemma is proved in the appendix.

Lemma 40. *$\mathcal{T}_M \cup \mathcal{T}'_M$ is not a model conservative extension of $\mathcal{T}_M$ iff M halts on the empty tape.*

We have thus shown the following.

Theorem 41. *The problem of checking whether an $\mathcal{EL}$-TBox $\mathcal{T}_1$ is a model conservative extension of an $\mathcal{EL}$-TBox $\mathcal{T}_2$ is undecidable.*

8. Conclusion

We have introduced different notions of entailment and inseparability between TBoxes and of conservative extensions of TBoxes. Concentrating on the lightweight description logic $\mathcal{EL}$, we have then studied the robustness of these notions and analyzed their interrelationship and computational properties. In particular, we have shown that a variety of '$\mathcal{EL}$-based' notions of entailment is ExpTime-complete, but that Σ-entailment w.r.t. SO is undecidable.

Our analysis leaves open a number of interesting questions, of which we discuss three. First, the following notion of Σ-entailment has been suggested in (10; 11; 17):

Definition 42. Let $\mathcal{QL}$ be a query language, Σ a signature, and $\mathcal{T}_1$, $\mathcal{T}_2$ TBoxes. Then $\mathcal{T}_1$ and $\mathcal{T}_2$ are *strongly Σ-inseparable* w.r.t. $\mathcal{QL}$ if for all TBoxes $\mathcal{T}$ with $\mathsf{sig}(\mathcal{T}) \cap \mathsf{sig}(\mathcal{T}_i) \subseteq \Sigma$ for all $i \in \{1, 2\}$, $\mathcal{T}_1 \cup \mathcal{T}$ and $\mathcal{T}_2 \cup \mathcal{T}$ are Σ-inseparable w.r.t. $\mathcal{QL}$.

This notion is relevant for importing a TBox into another one: if $\mathcal{T}_1$ and $\mathcal{T}_2$ are strongly Σ-inseparable, then it is safe to import $\mathcal{T}_1$ instead of $\mathcal{T}_2$ into *any* TBox $\mathcal{T}$ if no non-Σ symbols from $\mathcal{T}_1$ and $\mathcal{T}_2$ are used in $\mathcal{T}$. Decidability and the exact complexity of strong Σ-inseparability are yet unknown for the case of general $\mathcal{EL}$-TBoxes.

Second, it would be interesting to carry out a more detailed analysis of how the two inputs $\mathcal{T}_1$ and $\mathcal{T}_2$ contribute to the complexity of deciding Σ-entailment. In particular, it would be of interest to know whether there is an algorithm that, given two general $\mathcal{EL}$-TBoxes $\mathcal{T}_1$ and $\mathcal{T}_2$,

(1) decides whether $\mathcal{T}_1 \cup \mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ and
(2) needs time polynomial in $\mathcal{T}_1$ and exponential in $\mathcal{T}_2$.

Note that we assume $\Sigma = \mathsf{sig}(\mathcal{T}_1)$, and that the second input consists only of $\mathcal{T}_2$, and not $\mathcal{T}_1 \cup \mathcal{T}_2$. Such a result would be in line with results on conservative extensions of $\mathcal{ALC}$ TBoxes obtained in (14). They would be quite relevant since the extension $\mathcal{T}_2$ is usually small compared to the extended TBox $\mathcal{T}_1$.

Finally, we point out that it would be worthwhile to develop decision procedures that can be used for efficient implementation. In (15; 16), polynomial time algorithms are developed for Σ-entailment between *acyclic* $\mathcal{EL}$-TBoxes, and it is demonstrated that these algorithms perform very well in practice. Blending these algorithms with the ones from the current paper may be an interesting start.

Acknowledgements

We thank the anonymous reviewers for their helpful comments.

References

[1] G. Antoniou and K. Kehagias. A note on the refinement of ontologies. *International Journal of Intelligent Systems*, 15:623–632, 2000.

[2] F. Baader, D. Calvanese, D. L.McGuinness, D. Nardi, P. Patel-Schneider. The Description Logic Handbook. Cambridge, 2007.

[3] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. LTCS-Report LTCS-05-01, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2005. See http://lat.inf.tu-dresden.de/research/reports.html.

[4] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI'05)*, pages 364–369. Professional Book Center, 2005.

[5] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI'99)*, pages 96–101. Morgan Kaufmann, 1999.

[6] B. Bloom and R. Paige. Transformational design and implementation of a new efficient solution to the ready simulation problem. *Sci. Comput. Program.*, 24(3):189–220, 1995.

[7] S. Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In *Proc. of the 16th European Conf. on Artificial Intelligence (ECAI-2004)*, pages 298–302. IOS Press, 2004.

[8] P. Byers and D. Pitt, Conservative extensions: a cautionary note. EATCS-Bulletin, 196–201, 1997

[9] E. Clarke and H. Schlingloff. Model checking. In *Handbook of Automated Reasoning*, volume II, chapter 24, pages 1635–1790. Elsevier Science, 2001.

[10] B. Cuenca Grau, I. Horrocks, Y. Kazakov and U. Sattler. Modular Reuse of Ontologies: Theory and Practice. *Journal of Artificial Intelligence Research*, Volume 31, pages 273-318, 2008.

[11] B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler. A logical framework for modularity of ontologies. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence (IJCAI'07)*. AAAI Press, 2007.

[12] B. Cuenca Grau, B. Parsia, E. Sirin, and A. Kalyanpur. Modularity and web ontologies. In *Proc. of the 10th Int. Conf. on Principles of Knowledge Repr. and Reasoning (KR'06)*, pages 198–209. AAAI Press, 2006.

[13] R. Diaconescu, J. Goguen and P. Stefaneas. Logical support for modularisation, In *Papers presented at the second annual Workshop on Logical Environments*, pages 83–130, Cambridge University Press, 1993.

[14] S. Ghilardi, C. Lutz, and F. Wolter. Did I damage my ontology? A case for conservative extensions in description logics. In *Proc. of the 10th Int. Conf. on Principles of Knowledge Repr. and Reasoning (KR'06)*, pages 187–197. AAAI Press, 2006.

[15] B. Konev, D. Walther and F. Wolter, The logical difference problem for description logic terminologies. In *Proceedings of the 4th International Joint Cconference on Automated Reasoning (IJCAR'08)*, Springer Lecture Notes, 2008

[16] B. Konev, C. Lutz, D. Walther, and F. Wolter. Semantic modularity and module extraction in description logic. In *Proceedings of 18th European Conference on Artificial Intelligence (ECAI'08)*, IOS Press, 2008

[17] R. Kontchakov, F. Wolter and M. Zakharyaschev. Can you tell the difference between DL-Lite ontologies. In *Proceedings of the 11th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'08)*, AAAI Press, 2008.

[18] C. Lutz, D. Walther, and F. Wolter. Conservative extensions in expressive description logics. In *Proc. of the 20th Int. Joint Conf. on Artificial Intelligence IJCAI-07*. AAAI Press, 2007.

[19] T. Maibaum. Conservative extensions, interpretations between theories and all that! In *Proceedings of 7th CAAP/FASE Conference on Theory and Practice of Software Development*. pages 40–66, 1997.

[20] N. Sioutos, S. de Coronado, M. Haber, F. Hartel, W. Shaiu, and L. Wright. NCI thesaurus: a semantic model integrating cancer-related clinical and molecular information. *Journal of Biomedical Informatics*, 40(1):30–43, 2006.

[21] V. Sofronie-Stokkermans. Interpolation in local theory extensions. In *Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJCAR'06)*, Springer Lecture Notes 4130, pages 235–250, Springer, 2006.

[22] K. Spackman. Managing clinical terminology hierarchies using algorithmic calculation of subsumption: Experience with SNOMED-RT. 2000. Fall Symposium Special Issue.

[23] L. J. Stockmeyer and A. K. Chandra. Provably difficult combinatorial games. *SIAM Journal on Computing*, 8(2):151–174, 1979.

[24] W. M. Turski and T. Maibaum. The specification of Computer Programs. Addison Wesley, 1987.

[25] P.A.S. Veloso. Yet another cautionary note on conservative extensions: a simple case with a computing flavour. EATCS-Bulletin, 188–193, 1992.

[26] P.A.S. Veloso and S.R.M. Veloso. Some remarks on conservative extensions. A socratic dialog. EATCS-Bulletin, 189–198, 1991.

A. Proofs for Section 3

Lemma 12. Let C be an $\mathcal{EL}$-concept and $\mathcal{T}$ a TBox. Then
(1) for all $E \in \Delta^{\mathcal{I}_{C,\mathcal{T}}}$, we have $E \in E^{\mathcal{I}_{C,\mathcal{T}}}$;
(2) $\mathcal{I}_{C,\mathcal{T}} \models \mathcal{T}$;
(3) $(\mathcal{I}_{C,\mathcal{T}}, D) \leq (\mathcal{I}_{C',\mathcal{T}}, D)$, for all $\mathcal{EL}$-concepts C' and all $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}}} \cap \Delta^{\mathcal{I}_{C',\mathcal{T}}}$.

Proof. (1) is straightforward by induction on the structure of E. The proof of (2) boils down to establishing the following claim.

Claim. For all $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}}}$ and all $E \in \mathsf{sub}(\mathcal{T})$: $D \in E^{\mathcal{I}_{C,\mathcal{T}}}$ iff $\mathcal{T} \models D \sqsubseteq E$.

The claim is proved by induction on the structure of E. We only consider the interesting case of the induction, i.e., $E = \exists r.F$.

"$\Rightarrow$". Let $D \in (\exists r.F)^{\mathcal{I}_{C,\mathcal{T}}}$. Then there is a $F' \in F^{\mathcal{I}_{C,\mathcal{T}}}$ with $(D, F') \in r^{\mathcal{I}_{C,\mathcal{T}}}$. We have $F' \in \Delta^{\mathcal{I}_{C,\mathcal{T}}}$ and can apply IH to F, yielding $\mathcal{T} \models F' \sqsubseteq F$. Since $(D, F') \in r^{\mathcal{I}_{C,\mathcal{T}}}$, we have $\mathcal{T} \models D \sqsubseteq \exists r.F'$, thus $\mathcal{T} \models D \sqsubseteq \exists r.F$.

"$\Leftarrow$". Let $\mathcal{T} \models D \sqsubseteq \exists r.F$. Then $(D, F) \in r^{\mathcal{I}_{C,\mathcal{T}}}$. From (1), we get $D \in (\exists r.F)^{\mathcal{I}_{C,\mathcal{T}}}$.

It is not hard to see that the claim implies (2): Let $D \sqsubseteq E \in \mathcal{T}$ and $F \in D^{\mathcal{I}_{C,\mathcal{T}}}$. By the claim, $\mathcal{T} \models F \sqsubseteq D$, and thus $\mathcal{T} \models F \sqsubseteq E$. Again by the claim, $F \in E^{\mathcal{I}_{C,\mathcal{T}}}$.

For (3), let $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}}} \cap \Delta^{\mathcal{I}_{C',\mathcal{T}}}$. Define a relation $S \subseteq \Delta^{\mathcal{I}_{C,\mathcal{T}}} \times \Delta^{\mathcal{I}_{C',\mathcal{T}}}$ by setting $S := \{(E, E) \mid E \in \Delta^{\mathcal{I}_{D,\mathcal{T}}}\}$. By construction, $(D, D) \in S$. It is easy to show that S is a simulation, hence $(\mathcal{I}_{C,\mathcal{T}}, D) \leq (\mathcal{I}_{C',\mathcal{T}}, D)$ as required. □

Lemma 13. Let C and D be $\mathcal{EL}$-concepts and $\mathcal{T}$ a TBox. Then the following holds:
(1) For all models $\mathcal{I}$ of $\mathcal{T}$ and all $d \in \Delta^{\mathcal{I}}$, the following conditions are equivalent:
  (a) $d \in C^{\mathcal{I}}$;
  (b) $(\mathcal{I}_{C,\mathcal{T}}, C) \leq (\mathcal{I}, d)$;
  (c) $(\mathcal{I}_{C,\mathcal{T}}, C) \leq^{\mathsf{full}} (\mathcal{I}, d)$.
(2) The following conditions are equivalent:
  (a) $\mathcal{T} \models C \sqsubseteq D$;
  (b) $C \in D^{\mathcal{I}_{C,\mathcal{T}}}$;
  (c) $(\mathcal{I}_{D,\mathcal{T}}, D) \leq (\mathcal{I}_{C,\mathcal{T}}, C)$.
(3) The following conditions are equivalent:
  (a) $\mathcal{T} \models C \sqsubseteq \exists u.D$;
  (b) $C \in (\exists u.D)^{\mathcal{I}_{C,\mathcal{T}}}$.

Proof. (1) (c) $\Rightarrow$ (b) is trivial and (b) $\Rightarrow$ (a) follows from Theorem 10 since $C \in C^{\mathcal{I}_{C,\mathcal{T}}}$. For (a) $\Rightarrow$ (c), let $\mathcal{I}$ be a model of $\mathcal{T}$ and $d \in C^{\mathcal{I}}$. Define a relation $S \subseteq \Delta^{\mathcal{I}_{C,\mathcal{T}}} \times \Delta^{\mathcal{I}}$ by setting $(D, e) \in S$ iff $e \in D^{\mathcal{I}}$. We show that S is a full simulation. Let $(D, e) \in S$. Assume $D \in A^{\mathcal{I}_{C,\mathcal{T}}}$, with A a concept name. This implies $\mathcal{T} \models D \sqsubseteq A$, and $e \in A^{\mathcal{I}}$ follows from $e \in D^{\mathcal{I}}$ and $\mathcal{I} \models \mathcal{T}$. Now assume $(D, D') \in r^{\mathcal{I}_{C,\mathcal{T}}}$. Then $\mathcal{T} \models D \sqsubseteq \exists r.D'$ and we obtain $e \in (\exists r.D')^{\mathcal{I}}$. Hence, there exists $e' \in \Delta^{\mathcal{I}}$ with $(e, e') \in r^{\mathcal{I}}$ and $e' \in D'^{\mathcal{I}}$, which implies $(D', e') \in S$. It follows that S is a simulation. By definition, we have $(C, d) \in S$. S is full because $D \in \Delta^{\mathcal{I}_{C,\mathcal{T}}}$ implies $\mathcal{T} \models C \sqsubseteq \exists u.D$. Hence there exists $e \in \Delta^{\mathcal{I}}$ with $e \in D^{\mathcal{I}}$ and this implies $(D, e) \in S$.

(2) (a) $\Rightarrow$ (b). Assume $\mathcal{T} \models C \sqsubseteq D$. Since $\mathcal{I}_{C,\mathcal{T}}$ is a model of $\mathcal{T}$ and $C \in C^{\mathcal{I}_{C,\mathcal{T}}}$, this implies $C \in D^{\mathcal{I}_{C,\mathcal{T}}}$. (b) $\Rightarrow$ (c) is an immediate consequence of (1). For (c) $\Rightarrow$ (a), let $\mathcal{I}$ be a model of $\mathcal{T}$ and $d \in C^{\mathcal{I}}$. By (1), $(\mathcal{I}_{C,\mathcal{T}}, C) \leq (\mathcal{I}, d)$. Together with $(\mathcal{I}_{D,\mathcal{T}}, D) \leq (\mathcal{I}_{C,\mathcal{T}}, C)$ and transitivity of "$\leq$", we get $(\mathcal{I}_{D,\mathcal{T}}, D) \leq (\mathcal{I}, d)$. Again by (1), we obtain $d \in D^{\mathcal{I}}$.

(3) (a) $\Rightarrow$ (b) follows from $\mathcal{I}_{C,\mathcal{T}} \models \mathcal{T}$ and $C \in C^{\mathcal{I}_{C,\mathcal{T}}}$. Conversely, let $C \in (\exists u.D)^{\mathcal{I}_{C,\mathcal{T}}}$. Then there is an $E \in D^{\mathcal{I}_{C,\mathcal{T}}}$. By (2), this yields $\mathcal{T} \models E \sqsubseteq D$. Since $E \in \Delta^{C,\mathcal{T}}$, we have $\mathcal{T} \models C \sqsubseteq \exists u.E$. Thus, $\mathcal{T} \models C \sqsubseteq \exists u.D$. $\square$

Lemma 15. Let $\mathcal{T}_1, \mathcal{T}_2$ be TBoxes and C_1, C_2 $\mathcal{EL}$-concepts and Σ a signature. Then

- $(\mathcal{T}_1, C_1) \sqsubseteq_\Sigma (\mathcal{T}_2, C_2)$ iff $(\mathcal{I}_{C_2,\mathcal{T}_2}, C_2) \leq_\Sigma (\mathcal{I}_{C_1,\mathcal{T}_1}, C_1)$;
- $(\mathcal{T}_1, C_1) \sqsubseteq^u_\Sigma (\mathcal{T}_2, C_2)$ iff $(\mathcal{I}_{C_2,\mathcal{T}_2}, C_2) \leq^{\mathsf{full}}_\Sigma (\mathcal{I}_{C_1,\mathcal{T}_1}, C_1)$.

Proof. We only prove the first equivalence since the proof of the second is similar (using Point 3 of Lemma 13 instead of Point 2) .

"$\Rightarrow$". Assume $(\mathcal{T}_1, C_1) \not\sqsubseteq_\Sigma (\mathcal{T}_2, C_2)$. Then there is an $\mathcal{EL}_\Sigma$-concept E such that $\mathcal{T}_2 \models C_2 \sqsubseteq E$ and $\mathcal{T}_1 \not\models C_1 \sqsubseteq E$. By Point 2 of Lemma 13, this yields $C_2 \in E^{\mathcal{I}_{C_2,\mathcal{T}_2}}$ and $C_1 \notin E^{\mathcal{I}_{C_1,\mathcal{T}_1}}$. Hence, by Theorem 10, $(\mathcal{I}_{C_2,\mathcal{T}_2}, C_2) \not\leq_\Sigma (\mathcal{I}_{C_1,\mathcal{T}_1}, C_1)$.

"$\Leftarrow$". Let $(\mathcal{I}_{C_2,\mathcal{T}_2}, C_2) \not\leq_\Sigma (\mathcal{I}_{C_1,\mathcal{T}_1}, C_1)$. By Theorem 10, there exists an $\mathcal{EL}_\Sigma$-concept E with $C_2 \in E^{\mathcal{I}_{C_2,\mathcal{T}_2}}$ but $C \notin E^{\mathcal{I}_{C_1,\mathcal{T}_1}}$. By Point 2 of Lemma 13, $\mathcal{T}_2 \models C_2 \sqsubseteq E$ and $\mathcal{T}_1 \not\models C_1 \sqsubseteq E$. $\square$

B. Disjunction Property and Robustness

In this section, we prove that $(\mathcal{EL}, \mathcal{QL}^u_{\mathcal{EL}})$ is robust under signature extensions and has the join modularity property. Note that this section comes after the section on proofs for Section 3 because we employ the canonical model construction and its properties. First, we show two auxiliary lemmas which will be useful in subsequent sections as well.

Lemma 43 (Disjunction Property)**.** *Let $\mathcal{T}$ be a TBox and let*

$$C = C_0 \sqcap \bigsqcap_{i \in I} \exists u.C_i, \quad D = \bigsqcup_{i \in J} D_i \sqcup \bigsqcup_{i \in K} \exists u.D_i,$$

where C_0, C_i, $i \in I$, D_i, $i \in J$, and D_i, $i \in K$, are $\mathcal{EL}$-concepts. If $\mathcal{T} \models C \sqsubseteq D$, then $\mathcal{T} \models C' \sqsubseteq D'$ for a $C' \in \{C_0\} \cup \{\exists u.C_i \mid i \in I\}$ and $D' \in \{D_i \mid i \in J\} \cup \{\exists u.D_i \mid i \in K\}$.

Proof. We first show this property for C a $\mathcal{EL}$-concept. Thus, assume that C is a $\mathcal{EL}$-concept, D as defined in the lemma, and $\mathcal{T} \models C \sqsubseteq D$. Take the canonical model $\mathcal{I}_{C,\mathcal{T}}$ of $\mathcal{T}$. By Lemma 12, $C \in C^{\mathcal{I}_{C,\mathcal{T}}}$. Since $\mathcal{T} \models C \sqsubseteq D$, we have $C \in E^{\mathcal{I}_{C,\mathcal{T}}}$ for some disjunct $E \in \{D_i \mid i \in J\} \cup \{\exists u.D_i \mid i \in K\}$. By Lemma 13, this implies $\mathcal{T} \models C \sqsubseteq E$, as required.

Next we show that, in general, from $\mathcal{T} \models C \sqsubseteq D$ follows $\mathcal{T} \models C_0 \sqsubseteq D$ or $\mathcal{T} \models \exists u.C_i \sqsubseteq D$, for some $i \in I$. Assume this is not the case. Take, for $i \in I \cup \{0\}$, a model $\mathcal{I}_i$ of $\mathcal{T}$ such that $x_i \in C_i^{\mathcal{I}_i} \setminus D^{\mathcal{I}_i}$. Take the disjoint union $\mathcal{I}$ of the models $\mathcal{I}_i$, $i \in I \cup \{0\}$. Then $\mathcal{I}$ is a model of $\mathcal{T}$ and $x_0 \in C^{\mathcal{I}} \setminus D^{\mathcal{I}}$. Hence $\mathcal{T} \not\models C \sqsubseteq D$ and we have derived a contradiction.

To prove the lemma it remains to consider the case $\mathcal{T} \models \exists u.C_i \sqsubseteq D$, for some $i \in I$. Fix an $i \in I$ with this property. Using a construction similar to the disjoint union construction above, it is not difficult to see that then $\mathcal{T} \models \top \sqsubseteq D_i$, for some $i \in J$, or $\mathcal{T} \models C_i \sqsubseteq \bigsqcup_{i \in K} \exists u.D_i$. In the first case, $\mathcal{T} \models C_0 \sqsubseteq D_i$, as required. In the second case, by the disjunction property for C a $\mathcal{EL}$-concept proved above, we obtain $\mathcal{T} \models C_i \sqsubseteq \exists u.D_j$, for some $j \in K$. But then $\mathcal{T} \models \exists u.C_i \sqsubseteq \exists u.D_j$, as required. □

Say that an $\mathcal{EL}^u$-concept C follows from a TBox $\mathcal{T}$ and a (possibly infinite) set Ψ of $\mathcal{EL}^u$-concepts, written $\mathcal{T} \cup \Psi \models C$, if for every model $\mathcal{I}$ of $\mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$: if $d \in D^{\mathcal{I}}$ for all $D \in \Psi$, then $d \in C^{\mathcal{I}}$. Observe that $\mathcal{T} \models C \sqsubseteq D$ if, and only if, $\mathcal{T} \cup \{C\} \models D$ and that this consequence is compact in the sense that $\mathcal{T} \cup \Psi \models D$ implies that there exists a finite subset Ψ' of Ψ such that $\mathcal{T} \cup \Psi' \models D$.

Lemma 44. *Let $\mathcal{T}$ be a TBox and Ψ a set of $\mathcal{EL}^u$-concepts. Then there exists a model $\mathcal{I}$ of $\mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$ such that, for all $\mathcal{EL}^u$-concepts C: $d \in C^{\mathcal{I}}$ iff $\mathcal{T} \cup \Psi \models C$.*

Proof. Follows immediately from Lemma 43 and compactness: suppose no such model exists. Let $\Gamma = \{C \mid \mathcal{T} \cup \Psi \models C,\ C \text{ a } \mathcal{EL}^u\text{-concept}\}$ and denote by $\overline{\Gamma}$ the set of all $\mathcal{EL}^u$-concepts not in Γ. Then there does not exist a model $\mathcal{I}$ of $\mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$ such that $d \in C^{\mathcal{I}}$, for all $C \in \Gamma$, and $d \in (\neg C)^{\mathcal{I}}$, for all $C \in \overline{\Gamma}$. By compactness, there exist finite subsets Γ_0 of Γ and $\overline{\Gamma}_0$ of $\overline{\Gamma}$ such that no such model exists for Γ_0 and $\overline{\Gamma}_0$. Hence

$$\mathcal{T} \models \bigsqcap_{C \in \Gamma_0} C \sqsubseteq \bigsqcup_{D \in \overline{\Gamma}_0} D.$$

By Lemma 43, $\mathcal{T} \models \bigsqcap_{C \in \Gamma_0} C \sqsubseteq D$, for some $D \in \overline{\Gamma}_0$. But then $\mathcal{T} \cup \Psi \models D$ and we have derived a contradiction. □

Theorem 45. *$(\mathcal{EL}, \mathcal{QL}^u_{\mathcal{EL}})$ is robust under signature extensions.*

Proof. The proof employs the Craig interpolation property of $\mathcal{EL}$. Suppose $\mathcal{T}_1 \sqsubseteq^u_\Sigma \mathcal{T}_2$ and $\Sigma' \supseteq \Sigma$ with $\Sigma' \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$. Let $\mathcal{T}_2 \models C \sqsubseteq D$ with $\mathsf{sig}(C \sqsubseteq D) \subseteq \Sigma'$. If D is an $\mathcal{EL}$-concept, then $\mathcal{T}_1 \models C \sqsubseteq D$ by robustness under signature extensions of $(\mathcal{EL}, \mathcal{QL}_{\mathcal{EL}})$. Assume now $D = \exists u.D'$ and $\mathcal{T}_2 \not\models C \sqsubseteq D'$. By Lemma 13 (see Lemma 16 for a similar observation), there are two cases:

(1) there exists $\exists r.C' \in \mathsf{sub}(C)$ such that $\mathcal{T}_2 \models C' \sqsubseteq D'$.

(2) there exists $\exists r.C' \in \mathsf{sub}(\mathcal{T}_2)$ such that $\mathcal{T}_2 \models C \sqsubseteq \exists u.C'$ and $\mathcal{T}_2 \models C' \sqsubseteq D'$.

If Point 1 applies, then, by robustness under vocabulary extensions of $(\mathcal{EL}, \mathcal{QL}_{\mathcal{EL}})$, $\mathcal{T}_1 \models C' \sqsubseteq D'$ and, therefore, $\mathcal{T}_1 \models \exists u.C' \sqsubseteq \exists u.D'$. Also, $\models C \sqsubseteq \exists u.C'$. So we obtain $\mathcal{T}_1 \models C \sqsubseteq \exists u.D'$.

Now assume Point 2 applies to $\exists r.C' \in \mathsf{sub}(\mathcal{T}_2)$. Replace, in C, all role names $r \in \Sigma' \setminus \Sigma$ by u, and all concept names $A \in \Sigma' \setminus \Sigma$ by $\top$, and denote the resulting concept by C^*. We have $\models C \sqsubseteq C^*$. It follows immediately from $\mathsf{sig}(C) \cap (\mathsf{sig}(\mathcal{T}_2) \cup \mathsf{sig}(C')) \subseteq \Sigma$ that $\mathcal{T}_2 \models C^* \sqsubseteq \exists u.C'$. Then there exists a subconcept C^*_0 of C^* in $\mathcal{EL}$ such that $\mathcal{T}_2 \models C^*_0 \sqsubseteq \exists u.C'$. Moreover, $\models C \sqsubseteq \exists u.C^*_0$. On the other hand, from $\mathcal{T}_2 \models C' \sqsubseteq D'$, we obtain $\mathsf{sig}(D') \subseteq \Sigma$ because $\Sigma' \cap (\mathsf{sig}(\mathcal{T}_2) \cup \mathsf{sig}(C')) \subseteq \Sigma$. Thus $\mathcal{T}_2 \models C^*_0 \sqsubseteq \exists u.D'$ and $\mathsf{sig}(C^*_0 \sqsubseteq \exists u.D') \subseteq \Sigma$. Hence $\mathcal{T}_1 \models C^*_0 \sqsubseteq \exists u.D'$ and from $\models C \sqsubseteq \exists u.C^*_0$ we obtain $\mathcal{T}_1 \models C \sqsubseteq \exists u.D'$. □

Theorem 46. *$(\mathcal{EL}, \mathcal{QL}_{\mathcal{EL}})$ and $(\mathcal{EL}, \mathcal{QL}^u_{\mathcal{EL}})$ have the join modularity property.*

Proof. We give a proof for $(\mathcal{EL}, \mathcal{QL}^u_{\mathcal{EL}})$; the proof for $(\mathcal{EL}, \mathcal{QL}_{\mathcal{EL}})$ is a minor modification of this proof and left to the reader.

Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be Σ-inseparable w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$, where Σ is a signature with $\mathsf{sig}(\mathcal{T}_1) \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$. Assume that $\mathcal{T}_i \not\models C \sqsubseteq D$, for some $C \sqsubseteq D \in \mathcal{QL}^u_{\mathcal{EL}}$ with $\mathsf{sig}(C \sqsubseteq D) \subseteq \Sigma$. We show $\mathcal{T}_1 \cup \mathcal{T}_2 \not\models C \sqsubseteq D$. Take the canonical model $\mathcal{I}_0 = \mathcal{I}_{C,\mathcal{T}_1}$ and let $d_0 = C \in \Delta^{\mathcal{I}_0}$. Then $d_0 \in C^{\mathcal{I}_0} \setminus D^{\mathcal{I}_0}$. Set $\Delta_0 = \Delta_{d_0} = \Delta^{\mathcal{I}_0}$. In the following, we construct an interpretation $\mathcal{I}^*$ of $\mathcal{T}_1 \cup \mathcal{T}_2$ refuting $C \sqsubseteq D$. We define inductively an infinite sequence $\mathcal{I}_1, \mathcal{I}_2, \dots$ of interpretations. The interpretation $\mathcal{I}^* = (\Delta^{\mathcal{I}^*}, \cdot^{\mathcal{I}^*})$ is then defined as the union of $\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_2, \dots$ as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}^*} &:= \bigcup_{i \geq 0} \Delta^{\mathcal{I}_i};\\
A^{\mathcal{I}^*} &:= \bigcup_{i \geq 0} A^{\mathcal{I}_i}, \text{ for all } A \in \mathsf{N_C};\\
r^{\mathcal{I}^*} &:= \bigcup_{i \geq 0} r^{\mathcal{I}_i}, \text{ for all } r \in \mathsf{N_R}.
\end{aligned}$$

Given an intepretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, recall that $d^{\Sigma,\mathcal{I},u}$ denotes the set of $\mathcal{EL}^u_\Sigma$-concepts E with $d \in E^{\mathcal{I}}$. For any Tbox $\mathcal{T}$ denote by $\mathcal{I}_{t_{\mathcal{I}}(d),\mathcal{T}}$ a model of $\mathcal{T}$ with d in its domain such that

$(*)$ $d \in E^{\mathcal{I}_{t_{\mathcal{I}}(d),\mathcal{T}}}$ iff $\mathcal{T} \cup d^{\Sigma,\mathcal{I},u} \models E$, for all $\mathcal{EL}^u$-concepts E.

By Lemma 44, such an interpretation always exists. Moreover, we may assume that d is not within the range of any $r^{\mathcal{I}_{t_{\mathcal{I}}(d),\mathcal{T}}}$ (if it is, one can use standard unravelling (see Section 5.2) to obtain a model with the required properties). Let $n \geq 0$ and assume the interpretation $\mathcal{I}_n$ with domain Δ_n has been defined. If n is even, then take for every $d \in \Delta_n \setminus \Delta_{n-1}$ (we set $\Delta_{-1} = \emptyset$) the interpretation $\mathcal{I}_d = \mathcal{I}_{t_{\mathcal{I}_n}(d),\mathcal{T}_2}$ with domain Δ_d such that $\Delta_n \cap \Delta_d = \{d\}$ and the Δ_d, $d \in \Delta_n \setminus \Delta_{n-1}$, are mutually disjoint. If n is odd, then take for every $d \in \Delta_n \setminus \Delta_{n-1}$ the interpretation $\mathcal{I}_d = \mathcal{I}_{t_{\mathcal{I}_n}(d),\mathcal{T}_1}$ with domain Δ_d such that $\Delta_n \cap \Delta_d = \{d\}$ and the Δ_d, $d \in \Delta_n \setminus \Delta_{n-1}$, are mutually disjoint. Now set

$$\begin{aligned}
\Delta_{n+1} &= \Delta_n \cup \textstyle\bigcup_{d \in \Delta_n \setminus \Delta_{n-1}} \Delta_d,\\
r^{\mathcal{I}_{n+1}} &= r^{\mathcal{I}_n} \cup \textstyle\bigcup_{d \in \Delta_n \setminus \Delta_{n-1}} r^{\mathcal{I}_d},\\
A^{\mathcal{I}_{n+1}} &= A^{\mathcal{I}_n} \cup \textstyle\bigcup_{d \in \Delta_n \setminus \Delta_{n-1}} A^{\mathcal{I}_d}.
\end{aligned}$$

For all $d \in \Delta^{\mathcal{I}^*}$ there exists a (uniquely) determined minimal natural number $n(d)$ with $d \in \Delta_{n(d)} \setminus \Delta_{n(d)-1}$. If $n(d) \neq 0$, then there exists a uniquely determined $d^* \in \Delta_{n(d)-1}$ with $d \in \Delta_{d^*}$. We set $d^* = d_0$ for $n(d) = 0$ and prove the following by induction on the construction of D. For all $d \in \Delta^{\mathcal{I}^*}$ and $\mathcal{EL}$-concepts D:

- if $n(d)$ is even then

(1) if $\mathsf{sig}(D) \cap \mathsf{sig}(\mathcal{T}_1) \subseteq \Sigma$, then $d \in D^{\mathcal{I}^*} \Leftrightarrow d \in D^{\mathcal{I}_d}$;

(2) if $\mathsf{sig}(D) \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$, then $d \in D^{\mathcal{I}^*} \Leftrightarrow d \in D^{\mathcal{I}_{d^*}}$;

- if $n(d)$ is odd then

(1) if $\mathsf{sig}(D) \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$, then $d \in D^{\mathcal{I}^*} \Leftrightarrow d \in D^{\mathcal{I}_d}$;

(2) if $\mathsf{sig}(D) \cap \mathsf{sig}(\mathcal{T}_1) \subseteq \Sigma$, then $d \in D^{\mathcal{I}^*} \Leftrightarrow d \in D^{\mathcal{I}_{d^*}}$.
The implications from right to left are trivial, so we consider the implications from left to right only. We concentrate on the case $n(d)$ even (the case $n(d)$ odd is proved in the same way) and prove the induction step for $D = \exists r.C$. First consider Point 1. So let $\mathsf{sig}(D) \cap \mathsf{sig}(\mathcal{T}_1) \subseteq \Sigma$ and assume $d \in D^{\mathcal{I}^*}$ with $n(d)$ even. There exists $c \in \Delta^{\mathcal{I}^*}$ such that $c \in C^{\mathcal{I}^*}$ and $(d, c) \in r^{\mathcal{I}^*}$. Assume first that $c \in \Delta_{n(d)}$. Then, by construction, $c \notin \Delta_{n(d)-1}$. Then $r \in \Sigma$ because for any $r \notin \mathsf{sig}(\mathcal{T}_1)$, $r^{\mathcal{I}^*} \cap (\Delta_{n(d)} \setminus \Delta_{n(d)-1})^2 = \emptyset$. We obtain $n(c) = n(d)$ and, by IH, $c \in C^{\mathcal{I}_c}$. We obtain $\mathcal{T}_2 \cup c^{\Sigma,\mathcal{I}_{n(d)},u} \models C$. By compactness and closure under conjunction of $c^{\Sigma,\mathcal{I}_{n(d)},u}$, there exists a concept C_0 in $c^{\Sigma,\mathcal{I}_{n(d)},u}$ with $\mathcal{T}_2 \models C_0 \sqsubseteq C$. Then $\mathcal{T}_2 \models \exists r.C_0 \sqsubseteq \exists r.C$. We have $\exists r.C_0 \in d^{\Sigma,\mathcal{I}_{n(d)},u}$ and so $\mathcal{T}_2 \cup d^{\Sigma,\mathcal{I}_{n(d)},u} \models \exists r.C$. But then $d \in D^{\mathcal{I}_d}$.

Now assume $c \notin \Delta_{n(d)}$. Then $c \in \Delta_d$, $c^* = d$, and $n(c) = n(d) + 1$. By induction hypothesis (for $n(c)$ odd), $c \in C^{\mathcal{I}^*}$ iff $c \in C^{\mathcal{I}_{c^*}} = C^{\mathcal{I}_d}$. Hence $d \in (\exists r.C)^{\mathcal{I}_d}$.

Consider now Point 2. Let $\mathsf{sig}(D) \cap \mathsf{sig}(\mathcal{T}_2) \subseteq \Sigma$ and $d \in D^{\mathcal{I}^*}$. There exists $c \in \Delta^{\mathcal{I}^*}$ such that $c \in C^{\mathcal{I}^*}$ and $(d, c) \in r^{\mathcal{I}^*}$. Assume first that $c \in \Delta_{d^*}$. Then $c^* = d^*$ and, by induction hypothesis, $c \in C^{\mathcal{I}_{d^*}}$. As we also have $(d, c) \in r^{\mathcal{I}_{d^*}}$, we obtain $d \in D^{\mathcal{I}_{d^*}}$.

Now assume $c \notin \Delta_{d^*}$. Then $c \in \Delta_d$. Then $r \in \Sigma$ because for any $r \notin \mathsf{sig}(\mathcal{T}_2)$, $r^{\mathcal{I}^*} \cap \Delta_d \times \Delta_d = \emptyset$. By induction hypothesis $c \in C^{\mathcal{I}_c}$. Hence $\mathcal{T}_1 \cup c^{\Sigma,\mathcal{I}_{n(d)+1},u} \models C$. By compactness and closure under conjunction of $c^{\Sigma,\mathcal{I}_{n(d)+1},u}$, there exists a concept C_0 in $c^{\Sigma,\mathcal{I}_{n(d)+1},u}$ with $\mathcal{T}_1 \models C_0 \sqsubseteq C$. Then $\mathcal{T}_1 \models \exists r.C_0 \sqsubseteq \exists r.C$. We have $d \in (\exists r.C_0)^{\mathcal{I}_d}$. Since $\mathsf{sig}(\exists r.C_0) \subseteq \Sigma$ it follows from Σ-inseparability w.r.t. $\mathcal{QL}^u_{\mathcal{EL}}$ of $\mathcal{T}_1$ and $\mathcal{T}_2$ and compactness that $\exists r.C_0 \in d^{\Sigma,\mathcal{I}_{n(d)},u}$. So $d \in (\exists r.C_0)^{\mathcal{I}_{d^*}}$. $\mathcal{I}_{d^*}$ is a model of $\mathcal{T}_1$. Hence $d \in (\exists r.C)^{\mathcal{I}_{d^*}}$.

It follows immediately that $\mathcal{I}^*$ is a model of $\mathcal{T}_1 \cup \mathcal{T}_2$: let $C_0 \sqsubseteq D_0 \in \mathcal{T}_i$. If $C_0^{\mathcal{I}^*} \setminus D_0^{\mathcal{I}^*} \neq \emptyset$, then there exists a an interpretation $\mathcal{I}_d$ of $\mathcal{T}_i$ with $C_0^{\mathcal{I}_d} \setminus D_0^{\mathcal{I}_d} \neq \emptyset$ which is a contradiction.

It remains to show that $d_0 \in C^{\mathcal{I}^*} \setminus D^{\mathcal{I}^*}$. $d_0 \in C^{\mathcal{I}^*}$ by the claim above and since $d_0 \in C^{\mathcal{I}_0}$. If D is a $\mathcal{EL}$-concept, then $d_0 \notin D^{\mathcal{I}^*}$ follows from $d_0 \notin D^{\mathcal{I}_0}$ and the claim above. Now suppose $D = \exists u.D_0$. Using the claim above it is readily proved by induction on n that $\Delta_n \cap D_0^{\mathcal{I}^*} = \emptyset$, for all $n \geq 0$. Hence $D^{\mathcal{I}^*} = \emptyset$, as required. ☐

C. Proofs for Section 4

Let P and Q be finite sets of pairs (r, E), where r is a role and E a $\mathcal{EL}$-concept. We say that Q *covers* P *w.r.t. a TBox* $\mathcal{T}$, in symbols $P \leq_{\mathcal{T}} Q$, if for all $\exists r.G \in \mathsf{sub}(\mathcal{T})$ and $(r, E) \in P$ with $\mathcal{T} \models E \sqsubseteq G$ there exists $(r, E') \in Q$ with $\mathcal{T} \models E' \sqsubseteq G$.

Lemma 47. *Let $\mathcal{T}$ be a TBox, $C_0 = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, and $C_1 = F_0 \sqcap \bigsqcap_{(r,E)\in Q} \exists r.E$ and assume $P \leq_{\mathcal{T}} Q$. Then the following holds:*
- *$K_{\mathcal{T}}(C_0) \subseteq K_{\mathcal{T}}(C_1)$.*
- *If C' is a $\mathcal{EL}$-concept with $\exists r.C_0 \in \mathsf{sub}(C')$ and C'' the resulting concept when $\exists r.C_0$ is replaced by $\exists r.C_1$ in C', then $K_{\mathcal{T}}(C') \subseteq K_{\mathcal{T}}(C'')$.*

Proof. We show Point 1. Point 2 can be proved by induction or directly using a similar construction and is left to the reader.

Let $H \in \mathsf{sub}(\mathcal{T}) \setminus K_{\mathcal{T}}(C_1)$. We have to show that $H \notin K_{\mathcal{T}}(C_0)$. There is a model $\mathcal{I}$ of $\mathcal{T}$ with $d_0 \in C_1^{\mathcal{I}} \setminus H^{\mathcal{I}}$. For each $(r, E) \in P$, take a copy $\mathcal{I}_{r,E}$ of the canonical model $\mathcal{I}_{E,\mathcal{T}}$ such that all these copies have disjoint domains, and their domains are disjoint from that of $\mathcal{I}$. In the copy $\mathcal{I}_{r,E}$, the point corresponding to E in the canonical model $\mathcal{I}_{E,\mathcal{T}}$ is denoted by $d_{r,E}$. Define a new interpretation $\mathcal{I}'$ as follows:

- take the union of $\mathcal{I}$ and the models $\mathcal{I}_{r,E}$, for all $(r,E) \in P$;
- for each $(r,E) \in P$ add the tuple $(d_0, d_{r,E})$ to $r^{\mathcal{I}'}$.

Observe that $d_0 \in C_0^{\mathcal{I}'}$. The following claims can be proved by induction on the structure of the $\mathcal{EL}$-concept D_0:

(a) for all $(r,E) \in P$, all $d \in \Delta^{\mathcal{I}_{r,E}}$, and all $\mathcal{EL}$-concepts D_0, $d \in D_0^{\mathcal{I}'}$ iff $d \in D_0^{\mathcal{I}_{r,E}}$.

(b) for all $d \in \Delta^{\mathcal{I}}$ and $D_0 \in \mathsf{sub}(\mathcal{T})$, $d \in D_0^{\mathcal{I}}$ iff $d \in D_0^{\mathcal{I}'}$.

The only interesting case is the direction from right to left in (b), when $D_0 = \exists r.D_0'$. Let $d \in (\exists r.D_0')^{\mathcal{I}'}$. Then there is a $d' \in {D'}_0^{\mathcal{I}'}$ such that $(d,d') \in r^{\mathcal{I}'}$. If $d' \in \Delta^{\mathcal{I}}$, we have $(d,d') \in r^{\mathcal{I}}$ and it remains to apply IH. Now let $d' \in \Delta^{\mathcal{I}_{r,E}}$ for some $(r,E) \in P$. Then $d = d_0$ and $d' = d_{r,E}$. By (a) above, $d_{r,E} \in {D'}_0^{\mathcal{I}'}$ implies $d_{r,E} \in {D'}_0^{\mathcal{I}_{r,E}}$. With Point 2 of Lemma 13, we get $\mathcal{T} \models E \sqsubseteq D_0'$. Since $P \leq_{\mathcal{T}} Q$, there is an $(r,E') \in Q$ such that $\mathcal{T} \models E' \sqsubseteq D_0'$. We have $d = d_0 \in C_1^{\mathcal{I}}$ and, therefore, there is a $d'' \in (E')^{\mathcal{I}}$ with $(d,d'') \in r^{\mathcal{I}}$. Hence, $d \in (\exists r.{D'}_0)^{\mathcal{I}}$.

Since $\mathcal{I}$ and all the $\mathcal{I}_{r,E}$ are models of $\mathcal{T}$ and by (a) and (b) above, it follows that $\mathcal{I}'$ is a model of $\mathcal{T}$. (b) implies $d_0 \in C_0^{\mathcal{I}'} \setminus H^{\mathcal{I}'}$ and we derive $H \notin K_{\mathcal{T}}(C_0)$. □

Lemma 48. *Assume $\Sigma \subseteq \mathsf{sig}(\mathcal{T}_2)$. Suppose there exists an $\mathcal{EL}_\Sigma$-concept C and a concept $D \in \mathsf{sub}(\mathcal{T}_2)$ such that*

(a) $\mathcal{T}_2 \models C \sqsubseteq D$;

(b) $(\mathcal{T}_1, C) \not\sqsubseteq_\Sigma (\mathcal{T}_2, D)$.

Then there exist C and D with properties (a) and (b) such that

(c) the outdegree of C is bounded by $|\mathcal{T}_2|$.

Proof. Let C be an $\mathcal{EL}_\Sigma$-concept and $D \in \mathsf{sub}(\mathcal{T}_2)$ such that Points (a) and (b) hold. If the outdegree of C is bounded by $|\mathcal{T}_2|$, C itself is as required. Assume that this is not the case. Then there exists a subconcept C_0 of C such that $C_0 = F \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, where F is a conjunction of concept names and $|P| > |\mathcal{T}_2|$. Let Q be a minimal subset of P such that $P \leq_{\mathcal{T}_2} Q$. Clearly, the cardinality of Q is bounded by $|\mathcal{T}_2|$. Now, replace in C the subconcept C_0 with $C_1 := F \sqcap \bigsqcap_{(r,E)\in Q} \exists r.E$ and call the result C'. We have $|C'| \leq |C|$ and, by Lemma 47, $K_{\mathcal{T}_2}(C) = K_{\mathcal{T}_2}(C')$. To obtain the desired concept C', we now execute the described contraction until the outdegree is bounded by $|\mathcal{T}_2|$. The resulting concept C' satisfies (a) because $K_{\mathcal{T}_2}(C) = K_{\mathcal{T}_2}(C')$. (b) holds for C' because $\emptyset \models C \sqsubseteq C'$. □

Lemma 18 Let $\mathcal{T}$ be a TBox and $C = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$. Then

$$K_{\mathcal{T}}(C) = K_{\mathcal{T}}(F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.(\bigsqcap_{D \in K_{\mathcal{T}}(E)} D)).$$

Proof. The condition of Point 1 of Lemma 47 is satisfied for $C_0 = C$ and $C_1 = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.(\bigsqcap_{D\in K_{\mathcal{T}}(E)} D)$ and vice versa. □

D. Proofs for Section 5

Lemma 22 Let $\mathcal{T}$ be a TBox, Σ a signature, and $\mathcal{A}$ an $\mathcal{EL}_\Sigma$-ABox.

(1) For every $\mathcal{EL}_\Sigma$-concept D and $a \in \mathsf{N_I}$, $(\mathcal{T}, \mathcal{A}) \models D(a)$ iff there exists an $\mathcal{EL}_\Sigma$-concept C such that $\mathcal{T} \models C \sqsubseteq D$ and $\mathcal{A} \models C(a)$.

(2) For every $\mathcal{EL}^u_\Sigma$-concept D and $a \in \mathsf{N_I}$, $(\mathcal{T}, \mathcal{A}) \models D(a)$ iff there exists an $\mathcal{EL}^u_\Sigma$-concept C such that $\mathcal{T} \models C \sqsubseteq D$ and $\mathcal{A} \models C(a)$.

Proof. The directions from right to left are trivial, so we concentrate on the other direction.

Point 1. Let D_0 be a $\mathcal{EL}_\Sigma$-concept, $a_0 \in \mathsf{N_I}$, and assume that $(\mathcal{T}, \mathcal{A}) \models D_0(a_0)$. Set, for every $a \in \mathsf{ob} = \mathsf{Obj}(\mathcal{A}) \cup \{a_0\}$,

$$t_\mathcal{A}(a) = \{C \mid \mathcal{A} \models C(a),\ C \text{ an } \mathcal{EL}_\Sigma\text{-concept}\}.$$

We show that $\mathcal{T} \cup t_\mathcal{A}(a_0) \models D_0$. Then, using compactness, we find a $\mathcal{EL}_\Sigma$-concept C such that $\mathcal{T} \models C \sqsubseteq D_0$ and $\mathcal{A} \models C(a_0)$, as required. Assume $\mathcal{T} \cup t_\mathcal{A}(a_0) \not\models D_0$. Take, for every $a \in \mathsf{ob}$, a model $\mathcal{I}_a$ of $\mathcal{T}$ with a point d_a such that for all $\mathcal{EL}$-concepts C: $d^{\mathcal{I}_a} \in C^{\mathcal{I}_a}$ iff $\mathcal{T} \cup t_\mathcal{A}(a) \models C$. Such models exist by Lemma 44. We may assume that they are mutually disjoint. Take the following union $\mathcal{I}$ of the models $\mathcal{I}_a$:

- $\Delta^\mathcal{I} = \bigcup_{a \in \mathsf{ob}} \Delta^{\mathcal{I}_a}$;
- $A^\mathcal{I} = \bigcup_{a \in \mathsf{ob}} A^{\mathcal{I}_a}$, for $A \in \mathsf{N_C}$;
- $r^\mathcal{I} = \bigcup_{a \in \mathsf{ob}} r^{\mathcal{I}_a} \cup \{(d_a, d_b) \mid r(a, b) \in \mathcal{A}\}$, for $r \in \mathsf{N_R}$;
- $a^\mathcal{I} = d_a$, for $a \in \mathsf{ob}$.

For all $\mathcal{EL}$-concepts C and all $a \in \mathsf{ob}$ the following holds for all $d \in \Delta^{\mathcal{I}_a}$:

$$d \in C^{\mathcal{I}_a} \text{ iff } d \in C^\mathcal{I}.$$

The proof is by induction on the construction of C. The only interesting case is $C = \exists r.D$ and the direction from right to left. Assume $d \in C^\mathcal{I} \cap \Delta^{\mathcal{I}_a}$. For $d \neq d_a$, $d \in C^{\mathcal{I}_a}$ follows immediately by IH. Assume $d = d_a$. Take d' with $(d, d') \in r^\mathcal{I}$ and $d' \in D^\mathcal{I}$. Again, if $d' \in \Delta^{\mathcal{I}_a}$, then the claim follows immediately from the IH. Now assume $d' \notin \Delta^{\mathcal{I}_a}$. Then $d' = b$ for some b with $r(a, b) \in \mathcal{A}$. By IH, $d' \in D^{\mathcal{I}_b}$. Hence $\mathcal{T} \cup t_\mathcal{A}(b) \models D$. By compactness, there exists a concept $E \in t_\mathcal{A}(b)$ such that $\mathcal{T} \models E \sqsubseteq D$. From $\mathcal{A} \models E(b)$ and $r(a, b) \in \mathcal{A}$ we obtain $\mathcal{A} \models \exists r.E(a)$. Therefore, $\exists r.E \in t_\mathcal{A}(a)$. But then $\mathcal{T} \cup t_\mathcal{A}(a) \models \exists r.D$ and we obtain $d_a \in C^{\mathcal{I}_a}$.

It follows that $\mathcal{I}$ is a model of $(\mathcal{T}, \mathcal{A})$ and $\mathcal{I} \not\models D_0(a_0)$. Hence $(\mathcal{T}, \mathcal{A}) \not\models D_0(a_0)$, and we have derived a contradiction.

Point 2. Let D_0 be a $\mathcal{EL}^u_\Sigma$-concept, $a_0 \in \mathsf{N_I}$, and assume that $(\mathcal{T}, \mathcal{A}) \models D_0(a_0)$. Set, for every $a \in \mathsf{ob} = \mathsf{Obj}(\mathcal{A}) \cup \{a_0\}$,

$$t^u_\mathcal{A}(a) = \{C \mid \mathcal{A} \models C(a),\ C \text{ an } \mathcal{EL}^u_\Sigma\text{-concept}\}.$$

We show that $\mathcal{T} \cup t^u_\mathcal{A}(a_0) \models D_0$. Then, using compactness, we find a $\mathcal{EL}^u_\Sigma$-concept C such that $\mathcal{T} \models C \sqsubseteq D_0$ and $\mathcal{A} \models C(a_0)$, as required. The construction is the same (except that now D_0 can be of the form $\exists u.D$ and the sets $t^u_\mathcal{A}(a)$ contain $\mathcal{EL}^u$-concepts). So we just provide the inductive proof of

$$d \in (\exists u.C)^{\mathcal{I}_a} \text{ iff } d \in (\exists u.C)^\mathcal{I},$$

for all $d \in \Delta^{\mathcal{I}_a}$ and $a \in \mathsf{ob}$. Assume $d \in (\exists u.C)^\mathcal{I} \cap \Delta^{\mathcal{I}_a}$. By IH, there exists $b \in \mathsf{ob}$ such that $C^{\mathcal{I}_b} \cap \Delta^{\mathcal{I}_b} \neq \emptyset$. Take such a b. Then $\mathcal{T} \cup t^u_\mathcal{A}(b) \models \exists u.C$. By compactness, there exists $E \in t^u_{\mathcal{A}}(b)$ with $\mathcal{T} \models E \sqsubseteq \exists u.C$. Assume first that E is an $\mathcal{EL}$-concept. Then $\mathcal{T} \models \exists u.E \sqsubseteq \exists u.C$. Moreover, $\mathcal{A} \models \exists u.E(a)$ because $\mathcal{A} \models E(b)$. We obtain $\exists u.E \in t^u_{\mathcal{A}}(a)$. Hence $\mathcal{T} \cup t^u_{\mathcal{A}}(a) \models \exists u.C$ and so $d \in (\exists u.C)^{\mathcal{I}_a}$. Now assume $E = \exists u.E'$. Then one can show similarly that $\exists u.E' \in t^u_{\mathcal{A}}(a)$ and so $\mathcal{T} \cup t^u_{\mathcal{A}}(a) \models \exists u.C$ which implies $d \in (\exists u.C)^{\mathcal{I}_a}$. $\square$

Let P and Q be finite sets of pairs (r, E), where r is a role and E a $\mathcal{EL}$-concept. Then Q *strongly covers* P *w.r.t. a TBox* $\mathcal{T}$, in symbols $P \leq^u_{\mathcal{T}} Q$, if $P \leq_{\mathcal{T}} Q$ and for all $\exists s.G \in \mathsf{sub}(\mathcal{T})$ and $(r, E) \in P$ with $\mathcal{T} \models E \sqsubseteq \exists u.G$ there exists $(r, E') \in Q$ with $\mathcal{T} \models E' \sqsubseteq \exists u.G$.

Lemma 49. *Let $\mathcal{T}$ be a TBox, $C_0 = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, $C_1 = F_0 \sqcap \bigsqcap_{(r,E)\in Q} \exists r.E$, and assume $P \leq^u_{\mathcal{T}} Q$. Then the following holds:*
- *$K^u_{\mathcal{T}}(C_0) \subseteq K^u_{\mathcal{T}}(C_1)$.*
- *If C' is a $\mathcal{EL}$-concept with $\exists r.C_0 \in \mathsf{sub}(C')$ and C'' the resulting concept when $\exists r.C_0$ is replaced by $\exists r.C_1$ in C', then $K^u_{\mathcal{T}}(C') \subseteq K^u_{\mathcal{T}}(C'')$.*

Proof. Similar to the proof of Lemma 47 and left to the reader. $\square$

Proposition 50. *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be TBoxes and Σ a signature. If C is an $\mathcal{EL}_\Sigma$-concept and $D \in K^u_{\mathcal{T}_2}(C)$ such that (a) there does not exist $D' \in \Delta^{\mathcal{I}_{C,\mathcal{T}_1}}$, $(\mathcal{T}_1, D') \sqsubseteq_\Sigma (\mathcal{T}_2, D)$, then there exist a $\mathcal{EL}_\Sigma$-concept C' and $D' \in K^u_{\mathcal{T}_2}(C')$ satisfying (a) and the outdegree of C' is bounded by $|\mathcal{T}_2|$.*

Proof. The argument is similar to the proof of Proposition 17. Assume C and D have property (a). If the outdegree of the C is bounded by $|\mathcal{T}_2|$, C itself is as required. Assume that this is not the case. Then there exists a subconcept C_0 of C such that $C_0 = F \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, and $|P| > |\mathcal{T}_2|$. Let Q be a minimal subset of P such that $P \leq^u_{\mathcal{T}_2} Q$. The cardinality of Q is bounded by $|\mathcal{T}_2|$. Now, replace in C the subconcept C_0 with $C_1 := F \sqcap \bigsqcap_{(r,E)\in Q} \exists r.E$ and call the result C'. We have $|C'| \leq |C|$ and, by Lemma 49, $K^u_{\mathcal{T}_2}(C) = K^u_{\mathcal{T}_2}(C')$. To obtain the desired concept C', we now execute the described contraction until the outdegree is bounded by $|\mathcal{T}_2|$. The resulting concept C' and D satisfy (a) because $K^u_{\mathcal{T}_2}(C) = K^u_{\mathcal{T}_2}(C')$ and $\emptyset \models C \sqsubseteq C'$. $\square$

Lemma 33. Let $\mathcal{T}$ be a TBox, $C = F_0 \sqcap \bigsqcap_{(r,E)\in P} \exists r.E$, where F_0 is a conjunction of concept names, and $D = \bigsqcap_{E \in K_{\mathcal{T}}(C)} E$. Then

$$K^u_{\mathcal{T}}(C) = K^u_{\mathcal{T}}(D) \cup \bigcup_{(r,E)\in P} K^u_{\mathcal{T}}(E).$$

Proof. The inclusion "$\supseteq$" is clear. Conversely, assume that $\exists r.H \in \mathsf{sub}(\mathcal{T})$ but $H \notin K^u_{\mathcal{T}}(D) \cup \bigcup_{(r,E)\in P} K^u_{\mathcal{T}}(E)$. We show $H \notin K^u_{\mathcal{T}}(C)$. The construction is similar to the proof of Lemma 47, so we only give a sketch. Take, for every $(r, E) \in P$, a copy $\mathcal{I}_{r,E}$ of the canonical model $\mathcal{I}_{E,\mathcal{T}}$ such that all these copies have disjoint domains. In the copy $\mathcal{I}_{r,E}$, the point corresponding to $E \in \Delta^{\mathcal{I}_{E,\mathcal{T}}}$ is denoted by $d_{r,E}$. We have $H^{\mathcal{I}_{r,E}} = \emptyset$ for all $(r, E) \in P$. Consider, in addition, the canonical model $\mathcal{I}_{D,\mathcal{T}}$ and assume it is disjoint from the models $\mathcal{I}_{r,E}$, $(r, E) \in P$. Again, $H^{\mathcal{I}_{D,\mathcal{T}}} = \emptyset$. Define a new model $\mathcal{I}$ by taking the union of the model $\mathcal{I}_{D,\mathcal{T}}$ and the models $\mathcal{I}_{r,E}$, $(r, E) \in P$, and adding $(D, d_{r,E})$ to the interpretation of $r^{\mathcal{I}}$ for every $(r, E) \in P$, and D to the interpretation of $B^{\mathcal{I}}$ for all conjuncts B of F_0 which are not in $\mathsf{sig}(\mathcal{T})$.

To prove that $H \notin K^u_{\mathcal{T}}(C)$, it is sufficient to show that $\mathcal{I}$ is a model of $\mathcal{T}$ refuting $C \sqsubseteq \exists u.H$. We clearly have $D \in C^{\mathcal{I}}$. Thus, it is sufficient to show

- for all $d \in \Delta^{\mathcal{I}_{D,\mathcal{T}}}$ and $G \in \mathsf{sub}(\mathcal{T})$: $d \in G^{\mathcal{I}}$ iff $d \in G^{\mathcal{I}_{D,\mathcal{T}}}$ and
- for all $(r, E) \in P$, $d \in \mathcal{I}_{r,E}$, and $G \in \mathsf{sub}(\mathcal{T})$: $d \in G^{\mathcal{I}}$ iff $d \in G^{\mathcal{I}_{r,E}}$.

This can be proved by induction on the construction of G (similarly to the proof of Lemma 47) and is left to the reader. □

E. Omitted Proofs for Section 6

Lemma 35. Player 1 has a winning strategy in G iff $\mathcal{T}_G \cup \mathcal{T}'_G$ is not a conservative extension of $\mathcal{T}_G$.

Proof. First assume that Player 1 has a winning strategy (V, E, ℓ) in G. We first define a mapping $m : V \to \{0, \dots, n\}$ as follows: if $(v, v') \in E$, $\ell(v) = (t, i)$, and $\ell(v') = (t', i')$, then $m(v')$ is the variable that was switched to reach t' from t (we assume that $m(v') = n$ means that no variable was switched). If $v \in V$ is the root, $m(v) = n$ (this is arbitrary). We associate a concept $C(v)$ with each node $v \in V$: if $\ell(v) = (t, i)$, then

$$C(v) := P_i \sqcap F_{m(v)} \sqcap \bigsqcap_{i \in t} V_i \sqcap \bigsqcap_{i \in (\Gamma_1 \cup \Gamma_2) \setminus t} \overline{V}_i$$

As a next step, we inductively associate another concept $W(v)$ with each node $v \in V$:

- if v is a leaf, then $W(v) := C(v)$;
- if v has successors $v_0, \dots, v_{\ell-1}$, then $W(v) = C(v) \sqcap \bigsqcap_{i<\ell} \exists r.W(v_i)$.

Let ε be the root of (V, E, ℓ) and define $W := W(\varepsilon)$. It is not too difficult to verify that $\mathcal{T}_G \cup \mathcal{T}'_G \models W \sqsubseteq B$. We show that $\mathcal{T}_G \not\models W \sqsubseteq B$, and thus $\mathcal{T}_G \cup \mathcal{T}'_G$ is not a conservative extension of $\mathcal{T}_G$. Define a model $\mathcal{I}$ as follows:

- $\Delta^{\mathcal{I}} := \{W\} \cup \{C \mid \exists r.C \in \mathsf{sub}(W)\}$;
- $A^{\mathcal{I}} := \{C \in \Delta^{\mathcal{I}} \mid A \text{ is a conjunct in } C\}$ for all $A \in \mathsf{N_C}$;
- $r^{\mathcal{I}} := \{(C, C') \mid \exists r.C' \text{ is a conjunct in } C\}$ for all $r \in \mathsf{N_R}$.

Here, "D being a conjunct of C" refers to top-level conjunctions and includes the case that $C = D$. It is easy to verify that $\mathcal{I}$ is a model of $\mathcal{T}_G$, and that $W \in W^{\mathcal{I}}$. Also, we have $B^{\mathcal{I}} := \emptyset$, and thus $\mathcal{T}_G \not\models W \sqsubseteq B$.

For the converse direction, we start with a preliminary. A model $\mathcal{I}$ of a TBox $\mathcal{T}$ is a *tree model* if the graph $(\Delta^{\mathcal{I}}, \bigcup_{r \in \mathsf{N_R}} r^{\mathcal{I}})$ is a tree. As in Section 7, for two interpretations $\mathcal{I}$ and $\mathcal{I}'$ and a signature Σ we write $\mathcal{I}|_\Sigma \equiv \mathcal{I}'|_\Sigma$ if $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}'}$ and $\sigma^{\mathcal{I}} = \sigma^{\mathcal{I}'}$ for all symbols $X \in \Sigma$. If

$(*)$ for every tree model $\mathcal{I}$ of $\mathcal{T}_G$, there is a model $\mathcal{I}'$ of $\mathcal{T}_G \cup \mathcal{T}'_G$ with $\mathcal{I}|_{\mathsf{sig}(\mathcal{T}_G)} = \mathcal{I}'|_{\mathsf{sig}(\mathcal{T}_G)}$

then $\mathcal{T}_G \cup \mathcal{T}'_G$ is a conservative extension of $\mathcal{T}_G$. To see this, let $\mathcal{T}_G \not\models C \sqsubseteq D$ with $\mathsf{sig}(C) \cup \mathsf{sig}(D) \subseteq \mathsf{sig}(\mathcal{T}_G)$. Then there is a model $\mathcal{I}$ of $\mathcal{T}_G$ and a $d \in C^{\mathcal{I}} \setminus D^{\mathcal{I}}$, and we can unravel $\mathcal{I}$ into a tree model $\mathcal{J}$ of $\mathcal{T}_G$ with root d and such that $d \in C^{\mathcal{J}} \setminus D^{\mathcal{J}}$. The existence of a model $\mathcal{J}'$ of $\mathcal{T}_G \cup \mathcal{T}_G'$ with $\mathcal{J}|_{\mathsf{sig}(\mathcal{T}_G)} = \mathcal{J}'|_{\mathsf{sig}(\mathcal{T}_G)}$ then shows that $\mathcal{T}_G \cup \mathcal{T}_G' \not\models C \sqsubseteq D$.

To prove the converse of Lemma 35, it thus suffices to show that if Player 1 does not have a winning strategy, then (∗) holds. Thus, suppose that Player 1 does not have a winning strategy, and let $\mathcal{I}$ be a tree model of $\mathcal{T}_G$. We define a sequence of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots$ whose limit $\mathcal{I}'$ is the desired interpretation, i.e., a model of $\mathcal{T}_G \cup \mathcal{T}_G'$ with $\mathcal{I}|_{\mathsf{sig}(\mathcal{T}_G)} = \mathcal{I}'|_{\mathsf{sig}(\mathcal{T}_G)}$.

To define $\mathcal{I}_0$, we start with $\mathcal{I}$ and redefine the interpretation of the concept names that occur in $\mathcal{T}_G'$, but not in $\mathcal{T}_G$; these are N, N', N'', $N_0, \ldots, N_{n-1}$, and X_ψ, with $\psi \in \mathsf{sub}(\varphi)$ not a literal. The interpretation of the new symbols by $\mathcal{I}'$ directly reflects the CIs in $\mathcal{T}_G'$:

$$\begin{aligned}
N_i^{\mathcal{I}_0} &= V_i^{\mathcal{I}} \cup \overline{V}_i^{\mathcal{I}} \text{ for all } i < n \\
(N')^{\mathcal{I}_0} &= F_0^{\mathcal{I}} \cup \cdots \cup F_{k-1}^{\mathcal{I}} \cup F_n^{\mathcal{I}} \\
(N'')^{\mathcal{I}_0} &= F_k^{\mathcal{I}} \cup \cdots \cup F_n^{\mathcal{I}} \\
X_{\vartheta \wedge \chi}^{\mathcal{I}_0} &= X_\vartheta^{\mathcal{I}_0} \cap X_\chi^{\mathcal{I}_0} \text{ for all } \vartheta \wedge \chi \in \mathsf{sub}(\varphi) \\
X_{\vartheta \vee \chi}^{\mathcal{I}_0} &= X_\vartheta^{\mathcal{I}_0} \cup X_\chi^{\mathcal{I}_0} \text{ for all } \vartheta \vee \chi \in \mathsf{sub}(\varphi) \\
N^{\mathcal{I}_0} &= (X_\varphi \sqcap P_1 \sqcap N'' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1})^{\mathcal{I}_0}. \\
&\quad \cup (X_\varphi \sqcap P_2 \sqcap N' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1})^{\mathcal{I}_0}.
\end{aligned}$$

The interpretation $\mathcal{I}_0$ is almost the desired one, except that the definition of $N^{\mathcal{I}_0}$ does not take into account all CIs in $\mathcal{T}_G'$ with N on the right-hand side. This problem is addressed by the interpretations $\mathcal{I}_1, \mathcal{I}_2, \ldots$, which are identical to $\mathcal{I}_0$ except for the interpretation of N:

$$\begin{aligned}
N^{\mathcal{I}_{i+1}} &= N^{\mathcal{I}_i} \cup (P_1 \sqcap N'' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1} \sqcap \exists r.N)^{\mathcal{I}_i} \\
&\quad \cup (P_2 \sqcap N' \sqcap N_0 \sqcap \cdots \sqcap N_{n-1} \sqcap \prod_{i \in \{0,\ldots,k-1,n\}} \exists r.(N \sqcap F_i))^{\mathcal{I}_i}.
\end{aligned}$$

Let $\mathcal{I}'$ be the limit of the sequence $\mathcal{I}_0, \mathcal{I}_1, \ldots$. By construction, $\mathcal{I}'$ is a model of $\mathcal{T}_G$. Additionally, all CIs in $\mathcal{T}_G'$ are easily seen to be satisfied by $\mathcal{I}'$, with the exception of

(I) $P_1 \sqcap N \sqcap \prod_{i \in \Gamma_I} V_i \sqcap \prod_{i \notin \Gamma_I} \overline{V}_i \sqsubseteq B$ and

(II) $P_1 \sqcap X_\varphi \sqcap \prod_{i \in \Gamma_I} V_i \sqcap \prod_{i \notin \Gamma_I} \overline{V}_i \sqsubseteq B$.

Assume that one of these CIs is not satisfied by $\mathcal{I}'$. We show that this implies the existence of a winning strategy for Player 1 in G, in contradiction to the assumption that there is no such strategy.

The case of (II) is simple. If (II) is not satisfied, there is a $d \in \Delta^{\mathcal{I}'}$ that satisfies the left-hand side of this CI and is not in $B^{\mathcal{I}'}$. Together with the CIs in $\mathcal{T}_G$, these two properties of d imply that d satisfies exactly one of V_i and $\overline{V}_i$, for all $i < n$, and that the corresponding valuation is Γ_I. Since $d \in X_\varphi^{\mathcal{I}'}$ and by construction of $\mathcal{I}'$, Γ_I satisfies φ. Thus, there is a trivial winning strategy for P_1 in G.

Now assume that (I) is violated. Then there is a $d_0 \in \Delta^{\mathcal{I}'}$ with $d_0 \in L^{\mathcal{I}'} \setminus B^{\mathcal{I}'}$, where L denotes the left-hand side of (I). Since $\mathcal{I}$ is a tree model, so is $\mathcal{I}'$. In the following, we use subsets $S \subseteq \Delta^{\mathcal{I}'}$ such that the restriction $\mathcal{I}'|_S$ of $\mathcal{I}'$ to domain S is a tree with root

d_0 to descibe partial winning strategies. More precisely, define the node-labeled graph G_S as $(S, r^{\mathcal{I}'|_S}, \ell)$, where $\ell(d) = (t_d, p_d)$ with

- t_d the valuation that makes variable i true if $d \in V_i^{\mathcal{I}'}$ and false if $d \in \overline{V}_i^{\mathcal{I}'}$, for all $i < n$;
- $p_d = 1$ if $d \in P_1^{\mathcal{I}'}$ and $p_d = 2$ if $d \in P_2^{\mathcal{I}'}$.

Regarding well-definedness of $\ell(d)$, note that since $d_0 \notin B^{\mathcal{I}'}$, the CIs in $\mathcal{T}_G$ ensure that $d \notin V_i^{\mathcal{I}'} \cap \overline{V}_i^{\mathcal{I}'}$, for all $i < n$. For the same reason, $d \notin P_1^{\mathcal{I}'} \cap P_2^{\mathcal{I}'}$. Moreover, we will choose the set S such that for all elements $d \in S$, we have $d \in V_i^{\mathcal{I}'} \cup \overline{V}_i^{\mathcal{I}'}$ and $d \in P_1^{\mathcal{I}'} \cup P_2^{\mathcal{I}'}$ (see Point (iii) below).

We inductively construct a finite sequence of subsets $S_0, \dots, S_m$ of $\Delta^{\mathcal{I}'}$ such that G_{S_m} is a (complete) winning strategy for Player 1 in G. During the construction, we ensure that for all $i \geq 0$,

(i) G_{S_i} satisfies Conditions (1) to (3) of winning strategies for Player 1 in G;
(ii) all elements of S_i are in $N^{\mathcal{I}'}$;
(iii) $S_i \subseteq V_j^{\mathcal{I}'} \cup \overline{V}_j^{\mathcal{I}'}$ for all $j < n$ and $S_i \subseteq P_1^{\mathcal{I}'} \cup P_2^{\mathcal{I}'}$.

We start with $S_0 = \{d_0\}$. Then, (i)-(iii) are satisfied since $d_0 \in L^{\mathcal{I}'}$. To define S_{i+1} from S_i, we proceed as follows. If $i > 0$ and all leaves of $\mathcal{I}'|_{S_i}$ are in $X_\varphi^{\mathcal{I}'}$, then S_i is the last element of the sequence. Otherwise, we do the following for all leaves d of $\mathcal{I}'|_{S_i}$ with $d \notin X_\varphi^{\mathcal{I}'}$. By (ii), $d \in N^{\mathcal{I}'}$. By construction of $\mathcal{I}'$ and since $d \notin X_\varphi^{\mathcal{I}'}$, this means that

$$\begin{aligned} d \in\ & (P_1 \sqcap N'' \sqcap N_0 \sqcap \dots \sqcap N_{n-1} \sqcap \exists r.N)^{\mathcal{I}'} \\ & \cup (P_2 \sqcap N' \sqcap N_0 \sqcap \dots \sqcap N_{n-1} \sqcap \bigsqcap_{i \in \{0,\dots,k-1,n\}} \exists r.(N \sqcap F_i))^{\mathcal{I}'}. \end{aligned}$$

If the former is the case, add an element $e \in N^{\mathcal{I}'}$ to S_i such that $(d, e) \in r^{\mathcal{I}'}$. Otherwise, add elements $e_0, \dots, e_{k-1}, e_n$ such that $e_j \in (N \sqcap F_j)^{\mathcal{I}'}$ and $(d, e_j) \in r^{\mathcal{I}'}$, for $j \in \{0, \dots, k-1, n\}$. Condition (ii) is clearly satisfied. By construction of $\mathcal{I}'$, e and the e_j are elements of

$$\begin{aligned} & (X_\varphi \sqcap P_1 \sqcap N'' \sqcap N_0 \sqcap \dots \sqcap N_{n-1})^{\mathcal{I}'} \\ & \cup (X_\varphi \sqcap P_2 \sqcap N' \sqcap N_0 \sqcap \dots \sqcap N_{n-1})^{\mathcal{I}'} \\ & \cup (P_1 \sqcap N'' \sqcap N_0 \sqcap \dots \sqcap N_{n-1} \sqcap \exists r.N)^{\mathcal{I}'} \\ & \cup (P_2 \sqcap N' \sqcap N_0 \sqcap \dots \sqcap N_{n-1} \sqcap \bigsqcap_{i \in \{0,\dots,k-1,n\}} \exists r.(N \sqcap F_i))^{\mathcal{I}'}. \end{aligned}$$

This implies that $S_i \subseteq P_1^{\mathcal{I}'} \cup P_2^{\mathcal{I}'}$. Additionally, it shows that e and the e_j are instances of $N_0, \dots, N_{n-1}$. By construction of $\mathcal{I}'$, it follows that (iii) is satisfied. To show that Condition (i) is satisfied as well, we need to argue that (a) $p_e = 2$ iff $p_d = 1$, (b) t_e is obtained from t_d by switching the truth value of a variable in Γ_1 or $t_e = t_d$, (c) t_{e_j} is obtained from t_d by switching the truth value of variable j for all $j < k$, and (d) $t_{e_n} = t_d$. Assume that $d \in P_1^{\mathcal{I}'}$ (the case that $d \in P_2^{\mathcal{I}'}$ is analogous). By the CIs in $\mathcal{T}_G$, this implies $e \in P_2^{\mathcal{I}'}$. This shows (a). Since $e \in P_2^{\mathcal{I}'}$ and e is contained in the above union, we have $e \in (N')^{\mathcal{I}'}$. By construction of $\mathcal{I}'$, there thus is a $j \in \{0, \dots, k-1, n\}$ such that $e \in F_j^{\mathcal{I}'}$. Together with the CIs in $\mathcal{T}_G$, this means that the truth value of variable j is different in t_e and t_d if $j < n$. Also by CIs in $\mathcal{T}_G$, $e \in F_i^{\mathcal{I}'}$ means that all truth values in t_d and t_e of variables ℓ with $\ell \neq j$ are identical, which establishes (b). The proofs of (c) and (d) are similar and left to the reader.

We have to argue that the construction of the sequence $S_0, S_1, \ldots$ terminates. By (ii), $d_0 \in N^{\mathcal{I}'}$. Let ℓ be minimal such that $d_0 \in N^{\mathcal{I}_\ell}$. It is easily verified that if $d \in S_i$ and $d \in N^{\mathcal{I}_j}$, then all successors of d are in $N^{\mathcal{I}_{j-1}}$, for all $i, j \geq 0$. Termination follows.

Finally, it remains to note that, by construction, the last element S_m of the constructed sequence satisfies Condition (4) of winning strategies. $\Box$

F. Omitted Proofs for Section 7

Lemma 40. $\mathcal{T}_M \cup \mathcal{T}'_M$ is not a model conservative extension of $\mathcal{T}_M$ iff M halts on the empty tape.

Proof. "$\Leftarrow$" Assume that M halts on the empty tape and let $c_0, \ldots, c_k$ be the halting computation of M. Extend this computation to an infinite sequence of computations by setting $c_\ell := c_k$ for all $\ell > k$. We define an interpretation $\mathcal{I}$ as follows:
- $\Delta^{\mathcal{I}} := \mathbb{N} \times \mathbb{N}$;
- $s^{\mathcal{I}} := \{((i,j),(i+1,j)) \mid i,j \geq 0\}$;
- $n^{\mathcal{I}} := \{((i,j),(i,j+1)) \mid i,j \geq 0\}$;
- $q^{\mathcal{I}} := \{(i,j) \mid i,j > 0 \text{ and the state in } c_{j-1} \text{ is } q\}$ for all $q \in Q$;
- $a^{\mathcal{I}} := \{(i,j) \mid i,j > 0 \text{ and tape cell } i-1 \text{ in } c_{j-1} \text{ is labeled } a\}$ for all $a \in \Gamma$;
- $\mathsf{head}^{\mathcal{I}} := \{(i,j) \mid i,j > 0 \text{ and the head position in } c_{j-1} \text{ is } i-1\}$;
- $\mathsf{before}^{\mathcal{I}} := \{(i,j) \mid (i',j) \in \mathsf{head}^{\mathcal{I}} \text{ for some } i' > i\}$;
- $\mathsf{after}^{\mathcal{I}} := \{(i,j) \mid (i',j) \in \mathsf{head}^{\mathcal{I}} \text{ for some } i' < i\}$;
- $D^{\mathcal{I}} = \emptyset$.

It is not hard to verify that $\mathcal{I}$ is a model of $\mathcal{T}_M$ (setting $c_\ell = c_k$ for all $\ell > k$ is justified by the fact that M does not allow any transitions in the halting state). Moreover, $\mathcal{I}$ cannot be extended to a model of $\mathcal{T}_M \cup \mathcal{T}'_M$: in any model $\mathcal{J}$ of $\mathcal{T}'_M$ which coincides on $\mathsf{sig}(\mathcal{T}_M)$ with $\mathcal{I}$ we would have $(0,0) \in (\exists n.\exists s.(N \sqcap q_0 \sqcap \mathsf{head}))^{\mathcal{J}}$, so we have to interpret u_0, A, and B such that $(0,0) \in (\exists u_0.(\exists n.\exists s.A \sqcap \exists s.\exists n.B))^{\mathcal{J}}$. To do this, we have to interpret A and B in $\mathcal{J}$ such that $(i,j) \in (A \sqcap B)^{\mathcal{J}}$ for some $i,j \geq 0$. Thus, we must ensure that $(i,j) \in D^{\mathcal{J}}$. This, however, is impossible since $D^{\mathcal{I}} = \emptyset$ is fixed. It follows that $\mathcal{T}_M \cup \mathcal{T}'_M$ is not a model conservative extension of $\mathcal{T}_M$.

"$\Rightarrow$". Assume that M does not halt on the empty tape and let $\mathcal{I}$ be a model of $\mathcal{T}_M$. We have to show that $\mathcal{I}$ can be extended to a model of $\mathcal{T}_M \cup \mathcal{T}'_M$. If $q_h^{\mathcal{I}} = \emptyset$, then we simply set $A^{\mathcal{I}} := B^{\mathcal{I}} := N^{\mathcal{I}} := u_0^{\mathcal{I}} := \emptyset$. If $q_h^{\mathcal{I}} \neq \emptyset$, let $N^{\mathcal{I}}$ be the smallest set such that $q_h^{\mathcal{I}} \subseteq N^{\mathcal{I}}$, $(\exists n.N)^{\mathcal{I}} \subseteq N^{\mathcal{I}}$, and $(\exists s.N)^{\mathcal{I}} \subseteq N^{\mathcal{I}}$. If the result is such that $(\exists n.\exists s.(N \sqcap q_0 \sqcap \mathsf{head}))^{\mathcal{I}} = \emptyset$, we are done. So assume the contrary. First assume that

(i) There are $d, d_1, d_2, d_3, d_4 \in \Delta^{\mathcal{I}}$ with $d n^{\mathcal{I}} d_1 s^{\mathcal{I}} d_2$ and $d s^{\mathcal{I}} d_3 n^{\mathcal{I}} d_4$ such that $d_2 \neq d_4$. Then we can set $u_0^{\mathcal{I}} := \Delta^{\mathcal{I}} \times \{d\}$, $A^{\mathcal{I}} := \{d_2\}$, and $B^{\mathcal{I}} := \{d_4\}$, and obtain a model of $\mathcal{T}'_M$. Now assume

(ii) There are $d_1, d_2, d_3, d_4 \in \Delta^{\mathcal{I}}$ with $d_1 n^{\mathcal{I}} d_2 s^{\mathcal{I}} d_4$, $d_1 s^{\mathcal{I}} d_3 n^{\mathcal{I}} d_4$, and $d_4 \in D^{\mathcal{I}}$.

Then we can set $u_0^{\mathcal{I}} := \Delta^{\mathcal{I}} \times \{d_1\}$ and $A^{\mathcal{I}} := B^{\mathcal{I}} := \{d_4\}$ to obtain a model of $\mathcal{T}'_M$. Now assume that neither (i) nor (ii) are the case. We show that this is impossible since it implies that M halts on the empty tape. Let $d_0 \in (\exists n.\exists s.(N \sqcap q_0 \sqcap \mathsf{head}))^{\mathcal{I}}$. Then there is a $d'_0 \in \Delta^{\mathcal{I}}$ and a $d \in (N \sqcap q_0 \sqcap \mathsf{head})^{\mathcal{I}}$ such that $d_0 n^{\mathcal{I}} d'_0 s^{\mathcal{I}} d$. For $d' \in \Delta^{\mathcal{I}}$, we say that d' is *reachable* from d *in* n *steps* if there exists a sequence $d_0, \ldots, d_n$ with $d_0 = d$, $d_n = d'$, and $(d_i, d_{i+1}) \in n^{\mathcal{I}} \cup s^{\mathcal{I}}$ for all $i < n$. We say that that d' is *reachable* from d if d' is reachable from d in n steps, for some $n \geq 0$. We first show the following:

Claim. Let d' be reachable from d. Then we have:
(1) there are $d_1, d_2, d_3 \in \Delta^{\mathcal{I}}$ such that $d_1 n^{\mathcal{I}} d_2 s^{\mathcal{I}} d'$ and $d_1 s^{\mathcal{I}} d_3 n^{\mathcal{I}} d'$;
(2) if $d' n^{\mathcal{I}} e$ and $d' n^{\mathcal{I}} e'$, then $e = e'$;
(3) if $d' s^{\mathcal{I}} e$ and $d' s^{\mathcal{I}} e'$, then $e = e'$;
Point 1 is proved by induction on the minimal n such that d' is reachable from d in n steps. For the induction start, we have $d' = d$. Recall that $d_0 n^{\mathcal{I}} d'_0 s^{\mathcal{I}} d$. By the CIs in $\mathcal{T}_M$, there are $d_1, d_2 \in \Delta^{\mathcal{I}}$ such that $d_0 s^{\mathcal{I}} d_1 n^{\mathcal{I}} d_2$. Since (i) does not hold, $d_2 = d$ and we are done. For the induction step, let d' be reachable from d in $n > 0$ steps. Then there is a d_1 such that d_1 is reachable from d in $n-1$ steps and $d_1 n^{\mathcal{I}} d'$ or $d_1 s^{\mathcal{I}} d'$. We only treat the first case since the second is analogous. By IH, there is a d_2 such that $d_2 s^{\mathcal{I}} d_1$. By the CIs in $\mathcal{T}_M$, there are d_3 and d_4 such that $d_2 n^{\mathcal{I}} d_3 s^{\mathcal{I}} d_4$. Since (i) does not holds, $d_4 = d'$ and we are done.

Now for Points 2 and 3. We only treat Point 2 explicitly since Point 3 is analogous. Let d' be reachable from d and let $e, e' \in \Delta^{\mathcal{I}}$ such that $d' n^{\mathcal{I}} e$ and $d' n^{\mathcal{I}} e'$. By Point 1, there is a d_1 such that $d_1 s^{\mathcal{I}} d'$. By the CIs in $\mathcal{T}_M$, there are d_2, d_3 such that $d_1 n^{\mathcal{I}} d_2 s^{\mathcal{I}} d_3$. Since (i) does not hold, we have $d_3 = e = e'$, and are done. This finishes the proof of the claim.

Set $R := \{d' \in \Delta^{\mathcal{I}} \mid d' \text{ is reachable from } d\}$. Points 2 and 3 of the claim together with the fact that (i) does not hold implies that we can easily find a bijection $\tau : R \to \mathbb{N} \times \mathbb{N}$ such that for all $e, e' \in R$, we have
- $e n^{\mathcal{I}} e'$ iff $\tau(e) = (i, j)$ and $\tau(e) = (i + 1, j)$ for some $i, j \in \mathbb{N}$;
- $e s^{\mathcal{I}} e'$ iff $\tau(e) = (i, j)$ and $\tau(e) = (i, j + 1)$ for some $i, j \in \mathbb{N}$.

Our aim is to read off a halting computation from M on the empty tape from $\mathcal{I}$, being guided by τ. To do this, we first show that (a) for all $q, q' \in Q$ with $q \neq q'$, $q^{\mathcal{I}} \cap q'^{\mathcal{I}} \cap R = \emptyset$, (b) for all $a, a' \in \Gamma$ with $a \neq a'$, $a^{\mathcal{I}} \cap a'^{\mathcal{I}} \cap R = \emptyset$, and (c) $\mathsf{before}^{\mathcal{I}} \cap R$, $\mathsf{after}^{\mathcal{I}} \cap R$, and $\mathsf{head}^{\mathcal{I}} \cap R$ are pairwise disjoint. Since the argument is the same in all three cases, we concentrate on (a). Assume $e \in q^{\mathcal{I}} \cap q'^{\mathcal{I}} \cap R$. By the GCIs in $\mathcal{T}_M$, $d' \in D^{\mathcal{I}}$. By Point 1 of the claim, there are $d_1, d_2, d_3 \in \Delta^{\mathcal{I}}$ such that $d_1 n^{\mathcal{I}} d_2 s^{\mathcal{I}} d'$ and $d_1 s^{\mathcal{I}} d_3 n^{\mathcal{I}} d'$. This is a contradiction to the fact that (ii) is false

We can now read off a halting computation from M in the obvious way: the i-th configuration is described by the elements $R_i := \{d \in R \mid \tau(d) = (j, i) \text{ for some } j \geq 0\}$. By the CIs in $\mathcal{T}_M$ and (a), there is a unique state $q \in Q$ such that $R_i \subseteq q^{\mathcal{I}}$. By the CIs in $\mathcal{T}_M$ and (b), for each $j \geq 0$, there is a unique $a \in \Gamma$ such that $\tau^{-1}(j, i) \in a^{\mathcal{I}}$. And by the CIs in $\mathcal{T}_M$ and (c), there is a unique $j \geq 0$ such that $\tau^{-1}(j, i) \in \mathsf{head}^{\mathcal{I}}$. Let us call the resulting sequence of configurations $c_0, c_1, \ldots$. By choice of d above and the CIs in $\mathcal{T}_M$, c_0 is the initial configuration of M on the empty tape. By the CIs in $\mathcal{T}_M$, c_{i+1} is a successor configuration of c_i for all $i \geq 0$. By definition of $N^{\mathcal{I}}$ and since $d \in N^{\mathcal{I}}$, it follows that we eventually reach a halting configuration. □