FREIE UNIVERSITÄT BOZEN
LIBERA UNIVERSITÀ DI BOLZANO
FREE UNIVERSITY OF BOZEN · BOLZANO

KRDB Research Centre
for Knowledge and Data

Faculty of Computer Science, Free University of Bozen-Bolzano, Piazza Domenicani 3, 39100 Bolzano, Italy
Tel: +39 04710 16000, fax: +39 04710 16009

KRDB Research Centre Technical Report:

Querying Attributed DL-Lite Ontologies using Provenance Semirings

Camille Bourgaux[1], Ana Ozaki[2]

Affiliation	1: Télécom ParisTech & DI ENS, CNRS, ENS, PSL University & Inria 2: KRDB Research Centre, Free University of Bozen-Bolzano
Corresponding authors	Camille Bourgaux: camille.bourgaux@ens.fr Ana Ozaki: ana.ozaki@unibz.it
Keywords	Description Logic, Provenance Semirings, Ontology-based Data Access, Complexity, Query Answering
Number	KRDB18-02
Date	November 13, 2018
URL	http://www.inf.unibz.it/krdb/

Abstract

Attributed description logic is a recently proposed formalism, targeted for graph-based representation formats, which enriches description logic concepts and roles with finite sets of attribute-value pairs, called annotations. One of the most important uses of annotations is to record provenance information. In this work, we first investigate the complexity of satisfiability and query answering for attributed DL-Lite ontologies. We then propose a new semantics, based on provenance semirings, for integrating provenance information with query answering. Finally, we establish complexity results for satisfiability and query answering under this semantics.

Acknowledgments

Partially supported by ANR-16-CE23-0007-01 (“DICOS”).

Querying Attributed DL-Lite Ontologies using Provenance Semirings

Camille Bourgaux
Télécom ParisTech & DI ENS, CNRS, ENS,
PSL University & Inria, France

Ana Ozaki
KRDB Research Centre,
Free University of Bozen-Bolzano, Italy

Abstract

Attributed description logic is a recently proposed formalism, targeted for graph-based representation formats, which enriches description logic concepts and roles with finite sets of attribute-value pairs, called annotations. One of the most important uses of annotations is to record provenance information. In this work, we first investigate the complexity of satisfiability and query answering for attributed DL-Lite$_\mathcal{R}$ ontologies. We then propose a new semantics, based on provenance semirings, for integrating provenance information with query answering. Finally, we establish complexity results for satisfiability and query answering under this semantics.

Introduction

Description logic (DL) (Baader et al. 2007) ontologies allow to express complex relationships between concepts and roles, but they are ill-equipped to represent and reason about multiple and heterogeneous types of meta-knowledge, such as the temporal validity of a fact, or its source. For instance, the YAGO ontology (Hoffart et al. 2013) attaches provenance metadata to its facts (e.g., source and confidence of the extraction) as well as temporal and geospatial information. Many practical applications therefore use *knowledge graphs*, which consist, like DL assertions, of directed labelled graphs but that also allow, unlike DLs, to add annotations to vertices and edges. Property Graph, the data model used in many graph databases (Rodriguez and Neubauer 2010), and Wikidata, the knowledge graph used by Wikipedia (Vrandečić and Krötzsch 2014), are prominent examples of such labelled graphs. To bridge the gap between DL and knowledge graphs, *attributed description logics* (Krötzsch et al. 2017; Krötzsch et al. 2018) have been recently introduced. They enrich DL concepts and roles with finite sets of attribute-value pairs, called *annotations*, and allow to express constraints on these annotations in the ontology inclusions. For example, the attributed DL assertion $\mathsf{spouse}(\mathsf{taylor}, \mathsf{burton})@[\mathsf{start} : 1975, \mathsf{end} : 1976]$ states that Liz Taylor was married to Richard Burton from 1975 to 1976, and the following role inclusion expresses that spouse is a symmetric relation, where the inverse statement has the same start and end dates:

$$\mathsf{spouse}@X \sqsubseteq \mathsf{spouse}^-@[\mathsf{start} : X.\mathsf{start}, \mathsf{end} : X.\mathsf{end}].$$

While the work by Krötzsch et al. studied the complexity of the satisfiability problem for several attributed DL languages, our focus in this paper is on *query answering* in attributed DL. The problem of querying DL ontologies using database-style queries (in particular, conjunctive queries) is an important research topic for which tractable DL languages have been tailored (Bienvenu and Ortiz 2015). We consider here the DL-Lite$_\mathcal{R}$ dialect of the *DL-Lite family* (Calvanese et al. 2007), which underlies the OWL 2 QL profile (Motik et al. 2009), and investigate attributed DL-Lite$_\mathcal{R}$.

One of the main motivations of attributed DLs is to integrate annotations carrying provenance information, which are very frequent in knowledge graphs[1]. Recording and tracking provenance information is an important topic in database theory, where *provenance semirings* (Green, Karvounarakis, and Tannen 2007) were introduced as an abstract tool to relate the result of a query with information about the original sources of the data and the ways in which the query was obtained. Such information comes in the form of a *provenance polynomial*. It has been useful for many applications, such as query answer explanation or querying of probabilistic databases (Senellart 2017; Cheney, Chiticariu, and Tan 2009; Suciu et al. 2011). Bienvenu, Deutch, and Suchanek (2012) argued that provenance would be useful for Web data, e.g., to establish the authorship or determine the trust in a given piece of data, or to help to guarantee the privacy of information. Provenance has also been investigated for non-relational databases and Semantic Web (see Conclusion for discussion of related work). In this work, we propose a new semantics for the attributed DL annotations, based on provenance semirings, so that queries can be annotated with provenance polynomials. To the best of our knowledge, this is the first work where provenance polynomials are embedded into both the syntax and the semantics of the query.

The first section introduces attributed DL-Lite$_\mathcal{R}$, following the formalism given by Krötzsch et al. (2017; 2018). We then define attributed conjunctive queries and study the complexity of satisfiability and query answering in attributed DL-Lite$_\mathcal{R}$. We next present our new semantics for the annotations to model provenance and analyse the complexity of satisfiability and query answering with this new model, considering

[1] E.g., in Wikidata *reference* (provenance) is one the most frequent types of annotations `https://www.wikidata.org`.

queries that can be annotated with provenance polynomials. In particular, we show that satisfiability and query answering in attributed DL-Lite$_\mathcal{R}$ are PSPACE-complete problems. For the semirings-based semantics and queries annotated with provenance polynomials, we establish that although satisfiability is EXPTIME-hard in the general case, the new semantics does not increase the complexity of query answering if the ontology contains only assertions and a restricted form of inclusions, which is close to the database setting considered by Green, Karvounarakis, and Tannen (2007). We also investigate various restrictions of the general setting. Our results are for *combined complexity*, when both the query and the ontology are considered as the input. Proofs are available in the appendix.

Attributed DL-Lite

Attributed DLs are defined over the usual DL signature with countable sets of *concept names* $\mathsf{N_C}$, *role names* $\mathsf{N_R}$, and *individual names* $\mathsf{N_I}$. We consider an additional set $\mathsf{N_U}$ of *set variables* and a set $\mathsf{N_V}$ of *object variables*. Annotation sets are defined as finite binary relations, understood as sets of attribute-value pairs. Attributes and values refer to domain elements and are syntactically denoted by individual names. To describe annotation sets, we use *specifiers*. The set $\mathbf{S}$ of specifiers contains the following expressions:

- set variables $X \in \mathsf{N_U}$;
- *closed specifiers* $[a_1 : v_1, \ldots, a_n : v_n]$; and
- *open specifiers* $\lfloor a_1 : v_1, \ldots, a_n : v_n \rfloor$,

where $a_i \in \mathsf{N_I}$ and v_i is either an individual name in $\mathsf{N_I}$, an object variable in $\mathsf{N_V}$, or an expression of the form $X.a$, with X a set variable in $\mathsf{N_U}$ and a an individual name in $\mathsf{N_I}$. We use $X.a$ to refer to the (finite, possibly empty) set of all values of attribute a in an annotation set X. A *ground specifier* is a closed or open specifier that only contains individual names. Intuitively, closed specifiers define specific annotation sets whereas open specifiers merely provide lower bounds (Krötzsch et al. 2017).

Syntax. A DL-Lite$^\mathcal{R}_@$ *role* (resp. *concept*) *assertion* is an expression $R(a, b)@S$ (resp. $A(a)@S$), with $R \in \mathsf{N_R}$ (resp. $A \in \mathsf{N_C}$), $a, b \in \mathsf{N_I}$, and $S \in \mathbf{S}$ a ground closed specifier. DL-Lite$^\mathcal{R}_@$ *role* and *concept inclusions* are of the form $X : S\ (P \sqsubseteq Q)$ and $X : S\ (B \sqsubseteq C)$ respectively, where $X \in \mathsf{N_U}$, $S \in \mathbf{S}$ is a closed or open specifier, and P, Q and B, C are respectively role and concept expressions defined by the following syntax, where $A \in \mathsf{N_C}$, $R \in \mathsf{N_R}$ and $S \in \mathbf{S}$:

$$P ::= R@S \mid R^-@S, \qquad Q ::= P \mid \neg P,$$
$$B ::= A@S \mid \exists P, \qquad C ::= B \mid \neg B.$$

We further require that all variables are *safe*. For set variables, this means that if $Y \in \mathsf{N_U}$ occurs on the right side of an inclusion (or in a specifier S such that the prefix of the inclusion is $X : S$ and X occurs on the right side), then the specifier of the left side expression is Y. For object variables, if they occur on the right side of an inclusion then they must also occur on the left side or in S such that $X : S$ and X occurs on the left. Note that if object variables occur in S with $X : S$ in the prefix and X on the right side, then X is the specifier on the left by the safety definition. If the prefix of an inclusion is $X : S$ and X does not occur in the role/concept expressions of the inclusion, we may ommit $X : S$.

A DL-Lite$^\mathcal{R}_@$ *ontology* is a set of DL-Lite$^\mathcal{R}_@$ assertions, role and concept inclusions. Also, we say that a DL-Lite$^\mathcal{R}_@$ ontology is *ground* if it does not contain variables. To simplify notation, we omit the specifier $\lfloor\rfloor$ (meaning "any annotation set") in role or concept expressions. In this sense, any DL-Lite$_\mathcal{R}$ axiom is also a DL-Lite$^\mathcal{R}_@$ axiom. Moreover, we omit prefixes of the form $X : \lfloor\ \rfloor$, which state that there is no restriction on X. The *size* of an ontology $\mathcal{O}$ (or a query, defined later), which we may denote with $|\mathcal{O}|$, is the length of the string that represents it.

Example 1. *Our running example's ontology $\mathcal{O}_{\mathsf{ex}}$ expresses that those who are married (role* spouse*) to someone are married (concept* Married*), annotated with the same sources from which the information has been extracted (attribute* src*):*

$$\exists \mathsf{spouse}@X \sqsubseteq \mathsf{Married}@\lfloor \mathsf{src} : X.\mathsf{src} \rfloor.$$

The assertion states that Zsa Zsa Gabor was married to Jack Ryan and it is annotated with the sources of this information:

$$\mathsf{spouse}(\mathsf{gabor}, \mathsf{ryan})@[\mathsf{src} : \mathsf{s_1}, \mathsf{src} : \mathsf{s_2}].$$

Semantics. An *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ of an attributed DL consists of a non-empty domain $\Delta^\mathcal{I}$ and a function $\cdot^\mathcal{I}$. Individual names $a \in \mathsf{N_I}$ are interpreted as elements $a^\mathcal{I} \in \Delta^\mathcal{I}$. To interpret annotation sets, we use the set $\Phi^\mathcal{I} := \{\Sigma \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I} \mid \Sigma \text{ is finite } \}$ of all finite binary relations over $\Delta^\mathcal{I}$. Each concept name $A \in \mathsf{N_C}$ is interpreted as a set $A^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Phi^\mathcal{I}$ of elements with annotations, and each role name $R \in \mathsf{N_R}$ is interpreted as a set $R^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I} \times \Phi^\mathcal{I}$ of pairs of elements with annotations. Each element (pair of elements) may appear with multiple different annotations.

$\mathcal{I}$ *satisfies* a concept assertion $A(a)@[a_1 : v_1, \ldots, a_n : v_n]$ if $(a^\mathcal{I}, \{(a_1^\mathcal{I}, v_1^\mathcal{I}), \ldots, (a_n^\mathcal{I}, v_n^\mathcal{I})\}) \in A^\mathcal{I}$. Role assertions are interpreted analogously. Expressions with free set or object variables are interpreted using variable assignments $\mathcal{Z}$ mapping object variables $x \in \mathsf{N_V}$ to elements $\mathcal{Z}(x) \in \Delta^\mathcal{I}$ and set variables $X \in \mathsf{N_U}$ to finite binary relations $\mathcal{Z}(X) \in \Phi^\mathcal{I}$. For convenience, we also extend variable assignments to individual names, setting $\mathcal{Z}(a) = a^\mathcal{I}$ for every $a \in \mathsf{N_I}$. A specifier $S \in \mathbf{S}$ is interpreted as a set $S^{\mathcal{I},\mathcal{Z}} \subseteq \Phi^\mathcal{I}$ of matching annotation sets. We set $X^{\mathcal{I},\mathcal{Z}} := \{\mathcal{Z}(X)\}$ for variables $X \in \mathsf{N_U}$. The semantics of closed specifiers is defined as:

- $[a : v]^{\mathcal{I},\mathcal{Z}} := \{\{(a^\mathcal{I}, \mathcal{Z}(v))\}\}$ where $v \in \mathsf{N_I} \cup \mathsf{N_V}$;
- $[a : X.b]^{\mathcal{I},\mathcal{Z}} := \{\{(a^\mathcal{I}, \delta) \mid (b^\mathcal{I}, \delta) \in \mathcal{Z}(X)\}\}$;
- $[a_1 : v_1, \ldots, a_n : v_n]^{\mathcal{I},\mathcal{Z}} := \{\bigcup_{i=1}^n F_i \mid F_i \in [a_i : v_i]^{\mathcal{I},\mathcal{Z}}\}$.

$S^{\mathcal{I},\mathcal{Z}}$ therefore is a singleton set for set variables and closed specifiers. For open specifiers, however, we define $\lfloor a_1 : v_1, \ldots, a_n : v_n \rfloor^{\mathcal{I},\mathcal{Z}}$ to be the set:

$$\{F \subseteq \Phi^\mathcal{I} \mid F \supseteq G \text{ for } \{G\} = [a_1 : v_1, \ldots, a_n : v_n]^{\mathcal{I},\mathcal{Z}}\}.$$

Now given $A \in \mathsf{N_C}$, $R \in \mathsf{N_R}$, and $S \in \mathbf{S}$, we define:

$(A@S)^{\mathcal{I},\mathcal{Z}} := \{\delta \mid (\delta, F) \in A^\mathcal{I} \text{ for some } F \in S^{\mathcal{I},\mathcal{Z}}\}$,

$(R@S)^{\mathcal{I},\mathcal{Z}} := \{(\delta, \epsilon) \mid (\delta, \epsilon, F) \in R^\mathcal{I} \text{ for some } F \in S^{\mathcal{I},\mathcal{Z}}\}$.

Further DL expressions are defined as usual: $(R^{-}@S)^{\mathcal{I},\mathcal{Z}} = \{(\gamma,\delta) \mid (\delta,\gamma) \in (R@S)^{\mathcal{I},\mathcal{Z}}\}$, $\neg P^{\mathcal{I},\mathcal{Z}} = (\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}) \setminus P^{\mathcal{I},\mathcal{Z}}$, $\exists P^{\mathcal{I},\mathcal{Z}} = \{\delta \mid \text{there is } (\delta,\epsilon) \in P^{\mathcal{I},\mathcal{Z}}\}$, $\neg C^{\mathcal{I},\mathcal{Z}} = \Delta^{\mathcal{I}} \setminus C^{\mathcal{I},\mathcal{Z}}$. $\mathcal{I}$ *satisfies* a concept inclusion $X : S\ (B \sqsubseteq C)$ if, for all variable assignments $\mathcal{Z}$ that satisfy $\mathcal{Z}(X) \in S^{\mathcal{I},\mathcal{Z}}$, we have $B^{\mathcal{I},\mathcal{Z}} \subseteq C^{\mathcal{I},\mathcal{Z}}$. Satisfaction of role inclusions is defined analogously. An interpretation $\mathcal{I}$ satisfies an ontology $\mathcal{O}$, or is a *model* of $\mathcal{O}$, if it satisfies all of its axioms. As usual, $\models$ denotes the induced logical entailment relation.

Example 2 (Example 1 cont'd)**.** *Consider an interpretation $\mathcal{I}$ with domain $\Delta^{\mathcal{I}} = \{\mathsf{gabor}, \mathsf{ryan}, \mathsf{src}, \mathsf{s}_1, \mathsf{s}_2\}$ and such that $\cdot^{\mathcal{I}}$ maps each individual name to itself and*

$$\mathsf{spouse}^{\mathcal{I}} = \{(\mathsf{gabor}, \mathsf{ryan}, \{(\mathsf{src}, \mathsf{s}_1), (\mathsf{src}, \mathsf{s}_2)\})\}$$
$$\mathsf{Married}^{\mathcal{I}} = \{(\mathsf{gabor}, \{(\mathsf{src}, \mathsf{s}_1), (\mathsf{src}, \mathsf{s}_2)\})\}.$$

The interpretation $\mathcal{I}$ is a model of $\mathcal{O}_{\mathsf{ex}}$.

Reasoning in DL-Lite$^{\mathcal{R}}_{@}$

In this section, we study the complexity of satisfiability and query answering over DL-Lite$^{\mathcal{R}}_{@}$ ontologies. Our first result is that the satisfiability problem, which is in NL for DL-Lite$_{\mathcal{R}}$ (Artale et al. 2009), is harder for DL-Lite$^{\mathcal{R}}_{@}$. The proof is by reduction from the word problem for polynomially space bounded deterministic Turing Machines (DTM). Annotations raise the complexity because they can encode configurations of a DTM, using expressions of the form $X.b$ to encode the synchronization of successive configurations.

Theorem 1. *In DL-Lite$^{\mathcal{R}}_{@}$, satisfiability is* PSPACE*-hard.*

To prove the PSPACE upper bound for satisfiability, we use *grounding* (Krötzsch et al. 2017), which is a classical technique that consists in eliminating variables from an ontology to transform it into an equisatisfiable ground ontology. The ground ontology can then be translated into an equisatisfiable DL-Lite$_{\mathcal{R}}$ ontology. The grounding leads to an exponential blowup of the ontology while the translation to DL-Lite$_{\mathcal{R}}$ is polynomial. Since satisfiability of DL-Lite$_{\mathcal{R}}$ ontologies is in NL (Artale et al. 2009), it follows (by (Savitch 1970)) that satisfiability of DL-Lite$^{\mathcal{R}}_{@}$ ontologies is in PSPACE.

Theorem 2. *In DL-Lite$^{\mathcal{R}}_{@}$, satisfiability is in* PSPACE*.*

We now turn our attention to the problem of querying DL-Lite$^{\mathcal{R}}_{@}$ ontologies. In the following we only define and deal with conjunctive queries without free variables, i.e., boolean conjunctive queries (BCQ), as the problem of finding certain answers to a query is reducible to BCQ entailment.

Definition 1 (Attributed Queries)**.** *An* attributed boolean conjunctive query *(BCQ$_@$) q is an expression of the form:*

$$\exists \boldsymbol{x}.X_1 : S_1, \ldots, X_n : S_n\ \varphi(\boldsymbol{x}) \quad (1)$$

where, for $1 \leq i \leq n$, X_i are the set variables occurring in $\varphi(\boldsymbol{x})$, $S_i \in \mathbf{S}$, and $\varphi(\boldsymbol{x})$ is a conjunction of atoms of the form $A(t)@S$ or $R(t,u)@S$, with $A \in \mathsf{N_C}$, $R \in \mathsf{N_R}$, $S \in \mathbf{S}$, and t, u individual names in $\mathsf{N_I}$ or variables in $\boldsymbol{x} \subseteq \mathsf{N_V}$.

We may write $E(\boldsymbol{t})@S$ to refer to an atom of any of the two forms ($E \in \mathsf{N_C} \cup \mathsf{N_R}$ and $\boldsymbol{t}$ is a tuple of elements from $\mathsf{N_I} \cup \boldsymbol{x}$ of the arity of E).

An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ satisfies a BCQ$_@$ *q, written $\mathcal{I} \models q$, if there exists a variable assignment $\mathcal{Z}$ such that:*

- *$\mathcal{Z}(X_i) \in S_i^{\mathcal{I},\mathcal{Z}}$ for all $1 \leq i \leq n$; and*
- *$(\mathcal{Z}(\boldsymbol{t}), F) \in E^{\mathcal{I}}$ for some $F \in S^{\mathcal{I},\mathcal{Z}}$, for every atom $E(\boldsymbol{t})@S$ occurring in q.*

A BCQ$_@$ q is entailed *by $\mathcal{O}$, written $\mathcal{O} \models q$, iff q is satisfied by every model of $\mathcal{O}$. A BCQ$_@$ that consists of a single atom is an* attributed boolean atomic query *(BAQ$_@$). We say that a BCQ$_@$ is* ground *if it contains only ground specifiers.*

BCQ$_@$ can express conditions on annotations, for instance require that there exists an annotation set where a given attribute is present or has a specific value.

Example 3. *We modify $\mathcal{O}_{\mathsf{ex}}$ to express that those who have a spouse are married, associated with the same annotations:*

$$\exists \mathsf{spouse}@X \sqsubseteq \mathsf{Married}@X.$$

We also add assertions stating that Zsa Zsa Gabor was married to Jack Ryan from 1975 *to* 1976*, while Liz Taylor was married to Richard Burton from* 1975 *to* 1976*, as well as the sources of this information:*

$$\mathsf{spouse}(\mathsf{gabor}, \mathsf{ryan})@[\mathsf{start} : 1975, \mathsf{end} : 1976, \mathsf{src} : \mathsf{s}_1],$$
$$\mathsf{spouse}(\mathsf{gabor}, \mathsf{ryan})@[\mathsf{start} : 1975, \mathsf{end} : 1976, \mathsf{src} : \mathsf{s}_2],$$
$$\mathsf{spouse}(\mathsf{taylor}, \mathsf{burton})@[\mathsf{start} : 1975, \mathsf{end} : 1976, \mathsf{src} : \mathsf{s}_3].$$

The following query expresses that Gabor and Taylor were married (to someone) with the same start and end dates:

$$q_{\mathsf{ex}} = \exists xy\, \mathsf{Married}(\mathsf{gabor})@\lfloor \mathsf{start} : x, \mathsf{end} : y \rfloor \wedge \mathsf{Married}(\mathsf{taylor})@\lfloor \mathsf{start} : x, \mathsf{end} : y \rfloor.$$

By the semantics of DL-Lite$^{\mathcal{R}}_{@}$, it follows that $\mathcal{O}_{\mathsf{ex}} \models q_{\mathsf{ex}}$. This other query expresses that a set of sources that includes s_1 and is associated with Gabor's married status is also associated with Taylor's married status:

$$q'_{\mathsf{ex}} = X : \lfloor \mathsf{src} : \mathsf{s}_1 \rfloor\ \mathsf{Married}(\mathsf{gabor})@X \wedge \mathsf{Married}(\mathsf{taylor})@\lfloor \mathsf{src} : X.\mathsf{src} \rfloor.$$

By the semantics of DL-Lite$^{\mathcal{R}}_{@}$, it follows that $\mathcal{O}_{\mathsf{ex}} \not\models q'_{\mathsf{ex}}$.

While BCQ entailment is NP-complete in DL-Lite$_{\mathcal{R}}$, it follows from Theorem 1 that BAQ$_@$ entailment is already PSPACE-hard. Indeed, satisfiability can be reduced to BAQ$_@$ entailment: $\mathcal{O}$ is unsatisfiable iff $\mathcal{O} \models A(a)$ where A and a are respectively a concept and an individual name that do not occur in $\mathcal{O}$. We show PSPACE-completeness of BCQ$_@$ entailment by describing how to decide $\mathcal{O} \models q$ for a BCQ$_@$ q, using only polynomial space w.r.t. the size of $\mathcal{O}$ and q.

The main ingredients to prove our result are grounding, translation to DL-Lite$_{\mathcal{R}}$, and also *query rewriting*, a prominent query answering technique for DL-Lite$_{\mathcal{R}}$ in which the query is rewritten w.r.t. the concept and role inclusions, to be evaluated over the assertions as in the classical database setting. However, as the ground version of $\mathcal{O}$ is of exponential size and the number of rewritten queries is exponential, we do not compute them but instead guess the DL-Lite$_{\mathcal{R}}$ translation $\mathsf{dl}(q_{\mathcal{Z}})$ of a grounded version $q_{\mathcal{Z}}$ of q together with one of its rewritings q'. We can verify in NP that q' is entailed by the DL-Lite$_{\mathcal{R}}$ translation of the assertions of $\mathcal{O}$, in PTIME that $\mathsf{dl}(q_{\mathcal{Z}})$ is the DL-Lite$_{\mathcal{R}}$ translation of a grounded version of q, and in PSPACE that q' is indeed a rewriting of $\mathsf{dl}(q_{\mathcal{Z}})$. For this last step, we propose a non-deterministic adaptation of the rewriting algorithm PerfectRef for DL-Lite$_{\mathcal{R}}$ by Calvanese et al. (2007) that takes as input $\mathsf{dl}(q_{\mathcal{Z}})$, q' and $\mathcal{O}$. The main idea is to rewrite $\mathsf{dl}(q_{\mathcal{Z}})$ by guessing at each step an atom of the query together with a positive inclusion that would appear in the DL-Lite$_{\mathcal{R}}$ translation of the grounding of $\mathcal{O}$, thus avoiding the computation of the grounding of $\mathcal{O}$.

Theorem 3. *In DL-Lite$^{\mathcal{R}}_{@}$, BCQ$_{@}$ entailment is in* PSPACE.

The result of Theorem 3, which is for combined complexity, contrasts with the EXPTIME-hardness w.r.t. *data complexity* (only w.r.t. the data size) for MARPL, an attributed logic based on Datalog (Marx, Krötzsch, and Thost 2017). Finally, we show lower complexity bounds in the case of *ground* ontologies. Indeed, when $\mathcal{O}$ is ground, one can build a DL-Lite$_{\mathcal{R}}$ ontology of polynomial size w.r.t. the size of $\mathcal{O}$ that entails the DL-Lite$_{\mathcal{R}}$ translation of a grounded version of q if and only if $\mathcal{O} \models q$.

Theorem 4. *For ground DL-Lite$^{\mathcal{R}}_{@}$ ontologies, satisfiability is in* PTIME *and BCQ$_{@}$ entailment is* NP*-complete.*

Querying using Provenance Semirings

In this section, we investigate attributed DL in light of provenance semirings (Green, Karvounarakis, and Tannen 2007) and enhance the semantics of DL-Lite$^{\mathcal{R}}_{@}$ to deal with provenance information. Semirings generalize formalisms such as why-provenance, lineages used in view maintenance, or the lineage used by the Trio uncertain management system (Senellart 2017). The main motivation is to use annotations to answer questions such as "Where does the result come from?". Assuming that facts are annotated with their sources, we want to know which combinations of sources lead to the entailment of a query. Such annotations may represent various types of information, such as trust, probability, multiplicity or data classification (see Example 8).

Example 4 (Example 3 cont'd)**.** *The result of the query q_{ex} over the ontology $\mathcal{O}_{\mathsf{ex}}$ can be obtained from source s_3 together with any of $\mathsf{s}_1, \mathsf{s}_2$. Provenance semirings can formalize this information in the form of a* provenance polynomial*:*

$$(\mathsf{s}_1 + \mathsf{s}_2) \times \mathsf{s}_3.$$

The intuitive meaning is that $+$ corresponds to *alternative* use of data and $\times$ to *joint* use of data. The goal of this section is to embed the formalism of provenance semirings into the semantics of DL-Lite$^{\mathcal{R}}_{@}$, so that we can associate annotations using provenance polynomials to queries (e.g., associate the annotation $\mathsf{src} : (\mathsf{s}_1 + \mathsf{s}_2) \times \mathsf{s}_3$ to the query q_{ex} of Example 3).

We define DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ as an order-sorted version of DL-Lite$^{\mathcal{R}}_{@}$. Elements of different sorts correspond to sets of *individual names* $\mathsf{N_I}$, *provenance sums* $\mathsf{N_S}$ and *provenance polynomials* $\mathsf{N_P}$. We represent provenance polynomials with the *positive algebra provenance semiring* for $\mathsf{N_I}$, defined as the commutative semiring of polynomials with variables in $\mathsf{N_I}$ and coefficients from $\mathbb{N}$, with operations defined as usual: $\mathcal{K} = (\mathbb{N}[\mathsf{N_I}], +, \times, 0, 1)$. We denote by $\mathsf{N_P}$ the set of polynomials of $\mathcal{K}$ and by $\mathsf{N_S}$ the subset of $\mathsf{N_P}$ containing the sums of the commutative monoid $(\mathbb{N}[\mathsf{N_I}], +, 0)$. We thus have $\mathsf{N_I} \subseteq \mathsf{N_S} \subseteq \mathsf{N_P}$. We may use the symbols $\sum$ and $\prod$ to denote sum and product of elements in $\mathsf{N_P}$ (which will then also be in $\mathsf{N_P}$). Elements of $\mathsf{N_S}$ are used as values in the ontology specifiers while elements of $\mathsf{N_P}$ appear as values in the query specifiers. Non-linear polynomials indicate the use of several assertions to derive a query, while provenance sums indicate that a query can be derived from different sources.

Role and concept inclusions in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ are defined similarly as in DL-Lite$^{\mathcal{R}}_{@}$, with the only difference that we allow elements from $\mathsf{N_S}$ to be values of attributes in the specifiers. Concept and role assertions are defined as in DL-Lite$^{\mathcal{R}}_{@}$. The fact that we do not allow values from $\mathsf{N_S}$ in the assertions does not change the expressivity of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, since inclusions can enforce the entailments of such assertions.

Example 5. *The following concept inclusion restricts that of Example 3 by further requiring that the fact that someone has a spouse has to be associated both with s_1 and with s_2 to conclude that this person is married.*

$$X : \lfloor \mathsf{src} : \mathsf{s}_1 + \mathsf{s}_2 \rfloor \quad (\exists \mathsf{spouse}@X \sqsubseteq \mathsf{Married}@X)$$

Provenance Semantics. We now introduce the semantics of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, based on provenance sums. A *provenance-interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is such that $\cdot^{\mathcal{I}}$ maps polynomials a and b in $\mathsf{N_P}$ to the same element $a^{\mathcal{I}} = b^{\mathcal{I}}$ in $\Delta^{\mathcal{I}}$ if and only if they are mathematically equal[2]. We denote by $\Delta^{\mathcal{I}}_{I}$ the domain of individuals and by $\Delta^{\mathcal{I}}_{S}$ the domain of provenance sums, which are the subsets of $\Delta^{\mathcal{I}}$ corresponding to the image of elements in $\mathsf{N_I}$ and $\mathsf{N_S}$, respectively. Thus $\Delta^{\mathcal{I}}_{I} \subseteq \Delta^{\mathcal{I}}_{S} \subseteq \Delta^{\mathcal{I}}$. To capture the semantics of provenance sums we develop a notion of closure. Intuitively, if a fact is annotated with n sources then it should also be annotated with the sum of *any subset* of these sources, since the fact can be retrieved alternatively by any source from this subset. For example, assume $(\boldsymbol{a}, F_1), \ldots, (\boldsymbol{a}, F_n)$ are in the interpretation of a concept or a role name E. If there is $(\mathsf{src}^{\mathcal{I}}, \mathsf{s}_i^{\mathcal{I}})$ in each F_i and these annotation sets only differ by such pairs, then for each subset of $\{\mathsf{s}_1, \ldots, \mathsf{s}_n\}$, the interpretation of E should have $(\boldsymbol{a}, F_{\mathsf{s}})$ with F_{s} differing from F_i only by the pair $(\mathsf{src}^{\mathcal{I}}, \mathsf{s}^{\mathcal{I}})$, where s is the sum of the elements of the subset.

We say that $G, H \in \Phi^{\mathcal{I}}$ are *differentiated by* p in F if

$$F = G \setminus \{(p, a) \mid (p, a) \in G\} = H \setminus \{(p, b) \mid (p, b) \in H\}.$$

In this case, we denote by $G +^{p} H$ the set

$$F \cup \{(p, (a+b)^{\mathcal{I}}) | \{a, b\} \subseteq \mathsf{N_P}, (p, a^{\mathcal{I}}) \in G, (p, b^{\mathcal{I}}) \in H\}.$$

A sum of possibly more than two annotation sets differentiated by p may be denoted by $\sum^{p}_{1 \le i \le n} G_i$ and is unique by the commutative law. For $E \in \mathsf{N_C} \cup \mathsf{N_R}$, and $\boldsymbol{a}$ a tuple of the arity of E, we say that $G +^{p} H$ is *non-primitive* for $\boldsymbol{a}$ and $E^{\mathcal{I}}$ if $\{(\boldsymbol{a}, G), (\boldsymbol{a}, H)\} \subseteq E^{\mathcal{I}}$. We denote by $E^{p}_{\mathcal{I},\boldsymbol{a},F}$ the set of annotation sets G pairwise differentiated by p in $F \in \Phi^{\mathcal{I}}$ such that $(\boldsymbol{a}, G) \in E^{\mathcal{I}}$ with G primitive for $\boldsymbol{a}$ and $E^{\mathcal{I}}$.

Definition 2 (Closure of $E^{\mathcal{I}}$)**.** *$E^{\mathcal{I}}$ is closed under sum if for all tuples $\boldsymbol{a}$ (in $\Delta^{\mathcal{I}}$ or $\Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ according to the arity of E),*

[2] According to associative, commutative and distributive laws. E.g., $(a+b)^{\mathcal{I}} = (b+a)^{\mathcal{I}}$ by the commutative law.

$\{(\boldsymbol{a}, \sum_{G\in\sigma}^{p} G) \mid \sigma \subseteq E^{p}_{\mathcal{I},\boldsymbol{a},F}, \sigma \neq \emptyset\} \subseteq E^{\mathcal{I}}$ *for every* $p \in \Delta^{\mathcal{I}}$ *and every* $F \in \Phi^{\mathcal{I}}$.

A provenance-interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is *well-founded* if $E^{\mathcal{I}}$ is closed under sum for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$. For all $E \in \mathsf{N_C} \cup \mathsf{N_R}$ and $\boldsymbol{a}$ with elements in $\Delta^{\mathcal{I}}$, we also require that the *support* of $E^{\mathcal{I}}$ and $\boldsymbol{a}$ defined as $\{F \mid (\boldsymbol{a}, F) \in E^{\mathcal{I}}\}$ is finite. This ensures that the sum in Definition 2 is finite. An interpretation of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ is a well-founded provenance-interpretation. We denote by $\mathbf{S}_\mathsf{S}$ the set of specifiers defined in the same way as $\mathbf{S}$ except that we use $\mathsf{N_S}$ instead of $\mathsf{N_I}$ when defining values of attributes. The semantics of specifiers in $\mathbf{S}_\mathsf{S}$ is defined as expected following the definition given in the Section 'Attributed DL-Lite' and we use the same notions of satisfiability and entailment. In Definition 2 we consider all subsets of $E^{p}_{\mathcal{I},\boldsymbol{a},F}$ rather than the sum of its elements. This is to ensure monotonicity of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$. Otherwise, given for example $A(a)@[p : a]$ and $A(a)@[p : b]$ we would lose the entailment $A(a)@[p : a + b]$ by adding $A(a)@[p : c]$.

Example 6. *Consider the ontology* $\mathcal{O}$ *with the assertions* $\mathsf{spouse}(\mathsf{gabor}, \mathsf{ryan})@[\mathsf{src} : \mathsf{s}_1]$, $\mathsf{spouse}(\mathsf{gabor}, \mathsf{ryan})@[\mathsf{src} : \mathsf{s}_2]$ *and the concept inclusion of Example 5. Let* $\mathcal{I}$ *have domain* $\Delta^{\mathcal{I}} = \{\mathsf{gabor}, \mathsf{ryan}, \mathsf{src}, \mathsf{s}_1, \mathsf{s}_2, \mathsf{s}_1 + \mathsf{s}_2\}$, *interpret each individual name by itself,* $(\mathsf{s}_1 + \mathsf{s}_2)^{\mathcal{I}} = \mathsf{s}_1 + \mathsf{s}_2$, *and*

$$\mathsf{spouse}^{\mathcal{I}} = \{(\mathsf{gabor}, \mathsf{ryan}, G), (\mathsf{gabor}, \mathsf{ryan}, H), (\mathsf{gabor}, \mathsf{ryan}, G +^{\mathsf{src}} H)\}$$

$$\mathsf{Married}^{\mathcal{I}} = \{(\mathsf{gabor}, G +^{\mathsf{src}} H)\} \text{ where } G = \{(\mathsf{src}, \mathsf{s}_1)\}, H = \{(\mathsf{src}, \mathsf{s}_2)\} \text{ and } G +^{\mathsf{src}} H = \{(\mathsf{src}, \mathsf{s}_1 + \mathsf{s}_2)\}.$$

$\mathsf{spouse}^{\mathcal{I}}$ *and* $\mathsf{Married}^{\mathcal{I}}$ *are closed under sum,* $\mathcal{I}$ *is a model of* $\mathcal{O}$ *and* $\mathcal{O} \models \mathsf{spouse}(\mathsf{gabor}, \mathsf{ryan})@\lfloor\mathsf{src} : \mathsf{s}_1 + \mathsf{s}_2\rfloor$.

We denote by $\mathbf{S}_\mathsf{P}$ the set of specifiers defined in the same way as $\mathbf{S}_\mathsf{S}$ except that we use $\mathsf{N_P}$ instead of $\mathsf{N_S}$ for the values of attributes. The semantics of specifiers in $\mathbf{S}_\mathsf{P}$ is as expected from the Section 'Attributed DL-Lite'. We assume that all polynomials occurring in a specifier in $\mathbf{S}_\mathsf{P}$ are of the form $\Sigma_{1\leq i\leq n_1}\Pi_{1\leq j\leq n_2} a_{i,j}$, where all $a_{i,j} \in \mathsf{N_I}$. Given an interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ and $\{F, G\} \subseteq \Phi^{\mathcal{I}}$, let $F \times G$ be:

$$\{(p,(a \times b)^{\mathcal{I}}) | \{a, b\} \subseteq \mathsf{N_P}, (p, a^{\mathcal{I}}) \in F, (p, b^{\mathcal{I}}) \in G\}.$$

Unlike $+^p$, $\times$ is not parameterized by an attribute because products combine different information, whereas sums represent alternative ways of obtaining the same information (i.e., tuple plus the same other attribute-value pairs). A product of annotation sets may be denoted by $\prod_{1\leq i\leq n} G_i$. We next define semiring attributed queries, which allow a ground specifier to be associated to the whole conjunction of atoms.

Definition 3 (Semiring Attributed Queries)**.** *A semiring attributed boolean conjunctive query (BCQ*$_{@,\mathbb{K}}$*) is an expression of the form:*

$$\exists \boldsymbol{x}. X_1 : S_1, \ldots, X_n : S_n \ (\varphi(\boldsymbol{x}))@S,$$

where S *is a ground specifier in* $\mathbf{S}_\mathsf{P}$, *for* $1 \leq i \leq n$, $X_i \in \mathsf{N_U}$ *are the set variables occurring in* $\varphi(\boldsymbol{x})$ *and* $S_i \in \mathbf{S}_\mathsf{S}$, *and*

$$\varphi(\boldsymbol{x}) = \bigwedge_{1\leq j\leq m} E_j(\boldsymbol{t_j})@T_j$$

where for $1 \leq j \leq m$, $T_j \in \mathbf{S}_\mathsf{S}$, $E_j \in \mathsf{N_C} \cup \mathsf{N_R}$ *and* $\boldsymbol{t_j}$ *is a tuple of elements from* $\mathsf{N_I} \cup \boldsymbol{x}$.

If $S = \lfloor\ \rfloor$, *we say that the BCQ*$_{@,\mathbb{K}}$ *is* plain.

Given a BCQ$_{@,\mathbb{K}}$ q, *let* q' *be the BCQ*$_@$ *that results from removing the outer specifier from* q. *Let* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ *be an interpretation and let* $\nu_{\mathcal{I}}(q')$ *be the set of all variable assignments* $\mathcal{Z}$ *that fufill the conditions of Definition 1 for* $\mathcal{I} \models q'$. $\mathcal{I}$ *satisfies* q, *written* $\mathcal{I} \models q$, *if there is a non-empty* $\chi \subseteq \nu_{\mathcal{I}}(q')$ *such that:*

1. *for any* $\mathcal{Z}, \mathcal{Z}' \in \chi$, *there exists* $X \in \mathsf{N_U}$ *occurring in* q *such that* $\mathcal{Z}(X) \neq \mathcal{Z}'(X)$ *or there exists* $x \in \boldsymbol{x}$ *such that* $\mathcal{Z}(x) \neq \mathcal{Z}'(x)$;
2. *for each* $\mathcal{Z} \in \chi$ *and* $1 \leq j \leq m$, *we have that* $(\mathcal{Z}(\boldsymbol{t_j}), F^{\mathcal{Z}}_j) \in E^{\mathcal{I},\mathcal{Z}}_j$ *for some* $F^{\mathcal{Z}}_j \in T^{\mathcal{I},\mathcal{Z}}_j$;
3. *there is* $p \in \Delta^{\mathcal{I}}$ *and* $G \in \Phi^{\mathcal{I}}$ *such that all* $H^{\mathcal{Z}} = \prod_{1\leq j\leq m} F^{\mathcal{Z}}_j$ *with* $\mathcal{Z} \in \chi$ *are differentiated by* p *in* G, *and* $\sum^{p}_{\mathcal{Z}\in\chi} H^{\mathcal{Z}} \in S^{\mathcal{I},\mathcal{Z}}$.

Essentially, Definition 3 says that: (1) there are different variable assignments which (2) satisfy the homomorphic conditions and (3) correspond to the interpretation of the outer specifier. Our semiring attributed queries can be easily extended so that the outer specifier has fresh and free object variables. In this case the answer to the query would be the set of provenance polynomials related with the respective attribute and the query. Semiring attributed queries can be used to query a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology using provenance polynomials, as we illustrate with the following example.

Example 7 (Example 3 cont'd)**.** *We now modify* q_{ex}, *so that we impose provenance constraints on the result:*

$$\exists xy\,(\mathsf{Married}(\mathsf{gabor})@\lfloor\mathsf{start}{:}\,x, \mathsf{end}{:}\,y\rfloor \wedge \mathsf{Married}(\mathsf{taylor})@\lfloor\mathsf{start}{:}\,x, \mathsf{end}{:}\,y\rfloor)@\lfloor\mathsf{src}{:}\,\gamma\rfloor$$

where γ *is the polynomial* $(\mathsf{s}_1 \times \mathsf{s}_3) + (\mathsf{s}_2 \times \mathsf{s}_3)$

By the semantics of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, *it follows that* $\mathcal{O}_{\mathsf{ex}} \models q_{\mathsf{ex}}$.

All shared attributes are taken into account when combining the annotations, while the non-shared attributes are irrelevant and lost in the product.

Example 8. *The query* $(\mathsf{Married}(a) \wedge \mathsf{Married}(b))@S$ *with*

$$S = \lfloor\mathsf{src}{:}\,\mathsf{s}_1 \times \mathsf{s}_2, \mathsf{classif}{:}\,\mathsf{public} \times \mathsf{confid}, \mathsf{mult}{:}\,2 \times 3\rfloor$$

is entailed by $\{\mathsf{Married}(a)@[\mathsf{src}{:}\,\mathsf{s}_1, \mathsf{classif}{:}\,\mathsf{public}, \mathsf{mult}{:}\,2], \mathsf{spouse}(b, c)@[\mathsf{src}{:}\,\mathsf{s}_2, \mathsf{classif}{:}\,\mathsf{confid}, \mathsf{mult}{:}\,3, \mathsf{time}{:}\,t]\}$ *and the inclusion of Example 3.*

The fact that a *and* b *are both married is obtained by combining sources* s_1 *and* s_2, *and by having access to both public and confidential information. Note that using inclusions to propagate annotations allows the query derived from assertions with multiplicities 2 and 3 to have multiplicity* 2×3, *as it would be under the bag semantics (Nikolaou et al. 2017).*

When interpreted over provenance-interpretations, ontologies in the DL-Lite$^{\mathcal{R}}_{@}$ fragment of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ (i.e., without sums) can entail queries with sums, as in Example 9.

Example 9. *Let* $\mathcal{O}$ *be the DL-Lite*$^{\mathcal{R}}_{@}$ *ontology*

$$\{A(a)@[p : a], A(a)@[p : b], A@X \sqsubseteq \exists R@X\}.$$

Then the query $\exists xy(R(x, y)@[p : a{+}b])@\lfloor\rfloor$ *follows from* $\mathcal{O}$ *only under the semirings-based semantics.*

Reasoning in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$

Unfortunately, Theorem 5 shows that provenance sums increase the complexity of the satisfiability problem. The proof is by reduction from the word problem for a polynomially space bounded Alternating Turing Machine (ATM) which is ExpTime-hard (Chandra, Kozen, and Stockmeyer 1981).

Theorem 5. *In DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, satisfiability is* ExpTime*-hard.*

The hardness result of Theorem 5 holds even for DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies without expressions of the form $\exists P$, where P is a role expression. Motivated by this negative result, we investigate restricted cases for query answering. We first show that for the class of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies which do not contain inclusions with expressions of the form $\exists P$ on the right side, we can check the entailment of BCQ$_{@,\mathbb{K}}$ via a transformation to ground and plain BCQ$_{@,\mathbb{K}}$s. Given such a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology $\mathcal{O}$, one can translate a BCQ$_{@,\mathbb{K}}$ q into a set of ground and plain BCQ$_{@,\mathbb{K}}$s $\mathsf{gr\_plain}(\mathcal{O}, q)$ such that $\mathcal{O} \models q$ iff there is some $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ that is entailed by an equisatisfiable ground ontology.

We can assume w.l.o.g. that if $E_j(\boldsymbol{t_j})@T_j$ occurs in q then $T_j \in \mathsf{N_U}$: if T_j is a specifier one can always replace it by a fresh $X \in \mathsf{N_U}$ and add $X : T_j$ to the prefix of q, that is:

$$q = \exists \boldsymbol{x}.\ X_1 : S_1, \ldots, X_m : S_m\ (\bigwedge_{1\leq j\leq m} E_j(\boldsymbol{t_j})@X_j)@S.$$

Assume $\star \in \mathsf{N_I}$ does not occur in $\mathcal{O}$ nor in q and let $\mathsf{N_{P_{min}}}$ be a fixed but arbitrary minimal subset of $\mathsf{N_P}$ such that for each $a \in \mathsf{N_P}$, $\mathsf{N_{P_{min}}}$ contains an element b such that a is mathematically equal to b. Let $\mathcal{I}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation with domain $\Delta^{\mathcal{I}} = \mathsf{N_{P_{min}}}$ and such that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N_{P_{min}}}$. We say that a variable assignment $\mathcal{Z}$ is *compatible with* q if $\mathcal{Z}(X_j) \in S_j^{\mathcal{I},\mathcal{Z}}$, $1 \leq j \leq m$. Let q' be the result of removing the outer specifier from q. Given a compatible $\mathcal{Z}$, a *$\mathcal{Z}$-image* $\bigwedge_{1\leq j\leq m} E_j(\boldsymbol{t_j})@T_j$ of q' is obtained by:

- replacing each X_j with $T_j = [a : b \mid (a, b) \in \mathcal{Z}(X_j)]$;
- replacing each object variable x by $\mathcal{Z}(x)$;
- if $\star$ occurs in some T_j, replacing $\star$ by $\star_{\mathcal{T}_j}$, where $\mathcal{T}_j$ is the set of attribute-value pairs in T_j that do not contain $\star$.

Given a ground specifier T, let $F_T := \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid a : b \text{ occurs in } T\} \in \Phi^{\mathcal{I}}$. We define $\mathsf{gr\_plain}(\mathcal{O}, q)$ as the set of ground plain BCQ$_{@,\mathbb{K}}$s:

$$q_{\mathsf{gp}} = (\bigwedge_{1\leq i\leq n} (\bigwedge_{1\leq j\leq m} E_j(\boldsymbol{t^i_j})@S^i_j))@\lfloor\,\rfloor$$

where the annotation sets $F_i = \prod_{1\leq j\leq m} F_{S^i_j}$ $(1 \leq i \leq n)$ are such that there exists p such that (i) the F_i are differentiated by p in some annotation set, (ii) each F_i contains some (p, a) with $a \in \mathsf{N_P}$, and (iii) $\sum^p_{1\leq i\leq n} F_i \in S^{\mathcal{I}}$. Also, $\bigwedge_{1\leq j\leq m} E_j(\boldsymbol{t^i_j})@S^i_j$ is a $\mathcal{Z}$-image of q' with attribute-value pairs built from elements of $\mathsf{N_S}$. By construction, q_{gp} does not contain variables.

Example 10 (Example 3 cont'd)**.** *The query below is a ground and plain version of the query in Example 7 which is entailed by $\mathcal{O}_{\mathsf{ex}}$.*

$$\mathsf{Married(gabor)}@[\mathsf{start} : 1975, \mathsf{end} : 1976, \mathsf{src} : \mathsf{s_1} + \mathsf{s_2}] \wedge$$
$$\mathsf{Married(taylor)}@[\mathsf{start} : 1975, \mathsf{end} : 1976, \mathsf{src} : \mathsf{s_3}].$$

One can show that, for DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies $\mathcal{O}$ without expressions of the form $\exists P$ on the right side of inclusions, $\mathcal{O} \models q$ iff there is $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ such that $\mathcal{O}_{\mathsf{gr}} \models q_{\mathsf{gp}}$, where $\mathcal{O}_{\mathsf{gr}}$ is an equisatisfiable ground ontology, obtained in a way similar to our construction of $\mathsf{gr\_plain}(\mathcal{O}, q)$ but imposing that the image of the variable assignments is over a finite set of individual names defined in terms of $\mathcal{O}$. In the case where $\mathcal{O}$ is ground, we further have a polynomial bound on the size of such q_{gp}.

Lemma 1. *Let q be a BCQ$_{@,\mathbb{K}}$ and let $\mathcal{O}$ be a ground DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology without expressions of the form $\exists P$ on the right side of inclusions. $\mathcal{O} \models q$ iff there is $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ such that (i) $\mathcal{O}_{\mathsf{gr}} \models q_{\mathsf{gp}}$, (ii) the size of q_{gp} is polynomial in $|q|$ and $|\mathcal{O}|$ and (iii) deciding $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ is in* PTime.

Lemma 1 does not hold for arbitrary DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies, as illustrated by Example 11.

Example 11. *Let $\mathcal{O}$ be the DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology $\{A \sqsubseteq \exists R, \exists R^- \sqsubseteq A@[p : b], \exists R^- \sqsubseteq \neg B, B(a), A(a)@[p : b]\}$. Then, $\mathcal{O}$ entails $q = \exists x(A(x)@\lfloor\,\rfloor)@\lfloor p : b + b\rfloor$, since there would be an R-successor in the anonymous part of the model, but there is no $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ such that $\mathcal{O} \models q_{\mathsf{gp}}$.*

We now use the polynomial bound in Lemma 1 to show an upper bound for a fragment, called *simple*, where we only allow inclusions of the form $E_1@S \sqsubseteq E_2@T$, with E_1 and E_2 concept/role names and S and T ground specifiers. We establish the complexity of BCQ$_{@,\mathbb{K}}$ entailment from simple ontologies. This case is close to the classical problem of query answering over databases, considered by Green, Karvounarakis, and Tannen (2007). Theorem 6 states that this complexity remains the same as in the database case.

Theorem 6. *BCQ$_{@,\mathbb{K}}$ entailment from a simple DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology is* NP*-complete.*

Proof. Let $\mathcal{O}$ be a simple DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology. We first show that one can decide in NP whether $E(\boldsymbol{a})@S$ is entailed from $\mathcal{O}$, where S is a ground specifier.

Claim 1. Deciding whether $\mathcal{O} \models E(\boldsymbol{a})@S$ is in NP.

Proof of Claim 1 We first guess the set $\mathcal{Q}$ of all atomic queries of the form $E_0(\boldsymbol{a})@T_0$ entailed by $\mathcal{O}$ such that $E_0@T_0$ occurs in $\mathcal{O}$ and an ordering for the entailment of such queries. If T_0 is an open specifier then replace it in $\mathcal{Q}$ by $T_{0,\star}$, defined as the ground closed specifier containing all attribute-value pairs in T_0 plus $\star_{\mathcal{S}} : \star_{\mathcal{S}}$ with $\mathcal{S}$ the set of attribute-value pairs in T_0. We make the usual assumption that individual names of the form $\star_{\mathcal{S}}$ do not occur in $\mathcal{O}$ and $E(\boldsymbol{a})@S$. Denote by $\mathcal{Q}_q$ the subset of $\mathcal{Q}$ containing all atomic queries which preceed q in the ordering. For each guessed query $q = E_0(\boldsymbol{a})@T_0$:

- Denote by F_T the set $\{(a, b) \mid a : b \text{ occurs in } T\}$ for any ground specifier T and let $E_0(\boldsymbol{a})@S_1, \ldots, E_0(\boldsymbol{a})@S_n$ be the assertions and atomic queries in $\mathcal{O} \cup \mathcal{Q}_q$ where E_0 and $\boldsymbol{a}$ occur.

- Guess a tree of annotation sets rooted either in F_{T_0} if T_0 is a closed specifier, or in a superset F of F_{T_0} if T_0 is an open specifier, where each non-leaf node F is the parent of children $G_1, \dots, G_m$ such that $F = \sum^p_{1 \le i \le m} G_i$, for some attribute p, and such that each leaf is either: one of $F_{S_1}, \dots, F_{S_n}$, or some F_T (or $F_{T_\star}$ if T is open) such that there exist $E_1@T_1 \sqsubseteq E_0@T$ and $E_1(\boldsymbol{a})@T_1$ (or $E_1(\boldsymbol{a})@T_{1,\star}$ if T_1 is open) in $\mathcal{O} \cup \mathcal{Q}_q$.

Check in polynomial time whether the trees satisfy the described conditions. The size of $\mathcal{Q}$ (and so the number of trees to guess and the size of the ordering) is bounded by the number of atomic queries $E_0(\boldsymbol{a})@T_0$ that can be built from concept/role expressions and individual names in $\mathcal{O}$, so it is polynomial in the size of $\mathcal{O}$.

To check whether $\mathcal{O} \models E(\boldsymbol{a})@S$, we check whether $E(\boldsymbol{a})@S \in \mathcal{Q}$ (assuming w.l.o.g. that $E@S$ occurs in $\mathcal{O}$). The size of each guessed tree is polynomial in the size of $\mathcal{O}$ since each leaf corresponds to an assertion/atomic query in $\mathcal{O}$ or $\mathcal{Q}$ (or an assertion/atomic query in $\mathcal{O}$ or $\mathcal{Q}$ together with an inclusion in $\mathcal{O}$) and they do not repeat in the tree. Thus, one can decide whether $\mathcal{O} \models E(\boldsymbol{a})@S$ in NP.

By Lemma 1, $\mathcal{O} \models q$ iff there exists $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ such that $\mathcal{O}_{\mathsf{gr}} \models q_{\mathsf{gp}}$. Moreover the size of q_{gp} is polynomial in the size of q and $\mathcal{O}$ and q_{gp} does not contain variables. We thus get the NP upper bound by guessing q_{gp} as well as certificates that $\mathcal{O}_{\mathsf{gr}} \models E(\boldsymbol{a})@S$ for each $E(\boldsymbol{a})@S$ in q_{gp}, using Claim 1 (indeed, $\mathcal{O}_{\mathsf{gr}}$ is also a simple ontology and is of polynomial size w.r.t. $\mathcal{O}$). The lower bound comes from the complexity of BCQ entailment in relational databases. □

One of the difficulties in showing Theorem 6 for arbitrary DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies is that one can express that elements in the anonymous part of the model are distinct, as illustrated in Example 11, and then our translation does not hold. In this case, $\mathsf{gr\_plain}(\mathcal{O}, q)$ needs to include queries with inequalities to distinguish anonymous elements, and entailment of BCQs with inequalities over DL-Lite$_{\mathcal{R}}$ ontologies easily leads to undecidability (e.g., see Theorem 13 in (Gutiérrez-Basulto et al. 2015)).

We now show an upper bound for *satisfiability* in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ by translating the ontology into an equisatisfiable ontology in a DL that we call DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$, which extends DL-Lite$_{\mathcal{R}}$ with *conjunctions on the left side of concept and role inclusions*. Our translation is double-exponential since in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ we need to ensure, e.g., that elements in the extension of $E@[\mathsf{src}\colon \mathsf{s}_1]$ and $E@[\mathsf{src}\colon \mathsf{s}_2]$ are also in the extension of $E@[\mathsf{src}\colon \mathsf{s}_1 + \mathsf{s}_2]$.

Theorem 7. *In DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, satisfiability is in* 2EXPTIME.

Sketch. We first ground the ontology and then translate it into DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$. We encode the semantics of provenance sums using a double-exponential number of concept and role inclusions with conjunctions on the left side. Since satisfiability in DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ is in PTIME (Artale et al. 2015) (Theorem 14), the 2EXPTIME upper bound follows. □

We next analyse entailment of *plain* BCQ$_{@,\mathbb{K}}$ w.r.t. DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies: the outer specifier is of the form $\lfloor\ \rfloor$ but inner specifiers can contain provenance sums (as in Ex. 9). We use the fact that BCQ entailment in DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ is in NP (Calì, Gottlob, and Pieris 2012, proof of Theorem 3.3).

Theorem 8. *In DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$, BCQ entailment is in* NP.

Theorem 9 establishes an upper bound for plain queries.

Theorem 9. *In DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, entailment of plain BCQ$_{@,\mathbb{K}}$ is in* N2EXPTIME.

Sketch. The proof uses the translation to DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ which leads to a double-exponential blowup of the ontology. Here, since queries are plain the translation is as for BCQ$_@$s. The result then follows from Theorem 8. □

Conclusion

We investigated the complexity of satisfiability and query answering in attributed DL-Lite$_{\mathcal{R}}$, for both the semantics introduced by Krötzsch et al. (2017) and a new semantics based on provenance semirings, which allows to embed provenance polynomials into the query. In particular, we show that these problems are PSPACE-complete for the classical semantics and that in the case of simple ontologies, even query answering under the semirings-based semantics has the same complexity as query answering in DL-Lite$_{\mathcal{R}}$. However, satisfiability of general DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies is EXPTIME-hard.

Related Work. Our attributed ontology language differs from DL-Lite$_{\mathcal{A}}$ (Calvanese et al. 2006), which allows to associate values to individuals or pairs of individuals, rather than to assertions, through binary or ternary relations called attribute concepts or attribute roles. In particular, while we can use the same attribute name to annotate different assertions about the same individual or pair of individuals, it would be ambiguous in DL-Lite$_{\mathcal{A}}$. For instance, we can express that Liz Taylor was married to Richard Burton from 1964 to 1974 and from 1975 to 1976 with $\mathsf{spouse}(\mathsf{taylor}, \mathsf{burton})@[\mathsf{start} : 1964, \mathsf{end} : 1974]$, $\mathsf{spouse}(\mathsf{taylor}, \mathsf{burton})@[\mathsf{start} : 1975, \mathsf{end} : 1976]$, while in DL-Lite$_{\mathcal{A}}$ we would need reification. The query $\mathsf{spouse}(\mathsf{taylor}, \mathsf{burton})@[\mathsf{start} : x, \mathsf{end} : y]$ that returns the start and end dates of the marriages would be more complex (namely, e.g., $\exists z\, \mathsf{spouse}_1(z, \mathsf{taylor}) \land \mathsf{spouse}_2(z, \mathsf{burton}) \land \mathsf{start}(z, x) \land \mathsf{end}(z, y)$). Another difference is the use in DL-Lite$_{\mathcal{A}}$ of two distinct alphabets and interpretation domains for the individuals and the values, following the distinction made in OWL between objects and values.

Regarding provenance, the topic has been extensively studied for relational databases (Cheney, Chiticariu, and Tan 2009), but has also drawn attention in other settings, e.g., for Datalog (Deutch et al. 2014), Datalog$^{+/-}$ (Lukasiewicz et al. 2014), and Semantic Web data, with numerous works proposing provenance models based on semirings for the evaluation of SPARQL queries over annotated RDF, see e.g., (Theoharis et al. 2011; Zimmermann et al. 2012; Geerts et al. 2016). In particular, Zimmermann et al. consider the possibility of having several annotations with different domains (fuzzy, temporal and provenance) and introduce an annotated version of SPARQL that manipulates explicitly annotations, while most work on provenance only implicitly propagates provenance annotations.

Future Work. Our next step will be the study of the data complexity and the design of practical algorithms for querying attributed DL-Lite ontologies. In particular, we would like to extend the classical DL-Lite rewriting approach to the attributed setting to avoid grounding the ontology. For instance, if an ontology only contains inclusions of the form $E@X \sqsubseteq F@X$, the rewriting algorithm for DL-Lite$_{\mathcal{R}}$ could be adapted to rewrite an attributed query where annotations sets are propagated in the rewriting process (e.g., $\mathsf{Married}(\mathsf{gabor})@\lfloor\mathsf{start}: 1975\rfloor$ in Example 3 could be rewritten into $\exists y\, \mathsf{spouse}(\mathsf{gabor}, y)@\lfloor\mathsf{start}: 1975\rfloor$).

Acknowledgements. We thank Stefan Borgwardt and an anonymous reviewer for pointing us to useful references. Partially supported by ANR-16-CE23-0007-01 ("DICOS").

References

Artale, A.; Calvanese, D.; Kontchakov, R.; and Zakharyaschev, M. 2009. The DL-Lite family and relations. *J. Artif. Intell. Res.* 36:1–69.

Artale, A.; Kontchakov, R.; Ryzhikov, V.; and Zakharyaschev, M. 2015. Tractable interval temporal propositional and description logics. In *Proceedings of AAAI*.

Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P., eds. 2007. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, second edition.

Bienvenu, M., and Ortiz, M. 2015. Ontology-mediated query answering with data-tractable description logics. In *Reasoning Web, Tutorial Lectures*, 218–307.

Bienvenu, M.; Deutch, D.; and Suchanek, F. M. 2012. Provenance for Web 2.0 data. In *Proceedings of Secure Data Management: 9th VLDB Workshop, SDM 2012*.

Botoeva, E.; Artale, A.; and Calvanese, D. 2010. Query rewriting in DL-Lite$^{(HN)}_{horn}$. In *Proceedings of DL*.

Calì, A.; Gottlob, G.; and Pieris, A. 2012. Towards more expressive ontology languages: The query answering problem. *Artif. Intell.* 193:87–128.

Calvanese, D.; De Giacomo, G.; Lembo, D.; Lenzerini, M.; Poggi, A.; and Rosati, R. 2006. Linking data to ontologies: The description logic DL-Lite$_A$. In *Proceedings of the OWLED*06 Workshop on OWL: Experiences and Directions*.

Calvanese, D.; Giacomo, G. D.; Lembo, D.; Lenzerini, M.; and Rosati, R. 2007. Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *J. of Automated Reasoning* 39(3):385–429.

Chandra, A. K.; Kozen, D. C.; and Stockmeyer, L. J. 1981. Alternation. *J. of the ACM* 28(1):114–133.

Cheney, J.; Chiticariu, L.; and Tan, W. C. 2009. Provenance in databases: Why, how, and where. *Foundations and Trends in Databases* 1(4):379–474.

Deutch, D.; Milo, T.; Roy, S.; and Tannen, V. 2014. Circuits for datalog provenance. In *Proceedings of ICDT*.

Geerts, F.; Unger, T.; Karvounarakis, G.; Fundulaki, I.; and Christophides, V. 2016. Algebraic structures for capturing the provenance of SPARQL queries. *J. ACM* 63(1):7:1–7:63.

Green, T. J.; Karvounarakis, G.; and Tannen, V. 2007. Provenance semirings. In *Proceedings of PODS*.

Gutiérrez-Basulto, V.; Ibáñez García, Y.; Kontchakov, R.; and Kostylev, E. V. 2015. Queries with negation and inequalities over lightweight ontologies. *Web Semant.* 35(P4):184–202.

Hoffart, J.; Suchanek, F. M.; Berberich, K.; and Weikum, G. 2013. YAGO2: A spatially and temporally enhanced knowledge base from wikipedia. *Artif. Intell.* 194:28–61.

Krötzsch, M.; Marx, M.; Ozaki, A.; and Thost, V. 2017. Attributed description logics: Ontologies for knowledge graphs. In *Proceedings of ISWC*.

Krötzsch, M.; Marx, M.; Ozaki, A.; and Thost, V. 2018. Attributed description logics: Reasoning on knowledge graphs. In *Proceedings of IJCAI*.

Krötzsch, M.; Rudolph, S.; and Hitzler, P. 2013. Complexities of Horn description logics. *ACM Trans. Comput. Logic* 14(1):2:1–2:36.

Lukasiewicz, T.; Martinez, M. V.; Predoiu, L.; and Simari, G. I. 2014. Information integration with provenance on the semantic web via probabilistic datalog+/-. In *URSW 2011-2013, Revised Selected Papers*.

Marx, M.; Krötzsch, M.; and Thost, V. 2017. Logic on MARS: Ontologies for generalised property graphs. In *Proceedings of IJCAI*.

Motik, B.; Cuenca Grau, B.; Horrocks, I.; Wu, Z.; Fokoue, A.; and Lutz, C., eds. 2009. *OWL 2 Web Ontology Language: Profiles*. W3C Recommendation. Available at `http://www.w3.org/TR/owl2-profiles/`.

Nikolaou, C.; Kostylev, E. V.; Konstantinidis, G.; Kaminski, M.; Grau, B. C.; and Horrocks, I. 2017. The bag semantics of ontology-based data access. In *Proceedings of IJCAI*.

Rodriguez, M. A., and Neubauer, P. 2010. Constructions from dots and lines. *Bulletin of the American Society for Information Science and Technology* 36(6):35–41.

Savitch, W. J. 1970. Relationships between nondeterministic and deterministic tape complexities. *Journal of Computer and System Sciences* 4(2):177 – 192.

Senellart, P. 2017. Provenance and probabilities in relational databases. *SIGMOD Record* 46(4):5–15.

Suciu, D.; Olteanu, D.; Ré, C.; and Koch, C. 2011. *Probabilistic Databases*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers.

Theoharis, Y.; Fundulaki, I.; Karvounarakis, G.; and Christophides, V. 2011. On provenance of queries on semantic web data. *IEEE Internet Computing* 15(1):31–39.

Vrandečić, D., and Krötzsch, M. 2014. Wikidata: A free collaborative knowledgebase. *Commun. ACM* 57(10).

Zimmermann, A.; Lopes, N.; Polleres, A.; and Straccia, U. 2012. A general framework for representing, reasoning and querying with annotated semantic web data. *J. Web Sem.* 11:72–95.

Notation in the Appendix

We introduce here the relevant notation and conventions used throughout this appendix. We often use:

- A for concept names;
- R for role names;
- E for concept/role names;
- K, L for concept/role expressions;
- F, G, H for annotation sets;
- S, T for specifiers.

For ground specifiers $\{S, T\} \subseteq \mathbf{S}_\mathsf{S}$, we write $S \Rightarrow T$ if T is an open specifier, and the set of attribute-value pairs $a : b$ in S is a superset of the set of attribute-value pairs in T.

In Appendix 'Proofs for Section 'Reasoning in DL-Lite$^{\mathcal{R}}_{@}$'', we frequently use an invididual name, called $\star$, to deal with elements of an interpretation of an ontology (or a query) which are not in the range of individual names that occur in the ontology. In Appendix 'Proofs for Section 'Reasoning in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$'', we use invididual names of the form $\star_{\mathcal{S}}$, where $\mathcal{S}$ is a set of attribute-value pairs, to deal with elements of an interpretation of an ontology (or a query) which are not in the range of individual names that occur in the ontology. The use of $\star_{\mathcal{S}}$ (instead of $\star$) is important to ensure that when we replace anonymous individuals by individuals of the form $\star_{\mathcal{S}}$ the resulting interpretation is still a well-founded provenance-interpretation.

To simplify the presentation, we also make the following assumptions:

1. whenever we speak about a ground (and plain) query q and a ground ontology $\mathcal{O}$, we assume w.l.o.g. that if an annotated concept or role name occurs in q then it also occurs in $\mathcal{O}$: if it does not occur we can add a tautology where it occurs. For instance, if $A(x)@[a : b]$ occurs in q then we assume that $A@[a : b] \sqsubseteq A@[a : b]$ is in $\mathcal{O}$. This is particularly important to ensure that the entailment relation is preserved by our DL translations in Lemmas 2 and 9.
2. whenever we speak about a query q and an ontology $\mathcal{O}$, we assume w.l.o.g. that all provenance sums occurring in q also occur in $\mathcal{O}$. If it is not the case we can add a tautology where they occur. This is in particular useful to define groundings.

We may denote with $|s|$ the cardinality of a finite set s.

Proofs for Section 'Reasoning in DL-Lite$^{\mathcal{R}}_{@}$'

In this section, we provide proofs for the theorems of the section 'Reasoning in DL-Lite$^{\mathcal{R}}_{@}$'.

Theorem 1. *In DL-Lite$^{\mathcal{R}}_{@}$, satisfiability is* PSPACE*-hard.*

Proof. The proof is by reduction from the word problem for a polynomially space bounded DTM. Let $\mathcal{M} = (Q, \Sigma, \Theta, q_0, q_\mathsf{f})$ be a DTM, where: Q is a finite set of *states*; Σ is a finite *alphabet* containing the *blank symbol* $\sqcup$; $\Theta : Q \times \Sigma \to Q \times \Sigma \times \{l, r\}$ is the *transition function*; and $\{q_0, q_\mathsf{f}\} \subseteq Q$ are the *initial* and *final* states.

Assume that $\mathcal{M}$ is polynomially space bounded on inputs $w_0 = \sigma_0 \dots \sigma_{n-1}$ of length n. As usual, a *configuration* of $\mathcal{M}$ is a word wqw' with $w, w' \in \Sigma^*$ and $q \in Q$, meaning that the tape contains the word ww', the machine is in state q and the head is on the position of the left-most symbol of w'. We assume w.l.o.g. that all configurations wqw' satisfy $|ww'| \leq m + 1$, where $m + 1$ is polynomial in n. The notion of *sucessive configurations* is defined as usual, in terms of the transition relation Θ. A *computation* of $\mathcal{M}$ on an input word w_0 is a sequence of successive configurations $\alpha_0, \alpha_1, \dots$, where $\alpha_0 = q_0 w_0$ is the *initial configuration* for the input $w_0 \in (\Sigma \setminus \{\sqcup\})^*$. Also, we assume w.l.o.g. that $\mathcal{M}$ does not attempt to move to the left (right) when it is on its left-most (right-most) tape position. We now construct a DL-Lite$^{\mathcal{R}}_{@}$ ontology $\mathcal{O}_{\mathcal{M},w_0}$ that is satisfiable iff $\mathcal{M}$ accepts w_0.

In the reduction, we use the following symbols (recall that attribute/attribute values are syntactically individual names):

- an individual name a and a concept name A annotated with attribute-value pairs encoding $\mathcal{M}$;
- attribute values $q \in Q$ to represent the states;
- an attribute s with values in Q;
- attribute values $0, \dots, m$ to encode the position of the head in the tape;
- an attribute h with values in $\{0, \dots, m\}$ to encode the head position;
- attribute values $\sigma \in \Sigma$ to represent the alphabet; and
- attributes $p_0, \dots, p_m$ with values in Σ to encode the tape.

The following assertion encodes the initial configuration:

$$A(a)@[s\colon q_0, h\colon 0, p_0\colon \sigma_0, \dots, p_{n-1}\colon \sigma_{n-1}, p_n\colon \sqcup, \dots, p_m\colon \sqcup].$$

To encode successive configurations, the main intuition is as follows. We read the content of the tape cell in the head position and create a new set of attribute-value pairs representing a successor configuration where, according to the transition, we modify the content of the cell in the previous head position and increment/decrement the previous head position in this new set of attribute-value pairs. We also copy all tape values from the previous configuration to the new configuration, except for the tape value at the previous head position. We can now encode our transitions $\Theta(q, \sigma) = (q', \tau, D)$ with concept inclusions of the form (we explain for $D = r$, the case $D = l$ is analogous), for each k with $0 \leq k \leq m$:

$$\Omega\ (A@X \sqsubseteq A@[s\colon q', h\colon k+1, p_k\colon \tau, P_{X\setminus k}])$$

where Ω is a shorthand for:

$$X\colon \lfloor s\colon q, h\colon k, p_k\colon \sigma \rfloor$$

and $P_{X\setminus k}$ abbreviates

$$p_0\colon X.p_0, \dots, p_{k-1}\colon X.p_{k-1}, p_{k+1}\colon X.p_{k+1}, \dots, p_m\colon X.p_m.$$

$\mathcal{M}$ accepts w_0 iff the final state is reachable. We formalize this with the following claim.

Claim. $\mathcal{O}_{\mathcal{M},w_0} \models A(a)@\lfloor s\colon q_\mathsf{f} \rfloor$ iff $\mathcal{M}$ accepts w_0.

Since one can reduce entailment of an assertion to (un)satisfiability, the above claim implies this theorem. $\square$

As defined by Krötzsch et al. (Theorem 4) (2017), the grounding of an attributed DL ontology consists of all assertions of the ontology together with grounded versions of inclusion axioms. In this paper we also allow object variables in specifiers and adapt the definition accordingly. We define the *grounding of a DL-Lite$^{\mathcal{R}}_{@}$ ontology* as follows. Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@}$ ontology and let $\mathsf{N}_\mathsf{I}^{\mathcal{O}}$ be the set of individual names occurring in $\mathcal{O}$, extended by the already mentioned fresh individual name $\star$. Let $\mathcal{I}$ be an interpretation over the domain $\Delta^{\mathcal{I}} = \mathsf{N}_\mathsf{I}^{\mathcal{O}}$ satisfying $a^{\mathcal{I}} = a$, for all $a \in \mathsf{N}_\mathsf{I}^{\mathcal{O}}$, and let $\mathcal{Z}$ be a variable assignment mapping object variables $x \in \mathsf{N}_\mathsf{V}$ to elements $\mathcal{Z}(x) \in \Delta^{\mathcal{I}}$ and set variables $X \in \mathsf{N}_\mathsf{U}$ to finite binary relations $\mathcal{Z}(X) \in \Phi^{\mathcal{I}}$. Consider a concept or role inclusion I of the form $X : S\ (K \sqsubseteq L)$. A variable assignment $\mathcal{Z}$ is said to be compatible with I if $\mathcal{Z}(X) \in S^{\mathcal{I},\mathcal{Z}}$. The *$\mathcal{Z}$-instance* $I_{\mathcal{Z}}$ of I is the concept or role inclusion $K' \sqsubseteq L'$ obtained by:

- replacing each X with $[a\colon b \mid (a,b) \in \mathcal{Z}(X)]$;
- replacing each assignment $a\colon X.b$ occurring in some specifier by all assignments $a\colon c$ such that $(b,c) \in \mathcal{Z}(X)$; and,
- replacing each object variable x by $\mathcal{Z}(x)$.

Then, the grounding of $\mathcal{O}$, denoted by $\mathsf{gr}(\mathcal{O})$, contains all $\mathcal{Z}$-instances $I_{\mathcal{Z}}$ for all concept or role inclusions I in $\mathcal{O}$ and all compatible variable assignments $\mathcal{Z}$. The resulting ontology $\mathsf{gr}(\mathcal{O})$ can be exponentially larger than $\mathcal{O}$, as there can be (at most) $2^{|\mathsf{N}_\mathsf{I}^{\mathcal{O}}|^2}$ $\mathcal{Z}$-instances for each inclusion.

Theorem 2. *In DL-Lite$^{\mathcal{R}}_{@}$, satisfiability is in* PSPACE.

Proof. The proof strategy consists on first translating a DL-Lite$^{\mathcal{R}}_{@}$ ontology into an equisatisfiable ground DL-Lite$^{\mathcal{R}}_{@}$ ontology and then translating it into DL-Lite$_{\mathcal{R}}$. Claims 1 and 2 below, are adaptations to DL-Lite$^{\mathcal{R}}_{@}$ of Theorems 4 and 3 respectively, by Krötzsch et al. (2017).

Claim 1. $\mathcal{O}$ is satisfiable iff $\mathsf{gr}(\mathcal{O})$ is satisfiable.

A ground ontology $\mathcal{O}_\mathsf{g}$ can be translated into a standard DL-Lite$_{\mathcal{R}}$ ontology $\mathsf{dl}(\mathcal{O}_\mathsf{g})$ as follows: replace every annotated concept/role name $E@S$ (or inverse role $R^{-}@S$) with a fresh concept/role name E_S (or inverse role R^{-}_S) in all the assertions and concept or role inclusions of $\mathcal{O}$, and extend the obtained DL-Lite$_{\mathcal{R}}$ ontology $\mathsf{dl}(\mathcal{O}_\mathsf{g})$ by all axioms $E_S \sqsubseteq E_T$ where E_S and E_T occur in translated axioms of $\mathsf{dl}(\mathcal{O}_\mathsf{g})$ and $S \Rightarrow T$.

Claim 2. $\mathcal{O}_\mathsf{g}$ is satisfiable iff $\mathsf{dl}(\mathcal{O}_\mathsf{g})$ is satisfiable.

Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@}$ ontology. By Claims 1 and 2, $\mathcal{O}$ is satisfiable iff $\mathsf{dl}(\mathsf{gr}(\mathcal{O}))$ is satisfiable. The grounding of $\mathcal{O}$ leads to an exponential blowup while the translation of a ground DL-Lite$^{\mathcal{R}}_{@}$ ontology to DL-Lite$_{\mathcal{R}}$ is polynomial. Since satisfiability of DL-Lite$_{\mathcal{R}}$ ontologies is in NL (Artale et al. 2009), it follows (by (Savitch 1970)) that satisfiability of DL-Lite$^{\mathcal{R}}_{@}$ ontologies is in PSPACE. ☐

The following lemma shows that given a ground DL-Lite$^{\mathcal{R}}_{@}$ ontology and a ground BCQ$_{@}$, there are polynomial size translations to a DL-Lite$_{\mathcal{R}}$ ontology and a boolean conjunctive query (BCQ) defined as usual. A ground BCQ$_{@}$ $q = \exists \boldsymbol{x}.\varphi(\boldsymbol{x})$ can be translated into a non-attributed BCQ $\mathsf{dl}(q) = \exists \boldsymbol{x}.\psi(\boldsymbol{x})$ where $\psi(\boldsymbol{x})$ is obtained from $\varphi(\boldsymbol{x})$ by replacing each atom $E(\boldsymbol{t})@S$ by $E_S(\boldsymbol{t})$.

Given a ground specifier S, the invididual name $\star$ and an interpretation $\mathcal{I}$ with $\star^{\mathcal{I}} \in \Delta^{\mathcal{I}}$, we define $F_S^{\star}$ as the annotation set in the singleton set $S^{\mathcal{I}}$ if S is a closed specifier, and $F_S^{\star} = \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid a\colon b \text{ occurs in } S\} \cup \{(\star^{\mathcal{I}}, \star^{\mathcal{I}})\}$ if S is an open specifier. We sometimes write $\mathcal{I}(\sigma)$ to denote the image of $\sigma \subseteq \mathsf{N}_\mathsf{I}$ in an interpretation $\mathcal{I}$.

Lemma 2. *Let $\mathcal{O}$ be a ground DL-Lite$^{\mathcal{R}}_{@}$ ontology and let q be a ground BCQ$_{@}$. Then, $\mathcal{O} \models q$ iff $\mathsf{dl}(\mathcal{O}) \models \mathsf{dl}(q)$.*

Proof. We prove this result similarly as Theorem 3 in (Krötzsch et al. 2017) and recall that we assume that all annotated concept/role names occurring in q also occur in $\mathcal{O}$.

Assume that $\mathsf{dl}(\mathcal{O}) \models \mathsf{dl}(q)$. Let $\mathcal{I}$ be a model of $\mathcal{O}$. We obtain a model $\mathcal{J}$ of $\mathsf{dl}(\mathcal{O})$ by interpreting each $a \in \mathsf{N}_\mathsf{I}$ by $a^{\mathcal{I}}$, and each E_S by $E@S^{\mathcal{I}}$.

Since $\mathsf{dl}(\mathcal{O}) \models \mathsf{dl}(q)$, $\mathcal{J}$ satisfies $\mathsf{dl}(q)$, i.e., there exists a mapping σ from object variables and individual names of $\mathsf{dl}(q)$ into $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$ such that $\sigma(a) = a^{\mathcal{J}}$ for every $a \in \mathsf{N}_\mathsf{I}$ and $\sigma(\boldsymbol{t}) \in E_S^{\mathcal{J}}$ for every $E_S(\boldsymbol{t})$ that occurs in $\mathsf{dl}(q)$. Let $\mathcal{Z}$ be a variable assignment such that $\mathcal{Z}(x) = \sigma(x)$ for every $x \in \mathsf{N}_\mathsf{V}$. Since for every $a \in \mathsf{N}_\mathsf{I}$, $\mathcal{Z}(a) = a^{\mathcal{I}} = \sigma(a)$ and q is ground, we obtain that $(\mathcal{Z}(\boldsymbol{t}), F) \in E^{\mathcal{I}}$ for some $F \in S^{\mathcal{I},\mathcal{Z}}$, for every atom $E(\boldsymbol{t})@S$ that occurs in q. The condition $\mathcal{Z}(X) \in S^{\mathcal{I},\mathcal{Z}}$ is always true since q is ground. Hence $\mathcal{I} \models q$. It follows that $\mathcal{O} \models q$.

For the other direction, assume that $\mathcal{O} \models q$. Let $\mathcal{J}$ be a model of $\mathsf{dl}(\mathcal{O})$. We obtain a model $\mathcal{I}$ of $\mathcal{O}$ over $\Delta^{\mathcal{I}} = \Delta^{\mathcal{J}} \cup \{\star^{\mathcal{I}}\}$, with $\star^{\mathcal{I}} \notin \Delta^{\mathcal{J}}$, by interpreting each $a \in \mathsf{N}_\mathsf{I}$ by $a^{\mathcal{J}}$, each concept/role name E by $E^{\mathcal{I}} = \{(\boldsymbol{a}, F_S^{\star}) \mid \boldsymbol{a} \in E_S^{\mathcal{J}} \text{ for some specifier } S\}$.

Since $\mathcal{O} \models q$, $\mathcal{I}$ satisfies q, that is, there exists a variable assignment $\mathcal{Z}$ such that $(\mathcal{Z}(\boldsymbol{t}), F) \in E^{\mathcal{I}}$ for some $F \in S^{\mathcal{I},\mathcal{Z}}$, for every atom $E(\boldsymbol{t})@S$ occurring in q. It follows that $\mathcal{Z}(a) = a^{\mathcal{I}} = a^{\mathcal{J}}$ for every $a \in \mathsf{N}_\mathsf{I}$ and $\mathcal{Z}(\boldsymbol{t}) \in E_S^{\mathcal{J}}$ for every atom $E_S(\boldsymbol{t})$ that occurs in $\mathsf{dl}(q)$. Thus, $\mathsf{dl}(\mathcal{O}) \models \mathsf{dl}(q)$. ☐

We now define the *grounding of a query* in the same way as for an ontology. Let q be a BCQ$_{@}$:

$$\exists \boldsymbol{x}.\ X_1 : S_1, \ldots, X_n : S_n\ \varphi(\boldsymbol{x}).$$

W.l.o.g., assume that if an atom $E(\boldsymbol{t})@S$ occurs in q then $S \in \mathsf{N}_\mathsf{U}$ (recall that if S is a specifier one can always replace it by a fresh set variable X and add $X : S$ to the prefix of q). A variable assignment $\mathcal{Z}$ is compatible with q if $\mathcal{Z}(X_i) \in S_i^{\mathcal{I},\mathcal{Z}}$, $1 \leq i \leq n$. The $\mathcal{Z}$-instance $q_{\mathcal{Z}}$ of q is obtained by replacing every set variable X_i and every assignment $a\colon X_i.b$ occurring in some specifier as we do for inclusion axioms of an ontology, as well as replacing every variable x occurring in some specifier by $\mathcal{Z}(x)$ in the whole query (also in the atoms). Thus, $q_{\mathcal{Z}}$ contains only variables that do not appear in any specifier of q. We call such $\mathcal{Z}$-instance $q_{\mathcal{Z}}$ a *grounded version of* q and denote by $\mathsf{gr}(q)$ the set of all grounded versions of q. Recall from Section "Notation in the Appendix" that we assume that individual names in q also occur in $\mathcal{O}$, and so in $\mathsf{N}_\mathsf{I}^{\mathcal{O}}$. Thus, the set $\mathsf{gr}(q)$ is non-empty.

To show that we can indeed use the grounding of a query q to check entailment of q, in particular, to deal with object variables occurring in specifiers, we use the classical notion of a *canonical model* of a DL-Lite$^{\mathcal{R}}_{@}$ ontology (Definition 5). As a first step we define the interpretation of a concept/role expression, which is then used in Definition 5.

Definition 4 (Interpretation of a DL-Lite$^{\mathcal{R}}_{@}$ concept/role)**.** *Given a DL-Lite$^{\mathcal{R}}_{@}$ ontology $\mathcal{O}$, a positive concept or role inclusion $I = X : S\ (K \sqsubseteq L) \in \mathcal{O}$, and a variable assignment $\mathcal{Z}$ mapping object variables to $\Delta^{\mathcal{M}_n}$ and set variables to $\Phi^{\mathcal{M}_n}$, we define $\mathcal{I}^{\mathcal{Z},\mathcal{M}_n}_{L,I} := \mathcal{J}$ as follows:*

- *if I is a concept inclusion and $L = A@T$ then:*
  $$\Delta^{\mathcal{J}} := \{\rho\} \cup \Delta^{\mathcal{M}_n}, \qquad A^{\mathcal{J}} = \{(\rho, H)\}, \qquad E^{\mathcal{J}} = \emptyset;$$
- *if I is a concept inclusion and $L = \exists R@T$ then:*
  $$\Delta^{\mathcal{J}} := \{\rho, \sigma\} \cup \Delta^{\mathcal{M}_n}, \quad R^{\mathcal{J}} = \{(\rho, \sigma, H)\}, \quad E^{\mathcal{J}} = \emptyset;$$
- *if I is a concept inclusion and $L = \exists R^-@T$ then:*
  $$\Delta^{\mathcal{J}} := \{\rho, \sigma\} \cup \Delta^{\mathcal{M}_n}, \quad R^{\mathcal{J}} = \{(\sigma, \rho, H)\}, \quad E^{\mathcal{J}} = \emptyset;$$
- *if I is a role inclusion and $L = R@T$ then:*
  $$\Delta^{\mathcal{J}} := \{\rho, \rho'\} \cup \Delta^{\mathcal{M}_n}, \quad R^{\mathcal{J}} = \{(\rho, \rho', H)\}, \quad E^{\mathcal{J}} = \emptyset;$$
- *if I is a role inclusion and $L = R^-@T$ then:*
  $$\Delta^{\mathcal{J}} := \{\rho, \rho'\} \cup \Delta^{\mathcal{M}_n}, \quad R^{\mathcal{J}} = \{(\rho', \rho, H)\}, \quad E^{\mathcal{J}} = \emptyset;$$

for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$ such that $E \neq A$ (case 1) or $E \neq R$ (last 4 cases), where $H \in T^{\mathcal{M}_n,\mathcal{Z}}$ if T is closed, $H = \bigcap_{F \in T^{\mathcal{M}_n,\mathcal{Z}}} F \cup \{(\star^{\mathcal{M}_n}, \star^{\mathcal{M}_n})\}$ if T is open, and $\sigma \notin \Delta^{\mathcal{M}_n}$. We write $\boldsymbol{\rho}$ to indicate ρ if I is a concept inclusion and (ρ, ρ') if I is a role inclusion.

Definition 5 (Canonical Model of a DL-Lite$^{\mathcal{R}}_{@}$ ontology)**.** *The canonical model $\mathcal{M}^{\mathcal{O}}_{\star}$ of a DL-Lite$^{\mathcal{R}}_{@}$ ontology $\mathcal{O}$ is the union of interpretations $\mathcal{M}_0, \mathcal{M}_1, \ldots$, with $\mathcal{M}_0$ defined as:*

$$\Delta^{\mathcal{M}_0} := \mathsf{N}^{\mathcal{O}}_{\mathsf{I}},\ a^{\mathcal{M}_0} = a \text{ for every } a \in \mathsf{N}^{\mathcal{O}}_{\mathsf{I}}$$
$$\text{and } E^{\mathcal{M}_0} := \{(\boldsymbol{a}^{\mathcal{M}_0}, F^{\star}_{S}) \mid E(\boldsymbol{a})@S \in \mathcal{O}\}$$

for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$. For the inductive definition of the sequence assume $\mathcal{M}_n$ is defined. Then obtain $\mathcal{M}_{n+1}$ by applying the following rule once:

- *if $I = X : S\ (K \sqsubseteq L)$ is a positive concept or role inclusion in $\mathcal{O}$, $\mathcal{Z}$ is a variable assignment compatible with I, and $\boldsymbol{a} \in K^{\mathcal{M}_n,\mathcal{Z}}$ but $\boldsymbol{a} \notin L^{\mathcal{M}_n,\mathcal{Z}}$ then take the interpretation $\mathcal{J} := \mathcal{I}^{\mathcal{Z},\mathcal{M}_n}_{L,I}$ and add it to $\mathcal{M}_n$ by identifying $\boldsymbol{\rho}$ with $\boldsymbol{a}$. In more detail, assume $\boldsymbol{a} = \boldsymbol{\rho}$ and define $\mathcal{M}_{n+1}$ by setting:*
  $$\Delta^{\mathcal{M}_{n+1}} := \Delta^{\mathcal{M}_n} \cup \Delta^{\mathcal{J}} \text{ and}$$
  $$E^{\mathcal{M}_{n+1}} := E^{\mathcal{M}_n} \cup E^{\mathcal{J}} \text{ for all } E \in \mathsf{N_C} \cup \mathsf{N_R}.$$

We assume that rule application is fair, that is, if a rule is applicable in a certain place, then it will indeed eventually be applied there. We obtain $\mathcal{M}^{\mathcal{O}}_{\star}$ by setting for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$:

$$\Delta^{\mathcal{M}^{\mathcal{O}}_{\star}} := \bigcup_{n \geq 0} \Delta^{\mathcal{M}_n} \text{ and } E^{\mathcal{M}^{\mathcal{O}}_{\star}} := \bigcup_{n \geq 0} E^{\mathcal{M}_n}.$$

We show in Theorem 10 the main properties of the canonical model. Before that, we show that we can modify an arbitrary interpretation $\mathcal{I}$ so that $\star^{\mathcal{I}}$ is the only element not in the range of individual names that occur in $\mathcal{O}$ which can occur in annotation sets (Definition 6), and that such modified interpretation does not change the entailment relation of BCQ$_{@}$s (Lemma 3) under certain conditions.

Definition 6. *Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation and $\mathcal{O}$ a DL-Lite$^{\mathcal{R}}_{@}$ ontology. Assume w.l.o.g. that there is $\star^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ such that $\star^{\mathcal{I}} \neq a^{\mathcal{I}}$ for all $a \in \mathsf{N}^{\mathcal{O}}_{\mathsf{I}} \setminus \{\star\}$. For an annotation set $F \in \Phi^{\mathcal{I}}$, we define $F_{\star}$ to be the annotation set obtained from F by replacing all $e \notin \mathcal{I}(\mathsf{N}^{\mathcal{O}}_{\mathsf{I}} \setminus \{\star\})$ in F by $\star^{\mathcal{I}}$ and define $\mathcal{I}^{\mathcal{O}}_{\star} = (\Delta^{\mathcal{I}^{\mathcal{O}}_{\star}}, \cdot^{\mathcal{I}^{\mathcal{O}}_{\star}})$ as follows:*

- *$\Delta^{\mathcal{I}^{\mathcal{O}}_{\star}} := \Delta^{\mathcal{I}}$; $a^{\mathcal{I}^{\mathcal{O}}_{\star}} := a^{\mathcal{I}}$ for all $a \in \mathsf{N_I}$; and*
- *$E^{\mathcal{I}^{\mathcal{O}}_{\star}} := \{(\boldsymbol{a}, F_{\star}), (\boldsymbol{a}, F_{\star} \cup \{(\star^{\mathcal{I}}, \star^{\mathcal{I}})\}) \mid (\boldsymbol{a}, F) \in E^{\mathcal{I}}\}$ for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$.*

Lemma 3. *Let $\mathcal{I}$ be an interpretation such that $\star^{\mathcal{I}}$ appears only in annotations sets, and, for all annotation sets F in $\mathcal{I}$ and all $(a, b) \in F$, either $\{a, b\} \subseteq \mathcal{I}(\mathsf{N}^{\mathcal{O}}_{\mathsf{I}} \setminus \{\star\})$ or $\{a, b\} \cap \mathcal{I}(\mathsf{N}^{\mathcal{O}}_{\mathsf{I}} \setminus \{\star\}) = \emptyset$. Then, for every BCQ$_{@}$ q without $\star$ and with concept, role and individual names occurring in a DL-Lite$^{\mathcal{R}}_{@}$ ontology $\mathcal{O}$ the following holds: $\mathcal{I} \models q$ iff $\mathcal{I}^{\mathcal{O}}_{\star} \models q$.*

Proof. $\mathcal{I} \models q$ iff there is $\mathcal{Z}$ as in Definition 1. By the conditions of $\mathcal{I}$ in this lemma and since $\star$ does not occur in q, for all $x \in \mathsf{N_V}$, either x does not occur in a specifier of q or $\mathcal{Z}(x) \in \mathcal{I}(\mathsf{N}^{\mathcal{O}}_{\mathsf{I}})$. Let $\mathcal{Z}'$ be a mapping from $\mathsf{N_U} \cup \mathsf{N_V}$ to $\mathcal{I}^{\mathcal{O}}_{\star}$ (Definition 6) defined as follows:

- $\mathcal{Z}'(x) = \mathcal{Z}(x)$, for every $x \in \mathsf{N_V}$; and
- $\mathcal{Z}'(X) = \mathcal{Z}(X)_{\star}$, for every $X \in \mathsf{N_U}$, where the annotation set $\mathcal{Z}(X)_{\star}$ is as in Definition 6.

By definition of $\mathcal{I}^{\mathcal{O}}_{\star}$, we can see that for every atom $E(\boldsymbol{t})@S$ occurring in q,

- $(\mathcal{Z}(\boldsymbol{t}), F) \in E^{\mathcal{I}}$ for some $F \in S^{\mathcal{I},\mathcal{Z}}$ iff $(\mathcal{Z}'(\boldsymbol{t}), F_{\star}) \in E^{\mathcal{I}^{\mathcal{O}}_{\star}}$ for some $F_{\star} \in S^{\mathcal{I}^{\mathcal{O}}_{\star},\mathcal{Z}'}$.

Moreover, $\mathcal{Z}'(X_i) \in S_i^{\mathcal{I}^{\mathcal{O}}_{\star},\mathcal{Z}'}$ for each $X_i : S_i$ in the prefix of q. The existence of such $\mathcal{Z}'$ happens iff $\mathcal{I}^{\mathcal{O}}_{\star} \models q$. □

In our proofs we use the following notion of homomorphism.

Definition 7. *Let $\mathcal{I}$ and $\mathcal{J}$ be DL-Lite$^{\mathcal{R}}_{@}$ (or DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$) interpretations. A* homomorphism *$h : \mathcal{I} \to \mathcal{J}$ is a function from $\Delta^{\mathcal{I}}$ to $\Delta^{\mathcal{J}}$ such that:*

- *$h(a^{\mathcal{I}}) = a^{\mathcal{J}}$ for all $a \in \mathsf{N_I}$ (or $a \in \mathsf{N_S}$), and, for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$,*
- *$(\boldsymbol{a}, F) \in E^{\mathcal{I}}$ implies $h((\boldsymbol{a}, F)) \in E^{\mathcal{J}}$,*

where $h((\boldsymbol{a}, F))$ is a shorthand for $(h(\boldsymbol{a}), \{(h(b), h(c)) \mid (b, c) \in F\})$ and $h(\boldsymbol{a})$ is a shorthand for $h(a)$ if $E \in \mathsf{N_C}$ and $\boldsymbol{a} = a$, and $(h(a), h(a'))$, if $E \in \mathsf{N_R}$ and $\boldsymbol{a} = (a, a')$. We write $\mathcal{I} \to \mathcal{J}$ if there is a homomorphism from $\mathcal{I}$ to $\mathcal{J}$.

The following lemma is an easy consequence of Definitions 7 and 1 (or 3).

Lemma 4. *Let $\mathcal{I}, \mathcal{J}$ be DL-Lite$^{\mathcal{R}}_{@}$ (or DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$) interpretations. If $\mathcal{I} \models q$ and $\mathcal{I} \to \mathcal{J}$ then $\mathcal{J} \models q$.*

We may write $\mathcal{M}^{\mathcal{O}}_{\star} \models \mathsf{gr}(q)$ (or $\mathcal{O} \models \mathsf{gr}(q)$) meaning that there is $q_{\mathcal{Z}} \in \mathsf{gr}(q)$ such that $\mathcal{M}^{\mathcal{O}}_{\star} \models \mathsf{gr}(q)$ (or $\mathcal{O} \models q_{\mathcal{Z}}$).

Theorem 10. *Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@}$ ontology and let q be a BCQ$_{@}$. Assume $\mathcal{O}$ is satisfiable and the individual names that occur in q occur in $\mathcal{O}$. Then, the following holds:*

1. *$\mathcal{O} \models q$ iff $\mathcal{M}^{\mathcal{O}}_{\star} \models q$;*
2. *$\mathcal{M}^{\mathcal{O}}_{\star} \models q$ iff $\mathcal{M}^{\mathcal{O}}_{\star} \models \mathsf{gr}(q)$.*

Proof. Recall that $\star$ is an individual name that does not occur in $\mathcal{O}$ and that we assume that all individual names that occur in q occur in $\mathcal{O}$, so that $\star$ does not occur in q. For Point 1, if $\mathcal{O}$ is satisfiable then, by construction, $\mathcal{M}^{\mathcal{O}}_{\star}$ is a model of $\mathcal{O}$. Thus, $\mathcal{O} \models q$ implies $\mathcal{M}^{\mathcal{O}}_{\star} \models q$. Conversely, assume $\mathcal{M}^{\mathcal{O}}_{\star} \models q$. By construction, if an interpretation $\mathcal{I}$ models $\mathcal{O}$ then $\mathcal{M}^{\mathcal{O}}_{\star} \to \mathcal{I}^{\mathcal{O}}_{\star}$, where $\mathcal{I}^{\mathcal{O}}_{\star}$ is as in Definition 6. Then, by Lemma 4, $\mathcal{I}^{\mathcal{O}}_{\star} \models q$. Moreover, as $\mathcal{M}^{\mathcal{O}}_{\star}$ only has annotation sets H such that for all $(a, b) \in H$ either $(a, b) = (\star^{\mathcal{M}^{\mathcal{O}}_{\star}}, \star^{\mathcal{M}^{\mathcal{O}}_{\star}})$ or $\{a, b\} \subseteq \mathcal{M}^{\mathcal{O}}_{\star}(\mathsf{N}^{\mathcal{O}}_{\mathsf{I}} \setminus \{\star\})$, we can assume that the image $\mathcal{J}$ of the homomorphism from q to $\mathcal{I}^{\mathcal{O}}_{\star}$ only contains annotation sets of this form. Consider the interpretation $\mathcal{K}$ that is the result of removing from $\mathcal{I}$ all annotations sets H such that there is $(a, b) \in H$ with $a = \star^{\mathcal{I}}$ or $b = \star^{\mathcal{I}}$ but $(a, b) \neq (\star^{\mathcal{I}}, \star^{\mathcal{I}})$. We then have that $\mathcal{J} \to \mathcal{K}^{\mathcal{O}}_{\star}$. As $\mathcal{J} \models q$, by Lemma 4, $\mathcal{K}^{\mathcal{O}}_{\star} \models q$. By Lemma 3, $\mathcal{K} \models q$. By definition, $\mathcal{K} \to \mathcal{I}$, and thus by Lemma 4, $\mathcal{I} \models q$.

Finally, for Point 2, assume $\mathcal{M}^{\mathcal{O}}_{\star} \models q$. Then there is a variable assignment $\mathcal{Z}$ as in Definition 1. Since all annotation sets of $\mathcal{M}^{\mathcal{O}}_{\star}$ contain only elements in $\mathcal{M}^{\mathcal{O}}_{\star}(\mathsf{N}^{\mathcal{O}}_{\mathsf{I}})$, the $\mathcal{Z}$-instance $q_{\mathcal{Z}}$ of q is in $\mathsf{gr}(q)$. Thus, $\mathcal{M}^{\mathcal{O}}_{\star} \models \mathsf{gr}(q)$. Conversely, by definition of $\mathsf{gr}(q)$, if there is some $q_{\mathcal{Z}} \in \mathsf{gr}(q)$ such that $\mathcal{M}^{\mathcal{O}}_{\star} \models q_{\mathcal{Z}}$ then the variable assignment $\mathcal{Z}$ satisfies the conditions of Definition 1 for the query q. So $\mathcal{M}^{\mathcal{O}}_{\star} \models q$. □

Given a DL-Lite$^{\mathcal{R}}_{@}$ ontology $\mathcal{O}$, let $\mathsf{N}^{\mathcal{O}}$ be the set of concept/role names occurring in $\mathcal{O}$. We define $\mathcal{O}_{\star}$ as the union of $\mathcal{O}$ and all concept/role inclusions of the form $E@S \sqsubseteq E@S_{\star}$ with $E \in \mathsf{N}^{\mathcal{O}}$ and S an open ground specifier occurring in $\mathcal{O}$, where $S_{\star}$ is the closed specifier with attribute-value pairs occurring in S plus $\star{:}\,\star$.

Lemma 5. *Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@}$ ontology and q a BCQ$_{@}$. We have that $\mathcal{O} \models q$ iff $\mathsf{gr}(\mathcal{O})_{\star} \models \mathsf{gr}(q)$.*

Proof. Recall that $\star$ does not occur in $\mathcal{O}$ by definition, and thus neither in q by assumption. Assume $\mathcal{O} \models q$. It is important to note that $\star$ may occur in some queries in $\mathsf{gr}(q)$, so that Point 1 of Theorem 10 does not hold for queries from $\mathsf{gr}(q)$. In the following, we therefore use $\mathsf{gr}(\mathcal{O})_{\star}$ instead of $\mathsf{gr}(\mathcal{O})$. By definition of $\mathsf{gr}(\mathcal{O})_{\star}$ and construction of $\mathcal{M}^{\mathcal{O}}_{\star}$, we have that $\mathcal{M}^{\mathcal{O}}_{\star} = \mathcal{M}^{\mathsf{gr}(\mathcal{O})_{\star}}_{\star}$. Suppose an intepretation $\mathcal{I}$ models $\mathsf{gr}(\mathcal{O})_{\star}$. By Point 1 of Theorem 10, $\mathcal{M}^{\mathcal{O}}_{\star} \models q$ and by Point 2, $\mathcal{M}^{\mathcal{O}}_{\star} \models \mathsf{gr}(q)$. Moreover, by our construction, $\mathcal{M}^{\mathcal{O}}_{\star} \to \mathcal{I}$: here we do not need to use $\mathcal{I}^{\mathcal{O}}_{\star}$ from Definition 6 because $\mathcal{I}$ is a model of $\mathsf{gr}(\mathcal{O})_{\star}$. Thus, by Lemma 4, $\mathcal{I} \models \mathsf{gr}(q)$. Since $\mathcal{I}$ was an arbitrary interpretation satisfying $\mathsf{gr}(\mathcal{O})_{\star}$, we have that $\mathsf{gr}(\mathcal{O})_{\star} \models \mathsf{gr}(q)$.

Now, assume $\mathsf{gr}(\mathcal{O})_{\star} \models \mathsf{gr}(q)$. By Claim 1 of Theorem 2, $\mathcal{O}$ is satisfiable iff $\mathsf{gr}(\mathcal{O})$ is satisfiable and our extension $\mathsf{gr}(\mathcal{O})_{\star}$ clearly does not change this relation. So if $\mathsf{gr}(\mathcal{O})_{\star}$ is unsatisfiable $\mathcal{O}$ trivially entails q. Then, assume $\mathsf{gr}(\mathcal{O})_{\star}$ is satisfiable. By construction, $\mathcal{M}^{\mathsf{gr}(\mathcal{O})_{\star}}_{\star}$ is a model of $\mathsf{gr}(\mathcal{O})_{\star}$. Then, $\mathcal{M}^{\mathsf{gr}(\mathcal{O})_{\star}}_{\star} \models \mathsf{gr}(q)$. By definition of $\mathsf{gr}(\mathcal{O})_{\star}$, we have that $\mathcal{M}^{\mathcal{O}}_{\star} = \mathcal{M}^{\mathsf{gr}(\mathcal{O})_{\star}}_{\star}$. This means that $\mathcal{M}^{\mathcal{O}}_{\star} \models \mathsf{gr}(q)$. By Points 1 and 2 of Theorem 10, $\mathcal{O} \models q$. □

We are now ready to show Theorem 3. The proof of Theorem 3 is based on grounding and translation to DL-Lite$_{\mathcal{R}}$. However, we cannot simply ground both the ontology and the query since the exponential expansion would give us an NEXPTIME upper bound in the combined complexity, whereas here we show a PSPACE upper bound.

Theorem 3. *In DL-Lite$^{\mathcal{R}}_{@}$, BCQ$_{@}$ entailment is in* PSPACE.

Proof. Let $\mathcal{O}$ be DL-Lite$^{\mathcal{R}}_{@}$ ontology and let q be a BCQ$_{@}$. We first check in PSPACE whether $\mathcal{O}$ is satisfiable (Theorem 2). If $\mathcal{O}$ is unsatisfiable, $\mathcal{O}$ trivially entails q. Otherwise, by Lemma 5, $\mathcal{O} \models q$ iff $\mathsf{gr}(\mathcal{O})_{\star} \models \mathsf{gr}(q)$. We show that $\mathsf{gr}(\mathcal{O})_{\star} \models \mathsf{gr}(q)$ can be checked in polynomial space w.r.t. $|\mathcal{O}|$ and $|q|$.

Let $\mathcal{A} = \{E_S(\boldsymbol{a}) \mid E(\boldsymbol{a})@S \in \mathcal{O}\}$ be the set of DL-Lite$_{\mathcal{R}}$ assertions in $\mathsf{dl}(\mathsf{gr}(\mathcal{O})_{\star})$. To decide in PSPACE whether $\mathsf{gr}(\mathcal{O})_{\star} \models \mathsf{gr}(q)$, we guess the DL-Lite$_{\mathcal{R}}$ version $\mathsf{dl}(q_{\mathcal{Z}})$ of a grounded version $q_{\mathcal{Z}} \in \mathsf{gr}(q)$ of q and a rewriting q' of $\mathsf{dl}(q_{\mathcal{Z}})$ w.r.t. $\mathcal{T}$, where $\mathcal{T}$ is the set of positive concept or role inclusions in $\mathsf{dl}(\mathsf{gr}(\mathcal{O})_{\star})$, such that $\mathcal{A} \models q'$. Recall that we assumed in Section "Notation in the Appendix" that $\mathsf{gr}(\mathcal{O})$ (and therefore $\mathsf{gr}(\mathcal{O})_{\star}$) contains the annotated concept/role names that occur in $q_{\mathcal{Z}}$.

Checking that $\mathsf{dl}(q_{\mathcal{Z}})$ is the DL-Lite$_{\mathcal{R}}$ translation of a grounded version of q can be done in PTIME, and checking that $\mathcal{A} \models q'$ can be done in NP. The number of rewriting steps from $\mathsf{dl}(q_{\mathcal{Z}})$ to q' is polynomial in the size of $\mathcal{T}$ (Calvanese et al. 2007), so potentially exponential in the size of $\mathcal{O}$. This gives an NEXPTIME upper bound for checking whether q' is a rewriting of $\mathsf{dl}(q_{\mathcal{Z}})$. We improve this upper bound by showing that it is possible to verify that q' is a rewriting of $\mathsf{dl}(q_{\mathcal{Z}})$ in PSPACE.

We consider a non-deterministic adaptation of the algorithm PerfectRef by Calvanese et al. (2007) that takes as an input $\mathsf{dl}(q_{\mathcal{Z}})$, q' and $\mathcal{O}$. We adopt similar definitions and terminology as in (Calvanese et al. 2007), Section 5.1., which we briefly recall here. The symbol "_" represents non-distinguished non-shared variables. A positive inclusion I is applicable to an atom $A(x)$ if I has A in its right-hand side. A positive inclusion I is applicable to an atom $R(x, y)$ if (i) $x =\_$ and the right-hand side of I is $\exists R$, or (ii) I is a role inclusion and its right-hand side is either R or R^-. For each role name occurring in $\mathcal{O}$ we add to $\mathcal{O}$ an equivalence $R^- \equiv \overline{R}$ (that is, $R^- \sqsubseteq \overline{R}$ and $\overline{R} \sqsubseteq R^-$), where $\overline{R}$ is a fresh role name. We can then assume w.l.o.g. that inverse roles only occur in such equivalences by replacing R^- in other places by $\overline{R}$. Let g be an atom and I be a positive inclusion that is applicable to g. The atom obtained from g by applying I, denoted by $gr(g, I)$, is defined as follows:

- if $g = A(x)$ and $I = A_1 \sqsubseteq A$, then $gr(g, I) = A_1(x)$;
- if $g = A(x)$ and $I = \exists R \sqsubseteq A$, then $gr(g, I) = R(x, \_)$;
- if $g = R(x, \_)$ and $I = A \sqsubseteq \exists R$, then $gr(g, I) = A(x)$;
- if $g = R(x, \_)$ and $I = \exists R_1 \sqsubseteq \exists R$, then $gr(g, I) = R_1(x, \_)$;
- if $g = R(x, y)$ and $I = R_1 \sqsubseteq R$ or $I = R_1^- \sqsubseteq R^-$, then $gr(g, I) = R_1(x, y)$;
- if $g = R(x, y)$ and $I = R_1 \sqsubseteq R^-$ or $I = R_1^- \sqsubseteq R$, then $gr(g, I) = R_1(y, x)$.

Let $q[g/g']$ denote the conjunctive query obtained from q by replacing the atom g with a new atom g'. Let τ be a function that takes as input a conjunctive query q and returns a new conjunctive query obtained by replacing each occurrence of an unbound variable in q with the symbol '_'; and let reduce be a function that takes as input a conjunctive query q and two atoms g_1, g_2 and returns a conjunctive query obtained by applying to q the most general unifier between g_1 and g_2. We are now ready to present our adaptation of the algorithm:

1. $q'' \leftarrow \mathsf{dl}(q_\mathcal{Z})$
2. $\Gamma \leftarrow 0$
3. while $\Gamma \leq \mathsf{MaxStep}$
   - guess
     - either an atom g of q'' and a positive inclusion $I_\text{g} \in \mathcal{T}$, i.e.,
       - either I_g is the DL-Lite$_\mathcal{R}$ translation of a grounded version of a positive inclusion I in $\mathcal{O}$
       - or $I_\text{g} = E_S \sqsubseteq E_{S_\star}$ with $E@S$ occurring in a grounded version of a positive inclusion I in $\mathcal{O}$ and $S_\star$ is the closed specifier with attribute-value pairs occurring in S plus $\star\colon\star$.
       - or $I_\text{g} = E_S \sqsubseteq E_T$ where T is an open specifier and the set of attribute-value pairs in S is a superset of the set of attribute-value pairs in T
     - or "unify" together with two atoms g_1, g_2 of q''
   - if a positive inclusion I_g applicable to g has been guessed, let $q'' \leftarrow q''[g/gr(g, I_\text{g})]$
   - if "unify" has been guessed with two atoms g_1, g_2 of q'' that unify, let $q'' \leftarrow \tau(\mathsf{reduce}(q'', g_1, g_2))$
   - if $q'' \equiv q'$ return true
   - $\Gamma \leftarrow \Gamma + 1$
4. return false

where $\mathsf{MaxStep} = |q| \cdot (N_{\text{gr}\star} + N_{\text{gr}\star}^2) + |q|$ with $N_{\text{gr}\star} = 3 \cdot |\mathcal{O}| \cdot 2^{|\mathsf{N}_\mathsf{I}^\mathcal{O}|^2} + |q|$.

We can encode Γ in binary in polynomial space and the size of q'' is always at most $|\mathsf{dl}(q_\mathcal{Z})|$ so the algorithm uses only polynomial space. Moreover, q' is in $\mathsf{PerfectRef}(\mathsf{dl}(q_\mathcal{Z}), \mathcal{T})$ iff there exists a sequence of rewriting steps from $\mathsf{dl}(q_\mathcal{Z})$ to q' such that each step consists either in applying a positive inclusion $I_\text{g} \in \mathcal{T}$ to an atom g of the query or in unifying two atoms with $\tau \circ \mathsf{reduce}$.

The number of rewriting steps is at most $|q| \cdot |\mathcal{T}| + |q|$. Indeed, each step either applies a positive inclusion of $\mathcal{T}$ to an atom of the current rewriting, whose size is bounded by $|q|$ (and a positive inclusion is not applied several times to the same atom), or unifies two atoms (and the number of unification steps is bounded by the number of atoms in the initial query). Moreover, there are at most $N = 2^{|\mathsf{N}_\mathsf{I}^\mathcal{O}|^2}$ possible grounded versions for each positive inclusion in $\mathcal{O}$, so $\mathsf{dl}(\mathsf{gr}(\mathcal{O})_\star)$ contains at most $N_{\text{gr}\star} = 3 \cdot |\mathcal{O}| \cdot N + |q|$ inclusions translated from $\mathsf{gr}(\mathcal{O})_\star$. That is, at most $|\mathcal{O}| \cdot N$ inclusions from $\mathsf{gr}(\mathcal{O})$, at most $2 \cdot |\mathcal{O}| \cdot N$ inclusions of the form $E@S \sqsubseteq E@S_\star$, and at most $|q|$ inclusions to add the annotated concept/role names that occur in $q_\mathcal{Z}$ in $\mathsf{gr}(\mathcal{O})_\star$. To model the behavior of the open specifiers occurring in $\mathsf{gr}(\mathcal{O})_\star$ we add further $N_{\text{gr}\star}^2$ inclusions. It follows that the size of $\mathcal{T}$ is bounded by $N_{\text{gr}\star} + N_{\text{gr}\star}^2$. ☐

Theorem 4. *For ground DL-Lite$^\mathcal{R}_@$ ontologies, satisfiability is in* PTIME *and BCQ$_@$ entailment is* NP*-complete.*

Proof. The satisfiability upper bound comes from Claim 2 in the proof of Theorem 2, since the translation of a ground ontology into an equisatisfiable DL-Lite$_\mathcal{R}$ ontology is in PTIME, as well as the satisfiability problem in DL-Lite$_\mathcal{R}$.

Let $\mathcal{O}$ be a DL-Lite$^\mathcal{R}_@$ ground ontology, and let q be a BCQ$_@$. By Lemma 5, $\mathcal{O} \models q$ iff $\mathsf{gr}(\mathcal{O})_\star \models \mathsf{gr}(q)$, i.e., iff $\mathcal{O}_\star \models \mathsf{gr}(q)$ since $\mathsf{gr}(\mathcal{O}) = \mathcal{O}$. It follows that $\mathcal{O} \models q$ iff there exists $q_\mathcal{Z} \in \mathsf{gr}(q)$ such that $\mathcal{O}_\star \models q_\mathcal{Z}$. Since $q_\mathcal{Z}$ has the same size as the original query q, we can add the annotated concept or role names that occur in $q_\mathcal{Z}$ to $\mathcal{O}_\star$ as explained in Section "Notation in the Appendix" while keeping $\mathcal{O}_\star$ of polynomial size w.r.t. $\mathcal{O}$ and q. Thus by Lemma 2, we have that $\mathcal{O} \models q$ iff $\mathsf{dl}(\mathcal{O}_\star) \models \mathsf{dl}(q_\mathcal{Z})$ and $\mathsf{dl}(\mathcal{O}_\star)$ is polynomial in the size of $\mathcal{O}$ and q. It is then possible to decide whether $\mathcal{O} \models q$ by guessing the DL-Lite$_\mathcal{R}$ translation $\mathsf{dl}(q_\mathcal{Z})$ of a grounded version $q_\mathcal{Z}$ of q together with a certificate that $\mathsf{dl}(\mathcal{O}_\star) \models \mathsf{dl}(q_\mathcal{Z})$.

Checking that $\mathsf{dl}(q_\mathcal{Z})$ is indeed a DL-Lite$_\mathcal{R}$ translation of a grounded version of q and certifying that $\mathsf{dl}(\mathcal{O}_\star) \models \mathsf{dl}(q_\mathcal{Z})$ can be done in polynomial time since BCQ entailment in DL-Lite$_\mathcal{R}$ is in NP. The NP lower bound comes from BCQ entailment for DL-Lite$_\mathcal{R}$ ontologies. ☐

Proof of Theorem 8

This section presents an alternative proof for Theorem 8 that extends the results by Botoeva, Artale, and Calvanese (2010) for DL-Lite$^\mathcal{R}_\mathsf{Horn}$ (the extension of DL-Lite$_\mathcal{R}$ with conjunctions of concepts on the left of inclusions) with *role conjunctions in role inclusions*. We use this result to show another result in the Section 'Reasoning in DL-Lite$^\mathcal{R}_{@,\mathbb{K}}$'.

Recall that DL-Lite$^{\mathcal{R},\sqcap}_\mathsf{Horn}$ is a DL which extends DL-Lite$_\mathcal{R}$ with conjunctions on the left side of concept and role inclusions. That is, expressions of the form:

$$I = K_1 \sqcap \cdots \sqcap K_n \sqsubseteq L,$$

with I a concept or a role inclusion. We start by defining the canonical model of a DL-Lite$^{\mathcal{R},\sqcap}_\mathsf{Horn}$ ontology. As we did before, for each role name occurring in $\mathcal{O}$ we add to $\mathcal{O}$ an equivalence $R^- \equiv \overline{R}$, where $\overline{R}$ is a fresh role name. We can then assume w.l.o.g. that inverse roles only occur in such equivalences by replacing R^- in other places by $\overline{R}$.

Definition 8 (Canonical Model of a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology)**.** *The canonical model $\mathcal{I}_\mathcal{O}$ of a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology $\mathcal{O}$ is the union of interpretations $\mathcal{I}_0, \mathcal{I}_1, \ldots$, with $\mathcal{I}_0$ defined as:*

$$\Delta^{\mathcal{I}_0} := \mathsf{N_I},\ a^{\mathcal{I}_0} = a \text{ for every } a \in \mathsf{N_I}$$
$$\text{and } E^{\mathcal{I}_0} := \{\boldsymbol{a}^{\mathcal{I}_0} \mid E(\boldsymbol{a}) \in \mathcal{O}\}$$

for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$. For the inductive definition of the sequence assume $\mathcal{I}_n$ is defined. Then obtain $\mathcal{I}_{n+1}$ by applying the one of the following rules:

- *if $I = A_1 \sqcap \cdots \sqcap A_k \sqsubseteq A$ is a positive concept inclusion in $\mathcal{O}$, and $a \in A_1^{\mathcal{I}_n}, \ldots, a \in A_k^{\mathcal{I}_n}$ but $a \notin A^{\mathcal{I}_n}$ then add a to $A^{\mathcal{I}_n}$,*
- *if $I = R_1 \sqcap \cdots \sqcap R_k \sqsubseteq R$ is a positive role inclusion in $\mathcal{O}$, and $(a, b) \in R_1^{\mathcal{I}_n}, \ldots (a, b) \in R_k^{\mathcal{I}_n}$ but $(a, b) \notin R^{\mathcal{I}_n}$ then add (a, b) to $R^{\mathcal{I}_n}$,*
- *$I = R_1 \sqsubseteq R^-$ or $I = R_1^- \sqsubseteq R$ is a positive role inclusion in $\mathcal{O}$, and $(a, b) \in R_1^{\mathcal{I}_n}$ but $(b, a) \notin R^{\mathcal{I}_n}$ then add (b, a) to $R^{\mathcal{I}_n}$,*
- *if $I = \exists R \sqsubseteq A$ is a positive concept inclusion in $\mathcal{O}$, and there exists b such that $(a, b) \in R^{\mathcal{I}_n}$ but $a \notin A^{\mathcal{I}_n}$ then add a to $A^{\mathcal{I}_n}$,*
- *if $I = A \sqsubseteq \exists R$ is a positive concept inclusion in $\mathcal{O}$, and $a \in A^{\mathcal{I}_n}$ but there is no b such that $(a, b) \in R^{\mathcal{I}_n}$ then add a fresh element b to $\Delta^{\mathcal{I}_n}$ and add (a, b) to $R^{\mathcal{I}_n}$,*
- *if $I = \exists R_1 \sqsubseteq \exists R$ is a positive concept inclusion in $\mathcal{O}$, and there exists b such that $(a, b) \in R_1^{\mathcal{I}_n}$ but there is no c such that $(a, c) \in R^{\mathcal{I}_n}$ then add a fresh element c to $\Delta^{\mathcal{I}_n}$ and add (a, c) to $R^{\mathcal{I}_n}$.*

We assume that rule application is fair, that is, if a rule is applicable in a certain place, then it will indeed eventually be applied there. We obtain $\mathcal{I}_\mathcal{O}$ by setting for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$:

$$\Delta^{\mathcal{I}_\mathcal{O}} := \bigcup_{n \geq 0} \Delta^{\mathcal{I}_n} \text{ and } E^{\mathcal{I}_\mathcal{O}} := \bigcup_{n \geq 0} E^{\mathcal{I}_n}.$$

Theorem 11. *Let $\mathcal{I}_\mathcal{O}$ be the canonical model of a satisfiable DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology $\mathcal{O}$. For any BCQ q, $\mathcal{I}_\mathcal{O} \models q$ iff $\mathcal{O} \models q$.*

Importantly, if $b \in \Delta^{\mathcal{I}_\mathcal{O}} \setminus \mathsf{N_I}$ then, by construction of the canonical model, b has been introduced when applying a concept inclusion whose right side is of the form $\exists R$. Thus, it can only be in the extension of concepts or roles derived from R or $\exists R^-$ w.r.t. the set of positive inclusions of $\mathcal{O}$. In particular, if $(a, b) \in R_1^{\mathcal{I}_\mathcal{O}}$ and $(a, b) \in R_2^{\mathcal{I}_\mathcal{O}}$, there must be some S such that $(a, b) \in S^{\mathcal{I}_\mathcal{O}}$ and $\mathcal{T} \models S \sqsubseteq R_1$, $\mathcal{T} \models S \sqsubseteq R_2$.

We show Theorem 8 by reduction to the problem of BCQ entailment from DL-Lite$^{\mathcal{R}}_{\mathsf{Horn}}$ ontologies, where conjunctions are allowed only in concept inclusions and for which BCQ entailment is NP-complete (Botoeva, Artale, and Calvanese 2010).

Given a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology $\mathcal{O}$, we denote by $\mathcal{T}$ the set of concept and role inclusions in $\mathcal{O}$ (TBox), and by $\mathcal{A}$ the set of assertions in $\mathcal{O}$ (ABox).

Let $\mathcal{A}_\mathsf{s}$ be the ABox obtained by saturating $\mathcal{A}$ w.r.t. role inclusions as follows. We start with $\mathcal{A}_\mathsf{s} = \mathcal{A}$ and repeat the following process until we reach a fix point: for every role inclusion $R_1 \sqcap \cdots \sqcap R_n \sqsubseteq R$ (resp. $R_1 \sqsubseteq R^-$), we check for every pair of individuals (a, b) of $\mathcal{A}$ whether $R_1(a, b), \ldots, R_n(a, b) \in \mathcal{A}_\mathsf{s}$ (resp. $R_1(a, b) \in \mathcal{A}_\mathsf{s}$) and $R(a, b) \notin \mathcal{A}_\mathsf{s}$ (resp. $R(b, a) \notin \mathcal{A}_\mathsf{s}$) and add $R(a, b)$ (resp. $R(b, a)$) to $\mathcal{A}_\mathsf{s}$ if it is the case. This process terminates in polynomial time and the size of $\mathcal{A}_\mathsf{s}$ is polynomial w.r.t. $\mathcal{T}$ and $\mathcal{A}$.

Let $\mathcal{T}_\mathsf{s}$ be the TBox obtained by adding to $\mathcal{T}$ all inclusions of the form $R \sqsubseteq R'$ such that $\mathcal{T} \models R \sqsubseteq R'$, then removing all role inclusions $R_1 \sqcap \cdots \sqcap R_n \sqsubseteq R$ such that $n > 1$. The construction of $\mathcal{T}_\mathsf{s}$ can also be done in polynomial time because there are a polynomial number of inclusions of the form $R \sqsubseteq R'$ such that R, R' occur in $\mathcal{T}$, and the check that $\mathcal{T} \models R \sqsubseteq R'$ is in polynomial time (Artale et al. 2009).

Lemma 6. *The DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology $\langle \mathcal{T}, \mathcal{A} \rangle$ and the DL-Lite$^{\mathcal{R}}_{\mathsf{Horn}}$ ontology $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle$ entail the same conjunctive queries.*

Proof. Let q be a conjunctive query. Since $\mathcal{T} \models \mathcal{T}_\mathsf{s}$ and $\langle \mathcal{T}, \mathcal{A} \rangle \models \mathcal{A}_\mathsf{s}$, clearly, the models of $\langle \mathcal{T}, \mathcal{A} \rangle$ are also models of $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle$. It follows that if $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle \models q$ then $\langle \mathcal{T}, \mathcal{A} \rangle \models q$.

In the other direction, assume that $\langle \mathcal{T}, \mathcal{A} \rangle \models q$ and let $\mathcal{I}$ be the canonical model of $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle$. Assume for a contradiction that $\mathcal{I} \not\models q$. It follows that $\mathcal{I}$ is not a model of $\langle \mathcal{T}, \mathcal{A} \rangle$. Since $\mathcal{A} \subseteq \mathcal{A}_\mathsf{s}$, $\mathcal{I}$ is a model of $\mathcal{A}$, so $\mathcal{I}$ is not a model of $\mathcal{T}$. Since the only inclusions of $\mathcal{T}$ that are not in $\mathcal{T}_\mathsf{s}$ are of the form $R_1 \sqcap \cdots \sqcap R_n \sqsubseteq R$ with $n > 1$, all violations of $\mathcal{T}$ by $\mathcal{I}$ are of the following form: there exists such an inclusion and two elements $a, b \in \Delta^\mathcal{I}$ such that $(a, b) \in R_1^\mathcal{I}, \ldots, (a, b) \in R_n^\mathcal{I}$ and $(a, b) \notin R^\mathcal{I}$. We show that this is impossible distinguishing between the cases where a, b are images of individual names or not (recall that if $a \in \mathsf{N_I}$, $a^\mathcal{I} = a$):

- If $a, b \in \mathsf{N_I}$, since $\mathcal{I}$ is the canonical model of $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle$, $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle \models R_i(a, b)$ for $1 \leq i \leq n$. Since $\mathcal{A}_\mathsf{s}$ is saturated w.r.t. the role inclusions in $\mathcal{T}$ (and thus in $\mathcal{T}_\mathsf{s}$), it follows that $R_1(a, b), \ldots, R_n(a, b) \in \mathcal{A}_\mathsf{s}$ and $R(a, b) \in \mathcal{A}_\mathsf{s}$. Thus $(a, b) \in R^\mathcal{I}$.
- If $a \notin \mathsf{N_I}$ or $b \notin \mathsf{N_I}$, by construction of the canonical model $\mathcal{I}$ of $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle$, there is some role R' such that $(a, b) \in R'^\mathcal{I}$ and $\mathcal{T}_\mathsf{s} \models R' \sqsubseteq R_1, \ldots, \mathcal{T}_\mathsf{s} \models R' \sqsubseteq R_n$. Thus $\mathcal{T} \models R' \sqsubseteq R_1, \ldots, \mathcal{T} \models R' \sqsubseteq R_n$ and since $\mathcal{T} \models R_1 \sqcap \cdots \sqcap R_n \sqsubseteq R$, then $\mathcal{T} \models R' \sqsubseteq R$, so by construction of $\mathcal{T}_\mathsf{s}$, $R' \sqsubseteq R \in \mathcal{T}_\mathsf{s}$. Since $\mathcal{I}$ is a model of $\mathcal{T}_\mathsf{s}$, it follows that $(a, b) \in R^\mathcal{I}$.

It follows that $\mathcal{I}$ is a model of $\langle \mathcal{T}, \mathcal{A} \rangle$, so $\mathcal{I} \models q$, i.e., $\langle \mathcal{T}_\mathsf{s}, \mathcal{A}_\mathsf{s} \rangle \models q$. □

Theorem 8. *In DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$, BCQ entailment is in* NP.

Proof. It follows from Lemma 6 that BCQ entailment from DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ can be reduced polynomially to BCQ entailment in DL-Lite$^{\mathcal{R}}_{\mathsf{Horn}}$ which is NP-complete (Botoeva, Artale, and Calvanese 2010). □

Proofs for Section 'Reasoning in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$'

In this section, we sometimes write $+$ or $\sum$ instead of $+^p$ or $\sum^p$ when p is clear from the context (when summing annotations sets differentiated by p in some $F \in \Phi^{\mathcal{I}}$).

Theorem 5. *In DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, satisfiability is* ExpTime*-hard.*

Proof. We reduce from the word problem for a polynomially space bounded alternating Turing Machine (ATM), which is ExpTime-hard (Chandra, Kozen, and Stockmeyer 1981). An ATM is a tuple $\mathcal{M} = (Q, \Sigma, \Theta, q_0)$, where $Q = Q_\exists \uplus Q_\forall$ is a finite set of states, partitioned into existential states $Q_\exists$ and universal states $Q_\forall$, Σ is a finite alphabet containing the blank symbol $\llcorner\!\lrcorner$, $q_0 \in Q$ is the initial state, and $\Theta \subseteq (Q \times \Sigma) \times (Q \times \Sigma) \times \{l, r\}$ is the transition relation.

We assume that $\mathcal{M}$ is polynomially space bounded on inputs $w_0 = \sigma_0 \ldots \sigma_{n-1}$ of length n. We use the same notion of configuration, computation and initial configuration given in the proof of Theorem 1 and also make the assumption that $\mathcal{M}$ does not attempt to move to the left (right) when it is on its left-most (right-most) tape position. We now recall here the acceptance condition of an ATM, following the terminology provided in (Krötzsch, Rudolph, and Hitzler 2013). A configuration $\alpha = wqw'$ is *accepting* iff:

- α is a universal configuration and all its successor configurations are accepting, or
- α is an existential configuration and at least one of its successor configurations is accepting.

By the definition above, universal configurations without any successors are accepting. We consider w.l.o.g. ATMs with only finite computations on any input (Chandra, Kozen, and Stockmeyer 1981). $\mathcal{M}$ *accepts* a word in Σ^* (using space polynomial in the size of the input) iff the initial configuration is accepting.

We construct a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology $\mathcal{O}_{\mathcal{M},w_0}$ that is satisfiable iff $\mathcal{M}$ accepts w_0. We use the following symbols in addition to those introduced in the proof of Theorem 1:

- attribute values $\theta \in \Theta$ to represent the transitions;
- an attribute acc_Θ with values which are either elements of Θ or a sum of elements of Θ;
- finally, an attribute acc with value 1 to mark accepting configurations.

We encode the initial configuration and transitions $\theta = (q, \sigma, q', \tau, D) \in \Theta$ as in the proof of Theorem 1. To encode the acceptance condition, we compute the set of accepting configurations *backwards*. First, we ensure that if a configuration is accepting then the predecessor configuration w.r.t. to a transition $\theta \in \Theta$ will have this information. To encode this, we add to $\mathcal{O}_{\mathcal{M},w_0}$ concept inclusions of the form (we show the case where $D = r$, the case with $D = l$ is analogous), for $\theta = (q, \sigma, q', \tau, D) \in \Theta$ and $1 \leq k + 1 \leq m$:

$$\Omega\ (A@X \sqsubseteq A@[\mathsf{acc}_\Theta : \theta, s\colon q, h\colon k, p_k\colon \sigma, P_{X\setminus k}])$$

where Ω is a shorthand for:

$$X : \lfloor \mathsf{acc}\colon 1, s\colon q', h\colon k+1, p_k\colon \tau \rfloor$$

and $P_{X\setminus k}$ is the abbreviation used in the proof of Theorem 1.

We now mark accepting configurations with the attribute-value pair $\mathsf{acc}\colon 1$. For universal configurations, we have, for each $q \in Q_\forall$ and $0 \leq k \leq m$, a concept inclusion:

$$\Omega\ (A@X \sqsubseteq A@[\mathsf{acc}\colon 1, s\colon q, h\colon k, p_k\colon \sigma, P_{X\setminus k}])$$

where Ω is a shorthand for:

$$X : \lfloor \mathsf{acc}_\Theta \colon \sum_{\substack{\theta \in \Theta,\\ \theta = (q,\sigma,q',\tau,D)}} \theta, s\colon q, h\colon k, p_k\colon \sigma \rfloor$$

and we omit the attribute acc_Θ and its value above if there is no suitable transition. For existential configurations, we have, for each $q \in Q_\exists$, $\theta = (q, \sigma, q', \tau, D) \in \Theta$ and $0 \leq k \leq m$, a concept inclusion:

$$\Omega\ (A@X \sqsubseteq A@[\mathsf{acc}\colon 1, s\colon q, h\colon k, p_k\colon \sigma, P_{X\setminus k}])$$

where Ω is a shorthand for:

$$X : \lfloor \mathsf{acc}_\Theta \colon \theta, s\colon q, h\colon k, p_k\colon \sigma \rfloor.$$

Claim. $\mathcal{O}_{\mathcal{M},w_0} \models A(a)@S$ iff $\mathcal{M}$ accepts w_0, where S is $[\mathsf{acc}\colon 1, s\colon q_0, h\colon 0, p_0\colon \sigma_0, \ldots, p_{n-1}\colon \sigma_{n-1}, p_n\colon \llcorner\!\lrcorner, \ldots, p_m\colon \llcorner\!\lrcorner]$.

One can reduce Boolean atomic query entailment to (un)satisfiability, so the claim implies this theorem. □

To prove Theorem 7, we first devise a grounding strategy for DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ and then translate it into DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$. We denote by $\mathsf{N}_\mathsf{S}^\mathcal{O}$ the subset of N_S with elements occurring in $\mathcal{O}$. Importantly, we assume throughout this section that $\mathsf{N}_\mathsf{S}^\mathcal{O}$ does not contain multiple provenance sums which are mathematically equal. This is w.l.o.g. since if $a+b$ and $b+a$ both occur in $\mathcal{O}$ then one can always replace the latter by the former (or vice-versa). Let $\mathsf{N}_{\mathsf{P}_{\mathsf{min}}}$ be a fixed but arbitrary minimal subset of N_P such that for each $a \in \mathsf{N}_\mathsf{P}$, $\mathsf{N}_{\mathsf{P}_{\mathsf{min}}}$ contains an element b such that a is mathematically equal to b. Define $\mathsf{N}_\mathsf{S}^{\mathcal{O}+}$ by:

$$\mathsf{N}_\mathsf{S}^\mathcal{O} \cup \{\textstyle\sum_{b \in \sigma} b \mid \sigma \subseteq \mathsf{N}_\mathsf{S}^\mathcal{O}\}$$

Assume w.l.o.g. that elements of $\mathsf{N}_\mathsf{S}^{\mathcal{O}+}$ are among elements of $\mathsf{N}_{\mathsf{P}_{\mathsf{min}}}$. Also, assume w.l.o.g. that we respect the multiplicity of elements in $\mathcal{O}$ and include their sum in $\mathsf{N}_\mathsf{S}^{\mathcal{O}+}$. This can be done by first replacing each element of N_S occurring in $\mathcal{O}$ with an alias, so that they are unique and then replacing the alias by the original element of N_S. We may write $K \equiv L$ as a bidirectional $\sqsubseteq$ as usual; and $K \sqsubseteq L \sqcap L'$ as a shorthand for $K \sqsubseteq L$ and $K \sqsubseteq L'$. To show our upper bound we first identify some model theoretical properties of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$.

Definition 9 (Interpretation of a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ concept/role)**.** *We define the interpretation of a concept or a role expression L in an inclusion I w.r.t. a variable assignment $\mathcal{Z}$ in the same way as in Definition 4, except that we use $(\star_\mathcal{S}^{\mathcal{M}_n}, \star_\mathcal{S}^{\mathcal{M}_n})$ instead of $(\star^{\mathcal{M}_n}, \star^{\mathcal{M}_n})$ in $L^{\mathcal{I}^{\mathcal{Z},\mathcal{M}_n}_{L,I}}$ where $\mathcal{S} = \{a\colon b \mid (a^{\mathcal{M}_n}, b^{\mathcal{M}_n}) \in H, \{a, b\} \subseteq \mathsf{N}_\mathsf{S}^{\mathcal{O}+}\}$.*

The canonical model $\mathcal{M}_\star^\mathcal{O}$ of a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology $\mathcal{O}$ is constructed similarly. However we need to ensure that $\mathcal{M}_\star^\mathcal{O}$ is a well-founded provenance interpretation.

Definition 10 (Canonical Model of a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology)**.** *The canonical model $\mathcal{M}^{\mathcal{O}}_{\star}$ of a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology $\mathcal{O}$ is the union of provenance-interpretations $\mathcal{M}_0, \mathcal{M}_1, \ldots$, defined as in Definition 5, except that $\Delta^{\mathcal{M}_0} := \mathsf{N}_{\mathsf{P}_{\min}}$, and for all $a \in \mathsf{N}_\mathsf{P}$, $a^{\mathcal{M}_0} = b$ with a mathematically equal to b. Assuming $\mathcal{M}_n$ is defined (using Definition 9 for $E^{\mathcal{I}^{\mathcal{Z},\mathcal{M}_n}_{K,I}}$) we extend $\mathcal{M}_{n+1}$ with the following step:*

- *add to $E^{\mathcal{M}_{n+1}}$ all tuples in*
$$\{(\boldsymbol{a}, \textstyle\sum_{G \in \sigma} G) \mid \sigma \subseteq E^p_{\mathcal{M}_n, \boldsymbol{a}, F} \neq \emptyset\},$$
for all $p \in \Delta^{\mathcal{M}_n}$ and all $F \in \Phi^{\mathcal{M}_n}$.

It follows from our construction that the number of possible annotation sets in $\mathcal{M}^{\mathcal{O}}_{\star}$ is finite, therefore it also satisfies the support condition for all $E \in \mathsf{N}_\mathsf{C} \cup \mathsf{N}_\mathsf{R}$ and tuples with elements in $\Delta^{\mathcal{M}^{\mathcal{O}}_{\star}}$.

We now define the *grounding of a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology*. Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology. Let $\mathcal{I}$ be an interpretation over the domain $\Delta^{\mathcal{I}} = \mathsf{N}^{\mathcal{O}+}_\mathsf{S} \cup \{\star\}$ and such that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N}^{\mathcal{O}+}_\mathsf{S} \cup \{\star\}$ (such an interpretation exists since $\mathsf{N}^{\mathcal{O}+}_\mathsf{S}$ contains at most one representative of equal provenance sums). This interpretation is not a provenance-interpretation because we restrict the domain but is going to be used to define the grounding of $\mathcal{O}$. Let $\mathcal{Z}$ be a variable assignment mapping object variables $x \in \mathsf{N}_\mathsf{V}$ to elements $\mathcal{Z}(x) \in \Delta^{\mathcal{I}} \setminus \{\star^{\mathcal{I}}\}$ and set variables $X \in \mathsf{N}_\mathsf{U}$ to binary relations $\mathcal{Z}(X) \in \Phi^{\mathcal{I}}$ such that $(a,b) \in \mathcal{Z}(X)$ implies that $a \notin \mathcal{I}(\mathsf{N}_\mathsf{S} \setminus \mathsf{N}_\mathsf{I})$. Let $\mathcal{O}'$ be the ground ontology with all $\mathcal{Z}$-instances $I_{\mathcal{Z}}$ for all concept or role inclusions I in $\mathcal{O}$ and all compatible variable assignments $\mathcal{Z}$, where the $\mathcal{Z}$-instance $I_{\mathcal{Z}}$ of I is as defined for DL-Lite$^{\mathcal{R}}_{@}$ ontologies. Let $\mathsf{gr}_+(\mathcal{O})$ be the result of replacing in $\mathcal{O}'$ each occurrence of $\star$ in a specifier S by $\star_S$, where S is the set of attribute-value pairs (without $\star$) occurring in S. Since here we use $\mathsf{N}^{\mathcal{O}+}_\mathsf{S}$ to construct $\mathsf{gr}_+(\mathcal{O})$, the grounding is double exponential in the size of $\mathcal{O}$.

Lemma 7. *$\mathcal{O}$ is satisfiable iff* $\mathsf{gr}_+(\mathcal{O})$ *is satisfiable.*

Proof. By definition of $\mathsf{gr}_+(\mathcal{O})$, it is straightforward to show that $\mathcal{O} \models \mathsf{gr}_+(\mathcal{O})$. We show the converse direction. Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be a model of $\mathsf{gr}_+(\mathcal{O})$. To show this lemma we use the following claim.

Claim 1. If $\mathcal{O}$ has a model then $\mathcal{O}$ has a model $\mathcal{M}$ such that, for all $\{G, H\} \subseteq \Phi^{\mathcal{M}}$ that occur in the interpretation of some concept or role by $\cdot^{\mathcal{M}}$:

1. if G, H are differentiated by some $p \in \Delta^{\mathcal{M}}$ (in $F \in \Phi^{\mathcal{M}}$) then all elements in tuples in G, H are in $\mathcal{M}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S})$;
2. for all $(a,b) \in H$, either $\{a,b\} \subseteq \mathcal{M}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S})$ or $\{a,b\} \cap \mathcal{M}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S}) = \emptyset$.

Proof of Claim 1. If $\mathcal{O}$ is satisfiable then the canonical model $\mathcal{M}^{\mathcal{O}}_{\star}$ for $\mathcal{O}$ is a model of $\mathcal{O}$. One can show by induction that $\mathcal{M}^{\mathcal{O}}_{\star}$ satisfies the properties of this claim.

By Claim 1, we can assume that $\mathcal{I}$ satisfies the conditions in Points 1 and 2. W.l.o.g., assume that for each new symbol $\star_S$ in $\mathsf{gr}_+(\mathcal{O})$, there is $\star^{\mathcal{I}}_S \in \Delta^{\mathcal{I}}$ such that $\star^{\mathcal{I}}_S \neq a^{\mathcal{I}}$ for all $a \in \mathsf{N}^{\mathcal{O}+}_\mathsf{S}$. Given an annotation set $F \in \Phi^{\mathcal{I}}$, we denote by $\mathcal{S}_F$ the maximal set of attribute-values pairs $a : b$ such that $\{a,b\} \subseteq \mathsf{N}^{\mathcal{O}+}_\mathsf{S}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in F$. We define $F_\star$ as the annotation set obtained from F by replacing all $a^{\mathcal{I}} \notin \mathcal{I}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S})$ in F by $\star^{\mathcal{I}}_{\mathcal{S}_F}$. Let $\mathcal{J}$ be the interpretation over the domain $\Delta^{\mathcal{J}} := \Delta^{\mathcal{I}}$ such that $E^{\mathcal{J}} := \{(\boldsymbol{a}, H_\star) \mid (\boldsymbol{a}, H) \in E^{\mathcal{I}}\}$ for all $E \in \mathsf{N}_\mathsf{C} \cup \mathsf{N}_\mathsf{R}$, and $a^{\mathcal{J}} := a^{\mathcal{I}}$ for all $a \in \mathsf{N}_\mathsf{S}$.

Claim 2. $\mathcal{J}$ is a well-founded provenance-interpretation.

Proof of Claim 2. We want to satisfy the condition of Definition 2. Since $\mathcal{I}$ satisfies the conditions of Claim 1, $\mathcal{J}$ is such that, for all $\{G, H\} \subseteq \Phi^{\mathcal{J}}$ that occur in the interpretation of some concept or role by $\cdot^{\mathcal{J}}$:

1. if G, H are differentiated by p in some $F \in \Phi^{\mathcal{J}}$ then all elements in tuples in G, H are in $\mathcal{J}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S})$;
2. for all $(a,b) \in H$, either $\{a,b\} \subseteq \mathcal{J}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S})$ or $\{a,b\} \cap \mathcal{J}(\mathsf{N}^{\mathcal{O}+}_\mathsf{S}) = \emptyset$.

On one hand we know that annotations sets with only elements in $\mathsf{N}^{\mathcal{O}+}_\mathsf{S}$ are not changed in $\mathcal{J}$. On the other hand, if they are changed then they are not differenciated by p anymore. It follows that for all $E \in \mathsf{N}_\mathsf{C} \cup \mathsf{N}_\mathsf{R}$, if $E^{\mathcal{I}}$ is closed under sum then $E^{\mathcal{J}}$ is closed under sum. This finishes the proof of Claim 2.

It remains to show that $\mathcal{J}$ is indeed a model of $\mathcal{O}$. Suppose for a contradiction that there is a concept inclusion I in $\mathcal{O}$ that is not satisfied by $\mathcal{J}$ (the case for role inclusions is analogous). Then we have a compatible variable assignment $\mathcal{Z}$ that leaves I unsatisfied. Let $\mathcal{Z}'$ be the variable assignment $X \mapsto \mathcal{Z}(X)_\star$ for all $X \in \mathsf{N}_\mathsf{U}$. Clearly, $\mathcal{Z}'$ is also compatible with I. But now we have $C^{\mathcal{J},\mathcal{Z}} = C^{\mathcal{I},\mathcal{Z}'}$ for all DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ concepts C, yielding the contradiction $\mathcal{I} \not\models I_{\mathcal{Z}'}$, with $I_{\mathcal{Z}'}$ in $\mathsf{gr}_+(\mathcal{O})$. Thus, $\mathcal{O}$ is satisfiable iff $\mathsf{gr}_+(\mathcal{O})$ is satisfiable. ☐

The ground ontology $\mathcal{O}_\mathsf{g} := \mathsf{gr}_+(\mathcal{O})$ is now translated into a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology $\mathsf{dl}_+(\mathcal{O}_\mathsf{g})$ as follows. First, replace every annotated concept/role name $E@S$ (or inverse role $R^-@S$) with a fresh concept/role name E_S (or inverse role R^-_S) in all the assertions and concept or role inclusions of $\mathcal{O}_\mathsf{g}$. We now would like to capture the semantics of $+$. Let $\mathcal{I}$ be the interpretation with $\Delta^{\mathcal{I}} = \mathsf{N}^{\mathcal{O}+}_\mathsf{S} \cup \{\diamond\}$ where $\diamond$ is some fresh individual name. Again this interpretation is not a provenance-interpretation because we restrict the domain but is going to be used to define the DL translation. Assume $b^{\mathcal{I}} = b$ for all $b \in \mathsf{N}^{\mathcal{O}+}_\mathsf{S}$ and assume there is a concept name A containing $\{(\diamond, F) \mid F \text{ has tuples in } (\mathsf{N}^{\mathcal{O}}_\mathsf{S})^2\}$ in its extension plus all annotation sets so that $A^{\mathcal{I}}$ is closed under sum, as in Definition 2 (this is possible by construction of $\mathsf{N}^{\mathcal{O}+}_\mathsf{S}$). Let $[S]$ be the set of closed specifiers S_F such that for $(\diamond, F) \in A^{\mathcal{I}}$ we have that $a\colon b$ occurs in S_F iff $(a,b) \in F$. In the following, we conversely denote by F_S the annotation set such that $(a,b) \in F_S$ iff $a\colon b$ occurs in S. Recall from Section "Querying using Provenance Semirings" that we denote by $A^p_{\mathcal{I},\diamond,F}$ the set of annotation sets G differentiated by p in $F \in \Phi^{\mathcal{I}}$ such that $(\diamond, G) \in A^{\mathcal{I}}$ with G primitive for $\diamond$ and $A^{\mathcal{I}}$. Note that the size of $A^p_{\mathcal{I},\diamond,F}$ is exponential w.r.t.

the size of $\mathcal{O}$, since the primitive annotation sets are built from elements of $\mathsf{N}_\mathsf{S}^{\mathcal{O}}$. For each $E \in \mathsf{N_C} \cup \mathsf{N_R}$ occurring in $\mathcal{O}$ and each $\sigma \subseteq A^p_{\mathcal{I},\diamond,F}$, where $p \in \mathsf{N}_\mathsf{S}^{\mathcal{O}}$ and F with tuples in $(\mathsf{N}_\mathsf{S}^{\mathcal{O}+})^2$, we add:

$$\prod_{F_U \in \sigma} E_U \sqsubseteq \prod_{F_V \in \tau(\sigma)} E_V; \qquad (2)$$

where $\tau(\sigma) = \{\sum_{G \in \upsilon}^{p} G \mid \upsilon \subseteq \sigma\}$. Finally, extend the obtained DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology $\mathsf{dl}_+(\mathcal{O}_\mathsf{g})$ by all axioms:

1. $E_S \sqsubseteq E_T$ where E_S and E_T occur in translated axioms of $\mathsf{dl}_+(\mathcal{O}_\mathsf{g})$ and $S \Rightarrow T$;

which again introduces at most double-exponentially many concept/role inclusions *in the size of* $\mathcal{O}$ (for each E occurring in $\mathcal{O}$ we introduce at most double-exponentially many inclusions of the form of Equation 2).

We denote by $\Gamma_{\mathcal{O},\star}$ the set of individual names of the form $\star_{\mathcal{S}}$ where $\mathcal{S}$ is a set of attribute-value pairs built from elements of $\mathsf{N}_\mathsf{S}^{\mathcal{O}+}$.

For a ground specifier S, we define $F_S^{\mathbb{K}}$ as the annotation set in the singleton set $S^{\mathcal{I}}$ if S is a closed specifier, and $F_S^{\mathbb{K}} = \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid a\colon b \text{ occurs in } S\} \cup \{(\star_{\mathcal{S}}^{\mathcal{I}}, \star_{\mathcal{S}}^{\mathcal{I}})\}$ where $\mathcal{S} = \{a\colon b \mid a\colon b \text{ occurs in } S,\ \{a,b\} \subseteq \mathsf{N}_\mathsf{S}^{\mathcal{O}+}\}$ if S is an open specifier.

Lemma 8. *Let $\mathcal{O}$ be a ground DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology. Then, $\mathcal{O}$ is satisfiable iff $\mathsf{dl}_+(\mathcal{O})$ is satisfiable.*

Proof. Given a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ model of $\mathcal{O}$, we obtain a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ interpretation $\mathcal{J}$ over $\Delta^{\mathcal{I}}$ by setting $E_S^{\mathcal{J}}$ as $E@S^{\mathcal{I}}$ for all E_S that occur in $\mathsf{dl}_+(\mathcal{O})$ and $a^{\mathcal{J}}$ to $a^{\mathcal{I}}$, for all $a \in \mathsf{N_I}$ (here we use only individual names in $\mathsf{N_I}$ since the semantics of individual names in $\mathsf{N_S}$ is captured by the conjunctions). By the semantics of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ and definition of $\mathsf{dl}_+(\mathcal{O})$, clearly $\mathcal{J} \models \mathsf{dl}_+(\mathcal{O})$. Conversely, let $\mathcal{J}$ be a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ model of $\mathsf{dl}_+(\mathcal{O})$. We assume w.l.o.g. that $\mathcal{J}$ maps distinct individuals names to distinct domain elements (unique name assumption). Indeed, it follows from the syntax and semantics of DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ that if a DL-Lite$^{\mathcal{R},\sqcap}_{\mathsf{Horn}}$ ontology has a model, then it has a model which respects the unique name assumption. We construct a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation $\mathcal{I}$ over $\Delta^{\mathcal{I}} := \Delta^{\mathcal{J}} \cup \mathsf{N_{Pmin}}$, with $\Delta^{\mathcal{J}} \cap \mathsf{N_{Pmin}} = \emptyset$ and define $a^{\mathcal{I}} := a^{\mathcal{J}}$, for all $a \in \mathsf{N_I} \setminus \Gamma_{\mathcal{O},\star}$, and $a^{\mathcal{I}} := b$ such that a and $b \in \mathsf{N_{Pmin}}$ are mathematically equal, for all $a \in (\mathsf{N_P} \setminus \mathsf{N_I}) \cup \Gamma_{\mathcal{O},\star}$. Now, let $E^{\mathcal{I}} := \{(\boldsymbol{a}, F_S^{\mathbb{K}}) \mid \boldsymbol{a} \in E_S^{\mathcal{J}} \text{ for some specifier } S\}$ if E occurs in $\mathcal{O}$ and $E^{\mathcal{I}} = \emptyset$, otherwise.

Claim *$\mathcal{I}$ is a well-founded provenance-interpretation.*

Proof of the Claim. It follows from the definition of $\mathsf{dl}_+(\mathcal{O})$ that the condition of Definition 2 is satisfied by Equation 2. Thus, for all $E \in \mathsf{N_C} \cup \mathsf{N_R}$ we have that $E^{\mathcal{I}}$ is closed under sum. The support condition is also satisfied since only finitely many annotation sets occur in $\mathcal{I}$. Point 1 of the translation is also necessary because of the semantics of specifiers. This finishes the proof of the Claim.

By definition of $\mathcal{I}$, in particular of $F_S^{\mathbb{K}}$, for all concept/role name E and all $S \in \mathbf{S}_\mathsf{S}$, $E@S^{\mathcal{I}} = E_S^{\mathcal{J}}$. Thus, with an inductive argument one can show that $\mathcal{I}$ is a model of $\mathcal{O}$. □

Lemmas 7 and 8 together show the correctness of our reduction. Since the translation is double exponential we obtain our upper bound in Theorem 7. We now want to show the correctness of the reduction for the query entailment problem (Theorem 9).

We first establish in Lemma 9 that ground and plain BCQ$_{@,\mathbb{K}}$s can be translated into BCQs. Recall that whenever we speak about a ground and plain query q and a ground ontology $\mathcal{O}$, we assume w.l.o.g. that if an annotated concept or role name occurs in q then it also occurs in $\mathcal{O}$. Note that in the case of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$, we may need to replace equal provenance sums that occur in the query by the representative of these sums that occurs in the ontology (e.g., if $a+b$ and $b+a$ both occur in q and $a+b$ occurs in $\mathcal{O}$, we replace $b+a$ with $a+b$ in the query).

Given a ground and plain BCQ$_{@,\mathbb{K}}$ q, we define $\mathsf{dl}_+(q)$ as $\mathsf{dl}(q')$, where q' is the result of removing the outer specifier of q (which is of the form $\lfloor\ \rfloor$) and the translation $\mathsf{dl}(\cdot)$ is as defined for ground BCQ$_@$s. Recall that we denote by $\Gamma_{\mathcal{O},\star}$ the set of individual names of the form $\star_{\mathcal{S}}$ where $\mathcal{S}$ is a set of attribute-value pairs built from elements of $\mathsf{N}_\mathsf{S}^{\mathcal{O}+}$.

Lemma 9. *Let $\mathcal{O}$ be a ground DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology and let q be a ground and plain BCQ$_{@,\mathbb{K}}$. $\mathcal{O} \models q$ iff $\mathsf{dl}_+(\mathcal{O}) \models \mathsf{dl}_+(q)$.*

Proof. Assume $\mathsf{dl}_+(\mathcal{O}) \models \mathsf{dl}_+(q)$. If an interpretation $\mathcal{I}$ models $\mathcal{O}$ then we obtain a model of $\mathsf{dl}_+(\mathcal{O})$ by setting $E_S^{\mathcal{J}}$ as $E@S^{\mathcal{I}}$ for all E_S that occur in $\mathsf{dl}_+(\mathcal{O})$ and $a^{\mathcal{J}}$ to $a^{\mathcal{I}}$, for all $a \in \mathsf{N_I}$ (as in Lemma 8, we use only individual names in $\mathsf{N_I}$ since the semantics of individual names in $\mathsf{N_S}$ is captured by the conjunctions). By the semantics of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ and by definition of $\mathsf{dl}_+(\mathcal{O})$, we have that $\mathcal{J} \models \mathsf{dl}_+(\mathcal{O})$. Then, by assumption, $\mathcal{J} \models \mathsf{dl}_+(q)$. Recall that every interpretation of DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ is also an interpretation of DL-Lite$^{\mathcal{R}}_{@}$ (but not the converse). By Definition 3, if a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation satisfies a BCQ$_@$ q then it satisfies the query q' that is the result of removing the outer specifier of q. If q is a plain BCQ$_{@,\mathbb{K}}$ then, clearly, if a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation satisfies q' then it satisfies q. Thus, by following the same lines of the proof of Lemma 2, we have that $\mathcal{I} \models q$, and so $\mathcal{O} \models q$.

Conversely, assume $\mathcal{O} \models q$. Let $\mathcal{M}$ be a model of $\mathsf{dl}_+(\mathcal{O})$. There exists a model $\mathcal{J}$ of $\mathsf{dl}_+(\mathcal{O})$ which respects the unique name assumption and such that $\mathcal{J} \to \mathcal{M}$. We obtain a model $\mathcal{I}$ of $\mathcal{O}$ over $\Delta^{\mathcal{I}} := \Delta^{\mathcal{J}} \cup \mathsf{N_{Pmin}}$, with $\Delta^{\mathcal{J}} \cap \mathsf{N_{Pmin}} = \emptyset$, and define $a^{\mathcal{I}} := a^{\mathcal{J}}$, for all $a \in \mathsf{N_I} \setminus \Gamma_{\mathcal{O},\star}$, and $a^{\mathcal{I}} := b$ such that a and $b \in \mathsf{N_{Pmin}}$ are mathematically equal, for all $a \in (\mathsf{N_P} \setminus \mathsf{N_I}) \cup \Gamma_{\mathcal{O},\star}$. Now let $E^{\mathcal{I}} = \{(\boldsymbol{a}, F_S^{\mathbb{K}}) \mid \boldsymbol{a} \in E_S^{\mathcal{J}} \text{ for some specifier } S\}$ if E occurs in $\mathcal{O}$ and $E^{\mathcal{I}} = \emptyset$, otherwise. As in Lemma 8 one can show that $\mathcal{I}$ is a well-founded provenance-interpretation. By assumption, $\mathcal{I} \models q$. As we argued above we can see $\mathcal{I}$ and q as a DL-Lite$^{\mathcal{R}}_{@}$ interpretation and a BCQ$_@$ (by removing the outer specifier $\lfloor\ \rfloor$). Thus, by following the same lines of the proof of Lemma 2, we have that $\mathcal{J} \models \mathsf{dl}_+(q)$. It follows that $\mathcal{M} \models \mathsf{dl}_+(q)$, and so $\mathsf{dl}_+(\mathcal{O}) \models \mathsf{dl}_+(q)$. □

We define the grounding $\mathsf{gr}_+(q)$ of a *plain* BCQ$_{@,\mathbb{K}}$ q in the same way as for BCQ$_@$s. The main difference is that here we consider $\Delta^{\mathcal{I}} = \mathsf{N}_{\mathsf{P}_{\mathsf{min}}}$ and impose the condition that $\mathcal{Z}$-instances are defined in terms of variable assignments $\mathcal{Z}$ over $\mathcal{I}$ with image in $\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+} \cup \{\star\}$ and such that $(a,b) \in \mathcal{Z}(X)$ implies that $a \notin \mathcal{I}(\mathsf{N}_{\mathsf{S}} \setminus \mathsf{N}_{\mathsf{I}})$. Recall that we assume that the elements from N_{S} that occur in q also occur in $\mathcal{O}$. As we did for DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontologies, we replace each occurrence of $\star$ in a specifier S by $\star_{\mathcal{S}} \in \Gamma_{\mathcal{O},\star}$. We thus have that $\mathsf{gr}_+(q)$ is a set of ground and plain BCQ$_{@,\mathbb{K}}$s.

For a ground specifier S, recall that we denote by $\mathcal{S}$ the set of attribute-value pairs (without $\star$) occurring in it. We define $\mathcal{O}_{\Gamma_\star}$ as the union of $\mathcal{O}$ and:

1. all concept/role inclusions $E@S \sqsubseteq E@T$ such that E occurs in $\mathcal{O}$, S is a ground (open/closed) specifier whose attribute-value pairs are either $\mathcal{S}$ or $\mathcal{S} \cup \{\star_{\mathcal{S}}:\star_{\mathcal{S}}\}$, and $S \Rightarrow T$;
2. all concept/role inclusions of the form $E@S \sqsubseteq E@S_{\star_{\mathcal{S}}}$ where $E@S$ occurs in $\mathcal{O}$ or in $\mathcal{O}_{\Gamma_\star}$ in Point 1, S is a ground (open/closed) specifier, and $S_{\star_{\mathcal{S}}}$ is the closed specifier with all attribute-value pairs in $\mathcal{S}$ plus $\star_{\mathcal{S}}:\star_{\mathcal{S}}$.

Point 1 introduces double exponentially many inclusions w.r.t. the number of attributes and values occurring in $\mathcal{O}$. Thus, even if we consider $\mathcal{O}' = \mathsf{gr}_+(\mathcal{O})$ we still have that $\mathcal{O}'_{\Gamma_\star}$ is double exponential in the size of $\mathcal{O}$.

We now adapt Definition 6 to use it together with our groundings in this section.

Definition 11. *Let $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ be an interpretation and let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology. Assume w.l.o.g. that for each $\star_{\mathcal{S}} \in \Gamma_{\mathcal{O},\star}$ there is $\star^{\mathcal{I}}_{\mathcal{S}} \in \Delta^{\mathcal{I}}$ such that $\star^{\mathcal{I}}_{\mathcal{S}} \neq a^{\mathcal{I}}$ for all $a \in \mathsf{N}_{\mathsf{S}}^{\mathcal{O}+}$. For an annotation set $F \in \Phi^{\mathcal{I}}$, we define $F_{\star_{\mathcal{S}}}$ to be the annotation set obtained from F by replacing all $a^{\mathcal{I}} \notin \mathcal{I}(\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+})$ in F by $\star^{\mathcal{I}}_{\mathcal{S}}$, where $\mathcal{S}$ is the set of of attribute-value pairs $a : b$ such that $\{a,b\} \subseteq \mathsf{N}_{\mathsf{S}}^{\mathcal{O}+}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in F$. Let $\mathcal{I}^{\mathcal{O}}_{\Gamma_\star} = (\Delta^{\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}}, \cdot^{\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}})$ be as follows:*

- *$\Delta^{\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}} := \Delta^{\mathcal{I}}$; $a^{\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}} := a^{\mathcal{I}}$ for all $a \in \mathsf{N}_{\mathsf{S}}$; and*
- *$E^{\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}} := \{(\boldsymbol{a}, F_{\star_{\mathcal{S}}}), (\boldsymbol{a}, F_{\star_{\mathcal{S}}} \cup \{(\star_{\mathcal{S}}, \star_{\mathcal{S}})\}) \mid (\boldsymbol{a}, F) \in E^{\mathcal{I}}\}$ for all $E \in \mathsf{N}_{\mathsf{C}} \cup \mathsf{N}_{\mathsf{R}}$.*

The most important observation regarding Definition 11 is that if $\mathcal{I}$ is a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation then $\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}$ is also a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation (i.e., it is a well-founded provenance-interpretation). We now state Lemma 10 which can be proved with an argument similar to the one in Lemma 3.

Lemma 10. *Let $\mathcal{I}$ be an interpretation such that elements of $\mathcal{I}(\Gamma_{\mathcal{O},\star})$ can only occur in annotations sets, and, for all annotation sets F in $\mathcal{I}$ and all $(a,b) \in F$, either $\{a,b\} \subseteq \mathcal{I}(\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+})$ or $\{a,b\} \cap \mathcal{I}(\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+}) = \emptyset$. Then, for every plain BCQ$_@$ q without any $\star_{\mathcal{S}} \in \Gamma_{\mathcal{O},\star}$ and with concept/role/individual names occurring in $\mathcal{O}$: $\mathcal{I} \models q$ iff $\mathcal{I}^{\mathcal{O}}_{\Gamma_\star} \models q$.*

Theorem 12. *Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology and let q be a plain BCQ$_{@,\mathbb{K}}$. Assume $\mathcal{O}$ is satisfiable and the provenance sums that occur in q occur in $\mathcal{O}$. Then, the following holds:*

1. *$\mathcal{O} \models q$ iff $\mathcal{M}^{\mathcal{O}}_\star \models q$;*
2. *$\mathcal{M}^{\mathcal{O}}_\star \models q$ iff $\mathcal{M}^{\mathcal{O}}_\star \models \mathsf{gr}_+(q)$.*

Proof. We adapt our proof of Theorem 10 to our notion of canonical model given in Definition 10. Recall that no $\star_{\mathcal{S}} \in \Gamma_{\mathcal{O},\star}$ occurs in $\mathcal{O}$ or in q, since we assume that all provenance sums that occur in q also occur in $\mathcal{O}$. For Point 1, if $\mathcal{O}$ is satisfiable then, by construction, $\mathcal{M}^{\mathcal{O}}_\star$ is a model of $\mathcal{O}$. Thus, $\mathcal{O} \models q$ implies $\mathcal{M}^{\mathcal{O}}_\star \models q$. Conversely, assume $\mathcal{M}^{\mathcal{O}}_\star \models q$. By construction, if an interpretation $\mathcal{I}$ models $\mathcal{O}$ then $\mathcal{M}^{\mathcal{O}}_\star \to \mathcal{I}^{\mathcal{O}}_{\Gamma_\star}$. Then, by Lemma 4, $\mathcal{I}^{\mathcal{O}}_{\Gamma_\star} \models q$. Moreover, as $\mathcal{M}^{\mathcal{O}}_\star$ only has annotation sets H such that for all $(a,b) \in H$ either $(a,b) = (\star_{\mathcal{S}}^{\mathcal{M}^{\mathcal{O}}_\star}, \star_{\mathcal{S}}^{\mathcal{M}^{\mathcal{O}}_\star})$ or $\{a,b\} \subseteq \mathcal{M}^{\mathcal{O}}_\star(\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+})$ for some $\mathcal{S}$ defined as a set of attribute-value pairs formed with elements from $\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+}$, we can assume that the image $\mathcal{J}$ of the homomorphism from q to $\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}$ only contains annotation sets of this form.

Consider the interpretation $\mathcal{K}$ that is the result of removing from $\mathcal{I}$ all annotations sets H such that there is $(a,b) \in H$ with $a = \star^{\mathcal{I}}_{\mathcal{S}}$ or $b = \star^{\mathcal{I}}_{\mathcal{S}}$ but $(a,b) \neq (\star^{\mathcal{I}}_{\mathcal{S}}, \star^{\mathcal{I}}_{\mathcal{S}})$, where $\star_{\mathcal{S}} \in \Gamma_{\mathcal{O},\star}$. We have that $\mathcal{J} \to \mathcal{K}^{\mathcal{O}_{\Gamma_\star}}_{\Gamma_\star}$ (Definitions 7 and 11) and, by construction, $\mathcal{K}^{\mathcal{O}_{\Gamma_\star}}_{\Gamma_\star}$ is a well-founded provenance-interpretation. As $\mathcal{J} \models q$, by Lemma 10 and since q is plain, $\mathcal{K} \models q$. By definition, $\mathcal{K} \to \mathcal{I}$, and thus by Lemma 4, $\mathcal{I} \models q$.

Finally, for Point 2, assume $\mathcal{M}^{\mathcal{O}}_\star \models q$. Then there is a non-empty set of variable assignments $\mathcal{Z}$ as in Definition 3. Since all annotation sets of $\mathcal{M}^{\mathcal{O}}_\star$ contain only elements in $\mathcal{M}^{\mathcal{O}}_\star(\mathsf{N}_{\mathsf{S}}^{\mathcal{O}+})$, each $\mathcal{Z}$-instance $q_{\mathcal{Z}}$ of q is in $\mathsf{gr}_+(q)$. Thus, $\mathcal{M}^{\mathcal{O}}_\star \models \mathsf{gr}_+(q)$. Conversely, by definition of $\mathsf{gr}_+(q)$, if there is some $q_{\mathcal{Z}} \in \mathsf{gr}_+(q)$ such that $\mathcal{M}^{\mathcal{O}}_\star \models q_{\mathcal{Z}}$ then the variable assignment $\mathcal{Z}$ satisfies the conditions of Definition 3 for the query q. So $\mathcal{M}^{\mathcal{O}}_\star \models q$. ☐

Lemma 11. *Let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology and q a plain BCQ$_{@,\mathbb{K}}$. We have that $\mathcal{O} \models q$ iff $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models \mathsf{gr}_+(q)$.*

Proof. The argument here is similar to our proof of Lemma 5. Assume $\mathcal{O} \models q$. Note that elements of $\Gamma_{\mathcal{O},\star}$ may occur in some queries in $\mathsf{gr}_+(q)$, so that Point 1 of Theorem 12 does not hold for queries in $\mathsf{gr}_+(q)$. So, in the following, we use $\mathcal{O}' = \mathsf{gr}_+(\mathcal{O})_{\Gamma_\star}$ instead of $\mathsf{gr}_+(\mathcal{O})$. By definition of $\mathcal{O}'$ and construction of $\mathcal{M}^{\mathcal{O}}_\star$, we have that $\mathcal{M}^{\mathcal{O}}_\star \to \mathcal{M}^{\mathcal{O}'}_\star$. Suppose an interpretation $\mathcal{I}$ models $\mathcal{O}'$. By Point 1 of Theorem 12, $\mathcal{M}^{\mathcal{O}}_\star \models q$ and by Point 2, $\mathcal{M}^{\mathcal{O}}_\star \models \mathsf{gr}_+(q)$. Moreover, by our construction, $\mathcal{M}^{\mathcal{O}}_\star \to \mathcal{M}^{\mathcal{O}'}_\star \to \mathcal{I}$: here we do not need to use $\mathcal{I}^{\mathcal{O}}_{\Gamma_\star}$ from Definition 11 because $\mathcal{I}$ is a model of $\mathcal{O}'$. Thus, by Lemma 4, $\mathcal{I} \models \mathsf{gr}_+(q)$. Since $\mathcal{I}$ was an arbitrary interpretation satisfying $\mathcal{O}'$, we have that $\mathcal{O}' \models \mathsf{gr}_+(q)$.

Now, assume $\mathcal{O}' \models \mathsf{gr}_+(q)$. By Lemma 8, $\mathcal{O}$ is satisfiable iff $\mathsf{gr}_+(\mathcal{O})$ is satisfiable and our extension $\mathcal{O}'$ does not change this relation. So if $\mathcal{O}'$ is unsatisfiable $\mathcal{O}$ trivially entails q. Then, assume $\mathcal{O}'$ is satisfiable. Since $\mathcal{O}' \models \mathsf{gr}_+(q)$, by definition of $\mathsf{gr}_+(q)$, then $\mathcal{O}' \models q$. Since individual names of the form $\star_{\mathcal{S}}$ do not occur in q, in fact $\mathsf{gr}_+(\mathcal{O}) \models q$. By construction of $\mathsf{gr}_+(\mathcal{O})$, $\mathcal{O} \models \mathsf{gr}_+(\mathcal{O})$. Thus, $\mathcal{O} \models q$. ☐

Lemmas 9 and 11 together show the correctness of our reduction for the query entailment problem. We use these results in Theorem 9. The rest of this appendix is devoted to show Lemma 1 and Theorem 6.

In the Section 'Reasoning in DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$' we provided a transformation of a BCQ$_{@,\mathbb{K}}$ q into a set $\mathsf{gr\_plain}(\mathcal{O}, q)$ of ground and plain queries such that a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology $\mathcal{O}$ without expressions of the form $\exists P$ on the right side entails q iff one of the queries of this set is entailed. For convenience of the reader, we recall this construction and provide more details.

Let $\mathcal{I}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation with domain $\Delta^{\mathcal{I}} = \mathsf{N_{P_{min}}}$ and such that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N_{P_{min}}}$. Let $F_S := \{(a^{\mathcal{I}}, b^{\mathcal{I}}) \mid a : b \text{ occurs in } S\} \in \Phi^{\mathcal{I}}$. We say that a set $\{S^1, \dots, S^n\}$ of ground closed specifiers from $\mathbf{S}_\mathsf{P}$ is a *decomposition* of S into n if there exists some p such that the F_{S^i} are differentiated by p in some annotation set, each F_{S^i} contains some attribute-value pair of the form (p, a) with $a \in \mathsf{N_P}$, and $\sum^p_{1 \le i \le n} F_{S^i} \in S^{\mathcal{I}}$. We say that a set $\{S^i_1, \dots, S^i_m\}$ of ground closed specifiers from $\mathbf{S}_\mathsf{S}$ *partitions* a ground closed specifier S^i if $\prod_{1 \le j \le m} F_{S^i_j} = F_{S^i}$.

Example 12. *It follows from our definitions that the set of ground closed specifiers* $\{[\mathsf{src}\colon \mathsf{s_1} \times \mathsf{s_3}], [\mathsf{src} : \mathsf{s_2} \times \mathsf{s_3}]\}$ *is a decomposition of* $[\mathsf{src} : (\mathsf{s_1} \times \mathsf{s_3}) + (\mathsf{s_2} \times \mathsf{s_3})]$.

Also, $\{[\mathsf{src}\colon \mathsf{s_1}], [\mathsf{src} : \mathsf{s_3}]\}$ *partitions* $[\mathsf{src}\colon \mathsf{s_1} \times \mathsf{s_3}]$ *and* $\{[\mathsf{src}\colon \mathsf{s_2}], [\mathsf{src} : \mathsf{s_3}]\}$ *partitions* $[\mathsf{src}\colon \mathsf{s_2} \times \mathsf{s_3}]$.

Given a BCQ$_{@,\mathbb{K}}$ q of the form

$$\exists \boldsymbol{x}.\, X_1 : S_1, \dots, X_m : S_m\ (\bigwedge_{1 \le j \le m} E_j(\boldsymbol{t_j})@X_j)@S \quad (3)$$

let q' be the result of removing the outer specifier from q. Let $\mathcal{I}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation with domain $\Delta^{\mathcal{I}} = \mathsf{N_{P_{min}}}$ and such that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N_{P_{min}}}$. Given a compatible $\mathcal{Z}$, a $\mathcal{Z}$-image $\bigwedge_{1 \le j \le m} E_j(\boldsymbol{t_j})@T_j$ of q' is obtained by:

- replacing each X_j with $T_j = [a\colon b \mid (a, b) \in \mathcal{Z}(X_j)]$;
- replacing each object variable x by $\mathcal{Z}(x)$;
- if $\star$ occurs in some T_j, replacing $\star$ by $\star_{\mathcal{T}_j}$ where $\mathcal{T}_j$ is the set of attribute-value pairs in T_j that do not contain $\star$.

We define $\mathsf{gr\_plain}(\mathcal{O}, q)$ as the set of ground plain BCQ$_{@,\mathbb{K}}$s:

$$q_{\mathsf{gp}} = (\bigwedge_{1 \le i \le n} (\bigwedge_{1 \le j \le m} E_j(\boldsymbol{t^i_j})@S^i_j\ ))@\lfloor\ \rfloor \quad (4)$$

where $\{S^1, \dots, S^n\}$ is a decomposition of S; for every $1 \le i \le n$, $\{S^i_1, \dots, S^i_m\}$ partitions S^i and:

$$\bigwedge_{1 \le j \le m} E_j(\boldsymbol{t^i_j})@S^i_j$$

is a $\mathcal{Z}$-image of q' such that all attribute-value pairs are built with elements from $\mathsf{N_S}$.

We start with the following lemma, on the complexity of recognizing that a query belongs to $\mathsf{gr\_plain}(\mathcal{O}, q)$.

Lemma 12. *Given an ontology $\mathcal{O}$, a BCQ$_{@,\mathbb{K}}$ q as in Equation 3 and a plain and ground BCQ$_{@,\mathbb{K}}$ q_{gp} of the form of the query in Equation 4 without variables and where the S^i_j are ground closed specifiers, we can decide whether $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ in polynomial time in $|q|$ and $|q_{\mathsf{gp}}|$.*

Proof. We argue that one can check in polynomial time (in $|q|$ and $|q_{\mathsf{gp}}|$) whether, for every $1 \le i \le n$, $\{S^i_1, \dots, S^i_m\}$ partitions some S^i such that $\{S^1, \dots, S^n\}$ form a decomposition of S, where S^i_j are ground closed specifiers of q_{gp}. Let $\mathcal{I}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretation with domain $\Delta^{\mathcal{I}} = \mathsf{N_{P_{min}}}$ and such that $a^{\mathcal{I}} = a$ for every $a \in \mathsf{N_{P_{min}}}$. Assumme w.l.o.g. that polynomials occurring in q_{gp} and q are among those in $\mathsf{N_{P_{min}}}$. For every $1 \le i \le n$, we define F_{S^i} as $\prod_{1 \le j \le m} F_{S^i_j}$ and $(*)$ if F_{S^i} are differentiated by p in some annotation set, we define $F^p_{q_{\mathsf{gp}}}$ as $\sum^p_{1 \le i \le n} F_{S^i}$. If $(*)$ is not satisfied then we are done, $q_{\mathsf{gp}} \notin \mathsf{gr\_plain}(\mathcal{O}, q)$. Clearly $(*)$ can be checked in polynomial time. Otherwise we iterate over all such p satisfying $(*)$, there are only polynomially many of them. It remains to check whether $F^p_{q_{\mathsf{gp}}} \in S^{\mathcal{I}}$.

The important point here is checking equality between the polynomials in $F^p_{q_{\mathsf{gp}}}$ and S. Recall that we assume that all polynomials occurring in a specifier in $\mathbf{S}_\mathsf{P}$ are expanded, i.e., they are of the form $\Sigma_{1 \le i \le n_1} \Pi_{1 \le j \le n_2} a_{i,j}$, where all $a_{i,j} \in \mathsf{N_I}$. So all polynomials in S are expanded but, by construction of $F^p_{q_{\mathsf{gp}}}$, it may be the case that some polynomials in $F^p_{q_{\mathsf{gp}}}$ are not expanded (they are of the form $\Sigma_{1 \le i \le n} \Pi_{1 \le j \le m} s_{i,j}$, where all $s_{i,j} \in \mathsf{N_S}$). Given one expanded polynomial p_1 from S and another polynomial p_2 from $F^p_{q_{\mathsf{gp}}}$, we can check equality of p_1 and p_2 in polynomial time in $|p_1|$ and $|p_2|$ by iteratively expanding p_2 and checking whether there is a 1 to 1 correspondence giving equality between the expanded terms (equality between expanded terms amounts to check occurrences of individual names in the terms). We bound the number of times we expand p_2 by the number of terms in p_1 (if p_2 has more terms than p_1 it cannot be equal). Thus, this procedure can be performed in polynomial time as well.

Finally, it is easy to see that checking that each $\bigwedge_{1 \le j \le m} E_j(\boldsymbol{t^i_j})@S^i_j$ is a $\mathcal{Z}$-image of q' for some compatible $\mathcal{Z}$ can be done in polynomial time by considering $\mathcal{Z}$ such that $\mathcal{Z}(X_j) = F_{S^i_j}$ and checking whether $F_{S^i_j} \in S^{\mathcal{I},\mathcal{Z}}_j$ for each j. Indeed, if it is not the case, there is no compatible $\mathcal{Z}$ such that $\bigwedge_{1 \le j \le m} E_j(\boldsymbol{t^i_j})@S^i_j$ is a $\mathcal{Z}$-image of q'. ☐

We now show Lemma 1. To simplify the notation, we used $\mathcal{O}_{\mathsf{gr}}$ in the main text, which here is defined as $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star}$. Also, to avoid the definition of $\mathsf{N}^{\mathcal{O}+}_\mathsf{S}$ and $\Gamma_{\mathcal{O},\star}$, we used $\mathsf{N_S}$ to define $\mathsf{gr\_plain}(\mathcal{O}, q)$, which then was infinite. However, we can *restrict* $\mathsf{gr\_plain}(\mathcal{O}, q)$ *to queries containing only attribute-value pairs are built with elements from* $\mathsf{N}^{\mathcal{O}+}_\mathsf{S} \cup \Gamma_{\mathcal{O},\star}$, since queries containing elements from $\mathsf{N_S} \setminus (\mathsf{N}^{\mathcal{O}+}_\mathsf{S} \cup \Gamma_{\mathcal{O},\star})$ cannot be entailed by $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star}$. From now on we consider that $\mathsf{gr\_plain}(\mathcal{O}, q)$ is restricted to such queries which can be entailed by $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star}$. We show in Lemma 1 that the size of such queries is polynomial in $|q|$ and $|\mathcal{O}|$ if $\mathcal{O}$ is ground.

Lemma 1 *Let q be a BCQ$_{@,\mathbb{K}}$ and let $\mathcal{O}$ be a DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ ontology without expressions of the form $\exists P$ on the right side of inclusions. $\mathcal{O} \models q$ iff there is $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ such that $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models q_{\mathsf{gp}}$. The size of such q_{gp} is polynomial in $|q|$ and $|\mathcal{O}|$, if $\mathcal{O}$ is ground. Moreover, deciding whether $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ is in* PTIME *if $\mathcal{O}$ is ground.*

Proof. Assume $\mathcal{O} \models q$ and let $\mathcal{M} = \mathcal{M}^{\mathcal{O}}_{\star}$ be the canonical model of $\mathcal{O}$ (Definition 10). By construction of $\mathcal{M}$, and since $\mathcal{O}$ does not contain any inclusion with an expression of the form $\exists P$ on the right side, $\Delta^{\mathcal{M}} = \mathsf{N}_{\mathsf{P}_{\mathsf{min}}}$. Moreover, for every $a \in \mathsf{N}_\mathsf{P}$, $a^{\mathcal{M}} = b$ with a equal to some b, by definition of $\mathcal{M}$. Assume $\mathcal{O}$ is satisfiable (otherwise $\mathcal{O} \models q_{\mathsf{gp}}$ trivially). Then, $\mathcal{M} \models \mathcal{O}$ and, since $\mathcal{O} \models q$, $\mathcal{M} \models q$. So there is a non-empty $\chi \subseteq \nu_{\mathcal{M}}(q')$ satisfying Conditions 1-3 of Definition 3, where q' is the result of removing the outer specifier from q and $\nu_{\mathcal{M}}(q')$ is the set of all variable assignments $\mathcal{Z}$ as in Definition 1. Assume that χ is such a set minimal w.r.t. set inclusion. We show that there is a correspondence between χ and some $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ entailed by $\mathcal{O}$.

Assume q is of the form of the query in Equation 3. By definition of χ (Condition 2): for all $1 \leq j \leq m$, there is $(\mathcal{Z}(\boldsymbol{t_j}), F^{\mathcal{Z}}_j) \in E^{\mathcal{M},\mathcal{Z}}_j$ for some $F^{\mathcal{Z}}_j \in X^{\mathcal{M},\mathcal{Z}}_j$. Then, for all $\mathcal{Z} \in \chi$, $\mathcal{M}$ satisfies the $\mathcal{Z}$-images

$$q_{\mathcal{Z}} = \bigwedge_{1 \leq j \leq m} E_j(\mathcal{Z}(\boldsymbol{t_j}))@S_{F^{\mathcal{Z}}_j}$$

of q' where $S_{F^{\mathcal{Z}}_j} = [a\colon b \mid (a, b) \in F^{\mathcal{Z}}_j]$. By construction of $\mathcal{M}$ and the definition of $q_{\mathcal{Z}}$, all attribute-value pairs occurring in $q_{\mathcal{Z}}$ are built from elements of $\mathsf{N}^{\mathcal{O}+}_\mathsf{S} \cup \Gamma_{\mathcal{O},\star}$.

Since χ satisfies Condition 3, there is $p \in \Delta^{\mathcal{M}}$ and $G \in \Phi^{\mathcal{M}}$ such that all $H^{\mathcal{Z}} = \prod_{1 \leq j \leq m} F^{\mathcal{Z}}_j$ with $\mathcal{Z} \in \chi$ are differentiated by p in G, and $\sum^{p}_{\mathcal{Z} \in \chi} H^{\mathcal{Z}} \in S^{\mathcal{M}}$. Moreover, $|\chi|$ is bounded by the maximum number of terms in a sum occurring in S. To see this, assume χ is not singleton, otherwise we are done. Now note that: for every $\mathcal{Z} \in \chi$, $H^{\mathcal{Z}}$ contains some $(p, a^{\mathcal{M}})$ with $a \in \mathsf{N}_\mathsf{P}$: if there is some $\mathcal{Z}_0 \in \chi$ such that $H^{\mathcal{Z}_0}$ does not contain any $(p, a^{\mathcal{M}})$ with $a \in \mathsf{N}_\mathsf{P}$, then $\sum^{p}_{\mathcal{Z} \in \chi} H^{\mathcal{Z}} = H^{\mathcal{Z}_0}$, so $\{\mathcal{Z}_0\} \subset \chi$ would fulfill Conditions 1-3, contradicting minimality of χ. All tuples of the form $(p, a^{\mathcal{M}}) \in \sum^{p}_{\mathcal{Z} \in \chi} H^{\mathcal{Z}}$ are such that a is a polynomial equal to a sum of $|\chi|$ elements from N_P. So if S contains $p\colon a$ (recall that, by definition of $\mathcal{M}$, $p^{\mathcal{M}} = p$ for all $p \in \mathsf{N}_\mathsf{I}$) we are done, and, if S is closed this must be the case, since $\sum^{p}_{\mathcal{Z} \in \chi} H^{\mathcal{Z}} \in S^{\mathcal{M}}$. Otherwise, if S is open and does not contain $p\colon a$ then $\sum^{p}_{\mathcal{Z} \in \chi} H^{\mathcal{Z}} \setminus \{(p, a^{\mathcal{M}}) \mid a \in \mathsf{N}_\mathsf{P}\} = G$ is also in $S^{\mathcal{M}}$, as well as all supersets of G. In particular, $H^{\mathcal{Z}} \in S^{\mathcal{M}}$ for every $\mathcal{Z} \in \chi$. By assumption χ is not singleton. Since any $\{\mathcal{Z}\} \subseteq \chi$ fulfills Conditions 1-3, this contradicts minimality of χ. So S contains some polynomial a equal to the sum of $|\chi|$ elements from N_P. Thus, $|\chi|$ is polynomial in $|S|$ and:

- the specifiers $S_{H^{\mathcal{Z}}} = [p\colon a \mid (p, a^{\mathcal{M}}) \in H^{\mathcal{Z}}]$ form a decomposition of S into $|\chi|$ and
- the specifiers $S_{F^{\mathcal{Z}}_j} = [p\colon a \mid (p, a^{\mathcal{M}}) \in F^{\mathcal{Z}}_j]$ partition $S_{H^{\mathcal{Z}}}$.

This means that $q^{\chi} = (\bigwedge_{\mathcal{Z} \in \chi} q_{\mathcal{Z}})@\lfloor\ \rfloor \in \mathsf{gr\_plain}(\mathcal{O}, q)$ and is such that $\mathcal{M} \models q^{\chi}$. Let q^{χ}_o be the query obtained from q^{χ} by replacing each specifier $S_{F^{\mathcal{Z}}_j}$ which contains $\star_{S_{F^{\mathcal{Z}}_j}}$ by the open specifier that contains the attribute-value pairs without $\star_{S_{F^{\mathcal{Z}}_j}}$ in $S_{F^{\mathcal{Z}}_j}$. By Point 1 of Theorem 12, $\mathcal{O} \models q^{\chi}_o$ and, by Lemma 11, $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models \mathsf{gr}_+(q^{\chi}_o)$. By definition of q^{χ}_o, for all $q \in \mathsf{gr}_+(q^{\chi}_o)$ and all interpretations $\mathcal{I}$, if $\mathcal{I} \models q$ then $\mathcal{I} \models q^{\chi}_o$, which means that $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models q^{\chi}_o$. The fact that $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models q^{\chi}$ follows from the facts that: specifiers $S_{F^{\mathcal{Z}}_j}$ in q^{χ} correspond to annotation sets $F^{\mathcal{Z}}_j$ in the canonical model $\mathcal{M}$; and, annotation sets $F^{\mathcal{Z}}_j$ in $\mathcal{M}$ are used to construct specifiers $S_{F^{\mathcal{Z}}_j}$ occurring in $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star}$, in particular, Points 1 and 2 of the extension $\cdot_{\Gamma_\star}$ ensure that, for all interpretations $\mathcal{I}$ satisfying $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star}$, if $\mathcal{I}$ satisfies q^{χ}_o then it satisfies q^{χ}.

For the other direction, we show the following claim.

Claim 1. For all DL-Lite$^{\mathcal{R}}_{@,\mathbb{K}}$ interpretations $\mathcal{I}$ and all $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$, if $\mathcal{I} \models q_{\mathsf{gp}}$ then $\mathcal{I} \models q$.

Proof of Claim 1. Assume q_{gp} is of the form:

$$(\bigwedge_{1 \leq i \leq n} (\bigwedge_{1 \leq j \leq m} E_j(\boldsymbol{t^i_j})@S^i_j\ ))@\lfloor\ \rfloor$$

and $\mathcal{I} \models q_{\mathsf{gp}}$. Assume w.l.o.g. that there are no two identical $\bigwedge_{1 \leq j \leq m} E_j(\boldsymbol{t^i_j})@S^i_j$ in q_{gp}. Let $\mathcal{J}$ be the result of simultaneously replacing each $a^{\mathcal{I}}$ with $a \in \mathsf{N}_\mathsf{P}$ by $b \in \mathsf{N}_{\mathsf{P}_{\mathsf{min}}}$ such that b is equal to a. There is an isomorphism between $\mathcal{I}$ and $\mathcal{J}$ so $\mathcal{J} \models q_{\mathsf{gp}}$. We first show that $\mathcal{J} \models q$.

By construction of q_{gp}, there is a set χ of variable assignments $\mathcal{Z}$ compatible with q' corresponding to the $\mathcal{Z}$-images

$$\bigwedge_{1 \leq j \leq m} E_j(\boldsymbol{t^i_j})@S^i_j$$

used to build q_{gp}. Moreover, for every $1 \leq i \leq n$, $\{S^i_1, \ldots, S^i_m\}$ partitions S^i; and $\{S^1, \ldots, S^n\}$ is a decomposition of S. Since $\mathcal{J} \models q_{\mathsf{gp}}$ and q_{gp} is built from $\mathcal{Z}$-images of q', we have that $\chi \subseteq \nu_{\mathcal{J}}(q')$. We show that χ fulfills Conditions (1) to (3) of Definition 3 for $\mathcal{J} \models q$. By definition of χ, since we assume that all $\mathcal{Z}$-images in q_{gq} are different, any two $\mathcal{Z}$, $\mathcal{Z}'$ in χ are distinct in at least one object or set variable in q', so χ satisfies Condition (1) of Definition 3. The other 2 conditions follow from the facts that $\mathcal{J} \models q_{\mathsf{gp}}$, q_{gp} is built from $\mathcal{Z}$-images of q', $\{S^i_1, \ldots, S^i_m\}$ are closed specifers and partitions S^i, and $\{S^1, \ldots, S^n\}$ is a decomposition of S. Since χ satisfies Conditions (1) to (3) of Definition 3, $\mathcal{J} \models q$. Now since $\mathcal{I}$ is isomorphic to $\mathcal{J}$, $\mathcal{I} \models q$.

Assume $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models q_{\mathsf{gp}}$. By Claim 1, $\mathsf{gr}_+(\mathcal{O})_{\Gamma_\star} \models q$. Since individual names of the form $\star_S$ do not occur in q, in fact $\mathsf{gr}_+(\mathcal{O}) \models q$. By construction of $\mathsf{gr}_+(\mathcal{O})$, $\mathcal{O} \models \mathsf{gr}_+(\mathcal{O})$. Thus, $\mathcal{O} \models q$.

It remains to show that if $\mathcal{O}$ is ground the size of each specifier is polynomial in $|\mathsf{N}^{\mathcal{O}}_\mathsf{S}|$. We show that the size of an annotation set in the canonical model of a ground ontology $\mathcal{O}$ is polynomial in $\mathcal{O}$. Therefore we only need to consider specifiers with attribute-value pairs which can be interpreted as such annotations. We consider the canonical model $\mathcal{M}$ of $\mathcal{O}$. Since $\mathcal{O}$ is ground the annotations sets in $\mathcal{M}$ can only be:

1. $F^{\mathbb{K}}_S$ if S is a ground specifier in $\mathcal{O}$;
2. a sum of primitive annotation sets in $\mathcal{M}$.

The sizes of the annotation sets in Point 1 are clearly bounded by the size of a specifier in $\mathcal{O}$, which is in turn bounded by

$2 \cdot |\mathsf{N}_\mathsf{S}^\mathcal{O}|$. The size of a sum of primitive annotation sets is bounded by the number of primitive annotation sets and their size. The number of primitive annotation sets is bounded by the number of specifiers in $\mathcal{O}$. Since the size of each primitive annotation set is polynomial in $|\mathsf{N}_\mathsf{S}^\mathcal{O}|$, the size of a sum of primitive annotation sets is also polynomial in $|\mathsf{N}_\mathsf{S}^\mathcal{O}|$.

The last part of this lemma follows from Lemma 12: if $\mathcal{O}$ is ground then we can decide whether $q_{\mathsf{gp}} \in \mathsf{gr\_plain}(\mathcal{O}, q)$ in PTIME. $\square$