Evolution of *DL-Lite* Knowledge Bases

Diego Calvanese, Evgeny Kharlamov⋆, Werner Nutt, and Dmitriy Zheleznyakov

KRDB Research Centre, Free University of Bozen-Bolzano, Italy
last_name@inf.unibz.it

Abstract. We study the problem of evolution for Knowledge Bases (KBs) expressed in Description Logics (DLs) of the *DL-Lite* family. *DL-Lite* is at the basis of OWL 2 QL, one of the tractable fragments of OWL 2, the recently proposed revision of the Web Ontology Language. We propose some fundamental principles that KB evolution should respect. We review known model and formula-based approaches for evolution of propositional theories. We exhibit limitations of a number of model-based approaches: besides the fact that they are either not expressible in DL-Lite or hard to compute, they intrinsically ignore the structural properties of KBs, which leads to undesired properties of KBs resulting from such an evolution. We also examine proposals on update and revision of DL KBs that adopt the model-based approaches and discuss their drawbacks. We show that known formula-based approaches are also not appropriate for *DL-Lite* evolution, either due to high complexity of computation, or because the result of such an action of evolution is not expressible in *DL-Lite*. Building upon the insights gained, we propose two novel formula-based approaches that respect our principles and for which evolution is expressible in *DL-Lite*. For our approaches we also developed polynomial time algorithms to compute evolution of *DL-Lite* KBs.

1 Introduction

Description Logics (DLs) provide excellent mechanisms for representing structured knowledge, and as such they constitute the foundations for the various variants of OWL, the standard ontology language of the Semantic Web[1]. DLs have traditionally been used for modeling at the intensional level the static and structural aspects of application domains [1]. Recently, however, the scope of ontologies has broadened, and they are now used also for providing support in the maintenance and evolution phase of information systems. Moreover, ontologies are considered to be the premium mechanism through which services operating in a Web context can be accessed, both by human users and by other services. Supporting all these activities, makes it necessary to equip DL systems with additional kinds of inference tasks that go beyond the traditional ones of satisfiability, subsumption, and query answering provided by current DL inference engines. The most notable one, and the subject of this paper, is that of *knowledge base evolution* [2], where the task is to incorporate new knowledge into an existing knowledge base (KB) so as to take into account changes that occur in the underlying domain of interest. In general, the new knowledge to incorporate is represented by a set of formulas denoting

⋆ The author is co-affiliated with INRIA Saclay.

[1] http://www.w3.org/TR/owl2-overview/

those properties that should be true after the KB has evolved. In the case where the new knowledge interacts in an undesirable way with the knowledge in the KB, e.g., by causing the KB or relevant parts of it to become unsatisfiable, the new knowledge cannot simply be added to the KB. Instead, suitable changes need to be made in the KB so as to avoid the undesirable interaction, e.g., by deleting parts of the KB that conflict with the new knowledge. Different choices are possible, corresponding to different semantics for KB evolution [3–8].

In the literature, two main types of KB evolution have been considered: namely revision and update [4]. Both have a precise formal grounding in terms of *postulates* [4, 5] and a number of update and revision operators were proposed in the literature [6, 5]. This work has been carried out for propositional logic, providing a thorough understanding of the various options, both wrt semantics and wrt computational properties.

Work relevant to KB evolution has been carried out initially in schema evolution in databases, cf., [9], and more recently for expressive DLs [8, 7]. However, for such richer representation formalisms, the picture is much less clear, and the various possibilities are far from being completely explored. *(i)* The fundamental distinction in DLs between TBox (for *terminological*, or intensional knowledge) and ABox (for *assertional*, or extensional knowledge), calls for distinguishing these two components (both in the existing and in the new knowledge) also in the study of evolution. *(ii)* Going from propositional letters to first-order predicates and interpretations, on the one hand calls for novel principles underlying the semantics of evolution, and on the other hand broadens the spectrum of possibilities for defining such semantics. *(iii)* The combination of constructs of the considered DL will obviously affect the complexity of computing the result of evolution, independently of the chosen semantics. *(iv)* While in propositional logic the result of an evolution step is always expressible in the same formalism, this does not hold in general for DLs [10, 11].

In this paper we address several of the points raised by the above observations, thus contributing substantially to a clarification of the problem

In line with Item (i), we carry out our investigation and establish our results by considering separately the role of the ABox and of the TBox in evolution.

Regarding Item (ii), we propose some fundamental principles that KB evolution should respect. We review known model and formula-based approaches for evolution of propositional theories [5, 6], and we lift them to the first-order case in two natural ways (by considering symmetric difference on symbol interpretations vs. interpretation atoms). Previous proposals for KB evolution, such as ABox updates under Winslett's semantics [10, 11], and the approaches proposed in [8], fit nicely into our classification.

Regarding Items (iii) and (iv), we concentrate our technical development on the *DL-Lite* family [12], which is at the basis of OWL 2 QL, one of the tractable profiles of OWL 2. We exhibit limitations of a number of model-based approaches for the logics of the *DL-Lite* family: besides the fact that evolution under such approaches is either not expressible in *DL-Lite* or hard to compute, they intrinsically ignore the structural properties of KBs, which leads to undesired properties of KBs resulting from such an evolution. We also examine proposals on update and revision of DL KBs that adopt the model-based approaches and discuss their drawbacks. We show that known formula-based approaches are also not appropriate for *DL-Lite* evolution, either due to high complexity of computation, or because the result of such an action of evolution is not expressible in *DL-Lite*. Building upon the insights gained, we propose two novel formula-based approaches that respect our principles and for which evolution is expressible in *DL-Lite*. For our approaches we also developed polynomial time algorithms to compute evolution of *DL-Lite* KBs.

2 Preliminaries and Problem Definition

Description Logics. We introduce some basic notions of Description Logics (DLs), more details can be found in [13]. A DL *knowledge base* (KB) $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ is the union of two sets of assertions, those representing the *intensional-level* of the KB, that is, the general knowledge, and constituting the *TBox* $\mathcal{T}$, and those providing information on the *instance-level* of the KB, and constituting the *ABox* $\mathcal{A}$. In our work we consider a family of DLs, *DL-Lite* [12], which form a tractable fragment of OWL 2.

All the logics of the *DL-Lite* family have the following constructs for (complex) *concepts* and *roles*: *(i)* $B ::= A \mid \exists R$, *(ii)* $C ::= B \mid \neg B$, *(iii)* $R ::= P \mid P^-$, where A and P stand for an *atomic concept* and *role*, respectively, which are just names. A *DL-Lite*$_{core}$ TBox consists of concept inclusion assertions $B \sqsubseteq C$. *DL-Lite*$_{\mathcal{FR}}$ extends *DL-Lite*$_{core}$ by allowing in a TBox role inclusion assertions $R_1 \sqsubseteq R_2$ and functionality assertions $(\mathsf{funct}\ R)$, in a way that if $R_1 \sqsubseteq R_2$ appears in a TBox, then neither $(\mathsf{funct}\ R_2)$ nor $(\mathsf{funct}\ R_2^-)$ appears in the TBox. This syntactic restriction keeps the logic tractable. ABoxes in *DL-Lite*$_{core}$ and *DL-Lite*$_{\mathcal{FR}}$ consist of membership assertions of the form $B(a)$ and $P(a, b)$. When we write in this paper *DL-Lite* without a subscript, specifying a concrete language, we mean *any* language of this family. The *DL-Lite* family has nice computational properties, for example, KB satisfiability has polynomial-time complexity in the size of the TBox and logarithmic-space in the size of the ABox [14, 15].

The semantics of DL-Lite KBs is given in the standard way, using first order interpretations, all over the same infinite countable domain Δ. An *interpretation* $\mathcal{I}$ is a function $\cdot^{\mathcal{I}}$ that assigns to each concept C a subset $C^{\mathcal{I}}$ of Δ, and to each role R a binary relation $R^{\mathcal{I}}$ over Δ in such a way that $(\neg B)^{\mathcal{I}} = \Delta \setminus B^{\mathcal{I}}$, $(\exists R)^{\mathcal{I}} = \{a \mid \exists a'.(a, a') \in R^{\mathcal{I}}\}$, and $(P^-)^{\mathcal{I}} = \{(a_2, a_1) \mid (a_1, a_2) \in P^{\mathcal{I}}\}$. We assume that Δ contains the constants and that $c^{\mathcal{I}} = c$, i.e., we adopt *standard names*. Alternatively, we view an interpretation as a set of atoms and say that $A(a) \in \mathcal{I}$ iff $a \in A^{\mathcal{I}}$ and $P(a, b) \in \mathcal{I}$ iff $(a, b) \in P^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a *model* of a membership assertion $B(a)$ if $a \in B^{\mathcal{I}}$, and of $P(a, b)$ if $(a, b) \in P^{\mathcal{I}}$, of an inclusion assertion $D_1 \sqsubseteq D_2$ if $D_1^{\mathcal{I}} \subseteq D_2^{\mathcal{I}}$, and of a functionality assertion $(\mathsf{funct}\ R)$ if the relation $R^{\mathcal{I}}$ is a function.

As usual, we use $\mathcal{I} \models F$ to denote that $\mathcal{I}$ is a model of an assertion F, and $\mathcal{I} \models \mathcal{K}$ to denote that $\mathcal{I} \models F$ for each assertion F in $\mathcal{K}$. We use *Mod*$(\mathcal{K})$ to denote the set of all models of $\mathcal{K}$. A KB is *satisfiable* if it has at least one model and it is coherent[2] if for every concept and role S occurring in $\mathcal{K}$ there is an $\mathcal{I} \in$ *Mod*$(\mathcal{K})$ such that $S^{\mathcal{I}} \neq \emptyset$. We use entailment on KBs $\mathcal{K} \models \mathcal{K}'$ in the standard sense. We say that an ABox $\mathcal{A}$ $\mathcal{T}$*-entails* an ABox $\mathcal{A}'$, denoted $\mathcal{A} \models_{\mathcal{T}} \mathcal{A}'$, if $\mathcal{T} \cup \mathcal{A} \models \mathcal{A}'$, and $\mathcal{A}$ is $\mathcal{T}$*-equivalent* to $\mathcal{A}'$, denoted $\mathcal{A} \equiv_{\mathcal{T}} \mathcal{A}'$, if $\mathcal{A} \models_{\mathcal{T}} \mathcal{A}'$ and $\mathcal{A}' \models_{\mathcal{T}} \mathcal{A}$. The deductive *closure of a TBox* $\mathcal{T}$ (of an ABox

[2] Coherence is often called *full satisfiability*.

$\mathcal{A}$), denoted $cl(\mathcal{T})$ (resp., $cl_{\mathcal{T}}(\mathcal{A})$), is the set of all TBox (resp., ABox) assertions F such that $\mathcal{T} \models F$ (resp., $\mathcal{T} \cup \mathcal{A} \models F$). It is easy to see that in *DL-Lite* $cl(\mathcal{T})$ (and $cl_{\mathcal{T}}(\mathcal{A})$) is computable in quadratic time in the size of $\mathcal{T}$ (resp., $\mathcal{T}$ and $\mathcal{A}$). In our work we assume that all TBoxes and ABoxes are closed.

Ontology Evolution. Let $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ be a *DL-Lite* KB and $\mathcal{N}$ a set of "new" (TBox and/or ABox) assertions. We want to study how to incorporate the assertions $\mathcal{N}$ into $\mathcal{K}$, that is, how $\mathcal{K}$ *evolves* [2] under $\mathcal{N}$. More practically, we want to develop *evolution operators* that take $\mathcal{K}$ and $\mathcal{N}$ as input and return, possibly in *polynomial time*, a *DL-Lite* KB $\mathcal{K}'$ that captures the evolution, and which we call *the evolution of* $\mathcal{K}$ *under* $\mathcal{N}$.

In the Semantic Web context, update and revision [4, 5], the two classical understandings of ontology evolution, are too restrictive from the intuitive and formal perspective: in many applications we know neither the status of the real world, nor how accurate $\mathcal{N}$ is wrt to the world. For example, if in $\mathcal{K}$ we store knowledge from Web sources, say, online newspapers that we collected using RSS feeds or Web crawling, then there is no chance to say how this information is related to the state of the real world. When a new portion of knowledge $\mathcal{N}$ arrives to $\mathcal{K}$ and conflicts with $\mathcal{K}$, then it might be unclear whether the conflict is due to outdated or wrong information in $\mathcal{K}$. This situation does not fit in the formalisms of update and revision and, therefore, we propose now some new postulates to be adopted in the context of evolution in the Semantic Web.

First, we assume that the KBs we are dealing with make sense, that is, they are coherent (and hence also satisfiable), and we want evolution to preserve this property:

EP1: Evolution should preserve coherence of the KB, that is, $\mathcal{K}'$ is coherent.

The same postulate is stipulated in [8]. Notice that in *DL-Lite*[3] coherence can be reduced to satisfiability. Moreover, when $\mathcal{N}$ may contain ABox assertions, one can enforce coherence by adding to $\mathcal{N}$ for each atomic concept A an assertion $A(d_A)$, and for each atomic role P an assertion $P(d_P, d'_P)$, where d_A, d_P, d'_P are fresh individuals.

For example, if our online newspapers KB $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ records that John is married to Mary and that a person can be married to at most one person, and if the new knowledge $\mathcal{N}$ says that John is married to Patty, then $\mathcal{K} \cup \mathcal{N}$ is unsatisfiable (and hence incoherent) and does not comply with **EP1**. This can be resolved by either *(i)* discarding the old information about John's marriage, that is, by changing $\mathcal{A}$, or *(ii)* weakening the constraint in $\mathcal{K}$ on the number of spouses, that is, by changing $\mathcal{T}$, or *(iii)* discarding $\mathcal{N}$. What to do depends on the application. In data-centric applications, the most valuable information is the (extensional) data and we would have to discard the constraint on the number of spouses from $\mathcal{T}$. In Web data integration, the constraints of $\mathcal{T}$ define the global schema and the data coming from different Web sources may be contradictive by nature. Thus, it makes more sense to discard one of the two assertions about John's spouses using, for example, the trust we have in the sources of the data. To formalize this consideration we introduce the notion of *protected part* of a KB, which is simply a subset $\mathcal{K}_{pr} \subseteq \mathcal{K}$ that is preserved by evolution. This is sanctioned by our second postulate:

EP2: Evolution should entail the new knowledge and preserve the protected part, that is, $\mathcal{K}' \models \mathcal{K}_{pr} \cup \mathcal{N}$

This postulate is different from the classical ones of update and revision where it is only

[3] Actually, in all logics enjoying the disjoint-union model property.

required that the new KB $\mathcal{K}'$ should entail the new knowledge $\mathcal{N}$. We observe, however, that evolution of $\mathcal{K}$ with a protected part $\mathcal{K}_{pr}$ wrt $\mathcal{N}$ is conceptually the same as evolution of $\mathcal{K}$ with the empty protected part wrt $\mathcal{K}_{pr} \cup \mathcal{N}$.

Another principle that is widely accepted [5, 4] is the one of *minimality of change*:

EP3: The change to $\mathcal{K}$ should be minimal, that is, $\mathcal{K}'$ is minimally different from $\mathcal{K}$.

There are different approaches to define minimality, suitable for particular applications, and the current belief is that there is no general notion of minimality that will "do the right thing" under all circumstances [6].

Based on these principles, we will study evolution operators. We will consider the classical update and revision operators coming from AI [5] and also operators proposed for DLs [15, 8], and try to adapt them to our needs. In the following we distinguish three types of evolution: *TBox evolution*, when $\mathcal{N}$ consists of TBox assertions only, and we denote it $\mathcal{N}_T$, *ABox evolution*, when $\mathcal{N}$ consists of ABox assertions only, and we denote it $\mathcal{N}_A$, *KB evolution*, when $\mathcal{N}$ includes both TBox and ABox assertions.

Running Example. In our online-newspapers KB we have structural knowledge that wives (W) are exactly those individuals who have husbands (hh) and that some wives are employed (E). Singles (S) cannot be husbands. Priests (P) are clerics (C) and clerics are singles. Both clerics and wives are receivers of rent subsidies (R). We also know that Adam (a) and Bob (b) are priests, Mary (m) is a wife who is employed and her husband is John (j). Also, Carl (c) is a catholic minister (M). This knowledge can be expressed in *DL-Lite* by the KB $\mathcal{K}_{ex}$, consisting of the following assertions:

$\mathcal{T}$: $W \sqsubseteq \exists hh$, $\exists hh \sqsubseteq W$, $E \sqsubseteq W$, $S \sqsubseteq \neg\exists hh^-$, $P \sqsubseteq C$, $C \sqsubseteq S$, $C \sqsubseteq R$, $W \sqsubseteq R$;
$\mathcal{A}$: $P(a)$, $P(b)$, $E(m)$, $hh(m, j)$, $M(c)$.

By crawling some Web sources we found out that John is now single (that is, $S(j)$), in the Oxford Dictionary we discovered that catholic ministers are superiors of some religious orders and hence clerics ($M \sqsubseteq C$), and from economic news we found out that the current crisis affects people receiving rent subsidies in that subsidies were canceled for wives ($W \sqsubseteq \neg R$) and for clerics ($C \sqsubseteq \neg R$), since the former may receive support from their husbands and the latter from their church. In the rest of the paper we will discuss how to incorporate this new knowledge into our KB.

3 Approaches to Evolution

A number of candidate semantics for evolution operators have been proposed in the literature [6, 3, 16, 15, 8]. They can be divided into two groups, *model-based approaches* (MBAs) and *formula-based approaches* (FBAs).

3.1 Model-Based Approaches

In model-based approaches (MBAs) the result of evolution of a KB $\mathcal{K}$ wrt new knowledge $\mathcal{N}$ is a set $\mathcal{K} \diamond \mathcal{N}$ of models. The general idea of MBAs is to choose as the result of evolution some models of $\mathcal{N}$ depending on their distance to the models of $\mathcal{K}$. Katsuno and Mendelzon [4] considered two ways of choosing these models of $\mathcal{N}$.

The idea of the first one, which we call *local*, is to go over all models $\mathcal{I}$ of $\mathcal{K}$ and for each $\mathcal{I}$ to take those models $\mathcal{J}$ of $\mathcal{N}$ that are minimally distant from $\mathcal{I}$. Formally,

$$\mathcal{K} \diamond \mathcal{N} = \bigcup_{\mathcal{I} \in \mathit{Mod}(\mathcal{K})} \operatorname*{arg\,min}_{\mathcal{J} \in \mathit{Mod}(\mathcal{N})} \mathit{dist}(\mathcal{I}, \mathcal{J}),$$

where $\mathit{dist}(\cdot, \cdot)$ is a function whose range is a partially ordered domain and $\arg\min$ stands for the *argument of the minimum,* that is, in our case, the set of models $\mathcal{J}$ for which the value of $\mathit{dist}(\mathcal{I}, \mathcal{J})$ reaches its minimum value, given $\mathcal{I}$. The distance function *dist* varies from approach to approach and commonly takes as values either numbers or subsets of some fixed set.

The idea of the second way, called *global*, is to choose those models $\mathcal{J}$ of $\mathcal{N}$ that are minimally distant from the entire set of models of $\mathcal{K}$. Formally,

$$\mathcal{K} \diamond \mathcal{N} = \operatorname*{arg\,min}_{\mathcal{J} \in \mathit{Mod}(\mathcal{N})} \mathit{dist}(\mathit{Mod}(\mathcal{K}), \mathcal{J}), \tag{1}$$

where $\mathit{dist}(\mathit{Mod}(\mathcal{K}), \mathcal{J}) = \min_{\mathcal{I} \in \mathit{Mod}(\mathcal{K})} \mathit{dist}(\mathcal{I}, \mathcal{J})$.

The classical MBAs were developed for propositional theories. In this context, an interpretation was identified with the set of propositional atoms that it makes true and two distance functions were introduced, respectively based on symmetric difference and on the cardinality of symmetric difference,

$$\mathit{dist}_{\subseteq}(\mathcal{I}, \mathcal{J}) = \mathcal{I} \ominus \mathcal{J} \qquad \text{and} \qquad \mathit{dist}_{\sharp}(\mathcal{I}, \mathcal{J}) = |\mathcal{I} \ominus \mathcal{J}|. \tag{2}$$

where the symmetric difference of two sets is defined as $\mathcal{I} \ominus \mathcal{J} = (\mathcal{I} \setminus \mathcal{J}) \cup (\mathcal{J} \setminus \mathcal{I})$. Distances under $\mathit{dist}_{\subseteq}$ are sets and are compared by set inclusion, that is, $\mathit{dist}_{\subseteq}(\mathcal{I}_1, \mathcal{J}_1) \leq \mathit{dist}_{\subseteq}(\mathcal{I}_2, \mathcal{J}_2)$ iff $\mathit{dist}_{\subseteq}(\mathcal{I}_1, \mathcal{J}_1) \subseteq \mathit{dist}_{\subseteq}(\mathcal{I}_2, \mathcal{J}_2)$. Distances under $\mathit{dist}_{\sharp}$ are natural numbers and are compared in the standard way.

One can extend these distances to DL interpretations in two different ways. One way is to consider interpretations $\mathcal{I}$, $\mathcal{J}$ as sets of *atoms*. Then $\mathcal{I} \ominus \mathcal{J}$ is again a set of atoms and we can define distances as in Equation (2). We denote these distances as $\mathit{dist}^{a}_{\subseteq}(\mathcal{I}, \mathcal{J})$ and $\mathit{dist}^{a}_{\sharp}(\mathcal{I}, \mathcal{J})$. While in the propositional case distances are always finite, this may not be the case for DL interpretations that are infinite. Another way is to define distances at the level of the concept and role *symbols* in the underlying signature Σ:

$$\mathit{dist}^{s}_{\subseteq}(\mathcal{I}, \mathcal{J}) = \{S \in \Sigma \mid S^{\mathcal{I}} \neq S^{\mathcal{J}}\}, \quad \text{and} \quad \mathit{dist}^{s}_{\sharp}(\mathcal{I}, \mathcal{J}) = |\{S \in \Sigma \mid S^{\mathcal{I}} \neq S^{\mathcal{J}}\}|.$$

Summing up across the different possibilities, we have three dimensions, which give eight possibilities to define a semantics of evolution according to MBAs by choosing: (1) the *local* or the *global* approach, (2) *atoms* or *symbols* for defining distances, and (3) *set inclusion* or *cardinality* to compare symmetric differences.

We denote each of these eight possibilities by a combination of three symbols, indicating the choice in each dimension. By $\mathcal{L}$ we denote local and by $\mathcal{G}$ global semantics. We attach the superscripts a or s to indicate whether distances are defined in terms of atoms or symbols. We use the subscripts $\subseteq$ or $\sharp$ to indicate whether distances are compared in terms of set inclusion or cardinality. For example, $\mathcal{L}^{a}_{\sharp}$ denotes the local semantics where the distances are expressed in terms of cardinality of sets of atoms.

Considering that in the propositional case a distinction between atom and symbol-based semantics is meaningless, we can also use our notation, without superscripts, to identify MBAs in that setting. Interestingly, the two classical local MBAs proposed by Winslett [6] and Forbus [17] correspond, respectively, to $\mathcal{L}_{\subseteq}$, and $\mathcal{L}_{\sharp}$, while the one by Borgida [18] is a variant of $\mathcal{L}_{\subseteq}$. The two classical global MBAs proposed by Satoh [5] and Dalal [19] correspond, respectively, to $\mathcal{G}_{\subseteq}$, and $\mathcal{G}_{\sharp}$.

Under each of our eight semantics, evolution results in a set of interpretations. In the propositional case each set of interpretations over finitely many symbols can be captured by a formula whose models are exactly those interpretations. In the case of DLs this is no more necessarily the case, since on the one hand, interpretations can be infinite and on the other hand logics may miss some connectives like disjunction or negation.

Let $\mathcal{D}$ be a DL and M one of the eight MBAs introduced above. We say $\mathcal{D}$ is *closed under evolution wrt* M (or evolution wrt M is *expressible* in $\mathcal{D}$) if for any KBs $\mathcal{K}$ and $\mathcal{N}$ written in $\mathcal{D}$, there is a KB $\mathcal{K}'$ written in $\mathcal{D}$ such that $\mathit{Mod}(\mathcal{K}') = \mathcal{K} \diamond \mathcal{N}$. We study now whether the logics of the *DL-Lite* family are closed under the various semantics.

Global Model-Based Approaches. We start with an example showing that wrt all four semantics $\mathcal{G}^s_{\subseteq}$, $\mathcal{G}^s_{\sharp}$, $\mathcal{G}^a_{\subseteq}$ and $\mathcal{G}^a_{\sharp}$, TBox evolution is not expressible in *DL-Lite*.

The observation underlying these results is that on the one hand, the minimality of change principle introduces implicit disjunction in the evolved KB. On the other hand, *DL-Lite* can be embedded into a slight extension of Horn logic [20] and therefore does not allow one to express genuine disjunction. Technically, this can be expressed by saying that every *DL-Lite* KB that entails a disjunction of *DL-Lite* assertions entails one of the disjuncts. The lemma gives a contrapositive formulation of this statement. Although *DL-Lite* does not have a disjunction operator, by abuse of notation we write $\mathcal{J} \models \phi \vee \psi$ as a shorthand for "$\mathcal{J} \models \phi$ or $\mathcal{J} \models \psi$" for *DL-Lite* assertions ϕ, ψ.

Lemma 1. *Let $\mathcal{M}$ be a set of interpretations. Suppose there are DL-Lite assertions ϕ, ψ such that (1) $\mathcal{J} \models \phi \vee \psi$ for every $\mathcal{J} \in \mathcal{M}$; (2) there are $\mathcal{J}_1$, $\mathcal{J}_2 \in \mathcal{M}$ such that $\mathcal{J}_1 \not\models \phi$ and $\mathcal{J}_2 \not\models \psi$. Then there is no DL-Lite KB $\mathcal{K}$ such that $\mathcal{M} = \mathit{Mod}(\mathcal{K})$.*

Example 2. Consider the KB $\mathcal{K}_{ex}$ of our running example and assume that the new information $\mathcal{N}_T = \{ W \sqsubseteq \neg R \}$ arrived. We explore evolution wrt the $\mathcal{G}^s_{\sharp}$ semantics of $\diamond$, which counts for how many symbols the interpretation changes.

Consider three assertions, (derived) from $\mathcal{K}$, that are essential for this example: $E \sqsubseteq W$, $E \sqsubseteq R$, and $E(m)$. One can show that the minimum of $\mathit{dist}^s_{\sharp}(\mathcal{I}, \mathcal{J})$ for $\mathcal{I} \in \mathit{Mod}(\mathcal{K})$ and $\mathcal{J} \in \mathit{Mod}(\mathcal{N}_T)$ equals 1. Let $\mathcal{J} \in \mathcal{K} \diamond \mathcal{N}_T$. Then there exists $\mathcal{I} \in \mathit{Mod}(\mathcal{K})$ such that $\mathit{dist}^s_{\sharp}(\mathcal{I}, \mathcal{J}) = 1$. Hence, there is only one symbol $S \in \{E, W, R\}$ whose interpretation has changed from $\mathcal{I}$ to $\mathcal{J}$, that is $S^{\mathcal{I}} \neq S^{\mathcal{J}}$. Observe that S cannot be E. Otherwise, W and R would be interpreted identically under $\mathcal{I}$ and $\mathcal{J}$, and W and R would not be disjoint under $\mathcal{J}$, since m is an instance of both, thus contradicting $\mathcal{N}_T$. Now, assume that W has not changed. Then $\mathcal{J} \models E \sqsubseteq W$, since this held already for $\mathcal{I}$. However, $\mathcal{J} \not\models E \sqsubseteq R$, since $m \in E^{\mathcal{J}}$, but $m \notin R^{\mathcal{J}}$, due to the disjointness of W and R with respect to $\mathcal{J}$. Similarly, if we assume that R has not changed, it follows that $\mathcal{J} \models E \sqsubseteq R$, but $\mathcal{J} \not\models E \sqsubseteq W$. By Lemma 1 we conclude that $\mathcal{K} \diamond \mathcal{N}_T$ is not expressible in *DL-Lite*.

Analogously one can also show inexpressibility for $\mathcal{G}^s_{\subseteq}$, $\mathcal{G}^a_{\subseteq}$, and $\mathcal{G}^a_{\sharp}$. ■

From the example we conclude our first inexpressibility result.

Theorem 3. *DL-Lite is not closed under TBox evolution wrt* $\mathcal{G}^s_\subseteq$, $\mathcal{G}^s_\sharp$, $\mathcal{G}^a_\subseteq$, *and* $\mathcal{G}^a_\sharp$.

With a similar argument one can show that the operator $\diamond_{M'}$ of Qi and Du [8] (and its stratified extension $\diamond_S$), is not expressible in *DL-Lite*. This operator is a variant of $\mathcal{G}^s_\sharp$ where in Equation (1) one considers only models $\mathcal{J} \in \mathit{Mod}(\mathcal{N})$ that satisfy $A^\mathcal{J} \neq \emptyset$ for every A occurring in $\mathcal{K} \cup \mathcal{N}$. The modification does not affect the inexpressibility, which can again be shown using Example 2. We note that $\diamond_{M'}$ was developed for KB revision with empty ABoxes and the inexpressibility comes from the non empty ABox.

Local Model-Based Approaches. We start with an example showing that both ABox and TBox evolution wrt the $\mathcal{L}^a_\subseteq$ and $\mathcal{L}^a_\sharp$ semantics are not expressible in *DL-Lite*.

Example 4. We turn again to our KB $\mathcal{K}_{ex}$ and consider the scenario where we are informed that John is now a single, formally $\mathcal{N}_A = \{S(j)\}$. Suppose we want to perform ABox evolution where the TBox of $\mathcal{K}_{ex}$ is protected. The TBox assertions essential for this example are $W \sqsubseteq \exists hh$, $\exists hh \sqsubseteq W$, and $P \sqsubseteq \neg\exists hh^-$, that is, an individual is a wife iff she has a husband, and a priests is not a husband. The essential ABox assertions are $W(m)$, $P(a)$, and $P(b)$. We show the inexpressibility of evolution wrt $\mathcal{L}^a_\subseteq$ using Lemma 1.

Under $\mathcal{L}^a_\subseteq$, in every $\mathcal{J} \in \mathcal{K} \diamond \mathcal{N}_A$ one of four situations holds: *(i)* Mary is not a wife, that is, $\mathcal{J} \not\models W(m)$, and both Adam and Bob are priests, that is, $\mathcal{J} \models P(a) \wedge P(b)$. Hence, $\mathcal{J} \models P(a) \vee P(b)$. *(ii)* Mary is a wife and her husband is different from Adam and Bob. Due to minimality of change, both Adam and Bob are still priests, as in Case (i), and again $\mathcal{J} \models P(a) \vee P(b)$. *(iii)* Mary is a wife and her husband is Adam. Then Bob, due to mininality of change, is still a priest. Hence, $\mathcal{J} \models P(a) \vee P(b)$. Moreover, the new husband cannot stay priest any longer and $\mathcal{J} \not\models P(a)$. *(iv)* Mary is a wife and her husband is Bob. Analogously to Case (iii), we have $\mathcal{J} \models P(a) \vee P(b)$ and $\mathcal{J} \not\models P(b)$. We are in the conditions of Lemma 1, that is, for every model $\mathcal{J} \in \mathcal{K} \diamond \mathcal{N}_A$ it holds that $\mathcal{J} \models P(a) \vee P(b)$, and there are $\mathcal{J}' \in \mathcal{K} \diamond \mathcal{N}'_A$ s.t. $\mathcal{J}' \not\models P(a)$ and $\mathcal{J}'' \in \mathcal{K} \diamond \mathcal{N}_A$ s.t. $\mathcal{J}'' \not\models P(b)$. Consequently, the set of models $\mathcal{K} \diamond \mathcal{N}_A$ is not expressible in *DL-Lite*.

To show that the example works also for $\mathcal{L}^a_\sharp$, we need extra arguments. Intuitively, if a model $\mathcal{I} \models \mathcal{K}_{ex}$ contains individuals that are single, but not clerics, then the models $\mathcal{J} \models \mathcal{N}_A$ closest to $\mathcal{I}$ in terms of $\mathit{dist}_\sharp$ are such that Mary, if she remains a wife, marries one of these individuals and Adam and Bob remain priests, since this involves the fewest changes of atoms. However, this is no more the case if we consider a model $\mathcal{I}_0 \models \mathcal{K}_{ex}$ where everyone, except John, is a priest, that is $P^{\mathcal{I}_0} = \Delta \setminus \{j\}$. Reasoning as before, one can see that among the models $\mathcal{J}$ of $\mathcal{N}_A$ closest to $\mathcal{I}_0$, there are some such that $\mathcal{J} \not\models P(a)$ and others such that $\mathcal{J} \not\models P(b)$, while all of them satisfy $\mathcal{J} \models P(a) \vee P(b)$. Then Lemma 1 implies that $\mathcal{K} \diamond \mathcal{N}_A$ under $\mathcal{L}^a_\sharp$ is not expressible in *DL-Lite*.

Now, we consider TBox evolution, which means that the ABox of $\mathcal{K}_{ex}$ is protected. Suppose we found out that ministers are clerics, formally $\mathcal{N}_T = \{M \sqsubseteq C\}$. The assertions of $\mathcal{K}_{ex}$ essential for this example are $C \sqsubseteq S$ and $M(c)$. Assume there is a representation $\mathcal{K}'$ of the $\mathcal{K}_{ex} \diamond \mathcal{N}_T$ under $\mathcal{L}^a_\subseteq$. Since $\mathcal{K}_1 = \mathcal{K}_{ex} \cup \mathcal{N}_T$ is fully satisfiable, one might expect that $\mathcal{K}' = \mathcal{K}_1$. It turns out this is not the case. Indeed, since every model $\mathcal{J} \in \mathcal{K}_{ex} \diamond \mathcal{N}_T$ is such that $\mathcal{J} \models \mathcal{N}_T \cup \{M(c)\}$, it holds that $c \in M^\mathcal{J} \subseteq C^\mathcal{J}$.

Moreover, if $\mathcal{I} \in Mod(\mathcal{K})$ is such that $c \notin S^{\mathcal{I}}$, then $c \notin S^{\mathcal{J}'}$ for any $\mathcal{J}' \in Mod(\mathcal{N}_T)$ minimally different from $\mathcal{I}$. At the same time $\mathcal{K}_1 \models S(c)$, hence, such a $\mathcal{J}'$ is not a model of $\mathcal{K}_1$ and $\mathcal{K}_1$ cannot be $\mathcal{K}'$. Since the inclusion $C \sqsubseteq S$ caused the problem above, it might be the case that $\mathcal{K}'$ is $\mathcal{K}_2 = \mathcal{K}_1 \setminus \{C \sqsubseteq S\}$. It turns out this is not the case either, since $\mathcal{K}_2$ has models that are not in $\mathcal{K} \diamond \mathcal{N}_T$. Can we resolve this by adding some assertion to $\mathcal{K}_2$? No, again. If one adds *any DL-Lite* TBox assertion to $\mathcal{K}_2$ that is not entailed by $\mathcal{K}_2$ or not $C \sqsubseteq S$, one gets a KB with models not in $\mathcal{K} \diamond \mathcal{N}_T$. Hence, no representation $\mathcal{K}'$ of $\mathcal{K} \diamond \mathcal{N}_T$ exists. Analogously, one can show that $\mathcal{K} \diamond \mathcal{N}_T$ under $\mathcal{L}^a_\sharp$ is also not expressible in *DL-Lite*. ■

This example proves our second inexpressibility result, which follows.

Theorem 5. *DL-Lite is not closed under evolution wrt $\mathcal{L}^a_\subseteq$ and $\mathcal{L}^a_\sharp$. This holds already for the special cases of TBox evolution and ABox evolution with protected TBox.*

De Giacomo et al. [21] considered ABox evolution with protected TBox wrt $\mathcal{L}^a_\subseteq$ semantics. They presented an algorithm to compute *DL-Lite*$_{\mathcal{FR}}$ KBs that represent $\mathcal{K} \diamond \mathcal{N}_A$ for *DL-Lite*$_{\mathcal{FR}}$ KBs $\mathcal{K}$ and $\mathcal{N}_A$. As a consequence of Theorem 5, their algorithm is not complete.

A strange effect of evolution under $\mathcal{L}^a_\sharp$ semantics is that new information may "erase" completely the previous KB.

Proposition 6. *Let $\mathcal{K}$ be a KB with at least one finite model and let $\mathcal{N}$ be a satisfiable KB such that all its models are infinite. Then under $\mathcal{L}^a_\sharp$ we have that $\mathcal{K} \diamond \mathcal{N} = \mathcal{N}$.*

Since the *DL-Lite* logics without role functionality have the finite model property, that is, every satisfiable KB in these logics has a finite model, the above situation cannot occur for them. At the same time, in every *DL-Lite* logic with role functionality there are KBs all of whose models are infinite and such an erasure can take place.

The properties of the $\mathcal{L}^s_\subseteq$ and $\mathcal{L}^s_\sharp$ semantics are still an open problem for us.

We now discuss conceptual problems with all the local semantics. Recall Example 4 for local MBAs $\mathcal{L}^a_\subseteq$ and $\mathcal{L}^a_\sharp$. We note two problems. First, the divorce of Mary from John had a strange effect on the priests Bob and Adam. The semantics questions their celibacy and we have to drop the information that they are priests. This is counterintuitive, since Mary and her divorce have nothing to do with any of these priests. Actually, the semantics also erases from the KB assertions about all other people belonging to concepts whose instances are not married, since potentially each of them is Mary's new husband. Second, a harmless clarification introduced to the TBox that ministers are in fact clerics strangely affects the whole class of clerics. The semantics of evolution "requires" one to allow marriages for clerics. This appears also strange, because intuitively the clarification on ministers does not contradict by any means the celibacy of clerics.

Also the four global MBAs have conceptual problems that were exhibited in Example 2. The restriction on rent subsidies that cuts the payments for wives introduces a counterintuitive choice for employed wives. Under the symbol-based global semantics, they must either *collectively* get rid of their husbands or *collectively* lose the subsidy. Under atom-based semantics the choice is an individual one.

Summing up on both the global and the local MBAs that we have considered, they focus on minimal change of *models* of KBs and, hence, introduce choices that cannot be captured in *DL-Lite*, which owes its good computational properties to the absence of disjunction. This mismatch with regard to the structural properties of KBs leads to counterintuitive and undesired results, like inexpressibility in *DL-Lite* and erasure of the entire KB. Therefore, we think that these semantics are not suitable for the evolution of *DL-Lite* KBs, whether or not they satisfy **EP1**-**EP3**, and now study evolution according to formula-based approaches.

3.2 Formula-Based Approaches

Under formula-based approaches, the objects of change are sets of formulas. Given a KB $\mathcal{K}$ and new knowledge $\mathcal{N}$, a natural way to define the result of evolution seems to choose a maximal subset $\mathcal{K}_m$ of $\mathcal{K}$ that is consistent with $\mathcal{N}$. The result of evolution in this case is a set of formulas $\mathcal{K} \diamond \mathcal{N} = \mathcal{K}_m \cup \mathcal{N}$ However, a problem with this is that in general such a $\mathcal{K}_m$ is not unique.

Let $\mathcal{M}(\mathcal{K}, \mathcal{N})$ be the set of all such maximal $\mathcal{K}_m$. In the past, researchers have proposed a number of approaches to combine all elements of $\mathcal{M}(\mathcal{K}, \mathcal{N})$ into one set of formulas, which is then added to $\mathcal{N}$ [5, 6]. The two main ones are known as *Cross-Product*, or *CP* for short, and *When In Doubt Throw It Out*, or *WIDTIO* for short. The corresponding sets $\mathcal{K}_{\mathit{CP}}$ and $\mathcal{K}_{\mathit{WIDTIO}}$ are defined as follows:

$$\mathcal{K}_{\mathit{CP}} := \Big\{ \bigvee_{\mathcal{K}_m \in \mathcal{M}(\mathcal{K},\mathcal{N})} (\bigwedge_{\phi \in \mathcal{K}_m} \phi) \Big\}. \qquad \mathcal{K}_{\mathit{WIDTIO}} := \bigcap_{\mathcal{K}_m \in \mathcal{M}(\mathcal{K},\mathcal{N})} \mathcal{K}_m,$$

In CP one adds to $\mathcal{N}$ the disjunction of all $\mathcal{K}_m$, viewing each $\mathcal{K}_m$ as the conjunction of its assertions, while in WIDTIO one adds to $\mathcal{N}$ those formulas present in all $\mathcal{K}_m$. In terms of models, every model of $\mathcal{K}_{\mathit{WIDTIO}}$ is also a model of $\mathcal{K}_{\mathit{CP}}$, whose models are exactly the interpretations satisfying *some* of the $\mathcal{K}_m$.

Example 7. We consider again our running example. Suppose, we obtain the new information that priests no longer obtain rental subsidies. This can be captured by the set of TBox assertions $\mathcal{N}_T = \{P \sqsubseteq \neg R\}$. We now incorporate this information into our KB, under both CP and WIDTIO semantics. Clearly, $\mathcal{K}_{ex} \cup \mathcal{N}_T$ is not coherent and to resolve the conflict one can drop either $P \sqsubseteq C$ or $C \sqsubseteq R$. Hence, $\mathcal{M}(\mathcal{K}_{ex}, \mathcal{N}_T) = \{\mathcal{K}_m^{(1)}, \mathcal{K}_m^{(2)}\}$, where $\mathcal{K}_m^{(1)} = \mathcal{K} \setminus \{P \sqsubseteq C\}$, and $\mathcal{K}_m^{(2)} = \mathcal{K} \setminus \{C \sqsubseteq R\}$. Consequently, the results of evolving $\mathcal{K}$ with respect to $\mathcal{N}_T$ under the two semantics are

$$\mathcal{N}_T \cup \mathcal{K}_{\mathit{CP}} = \mathcal{N}_T \cup ((\mathcal{K} \setminus \{P \sqsubseteq C\}) \vee (\mathcal{K} \setminus \{C \sqsubseteq R\})) \tag{3}$$

$$\mathcal{N}_T \cup \mathcal{K}_{\mathit{WIDTIO}} = \mathcal{N}_T \cup \left(\mathcal{K}_m^{(1)} \cap \mathcal{K}_m^{(2)}\right) = \mathcal{N}_T \cup \mathcal{K}_{ex} \setminus \{P \sqsubseteq C, C \sqsubseteq R\},$$

where in (3) we have combined DL notation with first order logic notation. ∎

Intuitively, CP does not lose information, but the price to pay is that the resulting KB can be exponentially larger than the original KB, since there can exist exponentially many $\mathcal{K}_m$. In addition, as the example shows, even if $\mathcal{K}$ is a *DL-Lite* KB, the resulting

$\mathcal{K}_{CP}$ may not be representable in *DL-Lite* anymore since it requires disjunction. This effect is also present if the new knowledge involves only ABox assertions.

WIDTIO, on the other extreme, is expressible in *DL-Lite*. However, it can lose many assertions, which may be more than one is prepared to tolerate. Even, if one deems this loss acceptable, one has to cope with the fact that it is generally difficult to decide whether an assertion belongs to $\mathcal{K}_{WIDTIO}$. This problem is already difficult if our KBs are TBoxes that are specified in the simplest variant of *DL-Lite*. We note that the following theorem can be seen as a sharpening of a result about WIDTIO for propositional Horn theories in [5], obtained with a different reduction than ours.

Theorem 8. *Given DL-Lite TBoxes $\mathcal{T}$ and $\mathcal{N}_T$ and an inclusion assertion $A \sqsubseteq B$, deciding whether $A \sqsubseteq B \in \bigcap_{\mathcal{K}_m \in \mathcal{M}(\mathcal{T},\mathcal{N}_T)} \mathcal{K}_m$, is* coNP*-complete. Hardness holds already for DL-Lite$_{core}$.*

Against this backdrop we conclude that neither CP nor WIDTIO are good for practical solutions. As a pragmatic alternative we will explore the approach to nondeterministically choose *some* $\mathcal{K}_m^{(0)}$ among the $\mathcal{K}_m$. We call this semantics *bold semantics.*

4 Bold Semantics

We define as *bold semantics* the approach to evolution where, given a KB $\mathcal{K} = \mathcal{T} \cup \mathcal{A}$ and new knowledge $\mathcal{N}$, we add to $\mathcal{N}$ a *maximal compatible* subset $\mathcal{K}_m^{(0)} \subseteq cl(\mathcal{K})$, that is, a set such that $\mathcal{N} \cup \mathcal{K}_m^{(0)}$ is coherent and such that that $\mathcal{K}_m^{(0)}$ is maximal wrt to this property. Note that now we choose a subset of the *deductive closure* of $\mathcal{K}$ and not of $\mathcal{K}$ alone. By abuse of notation, we will use a binary operator to denote any result of bold evolution and write

$$\mathcal{K} \diamond_b \mathcal{N} = \mathcal{N} \cup \mathcal{K}_m^{(0)},$$

although $\mathcal{K} \diamond_b \mathcal{N}$ is not uniquely defined.

Example 9. Consider the KB and the update request from Example 7. As shown there, $\mathcal{M}(\mathcal{K}_{ex}, \mathcal{N}_T) = \{\mathcal{K}_m^{(1)}, \mathcal{K}_m^{(2)}\}$. According to bold semantics the result of the update is a KB $\mathcal{K}' = \mathcal{N} \cup \mathcal{K}_m^{(0)}$ for some $\mathcal{K}_m^{(0)} \in \mathcal{M}(\mathcal{K}_{ex}, \mathcal{N}_T)$. Thus, the result of the update is either $\mathcal{N}_T \cup \mathcal{K}_{ex} \setminus \{P \sqsubseteq C\}$ or $\mathcal{N}_T \cup \mathcal{K}_{ex} \setminus \{C \sqsubseteq R\}$. Whether to select one or the other of these two options depends on preferences, which we do not consider here. ■

Choosing an arbitrary $\mathcal{K}_m$ has the advantage that $\mathcal{K} \diamond_b \mathcal{N}$ can be computed in polynomial time. In Fig. 1 we present a nondeterministic algorithm that, given a KB $\mathcal{K}$ and new knowledge $\mathcal{N}$, returns a set $\mathcal{K}_m \subseteq cl(\mathcal{K})$ that is a maximal compatible set of assertions for $\mathcal{K}$ and $\mathcal{N}$. The algorithm loops as many times as there are assertions in $cl(\mathcal{K})$. The number of such assertions is at most quadratic in the number of constants, atomic concepts, and roles. The crucial step is the check for coherence, which is performed once per loop. If this test is polynomial in the size of the input then the entire runtime of the algorithm is polynomial. For *DL-Lite*$_{\mathcal{FR}}$ TBoxes $\mathcal{T}$, coherence can be checked in time quadratic in the number of assertions in the TBox, that is, $O(|\mathcal{T}|^2)$. Satisfiability of an ABox $\mathcal{A}$ with respect to $\mathcal{T}$ can be checked in time $O(|\mathcal{T}|^2 \times |\mathcal{A}|)$, where $|\mathcal{A}|$ is the number of assertions of $\mathcal{A}$. The $O(|\mathcal{T}|^2)$ complexity can be shown by reduction to satisfiability of sets of propositional Horn clauses (see [22] for details).

INPUT:	KBs $\mathcal{K}$ and $\mathcal{N}$
OUTPUT:	a set $\mathcal{K}_m \subseteq cl(\mathcal{K})$ of TBox and ABox assertions
[1]	$\mathcal{K}_m := \mathcal{N}$; $\mathcal{S} := cl(\mathcal{K})$
[2]	**repeat**
[3]	**choose some** $\phi \in \mathcal{S}$; $\mathcal{S} := \mathcal{S} \setminus \{\phi\}$
[4]	**if** $\{\phi\} \cup \mathcal{K}_m$ is coherent **then** $\mathcal{K}_m := \mathcal{K}_m \cup \{\phi\}$
[5]	**until** $\mathcal{S} = \emptyset$

Fig. 1. Algorithm *BoldEvol*$(\mathcal{K}, \mathcal{N})$ for nondeterministic computation of $\mathcal{K}_m$

Theorem 10. *The algorithm BoldEvol runs in polynomial time and computes evolution wrt bold semantics, that is,* $\mathcal{K} \diamond_b \mathcal{N} = BoldEvol(\mathcal{K}, \mathcal{N})$.

This shows that bold semantics has the great advantage that evolution can be computed in polynomial time. However, its nondeterminism is a disadvantage. Clearly, we can *avoid* nondeterminism if we impose a linear order on the assertions in $cl(\mathcal{K})$, and let *BoldEvol* choose them in this order. The question how to define such an order depends on the characteristics of the application, and we cannot discuss it here.

One may wonder whether it is possible to efficiently compute a $\mathcal{K}_m$ with maximal *cardinality*. (Recall that our algorithm is only guaranteed to compute a $\mathcal{K}_m$ that is maximal wrt set inclusion.) Unfortunately, it turns out, using various reductions from the Independent Set problem, that under this requirement computation is hard, even for $\mathcal{K}$ and $\mathcal{N}$ that consist only of TBox or only of ABox assertions, except when both, $\mathcal{K}$ and $\mathcal{N}$, are ABoxes, in which case no conflicts can arise.

Theorem 11. *Given DL-Lite KBs $\mathcal{K}$ and $\mathcal{N}$ and an integer $k > 0$, to decide whether there exists a subset $\mathcal{K}_0 \subseteq \mathcal{K}$ such that $\mathcal{K}_0 \cup \mathcal{N}$ is coherent and $|\mathcal{K}_0| \geq k$ is* NP*-complete.* NP*-hardness already holds for DL-Lite$_{core}$ if (1) both $\mathcal{K}$ and $\mathcal{N}$ are TBoxes, or (2) $\mathcal{K}$ is an ABox and $\mathcal{N}$ is a TBox, or (3) $\mathcal{K}$ is a TBox and $\mathcal{N}$ is an ABox.*

In the next section we will see that nondeterminism is not present in ABox evolution with a protected TBox and that there is always a single maximal compatible ABox.

5 ABox Evolution

We study ABox evolution assuming that the new knowledge $\mathcal{N}_A$ is satisfiable with the old TBox $\mathcal{T}$, may only conflict with the old ABox $\mathcal{A}$, and that $\mathcal{T}$ is protected.

ABox Evolution under Bold Semantics. In *DL-Lite*, unsatisfiability of a KB is caused either by a single ABox assertion, which will be a membership assertion for an unsatisfiable concept or role, or by a pair of assertions contradicting either a disjointness or a functionality assertion of the TBox.

Lemma 12. *Let $\mathcal{T} \cup \mathcal{A}$ be a DL-Lite KB. If $\mathcal{T} \cup \mathcal{A}$ is unsatisfiable, then there is a subset $\mathcal{A}_0 \subseteq \mathcal{A}$ with at most two elements, such that $\mathcal{T} \cup \mathcal{A}_0$ is unsatisfiable.*

The lemma implies that if $\mathcal{T} \cup \mathcal{N}_A \cup \mathcal{A}$ is unsatisfiable, then there are two assertions $\phi \in \mathcal{N}_A$ and $\psi \in \mathcal{A}$ such that $\mathcal{T} \cup \{\phi, \psi\}$ is unsatisfiable. In other words, whether or not $\psi \in \mathcal{A}$ needs to be eliminated from $\mathcal{A}$ as a result of evolution depends on ψ alone. As a consequence, ABox evolution wrt bold semantics is deterministic.

INPUT: TBox $\mathcal{T}$, and ABoxes $\mathcal{A}$, $\mathcal{D}$, each satisfiable with $\mathcal{T}$
OUTPUT: finite set of membership assertions $\mathcal{A}^w$
$\mathcal{A}^w := \mathcal{A}$
for each $B_1(c) \in \mathcal{D}$ **do**
  $\mathcal{A}^w := \mathcal{A}^w \setminus \{B_1(c)\}$ **and**
  for each $B_2 \sqsubseteq B_1 \in cl(\mathcal{T})$ **do** $\mathcal{A}^w := \mathcal{A}^w \setminus \{B_2(c)\}$
    if $B_2(c) = \exists R(c)$ **then**
      for each $R(c,d) \in \mathcal{A}^w$ **do** $\mathcal{D} := \mathcal{D} \cup \{R(c,d)\}$
for each $R_1(a,b) \in \mathcal{D}$ **do**
  $\mathcal{A}^w := \mathcal{A}^w \setminus \{R_1(a,b)\}$ **and**
  for each $R_2 \sqsubseteq R_1 \in cl(\mathcal{T})$ **do** $\mathcal{A}^w := \mathcal{A}^w \setminus \{R_2(a,b)\}$

Fig. 2. Algorithm *Weeding*$(\mathcal{T}, \mathcal{A}, \mathcal{D})$ for *DL-Lite*$_{\mathcal{FR}}$

INPUT: TBox $\mathcal{T}$, and ABoxes $\mathcal{A}$, $\mathcal{N}_A$, each satisfiable with $\mathcal{T}$
OUTPUT: finite set of membership assertions $\mathcal{A}_m$
$\mathcal{A}_0 := cl_{\mathcal{T}}(\mathcal{A} \cup \mathcal{N}_A), \mathcal{N}_A := cl_{\mathcal{T}}(\mathcal{N}_A),\ CA := \emptyset$
for each $B \sqsubseteq \neg B' \in cl(\mathcal{T})$ **do**
  if $\{B(c), B'(c)\} \subseteq \mathcal{A}_0$ **then**
    if $B(c) \notin \mathcal{N}_A$ **then** $CA := CA \cup \{B(c)\}$
      otherwise $CA := CA \cup \{B'(c)\}$
for each $(\mathsf{funct}\ R) \in \mathcal{T}$ **do**
  if $\{R(a,b), R(a,c)\} \subseteq \mathcal{A}_0$ **then**
    if $R(a,b) \notin \mathcal{N}_A$ **then** $CA := CA \cup \{R(a,b)\}$
      otherwise $CA := CA \cup \{R(a,c)\}$
$\mathcal{A}_m :=$ *Weeding*$(\mathcal{T}, cl_T(\mathcal{A}), CA)$

Fig. 3. Algorithm *FastEvol*$(\mathcal{A}, \mathcal{N}_A, \mathcal{T})$ for *DL-Lite*$_{\mathcal{FR}}$

Theorem 13. *The result of ABox evolution* $(\mathcal{T} \cup \mathcal{A}) \diamond_b \mathcal{N}_A$ *is uniquely defined.*

In principle, *BoldEvol* can be used to compute ABox evolution and regardless of the order in which it selects the assertions, it will always return the same result, due to Theorem 13. A drawback of *BoldEvol* is that it performs a coherence check during each loop, which is not needed in that form, since ABox evolution does not affect coherence of the TBox. We exhibit now a new algorithm *FastEvol* that replaces the coherence check with implicit satisfiability checks.

The algorithm *FastEvol* computes the set $\mathcal{A}_m \subseteq cl_{\mathcal{T}}(\mathcal{A})$ of all assertions that do not conflict with $\mathcal{T}$ and $\mathcal{N}$ and is based on Lemma 12. It exploits the algorithm *Weeding* (see Fig. 2), which takes as input $\mathcal{T}$, $\mathcal{A}$, and a set $\mathcal{D}$ of membership assertions to be "deleted" from $\mathcal{A}$. For every assertion $\phi \in \mathcal{D}$, *Weeding* deletes from $\mathcal{A}$ ϕ and also all the assertions that $\mathcal{T}$-entail ϕ. The algorithm *FastEvol* (see Fig. 3) takes as input $\mathcal{T}$, $\mathcal{A}$, and $\mathcal{N}_A$. It detects assertions in the closure of $\mathcal{A} \cup \mathcal{N}_A$ that *conflict* with the new data and stores them in CA. Finally, it resolves the conflicts by deleting CA from $\mathcal{A}$ using *Weeding*.

Theorem 14. *The algorithm FastEvol computes ABox evolution wrt bold semantics, that is,* $(\mathcal{T} \cup \mathcal{A}) \diamond_b \mathcal{N}_A = \mathcal{T} \cup \mathcal{N}_A \cup$ *FastEvol*$(\mathcal{K}, \mathcal{N}_A)$*, and runs in polynomial time.*

Note that, although *FastEvol* may look similar to the algorithm *ComputeUpdate* in [21], it is actually different. Our algorithm always keeps at least as many assertions as

ComputeUpdate. In some cases, however, *ComputeUpdate* drops an existential restriction of the form $\exists R(a)$, although it would not cause a contradiction.

Careful Semantics. We start with an example illustrating drawbacks of bold semantics. Apparently, the drawbacks come from the minimality of evolution principle **EP3**.

Example 15. Coming back to $\mathcal{K}_{ex}$, consider evolution wrt bold semantics for the news that John is getting single, formally, $\mathcal{N}_A = \{S(j)\}$. One can see that the only assertion to be dropped from $\mathcal{K}_{ex}$ is that John is the husband of Mary, that is, $\mathcal{K}_{ex} \diamond_b \mathcal{N}_A = \mathcal{K}_{ex} \cup \mathcal{N}_A \setminus \{hh(m, j)\}$. This implies that $\mathcal{K}_{ex} \diamond_b \mathcal{N}_A \models W(m)$ and, consequently, Mary *still has a husband* who is not John, despite the divorce with John, that is, $\mathcal{K}_{ex} \diamond_b \mathcal{N}_A \models \phi$, where $\phi = \exists x(hh(m, x) \wedge (x \neq j))$. The only option that bold semantics offers to Mary is to find another husband immediately after the divorce. It does not consider it an option for her to become single. We are interested in a semantics that allows for both possibilities. Note that the entailment $\mathcal{K}_{ex} \diamond_b \mathcal{N}_A \models \phi$ is *unexpected* in the sense that neither $\mathcal{K}_{ex}$ nor $\mathcal{N}_A$ entail ϕ, that is, $\mathcal{K}_{ex} \not\models \phi$ and $\mathcal{N}_A \not\models \phi$ hold. ■

As the example shows, the situation when the result of evolution entails *unexpected information*, that is, information coming neither from the original KB, nor from the new knowledge, may be counterintuitive. In our example, the unexpected information is the formula $\exists x(hh(m, x) \wedge (x \neq j))$, which has a specific form: it restricts the possible values in the second component of the role hh. Our next semantics prohibits these role restrictions from being unexpectedly entailed from the result of evolution.

We say that a formula is *role-constraining*, or an *RCF* for short, if it is of the form $\exists x(R(a, x) \wedge (x \neq c_1) \wedge \cdots \wedge (x \neq c_n))$, where a and all c_i are constants. Let $\mathcal{T}$ be a TBox, and $\mathcal{A}, \mathcal{N}_A$ be ABoxes. A subset $\mathcal{A}_1 \subseteq \mathcal{A}$ is *careful* if for every RCF φ, whenever $\mathcal{A}_1 \cup \mathcal{N}_A \models_{\mathcal{T}} \varphi$ holds, either $\mathcal{A}_1 \models_{\mathcal{T}} \varphi$ or $\mathcal{N}_A \models_{\mathcal{T}} \varphi$ holds.

Theorem 16. *Let $\mathcal{T}$ be a DL-Lite TBox and $\mathcal{A}$, $\mathcal{N}_A$ DL-Lite ABoxes, and suppose that both $\mathcal{T} \cup \mathcal{A}$ and $\mathcal{T} \cup \mathcal{N}_A$ are satisfiable. Then, the set*

$$\{\mathcal{A}_0 \subseteq cl_{\mathcal{T}}(\mathcal{A}) \mid \mathcal{A}_0 \textit{ is careful and } \mathcal{A}_0 \cup \mathcal{N}_A \textit{ is } \mathcal{T}\textit{-satisfiable}\}$$

has a unique maximal element wrt set inclusion.

We can exploit the maximal set $\mathcal{A}_m^c$ of assertions (where c stands for careful), whose uniqueness is guaranteed by Theorem 16, to define the *careful evolution*:

$$(\mathcal{T} \cup \mathcal{A}) \diamond_c \mathcal{N}_A \ := \ \mathcal{T} \cup \mathcal{N}_A \cup \mathcal{A}_m^c. \tag{4}$$

One can see that, by its definition, careful semantics satisfies the principles **EP1**, **EP2**, and **EP3**, where for **EP3** the minimality should take into account carefulness. We exhibit now the algorithm *CarefulEvol*, which computes the uniquely determined set $\mathcal{A}_m^c$ of Equation (4). The *preclosure of $\mathcal{A}$ wrt $\mathcal{T}$*, denoted $Precl_{\mathcal{T}}(\mathcal{A})$, is a subset of $cl_{\mathcal{T}}(\mathcal{A})$ obtained as follows: one removes from $cl_{\mathcal{T}}(\mathcal{A})$ all the assertions of the form $\exists R(a)$, whenever there is an assertion of the form $R(a, c)$ in $cl_{\mathcal{T}}(\mathcal{A})$, for some constant c. The preclosure is needed to detect unexpected RCFs. The algorithm *CarefulEvol*

INPUT: TBox $\mathcal{T}$, and ABoxes $\mathcal{A}, \mathcal{N}_A$, each satisfiable with $\mathcal{T}$
OUTPUT: finite set of membership assertions $\mathcal{A}^c_m$

$\mathcal{A}^c_m := \mathit{FastEvol}(\mathcal{T}, \mathcal{A}, \mathcal{N}_A)$, $\mathit{UF} := \emptyset$
for each $\exists R(a) \in \mathit{Precl}_{\mathcal{T}}(\mathcal{N}_A)$ **do**
  if $R(a,b) \notin \mathcal{A}^c_m$ for every b **then**
    for each $\exists R^- \sqsubseteq \neg C \in \mathit{cl}(\mathcal{T})$ **do**
      for each $C(d) \in \mathit{cl}_{\mathcal{T}}(\mathcal{A}) \setminus \mathit{cl}_{\mathcal{T}}(\mathcal{N}_A)$ **do**
        $\mathit{UF} := \mathit{UF} \cup \{C(d)\}$
for each $\exists R(a) \in \mathcal{A}^c_m \setminus \mathit{Precl}_{\mathcal{T}}(\mathcal{N}_A)$ **do**
  if $R(a,b) \notin \mathcal{A}^c_m$ for every b **then**
    if there is a concept C in $\mathcal{T} \cup \mathcal{A} \cup \mathcal{N}_A$ s.t.
      $(\exists R^- \sqsubseteq \neg C) \in \mathit{cl}(\mathcal{T})$ **and** $C(d) \in \mathit{cl}_{\mathcal{T}}(\mathcal{N}_A) \setminus \mathit{cl}_{\mathcal{T}}(\mathcal{A})$ for some d **then**
        $\mathit{UF} := \mathit{UF} \cup \{\exists R(a)\}$
$\mathcal{A}^c_m := \mathit{Weeding}(\mathcal{T}, \mathcal{A}^c_m, \mathit{UF})$

Fig. 4. Algorithm *CarefulEvol*$(\mathcal{T}, \mathcal{A}, \mathcal{N}_A)$ for *DL-Lite*$_{\mathcal{FR}}$

(see Fig. 4) takes as input ABoxes $\mathcal{A}$, $\mathcal{N}_A$, and a TBox $\mathcal{T}$. It first computes the evolution wrt bold semantics. Then, it computes the set *UF* of assertions that cause *unexpectedness* in *FastEvol*$(\mathcal{T}, \mathcal{A}, \mathcal{N}_A)$ and belong to $\mathit{cl}_{\mathcal{T}}(\mathcal{A})$. Then it removes *UF* from *FastEvol*$(\mathcal{T}, \mathcal{A}, \mathcal{N}_A)$ by means of *Weeding*.

Theorem 17. *The algorithm CarefulEvol computes ABox evolution wrt careful semantics, that is,* $(\mathcal{T} \cup \mathcal{A}) \diamond_c \mathcal{N}_A = \mathcal{T} \cup \mathcal{N}_A \cup \mathit{CarefulEvol}(\mathcal{T}, \mathcal{A}, \mathcal{N}_A)$ *and runs in polynomial time.*

Again, *CarefulEvol* differs from *ComputeUpdate* in [21]. Sometimes the first may drop an existential restriction $\exists R(a)$ and the second keep it, while sometimes it may be the other way round.

We defer a detailed discussion on how bold and careful semantics are related to the classical update and revision postulates of [4] to an extended version of this paper.

6 Conclusion

We studied evolution of *DL-Lite* KBs. There are two main families of approaches to evolution: model-based and formula-based ones. We singled out and investigated a three-dimensional space of model-based approaches, and proved that most of them are not appropriate for *DL-Lite* due to their counterintuitive behavior and the inexpressibility of evolution results. Thus, we examined formula-based approaches, showed that the classical ones are again inappropriate for *DL-Lite*, and proposed a novel bold semantics. We showed that this semantics can be computed in polynomial time, but the result is, in general, non-deterministic. Then, we studied ABox evolution under bold semantics and showed that the result in this case is unique. We developed a polynomial time algorithm for DL-Lite KB evolution under this semantics, and an alternative optimized one for ABox evolution. We presented a conceptual drawback of ABox evolution under bold semantics and introduced careful semantics, which repairs the drawback. For this approach we proved that the evolution result is unique and developed a polynomial time algorithm to compute it.

Acknowledgements. The authors are partially supported by the EU projects ACSI (FP7-ICT-257593) and Ontorule (FP7-ICT-231875). The second author is also supported by the European Research Council grant Webdam (under FP7), agreement n. 226513.

References

1. Borgida, A., Brachman, R.J.: Conceptual modeling with description logics. [13] chapter 10 349–372
2. Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D., Antoniou, G.: Ontology change: Classification and survey. Knowledge Engineering Review **23**(2) (2008) 117–152
3. Abiteboul, S., Grahne, G.: Update semantics for incomplete databases. In: Proc. of VLDB'85. (1985)
4. Katsuno, H., Mendelzon, A.: On the difference between updating a knowledge base and revising it. In: Proc. of KR'91. (1991) 387–394
5. Eiter, T., Gottlob, G.: On the complexity of propositional knowledge base revision, updates and counterfactuals. Artificial Intelligence **57** (1992) 227–270
6. Winslett, M.: Updating Logical Databases. Cambridge University Press (1990)
7. Flouris, G.: On belief change in ontology evolution. AI Communications **19**(4) (2006)
8. Qi, G., Du, J.: Model-based revision operators for terminologies in description logics. In: Proc. of IJCAI 2009. (2009) 891–897
9. Peters, R.J., Özsu, M.T.: An axiomatic model of dynamic schema evolution in objectbase systems. ACM Trans. on Database Systems **22**(1) (1997) 75–114
10. De Giacomo, G., Lenzerini, M., Poggi, A., Rosati, R.: On instance-level update and erasure in description logic ontologies. J. of Logic and Computation, Special Issue on Ontology Dynamics **19**(5) (2009) 745–770
11. Liu, H., Lutz, C., Milicic, M., Wolter, F.: Updating description logic ABoxes. In: Proc. of KR 2006. (2006) 46–56
12. Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Rosati, R.: Tractable reasoning and efficient query answering in description logics: The *DL-Lite* family. J. of Automated Reasoning **39**(3) (2007) 385–429
13. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.F., eds.: The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press (2003)
14. Artale, A., Calvanese, D., Kontchakov, R., Zakharyaschev, M.: The *DL-Lite* family and relations. J. of Artificial Intelligence Research **36** (2009) 1–69
15. Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., Rosati, R.: Linking data to ontologies. J. on Data Semantics **X** (2008) 133–173
16. Ginsberg, M.L., Smith, D.E.: Reasoning about action I: A possible worlds approach. Technical Report KSL-86-65, Knowledge Systems, AI Laboratory (1987)
17. Forbus, K.D.: Introducing actions into qualitative simulation. In: Proc. of IJCAI'89. (1989)
18. Borgida, A.: Language features for flexible handling of exceptions in information systems. ACM Trans. on Database Systems **10**(4) (1985) 565–603
19. Dalal, M.: Investigations into a theory of knowledge base revision. In: Proc. of AAAI'88. (1988) 475–479
20. Calvanese, D., Kharlamov, E., Nutt, W.: A proof theory for *DL-Lite*. In: Proc. of DL 2007. Volume 250 of CEUR, `ceur-ws.org`. (2007) 235–242
21. De Giacomo, G., Lenzerini, M., Poggi, A., Rosati, R.: On the update of description logic ontologies at the instance level. In: Proc. of AAAI 2006. (2006) 1271–1276
22. Zheleznyakov, D., Calvanese, D., Kharlamov, E., Nutt, W.: Updating TBoxes in *DL-Lite*. In: Proc. of DL 2010. Volume 573 of CEUR, `ceur-ws.org`. (2010) 102–113