Can You Tell the Difference? Contrastive Explanations for ABox Entailments

Patrick Koopmann[1], **Yasir Mahmood**[2], **Axel-Cyrille Ngonga Ngomo**[2], **Balram Tiwari**[2]

[1]Knowledge in Artificial Intelligence, Vrije Universiteit Amsterdam, The Netherlands
[2]Data Science Group, Heinz Nixdorf Institute, Paderborn University, Germany
p.k.koopmann@vu.nl, {yasir.mahmood, axel.ngonga, balram.tiwari}@uni-paderborn.de,

Abstract

We introduce the notion of contrastive ABox explanations to answer questions of the type "Why is a an instance of C, but b is not?". While there are various approaches for explaining positive entailments (why is $C(a)$ entailed by the knowledge base) as well as missing entailments (why is $C(b)$ not entailed) in isolation, contrastive explanations consider both at the same time, which allows them to focus on the relevant commonalities and differences between a and b. We develop an appropriate notion of contrastive explanations for the special case of ABox reasoning with description logic ontologies, and analyze the computational complexity for different variants under different optimality criteria, considering lightweight as well as more expressive description logics. We implemented a first method for computing one variant of contrastive explanations, and evaluated it on generated problems for realistic knowledge bases.

Code — https://doi.org/10.5281/zenodo.17603219
Extended version — https://arxiv.org/abs/2511.11281

Introduction

A key advantage of knowledge representation systems is that they enable transparent and explainable decision-making. For example, with an ontology formalized in a description logic (DL) (Baader et al. 2017; Hitzler, Krötzsch, and Rudolph 2010) we can infer implicit information from data (then called an *ABox*) through logical reasoning, and all inferences are based on explicit statements in the ontology and data. However, due to the expressive power of DLs and the complexity of realistic ontologies, inferences obtained through reasoning may not always be immediately understandable. Consequently, in recent years, significant attention has been devoted to explaining *why* or *why not* something is entailed by a DL knowledge base (KB). The *why* question is typically answered through *justifications* (Schlobach, Cornet et al. 2003; Horridge 2011). For a KB $\mathcal{K}$ (consisting of ontology and ABox statements) and an entailed axiom α, a justification is a subset minimal $\mathcal{J} \subseteq \mathcal{K}$ such that $\mathcal{J} \models \alpha$. Other techniques for explaining *why* questions include *proofs* (Alrabbaa et al. 2022b) and *Craig interpolants* (Schlobach 2004). To answer a *why not* question, we can use *abductive reasoning* to determine what is missing in $\mathcal{K}$ to derive α (Elsenbroich, Kutz, and Sattler 2006; Peirce 1878). Research in this area for DLs encompasses *ABox abduction* (Del-Pinto and Schmidt 2019; Koopmann 2021), *TBox abduction* (Wei-Kleiner, Dragisic, and Lambrix 2014; Du, Wan, and Ma 2017; Haifani et al. 2022), *KB abduction* (Elsenbroich, Kutz, and Sattler 2006; Koopmann et al. 2020) and *concept abduction* (Bienvenu 2008), depending on the type of entailment to be explained.

If we query a KB for a set of objects, we may wonder why some object occurs in the answer but another does not. In this context, addressing the *why* and *why not* questions jointly can provide more clarity than considering them in isolation. To illustrate this, consider a simplified KB for a hiring process that determines which candidates are considered for a job interview. The KB uses a TBox with axioms

$$(a)\ \mathit{Qualified} \sqcap \exists \mathit{publishedAt}.\mathit{Journal} \sqsubseteq \mathit{Interviewed},$$
$$(b)\ \exists \mathit{leads}.\mathit{Group} \sqcup \exists \mathit{hasFunding}.\top \sqsubseteq \mathit{Qualified},$$
$$(c)\ \mathit{PostDoc} \sqcap \exists \mathit{leads}.\mathit{Group} \sqsubseteq \bot$$

stating that (a) someone who is qualified and has published at a journal gets interviewed, (b) someone who leads a group or has funding is qualified and (c) postdocs cannot lead groups. Further, we have an ABox with assertions

$$(1)\ \mathit{publishedAt}(\mathit{alice}, \mathit{aij}), \quad (2)\ \mathit{publishedAt}(\mathit{bob}, \mathit{aaai}),$$
$$(3)\ \mathit{Journal}(\mathit{aij}), \quad (4)\ \mathit{leads}(\mathit{alice}, \mathit{kr}), \quad (5)\ \mathit{Group}(\mathit{kr}),$$
$$(6)\ \mathit{hasFunding}(\mathit{alice}, \mathit{nsf}), \quad (7)\ \mathit{PostDoc}(\mathit{bob})$$

stating that (1) Alice published at AIJ, (2) Bob published at AAAI, (3) AIJ is a journal, (4–5) Alice leads the group KR, and (6) receives funding from the NSF, and (7) Bob is a Postdoc. This knowledge base entails $\mathit{Interviewed}(\mathit{alice})$, but not $\mathit{Interviewed}(\mathit{bob})$. We can explain why Alice was interviewed with an ABox justification, e.g. $\{(1), (3), (4), (5)\}$ ("*She published at the journal AIJ and leads the KR group*"). To explain why Bob is not interviewed, we may use ABox abduction and obtain an answer with a fresh individual e:

$$\{ \quad \mathit{Journal}(\mathit{aaai}), \qquad \mathit{hasFunding}(\mathit{bob}, e) \quad \}$$

("*If AAAI was a journal, and Bob received funding, he would have been interviewed.*") For the question "*Why was Alice interviewed, but not Bob?*", those explanations are not ideal, since they consider different reasons for being qualified (funding vs. leading a group). A better *contrastive explanation* would be: "*Alice's publication is at a journal and Bob's is not, and only Alice receives funding.*"

Formally, a contrastive explanation problem consists of a concept (*Interviewed*), a *fact* individual that is an instance of the concept (*alice*), and a *foil* individual that is not an instance (*bob*). Such contrastive ABox explanation problems are also motivated in the context of *concept learning* (Lehmann and Hitzler 2010; Funk et al. 2019; Heindorf et al. 2022), where the aim is to learn a concept from positive and negative examples. Contrastive explanations allow to explain the learned concepts in the light of a positive and a negative example.

The notion of contrastive explanations appears first in the work of Lipton (1990). The main theme of Lipton's work is to express an inquirer's *preference* or reflect their demand regarding the *context* in which an explanation is requested (e.g., explain why Bob was not interviewed *in the context of Alice*, who was interviewed). Contrastive explanations have since been considered for answer set programming (Eiter, Geibinger, and Oetsch 2023) with aim of explaining why some atoms are in an answer set instead of others. The idea has also been used to explain classification results of machine learning models (Dhurandhar et al. 2018; Ignatiev et al. 2020; Stepin et al. 2021; Miller 2021; Marques-Silva and Ignatiev 2022). A related concept are *counter-factual explanations* used in machine learning (Verma, Dickerson, and Hines 2020; Dandl et al. 2020). What these approaches have in common is that they look at similarities and differences at the same time, and use syntactic *patterns* to highlight the differences. We use a similar idea in the context of ABox reasoning by using *ABox patterns* which are instantiated differently for the fact and the foil.

Contributions Our notion of contrastive ABox explanations (CEs) is quite general, and even allows for contradictions with the KB. We distinguish between a syntactic and a semantic version, consider different optimality criteria and analyze them theoretically for different DLs ranging from light-weight $\mathcal{EL}$ to the more expressive $\mathcal{ALCI}$ (see Table 1 for an overview). Our contributions are three fold:

1. We introduce contrastive ABox explanations along with several variants and optimality criteria.
2. We characterize the complexity for various reasoning problems spanning five dimensions: variants, preference measures, types of optimality, DLs, and concept types.
3. We implemented a first practical method and evaluated it on realistic ontologies.

Description Logics

We recall the relevant DLs (Baader et al. 2017). Let $\mathsf{N_I}$, $\mathsf{N_C}$, and $\mathsf{N_R}$ denote countably infinite, mutually disjoint sets of *individual*, *concept* and *role names*, respectively. $\mathcal{ALCI}$ *concepts* are concept names or built following the syntax rules in Table 2, where R stands for a role name $r \in \mathsf{N_R}$ or its inverse r^-. We define further concepts as syntactic sugar: *top* $\top = A \sqcup \neg A$, *bottom* $\bot = A \sqcap \neg A$, *disjunction* $C \sqcup D = \neg(\neg C \sqcap \neg D)$ and *value restriction* $\forall r.C = \neg\exists r.\neg C$. An concept without inverse roles is in $\mathcal{ALC}$, if it furthermore only uses $\top, \sqcap, \exists$ and $\bot$ it is in $\mathcal{EL}_\bot$, and without $\bot$ it is in $\mathcal{EL}$.

optimality	fresh ind.	$\mathcal{EL}_\bot$ $\subseteq$ / $\leq$	$\mathcal{ALC}, \mathcal{ALCI}$ $\subseteq$ / $\leq$
diff-min		$\leq$ P[T11] / coNP-c[T12]	ExpTime-c[T11,T12]
conf-min	yes	ExpTime-c[T17]	coNExpTime-c[T17]
	no	coNP-c[T18]	ExpTime-c[T18]
com-max		open /coNP-c[T19]	ExpTime-c[T19]

Table 1: Complexity results for verifying minimality of CEs.

Construct	Syntax	Semantics
Conjunction	$C \sqcap D$	$C^\mathcal{I} \cap D^\mathcal{I}$
Existential restriction	$\exists R.C$	$\{x \mid \exists\, y \in C^\mathcal{I}, \langle x, y\rangle \in R^\mathcal{I}\}$
Negation	$\neg C$	$\Delta^\mathcal{I} \setminus C^\mathcal{I}$

Table 2: Syntax and semantics for $\mathcal{ALCI}$ concepts.

A *general concept inclusion* (GCI) is an expression of the form $C \sqsubseteq D$ for concepts C, D. A *TBox* is a finite set of GCIs. An *assertion* is an expression of the form $A(a)$ (*concept assertion*) or $r(a, b)$ (*role assertion*), where $a, b \in \mathsf{N_I}$, $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. An *ABox* is a finite set of assertions. A KB is a tuple $\langle\mathcal{T}, \mathcal{A}\rangle$ of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$, treated as $\mathcal{T} \cup \mathcal{A}$. GCIs and assertions are called *axioms*. A TBox/KB is in $\mathcal{ALCI}$/$\mathcal{ALC}$/$\mathcal{EL}_\bot$/$\mathcal{EL}$ if all concepts in it are.

The semantics of $\mathcal{ALCI}$ is defined in terms of interpretations. An *interpretation* $\mathcal{I}$ is a tuple $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$, where $\Delta^\mathcal{I}$ is a non-empty set called the *domain* of $\mathcal{I}$, and $\cdot^\mathcal{I}$ is the *interpretation function* that maps every individual name $a \in \mathsf{N_I}$ to an element $a^\mathcal{I} \in \Delta^\mathcal{I}$, every concept name $C \in N_C$ to a set $C^\mathcal{I} \subseteq \Delta^\mathcal{I}$, and every role name $r \in \mathsf{N_R}$ to a binary relation $r^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The interpretation function is extended to inverse roles using $(r^-)^\mathcal{I} = \{\langle x, y\rangle \mid \langle y, x\rangle \in r^\mathcal{I}\}$ and to concepts following Table 2.

Let $C \sqsubseteq D$ be a GCI and $\mathcal{I}$ be an interpretation. Then, $\mathcal{I}$ *satisfies* $C \sqsubseteq D$ (denoted by $\mathcal{I} \models C \sqsubseteq D$), if $C^\mathcal{I} \subseteq D^\mathcal{I}$. Similarly, $\mathcal{I}$ satisfies a concept assertion $A(a)$ if $a^\mathcal{I} \in A^\mathcal{I}$ and a role assertion $r(a, b)$ if $\langle a^\mathcal{I}, b^\mathcal{I}\rangle \in r^\mathcal{I}$. $\mathcal{I}$ is a *model* of $\mathcal{K}$ ($\mathcal{I} \models \mathcal{K}$), if $\mathcal{I}$ satisfies every axiom in $\mathcal{K}$. Finally, $\mathcal{K}$ *entails* α ($\mathcal{K} \models \alpha$) if $\mathcal{I} \models \alpha$ for every model $\mathcal{I}$ of $\mathcal{K}$. If $\mathcal{K} \models C(a)$, we call a an *instance of* C.

Contrastive Explanations

We are interested in *contrastive ABox explanation problems* (*CPs*), which are formally defined as tuples $P = \langle\mathcal{K}, C, a, b\rangle$ consisting of a KB $\mathcal{K}$, a concept C and two individual names a, b, s.t. $\mathcal{K} \models C(a)$ and $\mathcal{K} \not\models C(b)$. Intuitively, a CP reads as *"Why is a an instance of C and b is not?"*. If $\mathcal{K}$ and C are expressed in a DL $\mathcal{L}$, we call P an $\mathcal{L}$ *CP*. We call a (or more generally $C(a)$) the *fact* and b ($C(b)$) the *foil* of the CP. Note that, because $\mathcal{K} \not\models C(b)$, $\mathcal{K}$ is always consistent.

Building upon the framework of Lipton (1990), we aim to contrast a and b by highlighting the differences between the assertions that support $C(a)$ and the missing assertions that would support $C(b)$. Since different individuals may be related to a than to b, we abstract away from concrete individual names and instead use *ABox patterns*. An ABox pattern is a set $q(\vec{x})$ of assertions that uses variables from $\vec{x}$ instead of individual names. Given a vector $\vec{c}$ of individual names with the same length as $\vec{x}$, $q(\vec{c})$ then denotes the assertions obtained after replacing variables following $x_i \mapsto c_i$. We want to highlight the *difference* between a and b using an ABox pattern $q_{\mathit{diff}}(\vec{x})$, paired with two vectors $\vec{c}$ and $\vec{d}$ such that $q_{\mathit{diff}}(\vec{c})$ is entailed by the KB, and adding $q_{\mathit{diff}}(\vec{d})$ would entail $C(b)$. In our example, $q_{\mathit{diff}}(x, y, z) = \{\mathit{Journal}(y), \mathit{hasFunding}(x, z)\}$ would be such a pattern, where for the fact *alice* we have $\vec{c} = \langle \mathit{alice}, \mathit{aij}, \mathit{nsf}\rangle$, and for the foil *bob* we could use $\langle \mathit{bob}, \mathit{aaai}, e\rangle$, where e is fresh.

To fully explain the entailment, we also need to include other facts that are relevant to the entailment, and which fact and foil have in common. In our example, the explanation only makes sense together with the *commonality* $q_{\mathit{com}}(x, y, z) = \{\mathit{publishedAt}(x, y)\}$. Specifically, our contrastive explanations use ABox patterns $q(\vec{x}) = q_{\mathit{com}}(\vec{x}) \cup q_{\mathit{diff}}(\vec{x})$, with $q_{\mathit{com}}(\vec{x})$ stating what holds for both instantiations, and $q_{\mathit{diff}}(\vec{x})$ what holds only for the fact. To avoid irrelevant assertions in $q(\vec{x})$, we furthermore require that $q(\vec{c})$ is an ABox justification in the classical sense.

A final aspect regards how to deal with contradictions. Assume that in our example, instead of Axiom (b), we used (b') $\exists \mathit{leads}.\mathit{Group} \sqsubseteq \mathit{Qualified}$. Most notions of abduction require the hypothesis to be consistent with the KB, which in the present case, due to Axiom (c), is impossible if we want to entail *Interviewed*(*bob*). For the present example, we might however still want to provide an explanation, for instance: "*If AAAI was a journal and Bob lead the KR group, he would have been interviewed, but he cannot lead a group since he is a postdoc*". This results in the following components in the contrastive explanation:

$$q'_{\mathit{com}} = \{\ \mathit{publishedAt}(x, y), \mathit{Group}(z)\ \},$$
$$q'_{\mathit{diff}} = \{\ \mathit{Journal}(y), \mathit{leads}(x, z)\ \},$$
$$\vec{c'} = \langle \mathit{alice}, \mathit{aij}, \mathit{kr}\rangle, \qquad \vec{d'} = \{\mathit{bob}, \mathit{aaai}, \mathit{kr}\}$$

To point out the issue with this explanation, we add a final component, the *conflict set* $\mathcal{C}$, which in this case would be $\{\mathit{PostDoc}(\mathit{bob})\}$. Removing conflicts from the KB results in an alternative scenario consistent with the proposed explanation. This aligns with what are commonly called *counterfactual accounts* (Eiter, Geibinger, and Oetsch 2023). Intuitively, we would want to avoid conflicts if possible, but we will see later that this is not always desirable.

We can now formalise our new notion of explanations.

Definition 1. *Let $P = \langle \mathcal{K}, C, a, b\rangle$ be a CP where $\mathcal{K} = \langle \mathcal{T}, \mathcal{A}\rangle$. A solution to P (the* contrastive ABox explanation/CE*) is a tuple*

$$\langle q_{\mathit{com}}(\vec{x}), q_{\mathit{diff}}(\vec{x}), \vec{c}, \vec{d}, \mathcal{C}\rangle$$

of ABox patterns $q_{\mathit{com}}(\vec{x})$, $q_{\mathit{diff}}(\vec{x})$, vectors $\vec{c}$ and $\vec{d}$ of individual names, and a set $\mathcal{C}$ of assertions, which for $q(\vec{x}) = q_{\mathit{com}}(\vec{x}) \cup q_{\mathit{diff}}(\vec{x})$ satisfies the following conditions:

C1 *$\mathcal{T}, q(\vec{c}) \models C(a)$ and $\mathcal{T}, q(\vec{d}) \models C(b)$,*
C2 *$\mathcal{K} \models q(\vec{c})$,*
C3 *$\mathcal{K} \models q_{\mathit{com}}(\vec{d})$,*
C4 *$q(\vec{c})$ is a $\subseteq$-minimal set satisfying **C1+C2**,*
C5 *$\mathcal{C} \subseteq \mathcal{A}$ is $\subseteq$-minimal such that $\mathcal{T}, (\mathcal{A} \setminus \mathcal{C}) \cup q(\vec{d}) \not\models \bot$.*

We call $\vec{c}$ the *fact evidence* and $\vec{d}$ the *foil evidence*. The patterns $q_{\mathit{com}}(\vec{x})$ and $q_{\mathit{diff}}(\vec{x})$ will be called *commonality* and *difference*. Intuitively, $q(\vec{x})$ describes a pattern that is responsible for a being an instance of C, with $q_{\mathit{com}}(\vec{x})$ describing what a and b have in common, and $q_{\mathit{diff}}(\vec{x})$ what b is lacking. By instantiating $\vec{x}$ with $\vec{c}$ we obtain a set of entailed assertions that entail $C(a)$ (**C1** and **C2**), and by instantiating it with $\vec{d}$, we obtain a set of assertions that entails $C(b)$ (**C1**), where $q_{\mathit{com}}(\vec{d})$ is already provided by the present ABox (**C3**), and $q_{\mathit{diff}}(\vec{d})$ is missing. Since $q_{\mathit{diff}}(\vec{d})$ can be inconsistent with the KB, $\mathcal{C}$ presents the conflicts and $(\mathcal{A} \setminus \mathcal{C}) \cup q(\vec{d})$ depicts an alternative consistent scenario in which $C(b)$ is entailed (**C5**). To avoid unrelated assertions in q or $\mathcal{C}$, we require them to be minimal (**C4** and **C5**).

Example 2. *For our example, the CP is $\langle \mathcal{K}, \mathit{Interviewed}, \mathit{alice}, \mathit{bob}\rangle$. A CE for this CP is*

$$E_1 = \langle q_{\mathit{com}}(x, y, z),\ q_{\mathit{diff}}(x, y, z),\ \vec{c},\ \vec{d},\ \emptyset\rangle,$$

where $\vec{c} = \langle \mathit{alice}, \mathit{aij}, \mathit{nsf}\rangle, \vec{d} = \langle \mathit{bob}, \mathit{aaai}, e\rangle$,

$$q_{\mathit{com}}(x, y, z) = \{\ \mathit{publishedAt}(x, y)\ \}, \quad \text{and}$$
$$q_{\mathit{diff}}(x, y, z) = \{\ \mathit{Journal}(y),\ \mathit{hasFunding}(x, z)\ \}.$$

Another CE would be

$$E_2 = \langle q'_{\mathit{com}}, q'_{\mathit{diff}}, \vec{c'}, \vec{d'}, \{\mathit{PostDoc}(\mathit{bob})\}\rangle,$$

with q'_{com} and q'_{diff}, $\vec{c'}$ and $\vec{d'}$ as described above. ◁

Syntactic and Semantic CEs

In the example, our definition also allows for the following trivial CE that has limited explanatory value:

$$E_t = \langle \emptyset, \{\mathit{Interviewed}(x)\}, \langle \mathit{alice}\rangle, \langle \mathit{bob}\rangle, \emptyset\rangle.$$

A natural restriction to avoid this are *syntactic CEs*:

Definition 3 (Syntactic and Semantic CEs)**.** *Let $P = \langle \mathcal{K}, C, a, b\rangle$ be a CP where $\mathcal{K} = \langle \mathcal{T}, \mathcal{A}\rangle$. A CE*

$$E = \langle q_{\mathit{com}}(\vec{x}), q_{\mathit{diff}}(\vec{x}), \vec{c}, \vec{d}, \mathcal{C}\rangle$$

for P is called syntactic *if $q_{\mathit{com}}(\vec{c}), q_{\mathit{diff}}(\vec{c}), q_{\mathit{com}}(\vec{d}) \subseteq \mathcal{A}$, and otherwise* semantic*.*

Syntactic explanations can only refer to what is explicit in the ABox. Semantic explanations can additionally refer to implicit information that is entailed.

Example 4. *Consider the KB $\mathcal{K} = \langle \mathcal{T}, \mathcal{A}\rangle$ with*

$$\mathcal{T} = \{\ \mathit{Prof} \sqsubseteq \mathit{Qualified}, \mathit{Qualified} \sqcap \mathit{Nominee} \sqsubseteq \mathit{Offered}\ \}$$
$$\mathcal{A} = \{\ \mathit{Prof}(\mathit{alice}), \mathit{Nominee}(\mathit{alice}), \mathit{Qualified}(\mathit{bob})\ \}$$

A syntactic explanation for $\langle \mathcal{K}, \mathit{Offered}, a, b\rangle$ is

$$E_3 = \langle \emptyset,\ \{\mathit{Prof}(x),\ \mathit{Nominee}(x)\},\ \langle \mathit{alice}\rangle, \langle \mathit{bob}\rangle, \emptyset\rangle.$$

A semantic explanation can highlight the commonality:

$$E_4 = \langle\{\mathit{Qualified}(x)\}, \{\mathit{Nominee}(x)\}, \langle \mathit{alice}\rangle, \langle \mathit{bob}\rangle, \emptyset\rangle$$

and thus give a more precise explanation for why Bob was not offered the job. (He didn't need to be a professor.) ◁

If the CP contains a concept name as concept, a semantic explanation can always be obtained by simply using that concept as difference, which is why this case is more interesting for complex concepts. Moreover, we can reduce semantic CEs to syntactic ones:

Lemma 5. *Let $P = \langle\langle\mathcal{T}, \mathcal{A}\rangle, C, a, b\rangle$ be an $\mathcal{L}$ CP. Then, one can compute in polynomial time, with access to an oracle that decides entailment for $\mathcal{L}$, an ABox $\mathcal{A}_e$ such that every semantic CE for P is a syntactic CE for $P' = \langle\langle\mathcal{T}, \mathcal{A}_e\rangle, C, a, b\rangle$ and vice versa.*

Proof. We simply need to add all entailed assertions of the form $A(a)$/$r(a, b)$ where A, r, a and b occur in the input. □

Because of Lemma 5, we focus on syntactic CEs for most of the paper. Furthermore, some reasoning problems with semantic CEs are trivial for CPs involving concept names but intractable when complex concepts are considered.

Optimality Criteria

The minimality required in **C4** and **C5** is necessary to avoid unrelated assertions in the CE. Even with these restrictions in place, there can be many CEs for a given CP. The idea of CEs is to choose the ABox pattern so that the difference is as small as possible, and the commonality as large as possible. Also, while we allow for conflicts, having less seems intuitively better. There are therefore different components one may want to optimize, and optimization may be done locally (wrt. the subset relation) or globally (wrt. cardinality).

Definition 6 (Preferred CEs)**.** *Let P be a CP and $E = \langle q_{com}(\vec{x}), q_{diff}(\vec{x}), \vec{c}, \vec{d}, \mathcal{C}\rangle$ a CE for P. Then,*

- *E is* difference-minimal *if no explanation E' has difference $q'_{diff}(\vec{x}')$ and foil evidence $\vec{d}'$, s.t. $q'_{diff}(\vec{d}') \subset q_{diff}(\vec{d})$.*
- *E is* conflict-minimal *if there is no CE E' with a conflict set $\mathcal{C}' \subset \mathcal{C}$.*
- *E is* commonality-maximal *if no CE E' has commonality $q'_{com}(\vec{x}')$ and foil evidence $\vec{d}'$, s.t. $q_{com}(\vec{d}) \subset q'_{com}(\vec{d}')$.*

We define each of the aforementioned optimality also w.r.t. the cardinality of given sets.

Minimizing differences aligns with the general aim of CEs—the smaller the difference, the easier to understand the explanation. Minimizing conflicts allows to deprioritize far-fetched explanations that contradict much of what is known about the foil—if possible, we would want to provide a CE without conflicts. From a practical viewpoint, difference minimality allows the smallest factual change and conflict minimality limits CEs whose difference conflicts with the known data about foil. Commonality-maximality is similarly motivated, and allows to force the CE to be even more focussed. With commonality-maximality, we obtain interesting semantic CEs even when the concept in the CP is a concept name: in Example 4, both E_4 and the trivial CE using $q_{diff} = \{\mathit{Offered}(x)\}$ are difference-minimal, but E_4 is also commonality-maximal and explains the CP better.

In Example 2, E_1 is conflict-minimal and E_2 is commonality-maximal, whereas both are difference-minimal. The trivial E_t is conflict- and difference-minimal, but not commonality-maximal.

Decision Problems For any CP, we can always construct an arbitrary CE based on an ABox justification for the fact, where for the foil evidence, we simply replace a by b. Finding CEs that are also good wrt. our optimality criteria is less trivial. As usual, it is more convenient to look at decision problems rather than at the computation problem, in particular at the *verification problem*: Given a CP P with CE E, is E optimal wrt. a given criterion? Table 1 gives the complexity for these problems for different DLs. The global versions reduce to bounded versions of the existence problem: is there an E with a conflict/difference/commonality that has at most/least n elements? ExpTime-hardness for $\mathcal{ALC}$ follows in all cases by a reduction to entailment (see full version). We discuss the other results in the following sections.

Difference-Minimal Explanations

The challenge in computing and verifying difference-minimal CEs is that we cannot fix the other components: it is possible that the difference can only be made smaller by completely changing the other components. To deal with this, we define a maximal structure that intuitively contains all possible CEs, on which we then minimize the different components one after the other, starting with the difference. What it means for a structure to "contain" a CE is captured formally by the following definition. We call $\langle q_{com}(\vec{x}), q_{diff}(\vec{x}), \vec{c}, \vec{d}, \mathcal{C}\rangle$ a *candidate CE* if it satisfies Definition 1 except for **C4** and **C5**.

Definition 7. *Let $E_p = \langle p_{com}(\vec{x}), p_{diff}(\vec{x}), \vec{c}_p, \vec{d}_p, \mathcal{C}_p\rangle$ and $E_q = \langle q_{com}(\vec{y}), q_{diff}(\vec{y}), \vec{c}_q, \vec{d}_q, \mathcal{C}_q\rangle$ be two candidate CEs for a CP $P = \langle\mathcal{K}, C, a, b\rangle$. A* homomorphism *from E_p to E_q is a mapping $\sigma : \vec{x} \to \vec{y}$ that ensures $p_{com}(\sigma(\vec{x})) \subseteq q_{com}(\vec{y})$ and $p_{diff}(\sigma(\vec{x})) \subseteq q_{diff}(\vec{y})$. We say that E_p* embeds into *E_q if there is such a homomorphism and additionally $\mathcal{C}_p \subseteq \mathcal{C}_q$.*

Fix a CP $P = \langle\mathcal{K}, C, a, b\rangle$. We define the *CE superstructure* $E_m = \langle q_{com}(\vec{x}_m), q_{diff}(\vec{x}_m), \vec{c}_m, \vec{d}_m, \mathcal{A}\rangle$ for P as follows:

- $\vec{x}_m$ contains one variable $x_{a',b'}$ for every $\langle a', b'\rangle \in \mathsf{N_I}(\mathcal{A}) \times \mathsf{N_I}(\mathcal{A})$,
- $\vec{c}_m$ contains a' for every $x_{a',b'}$ in $\vec{x}$,
- $\vec{d}_m$ contains b' for every $x_{a',b'}$ in $\vec{x}$,
- $q_{com}(\vec{x}_m) = \{A(x_{a',b'}) \mid A(a'), A(b') \in \mathcal{A}\} \cup \{r(x_{a_0,b_0}, x_{a_1,b_1}) \mid r(a_0, a_1), r(b_0, b_1) \in \mathcal{A}\}$,
- $q_{diff}(\vec{x}_m) = \{A(x_{a',b'}) \mid A(a') \in \mathcal{A},\ b' \in \mathsf{N_I}(\mathcal{A}),\ A(b') \notin \mathcal{A}\} \cup \{r(x_{a_0,b_0}, x_{a_1,b_1}) \mid r(a_0, a_1) \in \mathcal{A},\ b_0, b_1 \in \mathsf{N_I}(\mathcal{A}), r(b_0, b_1) \notin \mathcal{A}\}$

We set $q_m = q_{com} \cup q_{diff}$. The variables $\vec{x}_m$ contain all possible combinations of mapping to an individual for the foil and for the fact, which is why they correspond to pairs of individual names. q_{com} and q_{diff} are then constructed based on the set of all assertions we can build over these variables. E_m indeed captures all syntactic CEs for P that are defined over the signature of the input.

Lemma 8. *Every syntactic CE for P without fresh individual names embeds into E_m.*

E_m is polynomial in size, and is *almost* a CE, modulo the minimality of $q(\vec{c})$ (**C4**) and of $\mathcal{C}_m$ (**C5**). Moreover, to satisfy **C5**, we also need $q_m(\vec{d}_m)$ to be consistent with $\mathcal{T}$, which may not be the case. To obtain a difference-minimal CE, we need to remove elements from q_m to make $\mathcal{T}, q_m(\vec{d}_m)$ consistent and $q_2(\vec{d}_m)$ subset-minimal without violating $\mathcal{T}, q_m(\vec{d}_m) \models C(b)$. The following lemma states how to safely remove elements from $q_m(\vec{d}_m)$. For $\vec{x} \subseteq \vec{x}_m$, denote by $q_m|_{\vec{x}}(\vec{x}_m)$ the restriction of $q_m(\vec{x}_m)$ to atoms that only use variables from $\vec{x}$.

Lemma 9. *Let $\vec{x} \subseteq \vec{x}_m$ be s.t 1) $x_{a,b} \in \vec{x}$, and 2) for every $x_{a',b'}$ in $\vec{x}_m$, we have some $x_{a'',b''} \in \vec{x}$ with $a' = a''$. Then, $\mathcal{K}, q_{diff}|_{\vec{x}}(\vec{d}_m) \models C(b)$.*

Lemma 9 tells us which atoms we should keep if we want to make sure the foil remains entailed. The following lemma tells us how to make $q_m(\vec{d}_m)$ consistent with $\mathcal{T}$:

Lemma 10. *Let $\vec{x} \subseteq \vec{x}_m$ be s.t. for every $x_{a_1,b_1}, x_{a_2,b_2} \in \vec{x}$, $a_1 \neq a_2$ implies $b_1 \neq b_2$. Then, $\mathcal{T}, q_m|_{\vec{x}}(\vec{d}_m) \not\models \bot$.*

Importantly, the conditions in Lemma 9 and 10 are compatible: we can always find a vector $\vec{x}$ that is *safe* in the sense that it satisfies the conditions in both lemmas, and thus ensures both consistency with $\mathcal{T}$ and entailment of $C(b)$. This can be used as follows to compute difference-minimal CEs, starting from the super-structure E_m.

P1 Obtain a query $q^{(0)} = q^{(0)}_{com} \cup q^{(0)}_{diff}$ from q_m by removing atoms until $\mathcal{T}, q^{(0)}(\vec{d}_m) \not\models \bot$. By doing so, ensure that for some safe vector $\vec{x}$, we have $q_m|_{\vec{x}} \subseteq q^{(0)}$. By Lemma 9, we then also have $\mathcal{K}, q^{(0)}_{diff}(\vec{d}_m) \models C(b)$.

P2 Compute a minimal subset $q_{diff}(\vec{x}_m)$ of $q^{(0)}_{diff}(\vec{x}_m)$ s.t. $\mathcal{T}, q^{(0)}_{com}(\vec{d}_m) \cup q_{diff}(\vec{d}_m) \models C(b)$. This can be done in polynomial time by checking each axiom in turn.

P3 To satisfy **C4**, compute a minimal subset $q_{com}(\vec{x}_m)$ of $q^{(0)}_{com}(\vec{x}_m)$ s.t. $\mathcal{T}, q_{com}(\vec{c}_m) \cup q_{diff}(\vec{c}_m) \models C(a)$. This also takes polynomial time by checking each axiom in turn.

P4 To satisfy **C5**, minimize $\mathcal{C}_m$ in the same way.

We obtain the following theorem.

Theorem 11. *For any DL $\mathcal{L}$-CPs, given an oracle that decides entailment in $\mathcal{L}$, we can 1) compute a difference-minimal syntactic CE, and 2) decide difference-minimality of a given syntactic CE in polynomial time.*

To see why 2) holds, let E be a syntactic CE. We may assume that E contains no fresh individuals, since we can always add occurrences of individuals to the KB in a way that does not affect relevant entailments. By Lemma 8, E then embeds in E_m. We now apply **P1** in the above procedure with the additional requirement that the pattern from E is also contained in $q^{(0)}$, and in **P2**, we try to construct a subset of the difference of E.

This establishes our results in Table 1 for local difference minimality. When minimizing the difference *globally*, we lose tractability, Indeed, it is NP-complete to decide whether a CE exists with the cardinality of the difference bounded by some $n \in \mathbb{N}$. This allows us to prove the following theorem.

Theorem 12. *Deciding whether a given syntactic or semantic CE for a given CP is difference-minimal w.r.t cardinality is* coNP*-complete for $\mathcal{EL}$ and $\mathcal{EL}_\bot$, and* ExpTime*-complete for $\mathcal{ALC}$.*

Conflict-Minimal Explanations

It seems natural to favor explanations with an empty or minimal conflict set. Unfortunately, it turns out that this makes computing CEs significantly harder, and may lead to explanations that are overall much more complex. The reason is that avoiding conflicts may require fresh individuals in the foil evidence:

Example 13. *Consider $P = \langle\langle\mathcal{T}, \mathcal{A}\rangle, C, a, b\rangle$ with*

$$\mathcal{T} = \{\ \exists r.\exists r.A \sqsubseteq C,\ \ B \sqsubseteq \neg A \sqcap \forall r.\neg A\ \},$$
$$\mathcal{A} = \{\ A(a),\ \ r(a, a),\ \ B(b)\ \}$$

A conflict-free syntactic CE for P is

$$\langle\emptyset,\ \{\ r(x, y),\ r(y, z),\ A(z)\ \},\ \langle a, a, a\rangle,\ \langle b, c, a\rangle,\ \emptyset\rangle$$

We need the fresh individual c since b cannot satisfy A nor have a as successor without creating a conflict. ◁

Indeed, we may even need *exponentially many* of such individual names. To show this, we reduce a problem to conflict-minimality that has been studied under the names *instance query emptiness* (Baader et al. 2016) and *flat signature-based ABox abduction* (Koopmann 2021). We reduce from the abduction problem as it simplifies transferring size bounds.

Definition 14. *A* signature-based (flat) ABox abduction problem *is a tuple $\langle\mathcal{K}, \alpha, \Sigma\rangle$ of a KB $\mathcal{K}$, an axiom α (the observation) and a set Σ of concept and role names. A* hypothesis *for this problem is a set $\mathcal{H}$ of assertions that uses only names from Σ, and for which $\mathcal{K} \cup \mathcal{H} \not\models \bot$ and $\mathcal{K} \cup \mathcal{H} \models \alpha$.*

Lemma 15. *For $\mathcal{L} \in \{\mathcal{EL}_\bot, \mathcal{ALC}, \mathcal{ALCI}\}$, let $\mathfrak{A} = \langle\mathcal{T}, C(b), \Sigma\rangle$ be a signature-based ABox abduction problem with an $\mathcal{L}$ TBox $\mathcal{T}$. Then, one can construct in polynomial time an $\mathcal{L}$-CP s.t. from every syntactic CE with empty conflict set, one can construct in polynomial time a subset-minimal hypothesis for $\mathfrak{A}$ and vice versa.*

Proof sketch. We give the idea for $\mathcal{ALCI}$ while the other reductions can be found in our technical report (Koopmann et al. 2025). We construct a new TBox $\mathcal{T}'$ that contains for every CI $C \sqsubseteq D \in \mathcal{T}$ the CI $C \sqsubseteq D \sqcup A_\bot$, where $A_\bot$ is fresh. In addition, $\mathcal{T}'$ contains $\exists r.A_\bot \sqsubseteq A_\bot$ and $A_\bot \sqsubseteq \forall r.A_\bot$ for every role r occurring in $\mathcal{T}$, and the axiom $B_\bot \sqcap A_\bot \sqsubseteq \bot$, where $B_\bot$ is also fresh. For any ABox $\mathcal{A}'$ not containing any of the fresh names, (I) $\mathcal{T}', \mathcal{A}' \not\models \bot$ and

(II) $\mathcal{T}, \mathcal{A}' \models \bot$ iff for some individual name b, $\mathcal{T}' \cup \mathcal{A}' \models A_\bot(b)$. We further define

$$\mathcal{A} =\{A(a), r(a,a) \mid A \in \mathsf{N_C} \cap \Sigma, r \in \mathsf{N_R} \cap \Sigma\} \cup \{A_\bot(a), B_\bot(b)\}.$$

The CP is now defined as $P = \langle\langle \mathcal{T}', \mathcal{A}\rangle, C \sqcup A_\bot, a, b\rangle$. □

Lemma 15 also allows us to reduce the abduction problem to the problem of computing conflict-minimal syntactic CEs. The reason is that it is easy to extend a given CP $\langle \mathcal{K}, C, a, b\rangle$ so that it has a syntactic CE with a non-empty conflict set: Simply add to $\mathcal{K}$ the following axioms, where A^* and B^* are fresh: $A^*(a)$, $B^*(b)$, $A^* \sqsubseteq C$, $A^* \sqcap B^* \sqsubseteq \bot$. Now a syntactic CE can be obtained by setting $q_{com}(\vec{x}) = \emptyset$, $q_{diff}(\vec{x}) = \{A^*(x)\}$ and $\mathcal{C} = \{B^*(b)\}$. Consequently, we can decide the existence of a hypothesis for a given abduction problem by computing a conflict-minimal CE and checking whether its conflict set is empty. This now allows us to import a range of complexity results from (Koopmann 2021). To obtain matching upper bounds, we extend the construction of the CE super-structure to now work on possible *types* of individuals in the foil evidence. Fix a CE with difference q_{diff} and foil evidence $\vec{d}$. Let $\mathcal{I}$ be a model of $\mathcal{K} \cup q_{diff}(\vec{d})$, and let $\mathcal{S}$ be the set of (sub-)concepts occurring in $\mathcal{K}$ and C. We assign to every $d \in \Delta^\mathcal{I}$ its *type* defined as $tp_\mathcal{I}(d) = a$ if $d = a^\mathcal{I}$ with a an individual that occurs in $\mathcal{K}$ or P, and otherwise as

$$tp_\mathcal{I}(d) = \{C \in \mathcal{S} \mid d \in C^\mathcal{I}\}.$$

We can use the types for an arbitrary model $\mathcal{I}$ of the KB to bound the number of fresh individuals in any given CE without introducing new conflicts. The resulting CE contains a variable $x_{c,t}$ for every individual c occurring in $\mathcal{A}$ and type t occurring in the range of $tp_\mathcal{I}$. Together with the corresponding lower bound from the abduction problem, we obtain:

Theorem 16. *There exists a family of $\mathcal{EL}_\bot$-CPs in s.t. the size of their conflict-minimal syntactic CEs is exponential in the size of the CP. At the same time, every $\mathcal{ALCI}$-CP has a subset and a cardinality conflict-minimal syntactic CE whose size is at most exponential in the size of the CP.*

We can construct the set of possible types for a given CP using a type-elimination structure, giving us a set of possible tuples for the CE in deterministic exponential time. Based on this, we can modify our method for computing difference-minimal CEs to also construct conflict-minimal CEs. For $\mathcal{ALC}$, using an oracle for entailment would yield membership in 2EXPTIME. To improve this, we observe that even in models for the constructed CP, the number of possible types is still exponentially bounded.

Theorem 17. *Deciding whether a given syntactic CE for a CP is (subset or cardinality) conflict-minimal, is*

- EXPTIME-*complete for $\mathcal{EL}_\bot$-CPs,*
- coNEXPTIME-*complete for $\mathcal{ALC}$- and $\mathcal{ALCI}$-CPs,*

where the complexity only depends on the size of the CP.

A straight-forward solution to this exponential explosion is to bound the number of individuals or to disallow fresh individuals altogether. This immediately yields a coNP-upper bound for the verification, but cannot regain tractability.

Theorem 18. *Deciding whether a given syntactic explanation without fresh individuals is (subset or cardinality) conflict-minimal is* coNP-*complete for $\mathcal{EL}_\bot$, but* EXPTIME-*complete for $\mathcal{ALC}$ and $\mathcal{ALCI}$.*

Commonality-Maximal Explanations

As illustrated in the extreme by the case of semantic CPs with concept names as concept, sometimes minimizing differences and conflicts is not sufficient, and we want to maximize the commonality instead to obtain a more focussed CE. This gives us an idea on how close the foil can get to the fact. For $\mathcal{EL}$, we prove that it is NP-complete to decide the existence of an explanation with commonality above a given threshold $n \in \mathbb{N}$. Proving the lower bound requires a different reduction as for Theorem 12.

Theorem 19. *Deciding whether a given syntactic CE for a CP is commonality-maximal is* coNP-*complete for $\mathcal{EL}/\mathcal{EL}_\bot$, but* EXPTIME-*complete for $\mathcal{ALC}/\mathcal{ALCI}$.*

Evaluation of a First Prototype

To understand how to compute CEs in practice, we implemented a first prototype for one of the variants.Theorem 11 shows that difference-minimal syntactic contrastive explanations can be computed in polynomial time, with an oracle for deciding entailment, while the other criteria are not tractable. Semantic explanations can be reduced to syntactic ones by computing all entailed concept assertions (Lemma 5), which is a standard functionality of OWL reasoning systems. Based on these observations, we developed a prototype to compute *difference-minimal syntactic CEs*.

A Practical Method for Computing CEs

To make our method for computing difference-minimal CEs practical, we refine the definition of the super structure. Fix a CP $P = \langle\langle \mathcal{T}, \mathcal{A}\rangle, C, a, b\rangle$. Our construction is now based on a subset $\mathcal{A}' \subseteq \mathcal{A}$, and a set $\boldsymbol{I} \subseteq \mathsf{N_I}$ of individuals to be used for the foil. We define $E_m = \langle q_{com}(\vec{x}), q_{diff}(\vec{x}), \vec{c}_m, \vec{d}_m, \mathcal{C}_m\rangle$, where now

- $\vec{x}$ contains a variable $x_{c,d}$ for every $\langle c, d\rangle \in \mathsf{N_I}(\mathcal{A}') \times \boldsymbol{I}$,
- $\vec{c}_m$ uses c for every $x_{c,d}$ in $\vec{x}$,
- $\vec{d}_m$ uses d for every $x_{c,d}$ in $\vec{x}$,
- $q_{com}(\vec{x}) = \{A(x_{c,d}) \mid A(c) \in \mathcal{A}', d \in \boldsymbol{I}, A(d) \in \mathcal{A}\} \cup \{r(x_{c,d}, x_{c',d'}) \mid r(c,c') \in \mathcal{A}', d, d' \in \boldsymbol{I}, r(d,d') \in \mathcal{A}\}$,
- $q_{diff}(\vec{x}) = \{A(x_{c,d}) \mid A(c) \in \mathcal{A}', d \in \boldsymbol{I}, A(d) \notin \mathcal{A}\} \cup \{r(x_{c,d}, x_{c',d'}) \mid r(c,c') \in \mathcal{A}', d, d' \in \boldsymbol{I}, r(d,d') \notin \mathcal{A}\}$.

For $\mathcal{A}' = \mathcal{A}$ and $\boldsymbol{I} = \mathsf{N_I}(\mathcal{A})$, E_m is identical to the maximal CE defined before, but too large. Instead, for $\mathcal{A}'$ we compute the union of all justifications of $C(a)$ with an optimized implementation. In $\boldsymbol{I}$, we include individuals that are "sufficiently close" to the foil, as well as some fresh individuals, using Lemma 9 and 10 to ensure that $\boldsymbol{I}$ is sufficiently large. To apply **P1**–**P4** efficiently, we modified the implementation of the justification algorithm presented in (Kalyanpur et al. 2007) to compute justifications with a fixed component:

Corpus	Signature Size			Number of Individuals			TBox size			ABox size		
	avg.	med.	range	avg.	med.	range	avg.	med.	range	avg.	med.	range
$\mathcal{EL}_\bot$	2,013	728	167 – 7,217	781	417	48 – 3,608	1,478	390	105 – 5,653	1,254	482	103 – 8,234
$\mathcal{ALCI}$	1,764	606	54 – 7,355	490	185	0 – 5,473	1,603	498	94 – 5,286	1,207	494	101 – 9,284

Table 3: Some details about the two corpora.

Corpus	#CPs		Commonality		Difference		Conflict		Fresh Individuals		Duration (sec.)	
	average	range	average	range	average	range	average	range	average	range	average	range
$\mathcal{EL}_\bot$	35.1	4 – 50	0.45	0 – 4	1.42	1 – 6	0.0	0 – 0	0.34	0 – 3	2.84	0.08 – 386.9
$\mathcal{ALCI}$	34.7	1 – 50	0.34	0 – 7	1.29	1 – 5	0.36	0 – 9	0.25	0 – 5	8.81	0.06 – 493.6

Table 4: Results for the two corpora. "#CP" states the number of CPs answered (out of 50) within the timeout (10 mins).
.

Definition 20. *Let $\mathcal{K}$ be a KB, $\mathcal{K}' \subseteq \mathcal{K}$ and α an axiom s.t. $\mathcal{K} \models \alpha$. A* justification for $\mathcal{K} \models \alpha$ with fixed component $\mathcal{K}'$ *is a subset-minimal $\mathcal{J} \subseteq (\mathcal{K} \setminus \mathcal{K}')$ s.t. $\mathcal{K}' \cup \mathcal{J} \models \alpha$.*

In each step, we compute such justifications where we fix the TBox and the components that are currently not modified, which allows to speed up the computation significantly.

Our implementation uses different optimizations for $\mathcal{ALCI}$ and for $\mathcal{EL}_\bot$. We implemented it in Java 8, using the OWL API 5.1.20 (Horridge and Bechhofer 2011), reasoning systems ELK (Kazakov, Krötzsch, and Simancik 2014) and HERMIT (Glimm et al. 2014), as well as the explanation library EVEE-LIB 0.3 (Alrabbaa et al. 2022a), which helped us in computing unions of justifications for $\mathcal{EL}_\bot$.

Benchmark

Ontologies. We used KBs from the OWL Reasoner Competition ORE 2015 (Parsia et al. 2017), namely from the tracks Materialization tracks for OWL DL and OWL EL, restricted respectively to $\mathcal{ALCI}$ and $\mathcal{EL}_\bot$. Those tracks focus on ABox reasoning, and contain KBs of varying shapes. Some KBs contained all entailed assertions, which limits their use for explanations. Therefore, we step-wisely removed from each KB all entailed assertions. KBs with more than 10,000 axioms were discared. The resulting corpora contained 46 ($\mathcal{EL}_\bot$) and 100 ($\mathcal{ALCI}$) KBs (see Table 3).

CPs. For each KB in the corpus, we performed 5 runs and constructed 10 CPs for each run. For each CP, we generated a random $\mathcal{EL}$ concept C of maximum size 5 using a random walk on the ABox starting from a randomly selected fact individual a. For this, we also considered entailed assertions. For the foil, we selected a random individual b that is not an instance of C and shares at least one concept with a.

Evaluation Results

Some of the KBs in our corpus had to be excluded: in the $\mathcal{EL}_\bot$/$\mathcal{ALCI}$ corpus, 13 were inconsistent, for 21/17 no problems could be generated under our constraints. Two more KBs from the $\mathcal{ALCI}$ corpus where removed because HERMIT threw an exception on those. Of the remaining KBs, 10/13 KBs did not allow to produce interesting CEs: for those KBs, every CE had an empty commonality and conflict, and exactly one axiom in the difference. The reason was the simple structure of the ABox, which simply allowed for no more contrasting entailments, e.g. because no role assertions were used. We exclude those KBs in the following evaluation, and focus on the remaining 14/67 ones.

Experiments were conducted on a server with 2× Intel Xeon E5-2630 v4 20 cores, 2.2GHz CPUs, along with 189 GB of available RAM running Debian 11 (Bullseye). The Java runtime environment was OpenJDK 11.0.28. The results are shown in Table 4. In general, the computed CEs tended to be simple, even though the concepts to be explained were of size up to 5. This can be explained with the simplicity of some ABoxes: if an individual has only one successor, even a complex concept of size 5 can only refer to those two individuals, and consequently the CP may use only one fact. Nonetheless, we see that every component of a CE is used, sometimes with several assertions, and also fresh individuals appear. We also see that conflicts are a relatively rare occasion, not happening at all in the $\mathcal{EL}_\bot$ corpus, which may indicate that computing conflict-free CEs could still be feasible in practice. What our evaluation also shows is that our current prototype takes surprisingly long to compute the answers. Reasons include our approach for making $q(\vec{x})$ consistent, selection of individuals, and that our construction often results in very large super-structures. This shows potential for more dedicated methods in the future.

Conclusion and Future Work

We introduced contrastive explanation problems and proposed CEs as a way to answer them. It turns out that minimizing difference is tractable and feasible in practice, while minimzing conflicts may lead to an exponential explosion. At the same time, conflicts do not seem to happen often for realistic ontologies. In the future we want to investigate dedicated algorithms for computing CEs more efficiently. We are also exploring a variant of CEs which use quantified variables in the fact and foil vectors. Moreover, one can address the counting and enumeration complexity for CEs. To conclude, we propose CEs as a tool to contrast positive and negative query answers in ontology mediated query answering.

Acknowledgments

We thank all anonymous reviewers for their valuable feedback. Research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant TRR 318/1 2021 – 438445824 and the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW) within projects WHALE (LFN 1-04) funded under the Lamarr Fellow Network programme and project SAIL, grant NW21-059D.

References

Alrabbaa, C.; Borgwardt, S.; Friese, T.; Koopmann, P.; Méndez, J.; and Popovic, A. 2022a. On the Eve of True Explainability for OWL Ontologies: Description Logic Proofs with Evee and Evonne. In Arieli, O.; Homola, M.; Jung, J. C.; and Mugnier, M., eds., *Proceedings of the 35th International Workshop on Description Logics (DL 2022) co-located with Federated Logic Conference (FLoC 2022), Haifa, Israel, August 7th to 10th, 2022*, volume 3263 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Alrabbaa, C.; Borgwardt, S.; Koopmann, P.; and Kovtunova, A. 2022b. Explaining ontology-mediated query answers using proofs over universal models. In *International Joint Conference on Rules and Reasoning*, 167–182. Springer.

Baader, F.; Bienvenu, M.; Lutz, C.; and Wolter, F. 2016. Query and Predicate Emptiness in Ontology-Based Data Access. *J. Artif. Intell. Res.*, 56: 1–59.

Baader, F.; Horrocks, I.; Lutz, C.; and Sattler, U. 2017. *An Introduction to Description Logic*. Cambridge University Press. ISBN 978-0-521-69542-8.

Bienvenu, M. 2008. Complexity of Abduction in the $\mathcal{EL}$ Family of Lightweight Description Logics. In Brewka, G.; and Lang, J., eds., *Principles of Knowledge Representation and Reasoning: Proceedings of the Eleventh International Conference, KR 2008, Sydney, Australia, September 16-19, 2008*, 220–230. AAAI Press.

Dandl, S.; Molnar, C.; Binder, M.; and Bischl, B. 2020. Multi-objective counterfactual explanations. In *International conference on parallel problem solving from nature*, 448–469. Springer.

Del-Pinto, W.; and Schmidt, R. A. 2019. ABox Abduction via Forgetting in $\mathcal{ALC}$. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 2768–2775.

Dhurandhar, A.; Chen, P.-Y.; Luss, R.; Tu, C.-C.; Ting, P.; Shanmugam, K.; and Das, P. 2018. Explanations based on the missing: Towards contrastive explanations with pertinent negatives. *Advances in neural information processing systems*, 31.

Du, J.; Wan, H.; and Ma, H. 2017. Practical TBox Abduction Based on Justification Patterns. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, 1100–1106.

Eiter, T.; Geibinger, T.; and Oetsch, J. 2023. Contrastive Explanations for Answer-Set Programs. In *European Conference on Logics in Artificial Intelligence*, 73–89. Springer.

Elsenbroich, C.; Kutz, O.; and Sattler, U. 2006. A case for abductive reasoning over ontologies. In *Proceedings of the OWLED 2006 Workshop on OWL: Experiences and Directions, Athens, Georgia, USA, November 10-11, 2006*, volume 216, 1–12. CEUR.

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 Kraus, S., ed., *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, 1682–1688. ijcai.org.

Glimm, B.; Horrocks, I.; Motik, B.; Stoilos, G.; and Wang, Z. 2014. HermiT: An OWL 2 Reasoner. *J. Autom. Reason.*, 53(3): 245–269.

Haifani, F.; Koopmann, P.; Tourret, S.; and Weidenbach, C. 2022. Connection-Minimal Abduction in $\mathcal{EL}$ via Translation to FOL. In Blanchette, J.; Kovács, L.; and Pattinson, D., eds., *Automated Reasoning - 11th International Joint Conference, IJCAR 2022*, volume 13385 of *Lecture Notes in Computer Science*, 188–207. Springer.

Heindorf, S.; Blübaum, L.; Düsterhus, N.; Werner, T.; Golani, V. N.; Demir, C.; and Ngomo, A. N. 2022. EvoLearner: Learning Description Logics with Evolutionary Algorithms. In Laforest, F.; Troncy, R.; Simperl, E.; Agarwal, D.; Gionis, A.; Herman, I.; and Médini, L., eds., *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, 818–828. ACM.

Hitzler, P.; Krötzsch, M.; and Rudolph, S. 2010. *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press. ISBN 9781420090505.

Horridge, M. 2011. *Justification based explanation in ontologies*. The University of Manchester (United Kingdom).

Horridge, M.; and Bechhofer, S. 2011. The OWL API: A Java API for OWL ontologies. *Semantic Web*, 2(1): 11–21.

Ignatiev, A.; Narodytska, N.; Asher, N.; and Marques-Silva, J. 2020. From Contrastive to Abductive Explanations and Back Again. In *AIxIA 2020 - Advances in Artificial Intelligence: XIX Int. Conf. of the Italian Association for AI*, volume 12414, 335–355. Springer.

Kalyanpur, A.; Parsia, B.; Horridge, M.; and Sirin, E. 2007. Finding All Justifications of OWL DL Entailments. In Aberer, K.; Choi, K.; Noy, N. F.; Allemang, D.; Lee, K.; Nixon, L. J. B.; Golbeck, J.; Mika, P.; Maynard, D.; Mizoguchi, R.; Schreiber, G.; and Cudré-Mauroux, P., eds., *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*, volume 4825 of *Lecture Notes in Computer Science*, 267–280. Springer.

Kazakov, Y.; Krötzsch, M.; and Simancik, F. 2014. The Incredible ELK — From Polynomial Procedures to Efficient Reasoning with $\mathcal{EL}$ Ontologies. *J. Autom. Reason.*, 53(1): 1–61.

Koopmann, P. 2021. Signature-Based Abduction with Fresh Individuals and Complex Concepts for Description Logics.

In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021*, 1929–1935.

Koopmann, P.; Del-Pinto, W.; Tourret, S.; and Schmidt, R. A. 2020. Signature-Based Abduction for Expressive Description Logics. In Calvanese, D.; Erdem, E.; and Thielscher, M., eds., *Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning, KR 2020, Rhodes, Greece, September 12-18, 2020*, 592–602.

Koopmann, P.; Mahmood, Y.; Ngomo, A.-C. N.; and Tiwari, B. 2025. Can You Tell the Difference? Contrastive Explanations for ABox Entailments. arXiv:2511.11281.

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

Lipton, P. 1990. Contrastive explanation. *Royal Institute of Philosophy Supplements*, 27: 247–266.

Marques-Silva, J.; and Ignatiev, A. 2022. Delivering Trustworthy AI through Formal XAI. In *Thirty-Sixth AAAI Conference on Artificial Intelligence*, 12342–12350. AAAI Press.

Miller, T. 2021. Contrastive explanation: a structural-model approach. *The Knowledge Engineering Review*, 36: e14.

Parsia, B.; Matentzoglu, N.; Gonçalves, R. S.; Glimm, B.; and Steigmiller, A. 2017. The OWL Reasoner Evaluation (ORE) 2015 Competition Report. *J. Autom. Reason.*, 59(4): 455–482.

Peirce, C. 1878. Deduction, induction and hypothesis: Popular Science Monthly, v. 13.

Schlobach, S. 2004. Explaining Subsumption by Optimal Interpolation. In Alferes, J. J.; and Leite, J. A., eds., *Logics in Artificial Intelligence, 9th European Conference, JELIA 2004, Lisbon, Portugal, September 27-30, 2004, Proceedings*, volume 3229 of *Lecture Notes in Computer Science*, 413–425. Springer.

Schlobach, S.; Cornet, R.; et al. 2003. Non-standard reasoning services for the debugging of description logic terminologies. In *Ijcai*, volume 3, 355–362.

Stepin, I.; Alonso, J. M.; Catala, A.; and Pereira-Fariña, M. 2021. A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence. *IEEE Access*, 9: 11974–12001.

Verma, S.; Dickerson, J.; and Hines, K. 2020. Counterfactual explanations for machine learning: A review. *arXiv preprint arXiv:2010.10596*, 2(1): 1.

Wei-Kleiner, F.; Dragisic, Z.; and Lambrix, P. 2014. Abduction framework for repairing incomplete $\mathcal{EL}$ ontologies: Complexity results and algorithms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.