EPiC Series in Computing

Volume 68, 2020, Pages 1–25

ANDREI-60. Automated New-era Deductive Reasoning Event in Iberia

Description Logics That Count, and What They Can and Cannot Count

Franz Baader and Filippo De Bortoli

Institute of Theoretical Computer Science, TU Dresden, Germany
franz.baader@tu-dresden.de, filippo.de_bortoli@tu-dresden.de

Abstract

Simple counting quantifiers that can be used to compare the number of role successors of an individual or the cardinality of a concept with a fixed natural number have been employed in Description Logics (DLs) for more than two decades under the respective names of number restrictions and cardinality restrictions on concepts. Recently, we have considerably extended the expressivity of such quantifiers by allowing to impose set and cardinality constraints formulated in the quantifier-free fragment of Boolean Algebra with Presburger Arithmetic (QFBAPA) on sets of role successors and concepts, respectively. We were able to prove that this extension does not increase the complexity of reasoning.

In the present paper, we investigate the expressive power of the DLs obtained in this way, using appropriate bisimulation characterizations and 0–1 laws as tools to differentiate between the expressiveness of different logics. In particular, we show that, in contrast to most classical DLs, these logics are no longer expressible in first-order predicate logic (FOL), and we characterize their first-order fragments. In most of our previous work on DLs with QFBAPA-based set and cardinality constraints we have employed finiteness restrictions on interpretations to ensure that the obtained sets are finite, as required by the standard semantics for QFBAPA. Here we dispense with these restrictions to ease the comparison with classical DLs, where one usually considers arbitrary models rather than finite ones, easier. It turns out that doing so does not change the complexity of reasoning.

1 Introduction

Description Logics (DLs) [6] are a well-investigated family of logic-based knowledge representation languages, which are frequently used to formalize ontologies for application domains such as biology and medicine [18]. To define the important notions of such an application domain as formal concepts, DLs state necessary and sufficient conditions for an individual to belong to a concept. These conditions can be Boolean combinations of atomic properties required for the individual (expressed by concept names) or properties that refer to relationships with other individuals and their properties (expressed as role restrictions). Adapting an example from [8, 3] to a different domain, the concept of a Computer Science author can be formalized by the concept description

$$\mathit{Person} \sqcap \exists \mathit{published}.(\mathit{Paper} \sqcap \forall \mathit{topic}.\mathit{CS}),$$

L. Kovacs, K. Korovin and G. Reger (eds.), ANDREI-60 (EPiC Series in Computing, vol. 68), pp. 1–25

which uses the concept names *Person*, *Paper* and *CS*, and the role names *published* and *topic*, as well as the concept constructors conjunction ($\sqcap$), existential restriction ($\exists r.C$), and value restriction ($\forall r.C$). It describes the set of all persons that have published a paper all of whose topics lie in the area of Computer Science. The concept inclusion (CI)

$$\mathit{CS\text{-}author} \sqsubseteq \mathit{Person} \sqcap \exists \mathit{published}.(\mathit{Paper} \sqcap \forall \mathit{topic}.\mathit{CS})$$

can then be used to state that Computer Science authors need to belong to this concept description.

Numerical constraints on the number of role successors (so-called number restrictions) have been used in DLs for more than three decades [10, 20, 19]. For example, using number restrictions, we can define prolific authors as those having published at least 100 papers:

$$\mathit{Prolific\text{-}author} \sqsubseteq \mathit{Person}\ \sqcap (\geqslant 100\, \mathit{published}.\mathit{Paper}).$$

The exact complexity of reasoning in $\mathcal{ALCQ}$, the DL that has all Boolean operations and number restrictions of the form $(\leqslant n\, r.C)$ and $(\geqslant n\, r.C)$ as concept constructors, was determined by Stephan Tobies [29, 31]: it is PSpace-complete without CIs and ExpTime-complete w.r.t. CIs, independently of whether the numbers occurring in the number restrictions are encoded in unary or binary. Note that, using unary coding of numbers, the number n is assumed to contribute n to the size of the input, whereas with binary coding the size of the number n is $\log n$. Thus, for large numbers, assuming binary coding (or coding w.r.t. any base larger than 1) is more realistic.

Numerical constraints have also been used in DLs to formulate cardinality restrictions on concepts (CRs) [5, 30]. For example, the CRs[1]

$$|\mathit{Conference} \sqcap \exists \mathit{uses}.\mathit{Easychair}| \geqslant 75000 \quad \text{and} \quad |\mathit{Person} \sqcap \exists \mathit{uses}.\mathit{Easychair}| \leqslant 3000000$$

state that at least 75 thousand conferences and at most 3 million persons use the conference management system Easychair.[2] Whereas number restrictions are local in the sense that they consider role successors of an individual under consideration (e.g. the papers published by a particular author), CRs are global, i.e., they consider all individuals in an interpretation. Cardinality restrictions can express CIs since, clearly, $C \sqsubseteq D$ is equivalent to $|C \sqcap \neg D| \leqslant 0$. They are, however, considerably more expressive. The higher expressivity of CRs over CIs can, for example, be seen from the fact that CIs in $\mathcal{ALCQ}$ are closed under disjoint union of models, but models of a CR like $|A| \leqslant 1$ are clearly not (see Section 2.3 for more details).

In addition, CRs increase the complexity of reasoning: for the DL $\mathcal{ALCQ}$, consistency w.r.t. CIs is ExpTime-complete [31], but consistency w.r.t. CRs is NExpTime-complete if the numbers occurring in the CRs are assumed to be encoded in binary [30]. With unary coding of numbers, consistency stays ExpTime-complete even w.r.t. CRs [30], but the above example considering 3 million conferences clearly shows that unary coding is not appropriate if numbers with large values are employed. It should be noted that both number restrictions and CRs can be expressed in $\mathcal{C}^2$, the two-variable fragment of first-order logic with counting quantifiers [15, 26], whose satisfiability problem is known to be NExpTime-complete [27].

The logic $\mathcal{C}^2$, and thus also number restrictions and CRs, are expressible in FOL. In contrast, the counting extensions considered in the present paper actually leave the realm of FOL. The classical number restrictions available in $\mathcal{ALCQ}$ can only be used to compare the number of role

[1] Note that the syntax we use here for CRs differs from the one introduced in [5] to make it more similar to the syntax used later on for our extensions of CRs.

[2] See https://www.easychair.org/, last accessed November 24, 2019.

successors of an individual with a *fixed* natural number. They cannot compare the numbers of different kinds of role successors to each other without relating them to a fixed number. This would, e.g., be required to describe persons that have developed more theorem provers than conference management systems, without fixing what these numbers actually are. To overcome this deficit, we have extended $\mathcal{ALCQ}$ by allowing the statement of constraints on role successors that are more general than the number restrictions of $\mathcal{ALCQ}$ [1]. To formulate these constraints, we have used the quantifier-free fragment of Boolean Algebra with Presburger Arithmetic (QFBAPA) [23], in which one can express Boolean combinations of set constraints and numerical constraints comparing the cardinalities of sets. In the resulting logic, called $\mathcal{ALCSCC}$, the above constraint regarding theorem provers and conference managements systems can be expressed using a cardinality constraint on the role *succ*essors:

$$\mathit{Person} \sqcap \mathit{succ}(|\mathit{developed} \cap \mathit{TP}| > |\mathit{developed} \cap \mathit{CMS}|). \tag{1}$$

In general, such a *succ*-expression considers the set of all role successors of a given individual, and requires certain subsets to satisfy the stated QFBAPA constraints. In our example, for a person Andrei to belong to this concept, the cardinality of the set of *developed*-successors of Andrei that belong to the concept TP (collecting all theorem provers) must be larger than the cardinality of the set of *developed*-successors of Andrei that belong to the concept CMS (collecting all conference management systems).

Adding such cardinality constraints strictly extends the expressive power of $\mathcal{ALCQ}$. In [1] it is shown that the constraint $\mathit{succ}(|r| = |s|)$, which describes individuals that have the same number of r-successors as s-successors, cannot be expressed in $\mathcal{ALCQ}$. In [4], the constraint $\mathit{succ}(|r \cap A| = |r \cap \neg A|)$, which describes individuals that have exactly as many r-successors belonging to A as r-successors not belonging to A, is shown to be not even expressible in first-order logic. Intuitively, both kinds of constraints can, e.g., be used to describe people that have the same number of sons and daughters, where in the first constraint one uses roles *son* and *daughter*, whereas in the second one uses the role *child* and the concept *Male*. In spite of this considerable increase of the expressive power, we were able to show in [1] that this does not increase the complexity of reasoning: like for $\mathcal{ALCQ}$, the complexity of the satisfiability problem in $\mathcal{ALCSCC}$ is PSpace-complete without CIs and ExpTime-complete w.r.t. CIs. While the PSpace result also follows from previous work [12] on modal logics with Presburger constraints, the ExpTime result was new.

Just like classical number restrictions, CRs can only relate the cardinality of a concept to a *fixed* number. In [8], we have introduced and investigated more general constraints on the cardinalities of concepts, which we called extended cardinality constraints. The main idea was again to use QFBAPA to formulate and combine these constraints. An example of a constraint expressible this way, but not expressible using CRs is

$$2 \cdot |\mathit{Paper} \sqcap \forall \mathit{topic}.\mathit{DL}| \leqslant |\mathit{Paper} \sqcap \forall \mathit{topic}.\mathit{AR}|$$

which states that papers with topic Automated Reasoning outnumber papers with topic Description Logic by a factor of at least two. In [8] it is shown that, in the DL $\mathcal{ALC}$, the complexity of reasoning w.r.t. extended cardinality constraints (NExpTime for binary coding of numbers) is the same as for reasoning w.r.t. CRs. In addition, the paper introduces a restricted version of this formalism, which can express CIs, but not CRs, and shows that this way the complexity can be lowered to ExpTime. The NExpTime upper bound for the general case actually also follows from the NExpTime upper bound in [32] for a more expressive logic with n-ary relations and function symbols, but the ExpTime result for the restricted case was new.

In [2, 3], we combined the work in [1] and [8] by considering extended cardinality constraints in $\mathcal{ALCSCC}$. This turned out to be non-trivial since the local cardinality constraints of $\mathcal{ALCSCC}$ may interact with the global ones in the extended cardinality constraints. Nevertheless, we were able to show that the complexity results (NExpTime-complete in general, and ExpTime-complete in the restricted case) hold not only for $\mathcal{ALC}$, but also for $\mathcal{ALCSCC}$.

The purpose of the present paper is twofold. On the one hand, we give a compact representation of the known complexity results for the DLs with extended counting facilities mentioned above, and transfer them to a setting where arbitrary rather than just finite models are considered (see below). On the other hand, we investigate the expressive power of these DLs in detail. A first step in this direction was already made in [4], where the expressive power of concept descriptions was examined using appropriate bisimulation relations. Basically, we showed there that $\mathcal{ALCSCC}$ is not expressible in FOL, and determined a sub-logic of $\mathcal{ALCSCC}$, called $\mathcal{ALCCQU}$, that is the first-order fragment of $\mathcal{ALCSCC}$. We also showed that $\mathcal{ALCCQU}$ is more expressive that $\mathcal{ALCQ}$. Here, we recall these results, and then extend them to TBoxes, CRs, and extended cardinality constraints, by adapting methods and ideas from [25]. As in [4], we consider variants of QFBAPA and $\mathcal{ALCSCC}$ that allow for possibly infinite sets and interpretations, respectively. This change has no influence on the complexity of reasoning, but it eases the comparison with classical DLs, for which one usually employs arbitrary models rather than finite ones when defining the semantics. It also adds flexibility since finiteness can be expressed in these logics, and thus one can actually switch between arbitrary model reasoning and finite model reasoning.

2 DLs with counting quantifiers

In this section, we formally introduce the DLs with extended counting facilities mentioned in the introduction, and recall the known complexity results for reasoning in these logics. As mentioned above, we will not restrict the semantics to finite models. For this reason, the results originally obtained for the "finite model" case need to be adapted. We start with introducing the infinite variant of QFBAPA upon which all our logics are based.

2.1 An infinite variant of QFBAPA

We recall the definition of QFBAPA$^\infty$ as introduced in [4].[3] In this logic one can build *set terms* by applying Boolean operations (intersection $\cap$, union $\cup$, and complement $\cdot^c$) to set variables as well as the constants $\emptyset$ and $\mathcal{U}$. Set terms s, t can then be used to state inclusion and equality constraints ($s = t, s \subseteq t$) between sets. For example, if *Vampire* and *Easychair* are set variables, then the set constraint

$$\mathit{Barstool} \cap \mathit{Easychair} = \emptyset$$

says that barstools are not easychairs.

Presburger Arithmetic (PA) expressions are built from numerical variables, integer constants, and set cardinalities $|s|$ using addition as well as multiplication with an integer constant. They can be used to form numerical constraints of the form $k = \ell$ and $k < \ell$, where k, ℓ are PA expressions. For example, the numerical constraint

$$|\mathit{Vampire}| > 10\cdot(|\mathit{Easychair}| + |\mathit{Barstool}|)$$

[3] A variant of QFBAPA$^\infty$ with a slightly different expressivity, but the same complexity of reasoning, was introduced in [22].

says that there are more than ten times as many vampires as there are easychairs and barstools together. A *QFBAPA$^\infty$ formula* is a Boolean combination of set and numerical constraints.

The semantics of set terms and set constraints is defined using *substitutions* σ that assign a set $\sigma(\mathcal{U})$ to $\mathcal{U}$ and subsets of $\sigma(\mathcal{U})$ to set variables.[4] The evaluation of set terms and set constraints by such a substitution is defined in the obvious way, using the standard notions of intersection, union, complement,[5] inclusion, and equality for sets. PA expressions are evaluated over $\mathbb{N}^\infty = \mathbb{N} \cup \{\infty\}$, i.e., the natural numbers[6] extended with a symbol for infinity. Thus, substitutions additionally assign elements of $\mathbb{N}^\infty$ to numerical variables. The cardinality expression $|s|$ is evaluated under σ as the cardinality of $\sigma(s)$ if this set is finite, and as ∞ if $\sigma(s)$ is not finite.[7] When evaluating PA expressions w.r.t. a substitution σ, we employ the usual way of adding, multiplying, and comparing natural numbers, extended by the following rules that deal with infinity: $\infty + N = N + \infty = \infty = \infty + \infty$ for all natural numbers N, $0 \cdot \infty = 0 = \infty \cdot 0$, $N \cdot \infty = \infty = \infty \cdot N$ for all positive integers N, $N < \infty$ and $\infty \not< N$ for all natural numbers N, and $\infty = \infty$ as well as $\infty \not< \infty$.

A *solution* σ of a QFBAPA$^\infty$ formula ϕ is a substitution that evaluates ϕ to true, using the above rules for evaluating set and numerical constraints and the usual interpretation of the Boolean operators occurring in ϕ. The formula ϕ is *satisfiable* if it has a solution.

Note that, in QFBAPA$^\infty$, we can enforce infinity of a set although we do not allow the use of ∞ as a constant. For instance, $|s| = \infty$ is not an admissible numerical constraint, but it is easy to see that the constraint $|s| + 1 = |s|$ can only be satisfied by a substitution that assigns an infinite set to the set term s.

The set constraints in QFBAPA$^\infty$ are actually syntactic sugar since they can be expressed using numerical constraints. In fact, the set constraint $s \subseteq t$ is equivalent to the numerical constraint $|s \cap t^c| \leqslant 0$. Note that, for finite sets, this could equivalently be expressed as $|s \cup t| = |t|$, but for infinite sets the latter constraint is not equivalent to $s \subseteq t$. Since set constraints are syntactic sugar and $>$ and $<$ can easily be simulated in $\mathbb{N}^\infty$ using $\geqslant$ and $\leqslant$, we can assume without loss of generality that any QFBAPA$^\infty$ formula is a Boolean combination of *atomic QFBAPA$^\infty$ formulae* of the form

$$N_0 + N_1|s_1| + \cdots + N_k|s_k| \leqslant M_0 + M_1|t_1| + \cdots + M_\ell|t_\ell|, \tag{2}$$

where the s_i, t_j are set terms and the N_i, M_j are natural numbers.

The logic *CQU* as introduced in [4] is obtained from QFBAPA$^\infty$ by restricting numerical constraints to be of the form $k \leqslant N$ and $k \geqslant N$, i.e., a CQU formula is a Boolean combination of set constraints and numerical constraints of this restricted form. By using the same arguments as above, we can show that any CQU formula is equivalent to a Boolean combinations of atomic formulae of the form (2) where $k = 0$ or $\ell = 0$. In addition, in this setting sums can be expressed using disjunction. For example, saying that $|s| + |t| \leqslant 1$ is equivalent to saying that $|s| \leqslant 0$ and $|t| \leqslant 1$, or $|s| \leqslant 1$ and $|t| \leqslant 0$. Thus, when it comes to expressive power, we can assume without loss of generality that formulae of CQU are Boolean combinations of numerical restrictions of the form $|s| \leqslant N$ or $|s| \geqslant N$.

It is actually not hard to see that the logic CQU as defined here and in [4] has the same expressivity as $\mathcal{C}^1$, the one-variable fragment of first-order logic with counting (see, e.g., [28]). The logic originally called CQU in [14] is the fragment where only conjunctions of atomic

[4] In QFBAPA, $\sigma(\mathcal{U})$ is constrained to be a *finite* set.

[5] The complement is built w.r.t. $\sigma(\mathcal{U})$, i.e., $\sigma(s^c) = \sigma(\mathcal{U}) \setminus \sigma(s)$.

[6] We assume here that the natural numbers contain 0. We call the natural numbers without 0 "positive integers."

[7] Note that we do not distinguish between different infinite cardinalities, such as countably infinite, uncountably infinite, etc.

restrictions of the form $|s| \leqslant N$ or $|s| \geqslant N$ can be used. However, when using CQU within our DLs with counting quantifiers, this difference is irrelevant since the Boolean operations are available anyway on the DL level.

2.2 Concept descriptions that count

We are now ready to define the DL $\mathcal{ALCSCC}^\infty$ and some of its sub-logics. Basically, $\mathcal{ALCSCC}^\infty$ provides us with Boolean operations on concepts and constraints on role successors, which are expressed in QFBAPA$^\infty$. In these constraints, role names and concept descriptions can be used as set variables, and there are no numerical variables allowed.

Definition 1 (Syntax of $\mathcal{ALCSCC}^\infty$)**.** *Given finite, disjoint sets N_C of concept names and N_R of role names, the set of $\mathcal{ALCSCC}^\infty$ concept descriptions over the signature (N_C, N_R) is inductively defined as follows:*

- $\top, \bot$*, and every concept name in N_C is an $\mathcal{ALCSCC}^\infty$ concept description over (N_C, N_R);*
- *if C, D are $\mathcal{ALCSCC}^\infty$ concept descriptions over the signature (N_C, N_R), then so are $C \sqcap D$, $C \sqcup D$, and $\neg C$;*
- *if Con is a set or numerical constraint of QFBAPA$^\infty$ using role names and already defined $\mathcal{ALCSCC}^\infty$ concept descriptions over the signature (N_C, N_R) as set variables, then succ(Con) is an $\mathcal{ALCSCC}^\infty$ concept description over (N_C, N_R).*

For example, the description (1) in the introduction is an $\mathcal{ALCSCC}^\infty$ concept description that uses the QFBAPA$^\infty$ numerical constraint $|\mathit{developed} \cap \mathit{TP}| > |\mathit{developed} \cap \mathit{CMS}|$, in which *developed*, *TP*, and *CMS* are viewed as set variables. Of course, successor constraints can also be nested, as in the $\mathcal{ALCSCC}^\infty$ concept description

$$\mathit{succ}(|\mathit{friend} \cap \mathit{succ}(|\mathit{developed} \cap \mathit{CMS}| \geqslant 1)| = |\mathit{friend} \cap \mathit{succ}(|\mathit{developed} \cap \mathit{TP}| \geqslant 1)|),$$

which describes all individuals having as many friends that have developed at least one conference management system as they have friends that have developed at least one theorem prover.

For the sake of simplicity, we will sometimes use "concept" in place of "concept description," and often dispense with explicitly mentioning the signature. As usual in DL, the semantics of $\mathcal{ALCSCC}^\infty$ is defined using the notion of an interpretation.

Definition 2 (Semantics of $\mathcal{ALCSCC}^\infty$)**.** *Given finite, disjoint sets N_C and N_R of concept and role names, respectively, an* interpretation *of N_C and N_R consists of a non-empty set $\Delta^\mathcal{I}$ and a mapping $\cdot^\mathcal{I}$ that maps every concept name $A \in N_C$ to a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$ and every role name $r \in N_R$ to a binary relation $r^\mathcal{I}$ over $\Delta^\mathcal{I}$. Given an individual $d \in \Delta^\mathcal{I}$ and a role name $r \in N_R$, we define*

$$r^\mathcal{I}(d) := \{e \in \Delta^\mathcal{I} \mid (d, e) \in r^\mathcal{I}\} \quad \textit{and} \quad \mathit{ars}^\mathcal{I}(d) := \bigcup_{r \in N_R} r^\mathcal{I}(d).$$

The function $\cdot^\mathcal{I}$ is inductively extended to $\mathcal{ALCSCC}^\infty$ concept descriptions over (N_C, N_R) by interpreting $\sqcap$, $\sqcup$, and $\neg$ respectively as intersection, union and complement as well as $\top$ as $\Delta^\mathcal{I}$ and $\bot$ as the empty set. Successor constraints are evaluated according to the semantics of QFBAPA$^\infty$: to determine whether $d \in \mathit{succ}(\mathit{Con})^\mathcal{I}$ or not

- *$\mathcal{U}$ is evaluated as $\mathit{ars}^\mathcal{I}(d)$ (i.e., the set of all role successors of d),*

- *$\emptyset$ as the empty set,*
- *roles r occurring in Con as $r^{\mathcal{I}}(d)$ (i.e., the set of r-successors of d),*
- *and concept descriptions D as $D^{\mathcal{I}} \cap ars^{\mathcal{I}}(d)$ (i.e., the set of role successors of d that belong to D). Note that, by induction, the sets $D^{\mathcal{I}}$ are well-defined.*

Then $d \in succ(Con)^{\mathcal{I}}$ iff the substitution obtained this way is a solution of the QFBAPA$^\infty$ formula Con.

The $\mathcal{ALCSCC}^\infty$ concept description C is satisfiable *if there is an interpretation $\mathcal{I}$ such that $C^{\mathcal{I}} \neq \emptyset$. The $\mathcal{ALCSCC}^\infty$ concept descriptions C, D are* equivalent *(written $C \equiv D$) if $C^{\mathcal{I}} = D^{\mathcal{I}}$ for all interpretations $\mathcal{I}$.*

The sub-logics $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, and $\mathcal{ALCCQU}$ of $\mathcal{ALCSCC}^\infty$ can be obtained from $\mathcal{ALCSCC}^\infty$ by restricting the successor constraints appropriately:

- The DL $\mathcal{ALCQ}$ is the fragment of $\mathcal{ALCSCC}^\infty$ in which only successor constraints of the form $succ(|C \cap r| \geqslant N)$ or $succ(|C \cap r| \leqslant N)$ are allowed, where N is a natural number, r is a role name, and C is an $\mathcal{ALCQ}$ concept description. These constraints are usually written as $(\geqslant N\, r.C)$ and $(\leqslant N\, r.C)$, and are called *qualified number restrictions.*
- The DL $\mathcal{ALCQ}t$ is the fragment of $\mathcal{ALCSCC}^\infty$ in which only successor constraints of the form $succ(|C \cap \tau| \geqslant N)$ or $succ(|C \cap \tau| \leqslant N)$ are allowed, where N is a natural number, τ is a safe role type, and C is an $\mathcal{ALCQ}t$ concept description. A *safe role type* is an intersection of role names r (positive occurrence) and complements r^c of role names (negative occurrence) such that every role name in N_R occurs either positively or negatively, and at least one role name occurs positively. Using the syntax for qualified number restrictions, these constraints can be written as $(\geqslant N\, \tau.C)$ and $(\leqslant N\, \tau.C)$.
- The DL $\mathcal{ALCCQU}$ is the fragment of $\mathcal{ALCSCC}^\infty$ in whose successor constraints only constraints of CQU are allowed.

By definition, $\mathcal{ALCQ}$ is a sub-logic of $\mathcal{ALCCQU}$, and $\mathcal{ALCCQU}$ is a sub-logic of $\mathcal{ALCSCC}^\infty$. In addition, $\mathcal{ALCQ}t$ is clearly a sub-logic of $\mathcal{ALCCQU}$. Moreover, it is shown in [4] that any $\mathcal{ALCCQU}$ concept description can be expressed by an equivalent $\mathcal{ALCQ}t$ concept description, and thus that $\mathcal{ALCCQU}$ and $\mathcal{ALCQ}t$ have the same expressive power.

We will see below that $\mathcal{ALCSCC}^\infty$ concept descriptions can in general not be expressed in FOL. In contrast, the concept descriptions of the three fragments introduced above can be expressed by first-order formulae with one free variable. This is well-known for $\mathcal{ALCQ}$ [9], and can be shown for $\mathcal{ALCQ}t$ by a simple adaptation of the first-order translation for $\mathcal{ALCQ}$, where safe role types are translated into first-order formulae with two variables in the obvious way. For $\mathcal{ALCCQU}$ this follows from its equivalence with $\mathcal{ALCQ}t$.

Proposition 3. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}\}$, then $\mathcal{L}$ is a fragment of FOL, i.e., for every $\mathcal{L}$ concept description C there exists a first-order formula with one free variable $C^\sharp(x)$ such that C and $C^\sharp(x)$ are equivalent in the sense that, for every interpretation $\mathcal{I}$, we have $C^{\mathcal{I}} = \{d \in \Delta^{\mathcal{I}} \mid \mathcal{I} \models C^\sharp(d)\}$.*

The logic $\mathcal{ALCSCC}^\infty$ and its sub-logics are *local* in the sense that the decision on whether a certain individual belongs to a concept depends only on this individual and other individuals connected via roles to it. For this reason, evaluating a concept in the disjoint union of interpretations corresponds to evaluating it separately in the single interpretations. To be more precise, given a family $(\mathcal{I}_\nu)_{\nu \in N}$ of interpretations, we define their *disjoint union* $\mathcal{I} = \bigoplus_{\nu \in N} \mathcal{I}_\nu$ as

$$\begin{aligned}
\Delta^{\mathcal{I}} &:= \{(d,\nu) \mid \nu \in N \text{ and } d \in \Delta^{\mathcal{I}_\nu}\},\\
A^{\mathcal{I}} &:= \{(d,\nu) \mid \nu \in N \text{ and } d \in A^{\mathcal{I}_\nu}\} \text{ for all } A \in N_C,\\
r^{\mathcal{I}} &:= \{((d,\nu),(e,\nu)) \mid \nu \in N \text{ and } (d,e) \in r^{\mathcal{I}_\nu}\} \text{ for all } r \in N_R.
\end{aligned}$$

The following is now easy to show, using the locality of $\mathcal{ALCSCC}^\infty$ concept descriptions mentioned above.

Lemma 4. *Let C be an $\mathcal{ALCSCC}^\infty$ concept description. Then we have*

$$C^{\mathcal{I}} = \{(d,\nu) \mid \nu \in N \text{ and } d \in C^{\mathcal{I}_\nu}\}.$$

2.3 TBoxes and cardinality boxes

In classical DLs, terminological knowledge is represented using so-called TBoxes, which are finite sets of CIs of the form $C \sqsubseteq D$ for concepts C, D. Cardinality boxes extend TBoxes by allowing for the formulation of cardinality constraints also on this level. To simplify the comparison with cardinality boxes, in which Boolean combinations of numerical constraints are allowed, we also consider Boolean TBoxes.

Definition 5. *Let $\mathcal{L}$ be one of the DLs $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, $\mathcal{ALCCQU}$, or $\mathcal{ALCSCC}^\infty$.*

1. *A* Boolean $\mathcal{L}$ TBox *is a Boolean combination of CIs $C \sqsubseteq D$, where C, D are $\mathcal{L}$ concept descriptions. An $\mathcal{L}$* TBox *is a conjunction of such CIs.*
2. *A* Boolean $\mathcal{L}$ CBox *is a Boolean combination of CRs of the form $|C| \leqslant N$ and $|C| \geqslant N$, where C is an $\mathcal{L}$ concept description and N is a natural number. An $\mathcal{L}$* CBox *is a conjunction of such CRs.*
3. *An $\mathcal{L}$* ECBox *is a Boolean combination of inequations of the form*

$$N_0 + N_1|C_1| + \cdots + N_k|C_k| \leqslant M_0 + M_1|D_1| + \cdots + M_\ell|D_\ell|, \tag{3}$$

where the C_i, D_j are $\mathcal{L}$ concept descriptions and the N_i, M_j are natural numbers.

4. *An $\mathcal{L}$* RCBox *is a conjunction of inequations of the form*

$$N_1|C_1| + \cdots + N_k|C_k| \leqslant M_1|D_1| + \cdots + M_\ell|D_\ell|, \tag{4}$$

where the C_i, D_j are $\mathcal{L}$ concept descriptions and the N_i, M_j are positive natural numbers.

We say that the interpretation $\mathcal{I}$ is a model *of*

1. *the CI $C \sqsubseteq D$ if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ holds,*
2. *the CR $|C| \leqslant N$ if $|C^{\mathcal{I}}| \leqslant N$, and of the CR $|C| \geqslant N$ if $|C^{\mathcal{I}}| \geqslant N$,*
3. *an inequation of the form (3) if*

$$N_0 + N_1|C_1^{\mathcal{I}}| + \cdots + N_k|C_k^{\mathcal{I}}| \leqslant M_0 + M_1|D_1^{\mathcal{I}}| + \cdots + M_\ell|D_\ell^{\mathcal{I}}|,$$

4. *an inequation of the form* (4) *if*

$$N_1|C_1^{\mathcal{I}}| + \cdots + N_k|C_k^{\mathcal{I}}| \leqslant M_1|D_1^{\mathcal{I}}| + \cdots + M_\ell|D_\ell^{\mathcal{I}}|.$$

The notion of a model is extended to Boolean combinations of such constraints in the obvious way.

Obviously, the CI $C \sqsubseteq D$ can be expressed by the CR $|C \sqcap \neg D| \leqslant 0$. Equivalently, one can express this restriction as $|C \sqcap \neg D| \leqslant |\bot|$, which shows that CIs are also expressible by ECBoxes and RCBoxes. Thus, TBoxes can be expressed using CBoxes, ECBox, or RCBoxes. CBoxes and RCBoxes are clearly expressible by ECBoxes. However, the expressiveness of CBoxes and RCBoxes appears to be orthogonal. While the former only allow us to compare concept cardinalities with a fixed number, this is exactly what is prohibited in RCBoxes. On the other hand, RCBoxes enable us to compare the cardinalities of different concepts whereas this is not possible in CBoxes.

In case the underlying DL $\mathcal{L}$ is expressible in FOL, $\mathcal{L}$ TBoxes and $\mathcal{L}$ CBoxes are clearly also expressible in FOL. Together with Proposition 3 this observation yields the following:

Corollary 6. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}\}$, then $\mathcal{L}$ (Boolean) TBoxes and CBoxes can be expressed in FOL, i.e., for every (Boolean) $\mathcal{L}$ TBox or CBox $\mathcal{T}$ there exists a first-order sentence $\mathcal{T}^\sharp$ such that $\mathcal{T}$ and $\mathcal{T}^\sharp$ are equivalent in the sense that they have the same interpretations as models.*

We can use disjoint unions to show inexpressibility results for some of our box formalisms.

Definition 7. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}, \mathcal{ALCSCC}^\infty\}$ and $\mathcal{T}$ be a (Boolean) TBox, CBox, RCBox, or ECBox. We say that the models of $\mathcal{T}$ are* closed under disjoint union *if the following holds: if the interpretations $\mathcal{I}_\nu$ for $\nu \in N$ are models of $\mathcal{T}$, then their disjoint union $\mathcal{I} = \bigoplus_{\nu \in N} \mathcal{I}_\nu$ is also a model of $\mathcal{T}$. The models of $\mathcal{T}$ are* invariant under disjoint union *if additionally the implication in the other direction holds, i.e., if the disjoint union $\mathcal{I} = \bigoplus_{\nu \in N} \mathcal{I}_\nu$ is a model of $\mathcal{T}$, then so are the interpretations $\mathcal{I}_\nu$ for $\nu \in N$.*

Using Lemma 4, the positive statements of the following proposition are easy to show.

Proposition 8. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}, \mathcal{ALCSCC}^\infty\}$, then*

1. *the models of $\mathcal{L}$ TBoxes are invariant under disjoint union;*
2. *the models of $\mathcal{L}$ RCBoxes are closed under disjoint union, but in general not invariant under disjoint union;*
3. *the models of $\mathcal{L}$ ECBoxes or CBoxes are in general not closed under disjoint union;*
4. *the models of Boolean $\mathcal{L}$ TBoxes are not closed under disjoint union.*

Regarding the negative statement in 2., consider the RCBox $|A| + |B| \leqslant |C|$ for concept names A, B, C. If we consider interpretation $\mathcal{I}_1$ and $\mathcal{I}_2$ in which $A^{\mathcal{I}_1}$ contains one element, $B^{\mathcal{I}_1}$ one element, $C^{\mathcal{I}_1}$ one element, $A^{\mathcal{I}_2}$ one element, $B^{\mathcal{I}_2}$ one element, and $C^{\mathcal{I}_2}$ three elements, then the disjoint union of $\mathcal{I}_1$ and $\mathcal{I}_2$ is a model of the RCBox, but $\mathcal{I}_1$ is not. Regarding 3., it should be clear that the models of $|A| \leqslant 1$ cannot be closed under disjoint union. Finally, it is also easy to see that the models of the Boolean TBox $(A \sqsubseteq \bot) \vee (B \sqsubseteq \bot)$ are not closed under disjoint union.

As an immediate consequence of the above lemma, we obtain the following inexpressibility results.

Proposition 9. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}, \mathcal{ALCSCC}^\infty\}$, then*

- *$\mathcal{L}$ TBoxes in general cannot express $\mathcal{L}$ RCBoxes, ECBoxes, CBoxes, and Boolean TBoxes;*

- *$\mathcal{L}$ RCBoxes in general cannot express $\mathcal{L}$ ECBoxes, CBoxes, and Boolean TBoxes.*

2.4 Reasoning in DLs that count

For a DL $\mathcal{L}$, the fundamental inference problems are satisfiability and subsumption of concepts:

- Given an $\mathcal{L}$ concept C, the *satisfiability problem* asks whether C is satisfiable, i.e., whether there is an interpretation $\mathcal{I}$ such that $C^\mathcal{I} \neq \emptyset$.

- Given $\mathcal{L}$ concepts C, D, the *subsumption problem* asks whether C is a subconcept of D (written $C \sqsubseteq D$), i.e., whether $C^\mathcal{I} \subseteq D^\mathcal{I}$ holds for all interpretations $\mathcal{I}$.

If the DL $\mathcal{L}$ can express $\bot$ as well as conjunction and negation of concepts, then subsumption and satisfiability can be reduced to each other in polynomial time since $C \sqsubseteq D$ holds iff $C \sqcap \neg D$ is unsatisfiable, and C is unsatisfiable iff $C \sqsubseteq \bot$ holds.

These two inference problems can also be considered w.r.t. the kinds of boxes introduced in the previous subsection. Let C, D be $\mathcal{L}$ concepts and $\mathcal{T}$ an $\mathcal{L}$ TBox, CBox, ECBox, or RCBox. Then we say that

- $\mathcal{T}$ is *consistent* if it has a model,

- C is *satisfiable w.r.t.* $\mathcal{T}$ if there is a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^\mathcal{I} \neq \emptyset$,

- C is *subsumed* by D *w.r.t.* $\mathcal{T}$ (written $C \sqsubseteq_\mathcal{T} D$) if $C^\mathcal{I} \subseteq D^\mathcal{I}$ holds for all models $\mathcal{I}$ of $\mathcal{T}$.

Again, it is well-known that these problems can be reduced to each other in polynomial time if $\bot, \top, \neg, \sqcap$, and qualified number restrictions of the form $(\geqslant 1\, r.C)$ are available. In fact, $\mathcal{T}$ is inconsistent iff $\top \sqsubseteq_\mathcal{T} \bot$ holds, and C is satisfiable w.r.t. $\mathcal{T}$ iff $\mathcal{T} \cup \{\top \sqsubseteq (\geqslant 1\, r.C)\}$ is consistent, where r is a new role name occurring neither in C nor in $\mathcal{T}$.

Since the prerequisites required for the reductions mentioned above are satisfied by the DLs $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, $\mathcal{ALCCQU}$, and $\mathcal{ALCSCC}^\infty$, and all our box formalisms can express CIs, we can restrict the attention to the satisfiability problem in case there is no box, and to the consistency problem in case there is a box, when investigating the complexity of reasoning.

Reasoning without a box in $\mathcal{ALCSCC}^\infty$ and its sub-logics

The satisfiability problem in $\mathcal{ALCQ}$ was shown to be PSpace-complete in [29]. In [1] it was proved that this result can be extended to $\mathcal{ALCSCC}$, and in [4] it was demonstrated that the same is true for $\mathcal{ALCSCC}^\infty$. Since we have a PSpace lower bound for $\mathcal{ALCQ}$, which is the least expressive DL considered in this paper, even for unary coding of numbers, as well as a PSpace upper bound for $\mathcal{ALCSCC}^\infty$, which is the most expressive one, even for binary coding, this determines the exact worst-case complexity of the satisfiability problem for all the DLs introduced above.

Theorem 10 ([29, 4])**.** *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}, \mathcal{ALCSCC}^\infty\}$, then satisfiability of $\mathcal{L}$ concepts is PSpace-complete independently of whether numbers are encoded in unary or binary.*

Reasoning w.r.t. CBoxes and ECBoxes

Consistency of $\mathcal{ALCQ}$ CBoxes was shown to be NExpTime-complete in [30] if binary coding of numbers is used, whereas for unary coding it stays in ExpTime. In [2, 3] we were able to prove a NExpTime upper bound for consistency of $\mathcal{ALCSCC}$ ECBoxes with numbers encoded in binary. Basically, the proof of this result takes the $\mathcal{ALCSCC}$ ECBoxes $\mathcal{E}$ and translates it into an exponentially larger QFBAPA formula $\delta_\mathcal{E}$ that is satisfiable iff $\mathcal{E}$ is consistent. Since satisfiability in QFBAPA is NP-complete for binary coding of numbers, this yields the NExpTime upper bound for $\mathcal{ALCSCC}$. This results can easily be transferred to $\mathcal{ALCSCC}^\infty$ by using the same translation, but then testing satisfiability of $\delta_\mathcal{E}$ in QFBAPA$^\infty$ rather than in QFBAPA. In [4] it is shown that the satisfiability problem in QFBAPA$^\infty$ is also in NP.

Theorem 11 ([30, 8, 2, 4])**.** *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}, \mathcal{ALCSCC}^\infty\}$, then consistency of $\mathcal{L}$ (Boolean) CBoxes and ECBoxes is NExpTime-complete if numbers are encoded in binary. For ECBoxes, NExpTime-hardness already holds for unary coding of numbers.*

The reason why the coding of numbers is irrelevant in the presence of ECBoxes is that one can use iterated multiplication to create large numbers from small ones (see [8] for a more detailed argument).

Reasoning w.r.t. TBoxes and RCBoxes

It is well-known that consistency of $\mathcal{ALCQ}$ TBoxes is an ExpTime-complete problem [31]. This result was extended in [1] to $\mathcal{ALCSCC}$ TBoxes, and in [4] it was argued that it also holds for $\mathcal{ALCSCC}^\infty$. RCBoxes were introduced in [8] to obtain a restriction of ECBoxes that lowers the complexity of the consistency problem from NExpTime to ExpTime. For $\mathcal{ALCSCC}$ (i.e., the case of finite models), it is shown in [2, 3] that the consistency problem for RCBoxes is ExpTime-complete. It would not be hard to demonstrate that the approach employed there to prove the ExpTime upper bound for the "finite model" case can be adapted to the infinite case as well. However, below we give a simpler proof of this result for $\mathcal{ALCSCC}^\infty$, which uses that fact that it is sufficient to consider solutions of inequations of the form (5) where the concepts C_i are either empty or have infinite cardinality.

Recall that an $\mathcal{ALCSCC}^\infty$ RCBox $\mathcal{R}$ is a system of inequations of the form

$$N_1|C_1| + \cdots + N_k|C_k| \leqslant N_{k+1}|C_{k+1}| + \cdots + N_{k+\ell}|C_{k+\ell}|, \tag{5}$$

where the C_i are $\mathcal{ALCSCC}^\infty$ concept descriptions and the N_i are positive integers. Our algorithm reduces consistency of $\mathcal{ALCSCC}^\infty$ RCBoxes to consistency of $\mathcal{ALCSCC}^\infty$ TBoxes. It receives an $\mathcal{ALCSCC}^\infty$ RCBox $\mathcal{R}$ as input and initializes the $\mathcal{ALCSCC}^\infty$ TBox $\mathcal{T}$ as $\mathcal{T} := \emptyset$. It then proceeds with the following steps:

1. Check whether the $\mathcal{ALCSCC}^\infty$ TBox $\mathcal{T}$ is consistent. If this is not the case, then terminate with *failure*. Otherwise, for all concepts C occurring in an inequation of $\mathcal{R}$, check whether $\mathcal{T}$ implies $C \sqsubseteq \bot$. If this is the case, then add $C \sqsubseteq \bot$ to $\mathcal{T}$. Then proceed with the next step.

2. For all inequations of the form (5) such that $C_j \sqsubseteq \bot$ belongs to $\mathcal{T}$ for all $k+1 \leqslant j \leqslant k+\ell$, add $C_i \sqsubseteq \bot$ to $\mathcal{T}$ for all $i, 1 \leqslant i \leqslant k$. If no new CI has been added to $\mathcal{T}$, then terminate with *success*. Otherwise, continue with the previous step.

Lemma 12. *The algorithm terminates after a polynomial number of iterations and it succeeds iff the RCBox $\mathcal{R}$ is consistent.*

Consistency of	TBox	RCBox	CBox	ECBox
$\mathcal{ALCQ}$	ExpTime-c.	ExpTime-c.	NExpTime-c.	NExpTime-c.
$\mathcal{ALCQ}t$	ExpTime-c.	ExpTime-c.	NExpTime-c.	NExpTime-c.
$\mathcal{ALCCQU}$	ExpTime-c.	ExpTime-c.	NExpTime-c.	NExpTime-c.
$\mathcal{ALCSCC}^\infty$	ExpTime-c.	ExpTime-c.	NExpTime-c.	NExpTime-c.

Table 1: Complexity results for consistency assuming binary coding of numbers

Proof. Termination after a polynomial number of iterations is an immediate consequence of the fact that only polynomially many CIs of the form $C \sqsubseteq \bot$ can be added to $\mathcal{T}$ since the concepts C for which such a CI can be added must occur in an inequation in $\mathcal{R}$.

Now, assume that $\mathcal{R}$ is consistent, and let $\mathcal{I}$ be a model of $\mathcal{R}$. By an induction on the number of iterations, it is easy to show that we must have $C^\mathcal{I} = \emptyset$ for all CIs added to $\mathcal{T}$ during the run of the algorithm. Consequently, in Step 1 the algorithm can never fail since $\mathcal{I}$ is a model of $\mathcal{T}$. Since the algorithm always terminates, it must thus succeed.

Next, assume that the algorithm succeeds with the final TBox $\mathcal{T}$. Then $\mathcal{T}$ is consistent, and for every concept C occurring in an inequation of $\mathcal{R}$ such that $C \sqsubseteq \bot$ does not belong to $\mathcal{T}$, there is a model $\mathcal{I}_C$ of $\mathcal{T}$ such that $C^{\mathcal{I}_C} \neq \emptyset$. By using closure under disjoint union of models of $\mathcal{ALCSCC}^\infty$ TBoxes, this implies that there is an interpretation $\mathcal{I}_\infty$ such that the following holds for all concepts C occurring in an inequation of $\mathcal{R}$:

- if $C \sqsubseteq \bot$ belongs to $\mathcal{T}$, then $C^{\mathcal{I}_\infty} = \emptyset$;
- if $C \sqsubseteq \bot$ does not belong to $\mathcal{T}$, then the cardinality of $C^{\mathcal{I}_\infty}$ is infinite.

It remains to shows that $\mathcal{I}_\infty$ is a model of $\mathcal{R}$. Thus, consider an inequation of the form (5) in $\mathcal{R}$. If there is a j with $k+1 \leqslant j \leqslant k+\ell$ such that $C_j^{\mathcal{I}_\infty}$ is infinite, then clearly this inequation is satisfied by $\mathcal{I}_\infty$. Otherwise, $C_j \sqsubseteq \bot$ belongs to $\mathcal{T}$ for all $k+1 \leqslant j \leqslant k+\ell$, and thus also $C_i \sqsubseteq \bot$ belongs to $\mathcal{T}$ for all i with $1 \leqslant i \leqslant k$. This shows that, again, the inequation is solved. □

Since consistency of $\mathcal{ALCSCC}^\infty$ TBoxes can be tested in exponential time [4], the overall complexity of our algorithm is ExpTime.

Proposition 13. *Consistency of $\mathcal{ALCSCC}^\infty$ RCBoxes is in ExpTime.*

Combining this result with the known lower bounds for TBox consistency, we thus obtain the following:

Theorem 14. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}, \mathcal{ALCSCC}^\infty\}$, then consistency of $\mathcal{L}$ (Boolean) TBoxes and RCBoxes is ExpTime-complete independently of whether numbers are encoded in unary or binary.*

To explain the ExpTime upper bound for *Boolean* TBoxes, note that one can reduce consistency of a Boolean TBox to exponentially many consistency tests for TBoxes. In fact, one can bring the Boolean TBox into disjunctive normal form and then test every disjunct for consistency. At first sight, such a disjunct is not a TBox since it may contain negated CIs, but one can replace negated CIs $\neg(C \sqsubseteq D)$ with CIs $\top \sqsubseteq (\geqslant 1\, r.(C \sqcap \neg D))$ for new roles r (see [7] for a justification).

The complexity results for "box consistency" in $\mathcal{ALCSCC}^\infty$ and its sub-logics are summarized in Table 1.

3 Expressivity of concept descriptions

The purpose of this section is to compare the expressive power of the concept description languages of the DLs $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, and $\mathcal{ALCSCC}^\infty$. Since we already know that $\mathcal{ALCQ}t$ and $\mathcal{ALCCQU}$ have the same expressiveness, we will not consider $\mathcal{ALCCQU}$ explicitly here. Our results, which have been presented first in [4], make use of appropriate bisimulation relations for the first-order expressible logics $\mathcal{ALCQ}$ and $\mathcal{ALCQ}t$.

Bisimulation relations for $\mathcal{ALCQ}$ and $\mathcal{ALCQ}t$

Let τ be a safe role type, $r_1, \dots, r_k$ the role names occurring positively in τ, and $s_1, \dots, s_\ell$ the role names occurring negatively, i.e., $\tau = r_1 \cap \dots \cap r_k \cap s_1^c \cap \dots \cap s_\ell^c$. For a given interpretation $\mathcal{I}$ and an element $d \in \Delta^\mathcal{I}$, we define

$$\tau^\mathcal{I}(d) := (r_1^\mathcal{I}(d) \cap \dots \cap r_k^\mathcal{I}(d)) \setminus (s_1^\mathcal{I}(d) \cup \dots \cup s_\ell^\mathcal{I}(d)).$$

Since τ is safe, we must have $k \geqslant 1$, and thus $\tau^\mathcal{I}(d) \subseteq r_1^\mathcal{I}(d) \subseteq ars^\mathcal{I}(d)$.

Definition 15 ($\mathcal{ALCQ}t$ bisimulation). *Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be interpretations of N_C and N_R. The relation $\rho \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is an $\mathcal{ALCQ}t$* bisimulation *between $\mathcal{I}_1$ and $\mathcal{I}_2$ if for all $A \in N_C$ and all safe role types τ over N_R the following three properties are satisfied:*

1. *$d_1 \,\rho\, d_2$ implies $d_1 \in A^{\mathcal{I}_1}$ iff $d_2 \in A^{\mathcal{I}_2}$;*
2. *if $d_1 \,\rho\, d_2$ and $D_1 \subseteq \tau^{\mathcal{I}_1}(d_1)$ is finite, then there is a set $D_2 \subseteq \tau^{\mathcal{I}_2}(d_2)$ such that ρ contains a bijection between D_1 and D_2;*
3. *if $d_1 \,\rho\, d_2$ and $D_2 \subseteq \tau^{\mathcal{I}_2}(d_2)$ is finite, then there is a set $D_1 \subseteq \tau^{\mathcal{I}_1}(d_1)$ such that ρ contains a bijection between D_1 and D_2.*

Two individuals $d_1 \in \Delta^{\mathcal{I}_1}$ and $d_2 \in \Delta^{\mathcal{I}_2}$ are called $\mathcal{ALCQ}t$ bisimilar *(written $(\mathcal{I}_1, d_1) \sim_{\mathcal{ALCQ}t} (\mathcal{I}_2, d_2)$) if there is an $\mathcal{ALCQ}t$ bisimulation ρ between $\mathcal{I}_1$ and $\mathcal{I}_2$ such that $d_1 \,\rho\, d_2$. These individuals are called $\mathcal{ALCQ}t$* equivalent *(written $(\mathcal{I}_1, d_1) \equiv_{\mathcal{ALCQ}t} (\mathcal{I}_2, d_2)$) if for all $\mathcal{ALCQ}t$ concept descriptions C we have $d_1 \in C^{\mathcal{I}_1}$ iff $d_2 \in C^{\mathcal{I}_2}$.*

The notion of an $\mathcal{ALCQ}$ bisimulation (called *counting bisimulation* in [25]) is obtained from the above definition by replacing safe role types τ over N_R with role names $r \in N_R$. $\mathcal{ALCQ}$ bisimilarity (written $(\mathcal{I}_1, d_1) \sim_{\mathcal{ALCQ}} (\mathcal{I}_2, d_2)$) and $\mathcal{ALCQ}$ equivalence (written $(\mathcal{I}_1, d_1) \equiv_{\mathcal{ALCQ}} (\mathcal{I}_2, d_2)$) are obtained by replacing $\mathcal{ALCQ}t$ in the above definition with $\mathcal{ALCQ}$. The next proposition states that concepts of $\mathcal{ALCQ}$ and $\mathcal{ALCQ}t$ are invariant under the respective notion of bisimulation. For $\mathcal{ALCQ}$, this was first shown in [25] and for $\mathcal{ALCQ}t$ in [4].

Proposition 16 ([25, 4]). *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$, then $(\mathcal{I}_1, d_1) \sim_\mathcal{L} (\mathcal{I}_2, d_2)$ implies $(\mathcal{I}_1, d_1) \equiv_\mathcal{L} (\mathcal{I}_2, d_2)$.*

This result is already sufficient for showing that $\mathcal{ALCQ}t$ is not expressible in $\mathcal{ALCQ}$.

Corollary 17 ([4]). *Let $N_R = \{r, s\}$ and $N_C = \{A\}$. There is no $\mathcal{ALCQ}$ concept description C such that C is equivalent to the $\mathcal{ALCQ}t$ concept description $succ(|A \cap r \cap s| \geqslant 1)$.*

In fact, if $succ(|A \cap r \cap s| \geqslant 1)$ was equivalent to an $\mathcal{ALCQ}$ concept description, then it would need to be invariant under $\mathcal{ALCQ}$ bisimulation as stated in the above proposition. However,

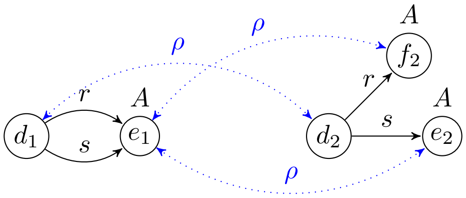

Figure 1: Two interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ and an $\mathcal{ALCQ}$ bisimulation ρ, which is not an $\mathcal{ALCQ}t$ bisimulation.

Fig. 1 shows two interpretations in which the individuals d_1 and d_2 are $\mathcal{ALCQ}$ bisimilar, but whereas d_1 belongs to $succ(|A \cap r \cap s| \geqslant 1)$, the individual d_2 does not.

The following theorem states that $\mathcal{ALCQ}$ and $\mathcal{ALCQ}t$ are exactly the fragments of first-order logic that are invariant under the respective notion of bisimulation. We say that a first-order formula $\phi(x)$ with one free variable x is *invariant under* $\sim_{\mathcal{L}}$ for $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ if $(\mathcal{I}_1, d_1) \sim_{\mathcal{L}} (\mathcal{I}_2, d_2)$ implies that $\mathcal{I}_1 \models \phi(d_1)$ iff $\mathcal{I}_2 \models \phi(d_2)$. For $\mathcal{ALCQ}$ this was first shown in [25] and the proof for $\mathcal{ALCQ}t$ can be obtained by adapting this proof. A detailed proof that closes some small gaps of the one in [25] can be found in the technical report [7].

Theorem 18 ([25, 4])**.** *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and $\phi(x)$ be a first-order formula with one free variable x. Then the following are equivalent:*

1. *there is an $\mathcal{L}$ concept description C such that C is equivalent to $\phi(x)$;*
2. *$\phi(x)$ is invariant under $\sim_{\mathcal{L}}$.*

Comparison with $\mathcal{ALCSCC}^\infty$

One might think that invariance of $\mathcal{ALCQ}t$ concept descriptions under $\mathcal{ALCQ}t$ bisimulation could be used to show that $\mathcal{ALCSCC}^\infty$ concepts cannot be expressed in $\mathcal{ALCQ}t$. This is, however, not the case since $\mathcal{ALCSCC}^\infty$ concepts are also invariant under $\mathcal{ALCQ}t$ bisimulation.

Proposition 19 ([4])**.** *If $(\mathcal{I}_1, d_1) \sim_{\mathcal{ALCQ}t} (\mathcal{I}_2, d_2)$ then $(\mathcal{I}_1, d_1) \equiv_{\mathcal{ALCSCC}^\infty} (\mathcal{I}_2, d_2)$.*

Here $\mathcal{ALCSCC}^\infty$ equivalence is defined in the obvious way, by considering all $\mathcal{ALCSCC}^\infty$ concept descriptions over the given signature. The main idea underlying the proof of this proposition is that all the PA expressions occurring in successor constraints can be transformed into the form

$$k = \sum_{i=1}^{\ell} N_i \cdot |\tau_i \cap C_i|,$$

where the N_i are natural numbers, the τ_i are safe role types, and the C_i are $\mathcal{ALCSCC}^\infty$ concept descriptions. Then, one can show that, for individuals that are $\mathcal{ALCQ}t$ bisimilar, expressions of the form $|\tau_i \cap C_i|$ evaluate to the same number or to ∞ on their role successors.

Combining Proposition 19 and Theorem 18 for $\mathcal{L} = \mathcal{ALCQ}t$, we can now conclude that $\mathcal{ALCQ}t$ is exactly the first order fragment of $\mathcal{ALCSCC}^\infty$.

Theorem 20 ([4])**.** *For an $\mathcal{ALCSCC}^\infty$ concept description C, the following are equivalent:*

1. *C is equivalent to an FOL formula with one free variable;*

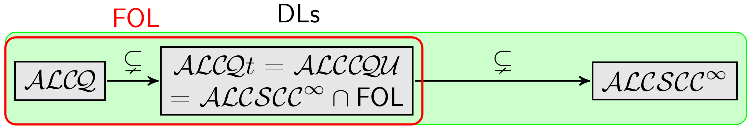
FOL
DLs
$\mathcal{ALCQ}$
$\subsetneq$
$\mathcal{ALCQ}t = \mathcal{ALCCQU}$
$= \mathcal{ALCSCC}^\infty \cap \mathsf{FOL}$
$\subsetneq$
$\mathcal{ALCSCC}^\infty$

Figure 2: The relative expressivity of the DLs $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, $\mathcal{ALCCQU}$, and $\mathcal{ALCSCC}^\infty$.

2. C is equivalent to an $\mathcal{ALCQ}t$ concept description.

The direction ($2 \Rightarrow 1$) is an immediate consequence of Proposition 3. For the other direction, assume that C is equivalent to the FOL formula $\phi(x)$. Then $\phi(x)$ is invariant under $\mathcal{ALCQ}t$ bisimulation by Proposition 19, and thus equivalent to an $\mathcal{ALCQ}t$ concept description by Theorem 18.

It remains to show that $\mathcal{ALCSCC}^\infty$ is more expressive than $\mathcal{ALCQ}t$. Note that, by Theorem 20, any $\mathcal{ALCSCC}^\infty$ concept that is not expressible in $\mathcal{ALCQ}t$ is also not expressible in FOL. The following proposition, which was first stated in [4], is an easy consequence of Proposition 30 in Section 4.1.

Proposition 21 ([4]). *The $\mathcal{ALCSCC}^\infty$ concept description $succ(|r \cap A| = |r \cap \neg A|)$ cannot be expressed in first-order logic.*

Fig. 2 summarizes the results obtained in this section.

4 Expressivity of boxes

Here we extend the bisimulation characterizations of the previous section to the box formalisms introduced in Section 2.3. For (Boolean) TBoxes and the DL $\mathcal{ALCQ}$, this was already done in [25]. First, we recall these results and extend them to $\mathcal{ALCQ}t$. As a consequence, we also obtain characterizations of the first-order fragments of $\mathcal{ALCSCC}^\infty$ TBoxes and Boolean TBoxes. Second, we show similar results for CBoxes and ECBoxes.

4.1 TBoxes and Boolean TBoxes in $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, and $\mathcal{ALCSCC}^\infty$

In order to deal with CIs, which make global statements about all individuals of an interpretation, we need to "globalize" the notion of a bisimulation [25].

Definition 22. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and $\mathcal{I}_1, \mathcal{I}_2$ be interpretations.*

- *The $\mathcal{L}$ bisimulation ρ between $\mathcal{I}_1$ and $\mathcal{I}_2$ is* global *if for every $d \in \Delta^{\mathcal{I}_1}$ there exists $e \in \Delta^{\mathcal{I}_2}$ such that $(d, e) \in \rho$ (and vice versa).*
- *The interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ are* globally $\mathcal{L}$ bisimilar *(written $\mathcal{I}_1 \sim^g_\mathcal{L} \mathcal{I}_2$) if there is a global $\mathcal{L}$ bisimulation ρ between $\mathcal{I}_1$ and $\mathcal{I}_2$.*
- *The interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ are* globally $\mathcal{L}$ equivalent *(written $\mathcal{I}_1 \equiv^g_\mathcal{L} \mathcal{I}_2$) if for every CI $C \sqsubseteq D$ with C and D $\mathcal{L}$ concept descriptions we have that $\mathcal{I}_1 \models C \sqsubseteq D$ iff $\mathcal{I}_2 \models C \sqsubseteq D$.*
- *The first-order sentence ϕ is* invariant under global $\mathcal{L}$ bisimulation *if $\mathcal{I}_1 \models \phi$ and $\mathcal{I}_1 \sim^g_\mathcal{L} \mathcal{I}_2$ imply $\mathcal{I}_2 \models \phi$.*

The following proposition is an easy consequence of the above definition and Proposition 16.

Proposition 23. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$, then $\mathcal{I}_1 \sim^g_{\mathcal{L}} \mathcal{I}_2$ implies $\mathcal{I}_1 \equiv^g_{\mathcal{L}} \mathcal{I}_2$.*

As an immediate consequence, we obtain invariance of (Boolean) $\mathcal{L}$ TBoxes (viewed as first-order sentences) under global $\mathcal{L}$ bisimulation.

Corollary 24. *Every (Boolean) $\mathcal{L}$ TBox (for $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$) is invariant under global $\mathcal{L}$ bisimulation.*

This result can be used to show that Boolean $\mathcal{ALCQ}$ TBoxes cannot express $\mathcal{ALCQ}t$ TBoxes.

Corollary 25. *Let $N_R = \{r, s\}$ and $N_C = \{A, B\}$. There is no Boolean $\mathcal{ALCQ}$ TBox that is equivalent to the $\mathcal{ALCQ}t$ TBox $\mathcal{T} = \{B \sqsubseteq succ(|A \cap r \cap s| \geqslant 1)\}$.*

To prove this corollary, we can basically reuse the interpretations $\mathcal{I}_1$, $\mathcal{I}_2$ and the $\mathcal{ALCQ}$ bisimulation ρ shown in Fig. 1, but where now additionally d_1, d_2 belong to the concept B, whereas the other elements do not belong to B. Then ρ is a global $\mathcal{ALCQ}$ bisimulation between $\mathcal{I}_1$ and $\mathcal{I}_2$. However, $\mathcal{I}_1$ is a model of $\mathcal{T}$, whereas $\mathcal{I}_2$ is not, which shows that $\mathcal{T}$ cannot be equivalent to a Boolean $\mathcal{ALCQ}$ TBox by Corollary 24.

Global $\mathcal{L}$ bisimulations can also be used to characterize the first-order sentences that are equivalent to Boolean $\mathcal{L}$ TBoxes. For $\mathcal{L} = \mathcal{ALCQ}$, this was already shown in [25]. A detailed proof for $\mathcal{L} = \mathcal{ALCQ}t$ can be found in [7].

Theorem 26 ([25, 7])**.** *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and ϕ be a first-order sentence. Then the following are equivalent:*

1. *There exists a Boolean $\mathcal{L}$ TBox $\mathcal{T}$ such that $\mathcal{T}$ is equivalent to ϕ.*
2. *The sentence ϕ is invariant under global $\mathcal{L}$ bisimulation.*

To distinguish TBoxes from Boolean TBoxes, one needs to use the fact that TBoxes are invariant under disjoint union, whereas Boolean TBoxes are not (see Proposition 8).

Theorem 27. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and ϕ be a first-order sentence. Then the following are equivalent:*

1. *There exists an $\mathcal{L}$ TBox $\mathcal{T}$ such that $\mathcal{T} \equiv \phi$.*
2. *The sentence ϕ is invariant under global $\mathcal{L}$ bisimulation and under disjoint unions.*

For $\mathcal{L} = \mathcal{ALCQ}$, this theorem was shown in [25] (see proof of Theorem 7 in [25]), and the adaptation of this proof to the case $\mathcal{L} = \mathcal{ALCQ}t$ is simple.

Using the fact that $\mathcal{ALCSCC}^\infty$ concept descriptions are invariant under $\mathcal{ALCQ}t$ bisimulation (see Proposition 19 above), it is easy to see that Proposition 23 can be extended to $\mathcal{ALCSCC}^\infty$ as follows.

Corollary 28. *$\mathcal{I}_1 \sim^g_{\mathcal{ALCQ}t} \mathcal{I}_2$ implies $\mathcal{I}_1 \equiv^g_{\mathcal{ALCSCC}^\infty} \mathcal{I}_2$.*

Combining this result with Theorems 26 and 27 for the case $\mathcal{L} = \mathcal{ALCQ}t$, we thus obtain the following characterizations of the first order fragments of (Boolean) $\mathcal{ALCSCC}^\infty$ TBoxes.

Theorem 29. *Let $\mathcal{T}$ be a (Boolean) $\mathcal{ALCSCC}^\infty$ TBox. Then the following are equivalent:*

1. *$\mathcal{T}$ is equivalent to a first-order sentence.*
2. *$\mathcal{T}$ is equivalent to a (Boolean) $\mathcal{ALCQ}t$ TBox.*

It remains to show that there are indeed $\mathcal{ALCSCC}^\infty$ TBoxes that cannot be expressed by a Boolean $\mathcal{ALCQ}t$ TBox, and thus are not expressible in FOL.

Proposition 30. *The $\mathcal{ALCSCC}^\infty$ TBox $\mathcal{T}_\star := \{\top \sqsubseteq succ(|r \cap A| = |r \cap \neg A|)\}$ cannot be expressed in first-order logic.*

Proof. It is sufficient to show that $\mathcal{T}_\star$ cannot be expressed as an equivalent $\mathcal{ALCQ}t$ TBox $\mathcal{T}$. Together with Theorem 29, this yields our statement. We fix $(N_C, N_R) := (\{A\}, \{r\})$ and assume by contradiction that such $\mathcal{T}$ exists over this signature. Note that, in this restricted signature, the only safe role type is the role r itself, and thus successor constraints are in fact qualified number restrictions for the role r.

Due to the semantic equivalence $(\leqslant K\, r.D) \equiv \neg(\geqslant (K+1)\, r.D)$, we can assume that every qualified number restriction occurring in $\mathcal{T}$ is of the form $(\geqslant K\, r.D)$ with K a natural number and D an $\mathcal{ALCQ}t$ concept description. Let N' be the largest natural number appearing in a qualified number restriction in $\mathcal{T}$. Then, we define $N := N' + 1$ and the sets $S_1 := \{1, \ldots, N\}$ and $S_2 := \{N+1, \ldots, 2N\}$.

The interpretation $\mathcal{I}$ over $(\{A\}, \{r\})$ of domain $\Delta^\mathcal{I} = \{0, 1, \ldots, 2N\}$ is defined by setting all elements of S_1 and S_2 as r-successors of 0 and $A^\mathcal{I} := S_1$. Then $\mathcal{I}$ is clearly a model of $\mathcal{T}_\star$, and hence of $\mathcal{T}$. We extend $\mathcal{I}$ to $\mathcal{I}'$ by adding $2N+1$ to the domain and to the interpretation of A, and connecting 0 with $2N+1$ via the role r. We show that $\mathcal{I}'$ is a model of $\mathcal{T}$, by proving the following facts:

1. For all $i \in \{1, \ldots, N\}$, $j \in \{1, \ldots, N, 2N+1\}$, and $\mathcal{ALCQ}t$ concepts D we have $i \in D^\mathcal{I}$ iff $j \in D^{\mathcal{I}'}$.
2. For all $i, j \in \{N+1, \ldots, 2N\}$, and $\mathcal{ALCQ}t$ concepts D we have $i \in D^\mathcal{I}$ iff $j \in D^{\mathcal{I}'}$.
3. For all $\mathcal{ALCQ}t$ concepts D containing only numbers $< N$ we have $0 \in D^\mathcal{I}$ iff $0 \in D^{\mathcal{I}'}$.

Before showing these facts, first note that they indeed imply that $\mathcal{I}'$ is a model of $\mathcal{T}$. In fact, assume that this is not the case. Then there is a CI $C_1 \sqsubseteq C_2$ in $\mathcal{T}$ such that $D := C_1 \sqcap \neg C_2$ is non-empty in $\mathcal{I}'$. Assume that $j \in D^{\mathcal{I}'}$. If $j \in \{1, \ldots, N, 2N+1\}$, then (1) implies that $1 \in D^\mathcal{I}$, contradicting the fact that $\mathcal{I}$ is a model of $\mathcal{T}$. Similarly, we can show that the case where $j \in \{N+1, \ldots, 2N\}$ leads to a contradiction. Finally, if $j = 0$, then we have $0 \in D^\mathcal{I}$ since D satisfies the restriction stated in (3). Again, this leads to a contradiction.

To show the three facts, first note that every element of $\{1, \ldots, N\}$ in $\mathcal{I}$ is $\mathcal{ALCQ}t$ bisimilar to every element of $\{1, \ldots, N, 2N+1\}$ in $\mathcal{I}'$. Similarly, every element of $\{N+1, \ldots, 2N\}$ in $\mathcal{I}$ is $\mathcal{ALCQ}t$ bisimilar to every element of $\{N+1, \ldots, 2N\}$ in $\mathcal{I}'$. By Proposition 16, this yields the facts (1) and (2).

We show (3) by induction on the structure of D. The only interesting case is the one where D is of the form $D = (\geqslant K\, r.E)$ for an $\mathcal{ALCQ}t$ concept E and a number $K < N$. We observe that all the elements of S_1 are pairwise $\mathcal{ALCQ}t$ bisimilar in $\mathcal{I}$, and the same is true for the elements of S_2. Combining this observation with Proposition 16, we obtain that, for the $\mathcal{ALCQ}t$ concept E, at least one of the following holds: (a) $S_1 \subseteq E^\mathcal{I}$, (b) $S_2 \subseteq E^\mathcal{I}$, (c) $S_1 \cup S_2 \subseteq (\neg E)^\mathcal{I}$.

If (a) holds, then $|S_1| = N > K$ yields $0 \in (\geqslant K\, r.E)^\mathcal{I}$. Due to the $\mathcal{ALCQ}t$ bisimilarity relationships between elements of $\mathcal{I}$ and $\mathcal{I}'$ mentioned above, Proposition 16 yields $S_1 \subseteq E^{\mathcal{I}'}$, and thus $0 \in (\geqslant K\, r.E)^{\mathcal{I}'}$ holds as well. The case (b) can be treated similarly. Finally, assume that (c) holds. The case $K = 0$ is trivial since then $(\geqslant 0\, r.E) \equiv \top$. If $K \geqslant 1$, then $0 \notin (\geqslant K\, r.E)^\mathcal{I}$ since none of the r-successors of 0 in $\mathcal{I}$ belong to E. From (1) and (2) we obtain that $\{1, \ldots, 2N+1\} \subseteq (\neg E)^{\mathcal{I}'}$, and can conclude that $0 \notin (\geqslant K\, r.E)^{\mathcal{I}'}$. This concludes the proof of (3).

Summing up, we have seen that both $\mathcal{I}$ and $\mathcal{I}'$ are models of $\mathcal{T}$. However, this contradicts our assumption that $\mathcal{T}$ is equivalent to $\mathcal{T}_\star$ since actually $\mathcal{I}$ is a model of $\mathcal{T}_\star$, but $\mathcal{I}'$ is not. □

Note that this proposition also implies Proposition 21. In fact, if $succ(|r \cap A| = |r \cap \neg A|)$ was expressible in FOL, then there would exist an $\mathcal{ALCQ}t$ concept C such that $C \equiv succ(|r \cap A| = |r \cap \neg A|)$. But then the TBox $\mathcal{T}_\star$ would be equivalent to the $\mathcal{ALCQ}t$ TBox $\{\top \sqsubseteq C\}$, which is expressible in FOL by Corollary 6.

4.2 Boolean CBoxes and ECBoxes in $\mathcal{ALCQ}$, $\mathcal{ALCQ}t$, and $\mathcal{ALCSCC}^\infty$

In order to deal with CRs rather than CIs, we need to extend our notion of a global bisimulation to one that can also compare cardinalities of sets on the global level. The following definition is inspired by the *first-order counting games* used in [16] to analyze extensions of first-order logic by certain counting quantifiers.

Definition 31. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and $\mathcal{I}_1$, $\mathcal{I}_2$ be interpretations.*

- *The $\mathcal{L}$ bisimulation $\rho \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is a* comparative $\mathcal{L}$ bisimulation *between $\mathcal{I}_1$ and $\mathcal{I}_2$ if ρ satisfies the following two properties:*
  1. *if $D_1 \subseteq \Delta^{\mathcal{I}_1}$ is finite, then there is a set $D_2 \subseteq \Delta^{\mathcal{I}_2}$ such that ρ contains a bijection between D_1 and D_2;*
  2. *if $D_2 \subseteq \Delta^{\mathcal{I}_2}$ is finite, then there is a set $D_1 \subseteq \Delta^{\mathcal{I}_1}$ such that ρ contains a bijection between D_1 and D_2.*
- *The interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ are* comparatively $\mathcal{L}$ bisimilar *(written $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$) if there is a comparative $\mathcal{L}$ bisimulation ρ between $\mathcal{I}_1$ and $\mathcal{I}_2$.*
- *The interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ are* comparatively $\mathcal{L}$ equivalent *(written $\mathcal{I}_1 \equiv_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$) if for all CRs $|C| \lesseqgtr N$ (with C an $\mathcal{L}$ concept, N a natural number, and $\lesseqgtr\, \in \{\leqslant, \geqslant\}$) we have $\mathcal{I}_1 \models |C| \lesseqgtr N$ iff $\mathcal{I}_2 \models |C| \lesseqgtr N$.*
- *The first-order sentence ϕ is* invariant under comparative $\mathcal{L}$ bisimulation *if $\mathcal{I}_1 \models \phi$ and $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$ imply $\mathcal{I}_2 \models \phi$.*

The following proposition states that CRs are indeed invariant under comparative bisimulation.

Proposition 32. *If $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$, then $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$ implies $\mathcal{I}_1 \equiv_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$.*

Proof. Assume that ρ is a comparative $\mathcal{L}$ bisimulation between the interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$. Let $|C| \geqslant N$ be a CR with C an $\mathcal{L}$ concept and N a natural number such that $|C^{\mathcal{I}_1}| \geqslant N$. Then $C^{\mathcal{I}_1}$ contains distinct elements $d_1, \ldots, d_N$, and the fact that ρ is a comparative bisimulation implies that there exist distinct elements $e_1, \ldots, e_N \in \Delta^{\mathcal{I}_2}$ such that $(d_i, e_i) \in \rho$ for $1 \leqslant i \leqslant N$. Thanks to Proposition 16, it follows that $e_1, \ldots, e_N \in C^{\mathcal{I}_2}$, and thus $|C^{\mathcal{I}_2}| \geqslant N$. We can prove that $|C^{\mathcal{I}_2}| \geqslant N$ implies $|C^{\mathcal{I}_1}| \geqslant N$ using an analogous argument. The case for CRs of the form $|C| \leqslant N$ follows from the semantic equivalence of $|C| \leqslant N$ and $\neg(|C| \geqslant N + 1)$. □

This proposition obviously implies that (Boolean) CBoxes are invariant under comparative bisimulation. We show next that this is true even for ECBoxes (which subsumes the case of (Boolean) CBoxes).

Corollary 33. *If $\mathcal{E}$ is an $\mathcal{L}$ ECBox with $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$ then $\mathcal{I}_1 \models \mathcal{E}$ iff $\mathcal{I}_2 \models \mathcal{E}$.*

Proof. First, we show that $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$ implies that $|C^{\mathcal{I}_1}| = |C^{\mathcal{I}_2}|$ holds for all $\mathcal{L}$ concepts C. In fact, if $|C^{\mathcal{I}_1}| = N$ is finite, then $\mathcal{I}_1$ satisfies the CRs $|C| \leqslant N$ and $|C| \geqslant N$. By Proposition 32, $\mathcal{I}_2$ must then satisfy these CRs as well, which shows that $|C^{\mathcal{I}_1}| = N = |C^{\mathcal{I}_2}|$. If $|C^{\mathcal{I}_1}|$ is infinite, then $\mathcal{I}_1$ satisfies the CRs $|C| \geqslant N$ for all $N \geqslant 0$, and by Proposition 32, $\mathcal{I}_2$ must satisfy all these CRs as well, which shows that $|C^{\mathcal{I}_2}|$ is also infinite.

Every extended cardinality constraint occurring in $\mathcal{E}$ is of the form

$$N_0 + N_1|C_1| + \cdots + N_k|C_k| \leqslant M_0 + M_1|D_1| + \cdots + M_\ell|D_\ell|$$

with N_i, M_j natural numbers and C_i, D_j $\mathcal{L}$ concepts. Since $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$ implies that $|C_i^{\mathcal{I}_1}| = |C_i^{\mathcal{I}_2}|$ and $|D_j^{\mathcal{I}_1}| = |D_j^{\mathcal{I}_2}|$ hold for $1 \leqslant i \leqslant k$ and $1 \leqslant j \leqslant \ell$ (as just shown), every cardinality constraint occurring in $\mathcal{E}$ is evaluated in the same way in $\mathcal{I}_1$ and $\mathcal{I}_2$. □

Next, we want to show that Boolean $\mathcal{L}$ CBoxes are exactly the first-order sentences that are invariant under comparative $\mathcal{L}$ bisimulation. In contrast to the Sections 3 and 4.1, where we have stated the corresponding results for concept descriptions and TBoxes (see Theorems 18 and 27) without proofs, here we will give a detailed proof. In fact, while the results for concept descriptions and TBoxes have been published before (in [25] for $\mathcal{ALCQ}$ and in [4] for $\mathcal{ALCQ}t$), the results for CBoxes are published for the first time in the present paper. Note that the proofs of Theorems 18 and 27 have a structure that is very similar to the proof given below.

The first step is to show that the converse of Proposition 32 holds as well if we restrict the statement to so-called ω-saturated interpretations [25, 11]. When defining ω-saturated interpretations, one assumes that every domain element of an interpretation $\mathcal{I}$ can be used as a constant symbol in formulae, where $d \in \Delta^{\mathcal{I}}$ interprets itself, i.e., $d^{\mathcal{I}} := d$. Let $\mathcal{I}$ be an interpretation of N_C and N_R. A (possibly infinite) set of first-order formulae Γ with free variables from a finite set $\{x_1, \ldots, x_n\}$, predicate symbols from $N_C \cup N_R$, and constant symbols from a finite subset of $\Delta^{\mathcal{I}}$ is called

- *realizable* in $\mathcal{I}$ if there is a variable assignment $a\colon \{x_1, \ldots, x_n\} \to \Delta^{\mathcal{I}}$ such that $\mathcal{I} \models \phi(a(x_{i_1}), \ldots, a(x_{i_k}))$ for every formula $\phi(x_{i_1}, \ldots, x_{i_k}) \in \Gamma$;
- *finitely realizable* in $\mathcal{I}$ if every finite subset Γ' of Γ is realizable in $\mathcal{I}$.

The interpretation $\mathcal{I}$ is *ω-saturated* if, for every such set Γ, finite realizability in $\mathcal{I}$ implies realizability in $\mathcal{I}$.

The following result from [11] implies that, though not every interpretations $\mathcal{I}$ is ω-saturated, one may without loss of generality assume that one has such an interpretation if one is only interested in the FOL sentences that the interpretation satisfies.

Theorem 34. *For every interpretation $\mathcal{I}$ there exists an ω-saturated interpretation $\mathcal{I}^\star$ that satisfies the same first-order sentences as $\mathcal{I}$.*

A further result that we will need in our proof of the converse of Proposition 32 is Hall's theorem [17]. Given a finite family $F = (S_1, \ldots, S_N)$ of sets, we say that F has a system of distinct representatives (SDR) if there are N *distinct* elements $s_1, \ldots, s_N$ such that $s_i \in S_i$ for $i = 1, \ldots, N$.

Theorem 35 (Hall)**.** *The family $F = (S_1, \ldots, S_N)$ has a system of distinct representatives iff for all index sets $I \subseteq \{1, \ldots, N\}$ we have $\left|\bigcup_{i \in I} S_i\right| \geqslant |I|$.*

The following lemma is an immediate consequence of Hall's theorem. It shows that the existence of an SDR can be characterized using a CBox.

Lemma 36. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$, $\mathcal{I}$ an interpretation, and $C_1, \ldots, C_N$ $\mathcal{L}$ concepts. Then the family $(C_1^{\mathcal{I}}, \ldots, C_N^{\mathcal{I}})$ has an SDR iff $\mathcal{I} \models \mathcal{C}_\star$ where $\mathcal{C}_\star$ is the CBox that is defined to be the conjunction of the following CRs:*

$$|C_{i_1} \sqcup \cdots \sqcup C_{i_k}| \geqslant k$$

where $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, N\}$ and $|\{i_1, \ldots, i_k\}| = k$.

Proposition 37. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and $\mathcal{I}_1$, $\mathcal{I}_2$ be ω-saturated interpretations. Then $\mathcal{I}_1 \equiv_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$ implies $\mathcal{I}_1 \sim_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$.*

Proof. Let $\mathcal{I}_1$, $\mathcal{I}_2$ be ω-saturated interpretations such that $\mathcal{I}_1 \equiv_{\mathcal{L}}^{\bowtie} \mathcal{I}_2$. To demonstrate that these two interpretations are also comparatively $\mathcal{L}$ bisimilar, it is sufficient to prove that the binary relation

$$Eq_{\mathcal{L}} := \{(d, e) \in \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2} \mid (\mathcal{I}_1, d) \equiv_{\mathcal{L}} (\mathcal{I}_2, e)\}$$

is a comparative $\mathcal{L}$ bisimulation between $\mathcal{I}_1$ and $\mathcal{I}_2$.

The fact that $Eq_{\mathcal{L}}$ is an $\mathcal{L}$ bisimulation between $\mathcal{I}_1$ and $\mathcal{I}_2$ is actually also needed in the proofs of Theorems 18 and 27. For the proofs of this fact, we thus refer the reader to [25] for the case $\mathcal{L} = \mathcal{ALCQ}$ and to [4, 7] for the case $\mathcal{L} = \mathcal{ALCQ}t$. Here, we concentrate on showing that condition (1) in Definition 31 is satisfied since condition (2) can be shown analogously.

Thus, let $d_1, \ldots, d_N$ be distinct individuals in $\Delta^{\mathcal{I}_1}$. To find distinct individuals $e_1, \ldots, e_N$ in $\Delta^{\mathcal{I}_2}$ such that $(d_i, e_i) \in Eq_{\mathcal{L}}$ for $1 \leqslant i \leqslant N$, we resort to the fact that $\mathcal{I}_2$ is ω-saturated. In particular, we define the set $\Gamma := \Gamma^{\neq} \cup \bigcup_{i=1}^{N} \Theta_i$ of first-order formulae, where

$$\begin{aligned}\Gamma^{\neq} &:= \{\textstyle\bigwedge_{1 \leqslant i < j \leqslant N} x_i \neq x_j\},\\ \Theta_i &:= \{C^\sharp(x_i) \mid C \text{ is an } \mathcal{L} \text{ concept and } d_i \in C^{\mathcal{I}_1}\}.\end{aligned}$$

Clearly, the variable assignment $a(x_i) := d_i$ for $1 \leqslant i \leqslant N$ shows that Γ is realizable in $\mathcal{I}_1$. If we could show that Γ is also realizable in $\mathcal{I}_2$ with variable assignment b, then setting $e_i := b(x_i)$ would clearly yield the distinct individuals $e_1, \ldots, e_N$ we are looking for.

Since $\mathcal{I}_2$ is ω-saturated, it is sufficient to show that each finite subset Γ' of Γ is realizable in $\mathcal{I}_2$. Without loss of generality, we can assume that Γ' contains $\Gamma^{\neq}$ since this set is finite. For $i = 1, \ldots, N$, we introduce the $\mathcal{L}$ concept descriptions

$$C_i := \bigsqcap\{C \mid C^\sharp(x_i) \in \Gamma' \cap \Theta_i\},$$

which are well-defined since Γ' is finite. Note that the first-order formula

$$\bigwedge_{i=1}^{N} C_i^\sharp(x_i) \wedge \bigwedge_{1 \leqslant i < j \leqslant N} x_i \neq x_j$$

is satisfied in $\mathcal{I}_1$ under the variable assignment $a(x_i) := d_i$. This shows that $d_1, \ldots, d_N$ is an SDR for the the family of sets $(C_1^{\mathcal{I}_1} \ldots, C_N^{\mathcal{I}_1})$. Then, by Lemma 36, we obtain that $\mathcal{I}_1$ is a model of $\mathcal{C}_\star$, where $\mathcal{C}_\star$ is the CBox defined in this lemma. Since $\mathcal{I}_1$ and $\mathcal{I}_2$ are comparatively $\mathcal{L}$ equivalent, $\mathcal{I}_2$ is also a model of $\mathcal{C}_\star$. By Lemma 36, this implies that the family of sets $(C_1^{\mathcal{I}_2} \ldots, C_N^{\mathcal{I}_2})$ also has an SDR, say $e_1, \ldots, e_N$. Clearly, setting $b(x_i) := e_i$ then shows that Γ' is realizable in $\mathcal{I}_2$. □

We are now ready to prove the main theorem of this subsection.

Theorem 38. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and ϕ be a first-order sentence. Then the following are equivalent:*

1. *There exists a Boolean $\mathcal{L}$ CBox $\mathcal{C}$ such that $\mathcal{C} \equiv \phi$.*
2. *The sentence ϕ is invariant under comparative $\mathcal{L}$ bisimulation.*

Proof. The direction ($1 \Rightarrow 2$) is a direct consequence of Corollary 33 since Boolean $\mathcal{L}$ CBoxes are a special case of $\mathcal{L}$ ECBoxes.

Let $\mathrm{Cons}(\phi)$ denote the set of Boolean $\mathcal{L}$ CBoxes entailed by the first-order sentence ϕ. We prove ($2 \Rightarrow 1$) by showing that (2) implies $\mathrm{Cons}(\phi) \models \phi$. In fact, if this is the case, then compactness of first-order logic yields a finite set of Boolean $\mathcal{L}$ CBoxes $\Gamma \subseteq \mathrm{Cons}(\phi)$ entailing ϕ. But then the conjunction $\mathcal{C} := \bigwedge \Gamma$ of the elements of Γ also belongs to $\mathrm{Cons}(\phi)$, and thus we have that $\mathcal{C}$ is a Boolean $\mathcal{L}$ CBox that is equivalent to ϕ.

We prove $\mathrm{Cons}(\phi) \models \phi$ by contradiction. Thus, assume that $\mathrm{Cons}(\phi) \not\models \phi$. Then $\mathrm{Cons}(\phi) \cup \{\neg\phi\}$ has a model $\mathcal{I}^-$, of which we can assume without loss of generality that it is ω-saturated (thanks to Theorem 34).

Now, let $\mathcal{G}$ denote the set of $\mathcal{L}$ CRs that are satisfied by $\mathcal{I}^-$. We claim that $\mathcal{G} \cup \{\phi\}$ has a model. In fact, otherwise first-order compactness would yield a finite subset $\mathcal{G}'$ of $\mathcal{G}$ such that $\mathcal{G}' \cup \{\phi\}$ also does not have a model. However, this would imply that $\phi \to \neg \bigwedge \mathcal{G}'$ is a tautology, which would yield $\neg \bigwedge \mathcal{G}' \in \mathrm{Cons}(\phi)$. This lead to a contradiction since now both $\bigwedge \mathcal{G}'$ and $\neg \bigwedge \mathcal{G}'$ would need to be satisfied by $\mathcal{I}^-$. Thus, we have shown that $\mathcal{G} \cup \{\phi\}$ has a model $\mathcal{I}^+$, of which can again assume that it is ω-saturated.

We observe that $\mathcal{I}^-$ and $\mathcal{I}^+$ both satisfy exactly the CRs occurring in $\mathcal{G}$, which implies that they are comparatively $\mathcal{L}$ equivalent. Since these two interpretations are also ω-saturated, Proposition 37 yields $\mathcal{I}^- \sim^{\bowtie}_{\mathcal{L}} \mathcal{I}^+$. This contradicts our assumption that (2) holds since we have $\mathcal{I}^+ \models \phi$, but $\mathcal{I}^- \not\models \phi$ Thus, we have shown that (2) implies $\mathrm{Cons}(\phi) \models \phi$, which concludes our proof. □

Since ECBoxes are invariant under comparative $\mathcal{L}$ bisimulation by Corollary 33, Theorem 38 yields the following characterization of the first-order fragment of ECBoxes for the DLs $\mathcal{ALCQ}$ and $\mathcal{ALCQ}t$.

Theorem 39. *Let $\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t\}$ and $\mathcal{E}$ be an $\mathcal{L}$ ECBox. Then the following are equivalent:*

1. *There exists a first-order sentence ϕ such that $\mathcal{E} \equiv \phi$.*
2. *$\mathcal{E}$ is equivalent to a Boolean $\mathcal{L}$ CBox $\mathcal{C}$.*

It remains to show that there are $\mathcal{ALCQ}$ ECBoxes that are not equivalent to a first-order sentence. Since it uses a technique different from the ones employed until now in this paper, we defer the proof of this result to the next section.

We close the current section by giving a characterization of the first-order fragment of $\mathcal{ALCSCC}^\infty$ ECBoxes.

Theorem 40. *Let $\mathcal{E}$ be an $\mathcal{ALCSCC}^\infty$ ECBox. Then the following are equivalent:*

1. *There exists a first-order sentence ϕ such that $\mathcal{E} \equiv \phi$.*
2. *$\mathcal{E}$ is equivalent to a Boolean $\mathcal{ALCQ}t$ CBox $\mathcal{C}$.*

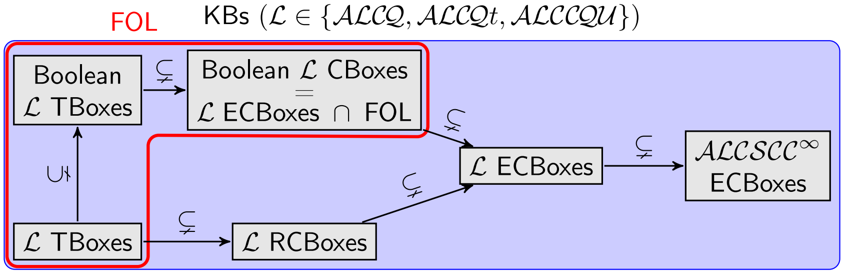
FOL
KBs ($\mathcal{L} \in \{\mathcal{ALCQ}, \mathcal{ALCQ}t, \mathcal{ALCCQU}\}$)
Boolean $\mathcal{L}$ TBoxes
$\subsetneq$
Boolean $\mathcal{L}$ CBoxes
=
$\mathcal{L}$ ECBoxes $\cap$ FOL
$\subsetneq$
$\mathcal{L}$ ECBoxes
$\subsetneq$
$\mathcal{ALCSCC}^\infty$ ECBoxes
$\subsetneq$
$\subsetneq$
$\subsetneq$
$\mathcal{L}$ TBoxes
$\mathcal{L}$ RCBoxes

Figure 3: The relative expressivity of boxes.

Proof. To prove (1 ⇒ 2), assume that ϕ is a first-order sentence equivalent to $\mathcal{E}$. It is easy to show that $\mathcal{ALCSCC}^\infty$ ECBoxes are invariant under comparative $\mathcal{ALCQ}t$ bisimulation. Therefore, ϕ is also invariant under comparative $\mathcal{ALCQ}t$ bisimulation. By Theorem 38, this implies that ϕ, and hence $\mathcal{E}$, is equivalent to a Boolean $\mathcal{ALCQ}t$ CBox $\mathcal{C}$.

(2 ⇒ 1) is an immediate consequence of the fact that Boolean $\mathcal{ALCQ}t$ CBoxes have a first-order translation (see Corollary 6). □

Fig. 3 summarizes the results obtained in this section and the next section.

5 ECBoxes and the 0-1 law for FOL

Let ϕ be a first-order sentence over a finite relational signature δ. We denote by $L_n(\delta)$ the set of interpretations over the signature δ with domain $\{1, \ldots, n\}$, and with $L_n(\phi)$ the number of these interpretations that are models of ϕ. We then set

$$\ell(\phi) := \lim_{n\to\infty} \frac{L_n(\phi)}{L_n(\delta)}. \tag{6}$$

Theorem 41 (0-1 law of FOL [13]). *For every first-order sentence ϕ, the limit $\ell(\phi)$ always exists and is equal to 0 or 1.*

One can use this theorem to prove that a sentence of a certain logic cannot be equivalent to a first-order sentence by showing that the corresponding limit either does not exist or is a number different from 0 or 1. An example for the former case would be a formula whose models are exactly the interpretations whose domain has even cardinality. We show now that ECBoxes can yield examples for the latter case.

Proposition 42. *The ECBox $\mathcal{E} := |A| \leqslant |\neg A|$ is not expressible as a first-order sentence.*

Proof. By contradiction, assume that $\mathcal{E}$ is equivalent to some first-order sentence ϕ. We restrict our attention to the relational signature $\delta := \{A\}$ since the only relation symbol contained in $\mathcal{E}$ is the concept name A. If we consider interpretations $\mathcal{I}$ with domain $\Delta^\mathcal{I} = \{1, \ldots, n\}$, then there are 2^n possible ways of interpreting $A^\mathcal{I}$, which shows that $L_n(\delta) = 2^n$. Among these interpretations, the ones where $|A^\mathcal{I}| = j$ for $0 \leqslant j \leqslant n$ are exactly $\binom{n}{j}$. Therefore, the number of interpretations with domain $\{1, \ldots, n\}$ over δ satisfying $\mathcal{E}$, and hence ϕ, is

$$L_n(\phi) = \textstyle\sum_{j=0}^{\lfloor n/2 \rfloor} \binom{n}{j}. \tag{7}$$

Let $\ell_n(\phi) := L_n(\phi)/L_n(\delta)$. We show that the sequence $L := (\ell_n(\phi))_{n\geqslant 1}$ is convergent and $\ell(\phi) := \lim_{n\to\infty} \ell_n(\phi) = 1/2$.[8] This yields a contradiction: by Theorem 41, it should hold that $\ell(\phi) = 0$ or $\ell(\phi) = 1$.

We split the sequence L into two subsequences $L_1 := (\ell_{2n}(\phi))_{n\geqslant 1}$ and $L_2 := (\ell_{2n+1}(\phi))_{n\geqslant 1}$. To show that L converges to $1/2$, it is sufficient to prove that both L_1 and L_2 have this limit. First, note that for $n \geqslant 1$ the following identities hold (which can, e.g., be shown by an application of Newton's binomial theorem):

$$2^{2n+1} = 2 \cdot \textstyle\sum_{j=0}^{n} \binom{2n+1}{j} \qquad (8) \qquad\qquad \textstyle\sum_{j=0}^{n} \binom{2n}{j} = \frac{1}{2} \cdot (2^{2n} + \binom{2n}{n}) \qquad (9)$$

By (8), our claim clearly holds for L_2. Indeed, for $n \geqslant 1$ we have

$$\ell_{2n+1}(\phi) = \frac{\sum_{j=0}^{n} \binom{2n+1}{j}}{2^{2n+1}} \overset{(8)}{=} \frac{\sum_{j=0}^{n} \binom{2n+1}{j}}{2 \cdot \sum_{j=0}^{n} \binom{2n+1}{j}} = \frac{1}{2}.$$

Regarding the other subsequence, note that the n-th term of L_1 corresponds to

$$\ell_{2n}(\phi) = \frac{\sum_{j=0}^{n} \binom{2n}{j}}{2^{2n}} \overset{(9)}{=} \frac{1}{2} + \frac{1}{2} \cdot \frac{\binom{2n}{n}}{4^n}.$$

We know that the following asymptotic equivalence holds [24]:

$$\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n}}. \qquad (10)$$

Hence, we deduce that

$$\lim_{n\to\infty} \ell_{2n}(\phi) = \frac{1}{2} + \frac{1}{2} \lim_{n\to\infty} \frac{\binom{2n}{n}}{4^n} = \frac{1}{2} + \frac{1}{2} \lim_{n\to\infty} \frac{1}{\sqrt{\pi n}} = \frac{1}{2}.$$

This yields the convergence of L_1 to $1/2$ as desired. □

6 Conclusion

In this paper, we have provided an almost complete picture of the complexity and expressivity of the DLs with extended counting facilities introduced in our previous work. Regarding expressivity, it would be interesting to see whether the results presented here for the "arbitrary model" setting also hold for the "finite model" case. We conjecture that this is the case, but it should be noted that the proofs of the bisimulation characterizations given here (Theorems 18, 20, 26, 27, 29, 38, 39, and 40) crucially depend on the application of compactness of FOL, which does not hold in the finite case. One idea for overcoming this problem could be to consider extensions of the model comparison games introduced in [21] instead of bisimulations. Regarding reasoning, the next step will be to design variants of our decision procedures that are more appropriate for implementation than the ones used to show the complexity results. The main idea for this is to combine SAT solvers with numerical methods, such as simplex, branch-and-bound, and column generation, similar to the decision procedure for CQU sketched in [14].

[8] This was already stated in [16], but without proof.

Acknowledgments

Franz Baader was partially supported by the German Research Foundation (DFG) within the Collaborative Research Center 248 (CPEC, grant 389792660), and Filippo De Bortoli was supported by the German Research Foundation (DFG) within the Research Training Group 1763 (QuantLA). The authors would also like to thank Ulrike Baumann for her help with computing limits of sequences of numbers.

References

[1] Franz Baader. A new description logic with set constraints and cardinality constraints on role successors. In Clare Dixon and Marcelo Finger, editors, *Proc. of the 11th Int. Symposium on Frontiers of Combining Systems (FroCoS'17)*, volume 10483 of *Lecture Notes in Computer Science*, pages 43–59, Brasília, Brazil, 2017. Springer-Verlag.

[2] Franz Baader. Expressive cardinality constraints on $\mathcal{ALCSCC}$ concepts. In *Proc. of the 34th ACM/SIGAPP Symposium On Applied Computing (SAC'19)*. ACM, 2019.

[3] Franz Baader. Expressive cardinality restrictions on concepts in a description logic with expressive number restrictions. *ACM SIGAPP Applied Computing Review*, 19:5–17, 2019.

[4] Franz Baader and Filippo De Bortoli. On the expressive power of description logics with cardinality constraints on finite and infinite sets. In Andreas Herzig and Andrei Popescu, editors, *Proc. of the 12th Int. Symposium on Frontiers of Combining Systems (FroCoS'19)*, volume 11715 of *Lecture Notes in Computer Science*. Springer-Verlag, 2019.

[5] Franz Baader, Martin Buchheit, and Bernhard Hollunder. Cardinality restrictions on concepts. *Artificial Intelligence*, 88(1–2):195–213, 1996.

[6] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[7] Franz Baader and Filippo De Bortoli. On the Complexity and Expressiveness of Description Logics with Counting. LTCS-Report 19-09, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany, 2019. https://tu-dresden.de/inf/lat/reports#BaBo-LTCS-19-09.

[8] Franz Baader and Andreas Ecke. Extending the description logic $\mathcal{ALC}$ with more expressive cardinality constraints on concepts. In *Proc. of the 3rd Global Conf. on Artificial Intelligence (GCAI'17)*, volume 50 of *EPiC Series in Computing*, pages 6–19. EasyChair, 2017.

[9] Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logic*. Cambridge University Press, 2017.

[10] Alexander Borgida, Ronald J. Brachman, Deborah L. McGuinness, and Lori Alperin Resnick. CLASSIC: A structural data model for objects. In *Proc. of the ACM SIGMOD Int. Conf. on Management of Data*, pages 59–67, 1989.

[11] Chen Chung Chang and H. Jerome Keisler. *Model theory*. Number v. 73 in Studies in logic and the foundations of mathematics. Elsevier, 3rd ed edition, 1990.

[12] Stéphane Demri and Denis Lugiez. Complexity of modal logics with Presburger constraints. *J. Applied Logic*, 8(3):233–252, 2010.

[13] Ronald Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41(1):50–58, 1976.

[14] Marcelo Finger and Glauber De Bona. Algorithms for deciding counting quantifiers over unary predicates. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*, pages 3878–3884. AAAI Press, 2017.

[15] Erich Grädel, Martin Otto, and Eric Rosen. Two-variable logic with counting is decidable. In *Proc. of the 12th IEEE Symp. on Logic in Computer Science (LICS'97)*, pages 306–317. IEEE

Computer Society Press, 1997.

[16] Stéphane Grumbach and Christophe Tollu. On the expressive power of counting. *Theoretical Computer Science*, 149(1):67–99, September 1995.

[17] P. Hall. On representatives of subsets. *Journal of the London Mathematical Society*, s1-10(1):26–30, 1935.

[18] Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. The role of ontologies in biological and biomedical research: A functional perspective. *Brief. Bioinform.*, 16(6):1069–1080, 2015.

[19] Bernhard Hollunder and Franz Baader. Qualifying number restrictions in concept languages. In *Proc. of the 2nd Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'91)*, pages 335–346, 1991.

[20] Bernhard Hollunder, Werner Nutt, and Manfred Schmidt-Schauß. Subsumption algorithms for concept description languages. In *Proc. of the 9th Eur. Conf. on Artificial Intelligence (ECAI'90)*, pages 348–353, London (United Kingdom), 1990. Pitman.

[21] Jean Christoph Jung, Fabio Papacchini, Frank Wolter, and Michael Zakharyaschev. Model comparison games for Horn description logics. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019*, pages 1–14. IEEE, 2019.

[22] Viktor Kuncak, Ruzica Piskac, and Philippe Suter. Ordered sets in the calculus of data structures. In Anuj Dawar and Helmut Veith, editors, *Computer Science Logic*, volume 6427 of *Lecture Notes in Computer Science*, pages 34–48. Springer, 2010.

[23] Viktor Kuncak and Martin Rinard. Towards efficient satisfiability checking for boolean algebra with presburger arithmetic. In Frank Pfenning, editor, *Automated Deduction – CADE-21*, Lecture Notes in Computer Science, pages 215–230. Springer Berlin Heidelberg, 2007.

[24] L. Lovász, J. Pelikán, and K. Vesztergombi. Binomial coefficients and pascal's triangle. In *Discrete Mathematics: Elementary and Beyond*, pages 43–64. Springer New York, New York, NY, 2003.

[25] Carsten Lutz, Robert Piro, and Frank Wolter. Description Logic TBoxes: Model-Theoretic Characterizations and Rewritability. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011*, pages 983–988, 2011. Long version available at https://arxiv.org/pdf/1104.2844.pdf.

[26] Leszek Pacholski, Wieslaw Szwast, and Lidia Tendera. Complexity of two-variable logic with counting. In *Proc. of the 12th IEEE Symp. on Logic in Computer Science (LICS'97)*, pages 318–327. IEEE Computer Society Press, 1997.

[27] Ian Pratt-Hartmann. Complexity of the two-variable fragment with counting quantifiers. *J. of Logic, Language and Information*, 14(3):369–395, 2005.

[28] Ian Pratt-Hartmann. On the computational complexity of the numerically definite syllogistic and related logics. *Bulletin of Symbolic Logic*, 14(1):1–28, 2008.

[29] Stephan Tobies. A PSPACE algorithm for graded modal logic. In Harald Ganzinger, editor, *Proc. of the 16th Int. Conf. on Automated Deduction (CADE'99)*, volume 1632 of *Lecture Notes in Artificial Intelligence*, pages 52–66. Springer-Verlag, 1999.

[30] Stephan Tobies. The complexity of reasoning with cardinality restrictions and nominals in expressive description logics. *J. of Artificial Intelligence Research*, 12:199–217, 2000.

[31] Stephan Tobies. *Complexity Results and Practical Algorithms for Logics in Knowledge Representation.* PhD thesis, LuFG Theoretical Computer Science, RWTH-Aachen, Germany, 2001.

[32] Kuat Yessenov, Ruzica Piskac, and Viktor Kuncak. Collections, cardinalities, and relations. In Gilles Barthe and Manuel V. Hermenegildo, editors, *Proc. of the 11th International Conference on Verification, Model Checking, and Abstract Interpretation (VMCAI'10)*, volume 5944 of *Lecture Notes in Computer Science*, pages 380–395. Springer-Verlag, 2010.