Extending $\mathcal{ALCQ}$ with Bounded Self-Reference

Daniel Gorín

Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg

Lutz Schröder [1]

Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg

Abstract

Self-reference has been recognized as a useful feature in description logics but is known to cause substantial problems with decidability. We have shown in previous work that the basic description logic $\mathcal{ALC}$ remains decidable, and in fact retains its low complexity, when extended with a bounded form of self-reference where only one variable (denoted me following previous work by Marx) is allowed, and no more than two relational steps are allowed to intercede between binding and use of me (this result is optimal in the sense that already allowing three steps leads to undecidability). Here, we extend these results to $\mathcal{ALCQ}$, i.e. $\mathcal{ALC}$ extended with qualified number restrictions, and analyse the expressivity of the arising logic, $\mathcal{ALCQ}\mathsf{me}_2$. In fact it turns out the expressive power of $\mathcal{ALCQ}\mathsf{me}_2$ is identical to that of $\mathcal{ALCHIQ}be$, the extension of $\mathcal{ALCQ}$ with role inverses, role hierarchies, safe Boolean combinations of roles, and a simple self-loop construct. However, while there is a straightforwardly defined polynomial translation from $\mathcal{ALCHIQ}be$ to $\mathcal{ALCQ}\mathsf{me}_2$, the translation from $\mathcal{ALCQ}\mathsf{me}_2$ to $\mathcal{ALCHIQ}be$ has an exponential blowup in the formula size. To establish the desired complexity bounds, we therefore provide a polynomial satisfiability-preserving encoding of $\mathcal{ALCQ}\mathsf{me}_2$ into $\mathcal{ALCHIQ}be$ and prove that the latter is decidable in EXPTIME.

Keywords: Description logic, hybrid logic, binding constructs, self-reference, complexity, expressivity, qualified number restrictions

Introduction

In a very broad sense, hybrid logics extend modal logics with special symbols, called *nominals*, that name individual states of models. The assignment of nominals to states can be either *static*, akin to that of *constant symbols* in first-order logic, or *dynamic*. The latter is typically achieved through the $\downarrow$-binder: if x is a nominal, then $\downarrow x.\phi$ is true at a state a if ϕ is true in a under the

[1] The research reported here forms part of the DFG project *Generic Algorithmic Methods in Modal and Hybrid Logics* (SCHR 1115-2)

assumption that x names a. The difference can be dramatic: while extending a modal logic with static hybrid machinery often results in only a moderate increase of computational complexity (if at all), adding $\downarrow$ to the basic modal logic immediately leads to undecidability. This has proven to be hard to repair: it persists when only one nominal is allowed to occur [12], also if one restricts to typically well-behaved classes of models [14] and even if one weakens the semantics of $\downarrow$ [2].

Decidability is regained in the uni-modal case on certain classes of models [14] and when restricting to the rather weak fragment that avoids the so-called $\Box\downarrow\Box$-pattern where the $\downarrow$ occurs under the scope of a $\Box$ and the bound nominal occurs under another $\Box$ [16]. In recent work [9], we have isolated another decidable fragment; it is obtained by allowing only one nominal to be bound by $\downarrow$ and ensuring that between every $\downarrow$ and the usage points of the bound nominal, no more than two modalities occur (when three modalities are allowed to occur, decidability is again lost); we refer to this restriction as limiting the *depth* of occurrences of the bound nominal. In [12], Marx proposes the usage of pronouns I and me as suggestive notation for the hybrid language were only one nominal can be bound (with I implicitly binding the nominal me); using this notation, an example of a depth 2 formula is given by

$$[\mathsf{hasParent}]\mathsf{I}.\neg\langle\mathsf{likes}\rangle(\mathsf{Club} \wedge \langle\mathsf{accepts}\rangle\mathsf{me}), \tag{1}$$

which, intuitively, describes those whose parents share Groucho's taste for clubs (observe, incidentally, that (1) falls inside the $\Box\downarrow\Box$-pattern, so it is outside the decidable fragment of [16]).

Despite its innocent look, the depth-2 fragment of the I-me hybrid logic turns out to be expressively interesting. In particular, the nominal-free language (that is, free of nominals other than me) with the universal modality is capable of expressing all formulas of the guarded fragment over the correspondence language; and this containment is strict: the latter has the finite model property while the former does not [9]. Nevertheless, the nominal-free fragment has the same complexity as basic $\mathcal{ALC}$: PSPACE-complete over acyclic TBoxes (even when the satisfaction operator $@_{\mathsf{me}}$ is added), and EXPTIME-complete over general TBoxes, equivalently in presence of the universal modality.

Given these results, it is interesting to understand the effect of *hybridizing* other modal logics in this restricted fashion. Here, we consider the case of *graded modal logic*, or in description logic parlance, the logic $\mathcal{ALCQ}$, whose extension with the I-me construct limited to depth 2 we denote $\mathcal{ALCQ}\mathsf{me}_2$. To avoid unnecessary duplication of concepts and notations, we will work using the language and terminology of description logics (DLs) throughout [5].

To get a taste of $\mathcal{ALCQ}\mathsf{me}_2$, consider an asymmetric social-networking platform such as Twitter as the domain of knowledge representation. In this context we can express the concept of a *celebrity* as

$$\mathsf{I}.\geq_N \mathsf{isFollowedBy}.\neg\exists\mathsf{isFollowedBy}.\mathsf{me}$$

for some suitable choice of $N \gg 0$. In words, a celebrity in Twitter is someone with a large base of followers that are not followed back.

It turns out that $\mathcal{ALCQ}\mathsf{me}_2$ is expressive enough to encode inverses (aka *past* modalities), union, intersection, and relative complementation of roles (i.e. relations), as well as role inclusions. This puts it, prima facie, close to the DL known as $\mathcal{ALCHIQ}b$. It is the main object of this work to make this comparison systematic.

In fact it is clear that $\mathcal{ALCQ}\mathsf{me}_2$ contains, via a straightforward semantics-preserving translation that has only a polynomial blowup in the formula size, a slight extension of $\mathcal{ALCHIQ}b$, namely by the self-loop construct $\exists R.\,\mathsf{Self}$ also found, e.g., in the DL $s\mathcal{ROIQ}$ [10]; we denote this extension by $\mathcal{ALCHIQ}be$. Conversely, we show that indeed there exists also a semantics-preserving translation from $\mathcal{ALCQ}\mathsf{me}_2$ to $\mathcal{ALCHIQ}be$, so that we can say in a precise sense that $\mathcal{ALCQ}\mathsf{me}_2$ and $\mathcal{ALCHIQ}be$ have the same expressive power; however, this converse translation has an exponential blowup in the formula size. In order to fix the computational complexity of $\mathcal{ALCQ}\mathsf{me}_2$, we therefore define a second translation into $\mathcal{ALCHIQ}be$. This translation has only a polynomial blowup; it preserves satisfiability but is not semantics-preserving in the proper sense, as it introduces new underdefined relation symbols and satisfaction of the original formula does not imply satisfaction of its translation (only *satisfiability* of the latter). We then show by reduction to $\mathcal{ALCIQ}b$ [18] that $\mathcal{ALCHIQ}be$ is decidable in PSPACE over the empty TBox, and in EXPTIME over general TBoxes, thus implying that the same bounds hold for $\mathcal{ALCQ}\mathsf{me}_2$, which is hence no harder than $\mathcal{ALCQ}$ (or, for that matter, basic $\mathcal{ALC}$). Our proof can be easily modified to show that, in general, adding the self-looping construct to a description logic is harmless in terms of computational complexity.

The material is organized as follows. We first introduce the description logics we will use as well as some basic terminology (for a deeper introduction, refer to [5]). In particular, we define the logics $\mathcal{ALCHIQ}be$ and $\mathcal{ALCQ}\mathsf{me}_2$. In Section 2 we discuss in a general fashion various ways in which one can compare the expressiveness of two description logics, in order to pave the ground for a comparison of the various translations that we define. We discuss the embedding of $\mathcal{ALCHIQ}be$ into $\mathcal{ALCQ}\mathsf{me}_2$ in Section 3, and the converse semantics-preserving translation in Section 4. Moreover, we establish a polynomial satisfiability-preserving translation of $\mathcal{ALCQ}\mathsf{me}_2$ into $\mathcal{ALCHIQ}be$ (Proposition 4.5). Finally, we present our complexity results in Section 5.

1 Description Logics and Self-Referential Constructs

Description logics (DLs) are typically described as arising from a combination of a number of more or less standard language features. The features we will be interested in are qualified number restrictions, role-hierarchies, inverses, safe Boolean combination of roles and the “self-looping” concept constructor. We now introduce a DL called $\mathcal{ALCHIQ}be$ that contains all of these features, and later consider some of its better known fragments.

Assume a vocabulary $\langle \mathsf{N_C}, \mathsf{N_R} \rangle$ composed of two disjoint and countably infinite sets $\mathsf{N_C} = \{A_1, A_2, \ldots\}$ and $\mathsf{N_R} = \{R_1, R_2, \ldots\}$ of *atomic* concepts and roles, respectively. The set of (complex) concepts C, D and roles R, S of $\mathcal{ALCHIQ}be$ are respectively given by

$$C, D := A_i \mid \neg C \mid C \sqcap D \mid \exists R.\mathsf{Self} \mid \geq_n R.C$$
$$R, S := R_i \mid R^- \mid R \sqcap S \mid R \sqcup S \mid R - S$$

where $A_i \in \mathsf{N_C}$; $R_i \in \mathsf{N_R}$ and n is a positive integer. The concept $\exists R.\mathsf{Self}$ is meant to denote the individuals related via R to themselves (but notice that Self in itself is not a concept).

For brevity, we will employ standard notation, such as $\bot \equiv A \sqcap \neg A$, for some $A \in \mathsf{N_C}$; $\top \equiv \neg\bot$; $\leq_n R.C \equiv \neg \geq_{n+1}.C$, for every $n \geq 0$; $\exists R.C \equiv \geq_1 R.C$ and $\forall R.C \equiv \neg\exists R.\neg C$. One can also define $\forall R.\mathsf{Self} \equiv \leq_1 R.\top \sqcap (\forall R.\bot \sqcup \exists R.\mathsf{Self})$. We will sometimes write $\geq_{n-1} R.C$ in cases where $n \geq 1$, and then assume that $\geq_0 R.C$ is a synonym for $\top$.

When measuring formula size, we assume numbers expressed in *binary*. We shall use *rank*(C) to mean the maximal number of nested qualified number restrictions (i.e., concepts of the form $\geq_n R.D$) occurring in C (this is also known as the *modal depth* of C).

An *interpretation* or *model* is a structure $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ where $\Delta^\mathcal{I}$ is a non-empty set (the *domain* of $\mathcal{I}$); $A^\mathcal{I} \subseteq \Delta^\mathcal{I}$ for each $A \in \mathsf{N_C}$; and, for $R_i \in \mathsf{N_R}$, $R_i^\mathcal{I} \subseteq \Delta^\mathcal{I} \times \Delta^\mathcal{I}$. We write $R^\mathcal{I}(a)$ for the set $\{b : (a, b) \in R^\mathcal{I}\}$. Let $\mathcal{I} = \langle \Delta^\mathcal{I}, \cdot^\mathcal{I} \rangle$ be an interpretation. The extension of complex concepts and roles under $\mathcal{I}$ is defined as:

$$\begin{aligned} (\neg C)^\mathcal{I} &= \Delta^\mathcal{I} - C^\mathcal{I} & (R - S)^\mathcal{I} &= R^\mathcal{I} - S^\mathcal{I} \\ (C \sqcap D)^\mathcal{I} &= C^\mathcal{I} \cap D^\mathcal{I} & (R \sqcap S)^\mathcal{I} &= R^\mathcal{I} \cap S^\mathcal{I} \\ \exists R.\mathsf{Self} &= \{a : (a, a) \in R^\mathcal{I}\} & (R \sqcup S)^\mathcal{I} &= R^\mathcal{I} \cup S^\mathcal{I} \\ (\geq_n R.C)^\mathcal{I} &= \{a : |R^\mathcal{I}(a) \cap C^\mathcal{I}| \geq n\} & (R^-)^\mathcal{I} &= \{(b, a) : (a, b) \in R^\mathcal{I}\} \end{aligned}$$

In description logics, a *TBox* is a set of "concept inclusion axioms" of the form $C \sqsubseteq D$. Sometimes the term "general TBox" is used to stress the fact that concepts occurring in it are arbitrary (in opposition to, say, *acyclic TBoxes*, where the dependency graph induced by the atomic concepts in the TBox must be acyclic). Apart from a TBox, in $\mathcal{ALCHIQ}be$, one has an *RBox* that contains "role inclusion axioms" of the form $R \sqsubseteq S$. Let $\mathcal{T}$ and $\mathcal{H}$ be a TBox and a RBox, respectively. An interpretation $\mathcal{I}$ *satisfies* $(\mathcal{T}, \mathcal{H})$ whenever $C^\mathcal{I} \subseteq D^\mathcal{I}$ for all $C \sqsubseteq D \in \mathcal{T}$ and $R^\mathcal{I} \subseteq S^\mathcal{I}$ for all $R \sqsubseteq S \in \mathcal{H}$. A concept C is *satisfiable over* $(\mathcal{T}, \mathcal{H})$ if $C^\mathcal{I} \neq \emptyset$ for some interpretation $\mathcal{I}$ that satisfies $(\mathcal{T}, \mathcal{H})$. Finally, C is *satisfiable* if it is satisfiable over the empty TBox and the empty RBox.

Many well-known description logics are obtained as fragments of the logic $\mathcal{ALCHIQ}be$ just introduced. E.g., $\mathcal{ALCHIQ}b$ results from dropping the $\exists R.\mathsf{Self}$ concept-constructor (which is part of the very expressive description logic $s\mathcal{ROIQ}$ [10]) and $\mathcal{ALCHIQ}$ is obtained if, additionally, one removes role-constructors $\sqcap$, $\sqcup$ and $-$. If, in any of these cases, attention is restricted to empty RBoxes, the logics obtained are respectively called $\mathcal{ALCIQ}be$, $\mathcal{ALCIQ}b$

and $\mathcal{ALCIQ}$; note that $\mathcal{ALCIQ}b$ played a key role in the proof of the EXPTIME upper bound for $\mathcal{SHIQ}$ [18]. We obtain the logic $\mathcal{ALCQ}$ by removing the inverse role constructor $\cdot^-$ from $\mathcal{ALCIQ}$. If we furthermore restrict n to 1 in qualified-number restrictions $\geq_n R$ (and hence to $n = 0$ in $\leq_n R$), the basic description logic $\mathcal{ALC}$ is obtained.

It is worth noticing that in [18], safe Boolean roles are defined as containing *role negation* $\neg R$, instead of *relativized negation* $R - S$ as defined above. However, the *safety* condition for role expressions used there is that they must contain at least one non-negated role in each clause of the disjunctive normal form. It is straightforward to verify that this is equivalent to the definition we gave. Moreover, it is interesting to observe that this restriction is not casual, as unsafe roles can be used, for instance, to express global cardinality constraints as in $\leq_5 R \sqcup \neg R.\, C$, which are known to increase complexity; e.g. $\mathcal{ALCIQ}$ with cardinality restrictions is NEXPTIME-complete [17].

Regarding computational complexity, the concept satisfiability problem for $\mathcal{ALCIQ}b$ is known to be PSPACE-complete over the empty TBox, and EXPTIME-complete over general TBoxes [18]. The latter problem remains EXPTIME-complete for $\mathcal{ALCHIQ}b$ [11].

Meet $\mathcal{ALCQ}\mathsf{me}_2$: Bounded self-referentiality for $\mathcal{ALCQ}$

In [12], Marx introduced the I-me construct as a convenient notation for the single-variable fragment of hybrid logic with the $\downarrow$-binder (cf. [3]). In this notation, I plays the role of a $\downarrow$-binder and me is the bound nominal. We want to extend the logic $\mathcal{ALCQ}$ with the I-me construct under the restriction that me is never separated from its binding I by more than two qualified number restrictions. We call the resulting logic $\mathcal{ALCQ}\mathsf{me}_2$.

At a syntactic level, we want to add to the concept language of $\mathcal{ALCQ}$ a new concept me and a concept constructor $\mathsf{I}.C$, where C must satisfy certain requirements with respect to the occurrences of me. Again, we assume two sets $\mathsf{N_C}$ and $\mathsf{N_R}$ of atomic concepts and roles; the concept language of $\mathcal{ALCQ}\mathsf{me}_2$ then corresponds to $\mathcal{F}_1$ in the following grammar:

$$\begin{aligned} \mathcal{F}_1 \ni C, C' &:= A_i \mid \mathsf{me} \mid \neg C \mid C \sqcap C' \mid \mathsf{I}.{\geq_n} R_i.\, C \mid {\geq_n} R_i.\, D \\ \mathcal{F}_2 \ni D, D' &:= A_i \mid \mathsf{me} \mid \neg D \mid D \sqcap D' \mid \mathsf{I}.{\geq_n} R_i.\, C \end{aligned} \tag{2}$$

Here also, $A_i \in \mathsf{N_C}$ and $R_i \in \mathsf{N_R}$, and n is a positive integer. We will say that me occurs *free* in a concept C if it is not under the scope of an I. Similarly, we say that a concept C occurs in a *relational context* in D whenever C occurs under the scope of a $\geq_n R_i$ or $\mathsf{I}.{\geq_n} R_i$ in D.

One can easily get an idea of why (2) works by verifying that, for instance, $\mathsf{I}.{\geq_n} R.{\geq_m} S.{\geq_l} T.\mathsf{me}$ is not well-formed. More generally, one can check that $\mathcal{F}_2 \subsetneq \mathcal{F}_1$ and that each free occurrence of me in C can occur in at most one relational context, if $C \in \mathcal{F}_1$, and in no relational context if $C \in \mathcal{F}_2$. This means that, in fact, a concept such as ${\geq_n} R.{\geq_m} S.\mathsf{me}$ is not well-formed, although $\mathsf{I}.{\geq_n} R.{\geq_m} S.\mathsf{me}$ is so.

While according to (2), I can only occur in front of a qualified number restriction, we shall liberally use it in other positions, by letting it commute through Boolean operations and putting $\mathsf{I}.A \equiv A$ for $A \in \mathsf{N_C}$, $\mathsf{I}.\mathsf{I}.C \equiv \mathsf{I}.C$, and $\mathsf{I}.\mathsf{me} \equiv \top$. That said, for definitions and proofs we will stick to the language defined by (2).

Let $\langle \Delta^{\mathcal{I}}, \cdot^{\mathcal{I}} \rangle$ be an interpretation. For every $a \in \Delta^{\mathcal{I}}$, we use $C^{\mathcal{I}}_a$ to denote the extension of an $\mathcal{ALCQ}\mathsf{me}_2$-concept C under the assumption that me stands for a. This is defined as follows:

$$
\begin{aligned}
A^{\mathcal{I}}_a &= A^{\mathcal{I}} & \mathsf{me}^{\mathcal{I}}_a &= \{a\} \\
(\neg C)^{\mathcal{I}}_a &= \Delta^{\mathcal{I}} - C^{\mathcal{I}}_a & (C \sqcap D)^{\mathcal{I}}_a &= C^{\mathcal{I}}_a \cap D^{\mathcal{I}}_a \\
(\mathsf{I}.{\geq_n} R.C)^{\mathcal{I}}_a &= \{b : b \in ({\geq_n} R.C)^{\mathcal{I}}_b\} & ({\geq_n} R.C)^{\mathcal{I}}_a &= \{b : |R^{\mathcal{I}}(b) \cap C^{\mathcal{I}}_a| \geq n\}
\end{aligned}
$$

We say that C is a *closed* concept if it has no free occurrences of me. One can easily show the following:

Proposition 1.1 *If C is a closed concept, then $C^{\mathcal{I}}_a = C^{\mathcal{I}}_b$ for all $a, b \in \Delta^{\mathcal{I}}$.*

We thus denote the extension of a closed concept C just by $C^{\mathcal{I}}$. From Proposition 1.1 it is clear that $\mathcal{ALCQ}\mathsf{me}_2$ is a conservative extension of $\mathcal{ALCQ}$.

In the context of $\mathcal{ALCQ}\mathsf{me}_2$, a TBox is a collection of axioms of the form $C \sqsubseteq D$, where C and D are closed (but otherwise arbitrary) concepts. An interpretation $\mathcal{I}$ *satisfies* a TBox $\mathcal{T}$ whenever $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all $C \sqsubseteq D \in \mathcal{T}$. A *closed* concept C is *satisfiable over* $\mathcal{T}$ if $C^{\mathcal{I}} \neq \emptyset$ for some interpretation $\mathcal{I}$ that satisfies $\mathcal{T}$. Finally, C is *satisfiable* if it is satisfiable over the empty TBox.

The object of the current work is to investigate the properties of $\mathcal{ALCQ}\mathsf{me}_2$. In the end, one wants to see how $\mathcal{ALCQ}\mathsf{me}_2$ stands with respect to its better-known fellow logics. In the following sections, we investigate then what is the *expressive power* and the *computational complexity* of $\mathcal{ALCQ}\mathsf{me}_2$.

2 Prerequisites for a Discussion of Relative Expressivity

The term "expressivity" or "expressive power" has come to mean different, although related, things. This has the unfortunate effect that it may lead to seemingly contradictory claims regarding the relative expressivity of two logics, both being correct once one pinpoints the precise notion of expressivity being used in each case. To prevent this type of misunderstandings, we will give a short account of the main notions of expressivity that can be usually found in the literature.

When comparing the expressive power of two logics, it certainly helps if they have a perfect match on the objects used as interpretations. In practice, though, this would be too strong a restriction: for instance, it would exclude all so-called *TBox-internalization results*, where a fresh atomic role is needed to behave as a universal modality (so the models of the logic doing the internalization need to interpret more symbols). If we want some liberty in this respect, we need to have the means to compare logics whose languages and/or models do not perfectly coincide. Having said that, we will avoid an overly abstract presentation by focusing on the case of "description logics" (a more abstract, very general presentation would require, for instance, moving to the framework of *institutions* [7]).

Let us make more precise what we mean by "description logics". These can be characterized by a *concept language*, built from atomic concepts and roles; and the possibility of building *theories*, in the form of TBoxes (sets of concept-related axioms), RBoxes (sets of role-related axioms), ABoxes (sets of axioms regarding named individuals), etc. Theories can be assumed to form a monoid, so one can talk about the *empty theory* or take the *union* of two theories. A model for a description logic is a non-empty domain plus an interpretation for each atomic concept, role, named-individual, etc.

In the various notions we introduce next, we will typically say that "a logic $\mathcal{L}_2$ has at least the same power (for some task) as a logic $\mathcal{L}_1$", and by this it should be understood, intuitively, that a certain logical task in $\mathcal{L}_1$ can be *uniformly reduced to the same task* on $\mathcal{L}_2$, perhaps *over an extended language*. This means that models for $\mathcal{L}_2$ will be *expansions* of models for $\mathcal{L}_1$, which leads to a natural way of mapping $\mathcal{L}_2$-models to $\mathcal{L}_1$-models (this can be seen as an specialization of the notion of *logic comorphism* [8]).

Definition 2.1 *Let $\mathcal{S}_1 = \langle \mathsf{N_{C1}}; \mathsf{N_{R1}} \rangle$ and $\mathcal{S}_2 = \langle \mathsf{N_{C2}}; \mathsf{N_{R2}} \rangle$ be two collections of atomic concepts and roles such that $\mathsf{N_{C1}} \subseteq \mathsf{N_{C2}}$ and $\mathsf{N_{R1}} \subseteq \mathsf{N_{R2}}$ hold (we write $\mathcal{S}_1 \subseteq \mathcal{S}_2$). The* forgetful mapping *β from $\mathcal{S}_2$-interpretations to $\mathcal{S}_1$-interpretations is defined by i) $\Delta^{\beta(\mathcal{I})} = \Delta^{\mathcal{I}}$; ii) $A^{\beta(\mathcal{I})} = A^{\mathcal{I}}$, for every $A \in \mathsf{N_{C1}}$ and iii) $R^{\beta(\mathcal{I})} = R^{\mathcal{I}}$, for every $R \in \mathsf{N_{R1}}$. Moreover, given a consistent $\mathcal{L}_2$-theory $\mathcal{T}$ over $\mathcal{S}_2$, we say that $\mathcal{T}$ is* oblivious *to $\mathcal{S}_1$ if for every model $\mathcal{I}_1$ for $\mathcal{S}_1$ there exists a model $\mathcal{I}_2$ satisfying $\mathcal{T}$ such that $\mathcal{I}_1 = \beta(\mathcal{I}_2)$.*

Intuitively, a theory is oblivious to $\mathcal{S}_1$ if it only refers to symbols that are not in $\mathcal{S}_1$. We are now ready to look at the notions of expressivity that will concern us.

Querying power. Let C be a concept of a description logic $\mathcal{L}_1$ over vocabulary $\mathcal{S}_1$. One can view C as a *query* over $\mathcal{S}_1$-interpretations by taking $C^{\mathcal{I}}$ to be the result of the query over $\mathcal{I}$. To a first approximation, we can say that $\mathcal{L}_2$ has at least the same querying power as $\mathcal{L}_1$ if there is an effective mapping α from $\mathcal{L}_1$-concepts to $\mathcal{L}_2$-concepts such that $C^{\mathcal{I}} = \alpha(C)^{\mathcal{I}}$. This assumes that concepts in the domain and image of α are built over $\mathcal{S}_1$. To lift this restriction, we shall say that $\mathcal{L}_2$ *has at least the same querying power as* $\mathcal{L}_1$, notation $\mathcal{L}_1 \leq_Q \mathcal{L}_2$, if there exist a mapping δ taking a vocabulary $\mathcal{S}_1$ to an $\mathcal{L}_2$-theory $\mathcal{T}$ oblivious to $\mathcal{S}_1$; and an effective mapping α such that, for all models $\mathcal{I}_2$ satisfying $\mathcal{T}$, $C^{\beta(\mathcal{I}_2)} = \alpha(C)^{\mathcal{I}_2}$. We write $\mathcal{L}_1 \leq_{Q,\delta} \mathcal{L}_2$ when we want to make explicit the mapping δ employed.

This definition essentially says that given a query C and a model $\mathcal{I}_1$ to be queried, we can find some appropriate $\mathcal{L}_2$- model $\mathcal{I}_2$ satisfying $\mathcal{T}$ (e.g., by an exhaustive search), and query it using $\alpha(C)$, instead.

Classification power. Let $\mathcal{T}$ be a theory for a description logic $\mathcal{L}_1$ over $\mathcal{S}_1$. Let us write $\mathcal{I}_1 \models_{\mathcal{L}_1} \mathcal{T}$ to denote that model $\mathcal{I}_1$ satisfies all the axioms of $\mathcal{T}$. We can then say that $\mathcal{T}$ splits the class of $\mathcal{S}_1$-models in two, implicitly defining a property on models. We then say that $\mathcal{L}_2$ *has at least the same classification power as* $\mathcal{L}_1$, notation $\mathcal{L}_1 \leq_C \mathcal{L}_2$, if there exists a mapping δ taking a vocabulary $\mathcal{S}_1$ to an $\mathcal{L}_2$-theory $\mathcal{T}_2$ oblivious to $\mathcal{S}_1$; and an effective mapping γ of $\mathcal{L}_1$-theories to $\mathcal{L}_2$-theories such that for every $\mathcal{L}_1$-theory $\mathcal{T}_1$ and every $\mathcal{L}_2$-model $\mathcal{I}_2$ of $\mathcal{T}_2$, it holds that $\beta(\mathcal{I}_2) \models_{\mathcal{L}_1} \mathcal{T}_1$ iff $\mathcal{I}_2 \models_{\mathcal{L}_2} \gamma(\mathcal{T}_1)$. In all cases, we assume γ to be defined axiom-wise, i.e., $\gamma(\mathcal{T}) = \bigcup_{a \in \mathcal{T}} \gamma'(a)$, where γ' maps axioms to (finite) theories; moreover, we identify γ with γ'. We write $\mathcal{L}_1 \leq_{C,\delta} \mathcal{L}_2$ when we want to make explicit the mapping δ employed.

It is possible to slightly weaken this notion as follows. We say that $\mathcal{L}_2$ *has at least weakly the same classification power as* $\mathcal{L}_1$, notation $\mathcal{L}_1 \leq_C^{\exists} \mathcal{L}_2$, if there exists a mapping δ taking a vocabulary $\mathcal{S}_1$ to an $\mathcal{L}_2$-theory $\mathcal{T}_2$ oblivious to $\mathcal{S}_1$; and an effective mapping γ of $\mathcal{L}_1$-theories to $\mathcal{L}_2$-theories such that for every $\mathcal{L}_1$-theory $\mathcal{T}_1$ and every $\mathcal{L}_1$-model $\mathcal{I}_1$, $\mathcal{I}_1 \models_{\mathcal{L}_1} \mathcal{T}_1$ iff *there exists* a model $\mathcal{I}_2$ of $\mathcal{T}_2$ such that $\beta(\mathcal{I}_2) = \mathcal{I}_1$ and $\mathcal{I}_2 \models_{\mathcal{L}_2} \gamma(\mathcal{T}_1)$. Observe that $\leq_C^{\exists}$ differs from $\leq_C$ in the quantification pattern of $\mathcal{L}_2$-models.

While $\mathcal{L}_1 \leq_{C,\delta} \mathcal{L}_2$ means that we can reduce the problem of deciding if an $\mathcal{L}_1$-model satisfies an $\mathcal{L}_1$-theory $\mathcal{T}_1$ to that of deciding if a certain $\mathcal{L}_2$-model (e.g., any model we find that satisfies the theory given by δ) satisfies the $\mathcal{L}_2$-theory $g(\mathcal{T}_1)$, if we have that $\mathcal{L}_1 \leq_C^{\exists} \mathcal{L}_2$, then the same problem is reduced to (potentially) *many* instances of the decision problem for $\mathcal{L}_2$ (i.e., one for each model $\mathcal{I}_2$ that satisfies the theory given by δ). One can regard the latter notion as a strong form of *satisfiability preserving* translation (see below).

Local and global reasoning power. The two criteria above could be encompassed under the term *descriptive power*, since they refer to the capacity of a logic to describe models or individuals in a model. The following criteria, on the other hand, refer to the *inferences* that can be made.

Consider first the set $V_{\mathcal{L}}$ of *valid concepts of* $\mathcal{L}$. We say that $\mathcal{L}_2$ *has at least the same local reasoning power as* $\mathcal{L}_1$, notation $\mathcal{L}_1 \leq_{R^l} \mathcal{L}_2$, if there exists an effective mapping α from $\mathcal{L}_1$-concepts to $\mathcal{L}_2$-concepts such that, for every $\mathcal{L}_1$-concept C, $C \in V_{\mathcal{L}_1}$ iff $\alpha(C) \in V_{\mathcal{L}_2}$. Similarly, for a theory $\mathcal{T}$, let $V_{\mathcal{L}}(\mathcal{T})$ denote the set of $\mathcal{L}$-consequences of $\mathcal{T}$, that is the set $\{C : \mathcal{I} \models_{\mathcal{L}} \mathcal{T} \Rightarrow C^{\mathcal{I}} = \Delta^{\mathcal{I}}, \forall \mathcal{I}\}$. We then say that $\mathcal{L}_2$ *has at least the same global reasoning power as* $\mathcal{L}_1$, notation $\mathcal{L}_1 \leq_{R_g} \mathcal{L}_2$, if there exist two effective mappings α and γ such that, for every theory $\mathcal{T}$ of $\mathcal{L}_1$, and every $\mathcal{L}_1$-concept C, $C \in V_{\mathcal{L}_1}(\mathcal{T})$ iff $\alpha(C) \in V_{\mathcal{L}_2}(\gamma(\mathcal{T}))$.

It is worth noticing that $\mathcal{L}_2$ having at least the same local reasoning power as $\mathcal{L}_1$ intuitively means that local reasoning $\mathcal{L}_1$ can be reduced to local reasoning in $\mathcal{L}_2$.

Proposition 2.2 *The following hold (where δ_0 maps a signature to the empty theory):*

(i) *If $\mathcal{L}_1 \leq_{Q,\delta_0} \mathcal{L}_2$, then $\mathcal{L}_1 \leq_{R^l} \mathcal{L}_2$,*

(ii) *If $\mathcal{L}_1 \leq_{Q,\delta_0} \mathcal{L}_2$ and $\mathcal{L}_1 \leq_{C,\delta_0} \mathcal{L}_2$, then $\mathcal{L}_1 \leq_{R^g} \mathcal{L}_2$.*

(iii) *If $\mathcal{L}_1 \leq_{R^g} \mathcal{L}_2$, then $\mathcal{L}_1 \leq_{R^l} \mathcal{L}_2$.*

Notice that for the last case, we use that γ preserves empty theories.

We say that $\mathcal{L}_1$ and $\mathcal{L}_2$ have *equal querying power* if each one has at least the same expressive power as the other. This extends to the other notions in the obvious way.

All the above notions are qualitative in nature. A quantitative comparison can be done by looking at the blowup incurred by the mapping on concepts and/or theories. So-called *succinctness* results (see, e.g., [1]) typically involve lower-bounds (at least exponential, usually) for the blowup incurred by the mapping on concepts for the notion of classification power. An analogous notion of succinctness for querying power was investigated in [6]. The mappings used for defining local and global reasoning power are usually called *satisfiability preserving translations*, and are used as a tool for proving complexity results; the blowup in this case impacts on the final complexity.

Proposition 2.3 *The following relations hold:*

(i) $\mathcal{ALCIQ}be \leq_Q \mathcal{ALCHIQ}be$ *and* $\mathcal{ALCHIQ}be \leq_Q \mathcal{ALCIQ}be$.

(ii) $\mathcal{ALCIQ}be \leq_C \mathcal{ALCHIQ}be$.

(iii) $\mathcal{ALCIQ}be \leq_{R^g} \mathcal{ALCHIQ}be$ *and* $\mathcal{ALCHIQ}e \leq_{R^g} \mathcal{ALCIQ}be$.

Proof Clearly, role-hierarchies add no querying power (since the model is fixed in this case), so it is no surprise that $\mathcal{ALCHIQ}be \leq_Q \mathcal{ALCIQ}be$. It is proved in [18] that $\mathcal{ALCHIQ} \leq_{R^g} \mathcal{ALCIQ}b$; essentially, every role S in a concept C is replaced by the conjunction $R_1 \sqcap R_2 \ldots \sqcap R_n \sqcap S$ of all roles $R_i \sqsubseteq S$. This is trivially extended to the case with $\exists R.\mathsf{Self}$. □

3 Lower Bounds for Expressivity

We want to show that $\mathcal{ALCQ}\mathsf{me}_2$ has enough expressive power to accommodate $\mathcal{ALCHIQ}be$ for all the tasks outlined in the previous section. Moreover, the blowup in all cases will be shown to be polynomial. We shall assume, throughout this section, that $\mathcal{ALCHIQ}be$ formulas are defined over a vocabulary $\mathcal{S}_1 = \langle \mathsf{N_{C1}}, \mathsf{N_{R1}} \rangle$; $\mathcal{ALCQ}\mathsf{me}_2$ concepts, on the other hand, will be built over $\mathcal{S}_2 = \langle \mathsf{N_{C2}}, \mathsf{N_{R2}} \rangle$, with $\mathcal{S}_1 \subseteq \mathcal{S}_2$. Further assumptions regarding $\mathcal{S}_2$ will be made as needed.

Let us start by considering the querying power of $\mathcal{ALCQ}\mathsf{me}_2$. Clearly, role-hierarchies can be disregarded here, so we concentrate on the features of $\mathcal{ALCIQ}be$. The first thing to observe is that the $\exists R.\mathsf{Self}$ concept corresponds to $\mathsf{I}.\exists R.\mathsf{me}$. To account for the inverse role of R_i we assume $\mathsf{N_{R2}}$ to contain a (fresh) role $\tilde{S}_{R_i^-}$ for each $R_i \in \mathsf{N_{R1}}$; its meaning is then defined by way of the TBox axiom:

$$(\mathsf{I}.\forall R_i.\exists \tilde{S}_{R_i^-}.\mathsf{me}\quad) \sqcap (\mathsf{I}.\forall \tilde{S}_{R_i^-}.\exists R_i.\mathsf{me}). \tag{3}$$

In any model that satisfies axiom (3), $\tilde{S}_{R_i^-}$ will be the inverse role of R_i. In order to deal, in addition, with safe Boolean combination of roles, we further assume that $\mathsf{N_{R2}}$ contains a fresh role $\tilde{S}_R$ for each complex role expression R built over atomic roles in $\mathsf{N_{R1}}$. We give them meaning via the following TBox axioms (where we identify $\tilde{S}_{R_i}$ with R_i, for $R_i \in \mathsf{N_{R1}}$):

$$(\mathsf{I}.\forall \tilde{S}_{R^-}.\exists \tilde{S}_R.\mathsf{me}) \sqcap (\mathsf{I}.\forall \tilde{S}_R.\exists \tilde{S}_{R^-}.\mathsf{me}) \quad (4)$$

$$\mathsf{I}.\forall \tilde{S}_{R\sqcup S}.(\exists \tilde{S}_{R^-}.\mathsf{me} \sqcup \exists \tilde{S}_{S^-}.\mathsf{me}) \quad (5)$$

$$\mathsf{I}.\forall \tilde{S}_{R\sqcap S}.(\exists \tilde{S}_{R^-}.\mathsf{me} \sqcap \exists \tilde{S}_{S^-}.\mathsf{me}) \quad (6)$$

$$\mathsf{I}.\forall \tilde{S}_{R-S}.(\exists \tilde{S}_{R^-}.\mathsf{me} \sqcap \neg\exists \tilde{S}_{S^-}.\mathsf{me}) \quad (7)$$

$$\mathsf{I}.\forall \tilde{S}_R.\exists \tilde{S}_{(R\sqcup S)^-}.\mathsf{me} \quad (8)$$

$$\mathsf{I}.\forall \tilde{S}_R.((\exists \tilde{S}_{S^-}.\mathsf{me} \sqcap \exists \tilde{S}_{(R\sqcap S)^-}.\mathsf{me}) \sqcup (\neg\exists \tilde{S}_{S^-}.\mathsf{me} \sqcap \exists \tilde{S}_{(R-S)^-}.\mathsf{me})) \quad (9)$$

Clearly, (4) makes inverses behave correctly for non-atomic roles as well; taking advantage of this, we have a sound definition for the symbols in (5)–(7) while (8)–(9) make it complete (one can add axioms for commutativity of $\sqcap$ and $\sqcup$, if needed). All this together immediately gives us:

Proposition 3.1 *$\mathcal{ALCHIQ}be \leq_Q \mathcal{ALCQ}\mathsf{me}_2$.*

We can move now to classification power. Because of the analysis above, it suffices to show how to encode role hierarchies in $\mathcal{ALCQ}\mathsf{me}_2$. Using the symbols $\tilde{S}_R$ already defined (again, identifying $\tilde{S}_{R_i}$ with $R_i \in \mathsf{N_{C1}}$), it is clear that a role inclusion axiom of the form $R \sqsubseteq S$ can be defined with the TBox axiom:

$$\mathsf{I}.\forall \tilde{S}_{R^-}.\exists \tilde{S}_S.\mathsf{me} \quad (10)$$

We then have

Proposition 3.2 *$\mathcal{ALCHIQ}be \leq_C \mathcal{ALCQ}\mathsf{me}_2$.*

Notice, though, that it is not possible to immediately conclude, from all these results, that $\mathcal{ALCHIQ}be \leq_{R^l} \mathcal{ALCQ}\mathsf{me}_2$. In essence, the mapping α induced in the proof of Proposition 3.1 requires the support of a non-empty TBox. This, in turn, means that neither can we conclude yet that $\mathcal{ALCHIQ}be \leq_{R^g} \mathcal{ALCQ}\mathsf{me}_2$. We turn, then, to proving that this reduction holds as well. For this, one needs a standard auxiliary result. For a relation $T \subseteq X \times X$, define $T^0 = Id_X$ and $T^{n+1} = T^n \cup (T \circ T^n)$; one then has the following:

Lemma 3.3 *Let C be a closed $\mathcal{ALCQ}\mathsf{me}_2$-concept over a vocabulary $\mathcal{S} = \langle \mathsf{N_C}, \mathsf{N_R} \rangle$ and let $a \in C^{\mathcal{I}}$ for some $\mathcal{I}$. Moreover, let $\mathcal{I}_{k,a}$ denote the restriction of $\mathcal{I}$ to the domain $\Delta^{\mathcal{I}_{k,a}} = \bigcup_{R_i \in \mathsf{N_R}} \{b : aR_i^k b\}$. Then we have that $a \in C^{\mathcal{I}_{k,a}}$ whenever $k \geq rank(C)$.*

With this lemma we can prove,

Theorem 3.4 *$\mathcal{ALCHIQ}be \leq_{R^g} \mathcal{ALCQ}\mathsf{me}_2$.*

Proof We need to define a mapping γ of axioms to finite $\mathcal{ALCQ}\mathsf{me}_2$-TBoxes for the $\mathcal{ALCHIQ}be$-TBox, and a mapping α from $\mathcal{ALCHIQ}be$-concepts to $\mathcal{ALCQ}\mathsf{me}_2$-concepts. For the former, we can just reuse the mapping from the proof of Proposition 3.2, with the proviso that each mapped axiom introduces also the finitely many definitional axioms (from the mapping δ of the vocabulary) for the symbols that occur in it. Of course, we cannot do this for α and the problem is how to locally define the symbols that are not mentioned in the $\mathcal{ALCHIQ}be$-TBox. We can then assume, without loss of generality, that we are working with an empty TBox. Let C be an $\mathcal{ALCHIQ}be$-concept, and let C' be the $\mathcal{ALCQ}\mathsf{me}_2$-concept obtained from the translation in Proposition 3.1. From Lemma 3.3, we know that it is enough to define the $\tilde{S}_R$ symbols occurring in C' until depth $rank(C') = m$. For this, we resort to a fresh role U, which we force to behave as a "universal role up to depth m" by means of the concept

$$D = \textstyle\bigsqcap_{i=1}^{k} \bigsqcap_{j=0}^{m} \forall^j U.\mathsf{I}.\forall R_i.(\exists U.\mathsf{me} \sqcap \mathsf{I}.\forall U.\exists U.\mathsf{me}) \tag{11}$$

where $R_1 \dots R_k$ are the atomic roles occurring in C' and $\forall^n R.D$ denotes D if $n = 0$, and $\forall R.\forall^{n-1} R.D$ otherwise. Having defined U this way, we can give the desired meaning, up to depth m, to every role $\tilde{S}_R$ occurring in C' using suitable concepts $E(R)$; e.g.,

$$E(R_i^-) = \textstyle\bigsqcap_{j=0}^{m} \forall^j U.(\mathsf{I}.\forall R_i.\exists \tilde{S}_{R_i^-}.\mathsf{me} \sqcap \mathsf{I}.\forall \tilde{S}_{R_i^-}.\exists R_i.\mathsf{me}). \tag{12}$$

The required concept then has the shape $C' \sqcap D \sqcap \bigwedge_R E(R)$ where R ranges over all complex roles occurring in C. □

4 A Tight Upper Bound for Expressive Power

We have seen in the previous section that $\mathcal{ALCQ}\mathsf{me}_2$ is at least as expressive as $\mathcal{ALCHIQ}be$. We now make this bound tight by showing that $\mathcal{ALCIQ}be$ is also as expressive as $\mathcal{ALCQ}\mathsf{me}_2$. It must be observed, though, that in this case we will incur an exponential blowup. We will later show that a weaker result can be obtained with a polynomial blowup.

We begin, then, by showing that $\mathcal{ALCQ}\mathsf{me}_2 \leq_{Q,\delta_0} \mathcal{ALCIQ}be$, where δ_0 maps a vocabulary to the empty TBox. From this it will be straightforward to derive the remaining bounds. The proof of the first result will go in two steps: we show that every $\mathcal{ALCQ}\mathsf{me}_2$-concept can be taken to a normal form from which an equivalent $\mathcal{ALCIQ}be$ concept can be derived. Both translation steps will cause an exponential blowup, but no additional symbols will be introduced.

Definition 4.1 *Assume a fixed, finite set of roles $\mathcal{R} = \{R_1 \dots R_n\}$. We let $P(\mathcal{R})$ denote the set of all maximal satisfiable conjunctive clauses over the set $\{\exists R_1.\mathsf{me}, \dots \exists R_n.\mathsf{me}\}$. Moreover, we say that $\mathsf{I}.\geq_n R.D$ is an $\mathcal{R}$-*covered con-cept *if D is of the form $\bigsqcup_{C_i \in P(\mathcal{R})} (C_i \sqcap D_i)$; in this case, the C_i are called* selectors*. Finally, a concept C is in* expanded form *if i) every $\mathsf{I}.\geq_n R.D$ occurring in C is an $\mathcal{R}$-covered concept, for some $\mathcal{R}$, and ii)* me *occurs in C only in the selectors of covered concepts and in concepts of the form $\mathsf{I}.\exists R.\mathsf{me}$.*

E.g., an $\{S\}$-covered concept is of the form $\mathsf{I}.\geq_n R.(((\exists S.\mathsf{me}) \sqcap D_1) \sqcup \neg((\exists S.\mathsf{me}) \sqcap D_2))$. The intuition behind this definition is that if $\mathsf{I}.\geq_n R.D$ is an $\mathcal{R}$-covered concept, then D describes a property D_i for the R-successors of the point of evaluation that is determined by the way they link back to it with respect to some set of roles $\mathcal{R}$. Notice that every R-successor of an element satisfying an $\mathcal{R}$-covered concept will satisfy one and only one of its selectors.

Lemma 4.2 *Every closed concept is equivalent to a closed concept in expanded form.*

Proof We define two mappings γ_1 and γ_2 such that for all closed $\mathcal{ALCQ}\mathsf{me}_2$-concept C, $\gamma_2(\gamma_1(C))$ is in expanded form and equivalent to C. Intuitively, γ_1 deals with those me that occur at depth 1 while γ_2 does so for those at depth 2. To achieve this, γ_2 turns every $(\mathsf{I}.)\geq_n R.D$ in $\gamma_1(C)$ into a covered concept. In what follows, we will use $C[x/y]$ to denote the substitution of all the *top-level* occurrences of x in C by y; here "top-level" means "not under the scope of any qualified number restriction" (e.g. $(\mathsf{me} \sqcap \geq_n R.\mathsf{me})[\mathsf{me}/A] = A \sqcap \geq_n R.\mathsf{me}$). For $x \in 2 = \{0, 1\}$ and a concept C, we let C^x denote C, if $x = 1$ and $\neg C$ otherwise.

We start by defining γ_1, whose only non-trivial case is $\gamma_1(\mathsf{I}.\geq_n R.C)$, given by

$$\bigsqcup_{x,y,z\in 2} \big(\mathsf{I}.\exists R.\mathsf{me}^x \sqcap \gamma_1(\mathsf{I}.C[\mathsf{me}/\top])^y \sqcap \gamma_1(\mathsf{I}.C[\mathsf{me}/\bot])^z \sqcap \mathsf{I}.\geq_{n-xy+xz} R.\gamma_1(C[\mathsf{me}/\bot])\big) \tag{13}$$

where $\mathsf{I}.C[\mathsf{me}/\top]$ (meant to be read $\mathsf{I}.(C[\mathsf{me}/\top])$) distributes I among Booleans as expected and, for $n - xy = 0$, we have $\geq_{n-xy+xz} R.D \equiv \top$ (even when $n - xy + xz = 1$). In essence, $\gamma_1(\mathsf{I}.\geq_n R.C)$ eliminates the top-level me in C, replacing it with $\bot$. For this to be valid, some adjustments need to be made: if $\exists R.\mathsf{me}$ and $\mathsf{I}.C[\mathsf{me}/\top]$ hold (at me), then we need one witness less (zero witnesses for the case $n = 1$), but if $\exists R.\mathsf{me}$ and $\mathsf{I}.C[\mathsf{me}/\bot]$ hold (at me) we need to account for an extra "false positive" (there is a corner case when $n, x, y, z = 1$, where zero witnesses are needed as well).

For γ_2 we require some additional notation. We use $C[x \in X/f(x)]$ to denote the uniform substitution in C of every concept $x \in X$ by $f(x)$. For $f, g \in 2^X$, we define the function $fg \in 2^X$ as $fg(x) = f(x) \cdot g(x)$. In addition, we let $F(C)$ be the set of all concepts of the form $\geq_m S.D$ occurring top-level in C without an $\mathsf{I}.$ in front; e.g., $F((\mathsf{I}.\geq_5 R.\mathsf{me}) \sqcap \neg\geq_3 R.A) = \{\geq_3 R.A\}$. Given $f \in 2^{F(C)}$, we say that f is $\mathcal{R}_C$-consistent if there are no distinct $\geq_n R.D$ and $\geq_m R.E$ in $F(C)$ such that $f(\geq_n R.D) \neq f(\geq_m R.E)$, and denote by $2_c^{F(C)}$ the set of all $\mathcal{R}_C$-consistent functions in $2^{F(C)}$. Now, each $f \in 2_c^{F(C)}$ induces these conjunctive clauses:

$$Cl_*(f) = \textstyle\bigsqcap_{\geq_m S.D \in F(C)} D[\mathsf{me}/*]^{f(\geq_m S.D)} \quad (* \in \{\top, \bot\}) \tag{14}$$

$$Cl_\exists(f) = \textstyle\bigsqcap_{\geq_m S.D \in F(C)} \exists S.\mathsf{me}^{f(\geq_m S.D)} \tag{15}$$

Moreover, given $f, g, h \in 2_c^{F(C)}$ we define $C[f, g, h]$, as:

$$C[f, g, h] = C[x = \geq_m S.D \in F(C)/\geq_{m-fh(x)+gh(x)} S.D[\mathsf{me}/\bot]] \tag{16}$$

Again, for $m - fh(x) = 0$, we assume $\geq_{m-fh(x)+gh(x)} S.D \equiv \top$. We can now define γ_2, the only non-trivial case being that for $\gamma_2(\mathsf{I}.\geq_n R.C)$, given by

$$\bigsqcup_{f,g\in 2_c^{F(C)}} \big[\gamma_2\big(Cl_\top(f) \sqcap Cl_\bot(g)\big) \sqcap \mathsf{I}.\geq_n R.\bigsqcup_{h\in 2_c^{F(C)}} (Cl_\exists(h) \sqcap \gamma_2(C[f, g, h]))\big] \tag{17}$$

The principle behind (17) is analogous to that used for γ_1. In this case we need an equivalent of x, y and z in (13) for each concept in $F(C)$, and we use functions f, g and h for this. Observe that $\gamma_2(\mathsf{I}.{\geq_n}R.C)$ is an $\mathcal{R}$-covered concept, where $\mathcal{R}$ is the set $\{S : {\geq_m}S.D \in F(C)\}$. Using this insights, it is not hard to verify that $\gamma_2(\gamma_1(C))$ is equivalent to C and in expanded form. □

Proposition 4.3 *$\mathcal{ALCQ}\mathsf{me}_2 \leq_Q \mathcal{ALCIQ}be$.*

Proof By Lemma 4.2 we can assume that closed $\mathcal{ALCQ}\mathsf{me}_2$-concepts are in expanded form. We can therefore define a translation δ, mapping concepts in this form to $\mathcal{ALCIQ}be$-concepts, by making it commute with the Booleans, stipulating $\delta(\mathsf{I}.\exists R.\mathsf{me}) = \exists R.\mathsf{Self}$ and $\delta({\geq_n}R.\mathsf{me}) = {\geq_n}R.\delta(\mathsf{me})$; and taking $\delta(\mathsf{I}.{\geq_n}R.((C_1 \sqcap D_1) \sqcup \ldots \sqcup (C_m \sqcap D_m))$, where the C_i are the selectors of the $\mathcal{R}$-covered concept, to be

$$\bigsqcup\nolimits_{n=\sum_{j=1}^{m} k_j} \bigsqcap\nolimits_{1 \leq i \leq m} {\geq_{k_j}}\theta(R, C_i).\delta(D_j) \tag{18}$$

where θ maps a role and selector to a safe role expression as follows:

$$\begin{aligned}\theta(R, \exists S_1.\mathsf{me} \sqcap \cdots \sqcap \exists S_l.\mathsf{me} \sqcap \neg\exists S_{l+1}.\mathsf{me} \sqcap \cdots \sqcap \neg\exists S_l.\mathsf{me}) = \\ (R \sqcap S_1^- \sqcap \cdots \sqcap S_l^-) - (S_{l+1}^- \sqcap \cdots \sqcap S_m^-)\end{aligned} \tag{19}$$

What $\delta(\mathsf{I}.{\geq_n}R.((C_1 \sqcap D_1) \sqcup \ldots \sqcup (C_m \sqcap D_m))$ does is to consider all the possible ways in which we can distribute the n required successors among the m equivalence classes given by the selectors. The number of cases is bounded by the number of *partitions of* n, which is exponential on n [4]. □

It is not hard to verify that from this proof one also gets the following:

Corollary 4.4 *$\mathcal{ALCQ}\mathsf{me}_2 \leq_C \mathcal{ALCIQ}be$ and $\mathcal{ALCQ}\mathsf{me}_2 \leq_{R^g} \mathcal{ALCIQ}be$.*

All the intermediate transformations in the above proofs incur exponential blowups. It is not hard, however, to verify that the overall blowup in formula size is still only exponential. In order to obtain a polynomial blowup, we need to relax the properties we require of the translation from preservation of *satisfaction* to preservation of *satisfiability*:

Proposition 4.5 *$\mathcal{ALCQ}\mathsf{me}_2 \leq_C^{\exists} \mathcal{ALCIQ}be$, with only a polynomial blowup.*

Proof We introduce a satisfiability preserving translation that adds fresh atomic concept and role symbols, and ensures that a model for the translated formula is turned into a model for the $\mathcal{ALCQ}\mathsf{me}_2$-formula simply by ignoring the interpretation of the extra symbols. We can view the translation as a two-step process. First, we map an $\mathcal{ALCQ}\mathsf{me}_2$-concept to an $\mathcal{ALCQ}b\mathsf{me}_2$-concept (i.e., a concept of $\mathcal{ALCQ}\mathsf{me}_2$ enriched with safe Boolean combinations of roles) in such a way that if ${\geq_n}R.C$ occurs in the resulting concept, then $C \neq \top$ implies $n = 1$. This is straightforward to achieve: a concept $(\mathsf{I}.){\geq_n}R.C$, with $n > 1$, occurring in a positive context (i.e., under an even number of negations) is mapped to ${\geq_n}(R \sqcap R').\top \sqcap (\mathsf{I}.)\forall R'.C$, for a fresh R'; while the same concept in a negative context is mapped to ${\geq_n}(R \sqcap R').\top \sqcup \neg(\mathsf{I}.)\forall(R - R').\neg C$, with R' fresh (these transformations need to be performed top-down).

We therefore need only eliminate all the occurrences of $\mathsf{I}.\forall R.C$ in the resulting concept. We can deal with the top-level occurrences of me in C replacing $\mathsf{I}.\forall R.C$ by

$$\mathsf{I}.\forall R.C[\mathsf{me}/D] \sqcap ((\exists R.\mathsf{Self} \sqcap D \sqcap \leq_1 R.D) \sqcup (\neg \exists R.\mathsf{Self} \sqcap \neg D \sqcap \forall R.\neg D)) \quad (20)$$

where D is fresh and $C[\mathsf{me}/D]$ is as in the proof of Lemma 4.2. Let $\mathsf{I}.\forall R.C$ occur in the concept resulting after this transformation; if me happens to occur free in C, then some $\forall S.D$ must occur in C with me free (top-level) in D. We can therefore substitute $\mathsf{I}.\forall R.C$ by

$$\mathsf{I}.\forall R.C[\forall S.D/\forall S.D[\mathsf{me}/E]] \sqcap E \sqcap (\forall (R \sqcap S^-).\leq_1 S.E) \sqcap (\forall (R - S^-).\forall S.\neg E) \quad (21)$$

where E is fresh. After all the occurrences of me are eliminated, the $\mathsf{I}.$ can be removed as well. It is not hard to see that these transformations incur in a polynomial blowup and that any model for the translated formula is trivially turned into a model for the original one by applying the forgetful mapping of Definition 2.1. □

5 Decidability and Complexity

It is time to discuss the decidability and complexity of the local satisfiability problem for $\mathcal{ALCQ}\mathsf{me}_2$, both over empty and over general TBoxes. Because Proposition 4.5 gives us an effective, polynomial, satisfiability preserving reduction of $\mathcal{ALCQ}\mathsf{me}_2$-concepts to $\mathcal{ALCIQ}be$-concepts, and because Proposition 3.1 gives us a reduction in the opposite direction, the complexities of $\mathcal{ALCQ}\mathsf{me}_2$ and $\mathcal{ALCIQ}be$ must match.

While the logic $\mathcal{ALCIQ}b$ is known to be PSPACE-complete for satisfiability over empty TBoxes and EXPTIME-complete for satisfiability over general TBoxes [18], not much appears to be known about the complexity of $\mathcal{ALCIQ}be$. We can conclude that it is decidable and at most in N2EXPTIME (for general TBoxes) from the fact that the extension of $s\mathcal{ROIQ}$ with Boolean combination of *simple roles* is N2EXPTIME-complete [13] and that in $\mathcal{ALCIQ}be$ all atomic roles would be "simple" in $s\mathcal{ROIQ}$ terminology. In this section we will then show that $\mathcal{ALCIQ}be$ is, in fact, not harder than $\mathcal{ALCIQ}b$.

The technique we will employ is a variation of the *internalized tableaux* of [9]. Although we will give a specialized proof for the case of $\mathcal{ALCIQ}be$, it should be clear that our construction can be used to show that adding the $\exists R.\mathsf{Self}$ concept to a description logic in general does not raise its computational complexity. The relevant property to have is the *self-loop free model property*, i.e. if a concept (resp. theory) is satisfiable, then it is satisfiable in a model without self-loops.

Lemma 5.1 *$\mathcal{ALCIQ}b$ has the self-loop free model property.*

Proof Given an interpretation $\mathcal{I}$, one can build an equivalent interpretation $\mathcal{J}$ that is self-loop free as follows. Take as domain of $\mathcal{J}$ two copies of the domain of $\mathcal{I}$, i.e., $\Delta^{\mathcal{J}} = 2 \times \Delta^{\mathcal{I}}$. Concepts and roles are preserved, except for self-loops, which are replaced by links to the corresponding element on the other copy; formally we have that $A^{\mathcal{J}} = 2 \times A^{\mathcal{I}}$ and $R^{\mathcal{J}}$ is given by

$$\begin{aligned} &\{((x,a),(x,b)) : x \in 2, (a,b) \in R^{\mathcal{I}}, a \neq b\} \\ \cup\, &\{((x,a),(y,a)) : x,y \in 2, x \neq y, (a,a) \in R^{\mathcal{I}}\} \end{aligned} \tag{22}$$

The proof that $\mathcal{I}$ and $\mathcal{J}$ are equivalent is left to the reader. □

Lemma 5.2 *TBox satisfiability in $\mathcal{ALCIQ}be$ can be polynomially reduced to TBox satisfiability in $\mathcal{ALCIQ}b$.*

Proof Let $\mathcal{T}$ be an $\mathcal{ALCIQ}be$ TBox over a vocabulary $\langle \mathsf{N_C}, \mathsf{N_R}\rangle$, and let Σ be the smallest set that contains every concept occurring in $\mathcal{T}$, is closed under subconcepts and single negations, and contains the concept $\exists R.\mathsf{Self}$ for every role expression R occurring in $\mathcal{T}$. Let $\mathsf{N'_C} = \mathsf{N_C} \cup \{B_C : C \in \Sigma\}$; we then define the $\mathcal{ALCIQ}b$ TBox $\mathcal{T}'$, over the vocabulary $\langle \mathsf{N'_C}, \mathsf{N_R}\rangle$. It contains an axiom $B_C \sqsubseteq B_D$ for each $C \sqsubseteq D$ in $\mathcal{T}$, plus the following definitions:

$$\begin{aligned} B_A &\equiv A \ (A \in \mathsf{N_C}) & B_{\neg C} &\equiv \neg B_C \\ B_{C \sqcap D} &\equiv B_C \sqcap B_D & B_{\exists R.\mathsf{Self}} &\equiv B_{\exists R^-.\mathsf{Self}} \\ B_{\exists (R \sqcap S).\mathsf{Self}} &\equiv B_{\exists R.\mathsf{Self}} \sqcap B_{\exists S.\mathsf{Self}} & B_{\exists (R \sqcup S).\mathsf{Self}} &\equiv B_{\exists R.\mathsf{Self}} \sqcup B_{\exists S.\mathsf{Self}} \\ B_{\exists (R - S).\mathsf{Self}} &\equiv B_{\exists R.\mathsf{Self}} \sqcap \neg B_{\exists S.\mathsf{Self}} \end{aligned}$$

$$B_{\geq_n R.C} \equiv ((B_C \sqcap B_{\exists R.\mathsf{Self}}) \sqcap \geq_{n-1} R.B_C) \sqcup (\neg(B_C \sqcap B_{\exists R.\mathsf{Self}}) \sqcap \geq_n R.B_C).$$

Recall that for $n = 1$, we identify $\geq_{n-1} R.C$ with $\top$. Clearly, $\mathcal{T}$ has size linear in the size of $(\mathcal{T}, \Sigma)$, which in turn is linear in $\mathcal{T}$. It only remains to see that $\mathcal{T}$ and $\mathcal{T}'$ are equisatisfiable. It is clear that any model $\mathcal{I}$ for $\mathcal{T}$ is expanded to a model $\mathcal{I}'$ for $\mathcal{T}'$ by setting $B_C^{\mathcal{I}'} = C^{\mathcal{I}}$, for all $C \in \Sigma$. For the other direction, let $\mathcal{I}'$ be a model for $\mathcal{T}'$ and notice that by Lemma 5.1, we can assume $\mathcal{I}'$ to contain no self-loops. We then obtain a model $\mathcal{I}$ by restricting $\mathcal{I}'$ to $\langle \mathsf{N_C}, \mathsf{N_R}\rangle$ and adding the necessary self-loops, i.e., setting $R^{\mathcal{I}} = R^{\mathcal{I}'} \cup \{(a,a) : a \in B^{\mathcal{I}'}_{\exists R.\mathsf{Self}}\}$. It is not hard to verify that $\mathcal{I}$ is a model for $\mathcal{T}$. □

Theorem 5.3 *The problem of concept satisfiability over general TBoxes for the logic $\mathcal{ALCIQ}be$ is EXPTIME-complete.*

Proof This follows directly from Lemma 5.2 (observing that satisfiability over general TBoxes can be reduced to TBox satisfiability) and EXPTIME-completeness of TBox satisfiability for $\mathcal{ALCIQ}b$ [18]. □

Theorem 5.4 *The concept satisfiability problem (over empty TBoxes) for the logic $\mathcal{ALCIQ}be$ is PSPACE-complete.*

Proof Given that satisfiability for $\mathcal{ALCIQ}be$ is PSPACE-complete [18], we only need to give a polynomial, satisfiability preserving translation from $\mathcal{ALCIQ}be$ to $\mathcal{ALCIQ}b$. This is done following the idea of the proof of Lemma 5.2: given an $\mathcal{ALCIQ}be$-concept C containing atomic roles $R_1 \ldots R_m$, one builds the $\mathcal{ALCHIQ}b$-concept $B_C \sqcap \forall^n(R_1 \sqcup \ldots \sqcup R_m).\mathit{Def}(B)$, with $n = \mathit{rank}(C)$, $\forall^0 R.D \equiv D$, $\forall^{n+1} R.D = \forall R.\forall^n R.D$, and $\mathit{Def}(B)$ is the conjunction of the definitional axioms for the fresh concepts B_D (with $D \in \Sigma$)

given in the proof of Lemma 5.2. Equisatisfiability follows by a standard argument. □

Corollary 5.5 *Satisfiability over general (resp. empty) TBoxes for $\mathcal{ALCQ}\mathsf{me}_2$ is EXPTIME-complete (resp. PSPACE-complete).*

6 Conclusions

Although it is known that decidability breaks easily when self-referential concepts are added to description logics, we have shown that, under controlled conditions, it is indeed feasible to extend $\mathcal{ALCQ}$ (i.e. $\mathcal{ALC}$ with qualified number restrictions, equivalently graded multi-modal logic) with a form of self-referential reasoning without affecting the complexity. Specifically, we have defined the logic $\mathcal{ALCQ}\mathsf{me}_2$, which includes the I-me construct that allows naming one state at a time for future reference but limits occurrences of me to (modal) depth at most 2 from the binding I. We have shown that $\mathcal{ALCQ}\mathsf{me}_2$ is expressively equivalent to the DL $\mathcal{ALCHIQ}be$, which includes role inverses and hierarchies, safe Boolean combinations of roles, and the self-loop construct $\exists R.\,\mathsf{Self}$. The translation of $\mathcal{ALCQ}\mathsf{me}_2$ into $\mathcal{ALCHIQ}be$, however, has an exponential blowup (it remains an open question whether this can be avoided). We have therefore given a second reduction of $\mathcal{ALCQ}\mathsf{me}_2$ to $\mathcal{ALCHIQ}be$ that is only satisfiability-preserving but has polynomial blowup. After subsequent analysis of the complexity of $\mathcal{ALCHIQ}be$, this has allowed us to prove that $\mathcal{ALCQ}\mathsf{me}_2$ is decidable in PSPACE over the empty TBox, and in EXPTIME over general TBoxes — the same bounds as for $\mathcal{ALCQ}$ or indeed basic $\mathcal{ALC}$.

In future research, we will study the addition of controlled binding constructions to richer DLs. Because $\mathcal{ALCHIQ}$ is embedded in $\mathcal{ALCQ}\mathsf{me}_2$, it is clear that adding nominals to the mix must put us at least in NEXPTIME, and we conjecture NEXPTIME-completeness for this language. It would also be interesting to know in which cases the interaction with transitive roles is safe. A further branch of investigation is the generalization of the results at depth 2 to extended description logics featuring, e.g., uncertainty or defaults, modelled generically in the framework of coalgebraic logic (see, e.g., [15]), thus improving a result at depth 1 obtained in our previous work [9].

Acknowledgements The authors wish to thank the referees of an earlier version of this paper for suggesting the possibility of proving Proposition 4.5. Moreover, we acknowledge discussions with Erwin R. Catesbeiana on the merits of consistency.

References

[1] Adler, M. and N. Immerman, *An n! lower bound on formula size*, ACM Trans. Comput. Logic **4** (2003), pp. 296–314.

[2] Areces, C., D. Figueira, S. Figueira and S. Mera, *The expressive power of memory logics*, Rev. Symb. Log. **2** (2011), pp. 290–318.

[3] Areces, C. and B. ten Cate, *Hybrid logics*, in: P. Blackburn, J. van Benthem and F. Wolter, editors, *Handbook of Modal Logic*, Elsevier, 2007 pp. 821–868.

[4] Ayoub, R., "An Introduction to the Analytic Theory of Numbers," Number 10 in Mathematical Surveys, American Mathematical Society, 1963.

[5] Baader, F., D. Calvanese, D. L. McGuinness, D. Nardi and P. F. Patel-Schneider, editors, "The Description Logic Handbook," Cambridge University Press, 2003.

[6] Figueira, S. and D. Gorín, *On the size of shortest modal descriptions*, in: *Advances in Modal Logic, vol. 8* (2010), pp. 120–139.

[7] Goguen, J. A. and R. M. Burstall, *Institutions: Abstract model theory for specification and programming*, J. ACM **39** (1992), pp. 95–146.

[8] Goguen, J. A. and G. Rosu, *Institution morphisms*, Formal Asp. Comput. **13** (2002), pp. 274–307.

[9] Gorín, D. and L. Schröder, *Narcissists are easy, stepmothers are hard*, in: *Foundations of Software Science and Computation Structures, FoSSaCS 2012*, LNCS **7213** (2012), pp. 240–254.

[10] Horrocks, I., O. Kutz and U. Sattler, *The even more irresistible* $\mathcal{SROIQ}$, in: *Proc. of the 10th Int. Conf. on Principles of Knowledge Represe ntation and Reasoning (KR2006)* (2006), pp. 57–67.

[11] Hustadt, U., B. Motik and U. Sattler, *A decomposition rule for decision procedures by resolution-based calculi*, in: F. Baader and A. Voronkov, editors, *Proc. of the 11th Int. Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2004)*, LNAI **3452** (2005), pp. 21–35.

[12] Marx, M., *Narcissists, stepmothers and spies*, in: *Proc. of the 2002 International Workshop on Description Logics (DL'02)*, CEUR **53**, 2002.

[13] Rudolph, S., M. Krötzsch and P. Hitzler, *Cheap boolean role constructors for description logics*, in: *Proc. of 11th European Conf. on Logics in Artificial Intelligence (JELIA)*, LNAI (2008), pp. 362–374.

[14] Schneider, T., "The Complexity of Hybrid Logics over Restricted Classes of Frames," Ph.D. thesis, Univ. of Jena (2007).

[15] Schröder, L. and D. Pattinson, *PSPACE bounds for rank-1 modal logics*, ACM Trans. Comput. Log. **10** (2009), pp. 13:1–13:33.

[16] ten Cate, B. and M. Franceschet, *On the complexity of hybrid logics with binders*, in: *Computer Science Logic, CSL 2005*, LNCS **3634**, 2005, pp. 339–354.

[17] Tobies, S., *The complexity of reasoning with cardinality restrictions in expressive description logics*, J. Artif. Intell. Res. **12** (2000), pp. 199–217.

[18] Tobies, S., "Complexity results and practical algorithms for logics in Knowledge Representation," Ph.D. thesis, RWTH Aachen (2001).