Islands of tractability for relational constraints: towards dichotomy results for the description logic $\mathcal{EL}$

A. Kurucz

Department of Computer Science
King's College London, UK
agi.kurucz@kcl.ac.uk

F. Wolter

Department of Computer Science
University of Liverpool, UK
frank@csc.liv.ac.uk

M. Zakharyaschev

Department of Computer Science and Information Systems
Birkbeck College London, UK
michael@dcs.bbk.ac.uk

Abstract

$\mathcal{EL}$ is a tractable description logic serving as the logical underpinning of large-scale ontologies. We launch a systematic investigation of the boundary between tractable and intractable reasoning in $\mathcal{EL}$ under relational constraints. E.g., we show that there are (modulo equivalence) exactly 3 universal constraints on a transitive and reflexive relation under which reasoning is tractable: being a singleton set, an equivalence relation, or the empty constraint. We prove a number of results of this type and discuss a spectrum of open problems including generalisations to the algebraic semantics for $\mathcal{EL}$ (semi-lattices with monotone operators).

Keywords: Description logic, tractability, frame condition.

1 Introduction

Standard modal logics are usually based on propositional logic and therefore cannot be tractable: unless P = NP, no algorithm is capable of checking validity (or satisfiability) for such a logic in polynomial time. In most cases, the computational complexity is even higher: with the notable exception of **S5**, basic modal logics like **K**, **K4**, **S4**, the Gödel–Löb logic **GL** and the Grzegorzcyk logic **Grz**, as well as their polymodal variants, are all PSpace-complete as far as the 'local' reasoning problem 'if φ is true in a world, then ψ is true in that world' is concerned. The 'global' reasoning problem 'if φ is true in all worlds, then ψ is true all worlds' is ExpTime-complete for all polymodal fusions of these logics and even unimodal $\mathbf{K}$.

Very few attempts have been made to understand the complexity of *sub-Boolean* modal logics, which do not have all propositional connectives or use them in a restricted way. For example, Hemaspaandra [9] considered satisfiability of the 'poor man's formulas,' built from literals, $\wedge$, $\Box$ and $\Diamond$, over various classes of frames. A complete classification of the complexity of modal satisfiability for finite sets of propositional connectives (without any constraints on frames) was obtained in [4].

In description logic (DL), the situation is quite different.[1] Until the mid-1990s, sub-Boolean DLs were the rule rather than exception, and mapping out the border between DLs with tractable and non-tractable reasoning problems was one of the main research goals [5]. This changed drastically in the second half of the 1990s when the focus was shifted to DLs with all Booleans (the so-called *expressive DLs*) due to the development of highly optimised tableau decision procedures and reasoning systems exhibiting satisfactory performance on real-world ontologies given in expressive DLs [10]. As a consequence, the DL-based web ontology language OWL, which became a W3C standard in 2003, was based solely on expressive DLs with (at least) ExpTime-hard TBox reasoning. Since then, however, two developments have led to a massive resurgence of interest in sub-Boolean and tractable DLs.

First, very large ontologies like SNOMED CT[2] (with $\geq 300,000$ axioms) have been designed and used in every day practice. These ontologies represent application domains at such a high level of abstraction that the full power of propositional connectives is not required. On the other hand, the enormous size of the ontologies makes tractability of reasoning a crucial factor. Second, realising the idea of employing ontologies for data access requires query answering to be tractable, at least in the size of the typically very large data sets. The two main families of tractable DLs currently evolving are $\mathcal{EL}$ and *DL-Lite*. $\mathcal{EL}$ is tailored towards representing large ontologies; it is the logical underpinning of the OWL 2 profile OWL 2 EL. *DL-Lite* is designed for ontology-based data access; it is the basis of OWL 2 QL.

In this paper, we focus on the DL $\mathcal{EL}$, where concepts are constructed using intersection $\sqcap$ and existential restriction $\exists r.C$ ($\wedge$ and $\Diamond_r \varphi$, in the modal logic parlance) interpreted over relational (or Kripke) models. The fundamental *subsumption problem for general TBoxes* in $\mathcal{EL}$—whether every model of an $\mathcal{EL}$ TBox (a set of concept inclusions $C \sqsubseteq D$) satisfies a given concept inclusion $C' \sqsubseteq D'$—is decidable in polynomial time. In modal logic, this inference corresponds to the *global consequence* relation 'if a set of implications $\varphi \to \psi$ between $\mathcal{EL}$-formulas is true in every world of a Kripke model, then an implication $\varphi' \to \psi'$ is true in every world of the model.' In algebraic terms, this problem is equivalent to the validity problem for *quasi-identities* in the variety of semi-lattices with monotone operators [13].

In DL applications, the intended models are rarely arbitrary; more often they have to satisfy certain constraints. Of particular importance are constraints imposed on the interpretation of relations. For example, the Gene Ontology GO[3] is an $\mathcal{EL}$

[1] We refer to differences between research communities and their activities rather than differences between modal and description logics. The view taken in this paper is that DLs form a class of modal logics [3].

[2] http://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html

[3] http://www.geneontology.org/

ontology with one transitive relation. SNOMED CT is an $\mathcal{EL}$-ontology interpreted over models where certain relations are included in each other (e.g., causative_agent is a subrelation of associated_with). Other standard OWL constraints (also familiar from modal logic) include (ir)reflexivity, (a)symmetry and functionality. The complexity of reasoning in $\mathcal{EL}$ under such concrete relational constraints is well understood [1,2,13]. For example, the subsumption problem for general TBoxes in $\mathcal{EL}$ is tractable for any finite set of constraints of the form

$$r_1(x_1, x_2) \wedge \cdots \wedge r_n(x_n, x_{n+1}) \to r_{n+1}(x_1, x_{n+1}) \tag{1}$$

(the order of the variables is essential). On the other hand, subsumption becomes ExpTime-complete in the presence of symmetry or functionality constraints [2].

Nevertheless, from a theoretical point of view, the selection of constraints on $\mathcal{EL}$ models investigated so far is rather *ad hoc* and narrow. In fact, no attempt has been made to *classify* constraints according to tractability of $\mathcal{EL}$-reasoning. The aim of this paper is to start filling in this gap by mapping out the border between tractability and intractability of TBox reasoning in $\mathcal{EL}$ under *arbitrary relational constraints*. Our initial findings indicate that informative dichotomy results can indeed be obtained. We establish transparent P/coNP dichotomies for finite classes of finite relational structures, universal classes of quasi-orders, and classes of Noetherian partial orders closed under substructures. Not every relational constraint is 'visible' to $\mathcal{EL}$: for example, as in modal logic, TBox reasoning over irreflexive relations coincides with TBox reasoning over arbitrary relations. To obtain basic insights into relational constraints 'visible' to $\mathcal{EL}$, we show that, for universal classes of relational constraints, there is no difference between modal definability and definability in $\mathcal{EL}$. On the other hand, a typical condition definable in modal logic but not in $\mathcal{EL}$ is the Church-Rosser property.

2 Description logic $\mathcal{EL}$

Fix two disjoint countably infinite sets NC of *concept names* and NR of *role names*. Let $R \subseteq \mathsf{NR}$. *$\mathcal{EL}$-concepts C over R* are defined inductively as follows:

$$C \quad ::= \quad \top \quad | \quad \bot \quad | \quad A \quad | \quad C_1 \sqcap C_2 \quad | \quad \exists r.C,$$

where $A \in \mathsf{NC}$, $r \in R$ and C, C_1, C_2 range over $\mathcal{EL}$-concepts over R. An *R-TBox* is a finite set of *concept inclusions* (CIs) $C \sqsubseteq D$, where C and D are $\mathcal{EL}$-concepts over R. An *R-interpretation* is a structure of the form $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}} \neq \emptyset$ is the *domain* of interpretation and $\cdot^{\mathcal{I}}$ is an *interpretation function* assigning to each concept name $A \in \mathsf{NC}$ a set $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and to each role name $r \in R$ a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. Complex concepts over R are interpreted in $\mathcal{I}$ as follows:

$$\top^{\mathcal{I}} = \Delta^{\mathcal{I}}, \qquad \bot^{\mathcal{I}} = \emptyset,$$
$$(C_1 \sqcap C_2)^{\mathcal{I}} = C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}, \qquad (\exists r.C)^{\mathcal{I}} = \{x \in \Delta^{\mathcal{I}} \mid \exists y \in C^{\mathcal{I}}\ (x, y) \in r^{\mathcal{I}}\}.$$

If $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, we say that $\mathcal{I}$ *satisfies* $C \sqsubseteq D$ and write $\mathcal{I} \models C \sqsubseteq D$. $\mathcal{I}$ is a *model* of a TBox $\mathcal{T}$, $\mathcal{I} \models \mathcal{T}$ in symbols, if it satisfies all the CIs in $\mathcal{T}$.

We now formalise the constraints on interpretations. An *R-frame* is a structure $\mathfrak{F} = (\Delta^{\mathfrak{F}}, \cdot^{\mathfrak{F}})$ where $\Delta^{\mathfrak{F}} \neq \emptyset$ and $\cdot^{\mathfrak{F}}$ is a map associating with each $r \in R$ a relation $r^{\mathfrak{F}} \subseteq \Delta^{\mathfrak{F}} \times \Delta^{\mathfrak{F}}$. An *R-constraint*, or an *R-frame condition*, is a class $\mathcal{K}$ of R-frames closed under isomorphic copies. For example, a constraint for $R = \{r_1, r_2, r_3\}$ can consist of R-frames $\mathfrak{F} = (\Delta^{\mathfrak{F}}, \cdot^{\mathfrak{F}})$ with arbitrary $r_1^{\mathfrak{F}}$, transitive $r_2^{\mathfrak{F}}$ and functional $r_3^{\mathfrak{F}}$.

We say that an R-interpretation $\mathcal{I}$ is *based on* an R-frame $\mathfrak{F}$ if $\Delta^{\mathfrak{F}} = \Delta^{\mathcal{I}}$ and $r^{\mathfrak{F}} = r^{\mathfrak{I}}$ for all $r \in R$. $\mathcal{I}$ *satisfies* an R-constraint $\mathcal{K}$ if $\mathcal{I}$ is based on some $\mathfrak{F} \in \mathcal{K}$. We say that $C \sqsubseteq D$ *follows from* $\mathcal{T}$ *with respect to* $\mathcal{K}$ and write $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$ if $\mathcal{I} \models C \sqsubseteq D$ for every model $\mathcal{I}$ of $\mathcal{T}$ satisfying $\mathcal{K}$. For singleton $\mathcal{K} = \{\mathfrak{F}\}$, we sometimes write $\mathcal{T} \models_{\mathfrak{F}} C \sqsubseteq D$.

A pair $(\mathcal{T}, C \sqsubseteq D)$, where $\mathcal{T}$ is an R-TBox and $C \sqsubseteq D$ an R-CI, is called an *R-consequence* in $\mathcal{EL}$. Given an R-frame condition $\mathcal{K}$, we define $\mathrm{Th}_T(\mathcal{K})$, the *TBox-theory of* $\mathcal{K}$, to be the set of all R-consequences $(\mathcal{T}, C \sqsubseteq D)$ such that $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$. The *subsumption problem for* $\mathcal{K}$ is the decision problem for the set $\mathrm{Th}_T(\mathcal{K})$: given an R-consequence $(\mathcal{T}, C \sqsubseteq D)$, decide whether $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$.

Example 2.1 In the extension $\mathcal{EL}^+$ of $\mathcal{EL}$ [1], along with a TBox one can also define an *RBox* containing inclusions of the form $r_1 \circ \dots \circ r_n \sqsubseteq r_{n+1}$, where $r_1, \dots, r_{n+1}$ are role names. In this case, we write $(\mathcal{T}, \mathcal{R}) \models C \sqsubseteq D$ if $\mathcal{I} \models C \sqsubseteq D$ holds whenever $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I}$ satisfies constraint (1) for every $r_1 \circ \dots \circ r_n \sqsubseteq r_{n+1} \in \mathcal{R}$. Reasoning with RBoxes $\mathcal{R}$ as defined above is clearly captured by the frame condition $\mathcal{K}_{\mathcal{R}}$ containing all $\mathsf{N}\mathsf{R}$-frames $\mathfrak{F}$ in which costraint (1) is valid for all $r_1 \circ \dots \circ r_n \sqsubseteq r_{n+1}$ in $\mathcal{R}$. According to [1,13], the subsumption problem for any such $\mathcal{K}_{\mathcal{R}}$ is decidable in polynomial time.

Example 2.2 It follows from Example 2.1 that the subsumption problem for the class of transitive frames is in P. Similarly, it is straightforward to extend existing proofs to show that the subsumption problem for the classes of reflexive or reflexive and transitive frames is also in P. On the other hand, the subsumption problem for the class of symmetric frames is ExpTime-complete [2].

3 TBox definability

To better understand the frame conditions in the context of $\mathcal{EL}$, let us take a look at frame classes that can be defined using TBoxes and compare them with modally definable frame classes. Thus, we take a brief detour into what is known in modal logic as *correspondence theory* [15].

Call R-frame conditions $\mathcal{K}_1$ and $\mathcal{K}_2$ *TBox-equivalent* if $\mathrm{Th}_T(\mathcal{K}_1) = \mathrm{Th}_T(\mathcal{K}_2)$. For example, the standard unravelling argument from modal logic shows that the TBox theory of the class of all frames coincides with the TBox theory of the class of all irreflexive frames. Similarly, the finite model property of the TBox theory of all frames [1] means that it is TBox-equivalent to the TBox theory of all finite frames.

Given a set Γ of R-consequences, denote by $\mathrm{Fr}\Gamma$ the class of R-frames $\mathfrak{F}$ such that $\mathcal{T} \models_{\mathfrak{F}} C \sqsubseteq D$ for all $(\mathcal{T}, C \sqsubseteq D) \in \Gamma$. An R-frame condition $\mathcal{K}$ is *TBox-definable* if $\mathcal{K} = \mathrm{Fr}\Gamma$ for a suitable set Γ of R-consequences. For example, the class of transitive $\{r\}$-frames is defined by $\Gamma = \{(\emptyset, \exists r.\exists r.A \sqsubseteq \exists r.A)\}$. Observe that in this definition the TBox is empty. Such R-frame conditions are called *concept* *definable*. Density is another example of a concept definable frame condition: it is defined by $\Gamma = \{(\emptyset, \exists r.A \sqsubseteq \exists r.\exists r.A)\}$.

The class of R-frames defined by $(\emptyset, C \sqsubseteq D)$ is clearly the class of R-frames validating the modal formula $C^\sharp \to D^\sharp$, where $\cdot^\sharp$ replaces each $A \in \mathsf{NC}$ with a propositional variable and each $\exists r$ with $\Diamond_r$. As all formulas of the form $C^\sharp \to D^\sharp$ are Sahlqvist, every concept definable class is first-order definable, and its first-order definition can be computed effectively [12]. More generally, a class $\mathcal{K}$ of R-frames is *modally definable* if there is a set Γ of modal formulas such that $\mathfrak{F} \in \mathcal{K}$ if $\mathfrak{F} \models \Gamma$. $\mathcal{K}$ is called *globally definable* if there is a set Γ of pairs (φ, ψ) of modal formulas such that $\mathfrak{F} \in \mathcal{K}$ iff $\mathfrak{F} \models \Box_u\varphi \to \Box_u\psi$, where $\Box_u$ is the universal modality [8]. One can easily show that every TBox definable class is globally definable.

Recall from modal logic that a *p-morphism* from an R-frame $\mathfrak{F}_1$ to an R-frame $\mathfrak{F}_2$ is a function $f\colon \Delta^{\mathfrak{F}_1} \to \Delta^{\mathfrak{F}_2}$ such that, for every $r \in R$, (i) $(v_1, v_2) \in r^{\mathfrak{F}_1}$ implies $(f(v_1), f(v_2)) \in r^{\mathfrak{F}_2}$ and (ii) if $(f(v_1), w) \in r^{\mathfrak{F}_2}$, then there is v_2 with $(v_1, v_2) \in r^{\mathfrak{F}_1}$ and $f(v_2) = w$. If there is a p-morphism from $\mathfrak{F}_1$ onto $\mathfrak{F}_2$, then $\mathfrak{F}_2$ is called a *p-morphic* image of $\mathfrak{F}_1$. An R-frame $\mathfrak{F}_1$ is called a *subframe* of an R-frame $\mathfrak{F}_2$ if $\Delta^{\mathfrak{F}_1} \subseteq \Delta^{\mathfrak{F}_2}$ and $r^{\mathfrak{F}_1}$ is the restriction of $r^{\mathfrak{F}_2}$ to $\Delta^{\mathfrak{F}_1}$, for every $r \in R$. A subframe $\mathfrak{F}_1$ of $\mathfrak{F}_2$ is said to be *generated* if whenever $u \in \Delta^{\mathfrak{F}_1}$ and $(u, v) \in r^{\mathfrak{F}_2}$, for some $r \in R$, then $v \in \Delta^{\mathfrak{F}_1}$. Finally, $u \in \Delta^{\mathfrak{F}}$ is a *root* of a frame $\mathfrak{F}$ if the subframe of $\mathfrak{F}$ generated by u coincides with $\mathfrak{F}$.

The following result is straightforward and left to the reader:

Lemma 3.1 *TBox definable frame conditions are closed under p-morphic images and disjoint unions.*

However, unlike modally definable frame classes, TBox definable classes are not necessarily closed under generated subframes.

Example 3.2 Let $\Gamma = (\{\top \sqsubseteq \exists r.\top\}, \top \sqsubseteq \bot)$. Then the $\{r\}$-frame condition $\mathrm{Fr}(\Gamma)$ contains the $\{r\}$-frame $\mathfrak{F}$, which is the disjoint union of an r-reflexive point and an r-irreflexive point, as no interpretation based on $\mathfrak{F}$ is a model of $\top \sqsubseteq \exists r.\top$. However, the subframe of $\mathfrak{F}$ generated by the r-reflexive point does not belong to $\mathrm{Fr}\Gamma$.

A *universal R-frame condition* is a class of R-frames axiomatisable by universal first-order sentences in the signature R. Equivalently, by [14], a universal frame condition is a first-order definable class of frames closed under taking (not necessarily generated) subframes. The vast majority of frame conditions considered in modal and description logics are universal: transitivity, reflexivity, symmetry, weak linearity, just to mention a few. Typical examples on non-universal conditions are the Church-Rosser property and density.

To characterise TBox definable universal frame conditions, with every R-frame $\mathfrak{F}$ we associate the 'TBox' $\mathcal{T}_S(\mathfrak{F})$ (here we slightly abuse notation as $\mathcal{T}_S(\mathfrak{F})$ is infinite whenever $\mathfrak{F}$ or R is infinite) containing the following CIs, where the A_u, for $u \in \Delta^{\mathfrak{F}}$, are distinct concept names:

- $A_u \sqsubseteq \exists r.A_v$, for $(u, v) \in r^{\mathfrak{F}}$, $r \in R$;
- $A_u \sqcap A_v \sqsubseteq \bot$, for $u \neq v$;
- $A_u \sqcap \exists r.A_v \sqsubseteq \bot$, for $(u, v) \notin r^{\mathfrak{F}}$, $r \in R$.

The meaning of $\mathcal{T}_S(\mathfrak{F})$ is explained by the following lemma (the standard proof of which is left to the reader):

Lemma 3.3 *Let $\mathfrak{F}$ be an R-frame with root w. Then, for every R-frame $\mathfrak{G}$, we have $\mathcal{T}_S(\mathfrak{F}) \not\models_{\mathfrak{G}} A_w \sqsubseteq \bot$ iff $\mathfrak{F}$ is a p-morphic image of a subframe of $\mathfrak{G}$.*

Using this lemma we obtain a characterisation of TBox definable universal frame conditions:

Theorem 3.4 *Let $\mathcal{K}$ be a universal class of R-frames, for some $R \subseteq \mathsf{NR}$. Then the following conditions are equivalent:*

(1) *$\mathcal{K}$ is TBox-definable;*

(2) *$\mathcal{K}$ is closed under p-morphic images and disjoint unions;*

(3) *$\mathcal{K}$ is modally definable;*

(4) *$\mathcal{K}$ is globally definable.*

Proof. By Lemma 3.1, (1) $\Rightarrow$ (2) and, as shown in [16], (2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4). To prove that (2) $\Rightarrow$ (1) it suffices to show that $\mathrm{FrTh}_T\mathcal{K} \subseteq \mathcal{K}$. So suppose that $\mathfrak{F} \in \mathrm{FrTh}_T\mathcal{K}$. We will have $\mathfrak{F} \in \mathcal{K}$ if we can show that all rooted generated subframes of $\mathfrak{F}$ are in $\mathcal{K}$ (because $\mathfrak{F}$ is a p-morphic image of the disjoint union of these frames). So let $\mathfrak{F}_w$ be a rooted subframe of $\mathfrak{F}$ with root w. If $\mathfrak{F}_w \notin \mathcal{K}$ then, by Lemma 3.3, $\mathcal{T}_S(\mathfrak{F}_w) \models_{\mathcal{K}} A_w \sqsubseteq \bot$. By compactness—as $\mathcal{K}$ is first-order definable—there exists a finite subset $\mathcal{T}$ of $\mathcal{T}_S(\mathfrak{F}_w)$ with $\mathcal{T} \models_{\mathcal{K}} A_w \sqsubseteq \bot$. But then $(\mathcal{T}, A_w \sqsubseteq \bot) \in \mathrm{FrTh}_T\mathcal{K}$ and $\mathcal{T} \not\models_{\mathfrak{F}_w} A_w \sqsubseteq \bot$, which is a contradiction. □

We conjecture that the equivalence of (1) and (4) in Theorem 3.4 can be generalised to arbitrary (not necessarily first-order definable) classes of R-frames closed under subframes. Note that without the subframe condition there are modally but not TBox definable classes of frames. One example is the *Church-Rosser property*

$$\forall x, y_1, y_2 (r(x, y_1) \land r(x, y_2) \to \exists z (r(y_1, z) \land r(y_2, z))),$$

which is modally definable by $\Diamond\Box p \to \Box\Diamond p$, but not TBox definable; cf. Section A.

It is beyond the scope of this paper to develop correspondence theory any further. The main conclusion, however, is clear: as far as TBox definability is concerned, $\mathcal{EL}$ is still a very powerful language, and one has to go beyond subframe conditions to find natural classes of frames definable in modal logic but not in $\mathcal{EL}$.

4 P/coNP dichotomy for tabular frame conditions

An R-frame condition $\mathcal{K}$ is called *tabular* if there is a number $n > 0$ such that $|\Delta^{\mathfrak{F}}| \leq n$ for all $\mathfrak{F} \in \mathcal{K}$. $\mathcal{K}$ is called *R-functional* if, for every $\mathfrak{F} \in \mathcal{K}$, every $r \in R$ and every $w \in \Delta^{\mathfrak{F}}$, we have $|\{v \in \Delta^{\mathfrak{F}} \mid (w, v) \in r^{\mathfrak{F}}\}| \leq 1$.

Our proofs of coNP-hardness in this and further sections are by reduction of the following *set splitting problem*, which is known to be NP-complete [7]:

– given a family I of subsets of a finite set S, decide whether there exists a *splitting* of (S, I), that is, a partition S_1, S_2 of S such that each set $G \in I$ is split by S_1 and S_2 in the sense that it is not the case that $G \subseteq S_i$ for $i \in \{1, 2\}$.

The following theorem gives a characterisation of those tabular frame conditions for which the subsumption problem is tractable.

Theorem 4.1 *Let $\mathcal{K}$ be a tabular R-frame condition for a finite $R \subseteq \mathsf{NR}$. Then either $\mathcal{K}$ is functional, in which case $Th_T(\mathcal{K})$ is in* P, *or $Th_T(\mathcal{K})$ is* coNP*-complete.*

Proof. Assume first that $\mathcal{K}$ is functional and that we are given an R-TBox $\mathcal{T}$ and and R-CI $C' \sqsubseteq D'$. For R-interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ based on a frame $\mathfrak{F}$, we say that $\mathcal{I}_1$ is *smaller* than $\mathcal{I}_2$ and write $\mathcal{I}_1 \leq \mathcal{I}_2$ if $A^{\mathcal{I}_1} \subseteq A^{\mathcal{I}_2}$ for all $A \in \mathsf{NC}$. Clearly, $\leq$ is a partial order on the set of interpretations based on $\mathfrak{F}$.

Suppose that $\mathcal{I}$ is an R-interpretation based on some frame in $\mathcal{K}$ and $w \in \Delta^{\mathcal{I}}$.

Lemma 4.2 *Given any R-concept C, one can decide in polynomial time in $|C|$ whether there exists an R-interpretation $\mathcal{J}$ such that $\mathcal{I} \leq \mathcal{J}$ and $w \in C^{\mathcal{J}}$. If such an interpretation exists, then one can construct, again in polynomial time in $|C|$, the smallest (with respect to $\leq$) R-interpretation $\mathcal{I}(w,C) \geq \mathcal{I}$ such that $w \in C^{\mathcal{I}(w,C)}$.*

A simple proof of this lemma is given in Section B.

Our polynomial time algorithm checking whether $\mathcal{T} \models_{\mathcal{K}} C' \sqsubseteq D'$ runs as follows. Let $\mathfrak{F}_1, \dots, \mathfrak{F}_n$ be a list of all frames in $\mathcal{K}$ (up to isomorphism). For each $\mathfrak{F}_i$ and each $w \in \mathfrak{F}_i$, we do the following:

1. Let $\mathcal{I}$ be the R-interpretation based on $\mathfrak{F}_i$ with $A^{\mathcal{I}} = \emptyset$ for all $A \in \mathsf{NC}$.
2. Compute $\mathcal{I} := \mathcal{I}(w, C')$ if it exists. If it does not exist, return 'yes' and stop.
3. Apply the following rule exhaustively: for $C \sqsubseteq D \in \mathcal{T}$ and $v \in \Delta^{\mathcal{I}}$, if $v \in C^{\mathcal{I}}$ and $\mathcal{I}(v, D)$ does not exist, return 'yes' and stop; otherwise, if $\mathcal{I}(v,D) \neq \mathcal{I}$, set $\mathcal{I} = \mathcal{I}(v,D)$.
4. If $w \in (D')^{\mathcal{I}}$, return 'yes.' Otherwise, return 'no.'

It is easy to see that $\mathcal{T} \models_{\mathcal{K}} C' \sqsubseteq D'$ iff the output is 'yes' for all $\mathfrak{F}_i$ and all $w \in \Delta^{\mathfrak{F}_i}$.

Suppose now that $\mathcal{K}$ is not R-functional. Then there exists $\mathfrak{F} \in \mathcal{K}$ with $w \in \Delta^{\mathfrak{F}}$ such that $|\{v \mid (w,v) \in r^{\mathfrak{F}}\}| \geq 2$. Let m be the maximal number for which there exist $r \in R$, $\mathfrak{F} \in \mathcal{K}$ and $w \in \Delta^{\mathfrak{F}}$ with $|\{v \mid (w,v) \in r^{\mathfrak{F}}\}| = m$. Fix such r, $\mathfrak{F}$ and w.

It should be clear that the complement of $\mathrm{Th}_T(\mathcal{K})$ is decidable in nondeterministic polynomial time. We show now that $\mathrm{Th}_T(\mathcal{K})$ is coNP-hard by reduction of the set splitting problem. Suppose we are given an instance (S, I) of this problem. It will be convenient for us to assume that the members of S are concept names. Consider the $\{r\}$-TBox $\mathcal{T}$ containing the following CIs:

(a) $B_i \sqcap B_j \sqsubseteq \bot$, for $1 \leq i < j \leq m$;

(b) $A \sqcap B_i \sqsubseteq \bot$, for $3 \leq i \leq m$ and $A \in S$;

(c) $\exists r.(B_i \sqcap \bigsqcap_{A \in G} A) \sqsubseteq \bot$, for $i = 1, 2$ and $G \in I$.

The meaning of these CIs will become clear from the following:

Claim 4.3 *There exists a splitting of (S, I) iff*

$$\mathcal{T} \not\models_{\mathcal{K}} B \sqcap \bigsqcap_{A \in S} \exists r.A \sqcap \bigsqcap_{1 \leq i \leq m} \exists r.B_i \ \sqsubseteq \ \bot.$$

Proof. Suppose S_1, S_2 is a splitting of (S, I). Let $w_1, \dots, w_m$ be the r-successors of w in $\mathfrak{F}$. Define an interpretation $\mathcal{I}$ based on $\mathfrak{F}$ by setting $B_i^{\mathcal{I}} = \{w_i\}$, $B^{\mathcal{I}} = \{w\}$,

$$A^{\mathcal{I}} = \begin{cases} \{w_1\}, & \text{if } A \in S_1; \\ \{w_2\}, & \text{if } A \in S_2. \end{cases}$$

The reader can check that $w \in (B \sqcap \bigsqcap_{A \in S} \exists r.A \sqcap \bigsqcap_{1 \le i \le m} \exists r.B_i)^{\mathcal{I}}$ and $\mathcal{I} \models \mathcal{T}$.

Conversely, suppose that there is a model $\mathcal{I}$ of $\mathcal{T}$ based on a frame $\mathfrak{F} \in \mathcal{K}$ and such that $v \in (B \sqcap \bigsqcap_{A \in S} \exists r.A \sqcap \bigsqcap_{1 \le i \le m} \exists r.B_i)^{\mathcal{I}}$. By the choice of m and (a), v has exactly m r-successors, say $w_1, \dots, w_m$, such that $w_i \in B_i^{\mathcal{I}}$. Now let

$$S_1 = \{A \in S \mid w_1 \in A^{\mathcal{I}}\}, \qquad S_2 = \{A \in S \setminus S_1 \mid w_2 \in A^{\mathcal{I}}\}.$$

By (b), $A^{\mathcal{I}} \cap \{w_1, w_2\} \neq \emptyset$ for any $A \in S$, and so S_1, S_2 is a partition of S. We show that S_1, S_2 is a splitting of (S, I). Indeed, let $G \in I$. By (c), there are $A_1, A_2 \in G$ such that $w_1 \notin A_1^{\mathcal{I}}$, $w_2 \in A_1^{\mathcal{I}}$ and $w_2 \notin A_2^{\mathcal{I}}$, $w_1 \in A_2^{\mathcal{I}}$, i.e., $A_1 \in S_2$ and $A_2 \in S_1$. □

As the set splitting problem is NP-complete, $\mathrm{Th}_T(\mathcal{K})$ is coNP-hard. □

Note that this proof of coNP-hardness goes through for many other constraints:

Theorem 4.4 *Let $\mathcal{K}$ be an R-frame condition such that there are $r \in R$ and $n \geq 2$ for which (i) no point in frames from $\mathcal{K}$ has $> n$ r-successors, and (ii) at least one point in a frame from $\mathcal{K}$ has ≥ 2 r-successors. Then $Th_T(\mathcal{K})$ is* coNP*-hard.*

5 P/coNP-hardness dichotomy for quasi-order constraints

The aim of this section is to start analysing of the border between tractability and intractability of subsumption for classes of *quasi-orders*, i.e., reflexive and transitive frames. Throughout, we assume that $R = \{r\}$ and omit R from our terminology. A *cluster* in a quasi-order $\mathfrak{F}$ is a set of the form $\{v \mid (u, v), (v, u) \in r^{\mathfrak{F}}\}$, for some $u \in \Delta^{\mathfrak{F}}$. Single-point clusters are called *simple*. A *partial order* is a quasi-order in which all clusters are simple. A quasi-order is called *Noetherian* if it is a partial-order without infinite ascending chains.

The main result to be proved in this section is the following:

Theorem 5.1 *Let $\mathcal{K} \neq \emptyset$ be a class of quasi-orders closed under isomorphic copies.*

(a) *If $\mathcal{K}$ is universal, then* $\mathrm{Th}_T(\mathcal{K})$ *is in P if one of the following holds:*

(a.1) *$\mathcal{K}$ is TBox-equivalent to the class of all quasi-orders;*

(a.2) *$\mathcal{K}$ is TBox-equivalent to the class of equivalence relations;*

(a.3) *$\mathcal{K}$ is TBox-equivalent to the singleton class consisting of a single-point frame.*

If none of (a.1)–(a.3) *holds then* $\mathrm{Th}_T(\mathcal{K})$ *is* coNP*-hard.*

(b) *If $\mathcal{K}$ is a class of Noetherian partial orders (e.g., a class of finite partial orders) closed under subframes, then* $\mathrm{Th}_T(\mathcal{K})$ *is in P if one of the following holds:*

(b.1) *$\mathcal{K}$ is TBox-equivalent to the class of all Noetherian partial orders;*

(b.2) *$\mathcal{K}$ is TBox-equivalent to the singleton class consisting of a single-point frame.*

If neither (b.1) *nor* (b.2) *holds then* $\mathrm{Th}_T(\mathcal{K})$ *is* coNP*-hard.*

Remark 5.2 Observe that there are uncountably many distinct $\mathrm{Th}_T(\mathcal{K})$, $\mathcal{K}$ a universal class of quasi-orders, and exactly three of them are in P. This follows from Theorem 3.4 and the fact that there are uncountably many distinct universal modally definable classes of quasi-orders [17]. The same applies to classes of Noetherian partial orders. To show this, one can again observe that there are uncountably many modally definable classes of Noetherian quasi-orders closed under subframes [17] and prove that they are non-TBox equivalent by using their finite model property [6] and the finite TBoxes $\mathcal{T}_S(\mathfrak{F})$ for finite rooted $\mathfrak{F}$.

We begin by proving the lower complexity bounds mentioned in Theorem 5.1.

5.1 *Proof of Theorem 5.1: lower bounds*

A *finite transitive tree* is a finite rooted partial order in which every point except the root has exactly one immediate predecessor. A finite *tree of clusters* is obtained from a finite transitive tree by replacing its points with finite clusters.

Lemma 5.3 *Let* $\mathcal{K}$ *be a nonempty class of quasi-orders closed under subframes. Suppose that there is a finite transitive tree* $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$ *such that* $|\Delta^{\mathfrak{F}}| \geq 3$ *and every proper subframe of* $\mathfrak{F}$ *is in* $\mathrm{FrTh}_T(\mathcal{K})$*. Then* $\mathrm{Th}_T\mathcal{K}$ *is* coNP*-hard.*

Proof. The proof is again by reduction of the set splitting problem. Suppose that we are given a family I of subsets of a finite set S. As before, we assume that the elements of S are concept names. Two cases are possible.

Case 1: $\mathfrak{F}$ *contains a point* w_1 *with exactly one successor* w_2*, which is a leaf.* Denote by $\mathfrak{F}'$ the tree obtained from $\mathfrak{F}$ by removing the leaf w_2. Then $\mathfrak{F}' \in \mathrm{FrTh}_T\mathcal{K}$. Denote by w the immediate predecessor of w_1 in $\mathfrak{F}'$; it must exist because $|\Delta^{\mathfrak{F}}| \geq 3$. Denote by w_0 the root of $\mathfrak{F}'$ and consider the TBox $\mathcal{T}$ containing the following CIs:

- $\mathcal{T}_S(\mathfrak{F}')$ defined in Section 3;
- $A \sqcap \exists r.A_{w'} \sqsubseteq \exists r.A_w$, for $(w,w') \in r^{\mathfrak{F}'}$, $w' \neq w_1$, $A \in S$;
- $A_w \sqsubseteq \exists r.(A \sqcap \exists r.A_{w_1})$ for $A \in S$;
- $\exists r.(A \sqcap \exists r.A_w) \sqcap \exists r.(A_{w_1} \sqcap \exists r.A) \sqsubseteq \bot$, for $A \in S$;
- $\displaystyle\bigsqcap_{A\in G} \exists r.(A \sqcap \exists r.A_w) \sqsubseteq \bot$, for $G \in I$;
- $\displaystyle\bigsqcap_{A\in G} \exists r.(A_{w_1} \sqcap \exists r.A) \sqsubseteq \bot$, for $G \in I$.

Intuitively, we distribute the $A \in S$ over w and w_1, which represent S_1 and S_2: if $\exists r.(A \sqcap \exists r.A_w) \neq \emptyset$ we put A in S_1, and if $\exists r.(A_{w_1} \sqcap \exists r.A) \neq \emptyset$ we put A in S_2.

Claim 5.4 *There exists a splitting of* (S,I) *iff* $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \bot$.

Proof. Let S_1, S_2 be a splitting of (S,I). Define an interpretation $\mathcal{I}$ based on $\mathfrak{F}'$ by taking $A_v^{\mathcal{I}} = \{v\}$ for $v \in \Delta^{\mathfrak{F}'}$, $w \in A^{\mathcal{I}}$ for $A \in S_1$, and $w_1 \in A^{\mathcal{I}}$ for $A \in S_2$. One can check that $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \not\models A_{w_0} \sqsubseteq \bot$, from which $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \bot$ as $\mathfrak{F}' \in \mathrm{FrTh}_T\mathcal{K}$.

Conversely, let $\mathcal{I}$ be a model of $\mathcal{T}$ based on a frame $\mathfrak{G} \in \mathcal{K}$ and let $d_0 \in A_{w_0}^{\mathcal{I}}$. Since $\mathfrak{F}'$ is a finite transitive tree, one can use Lemma 3.3 to show that there is an embedding f of $\mathfrak{F}'$ into $\mathfrak{G}$ such that $f(w_0) = d_0$, $(v, v') \in r^{\mathfrak{F}'}$ iff $(f(v), f(v')) \in r^{\mathfrak{G}}$, and $f(v) \in A_v^{\mathcal{I}}$, for all $v, v' \in \Delta^{\mathfrak{F}'}$. We claim that, for every $A \in S$, we have either $d_0 \in (\exists r.(A \sqcap \exists r.A_w))^{\mathcal{I}}$ or $d_0 \in (\exists r.(A_{w_1} \sqcap \exists r.A))^{\mathcal{I}}$. Indeed, suppose that this is not the case for some $A \in S$. Take the point $d = f(w) \in A_w$ with $(d_0, d) \in r^{\mathfrak{G}}$. By the definition of $\mathcal{T}$, we have $d \in (\exists r.(A \sqcap \exists r.A_{w_1}))^{\mathcal{I}}$, and so, in view of reflexivity of $r^{\mathfrak{G}}$ and our assumption, there must exist points d' and d'' such that $(d, d'), (d', d'') \in r^{\mathfrak{G}}$, $(d', d), (d'', d') \notin r^{\mathfrak{G}}$; $d' \in A^{\mathcal{I}}$, $d'' \in A_{w_1}$; and $d' \notin (\exists r.A_w)^{\mathcal{I}}$. As $d' \notin (\exists r.A_w)^{\mathcal{I}}$, by the definition of $\mathcal{T}$, we must have $d' \notin (\exists r.A_{w'})^{\mathcal{I}}$, for all w' with $w' \neq w_1$. Consider now the map $f' : \Delta^{\mathfrak{F}} \to \Delta^{\mathfrak{G}}$ defined by taking

$$f'(u) = \begin{cases} f(u), & \text{if } u \notin \{w_1, w_2\}; \\ d', & \text{if } u = w_1; \\ d'', & \text{if } u = w_2. \end{cases}$$

Clearly, f' is an embedding of $\mathfrak{F}$ into $\mathfrak{G}$, contrary to $\mathfrak{F} \notin \mathrm{FrTh}_{\mathcal{T}}\mathcal{K}$ and $\mathcal{K}$ being closed under subframes.

Thus, we have shown that, for every $A \in S$, either (i) $d_0 \in (\exists r.(A \sqcap \exists r.A_w))^{\mathcal{I}}$ or (ii) $d_0 \in (\exists r.(A_{w_1} \sqcap \exists r.A))^{\mathcal{I}}$, but not both, as stated in the definition of $\mathcal{T}$. Define S_1 and S_2 by putting A in the former if (i) holds and in the latter if (ii) holds. The last two items in the definition of $\mathcal{T}$ guarantee that S_1, S_2 is a splitting of (S, I). □

The complement of Case 1 is the following:

Case 2: *$\mathfrak{F}$ contains a point w with at least two successors, with all successors of w being leaves.* Take a proper successor w_3 of w and denote by $\mathfrak{F}'$ the frame obtained from $\mathfrak{F}$ by removing w_3. Let w_1 be one of the remaining successors of w in $\mathfrak{F}'$. Denote by $\mathfrak{F}''$ the frame obtained from $\mathfrak{F}'$ by adding a fresh successor w_2 to w_1. Clearly, both $\mathfrak{F}'$ and $\mathfrak{F}''$ are finite transitive trees; as before, we denote by w_0 the root of $\mathfrak{F}''$. Two cases are possible now.

Case 2.1: $\mathfrak{F}'' \in \mathrm{FrTh}_{\mathcal{T}}\mathcal{K}$. To encode set splitting for (S, I), we need additional concept names $\bar{A}$, for $A \in S$. This time the intuition behind the encoding is as follows: $A \in S_1$ will be encoded by $\exists r(A' \sqcap \exists r.\bar{A}')$ and $A \in S_2$ by $\exists r.(\bar{A}' \sqcap \exists r.A')$, where $A' = A_{w_1} \sqcap A$ and $\bar{A}' = A_{w_1} \sqcap \bar{A}$. Let $\mathcal{T}$ be the TBox with the following CIs:

- $\mathcal{T}_S(\mathfrak{F}'')$;
- $A_w \sqsubseteq \exists r.A'$, for $A \in S$;
- $A_w \sqsubseteq \exists r.\bar{A}'$, for $A \in S$;
- $\exists r.(A' \sqcap \exists r.\bar{A}') \sqcap \exists r.(\bar{A}' \sqcap \exists r.A') \sqsubseteq \bot$. for $A \in S$;
- $\displaystyle\bigsqcap_{A \in G} \exists r.(A' \sqcap \exists r.\bar{A}') \sqsubseteq \bot$, for $G \in I$;
- $\displaystyle\bigsqcap_{A \in G} \exists r.(\bar{A}' \sqcap \exists r.A') \sqsubseteq \bot$, for $G \in I$.

Claim 5.5 *There exists a splitting of (S, I) iff $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \bot$.*

Proof. Suppose S_1, S_2 is a splitting of (S, I). Define an interpretation $\mathcal{I}$ based on $\mathfrak{F}''$ by taking $A_v^{\mathcal{I}} = \{v\}$ for $v \in \Delta^{\mathfrak{F}''} \setminus \{w_1, w_2\}$, $A_{w_1}^{\mathcal{I}} = \{w_1, w_2\}$, $w_1 \in A^{\mathcal{I}}$ and $w_2 \in \bar{A}^{\mathcal{I}}$ for $A \in S_1$, $w_2 \in A^{\mathcal{I}}$ and $w_1 \in \bar{A}^{\mathcal{I}}$ for $A \in S_2$. It is readily checked that

$\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \bot$. Thus, $\mathcal{T} \not\models_{\mathcal{K}} A_{w_0} \sqsubseteq \bot$.

Conversely, let $\mathcal{I}$ be a model of $\mathcal{T}$ based on a frame $\mathfrak{G} \in \mathcal{K}$ and $d_0 \in A_{w_0}^{\mathcal{I}}$. Since $\mathfrak{F}'$ is a finite transitive tree, there is an embedding f of $\mathfrak{F}'$ into $\mathfrak{G}$ such that $f(w_0) = d_0$, $(v, v') \in r^{\mathfrak{F}'}$ iff $(f(v), f(v')) \in r^{\mathfrak{G}}$ and $f(v) \in A_v^{\mathcal{I}}$ for all $v, v' \in \Delta^{\mathfrak{F}'}$. We claim that, for every $A \in S$, either $d_0 \in (\exists r.(A' \sqcap \exists r.\bar{A}'))^{\mathcal{I}}$ or $d_0 \in (\exists r.(\bar{A}' \sqcap \exists r.A'))^{\mathcal{I}}$. Indeed, assume this is not so for $A \in S$. Let $d = f(w) \in A_w$ with $(d_0, d) \in r^{\mathfrak{G}}$. Then there are $r^{\mathfrak{G}}$-incomparable $d_1, d_2 \in A_{w_1}^{\mathcal{I}}$ such that $(d, d_1), (d, d_2) \in r^{\mathfrak{G}}$. Now we modify f to a map f' from $\mathfrak{F}$ into $\mathfrak{G}$ by taking $f'(w_1) = d_1$ and $f'(w_3) = d_2$, where w_3 is the point removed from $\mathfrak{F}$ in the definition of $\mathfrak{F}'$. Clearly, f' is an embedding of $\mathfrak{F}$ into $\mathfrak{G}$, contrary to $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$ and $\mathcal{K}$ being closed under subframes. □

Case 2.2: $\mathfrak{F}'' \notin \mathrm{FrTh}_T\mathcal{K}$. As $\mathfrak{F}' \in \mathrm{FrTh}_T\mathcal{K}$, we can deal with $\mathfrak{F}''$ in precisely the same way as in Case 1. □

To complete the proof of the lower bounds we require two more lemmas.

Lemma 5.6 *Let $\mathcal{K}$ be a nonempty class of Noetherian partial orders closed under subframes. If neither* (b.1) *nor* (b.2) *holds, then there is a finite transitive tree $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$ such that $|\Delta^{\mathfrak{F}}| \geq 3$ and every proper subframe of $\mathfrak{F}$ is in* $\mathrm{FrTh}_T(\mathcal{K})$.

Proof. Since (b.1) does not hold, we have $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$, for some $\mathcal{T}$ and C, D, such that $\mathcal{T} \not\models_{\mathcal{K}'} C \sqsubseteq D$, where $\mathcal{K}'$ is the class of all Noetherian partial orders. The polynomial time algorithm for $\mathrm{Th}_T(\mathcal{K}')$ presented below shows that we can find a finite interpretation $\mathcal{I}$ based on a Noetherian partial order such that $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \not\models C \sqsubseteq D$. (This can also be proved using the finite model property of **Grz**.) Further, using the unravelling argument, we can find a finite transitive tree $\mathfrak{F}$ such that $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$ but $\mathfrak{F}' \in \mathrm{FrTh}_T\mathcal{K}$ for all proper subtrees $\mathfrak{F}'$ of $\mathfrak{F}$.

If $\mathfrak{F}$ is a single-point frame then $\mathfrak{F}$ is a p-morphic image of any quasi-order, and so we must have $\mathcal{K} = \emptyset$, which is a contradiction. Suppose next that $\mathfrak{F}$ is a two-point chain. Then $\mathfrak{F}$ is a subframe of any rooted Noetherian frame with at least two points, and so $\mathcal{K}$ is TBox-equivalent to a single-point frame, contrary to our assumption that (b.2) does not hold. It follows that $|\Delta^{\mathfrak{F}}| \geq 3$. □

The proof of the next lemma is similar; it is given in Section C.

Lemma 5.7 *Let $\mathcal{K} \neq \emptyset$ be a universal class of quasi-orders. If none of* (a.1)–(a.3) *holds, then $\mathcal{K}$ is either tabular and non-functional or there is a finite transitive tree $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$ such that $|\Delta^{\mathfrak{F}}| \geq 3$ and every proper subframe of $\mathfrak{F}$ is in* $\mathrm{FrTh}_T(\mathcal{K})$.

5.2 Proof of Theorem 5.1: upper bounds

Recall from Theorem 4.1 that if $\mathcal{K}$ is tabular and not functional, then $\mathrm{Th}_T\mathcal{K}$ is coNP-hard. Thus, to prove Theorem 5.1 it remains to show that $\mathrm{Th}_T\mathcal{K}$ is in P whenever $\mathcal{K}$ is (i) the singleton class with a singe-point frame, (ii) the class of equivalence relations, (iii) the class of Noetherian partial orders or (iv) the class of quasi-orders. Case (i) is trivial as we clearly have $\exists r.C \equiv C$, and so deciding $\mathrm{Th}_T\mathcal{K}$ reduces to propositional Horn logic. For (iv), we can use a straightforward modification of the polynomial time algorithm for transitive frames [1]. Let us consider (ii) and (iii).

Without loss of generality we will assume that all TBoxes $\mathcal{T}$ in this section are *normalised*: in every $C \sqsubseteq D \in \mathcal{T}$, D is either a concept name or of the form $\exists r.A$, for a concept name A, and in every subconcept $\exists r.E$ of C, E is a concept name. Moreover, when deciding whether $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$ we can assume that C is a concept name. An easy reduction of the general subsumption problem to this case by adding 'abbreviations' $A \equiv C$ to TBoxes can be found in [1].

Suppose that we are given a normalised TBox $\mathcal{T}$ and a concept name A_0. Consider first the case when $\bot$ does not occur in $\mathcal{T}$.

Denote by *sub*$(\mathcal{T})$ the set of subconcepts of concepts in $\mathcal{T}$. The following *canonical interpretation* $\mathcal{I}_{\mathcal{T},A_0}$ is constructed in [1,11]. First, define $\mathcal{I}_0$ by taking

$$\Delta^{\mathcal{I}_0} \;=\; \{d_{A_0}\} \cup \{d_A \mid \exists r.A \in sub(\mathcal{T})\},$$

where the d_A and d_{A_0} are fresh objects. Set $d \in A^{\mathcal{I}_0}$ iff $d = d_A$, for all $d_A \in \Delta^{\mathcal{I}_0}$, and $r^{\mathcal{I}_0} = \emptyset$. Next, we apply exhaustively the following two rules to $\mathcal{I} := \mathcal{I}_0$:

- for $C \sqsubseteq A \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}_0}$, if $d \in C^{\mathcal{I}}$ and $d \notin A^{\mathcal{I}}$, then update $\mathcal{I}$ by setting $A^{\mathcal{I}} := A^{\mathcal{I}} \cup \{d\}$ and leaving the interpretation of all remaining symbols unchanged;
- for $C \sqsubseteq \exists r.A \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}_0}$, if $d \in C^{\mathcal{I}}$ and $d \notin (\exists r.A)^{\mathcal{I}}$, then update $\mathcal{I}$ by setting $r^{\mathcal{I}} := r^{\mathcal{I}} \cup \{(d, d_A)\}$ and leaving it unchanged for the remaining symbols.

The resulting interpretation is denoted by $\mathcal{I}_{\mathcal{T},A_0}$. Clearly, it can be constructed in polynomial time. To characterise $\mathcal{I}_{\mathcal{T},A_0}$, we require the notion of simulation between interpretations. A relation $S \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is a *simulation* between interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ if the following conditions hold:

(i) for all concept names A and all $(e_1, e_2) \in S$, if $e_1 \in A^{\mathcal{I}_1}$ then $e_2 \in A^{\mathcal{I}_2}$;

(ii) for all role names r, all $(e_1, e_2) \in S$ and all $e_1' \in \Delta^{\mathcal{I}_1}$ with $(e_1, e_1') \in r^{\mathcal{I}_1}$, there exists $e_2' \in \Delta^{\mathcal{I}_2}$ such that $(e_2, e_2') \in r^{\mathcal{I}_2}$ and $(e_1', e_2') \in S$.

For interpretations $\mathcal{I}_1, \mathcal{I}_2$ with $d_1 \in \Delta^{\mathcal{I}_1}$, $d_2 \in \Delta^{\mathcal{I}_2}$, we write $(\mathcal{I}_1, d_1) \leq (\mathcal{I}_2, d_2)$ and say that $(\mathcal{I}_1, d_1)$ is *simulated* by $(\mathcal{I}_2, d_2)$ if there is a simulation S between $\mathcal{I}_1$ and $\mathcal{I}_2$ such that $(d_1, d_2) \in S$. We require the following results from [11]:

Lemma 5.8 **(sim)** *If* $(\mathcal{I}_1, d_1) \leq (\mathcal{I}_2, d_2)$ *and* $d_1 \in C^{\mathcal{I}_1}$ *then* $d_2 \in C^{\mathcal{I}_2}$*, for any* C*.*

(can) $\mathcal{I}_{\mathcal{T},A_0} \models \mathcal{T}$ *and, for all* $\mathcal{I}$ *with* $\mathcal{I} \models \mathcal{T}$ *and all* $d_A \in \Delta^{\mathcal{I}_{\mathcal{T},A_0}}$, $d \in A^{\mathcal{I}}$*, we have* $(\mathcal{I}_{\mathcal{T},A_0}, d_A) \leq (\mathcal{I}, d)$*.*

It follows, in particular, that for all concepts D, $\mathcal{T} \models A_0 \sqsubseteq D$ iff $d_{A_0} \in D^{\mathcal{I}_{\mathcal{T},A_0}}$. Thus, we immediately obtain a polynomial time decision procedure for $\mathrm{Th}_T\mathcal{K}$, $\mathcal{K}$ the class of all frames, because $d_{A_0} \in D^{\mathcal{I}_{\mathcal{T},A_0}}$ can be checked in polynomial time.

We now generalise this procedure to the class of equivalence relations. Set

$$\mathfrak{E}_n \;=\; (\{1, \dots, n\}, r^{\mathfrak{E}_n} \;=\; \{1, \dots, n\} \times \{1, \dots, n\}), \quad \mathfrak{E}_\omega = (\omega, r^{\mathfrak{E}_\omega} = \omega \times \omega).$$

Lemma 5.9 *Given* A_0 *and* $\mathcal{T}$ *not containing* $\bot$*, one can construct in polynomial time, starting from* $\mathcal{I}_{\mathcal{T},A_0}$*, an interpretation* $\mathcal{I}^e_{\mathcal{T},A_0}$ *based on some* $\mathfrak{E}_n$ *such that* (i) $\mathcal{I}^e_{\mathcal{T},A_0} \models \mathcal{T}$ *and* $d_{A_0} \in A_0^{\mathcal{I}^e_{\mathcal{T},A_0}}$*, and* (ii) *if* $\mathcal{J}$ *is an interpretation based on* $\mathfrak{E}_\omega$ *with* $d \in A_0^{\mathcal{J}}$*, then* $(\mathcal{I}^e_{\mathcal{T},A_0}, d_{A_0}) \leq (\mathcal{J}, d)$*.*

Proof. Given an interpretation $\mathcal{I}$ and $d \in \Delta^{\mathcal{I}}$, we define a new interpretation $\mathcal{I}^d_\sim$ which coincides with $\mathcal{I}$ except that $(e_1, e_2) \in r^{\mathcal{I}^d_\sim}$, for all e_1, e_2 reachable from d via an $r^{\mathcal{I}}$-path $d_1, \dots, d_n$ with $d = d_1$ and $(d_i, d_{i+1}) \in r^{\mathcal{I}}$ for $i < n$. We now apply exhaustively the following rules to $\mathcal{I} = \mathcal{I}_{\mathcal{T},A_0}$:

(s1) if $\mathcal{I} \neq \mathcal{I}_{\sim}^{d_{A_0}}$ then set $\mathcal{I} := \mathcal{I}_{\sim}^{d_{A_0}}$;

(s2) for $C \sqsubseteq A \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$, if $d \in C^{\mathcal{I}}$ and $d \notin A^{\mathcal{I}}$ then update $\mathcal{I}$ by setting $A^{\mathcal{I}} := A^{\mathcal{I}} \cup \{d\}$ and leaving the interpretation of all remaining symbols unchanged;

(s3) for $C \sqsubseteq \exists r.A \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$, if $d \in C^{\mathcal{I}}$ and $d \notin (\exists r.A)^{\mathcal{I}}$ then update $\mathcal{I}$ by setting $r^{\mathcal{I}} := r^{\mathcal{I}} \cup \{(d, d_A)\}$ and leaving it unchanged for the remaining symbols.

Denote by $\mathcal{I}^e_{\mathcal{T},A_0}$ the restriction of the resulting interpretation to the subframe generated by d_{A_0}. Clearly, it can be constructed in polynomial time. One can show that $\mathcal{I}^e_{\mathcal{T},A_0}$ is as required (for details see Section D). □

Using Lemma 5.9, we can decide whether $\mathcal{T} \models_{\mathfrak{C}_\omega} A_0 \sqsubseteq D$ by checking, in polynomial time, whether $d_{A_0} \in D^{\mathcal{I}^e_{\mathcal{T},A_0}}$. If $\mathcal{T}$ contains $\bot$, we replace every occurrence of $\bot$ in $\mathcal{T}$ by the concept name $A_\bot$ and denote the resulting TBox by $\mathcal{T}^\bot$. By Lemma 5.9, the following conditions are equivalent:

– $\mathcal{T}^\bot \models_{\mathfrak{C}_\omega} A_0 \sqsubseteq \exists r^n.A_\bot$ for some n;

– $A_\bot^{\mathcal{I}^e_{\mathcal{T}^\bot,A_0}} \neq \emptyset$;

– $\mathcal{T} \models A_0 \sqsubseteq \bot$.

Thus, $\mathcal{T} \models A_0 \sqsubseteq D$ iff $\mathcal{T}^\bot \models A_0 \sqsubseteq D$ or $A_\bot^{\mathcal{I}^e_{\mathcal{T}^\bot,A_0}} \neq \emptyset$, and both conditions can be checked in polynomial time.

Consider now Noetherian partial orders.

Lemma 5.10 *Given A_0 and $\mathcal{T}$ not containing $\bot$, one can construct in polynomial time, starting from $\mathcal{I}_{\mathcal{T},A_0}$, an interpretation $\mathcal{I}^N_{\mathcal{T},A_0}$ based on a finite partial order with root $d^*_{A_0}$ such that* (i) *$\mathcal{I}^N_{\mathcal{T},A_0} \models \mathcal{T}$ and $d^*_{A_0} \in A_0^{\mathcal{I}^N_{\mathcal{T},A_0}}$, and* (ii) *if $\mathcal{J}$ is based on a partial order and $d \in A_0^{\mathcal{J}}$, then $(\mathcal{I}^N_{\mathcal{T},A_0}, d^*_{A_0}) \leq (\mathcal{J}, d)$.*

Proof. Let $\mathcal{I}^+_{\mathcal{T},A_0}$ be the interpretation obtained from $\mathcal{I}_{\mathcal{T},A_0}$ by adding a copy $d^*_{A_0}$ of d_{A_0} to its domain. More precisely, we set $(d^*_{A_0}, d) \in r^{\mathcal{I}^+_{\mathcal{T},A_0}}$ whenever $(d_{A_0}, d) \in r^{\mathcal{I}_{\mathcal{T},A_0}}$ or $d = d^*_{A_0}$ (note that $d^*_{A_0}$ has no proper predecessors). We set $d_{\{A\}} = d_A$, for all $d_A \in \Delta^{\mathcal{I}_{\mathcal{T},A_0}}$, and define two operators on interpretations $\mathcal{I}$ whose domains consist of points d_X, where X is a nonempty set of concept names, and the point $d^*_{A_0}$.

First, define $\mathcal{I}^*$ by replacing $r^{\mathcal{I}}$ with its transitive and reflexive closure $r^{\mathcal{I}^*}$. Second, if $r^{\mathcal{I}}$ is transitive and reflexive and $d \in \Delta^{\mathcal{I}}$, then define $\mathcal{I}_d$ by removing the cluster $[d] = \{d' \in \Delta^{\mathcal{I}} \mid (d, d'), (d', d) \in r^{\mathcal{I}}\}$ generated by d from $\mathcal{I}$, replacing it with a single point d_X, where $X = \bigcup_{d_Y \in [d]} Y$, and setting $d_X \in A^{\mathcal{I}_d}$ iff $d' \in A^{\mathcal{I}}$, for some $d' \in [d]$. (This operation has no effect for singleton $[d]$.) Now, we apply exhaustively the following rules to $\mathcal{I} = \mathcal{I}_{\mathcal{T},A_0}$:

(r1) if $r^{\mathcal{I}}$ is transitive and reflexive and $\mathcal{I} \neq \mathcal{I}_d$, for some $d \in \Delta^{\mathcal{I}}$, then set $\mathcal{I} := \mathcal{I}_d$;

(r2) if $\mathcal{I} \neq \mathcal{I}^*$ then set $\mathcal{I} := \mathcal{I}^*$;

(r3) for $C \sqsubseteq A \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$, if $d \in C^{\mathcal{I}}$ and $d \notin A^{\mathcal{I}}$ then update $\mathcal{I}$ by setting $A^{\mathcal{I}} := A^{\mathcal{I}} \cup \{d\}$ and leaving the interpretation of all remaining symbols unchanged;

(r4) for $C \sqsubseteq \exists r.A \in \mathcal{T}$ and $d \in \Delta^{\mathcal{I}}$, if $d \in C^{\mathcal{I}}$ and $d \notin (\exists r.A)^{\mathcal{I}}$ then update $\mathcal{I}$ by setting $r^{\mathcal{I}} := r^{\mathcal{I}} \cup \{(d, d_X)\}$ for the (unique) X with $A \in X$ and leaving the interpretation of all remaining symbols unchanged.

Denote by $\mathcal{I}^N_{\mathcal{T},A_0}$ the restriction of the resulting interpretation to the subframe generated by $d^*_{A_0}$. One can show that $\mathcal{I}^N_{\mathcal{T},A_0}$ is as required; see Section D. □

We now apply Lemma 5.10, in the same way as Lemma 5.9, to obtain decidability of $\mathrm{Th}_T\mathcal{K}$, $\mathcal{K}$ the class of Noetherian partial orders, for TBoxes with and without $\bot$.

6 Future directions

Our primary aim in this paper was to start investigating—from a purely theoretical standpoint—the difference between tractable and intractable relational constraints in the context of the sub-Boolean DL $\mathcal{EL}$ (a finer classification of the intractable constraints could also be very interesting). As a next step, one can consider classes of transitive frames or general frame conditions closed under subframes. We note, however, that even for classes of *irreflexive* transitive frames without infinite ascending chains (aka Noetherian transitive frames) closed under subframes, the dichotomy appears to be much more involved than for Noetherian partial orders. For example, using the technique developed above one can show that $\mathrm{Th}_T\mathcal{K}$ is in P not only for the class $\mathcal{K}$ of all such frames (the $\mathcal{EL}$ analogue of **GL**) but also for the class of irreflexive transitive frames of depth $\leq n$, for any $n < \omega$. We conjecture that there are other 'polynomial classes' of Noetherian transitive frames.

Although DLs come equipped with the intended semantics, generalisations to the algebraic setting would also be of interest. Earlier in the paper, we have given first 'correspondence' results for $\mathcal{EL}$, aiming to demonstrate the type of relational constraints 'visible' to $\mathcal{EL}$. It turned out that essentially all 'standard' modal conditions were TBox definable. Here are two more illustrative examples:

- $\Gamma_0 = \{(\emptyset, \exists r.\exists r.A \sqsubseteq \exists r.A), (\mathcal{T}_S(\circ), A_w \sqsubseteq \bot)\}$ defines the class of Noetherian transitive frames ($\circ$ is a single reflexive point w);
- $\Gamma_1 = \{(\emptyset, \exists r.\exists r.A \sqsubseteq \exists r.A), (\emptyset, A \sqsubseteq \exists r.A), (\mathcal{T}_S(\circledcirc\circledcirc), A_w \sqsubseteq \bot)\}$ defines the class of Noetherian partial orders ($\circledcirc\circledcirc$ is a two-point cluster containing w).

Despite the insights provided by such results, their applicability is somewhat limited. The main problem is that correspondence alone does not build a bridge between the algebraic/syntactic and the first-order views of modal logic. Ideally, correspondence results should come together with *completeness* results, like in Sahlqvist's theorem [12]. For instance, we would like to know whether the Γ_i above actually *axiomatise* (in some equational or Hilbert-style calculus) the classes of frames they define. Unfortunately, but not surprisingly, the gap between correspondence and completeness in $\mathcal{EL}$ is even wider than in classical modal logic. To be a bit more precise, we can regard $\mathcal{EL}$-concepts to be *terms* in the language of bounded semi-lattices with monotone operators (see, e.g., [13]). Then every CI $C \sqsubseteq D$ can be identified with the *identity* $C \sqcap D = D$, and every R-consequence $(\mathcal{T}, C' \sqsubseteq D')$ with the *quasi-identity*

$$\bigsqcap_{C \sqsubseteq D \in \mathcal{T}} C \sqcap D = D \quad \Rightarrow \quad C' \sqcap D' = D'.$$

Now, we call a set Γ of R-consequences *complete* (for relational models) if, for every R-consequence $q = (\mathcal{T}, C \sqsubseteq D)$, we have $\mathcal{T} \models_{\mathrm{Fr}\Gamma} C \sqsubseteq D$ iff q is valid in all bounded semi-lattices with monotone operators validating Γ. $\Gamma = \emptyset$ was shown to be complete in [13] by reduction of TBox reasoning in $\mathcal{EL}$ to validity of quasi-identities in semi-lattices with distributive operators. It is also shown in [13] that $(\emptyset, \{\exists r_1. \cdots \exists r_n.A \sqsubseteq \exists r_{n+1}.A\})$ is complete for the R-frames defined in (1). However, numerous completeness questions (e.g., for Γ_0 and Γ_1 above) remain open.

The P/NP dichotomy problem can be extended to the algebraic setting. It is to be noted, however, that there are 'many more' quasi-varieties of semilattices with monotone operators than TBox non-equivalent relational constraints. In contrast to modal logic, this is already the case for tabular logics. Indeed, consider the 3-element set-semilattice $\{\emptyset, \{a\}, \{a, b\}\}$ with $\Diamond_r(\emptyset) = \emptyset$, $\Diamond_r(\{a\}) = \emptyset$ and $\Diamond_r(\{a, b\}) = \{a\}$ induced by the 2-element irreflexive $\{r\}$-frame $\mathfrak{F} = (\{a, b\}, r^{\mathfrak{F}} = \{(a, b)\})$. This semilattice $\mathfrak{A}$ validates $(\emptyset, \exists r.A \sqsubseteq A)$. One can readily show that (i) the TBox theory corresponding to the quasi-variety generated by $\mathfrak{A}$ is 'incomplete' for relational models as it is not TBox equivalent to any TBox theory of any class of $\{r\}$-frames, and (ii) that $\{(\emptyset, \exists r.A \sqsubseteq A)\}$ is not complete either. Thus, even simple Sahlqvist inequalities such as $\Diamond x \leq x$ become incomplete when added as axioms to the theory of bounded semilattices with monotone operators. It follows that we cannot obtain dichotomy results for (even tabular) quasi-varieties of semi-lattices with monotone operators as immediate consequences of the results presented in this paper.

References

[1] Baader, F., S. Brandt and C. Lutz, *Pushing the EL envelope*, in: L. P. Kaelbling and A. Saffiotti, editors, *IJCAI* (2005), pp. 364–369.

[2] Baader, F., S. Brandt and C. Lutz, *Pushing the EL envelope further*, in: K. Clark and P. F. Patel-Schneider, editors, *In Proc. of OWLED*, 2008.

[3] Baader, F. and C. Lutz, *Description logic*, in: P. Blackburn, J. van Benthem and F. Wolter, editors, *The Handbook of Modal Logic*, Elsevier, 2007 pp. 757–820.

[4] Bauland, M., E. Hemaspaandra, H. Schnoor and I. Schnoor, *Generalized modal satisfiability*, in: *STACS*, 2006, pp. 500–511.

[5] Donini, F. M., M. Lenzerini, D. Nardi and W. Nutt, *The complexity of concept languages*, in: *KR*, 1991, pp. 151–162.

[6] Fine, K., *Logics containing* **K4***, part II*, Journal of Symbolic Logic **50** (1985), pp. 619–651.

[7] Garey, M. R. and D. S. Johnson, "Computers and Intractability: A Guide to the Theory of NP-Completeness," W. H. Freeman & Co., 1979.

[8] Goranko, V. and S. Passy, *Using the universal modality: gains and questions*, J. Log. Comput. (1992), pp. 5–30.

[9] Hemaspaandra, E., *The complexity of poor man's logic*, J. Log. Comput. **11** (2001), pp. 609–622.

[10] Horrocks, I., *Using an expressive description logic: Fact or fiction?*, in: *KR*, 1998, pp. 636–649.

[11] Lutz, C. and F. Wolter, *Deciding inseparability and conservative extensions in the description logic el*, J. Symb. Comput. **45** (2010), pp. 194–228.

[12] Sahlqvist, H., *Completeness and correspondence in the first and second order semantics for modal logic*, in: S. Kanger, editor, *Proc. of the 3rd Scandinavian Logic Symp.*, North-Holland, 1975 pp. 110–143.

[13] Sofronie-Stokkermans, V., *Locality and subsumption testing in EL and some of its extensions*, in: C. Areces and R. Goldblatt, editors, *Advances in Modal Logic* (2008), pp. 315–339.

[14] Tarski, A., *Contributions to the theory of models I, II*, Indag. Mathematicae **16** (1954), pp. 572–588.

[15] van Benthem, J., *Correspondence theory*, in: D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, vol.2*, Reidel, Dordrecht, 1984 pp. 167–247.

[16] Wolter, F., *The structure of lattices of subframe logics*, Annals of Pure and Applied Logic **86** (1997), pp. 47–100.

[17] Zakharyaschev, M., *Canonical formulas for K*4*. Part II: Cofinal subframe logics*, Journal of Symbolic Logic **61** (1996), pp. 421–449.

A Church-Rosser property is not TBox definable

To show that the Church-Rosser property is not TBox definable, we prove a more general closure property. Call a subframe $\mathfrak{F}'$ of $\mathfrak{F}$ *downward closed* if whenever $v \in \Delta^{\mathfrak{F}'}$ and $(v', v) \in r^{\mathfrak{F}}$ then $v' \in \Delta^{\mathfrak{F}'}$.

Lemma A.1 *TBox definable $\{r\}$-frame conditions are closed under downward closed subframes of Noetherian partial orders.*

Proof. Suppose that $\mathfrak{F} \in \mathrm{Fr}\Gamma$ is a Noetherian partial order and $\mathfrak{F}'$ is a downward closed subframe of $\mathfrak{F}$. Assume also that $\mathcal{I}'$ is based on $\mathfrak{F}'$, $\mathcal{I}' \models \mathcal{T}$ and $\mathcal{I}' \not\models C' \sqsubseteq D'$. We have to show that there exists a model $\mathcal{I}$ based on $\mathfrak{F}$ such that $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \not\models C' \sqsubseteq D'$. We construct $\mathcal{I}$ by extending $\mathcal{I}'$ to $\mathfrak{F}$ in the following way:

$$A^{\mathcal{I}} = A^{\mathcal{I}'} \cup (\Delta^{\mathfrak{F}} \setminus \Delta^{\mathfrak{F}'}), \quad \text{for all } A \text{ with } \mathcal{I}' \models \top \sqsubseteq \exists r.A.$$

For the remaining concept names A, we set $A^{\mathcal{I}} = A^{\mathcal{I}'}$. Using the condition that $\mathfrak{F}$ is Noetherian, one can prove by induction that, for all concepts C and all $v \in \Delta^{\mathfrak{F}} \setminus \Delta^{\mathfrak{F}'}$,

$$v \in C^{\mathcal{I}} \quad \text{iff} \quad \mathcal{I}' \models \top \sqsubseteq \exists r.C.$$

It follows that $v \in C^{\mathcal{I}'}$ iff $v \in C^{\mathcal{I}}$, for all $v \in \Delta^{\mathfrak{F}'}$. Moreover, suppose that there exists $v \in \Delta^{\mathfrak{F}} \setminus \Delta^{\mathfrak{F}'}$ such that $v \in C^{\mathcal{I}} \setminus D^{\mathcal{I}}$, for some C, D. Then $w \in C^{\mathcal{I}'}$, for all $w \in \Delta^{\mathfrak{F}'}$ without proper r-successors in $\Delta^{\mathfrak{F}'}$, and there exists such a w_0 with $w_0 \in D^{\mathcal{I}'}$. It follows that $\mathcal{I} \models \mathcal{T}$ and $\mathcal{I} \not\models C' \sqsubseteq D'$. □

The Church-Rosser property is not TBox definable because it is not closed under downward closed subframes of Noetherian partial orders.

B Proof of Lemma 4.2

Lemma B.1 *Given any R-concept C, one can decide in polynomial time in $|C|$ whether there exists an R-interpretation $\mathcal{J}$ such that $\mathcal{I} \leq \mathcal{J}$ and $w \in C^{\mathcal{J}}$. If such an interpretation does exist, then one can construct, again in polynomial time in $|C|$, the smallest (with respect to $\leq$) R-interpretation $\mathcal{I}(w, C) \geq \mathcal{I}$ such that $w \in C^{\mathcal{I}(w,C)}$.*

Proof. If $w \notin C^{\mathcal{I}}$, we 'saturate' $\mathcal{I}$ in the following way. Let $e(w)$ be the set of all conjuncts of C and $e(u) = \emptyset$ for $u \neq w$. If $\exists r.D \in e(u)$ and $(u, v) \in r^{\mathcal{I}}$, for some v, we remove $\exists r.D$ from $e(u)$ and add all the conjuncts of D to $e(v)$. If there is no such v, then the required interpretation does not exist. Otherwise, we repeat the construction. After at most $|C|$ steps, every $e(u)$ will either be empty or contain only atomic concepts. Then we define $\mathcal{I}(w, C)$ by taking $A^{\mathcal{I}(w,C)} = A^{\mathcal{I}} \cup \{u \mid A \in e(u)\}$, for every concept name A. □

C Proof of Lemma 5.7

Lemma C.1 *Let $\mathcal{K} \neq \emptyset$ be a universal class of quasi-orders. If none of the conditions* (a.1)–(a.3) *of Theorem 5.1 holds, then $\mathcal{K}$ is tabular and non-functional or there* *is a finite transitive tree* $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$ *such that* $|\Delta^{\mathfrak{F}}| \geq 3$ *and every proper subframe of* $\mathfrak{F}$ *is in* $\mathrm{FrTh}_T(\mathcal{K})$.

Proof. As (a.1) does not hold, there are $\mathcal{T}, C, D$ such that $\mathcal{T} \models_{\mathcal{K}} C \sqsubseteq D$ and $\mathcal{T} \not\models_{\mathcal{K}'} C \sqsubseteq D$ for the class $\mathcal{K}'$ of all quasi-orders. Using the finite model property of **S4**, one can readily show that there exists a finite interpretation $\mathcal{I}$ based on a quasi-order such that $\mathcal{I} \models \mathcal{T}$ but $\mathcal{I} \not\models C \sqsubseteq D$. Moreover, the unravelling argument provides us with a finite tree of clusters $\mathfrak{G}$ with $\mathfrak{G} \notin \mathrm{FrTh}_T\mathcal{K}$. By replacing every cluster in $\mathfrak{G}$ with an infinite ascending chain, we obtain an infinite $\mathfrak{G}' \notin \mathrm{FrTh}_T\mathcal{K}$ all rooted finite subframes of which are transitive trees. But then, using the fact that $\mathcal{K}$ is universal and employing Tarski's finite embedding property [14] (see also [6,17]), we can show that there is a finite transitive tree $\mathfrak{F}$ with $\mathfrak{F} \notin \mathrm{FrTh}_T\mathcal{K}$. Take a minimal $\mathfrak{F}$ of this kind. Now, if $\mathfrak{F}$ contains only one point then $\mathfrak{F}$ is a p-morphic image of any quasi-order, and therefore $\mathcal{K} = \emptyset$, which is a contradiction. If $\mathfrak{F}$ is a rooted frame with two points then $\mathfrak{F}$ is a subframe of every rooted quasi-order with at least two clusters. Thus, $\mathcal{K}$ can only be a class of equivalence relations. As (a.3) does not hold, $\mathcal{K}$ cannot consist of one single-point frame. It follows that either $\mathcal{K}$ is tabular and non-functional or $\mathrm{Th}_T\mathcal{K}$ is the TBox theory of all equivalence relations, contrary to our assumption that (a.2) does not hold. The only remaining case is $|\Delta^{\mathfrak{F}}| \geq 3$. □

D Proofs of Lemmas 5.9 and 5.10

Lemma D.1 *Given* A_0 *and* $\mathcal{T}$ *not containing* $\bot$*, one can construct in polynomial time, starting from* $\mathcal{I}_{\mathcal{T},A_0}$*, an interpretation* $\mathcal{I}^e_{\mathcal{T},A_0}$ *based on some* $\mathfrak{E}_n$ *such that*

(i) $\mathcal{I}^e_{\mathcal{T},A_0} \models \mathcal{T}$ *and* $d_{A_0} \in A_0^{\mathcal{I}^e_{\mathcal{T},A_0}}$*, and*

(ii) *if* $\mathcal{J}$ *is an interpretation based on* $\mathfrak{E}_\omega$ *with* $d \in A_0^{\mathcal{J}}$*, then* $(\mathcal{I}^e_{\mathcal{T},A_0}, d_{A_0}) \leq (\mathcal{J}, d)$.

Proof. Let $\mathcal{I}_{\mathcal{T},A_0} = \mathcal{I}_0, \mathcal{I}_1, \dots$ be a sequence obtained from $\mathcal{I}_{\mathcal{T},A_0}$ by applying the rules (s1), (s2), (s3). We show by induction on $n \geq 0$ that if $\mathcal{J}$ is based on $\mathfrak{E}_\omega$, $\mathcal{J} \models \mathcal{T}$ and $A_0^{\mathcal{J}} \neq \emptyset$, then the relation

$$S = \bigcup_{d_A \in \Delta^{\mathcal{I}_n}} \{(d_A, d) \mid d \in A^{\mathcal{J}}\}$$

is a simulation between $\mathcal{I}_n$ and $\mathcal{J}$. For $\mathcal{I}_0$ this follows from Lemma 5.8 (can). Now suppose that the claim holds for $\mathcal{I}_n$. Observe that $\Delta^{\mathcal{I}_n} = \Delta^{\mathcal{I}_{n+1}}$, and so the relation S does not depend on n.

Case 1: $\mathcal{I}_{n+1} = \mathcal{I}^{d_{A_0}}_{\sim}$ for $\mathcal{I} = \mathcal{I}_n$. By IH, S is a simulation between $\mathcal{I}_n$ and $\mathcal{J}$. As the interpretation of concept names coincides for $\mathcal{I}_n$ and $\mathcal{I}_{n+1}$, it is sufficient to show that, for $(d_A, d_B) \in r^{\mathcal{I}_{n+1}}$ and $(d_A, d') \in S$, there exists $d'' \in \Delta^{\mathcal{J}}$ such that $(d_B, d'') \in S$. This follows from IH if $(d_A, d_B) \in r^{\mathcal{I}_n}$. Otherwise, d_A, d_B are both reachable from d_{A_0} in $\mathcal{I}_n$. In view of $A_0^{\mathcal{J}} \neq \emptyset$ and IH, there exists d such that $(d_{A_0}, d) \in S$. Since S is a simulation between $\mathcal{I}_n$ and $\mathcal{J}$ and d_B is reachable from d_{A_0}, there exists d'' with $(d_B, d'') \in S$, as required.

Case 2: $\mathcal{I}_{n+1}$ is obtained from $\mathcal{I}_n$ using (s2). This case follows from $\mathcal{J} \models \mathcal{T}$.

Case 3: $\mathcal{I}_{n+1}$ is obtained from $\mathcal{I}_n$ using (s3). Let $C \sqsubseteq \exists r.B \in \mathcal{T}$, $d_0 \in C^{\mathcal{I}_n}$ and $r^{\mathcal{I}_{n+1}} = r^{\mathcal{I}_n} \cup \{(d_0, d_B)\}$. By IH, it is sufficient to show that if $(d_0, d) \in S$, then there exists d' with $(d_B, d') \in S$. Suppose $(d_0, d) \in S$. Since $d_0 \in C^{\mathcal{I}_n}$ and S is a simulation between $\mathcal{I}_n$ and $\mathcal{J}$, we obtain $d \in C^{\mathcal{J}}$ (Lemma 5.8). Since $\mathcal{J} \models \mathcal{T}$, there exists $d' \in \Delta^{\mathcal{J}}$ such that $d' \in B^{\mathcal{J}}$. But then $(d_B, d') \in S$, as required. □

Lemma D.2 *Given A_0 and $\mathcal{T}$ not containing $\bot$, one can construct in polynomial time, starting from $\mathcal{I}_{\mathcal{T},A_0}$, an interpretation $\mathcal{I}^N_{\mathcal{T},A_0}$ based on a finite partial order with root $d^*_{A_0}$ such that*

(i) $\mathcal{I}^N_{\mathcal{T},A_0} \models \mathcal{T}$ *and* $d^*_{A_0} \in A_0^{\mathcal{I}^N_{\mathcal{T},A_0}}$, *and*

(ii) *if $\mathcal{J}$ is based on a partial order and $d \in A_0^{\mathcal{J}}$, then* $(\mathcal{I}^N_{\mathcal{T},A_0}, d^*_{A_0}) \leq (\mathcal{J}, d)$.

Proof. Let $\mathcal{I}_{\mathcal{T},A_0} = \mathcal{I}_0, \mathcal{I}_1, \dots$ be a sequence obtained from $\mathcal{I}^+_{\mathcal{T},A_0}$ by applying the rules (r1), (r2), (r3), (r4). For a Noetherian partial order $\mathcal{J}$ and a concept name A, we set

$$m(A)^{\mathcal{J}} = \big\{d \in A^{\mathcal{J}} \mid \forall d' \; [(d' \in A^{\mathcal{J}} \land (d, d') \in r^{\mathcal{J}}) \Rightarrow d = d']\big\}$$

and call the elements of $m(A)^{\mathcal{J}}$ *maximal* in $A^{\mathcal{J}}$. We show by induction on $n \geq 0$ that, for every interpretation $\mathcal{J}$ based on a Noetherian partial order and such that $\mathcal{J} \models \mathcal{T}$,

$$S_n = \{(d^*_{A_0}, d) \mid d \in A_0^{\mathcal{J}}\} \cup \bigcup_{d_X \in \Delta^{\mathcal{I}_n}} \{(d_X, d) \mid \exists A \in X \; d \in m(A)^{\mathcal{J}}\}$$

is a simulation between $\mathcal{I}_n$ and $\mathcal{J}$, and for every $d_X \in \Delta^{\mathcal{I}_n}$, $m(A)^{\mathcal{J}} = m(B)^{\mathcal{J}}$ for all $A, B \in X$. For $\mathcal{I}_0$ this is readily shown using Lemma 5.8 (can) and the fact that $\mathcal{J}$ is a Noetherian partial order.

Case 1: $\mathcal{I}_{n+1} = \mathcal{I}_d$ for $\mathcal{I} = \mathcal{I}_n$. Let $X = \bigcup_{d_Y \in [d]} Y$. We first show that $m(A)^{\mathcal{J}} = m(B)^{\mathcal{J}}$ for all $A, B \in X$. Suppose that $d \in m(A)^{\mathcal{J}}$. Let $A \in X_1$, $B \in X_2$ be such that $d_{X_1}, d_{X_2} \in [d]$. Then $(d_{X_1}, d) \in S_n$. Since S_n is a simulation and $(d_{X_1}, d_{X_2}), (d_{X_2}, d_{X_1}) \in r^{\mathcal{I}_n}$, there exist d', d'' with $(d, d'), (d', d'') \in r^{\mathcal{J}}$ and $(d_{X_2}, d'), (d_{X_1}, d'') \in S_n$. By IH, $d' \in m(B)^{\mathcal{J}}$ and $d'' \in m(A)^{\mathcal{J}}$. Then $d = d''$ and, therefore, $d = d'$ and $d \in m(B)^{\mathcal{J}}$, as required. It is now straightforward to show that S_{n+1} is a simulation between $\mathcal{I}_{n+1}$ and $\mathcal{J}$.

Case 2: $\mathcal{I}_{n+1} = \mathcal{I}^*_n$. This case is straightforward in view of transitivity of $\mathcal{J}$.

Case 3: $\mathcal{I}_{n+1}$ is obtained from $\mathcal{I}_n$ using (r3). This case follows from $\mathcal{J} \models \mathcal{T}$.

Case 4. $\mathcal{I}_{n+1}$ is obtained from $\mathcal{I}_n$ using (r4). Let $C \sqsubseteq \exists r.B \in \mathcal{T}$, $d_0 \in C^{\mathcal{I}_n}$ and $r^{\mathcal{I}_{n+1}} = r^{\mathcal{I}_n} \cup \{(d_0, d_X)\}$, where $B \in X$. By IH, it is sufficient to show that if $(d_0, d) \in S_{n+1}$, then there exists d' with $(d, d') \in r^{\mathcal{J}}$ and $(d_X, d') \in S_{n+1}$. Suppose that $(d_0, d) \in S_{n+1}$. Then $(d_0, d) \in S_n$. Since $d_0 \in C^{\mathcal{I}_n}$ and S_n is a simulation between $\mathcal{I}_n$ and $\mathcal{J}$, we obtain $d \in C^{\mathcal{J}}$ by Lemma 5.8. Since $\mathcal{J} \models \mathcal{T}$, there exists $d' \in \Delta^{\mathcal{J}}$ such that $d' \in B^{\mathcal{J}}$ and $(d, d') \in r^{\mathcal{J}}$. Since $\mathcal{J}$ is Noetherian, we may assume that $d' \in m(B)^{\mathcal{J}}$. But then $(d_X, d') \in S_{n+1}$, as required. □