Proceedings of the 23rd International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Fitting Horn DL Ontologies to ABox and Query Examples: A Tale of Simulation Quantifiers and Finite Models

Marvin Grosser, **Carsten Lutz**
Leipzig University
ScaDS.AI Center Dresden/Leipzig
{grosser, clu}@informatik.uni-leipzig.com

Abstract

We study the problem of fitting a description logic (DL) ontology to a given set of positive and negative examples in the form of an ABox and a Boolean query. While previous work has investigated this problem for the expressive DLs $\mathcal{ALC}$ and $\mathcal{ALCI}$, we here focus on the Horn DLs $\mathcal{EL}$ and $\mathcal{ELI}$, as well as their extensions with the bottom concept. As the query language, we consider atomic queries (AQs), conjunctive queries (CQs), and unions thereof (UCQs). We provide characterizations of the existence of a fitting ontology based on simulations, use them to develop decision procedures, and clarify the exact computational complexity. For AQs, the problem is in PTIME for both $\mathcal{EL}$ and $\mathcal{ELI}$. For CQs and UCQs, it is Σ_2^p-complete for $\mathcal{EL}$ and EXPTIME-complete for $\mathcal{ELI}$. Adding the bottom concept does not change any of these complexities. Interestingly, moving from $\mathcal{ALC}$ and $\mathcal{ALCI}$ to $\mathcal{EL}$ and $\mathcal{ELI}$ introduces additional technical challenges rather than simplifying the matter.

1 Introduction

In knowledge representation and databases, fitting problems aim to support human engineers in constructing ontologies and queries. Given as input a collection of positively and negatively labeled examples, a fitting algorithm outputs a formal object that fits all examples or reports that no such object exists. The study of fitting problems originated in databases, where the formal object to be constructed is a query (Zloof 1975). This line of work is known as query-by-example, and sometimes as reverse query engineering, and it continues to be studied today, see for instance (Cohen and Weiss 2016; Barceló and Romero 2017; ten Cate et al. 2023a) and references therein. Although not the focus of this article, fitting problems are also closely related to machine learning, where fitting a model to a set of examples is linked to PAC-style generalization guarantees (Shalev-Shwartz and Ben-David 2014).

Fitting problems have also attracted considerable interest in ontology-based KR, particularly in the context of description logics (DLs). There are at least two ways in which fitting problems can support ontology engineering. On the one hand, they can help an engineer construct *concepts* that can then serve as building blocks in an ontology. On the other hand, fitting can be used to construct entire *ontologies*, either for stand-alone use or as a module in a larger ontology under development.

Interest in concept fitting first emerged under the label of *DL concept learning* although, unlike in machine learning, generalization guarantees were typically not considered; see (ten Cate et al. 2023b) for an exception. This research line led to the development of practical systems, most prominently DL Learner (Lehmann and Hitzler 2010; Bühmann et al. 2018). Starting with (Funk et al. 2019), the focus shifted to algorithms for concept fitting that are sound, complete and terminating, as well as to investigations of the decidability and computational complexity of concept fitting in the presence of an ontology; see for instance (Jung et al. 2022) and the references therein.

Ontology fitting, by contrast, has attracted attention only relatively recently. A key design choice concerns the form of the examples. One natural option, adopted in (Hosemann et al. 2025), is to use interpretations (logical structures) as examples. Another, used in (Funk, Grosser, and Lutz 2025), is to use pairs $(\mathcal{A}, q)$ where $\mathcal{A}$ is an ABox and q is a (Boolean) query. For a positive example, a fitting ontology $\mathcal{O}$ must then satisfy $\mathcal{A} \cup \mathcal{O} \models q$ while for a negative example it must satisfy $\mathcal{A} \cup \mathcal{O} \not\models q$. The latter setting is particularly relevant in the context of ontology-mediated querying (Bienvenu and Ortiz 2015; Xiao et al. 2018).

In this article, we study the problem of fitting ontologies to ABox-query examples. Whereas previous work has focused on expressive DLs such as $\mathcal{ALC}$ and $\mathcal{ALCI}$ as the ontology language, we instead consider Horn DLs, especially $\mathcal{EL}$, $\mathcal{ELI}$, and their extensions with the bottom concept '$\bot$'. These DLs are of special interest because they are computationally more tractable than their expressive counterparts. In particular, $\mathcal{EL}$ admits subsumption and querying in polynomial time for tree-shaped queries, and both $\mathcal{EL}$ and $\mathcal{ELI}$ are supported by so-called consequence-based reasoning algorithms that are highly efficient in practice (Cucala, Grau, and Horrocks 2019). This has contributed to an extension of $\mathcal{EL}$ becoming one of three profiles of the OWL 2 ontology language standard (Motik et al. 2009). As query languages, we consider atomic queries (AQs), conjunctive queries (CQs), and unions thereof (UCQs). In addition, we consider a fitting problem in which the examples consist of an ABox only and where we seek an ontology that is consistent with the positive examples and inconsistent with the negative ones. For all these fitting problems, we provide effective characterizations and determine the precise complexity of deciding whether a fitting ontology exists. Our algorithms also produce explicit fitting ontologies when they exist.

Our characterizations resemble those obtained in (Funk, Grosser, and Lutz 2025) for the expressive DLs $\mathcal{ALC}$ and $\mathcal{ALCI}$, but they also exhibit notable differences. To begin with, the characterizations for $\mathcal{ALC}$ and $\mathcal{ALCI}$ are based on homomorphisms which ours replace with simulations. This introduces technical challenges, especially for CQs and UCQs as the query language, intuitively because homomorphisms are more local in nature than simulations. For example, a reflexive loop must map homomorphically to a reflexive loop, but it may be simulated by an infinite path. When turning our characterizations into fitting algorithms, we are therefore led to employ the extension of $\mathcal{EL}$ and $\mathcal{ELI}$ with simulation quantifiers as studied for the $\mathcal{EL}$ case in (Lutz, Piro, and Wolter 2010). In the case of $\mathcal{ELI}$, an additional complication arises: the non-local nature of simulations clashes, in a certain sense, with the more local nature of $\mathcal{ELI}$-concepts used in ontologies. We address this issue by working with finite interpretations rather than with unrestricted ones in our characterizations, and by considering finite model reasoning in $\mathcal{ELI}$ extended with simulation quantifiers.

Table 1 summarizes our complexity results and compares them with the known complexities of query entailment in $\mathcal{EL}$ and $\mathcal{ELI}$. All results are completeness results, except the PTIME ones (for which we conjecture completeness under logspace-reductions as well). The lower bounds for (U)CQs already hold for rooted queries, i.e., (U)CQs in which every connected component contains an answer variable. The complexities are generally lower than for $\mathcal{ALC}$ and $\mathcal{ALCI}$. For example, $\mathcal{ALC}$- and $\mathcal{ALCI}$-ontology fitting is 2EXPTIME-complete for CQs and UCQs, and NP-complete for AQs (Funk, Grosser, and Lutz 2025). The complexity also differs from that of fitting ontologies to examples that take the form of interpretations. That problem is EXPTIME-complete for all four DLs listed in Table 1 (Hosemann et al. 2025).

We also discuss, throughout the paper, the size and nature of fitting ontologies. For AQs and for consistency-based fitting, the existence of a fitting ontology coincides with the existence of a fitting ontology that does not contain auxiliary symbols (i.e., concept or role names not present in the examples). For conjunctive queries and UCQs, this is not the case. In particular, a fresh role name is required to deal with disconnected queries. We also observe that, in most cases, admitting auxiliary symbols enables the construction of fitting ontologies of polynomial size while, without such symbols, they are of single exponential size. The sole exception is the case of $\mathcal{ELI}$ and (U)CQs, where we do not obtain any bounds on the size of fitting ontologies at all.

Proofs are provided in the appendix of (Grosser and Lutz 2026).

	$\mathcal{EL}_{(\bot)}$	$\mathcal{ELI}_{(\bot)}$
AQ-entailment	PTIME	EXPTIME
(U)CQ-entailment	NP	EXPTIME
ontology fitting, consistency	PTIME	PTIME
ontology fitting, AQs	PTIME	PTIME
ontology fitting, (U)CQ	Σ_2^{P}	EXPTIME

Table 1: Complexities of query entailment (top) from (Baader, Brandt, and Lutz 2005; Rosati 2007; Baader, Brandt, and Lutz 2008; Eiter et al. 2008) and of ontology fitting from this article (bottom).

2 Preliminaries

2.1 Description Logic

Let $\mathsf{N_C}$, $\mathsf{N_R}$, and $\mathsf{N_I}$ be countably infinite sets of *concept names*, *role names*, and *individual names*. An *inverse role* takes the form r^- with r a role name, and a *role* is a role name or an inverse role. If $r = s^-$ is an inverse role, then we set $r^- = s$. An $\mathcal{ELI}_\bot$*-concept* C is built according to the syntax rule

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

where A ranges over concept names and r over roles. An $\mathcal{EL}_\bot$*-concept* is an $\mathcal{ELI}_\bot$-concept that does not use inverse roles, an $\mathcal{EL}$*-concept* is an $\mathcal{EL}_\bot$-concept that does not use '$\bot$', and likewise for $\mathcal{ELI}$*-concepts*. The *role depth* of a concept C, denoted $\mathsf{rd}(C)$, is the maximum nesting depth of the operator $\exists r.D$ in it. We shall occasionally also mention $\mathcal{ALC}$-concepts, which extend $\mathcal{EL}$-concepts with negation $\neg C$, and likewise for $\mathcal{ALCI}$ and $\mathcal{ELI}$.

Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot, \mathcal{ALC}, \mathcal{ALCI}\}$. An $\mathcal{L}$-*ontology* is a finite set of *concept inclusions (CIs)* $C \sqsubseteq D$, where C, D are $\mathcal{L}$-concepts. An *ABox* is a finite and nonempty set of *concept assertions* $A(a)$ and *role assertions* $r(a, b)$ where $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $a, b \in \mathsf{N_I}$. We use $\mathsf{ind}(\mathcal{A})$ to denote the set of individual names used in $\mathcal{A}$.

The semantics of concepts is defined as usual in terms of interpretations $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ with $\Delta^\mathcal{I}$ the (non-empty) *domain* and $\cdot^\mathcal{I}$ the *interpretation function*, see (Baader et al. 2017) for details. A *pointed interpretation* is a pair $(\mathcal{I}, d)$ with $\mathcal{I}$ an interpretation and $d \in \Delta^\mathcal{I}$. An interpretation $\mathcal{I}$ *satisfies* a CI $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$, a concept assertion $A(a)$ if $a \in A^\mathcal{I}$, and a role assertion $r(a, b)$ if $(a, b) \in r^\mathcal{I}$; we thus make the *standard names assumption*. We say that $\mathcal{I}$ is a *model* of an ontology $\mathcal{O}$, written $\mathcal{I} \models \mathcal{O}$, if it satisfies all CIs in it, and likewise for ABoxes. An ontology is *satisfiable* if it has a model and an ABox is *consistent* with an ontology $\mathcal{O}$ if $\mathcal{A}$ and $\mathcal{O}$ have a common model.

2.2 Simulations, Products, etc.

A *homomorphism* from an interpretation $\mathcal{I}_1$ to an interpretation $\mathcal{I}_2$ is a mapping $h\colon \Delta^{\mathcal{I}_1} \to \Delta^{\mathcal{I}_2}$ such that $d \in A^{\mathcal{I}_1}$ implies $h(d) \in A^{\mathcal{I}_2}$ and $(d, e) \in r^{\mathcal{I}_1}$ implies $(h(d), h(e)) \in r^{\mathcal{I}_2}$ for all $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and $d, e \in \Delta^{\mathcal{I}_1}$. We write $(\mathcal{I}_1, d_1) \to (\mathcal{I}_2, d_2)$ to denote the existence of a homomorphism h from $\mathcal{I}_1$ to $\mathcal{I}_2$ with $h(d_1) = d_2$.

For $k \in \mathbb{N} \cup \{\omega\}$, a k-$\mathcal{EL}$*-simulation* from $\mathcal{I}_1$ to $\mathcal{I}_2$ is a relation $S \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2} \times \{i \in \mathbb{N} \mid i \leq k\}$ such that for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$:

(Atom) if $d_1 \in A^{\mathcal{I}_1}$ and $(d_1, d_2, i) \in S$, then $d_2 \in A^{\mathcal{I}_2}$;

(Rel) if $(d_1, e_1) \in r^{\mathcal{I}_1}$ and $(d_1, d_2, i) \in S$ with $i \geq 1$, then there exists $(d_2, e_2) \in r^{\mathcal{I}_2}$ with $(e_1, e_2, i-1) \in S$.[1]

[1] here we take $\omega - 1$ to mean ω

An $\mathcal{ELI}$*-simulation* is defined in the same way, but is required to satisfy the additional condition

(iRel) if $(e_1, d_1) \in r^{\mathcal{I}_1}$ and $(d_1, d_2, i) \in S$ with $i \geq 1$, then there exists $(e_2, d_2) \in r^{\mathcal{I}_2}$ with $(e_1, e_2, i-1) \in S$.

k-$\mathcal{EL}_\bot$- and k-$\mathcal{ELI}_\bot$*-simulations* S are defined in the same way, but additionally need to be total:

(Tot) for all $d_1 \in \Delta^{\mathcal{I}_1}$, there is a $d_2 \in \Delta^{\mathcal{I}_2}$ with $(d_1, d_2, k) \in S$.

Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}, \mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$. For $k \in \mathbb{N} \cup \{\omega\}$, $d_1 \in \Delta^{\mathcal{I}_1}$, and $d_2 \in \Delta^{\mathcal{I}_2}$, we write $(\mathcal{I}_1, d_1) \preceq^k_{\mathcal{L}} (\mathcal{I}_2, d_2)$ if there is a k-$\mathcal{L}$-simulation S from $\mathcal{I}_1$ to $\mathcal{I}_2$ with $(d_1, d_2, k) \in S$.

An $\mathcal{L}$*-simulation* from $\mathcal{I}_1$ to $\mathcal{I}_2$ is a relation $S \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ such that $\{(d, e, \omega) \mid (d, e) \in S\}$ is an ω-$\mathcal{L}$-simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$. We denote the existence of an $\mathcal{L}$-simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$ by $\mathcal{I}_1 \preceq_{\mathcal{L}} \mathcal{I}_2$. Note that for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, the empty relation is an $\mathcal{L}$-simulation from $\mathcal{I}_1$ to $\mathcal{I}_2$, for all interpretations $\mathcal{I}_1, \mathcal{I}_2$. This is not the case for $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$, where $\mathcal{L}$-simulations have to be total. We write $(\mathcal{I}_1, d_1) \preceq_{\mathcal{L}} (\mathcal{I}_2, d_2)$ if there is an $\mathcal{L}$-simulation S from $\mathcal{I}_1$ to $\mathcal{I}_2$ that contains (d_1, d_2). We shall often exploit that, clearly, every homomorphism is an $\mathcal{ELI}_\bot$-simulation.

We state the basic and well-known connection between DL concepts and k-simulations.

Lemma 1. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, $k \geq 0$, and C an $\mathcal{L}$-concept with $\mathsf{rd}(C) \leq k$. Then $d_1 \in C^{\mathcal{I}_1}$ and $(\mathcal{I}_1, d_1) \preceq^k_{\mathcal{L}} (\mathcal{I}_2, d_2)$ implies $d_2 \in C^{\mathcal{I}_2}$.*

The next lemma asserts the existence of characteristic concepts, a construction is in the appendix.

Lemma 2. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, let $\mathcal{I}$ be an interpretation that interprets only finitely many role and concept names as non-empty, $d \in \Delta^{\mathcal{I}}$, and $k \geq 0$. Then there exists a* characteristic $\mathcal{L}$-concept $C^{\mathcal{L},k}_{\mathcal{I},d}$ of role depth k for d in $\mathcal{I}$ *such that for all interpretations $\mathcal{J}$ and $d' \in \Delta^{\mathcal{J}}$:*

$$(\mathcal{I}, d) \preceq^k_{\mathcal{L}} (\mathcal{J}, d') \quad \textit{if and only if} \quad d' \in (C^{\mathcal{L},k}_{\mathcal{I},d})^{\mathcal{J}}.$$

The next lemma establishes a crucial link between bounded and unbounded simulations. For a proof see Lemma 15 in (Hosemann et al. 2025).

Lemma 3. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$ and $\mathcal{I}, \mathcal{J}$ be finite interpretations. Then*

$$(\mathcal{I}, d) \preceq_{\mathcal{L}} (\mathcal{J}, e) \textit{ if and only if } (\mathcal{I}, d) \preceq^k_{\mathcal{L}} (\mathcal{J}, e)$$

for all $d \in \Delta^{\mathcal{I}}$, $e \in \Delta^{\mathcal{J}}$ and $k = |\Delta^{\mathcal{I}}| \cdot |\Delta^{\mathcal{J}}|$.

While it is NP-complete to decide whether there is a homomorphism from one given finite interpretation to another, the corresponding problem for simulations can be solved in PTIME, see e.g. (Henzinger, Henzinger, and Kopke 1995).

Lemma 4. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Given two finite pointed interpretations $(\mathcal{I}_1, d_1)$ and $(\mathcal{I}_2, d_2)$, it can be decided in* PTIME *whether $(\mathcal{I}_1, d_1) \preceq_{\mathcal{L}} (\mathcal{I}_2, d_2)$.*

We will sometimes also consider simulations from an ABox to an ABox, or from an ABox to an interpretation. In such cases, we view an ABox $\mathcal{A}$ as an interpretation with $\Delta^{\mathcal{A}} = \mathsf{ind}(\mathcal{A})$, $r^{\mathcal{A}} = \{(a, b) \mid r(a, b) \in \mathcal{A}\}$, and $A^{\mathcal{A}} = \{a \mid A(a) \in \mathcal{A}\}$ for all $r \in \mathsf{N_R}$ and $A \in \mathsf{N_C}$.

The *direct product* of two pointed interpretations $(\mathcal{I}_1, d_1)$ and $(\mathcal{I}_2, d_2)$ is the pointed interpretation $(\mathcal{I}_1 \times \mathcal{I}_2, (d_1, d_2))$ where

$$\begin{aligned}
\Delta^{\mathcal{I}_1 \times \mathcal{I}_2} &= \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2} \\
A^{\mathcal{I}_1 \times \mathcal{I}_2} &= A^{\mathcal{I}_1} \times A^{\mathcal{I}_2} \quad \text{for all } A \in \mathsf{N_C} \\
r^{\mathcal{I}_1 \times \mathcal{I}_2} &= \{((d_1, d_2), (e_1, e_2)) \mid (d_1, e_1) \in r^{\mathcal{I}_1} \\
&\qquad \text{and } (d_2, e_2) \in r^{\mathcal{I}_2}\} \quad \text{for all } r \in \mathsf{N_R}.
\end{aligned}$$

The product construction extends naturally to non-empty finite sequences S of pointed instances. We denote the resulting pointed interpretation by $(\prod S, d_\Pi)$. By convention, $\prod \emptyset$ is the pointed interpretation $(\mathcal{I}, d)$ with $d = ()$, $\Delta^{\mathcal{I}} = \{d\} = A^{\mathcal{I}}$ for all $A \in \mathsf{N_C}$ and $r^{\mathcal{I}} = \{(d, d)\}$ for all $r \in \mathsf{N_R}$. We also refer to this $\mathcal{I}$ as *the maximal interpretation*. The following is well-known (Hell and Nesetril 2004).

Lemma 5. *Let $S = (\mathcal{I}_i, d_i)_{i \in I}$ be a sequence of pointed interpretations. Then*

1. *$(\prod S, d_\Pi) \to (\mathcal{I}, d)$ for all $(\mathcal{I}, d) \in S$;*
2. *for all $\bar{e} = (e_i)_{i \in I} \in \Delta^{\prod S}$ and all $\mathcal{ELI}$-concepts C: $\bar{e} \in C^{\prod S}$ iff $e_i \in C^{\mathcal{I}_i}$ for all $i \in I$;*
3. *If $(\mathcal{J}, e)$ is a pointed interpretation with $(\mathcal{J}, e) \to (\mathcal{I}, d)$ for all $(\mathcal{I}, d) \in S$, then $(\mathcal{J}, e) \to (\prod S, d_\Pi)$.*

Let $\mathcal{I}$ be an interpretation and $d \in \Delta^{\mathcal{I}}$. A *path* in $\mathcal{I}$ is a sequence $p = d_1 r_1 \cdots d_{n-1} r_{n-1} d_n$ of domain elements d_i from $\Delta^{\mathcal{I}}$ and roles r_i such that $(d_i, d_{i+1}) \in r_i^{\mathcal{I}}$ for $1 \leq i < n$. We say that the path *starts* at d_1 and use $\mathsf{tail}(p)$ to denote d_n. The $\mathcal{ELI}$*-unraveling* of $\mathcal{I}$ at d is the interpretation $\mathcal{I}_d^{\downarrow\mathcal{ELI}}$ defined as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}_d^{\downarrow\mathcal{ELI}}} &= \text{set of all paths in } \mathcal{I} \text{ starting at } d \\
A^{\mathcal{I}_d^{\downarrow\mathcal{ELI}}} &= \{p \mid \mathsf{tail}(p) \in A^{\mathcal{I}}\} \\
r^{\mathcal{I}_d^{\downarrow\mathcal{ELI}}} &= \{(p, p') \mid p' = pre \text{ for some } e\} \cup \\
&\quad\ \{(p', p) \mid p' = pr^- e \text{ for some } e\}.
\end{aligned}$$

The $\mathcal{EL}$*-unraveling* $\mathcal{I}_d^{\downarrow\mathcal{EL}}$ of $\mathcal{I}$ at d is defined in the same way, except that paths must not contain inverse roles. An $\mathcal{EL}_\bot$-unraveling is simply an $\mathcal{EL}$-unraveling, and likewise for $\mathcal{ELI}_\bot$-unravelings. The following properties are well-known.

Lemma 6. *Let $\mathcal{I}$ be an interpretation, $d \in \Delta^{\mathcal{I}}$, and $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. Then*

1. *$(\mathcal{I}_d^{\downarrow\mathcal{L}}, d) \to (\mathcal{I}, d)$ via the homomorphism $\pi \mapsto \mathsf{tail}(\pi)$;*
2. *$(\mathcal{I}, d) \preceq_{\mathcal{L}} (\mathcal{I}_d^{\downarrow\mathcal{L}}, d)$;*
3. *for all $\mathcal{L}$-concepts C and $p \in \Delta^{\mathcal{I}_d^{\downarrow\mathcal{L}}}$: $p \in C^{\mathcal{I}_d^{\downarrow\mathcal{L}}}$ iff $\mathsf{tail}(p) \in C^{\mathcal{I}}$.*

2.3 Queries

A *conjunctive query (CQ)* takes the form $q = \exists \overline{x}\, \varphi(\overline{x})$ where $\overline{x}$ is a tuple of variables and φ a conjunction of *atoms* $A(t)$ and $r(t, t')$, with $A \in \mathsf{N_C}$, $r \in \mathsf{N_R}$, and t, t' variables from $\overline{x}$ or individuals from $\mathsf{N_I}$. With $\mathsf{var}(q)$ and $\mathsf{ind}(q)$, we denote the set of variables and individual names in $\overline{x}$. We take the liberty to view q as a set of atoms, writing e.g. $\alpha \in q$ to indicate that α is an atom in q. We may also write $r^-(x,y) \in q$ in place of $r(y,x) \in q$. An *atomic query (AQ)* is a CQ of the simple form $A(a)$, with A a concept name and a an individual. A *union of conjunctive queries (UCQ)* q is a disjunction of CQs. A CQ is *rooted* if every variable is reachable from an individual name in the undirected graph $G_q = (\mathsf{ind}(q) \cup \mathsf{var}(q), \{\{t,t'\} \mid r(t,t') \in q\})$. A UCQ is *rooted* if every CQ in it is.

A CQ q gives rise to an interpretation $\mathcal{I}_q$ with $\Delta^{\mathcal{I}_q} = \mathsf{ind}(q) \cup \mathsf{var}(q)$, $A^{\mathcal{I}_q} = \{t \mid A(t) \in q\}$, and $r^{\mathcal{I}_q} = \{(t,t') \mid r(t,t') \in q\}$ for all $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. With a homomorphism from a CQ q to an interpretation $\mathcal{I}$, we mean a homomorphism from $\mathcal{I}_q$ to $\mathcal{I}$ that is the identity on all individual names. For an interpretation $\mathcal{I}$ and a UCQ q, we write $\mathcal{I} \models q$ if there is a homomorphism h from a CQ in q to $\mathcal{I}$. For an ABox $\mathcal{A}$ and ontology $\mathcal{O}$, we write $\mathcal{A} \cup \mathcal{O} \models q$ if $\mathcal{I} \models q$ for all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{O}$.

We shall also consider queries that in essence are DL concepts. Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. A *rooted $\mathcal{L}$-query* q takes the form $C(a)$ with C an $\mathcal{L}$-concept and $a \in \mathsf{N_I}$. We write $\mathcal{A} \cup \mathcal{O} \models q$ if $a \in C^\mathcal{I}$ for all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{O}$. An *existential $\mathcal{L}$-query* q takes the form $\exists x\, C(x)$ with C an $\mathcal{L}$-concept. We set $\mathcal{A} \cup \mathcal{O} \models q$ if $C^\mathcal{I} \neq \emptyset$ for all models $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{O}$. When speaking only of an $\mathcal{L}$-query, we mean an $\mathcal{L}$-query that is either rooted or existential. There is also a finite version of $\mathcal{L}$-query entailment, denoted by '$\models^{\mathsf{fin}}$'. While the unrestricted and the finite version coincide for the DLs $\mathcal{L}$ listed above, this is not the case for some of their extensions introduced later on. It is easy to see that every $\mathcal{EL}$- and $\mathcal{ELI}$-query can be expressed as a (tree-shaped) CQ.

Note that all queries introduced above are Boolean, that is, they evaluate to true or false instead of producing answers. This is without loss of generality since we admit individual names in queries.

We use $||O||$ to denote the *size* of a syntactic object O such as an ontology or an ABox. It is defined as the length of the encoding of O as a word over a suitable alphabet.

2.4 ABox Examples and the Fitting Problem

Let $\mathcal{Q}$ be a query language such as $\mathcal{Q} = \text{AQ}$ or $\mathcal{Q} = \text{CQ}$. An *ABox-$\mathcal{Q}$ example* is a pair $(\mathcal{A}, q)$ with $\mathcal{A}$ an ABox and q a query from $\mathcal{Q}$ such that all individual names that appear in q are from $\mathsf{ind}(\mathcal{A})$.

By a *collection of labeled examples* we mean a pair $E = (E^+, E^-)$ of finite sets of examples. The examples in E^+ are the *positive examples* and the examples in E^- are the *negative examples*. We say that $\mathcal{O}$ *fits* E if $\mathcal{A} \cup \mathcal{O} \models q$ for all $(\mathcal{A}, q) \in E^+$ and $\mathcal{A} \cup \mathcal{O} \not\models q$ for all $(\mathcal{A}, q) \in E^-$. The following example illustrates this central notion.

Example 1. *Let $E = (E^+, E^-)$ be the collection of labeled ABox-CQ examples where E^+ contains the two examples*

$$(\ \{\ \mathsf{NewHire(jane)}\ \},\ \exists x\, \mathsf{mentor}(\mathsf{jane}, x) \wedge \mathsf{Emp}(x)\ )$$

$$(\ \{\ \mathsf{NewHire(alex)},\ \mathsf{RemoteWorker(alex)}\ \},\quad \exists x\ \mathsf{mentor}(\mathsf{alex}, x) \wedge \mathsf{SeniorEmp}(x)\ ).$$

and E^- contains the single example

$$(\{\ \mathsf{NewHire(bob)}\ \},\ \exists x\ \mathsf{mentor}(\mathsf{bob}, x) \wedge \mathsf{SeniorEmp}(x)\ ).$$

Then, the $\mathcal{EL}$-ontology $\mathcal{O}$ that contains the CIs

$$\mathsf{NewHire} \sqsubseteq \exists \mathsf{mentor.Emp}$$
$$\mathsf{NewHire} \sqcap \mathsf{RemoteWorker} \sqsubseteq \exists \mathsf{mentor.SeniorEmp}$$

fits all examples in E. If we further add the CI $\mathsf{SeniorEmp} \sqsubseteq \mathsf{Emp}$*, the ontology still fits.*

Let $\mathcal{L}$ be an ontology language, such as $\mathcal{L} = \mathcal{ELI}_\bot$, and $\mathcal{Q}$ a query language. Then *$(\mathcal{L},\mathcal{Q})$-ontology fitting* is the problem to decide, given as input a collection of labeled ABox-$\mathcal{Q}$ examples E, whether E admits a fitting $\mathcal{L}$-ontology. We generally assume that the ABoxes used in E have pairwise disjoint sets of individual names. It is not hard to verify that this is without loss of generality because consistently renaming individual names in a collection of examples has no impact on the existence of a fitting ontology.

There is a natural variation of $(\mathcal{L}, \mathcal{Q})$-ontology fitting where one additionally requires the fitting ontology to be consistent with all ABoxes that occur in examples. We then speak of *consistent $(\mathcal{L}, \mathcal{Q})$-ontology fitting*. The following observation from (Funk, Grosser, and Lutz 2025) shows that it suffices to design algorithms for $(\mathcal{L}, \mathcal{Q})$-ontology fitting as originally introduced.

Proposition 1. *Let $\mathcal{L}$ be any ontology language and $\mathcal{Q} \in \{AQ, CQ, UCQ\}$. Then there is a* PTIME*-reduction from consistent $(\mathcal{L}, \mathcal{Q})$-ontology fitting to $(\mathcal{L}, \mathcal{Q})$-ontology fitting.*

3 Consistency-Based Fitting

We start with a version of ontology fitting that is based on ABox consistency rather than on querying. Besides being natural in its own right, it also serves to showcase some central techniques used in this paper. Here, an example is simply an ABox, and an ontology $\mathcal{O}$ *fits* a collection $E = (E^+, E^-)$ if $\mathcal{A}$ is consistent with $\mathcal{O}$ for all $\mathcal{A} \in E^+$ and inconsistent with $\mathcal{O}$ for all $\mathcal{A} \in E^-$. We refer to the resulting decision problem as *consistent $\mathcal{L}$-ontology fitting*.

We only consider $\mathcal{EL}_\bot$ and $\mathcal{ELI}_\bot$ because the cases of $\mathcal{EL}$ and $\mathcal{ELI}$ are uninteresting. In fact, it is easy to see that a collection of labeled ABox examples E has a fitting $\mathcal{EL}$-ontology if and only if E does not contain negative examples, and the same is true for $\mathcal{ELI}$. The subsequent theorem provides a characterization of the existence of fitting $\mathcal{EL}_\bot$- and $\mathcal{ELI}_\bot$-ontologies.

Theorem 1. *Let $E = (E^+, E^-)$ be a collection of labeled ABox examples with $E^+ \neq \emptyset$ and $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$. Then the following are equivalent, where $\mathcal{A}^+ = \biguplus E^+$:*

1. *E admits a fitting $\mathcal{L}$-ontology;*
2. *$\mathcal{A} \not\preceq_\mathcal{L} \mathcal{A}^+$ for every $\mathcal{A} \in E^-$.*

The next example shows that the existence of a fitting $\mathcal{EL}_\bot$-ontology does not, in general, coincide with the existence of a fitting $\mathcal{ELI}_\bot$-ontology. It is interesting to contrast this with the case of consistency-based fitting for $\mathcal{ALC}$ and $\mathcal{ALCI}$, where the existence of a fitting ontology coincides (Funk, Grosser, and Lutz 2025).

Example 2. *Consider the collection of ABox examples $E = (E^+, E^-)$ where E^+ contains the single example*

$$\mathcal{A}_1 = \{r(a_1, b_1), r(a_2, b_2), A_1(a_1), A_2(a_2)\}$$

and E^- contains the single example

$$\mathcal{A}_2 = \{r(a_1, b), r(a_2, b), A_1(a_1), A_2(a_2)\}.$$

Then there is a fitting $\mathcal{ELI}_\bot$-ontology, e.g. the one that contains the single CI

$$\exists r^-.A_1 \sqcap \exists r^-.A_2 \sqsubseteq \bot.$$

In contrast, it follows from Theorem 1 that there is no fitting $\mathcal{EL}_\bot$-ontology. Just observe that $\{(a_i, a_i), (b, b_i) \mid i \in \{1, 2\}\}$ is an $\mathcal{EL}_\bot$-simulation from $\mathcal{A}_2$ to $\mathcal{A}_1$ (but not an $\mathcal{ELI}_\bot$-simulation).

We work towards a proof of Theorem 1. We give the following definition also for $\mathcal{EL}$ and $\mathcal{ELI}$ as it will be reused in subsequent sections.

Definition 1. *Let $\mathcal{A}$ be an ABox, $\mathcal{I}$ an interpretation, and S an $\mathcal{L}$-simulation from $\mathcal{A}$ to $\mathcal{I}$, for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Define interpretation $\mathcal{I}_0$ as*

$$\Delta^{\mathcal{I}_0} := \mathsf{ind}(\mathcal{A})$$
$$A^{\mathcal{I}_0} := \{a \mid (a, d) \in S \text{ implies } d \in A^{\mathcal{I}}\} \quad \text{for all } A \in \mathsf{N_C}$$
$$r^{\mathcal{I}_0} := \{(a, b) \mid r(a, b) \in \mathcal{A}\} \quad \text{for all } r \in \mathsf{N_R}.$$

For every $a \in \mathsf{ind}(\mathcal{A})$, let $\mathcal{J}_a$ be obtained from

$$\prod_{(a,d)\in S} (\mathcal{I}, d)$$

by renaming the distinguished element d_Π to a. Assume that this is the only element that $\mathcal{J}_a$ shares with $\mathcal{I}_0$ and that $\Delta^{\mathcal{J}_a} \cap \Delta^{\mathcal{J}_b} = \emptyset$ whenever $a \neq b$. The interpretation $\mathcal{I}_{\mathcal{A},S,\mathcal{L}}$ is obtained by taking the (non-disjoint) union of $\mathcal{I}_0$ and the unraveled interpretations $\mathcal{J}_a^{\downarrow\mathcal{L}}$, $a \in \mathsf{ind}(\mathcal{A})$.

Lemma 7. *Let $\mathcal{A}$ be an ABox, $\mathcal{I}$ an interpretation, $\mathcal{O}$ an $\mathcal{L}$-ontology and S an $\mathcal{L}$-simulation from $\mathcal{A}$ to $\mathcal{I}$, with $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Then $\mathcal{I} \models \mathcal{O}$ implies $\mathcal{I}_{\mathcal{A},S,\mathcal{L}} \models \mathcal{O}$.*

We are now ready to prove Theorem 1.

Proof. (of Theorem 1.) '1 ⇒ 2'. Assume that there is an $\mathcal{L}$-ontology $\mathcal{O}$ that fits E and that, to the contrary of what we have to show, there is an $\mathcal{A}_0 \in E^-$ such that $\mathcal{A}_0 \preceq_\mathcal{L} \mathcal{A}^+$ via some simulation S. Since $\mathcal{O}$ fits E, every $\mathcal{A} \in E^+$ is consistent with $\mathcal{O}$. Since $\mathcal{L}$-ontologies are invariant under disjoint union, this implies that also $\mathcal{A}^+$ is consistent with $\mathcal{O}$. Let $\mathcal{I}$ be a model of $\mathcal{A}^+$ and $\mathcal{O}$. Then clearly S is a simulation from $\mathcal{A}_0$ to $\mathcal{I}$. Consider the interpretation $\mathcal{I}_{\mathcal{A}_0,S,\mathcal{L}}$ which by construction is a model of $\mathcal{A}_0$. By Lemma 7, it is also a model of $\mathcal{O}$. This means that $\mathcal{A}_0$ is consistent with $\mathcal{O}$, in contradiction to $\mathcal{O}$ fitting E.

'2 ⇒ 1'. Assume that $\mathcal{A} \not\preceq_\mathcal{L} \mathcal{A}^+$ for every $\mathcal{A} \in E^-$. Note that since $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$, we are concerned with total simulations. This means that for every $\mathcal{A} \in E^-$ there is an $a_\mathcal{A} \in \mathsf{ind}(\mathcal{A})$ such that $(\mathcal{A}, a_\mathcal{A}) \not\preceq_\mathcal{L} (\mathcal{A}^+, b)$ for all $b \in \mathsf{ind}(\mathcal{A}^+)$. Lemma 3, gives

$$(\mathcal{A}, a_\mathcal{A}) \not\preceq^\ell_\mathcal{L} (\mathcal{A}^+, b), \tag{$*$}$$

where $\ell = |\mathcal{A}^+| \cdot \max(\{|\mathcal{A}| \mid \mathcal{A} \in E^-\})$. Now consider the $\mathcal{L}$-ontology

$$\mathcal{O} := \bigcup_{\mathcal{A}\in E^-} \{C^{\mathcal{L},\ell}_{\mathcal{A},a_\mathcal{A}} \sqsubseteq \bot\},$$

where $C^{\mathcal{L},\ell}_{\mathcal{A},a_\mathcal{A}}$ is the characteristic $\mathcal{L}$-concept of role depth ℓ for $a_\mathcal{A}$ in $\mathcal{A}$ viewed as an interpretation. We show that $\mathcal{O}$ fits E. Clearly, every $\mathcal{A} \in E^-$ is inconsistent w.r.t. $\mathcal{O}$, and thus $\mathcal{O}$ fits all negative examples. It remains to show that every $\mathcal{A} \in E^+$ is consistent w.r.t. $\mathcal{O}$. To this end, it clearly suffices to show that $\mathcal{A}^+$ is consistent w.r.t. $\mathcal{O}$. Consider $\mathcal{A}^+$ viewed as an interpretation, which trivially is a model of $\mathcal{A}^+$. We show that $\mathcal{A}^+$ is also a model of $\mathcal{O}$, that is, $(C^{\mathcal{L},\ell}_{\mathcal{A},a_\mathcal{A}})^{\mathcal{A}^+} = \emptyset$ for all $\mathcal{A} \in E^-$. In fact, this already follows from ($*$), because by choice of $C^\ell_{\mathcal{A},a_\mathcal{A}}$ the existence of some $b \in (C^{\mathcal{L},\ell}_{\mathcal{A},a_\mathcal{A}})^{\mathcal{A}^+}$ implies $(\mathcal{A}, a_\mathcal{A}) \preceq^\ell_\mathcal{L} (\mathcal{A}^+, b)$. ❑

In the following, we highlight some properties of the fitting ontologies constructed in the proof of Theorem 1 and discuss alternative ways of constructing them. For simplicity, we restrict our attention to $\mathcal{EL}_\bot$. However, all observations extend straightforwardly also to $\mathcal{ELI}_\bot$.

The fitting $\mathcal{EL}_\bot$-ontology $\mathcal{O}$ constructed for a collection of labeled ABox examples E in the proof of Theorem 1 is derived only from the negative examples E^-. It contains only concept and role names from E^-, thus no auxiliary concept or role names, that is, symbols that do not occur in E. $\mathcal{O}$ is of size single exponential in $|E|$, but can be reduced to polynomial size at the expense of introducing an auxiliary concept name $X^k_{\mathcal{A},a}$ for every $\mathcal{A} \in E^-$, $a \in \mathsf{ind}(\mathcal{A})$ and $k \in \{0, \dots, \ell\}$. The resulting ontology $\mathcal{O}'$ contains the following CIs for all $C^{\ell\mathcal{L}}_{\mathcal{A},a} \sqsubseteq \bot \in \mathcal{O}$ and $1 \leq k \leq \ell$:

$$\bigsqcap_{A(a)\in\mathcal{A}} A \sqcap \bigsqcap_{r(a,b)\in\mathcal{A}} \exists r.\, X^{k-1}_{\mathcal{A},b} \quad \sqsubseteq \quad X^k_{\mathcal{A},a}$$
$$\bigsqcap_{A(a)\in\mathcal{A}} A \sqsubseteq X^0_{\mathcal{A},a} \qquad X^\ell_{\mathcal{A},a_\mathcal{A}} \quad \sqsubseteq \quad \bot.$$

Note that the sole purpose of the auxiliary concept names is to represent the characteristic concepts $C^{\mathcal{L},\ell}_{\mathcal{A},a_\mathcal{A}}$ succinctly. In fact, $\mathcal{O}'$ is a conservative extension of $\mathcal{O}$ in the sense that every model of $\mathcal{O}'$ is a model of $\mathcal{O}$ and, conversely, every model of $\mathcal{O}$ can be extended to a model of $\mathcal{O}'$ by interpreting the auxiliary concept names. Consequently, $\mathcal{O}'$ fits E.

Interestingly, $\mathcal{EL}_\bot$-ontologies that fit E can also be constructed in a completely different way, based on the positive examples. This alternative construction, which follows ideas from (Funk, Grosser, and Lutz 2025), requires the use of auxiliary concept names, and yields an ontology of polynomial size. To define a fitting ontology $\mathcal{O}_\mathsf{alt}$ of this kind, we introduce an auxiliary concept name $\overline{V_a}$ for each individual $a \in \mathsf{ind}(\mathcal{A}^+)$. Informally, $\overline{V_a}$ is satisfied at an element d in an interpretation $\mathcal{I}$ if $(\mathcal{I}, d) \not\preceq_{\mathcal{EL}_\bot} (\mathcal{A}^+, a)$. Let Σ be the set of concept and role names used in E^-. The ontology $\mathcal{O}_\mathsf{alt}$ then contains the following CIs for all $a \in \mathsf{ind}(\mathcal{A}^+)$ and $A, r \in \Sigma$:

$$A \sqsubseteq \overline{V_a} \quad \text{if } A(a) \notin \mathcal{A}^+,$$
$$\exists r. \bigsqcap_{r(a,b)\in\mathcal{A}^+} \overline{V_b} \sqsubseteq \overline{V_a}, \qquad \bigsqcap_{b\in\mathsf{ind}(\mathcal{A}^+)} \overline{V_b} \sqsubseteq \bot.$$

Lemma 8. *If any $\mathcal{EL}_\bot$-ontology fits E, then $\mathcal{O}_\mathsf{alt}$ fits E.*

It is interesting to note that none of the constructed ontologies uses existential quantifiers on the right-hand side of concept inclusions.

By Lemma 4, Theorem 1 yields the following result.

Theorem 2. *Let $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$. The consistent $\mathcal{L}$-ontology fitting problem is in* PTIME.

4 Atomic Queries

We consider atomic queries and present a characterization that resembles the one for consistency-based fitting, but has an additional ingredient. Informally, each positive example $(\mathcal{A}, q)$ now acts like a certain kind of implication on the negative examples, similar to an existential rule with q as the (atomic unary) rule head and $\mathcal{A}$ as the rule body that, however, must be matched in terms of a simulation rather than a homomorphism. We now make this precise.

Let $E = (E^+, E^-)$ be a collection of labeled ABox-AQ-examples. Note that the case where $E^- = \emptyset$ is trivial, since then the ontology $\{\top \sqsubseteq A \mid (\mathcal{A}, A(a)) \in E^+\}$ always fits. In the following, we therefore assume that there is at least one negative example. Let $\mathcal{A}^- = \biguplus_{(\mathcal{A},Q(a))\in E^-} \mathcal{A}$. A *completion* of $\mathcal{A}^-$ is an ABox $\mathcal{C}$ that satisfies

$$\mathcal{A}^- \subseteq \mathcal{C} \subseteq \mathcal{A}^- \cup \{Q(b) \mid (\mathcal{A}, Q(a)) \in E^+, b \in \mathsf{ind}(\mathcal{A}^-)\}.$$

In contrast to consistency-based fitting, also $\mathcal{EL}$ and $\mathcal{ELI}$ without the $\bot$ concept are natural choices for the ontology language.

Theorem 3. *Let $E = (E^+, E^-)$ be a collection of labeled ABox-AQ-examples with $E^- \neq \emptyset$, and $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Then the following are equivalent:*

1. *E admits a fitting $\mathcal{L}$-ontology*
2. *there exists a completion $\mathcal{C}$ of $\mathcal{A}^-$ such that*
   (a) *$Q(a) \notin \mathcal{C}$ for each $(\mathcal{A}, Q(a)) \in E^-$;*
   (b) *if $(\mathcal{A}, Q(a)) \in E^+$ and $(\mathcal{A}, a) \preceq_\mathcal{L} (\mathcal{C}, b)$, then $Q(b) \in \mathcal{C}$.*

Point (b) of Theorem 3 formalizes what we mean when saying 'positive examples act as an implication'. It is not hard to find a variation of Example 2 which shows that, also in the AQ case, the existence of a fitting $\mathcal{EL}_\bot$-ontology does not coincide with the existence of a fitting $\mathcal{ELI}_\bot$-ontology, and likewise for $\mathcal{EL}$ and $\mathcal{ELI}$. As illustrated by the next example, the presence of the bottom concept makes a difference, too.

Example 3. *Consider the collection of ABox examples $E = (E^+, E^-)$ where E^+ contains the single example $(\{A_1(a), B(b)\}), A_2(a))$ and E^- contains the single example $(\{A_1(a)\}, A_2(a))$. Then $\{B \sqsubseteq \bot\}$ is a fitting $\mathcal{EL}_\bot$-ontology, but there is no fitting $\mathcal{ELI}$-ontology. In fact, for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, every completion $\mathcal{C}$ of $\mathcal{A}^-$ that satisfies Point (b) of Theorem 3 must satisfy $A_2(a) \in \mathcal{C}$ and thus violate Point (a).*

The discussion about different ways to construct fitting ontologies as well as their size and use of auxiliary symbols given in Section 3 also applies here. One subtlety is that the positive and negative examples swap rôles. In particular, the fitting ontology constructed in the proof of Theorem 3, which parallels the proof of Theorem 1, is derived from the positive examples and does neither use auxiliary symbols nor existential quantifiers on the right-hand side of CIs. It is of single exponential size and can be made polynomial at the expense of introducing auxiliary symbols. Again, there is an alternative construction that, in the case of AQs, starts from a completion $\mathcal{C}$ of $\mathcal{A}^-$ that satisfies Conditions (a) and (b) from Theorem 3. For simplicity, we restrict our attention to $\mathcal{EL}_\bot$. Let Σ be the set of concept and role names used in E^+. The ontology $\mathcal{O}_{\mathsf{alt}}$ then contains the following CIs for all $a \in \mathsf{ind}(\mathcal{A}^+)$ and $A, r \in \Sigma$:

$$A \sqsubseteq \overline{V_a} \quad \text{if } A(a) \notin \mathcal{C},$$

$$\exists r. \bigsqcap_{r(a,b)\in\mathcal{C}} \overline{V_b} \sqsubseteq \overline{V_a}, \qquad \bigsqcap_{b\in\mathsf{ind}(\mathcal{C})} \overline{V_b} \sqsubseteq \bot$$

$$\bigsqcap_{a\in\mathsf{ind}(\mathcal{C}), A(a)\notin\mathcal{C}} \overline{V_a} \sqsubseteq A \quad \text{if } (\mathcal{A}, A(a)) \in E^+.$$

Using a naive implementation of Theorem 3 to decide $(\mathcal{L}, \text{AQ})$-ontology fitting involves guessing a completion $\mathcal{C}$ of $\mathcal{A}^-$ and thus results in an NP upper bound. However, instead of guessing $\mathcal{C}$ it is possible to identify a unique candidate for $\mathcal{C}$ in polynomial time, thus improving the upper bound to PTIME. This parallels the development for $\mathcal{ALC}$ and $\mathcal{ALCI}$ in (Funk, Grosser, and Lutz 2025).

Definition 2. *Let $E = (E^+, E^-)$ be a collection of labeled ABox-AQ examples with $E^- \neq \emptyset$. We use $\mathcal{A}^\pm$ to denote the ABox that can be obtained from $\mathcal{A}^- = \biguplus_{(\mathcal{A},Q(a))\in E^-} \mathcal{A}$ by exhaustively applying the following rule:*

(R) *if $(\mathcal{A}, Q(a)) \in E^+$ and $(\mathcal{A}, a) \preceq_\mathcal{L} (\mathcal{A}^\pm, b)$, set $\mathcal{A}^\pm = \mathcal{A}^\pm \cup \{Q(b) \mid (a, b) \in S\}$.*

Based on Lemma 4, it is easy to see that $\mathcal{A}^\pm$ can be constructed in polynomial time. Moreover, it is not hard to show the following.

Proposition 2. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$ and $E = (E^+, E^-)$ be a collection of labeled ABox-AQ examples with $E^- \neq \emptyset$. Then the following are equivalent:*

1. *E admits no fitting $\mathcal{L}$-ontology;*
2. *$Q(a) \in \mathcal{A}^\pm$ for some $(\mathcal{A}, Q(a)) \in E^-$.*

Proposition 2 clearly yields the following.

Theorem 4. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Then $(\mathcal{L}, AQ)$-ontology fitting is in* PTIME.

5 UCQs and CQs

We start with characterizations of $(\mathcal{L}, \text{UCQ})$-ontology fitting, for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$, which also cover the case of CQs. These characterizations are more intricate than in the AQ case. In particular, the finite completions used for the AQ characterizations have to be replaced with potentially infinite ones, which we choose to represent as interpretations. The transition to infinite completions is connected to the fact that the fitting ontologies constructed in the AQ-case do not use existential quantifiers on the right-hand side of CIs while this cannot be achieved for CQs and UCQs. Example 1 may serve as an illustration.

Since already for AQs the existence of a fitting ontology does not coincide for any two of our four languages $\mathcal{L} \in$

$\{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$, the same is clearly true for CQs and UCQs. In the following characterization, we assume that there is at least one negative example. Note that the case without negative examples is again not too interesting: for $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$, $\{\top \sqsubseteq \bot\}$ is a fitting ontology; for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$, either the ontology $\{\top \sqsubseteq A \mid A \in \mathsf{N_C}$ occurs in $E\} \cup \{\top \sqsubseteq \exists r.\top \mid r \in \mathsf{N_R}$ occurs in $E\}$ fits or there is no fitting ontology.

Theorem 5. *Let $E = (E^+, E^-)$ be a collection of labeled ABox-UCQ examples with $E^- \neq \emptyset$ and $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Then the following are equivalent:*

1. *there is an $\mathcal{L}$-ontology that fits E;*
2. *there exists a finite interpretation $\mathcal{I}$ such that*
   (a) *$\mathcal{I} = \biguplus_{e \in E^-} \mathcal{I}_e$ where, for each $e = (\mathcal{A}, q) \in E^-$, $\mathcal{I}_e$ is a model of $\mathcal{A}$ with $\mathcal{I}_e \not\models q$;*
   (b) *for all $(\mathcal{A}, q) \in E^+$: if S is an $\mathcal{L}$-simulation from $\mathcal{A}$ to $\mathcal{I}$, then $\mathcal{I}_{\mathcal{A},S,\mathcal{L}} \models q$.*

The subsequent example illustrates Theorem 5.

Example 4. *Consider the collection $E = (E^+, E^-)$ where*

$$\begin{aligned} E^+ &= \{ (\{A(a)\}, \exists x\, r(a,x) \wedge A(x)), \\ &\qquad (\{r(a,a)\}, B(a)) \} \\ E^- &= \{ (\{A(a)\}, B(a)) \}. \end{aligned}$$

E does not admit a fitting $\mathcal{EL}$-ontology. Indeed, any fitting $\mathcal{EL}$-ontology $\mathcal{O}$ must satisfy $\mathcal{O} \models A \sqsubseteq \exists r.A$ because of the first positive example. Thus in every model $\mathcal{I}$ of $\mathcal{O}$ and for every $d \in A^\mathcal{I}$, we have $(\mathcal{A}_2, a) \preceq_{\mathcal{EL}} (\mathcal{I}, d)$ with $\mathcal{A}_2 = \{r(a,a)\}$ the ABox from the second positive example. It can be seen that this implies $d \in B^\mathcal{I}$; c.f. Point (b) of Theorem 5. It follows that $\mathcal{O} \models A \sqsubseteq B$ and consequently $\mathcal{O}$ violates the negative example.

In contrast, the $\mathcal{ELI}$-ontology $\{A \sqsubseteq \exists r.A,\ \exists r^-.\top \sqsubseteq B\}$ fits E. Note in particular that, for a model $\mathcal{I}$ of $\mathcal{O}$ and $d \in A^\mathcal{I}$, $(\mathcal{A}_2, a) \preceq_{\mathcal{ELI}} (\mathcal{I}, d)$ needs not *hold because there may not be an infinite outgoing r^--path from d in $\mathcal{I}$.*

To further discuss Theorem 5, we compare it to the characterizations for $(\mathcal{ALC}, \text{UCQ})$ and $(\mathcal{ALCI}, \text{UCQ})$ from (Funk, Grosser, and Lutz 2025). They differ in three aspects. First, the characterizations for $\mathcal{ALC}$ and $\mathcal{ALCI}$ use homomorphisms in place of simulations. Second, our definition of $\mathcal{I}_{\mathcal{A},S,\mathcal{L}}$ involves direct products while the one in (Funk, Grosser, and Lutz 2025) does not. And third, the characterizations for $\mathcal{ALC}$ and $\mathcal{ALCI}$ do not require the interpretation $\mathcal{I}$ in Point 2 to be finite. It is worthwhile to discuss these differences in some more detail.

The first difference arises because of the limited expressive power of $\mathcal{EL}$ and $\mathcal{ELI}$, which is tightly linked to simulations. It is interesting to note that the characterizations for $\mathcal{ALC}$ and $\mathcal{ALCI}$ are in terms of homomorphisms, *not* in terms of bisimulations, which are the counterpart of simulations for these DLs. This is related to 'second-order effects' which sometimes make it possible to express CQs (or in this case: ABoxes from positive examples) in $\mathcal{ALC}$ and $\mathcal{ALCI}$, see (Feier, Lutz, and Wolter 2018). The second difference is due to the fact that in contrast to homomorphisms, simulations are not functional, and thus we may have $(a, d_1), (a, d_2) \in S$ for distinct d_1, d_2, which we reconcile via products. The root of the third difference is the non-local nature of simulations. For instance, a reflexive loop must map homomorphically to a reflexive loop, but it may be simulated by an infinite path. The example below illustrates why this prevents us from using potentially infinite interpretations $\mathcal{I}$ in Theorem 5. We remark that the third difference is crucial only for the $\mathcal{ELI}$ case. For $\mathcal{EL}$, in contrast, it follows from the technical development in Section 5.2 that the qualifier 'finite' can be dropped from Point 2 of Theorem 5.[2]

Example 5. *Consider the collection E' obtained from the collection E in Example 4 by adding a third positive example*

$$(\{B(c), r(b,c)\}, B(b)),$$

which enforces that any fitting $\mathcal{ELI}$-ontology $\mathcal{O}$ must satisfy $\mathcal{O} \models \exists r.B \sqsubseteq B$. Then E' no longer admits a fitting $\mathcal{ELI}$-ontology. Informally, the reason is that $\mathcal{ELI}$-ontologies cannot capture the infinite nature of simulations. More precisely, it can be shown that for $\mathcal{O}$ to satisfy the second positive example of E, there must be a finite *subset $\mathcal{A}$ of the $\mathcal{ELI}$-unraveling of the ABox $\{r(a,a)\}$ from that example such that $\mathcal{A} \cup \mathcal{O} \models B(a)$. Now let $\mathcal{I}$ be a model of a fitting $\mathcal{ELI}$-ontology $\mathcal{O}$ and let $d \in A^\mathcal{I}$. Since $\mathcal{A}$ is finite and acyclic, there must be an element e on the infinite outgoing r-chain from d in $\mathcal{I}$ such that $(\mathcal{A}, a) \preceq_{\mathcal{ELI}} (\mathcal{I}, e)$. Thus we must have $e \in B^\mathcal{I}$. Since the additional positive example enforces $\mathcal{O} \models \exists r.B \sqsubseteq B$, this yields $d \in B^\mathcal{I}$, thus $\mathcal{O} \models A \sqsubseteq B$ and $\mathcal{O}$ violates the negative example.*

Let us connect this to Theorem 5. The infinite interpretation $\mathcal{I}$ that is an infinite r-path with root a and $A^\mathcal{I} = \Delta^\mathcal{I}$ satisfies Points (a) and (b) of that theorem despite the fact that there is no $\mathcal{ELI}$-ontology that fits E'. However, there is no finite interpretation that exhibits the same behaviour. Intuitively, the mismatch between the infinite nature of simulations and the finite nature of $\mathcal{ELI}$-ontologies only manifests on infinite interpretations while on finite interpretations it is mitigated by Lemma 3.

Up to some minor differences, the discussion about different ways to construct fitting ontologies given in Section 3 also applies to Theorem 5. Notably, any collection of examples that only contains rooted queries admits a fitting ontology without auxiliary symbols whenever it admits one at all, whereas non-rooted queries may force a fitting ontology to use an auxiliary role. In the case of $\mathcal{EL}_{(\bot)}$, using additional auxiliary symbols again allows us to construct fitting ontologies of polynomial size. For $\mathcal{ELI}_{(\bot)}$, the size of fitting ontologies remains open.

5.1 Decomposing Conjunctive Queries

To obtain decision procedures based on the characterizations given in Theorem 5, we first introduce (a variation of) some well-known technical machinery that revolves around forest-shaped models and the decomposition of a CQ into tree-shaped parts, going back to the *fork rewritings* of (Lutz 2008).

A model $\mathcal{I}$ of an ABox $\mathcal{A}$ is an *$\mathcal{EL}$-forest model* if

[2] See in particular the remark after Theorem 6.

1. the directed graph
$$G_\mathcal{I} = (\Delta^\mathcal{I}, \bigcup_{r \in \mathsf{N_R}} r^\mathcal{I} \setminus (\mathsf{ind}(\mathcal{A}) \times \mathsf{ind}(\mathcal{A})))$$
is a forest (a disjoint union of trees) and
2. $r^\mathcal{I} \cap (\mathsf{ind}(\mathcal{A}) \times \mathsf{ind}(\mathcal{A})) = \{(a,b) \mid r(a,b) \in \mathcal{A}\}$.

$\mathcal{ELI}$*-forest models* are defined likewise, but based on the undirected version of the graph $G_\mathcal{I}$ in Point 1. In other words, in $\mathcal{EL}$-forest models all edges in trees must point away from the root of the tree while this is not the case for $\mathcal{ELI}$-forest models. An $\mathcal{EL}_\bot$-forest model is simply an $\mathcal{EL}$-forest model, and likewise for $\mathcal{ELI}_\bot$. The following is well-known (Lutz 2008).

Lemma 9. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$, $\mathcal{O}$ be an $\mathcal{L}$-ontology, $\mathcal{A}$ an ABox, and q a UCQ. If $\mathcal{A} \cup \mathcal{O} \not\models q$, then there is an $\mathcal{L}$-forest model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{O}$ that satisfies $\mathcal{I} \not\models q$.*

Let $\mathcal{A}$ be an ABox and p a CQ. An $\mathcal{A}$*-variation* of p is a CQ p' that can be obtained from p by consistently replacing zero or more variables with individual names from $\mathsf{ind}(\mathcal{A})$ and possibly identifying variables. We say that p' is $\mathcal{L}$*-forest*, for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$, if the following conditions are satisfied:

1. if $r(a,b) \in p'$ with $a, b \in \mathsf{N_I}$, then $r(a,b) \in \mathcal{A}$;
2. $\mathcal{I}_{p'}$ is an $\mathcal{L}$-forest model of the ABox $\mathcal{A} \cap p'$.

Informally, it follows from the above definition that every connected component $\widehat{p}$ of $p' \setminus \mathcal{A}$ is a tree in the directed sense if $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$ and in the undirected sense otherwise. Consequently, $\widehat{p}$ can be viewed as an $\mathcal{L}$-concept $C_{\widehat{p}}$. Some of the trees $\widehat{p}$ may have an individual name a as their root and thus give rise to a rooted $\mathcal{L}$-query $C_{\widehat{p}}(a)$. Other trees may not contain individual names and thus give rise to an existential $\mathcal{L}$-query $\exists x\, C_{\widehat{p}}(x)$.[3] We call these $\mathcal{L}$-queries the *reduction queries* of p' and denote them by $\mathsf{red}(p')$.

$\mathcal{A}$-variations are tightly connected to forest models. Indeed a homomorphism h from a CQ q into an $\mathcal{L}$-forest model of some ABox $\mathcal{A}$ induces an $\mathcal{L}$-forest $\mathcal{A}$-variation in an obvious way: for all variables x and y, (i) replace x with a if $h(x) = a \in \mathsf{ind}(\mathcal{A})$ and then (ii) identify any remaining variables x and y if $h(x) = h(y)$.

Lemma 10. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$, $\mathcal{A}$ be an ABox, $\mathcal{I}$ a model of $\mathcal{A}$, and p a CQ. Then*

1. *If $\mathcal{I}$ is an $\mathcal{L}$-forest model of $\mathcal{A}$, then $\mathcal{I} \models p$ if and only if there exists an $\mathcal{L}$-forest $\mathcal{A}$-variation p' of p such that $\mathcal{I} \models p'$;*
2. *for every $\mathcal{L}$-forest $\mathcal{A}$-variation p' of p: $\mathcal{I} \models p'$ if and only if $\mathcal{I} \models q$ for all $q \in \mathsf{red}(p')$.*

5.2 Algorithms and Complexity: $\mathcal{EL}$ and $\mathcal{EL}_\bot$

We use Theorem 5 to develop decision procedures for $(\mathcal{L}, \text{UCQ})$-ontology fitting with $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$, and also prove matching lower bounds. Clearly, the same procedures can then also be used for $(\mathcal{L}, \text{CQ})$-ontology fitting. The general idea is to check the existence of a finite interpretation $\mathcal{I}$ that satisfies Points (a) and (b) of Theorem 5 by independently checking the existence of each finite interpretation $\mathcal{I}_e$ from Point (a). A challenge is posed by Point (b) due to the infinite nature of simulations illustrated by Example 5. We thus first identify a suitable reformulation of Theorem 5.

The reformulation of Theorem 5 uses ontologies formulated in the extension of $\mathcal{EL}_\bot$ and $\mathcal{ELI}_\bot$ with simulation quantifiers, as studied for $\mathcal{EL}_\bot$ in (Lutz, Piro, and Wolter 2010). Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{ELI}\}$. An $\mathcal{L}_\bot^{\mathsf{sim}}$-concept C is defined like an $\mathcal{L}_\bot$-concept, but additionally admits quantifiers of the form $\exists^{\mathsf{sim}}_\mathcal{L}(\mathcal{I}, d)$ where $\mathcal{I}$ is a finite interpretation with $d \in \Delta^\mathcal{I}$ that interprets only finitely many role and concept names (all others are considered empty). An $\mathcal{L}_\bot^{\mathsf{sim}}$-ontology is then defined as expected. The semantics of the additional quantifier is defined by
$$(\exists^{\mathsf{sim}}_\mathcal{L}(\mathcal{J}, e))^\mathcal{I} = \{d \in \Delta^\mathcal{I} \mid (\mathcal{J}, e) \preceq_\mathcal{L} (\mathcal{I}, d)\}.$$

$\mathcal{L}_\bot^{\mathsf{sim}}$*-query entailment* is the problem to decide, given an $\mathcal{L}_\bot^{\mathsf{sim}}$-ontology $\mathcal{O}$, an ABox $\mathcal{A}$, and an $\mathcal{L}$-query q, whether $\mathcal{A} \cup \mathcal{O} \models q$. *Finite $\mathcal{L}$-query entailment* is defined likewise, based on '$\models^{\mathsf{fin}}$'. The following is a consequence of Theorems 10 and 11 of (Lutz, Piro, and Wolter 2010).

Theorem 6. *Finite $\mathcal{EL}_\bot^{\mathsf{sim}}$-query entailment is in* PTIME.

Actually, Theorem 6 is proved in (Lutz, Piro, and Wolter 2010) for the unrestricted case rather than for the finite case, but it is shown in the same article that the two cases coincide.

We now state the announced reformulation, which will enable us to implement $(\mathcal{L}, \text{UCQ})$-ontology fitting, for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$, in terms of finite $\mathcal{EL}_\bot^{\mathsf{sim}}$-query entailment.

Theorem 7. *Let $E = (E^+, E^-)$ be a collection of labeled ABox-UCQ examples with $E^- \neq \emptyset$ and $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. Then for each $e \in E^+$, there is a set Γ_e of $\mathcal{L}_\bot^{\mathsf{sim}}$-ontologies, each of size linear in $||e||$, such that the following are equivalent:*

1. *there is an $\mathcal{L}$-ontology that fits E;*
2. *there is a finite interpretation $\mathcal{I}$ such that*
   (a) *$\mathcal{I} = \biguplus_{e \in E^-} \mathcal{I}_e$ where, for each $e = (\mathcal{A}, q) \in E^-$, $\mathcal{I}_e$ is a model of $\mathcal{A}$ such that $\mathcal{I}_e \not\models p'$, for every $\mathcal{L}$-forest $\mathcal{A}$-variation p' of a CQ in q;*
   (b) *for all $e \in E^+$, there exists some $\mathcal{O}_e \in \Gamma_e$ such that $\mathcal{I} \models \mathcal{O}_e$.*

Proof. Let $e = (\mathcal{A}, q) \in E^+$ and u be a fresh role name. The set Γ_e contains the following ontologies:

- for every $\mathcal{L}$-forest $\mathcal{A}$-variation p' of a CQ in q and every way to choose an individual $a_{q'} \in \mathsf{ind}(\mathcal{A})$ for every existential reduction query q' of p', the ontology $\mathcal{O}$ that contains the following CIs:
$$\exists^{\mathsf{sim}}_\mathcal{L}(\mathcal{A}, a) \sqsubseteq C \quad \text{for every } C(a) \in \mathsf{red}(p');$$
$$\exists^{\mathsf{sim}}_\mathcal{L}(\mathcal{A}, a_{q'}) \sqsubseteq \exists u.C \quad \text{for every } q' = \exists x\, C(x) \in \mathsf{red}(p').$$
- if $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ELI}_\bot\}$, then for every $a \in \mathsf{ind}(\mathcal{A})$, the ontology $\mathcal{O}_a = \{\exists^{\mathsf{sim}}_\mathcal{L}(\mathcal{A}, a) \sqsubseteq \bot\}$.

It is shown in the appendix that Γ is as required. ❑

[3] For $\mathcal{L} = \mathcal{ELI}$, the root is not uniquely determined. We assume that it is chosen arbitrarily.

The new formulation of Point (b) is helpful in the following way. Before independently checking the existence of the interpretations $\mathcal{I}_e$ from Point (a) of Theorem 7, we choose an ontology $\mathcal{O}_e \in \Gamma_e$ for every $e \in E^+$ and make sure that each $\mathcal{I}_e$ satisfies all chosen $\mathcal{O}_e$. The existence check for a single $\mathcal{I}_e$ then amounts to a suitable collection of non-entailment checks of the form $\mathcal{A} \cup \bigcup_{e \in E^+} \mathcal{O}_e \not\models^{\mathsf{fin}} q'$.

Theorem 8. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$. Then $(\mathcal{L}, UCQ)$-ontology fitting is in Σ^{P}_2.*

To prove Theorem 8, let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$. We use the following algorithm. Given a collection $E = (E^+, E^-)$ of labeled ABox-UCQ examples with $E^- \neq \emptyset$, the algorithm first guesses, for each positive example $e \in E^+$, an ontology $\mathcal{O}_e$ from the set Γ_e. It then co-guesses, for every negative example $(\mathcal{A}, q) \in E^-$ and every CQ p in q, an $\mathcal{L}$-forest $\mathcal{A}$-variation p' of p and uses Theorem 6 to verify in PTIME that $\mathcal{A} \cup \bigcup_{e \in E^+} \mathcal{O}_e \not\models^{\mathsf{fin}} q'$ for some $q' \in \mathsf{red}(p')$ (of which there are at most $|\mathsf{ind}(\mathcal{A}) \cup \mathsf{Var}(q)|$ many). The following is proved in the appendix.

Lemma 11. *The algorithm accepts its input E if and only if there is an $\mathcal{L}$-ontology that fits E.*

The proof of Lemma 11 crucially exploits that $\mathcal{EL}^{\mathsf{sim}}_\bot$ is a Horn DL. More precisely, it is shown in (Lutz, Piro, and Wolter 2010) that for every ABox $\mathcal{A}$ and $\mathcal{EL}^{\mathsf{sim}}_\bot$-ontology $\mathcal{O}$, there is a finite *universal model* $\mathcal{I}_{\mathcal{A} \cup \mathcal{O}}$, that is, $\mathcal{I}_{\mathcal{A} \cup \mathcal{O}}$ is a model of $\mathcal{A}$ and $\mathcal{O}$ that admits an $\mathcal{L}$-simulation S into every model $\mathcal{J}$ of $\mathcal{A}$ and $\mathcal{O}$ such that S is the identity on $\mathsf{ind}(\mathcal{A})$.This universal model serves as the 'glue' that allows us to recover a single interpretation $\mathcal{I}_e$ from the different checks that our algorithm makes for every $\mathcal{A}$-variation p' of a negative example $e = (\mathcal{A}, q)$ and CQ p in q.

We also show that the complexity of our algorithm is optimal. The following theorem is proved by reduction from 2-COLORING-EXTENSION.

Theorem 9. *Let $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$. Then $(\mathcal{L}$, rooted CQ$)$-ontology fitting is Σ^{P}_2-hard.*

In summary, we have shown that for $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot\}$ and $\mathcal{Q} \in \{\text{CQ}, \text{UCQ}\}$, $(\mathcal{L}, \mathcal{Q})$-ontology fitting is Σ^{P}_2-complete, in the rooted and in the non-rooted case.

5.3 Algorithms and Complexity: $\mathcal{ELI}$ and $\mathcal{ELI}_\bot$

We now use Theorem 7 to develop decision procedures for $(\mathcal{L}, \text{UCQ})$-ontology fitting with $\mathcal{L} \in \{\mathcal{ELI}, \mathcal{ELI}_\bot\}$, and also prove matching lower bounds. To use Theorem 7, we have to work with $\mathcal{ELI}^{\mathsf{sim}}_\bot$-ontologies. However, in contrast to the case of $\mathcal{EL}^{\mathsf{sim}}_\bot$, we are not aware that this DL has ever been studied in the literature. We therefore study it here. A first observation is that, in contrast to $\mathcal{EL}^{\mathsf{sim}}_\bot$, finite and infinite query entailment no longer coincide.

Example 6. *Let the $\mathcal{ELI}^{\mathsf{sim}}_\bot$-ontology $\mathcal{O}$ contain the CIs*

$$A \sqsubseteq \exists r.A \qquad \exists^{\mathsf{sim}}_{\mathcal{ELI}}(\mathcal{I}, d) \sqsubseteq \bot$$

where $\mathcal{I}_{\circlearrowleft}$ is the interpretation with $\Delta^{\mathcal{I}_{\circlearrowleft}} = \{d\}$ and $r^{\mathcal{I}_{\circlearrowleft}} = \{(d, d)\}$. For $\mathcal{A} = \{A(a)\}$, we then have

$$\mathcal{A} \cup \mathcal{O} \not\models B(a) \text{ but } \mathcal{A} \cup \mathcal{O} \models^{\mathsf{fin}} B(a).$$

The above example is closely related to the failure of the finite model property for the two-way μ-calculus (Bojańczyk 2002). Our main objective is to establish the following.

Theorem 10. *$\mathcal{ELI}^{\mathsf{sim}}_\bot$-query entailment is* EXPTIME*-complete, both in the finite and in the unrestricted case.*

We only need finite entailment for our purposes, but clarifying the unrestricted case does not require extra effort. Before proving Theorem 10, we observe that it yields the following.

Theorem 11. *Let $\mathcal{L} \in \{\mathcal{ELI}, \mathcal{ELI}_\bot\}$. Then $(\mathcal{L}, UCQ)$-ontology fitting is in* EXPTIME.

The algorithm is essentially the same as for $\mathcal{EL}$ and $\mathcal{EL}_\bot$, except that we enumerate all possibilities instead of guessing or co-guessing. Given a collection $E = (E^+, E^+)$ of labeled ABox-UCQ examples with $E^- \neq \emptyset$, the algorithm uses an outer loop to iterate over all ways to choose for each positive example $e \in E^+$, an ontology $\mathcal{O}_e$ from the set Γ_e. In an inner loop, it iterates over all ways to choose, for every negative example $(\mathcal{A}, q) \in E^-$ and every CQ p in q, an $\mathcal{L}$-forest $\mathcal{A}$-variation p' of p. It then uses Theorem 10 to verify in EXPTIME that $\mathcal{A} \cup \bigcup_{e \in E^-} \mathcal{O}_e \not\models q'$ for some $q' \in \mathsf{red}(p')$. Clearly, we obtain an EXPTIME algorithm overall. The correctness proof is almost identical to that of Lemma 11, details are omitted.

We now return to the proof of Theorem 10. With *$\mathcal{ELI}$-subsumption w.r.t. $\mathcal{ELI}^{\mathsf{sim}}_\bot$-ontologies*, we mean the problem to decide, given two $\mathcal{ELI}_\bot$-concepts C and D and an $\mathcal{ELI}^{\mathsf{sim}}_\bot$-ontology $\mathcal{O}$, whether $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{O}$. If this is the case, then we write $\mathcal{O} \models C \sqsubseteq D$ and say that *C is subsumed by D w.r.t. $\mathcal{O}$*. There is also a finite version where only finite models $\mathcal{I}$ are considered, noted $\mathcal{O} \models^{\mathsf{fin}} C \sqsubseteq D$. The following is essentially a consequence of the results in (Vardi 1998) and (Bojańczyk 2002).

Theorem 12. *$\mathcal{ELI}_\bot$-subsumption w.r.t. $\mathcal{ELI}^{\mathsf{sim}}_\bot$-ontologies is* EXPTIME*-complete, both in the finite and in the unrestricted case.*

Proof. (sketch) The upper bound can be obtained as follows. Let $\mathcal{L}_\mu$ denote the two-way μ-calculus with simultaneous fixed points, see e.g. (Vardi 1998) for the conventional two-way μ-calculus and (Bradfield and Stirling 2007) for μ-calculi with simultaneous fixed points. Given $\mathcal{ELI}_\bot$-concepts C, D and an $\mathcal{ELI}^{\mathsf{sim}}_\bot$-ontology $\mathcal{O}$, one may construct in polynomial time $\mathcal{L}_\mu$-formulas $\varphi_C, \varphi_D, \varphi_O$ such that $\mathcal{O} \models^{(\mathsf{fin})} C \sqsubseteq D$ if and only if $\varphi_O \wedge \varphi_C \wedge \neg\varphi_D$ is (finitely) unsatisfiable It is folklore that the standard translation of an $\mathcal{L}_\mu$-formula without simultaneous fixed points into a two-way parity automaton given in (Vardi 1998) and also used in (Bojańczyk 2002) can be generalized to simultaneous fixed points while remaining in polynomial time. It remains to decide emptiness of the resulting automaton, which is possible in EXPTIME both in the unrestricted (Vardi 1998) and in the finite case (Bojańczyk 2002).

The lower bounds are inherited from subsumption in $\mathcal{ELI}$, which is EXPTIME complete. (Baader, Brandt, and Lutz 2008). This also covers the finite case since $\mathcal{ELI}$ enjoys the finite model property regarding subsumption. ❏

We next prove Theorem 10 using Theorem 12, thus completing the proof of . Note that $\mathcal{ELI}_\bot^{\mathsf{sim}}$-query entailment is formulated in terms of ABoxes while there are no ABoxes in the problem of $\mathcal{ELI}_\bot$-subsumption w.r.t. $\mathcal{ELI}_\bot^{\mathsf{sim}}$-ontologies. It is nevertheless possible to reduce the former to the latter in polynomial time. The central idea is to replace ABoxes by their approximations expressed through a simulation quantifier. This is summarized by the following lemma.

Lemma 12. *Let $\mathcal{L} \in \{\mathcal{ELI}, \mathcal{ELI}_\bot\}$, $\mathcal{A}$ be an ABox, and $\mathcal{O}$ an $\mathcal{L}_\bot^{\mathsf{sim}}$-ontology. Further let $\mathcal{O}_\mathcal{A}$ be the ontology that contains the following CIs, for all $a \in \mathsf{ind}(\mathcal{A})$:*

$$\top \sqsubseteq \exists u.\, S_{\mathcal{A},a} \qquad S_{\mathcal{A},a} \sqsubseteq \exists^{\mathsf{sim}}_{\mathcal{L}}(\mathcal{A}, a),$$

where u is a fresh role name and all $S_{\mathcal{A},a}$ are fresh concept names. Then the following hold for all $\mathcal{ELI}_\bot$-concepts C and $a \in \mathsf{ind}(\mathcal{A})$:

1. *$\mathcal{O} \cup \mathcal{O}_\mathcal{A} \models^{\mathsf{fin}} S_{\mathcal{A},a} \sqsubseteq C$ if and only if $\mathcal{A} \cup \mathcal{O} \models^{\mathsf{fin}} C(a)$;*
2. *$\mathcal{O} \cup \mathcal{O}_\mathcal{A} \cup \{C \sqsubseteq \bot\} \models^{\mathsf{fin}} S_{\mathcal{A},a} \sqsubseteq \bot$ for all $a \in \mathsf{ind}(\mathcal{A})$ if and only if $\mathcal{A} \cup \mathcal{O} \models^{\mathsf{fin}} \exists x\, C(x)$.*

and the same is true in the unrestricted case.

We also prove a matching lower bound, by reduction from the complement of AQ entailment in $\mathcal{ELI}$ which is ExpTime-hard (Baader, Brandt, and Lutz 2008).

Theorem 13. *Let $\mathcal{L} \in \{\mathcal{ELI}, \mathcal{ELI}_\bot\}$. The ($\mathcal{L}$, rooted CQ)-ontology fitting problem is* ExpTime*-hard.*

W have thus shown that for $\mathcal{L} \in \{\mathcal{ELI}, \mathcal{ELI}_\bot\}$ and $\mathcal{Q} \in \{\text{CQ}, \text{UCQ}\}$, $(\mathcal{L}, \mathcal{Q})$-ontology fitting is Σ_2^{P}-complete, in the rooted and in the non-rooted case.

6 Conclusion

We have obtained a relatively complete picture of ontology fitting problems for the most common Horn DLs based on ABoxes and queries. It would, however, be interesting to extend our study to other relevant Horn DLs including $\mathcal{ELO}$, Horn-$\mathcal{ALC}$, and Horn-$\mathcal{ALCI}$. Existential rule languages such as guarded existential rules would be another interesting subject to study.

One may also be interested in the case of full UCQs, that is, UCQs that do not use existentially quantified variables. We conjecture that it is not too difficult to extend the techniques that we have introduced for AQs to full UCQs, and that ($\mathcal{L}$, full UCQ)-ontology fitting is in PTime for all $\mathcal{L} \in \{\mathcal{EL}, \mathcal{EL}_\bot, \mathcal{ELI}, \mathcal{ELI}_\bot\}$. In the case of $\mathcal{ALC}$ and $\mathcal{ALCI}$, such an extension has been presented in (Funk, Grosser, and Lutz 2025).

It should be straightforward to adapt the characterizations and decidability results presented in this paper to the DLs $\mathcal{EL}_\bot^{\mathsf{sim}}$ and $\mathcal{ELI}_\bot^{\mathsf{sim}}$. An interesting nuance is that the qualifier 'finite' in Theorem 5 can then be dropped for both $\mathcal{EL}_\bot^{\mathsf{sim}}$ *and* $\mathcal{ELI}_\bot^{\mathsf{sim}}$.

Another question left open is the size of fitting ontologies in $(\mathcal{ELI}, \text{UCQ})$-ontology fitting, where we have obtained no bounds at all. In particular, it would be interesting to understand whether polynomial size fittings always exist if auxiliary symbols are admitted.

Acknowledgements

The third author was supported by DFG project LU 1417/4-1. The authors acknowledge the financial support by the Federal Ministry of Research, Technology and Space of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research 'Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig', project identification number: ScaDS.AI.

References

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press.

Baader, F.; Brandt, S.; and Lutz, C. 2005. Pushing the $\mathcal{EL}$ envelope. In *Proc. of IJCAI*, 364–369. Professional Book Center.

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

Barceló, P., and Romero, M. 2017. The complexity of reverse engineering problems for conjunctive queries. In *Proc. of ICDT*, volume 68 of *LIPIcs*, 7:1–7:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Bienvenu, M., and Ortiz, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Proc. of Reasoning Web*, volume 9203 of *LNCS*, 218–307. Springer.

Bojańczyk, M. 2002. Two-way alternating automata and finite models. In *Proc. of ICALP*, 833–844. Springer.

Bradfield, J. C., and Stirling, C. 2007. Modal mu-calculi. In *Handbook of Modal Logic*, volume 3. North-Holland. 721–756.

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

Cohen, S., and Weiss, Y. Y. 2016. The complexity of learning tree patterns from example graphs. *ACM Trans. Database Syst.* 41(2):14:1–14:44.

Cucala, D. T.; Grau, B. C.; and Horrocks, I. 2019. 15 years of consequence-based reasoning. In *Description Logic, Theory Combination, and All That - Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday*, volume 11560 of *LNCS*, 573–587. Springer.

Eiter, T.; Gottlob, G.; Ortiz, M.; and Simkus, M. 2008. Query answering in the description logic horn-$\mathcal{SHIQ}$. In *Proc. of JELIA*, volume 5293 of *LNCS*, 166–179. Springer.

Feier, C.; Lutz, C.; and Wolter, F. 2018. From conjunctive queries to instance queries in ontology-mediated querying. In *Proc. of IJCAI*. AAAI Press.

Funk, M.; Jung, J. C.; Lutz, C.; Pulcini, H.; and Wolter, F. 2019. Learning description logic concepts: When can positive and negative examples be separated? In *Proc. of IJCAI 2019*, 1682–1688. ijcai.org.

Funk, M.; Grosser, M.; and Lutz, C. 2025. Fitting description logic ontologies to ABox and query examples. In *Proc. of KR*.

Grosser, M., and Lutz, C. 2026. Fitting horn dl ontologies to abox and query examples: A tale of simulation quantifiers and finite models. arXiv. 2604.26976.

Hell, P., and Nesetril, J. 2004. *Graphs and homomorphisms*, volume 28 of *OLSMA*. Oxford University Press.

Henzinger, M. R.; Henzinger, T. A.; and Kopke, P. W. 1995. Computing simulations on finite and infinite graphs. In *Proc. of FOCS*, 453–462. IEEE Computer Society.

Hosemann, S.; Jung, J. C.; Lutz, C.; and Rudolph, S. 2025. Fitting Ontologies and Constraints to Relational Structures. In *Proc. of KR*, 407–416.

Jung, J. C.; Lutz, C.; Pulcini, H.; and Wolter, F. 2022. Logical separability of labeled data examples under ontologies. *Artif. Intell.* 313:103785.

Lehmann, J., and Hitzler, P. 2010. Concept learning in description logics using refinement operators. *Mach. Learn.* 78(1-2):203–250.

Lutz, C.; Piro, R.; and Wolter, F. 2010. Enriching $\mathcal{EL}$-concepts with greatest fixpoints. In *Proc. of ECAI*. IOS Press. 41–46.

Lutz, C. 2008. Two upper bounds for conjunctive query answering in $\mathcal{SHIQ}$. In *Proc. of DL*, volume 353 of *CEUR*. CEUR-WS.org.

Motik, B.; Grau, B. C.; Horrocks, I.; Wu, Z.; Fokoue, A.; and Lutz, C., eds. 2009. *OWL 2 Web Ontology Language Profiles (Second Edition)*. W3C Recommendation.

Rosati, R. 2007. On conjunctive query answering in $\mathcal{EL}$. In *Proc. of DL*, volume 250 of *CEUR*.

Shalev-Shwartz, S., and Ben-David, S. 2014. *Understanding Machine Learning - From Theory to Algorithms*. Cambridge University Press.

ten Cate, B.; Funk, M.; Jung, J. C.; and Lutz, C. 2023a. Fitting algorithms for conjunctive queries. *SIGMOD Rec.* 52(4):6–18.

ten Cate, B.; Funk, M.; Jung, J. C.; and Lutz, C. 2023b. SAT-based PAC learning of description logic concepts. In *Proc. of IJCAI*, 3347–3355. ijcai.org.

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

Xiao, G.; Calvanese, D.; Kontchakov, R.; Lembo, D.; Poggi, A.; Rosati, R.; and Zakharyaschev, M. 2018. Ontology-based data access: A survey. In *Proc. of IJCAI*, 5511–5519. ijcai.org.

Zloof, M. M. 1975. Query by example. In *Proc. of AFIPS*, volume 44, 431–438. AFIPS Press.