Finding High-Level Explanations for Subsumption w.r.t. Combinations of CBoxes in $\mathcal{EL}$ and $\mathcal{EL}^+$

Dennis Peuter and Viorica Sofronie-Stokkermans*

University Koblenz-Landau, Koblenz, Germany

Abstract. We address the problem of finding high-level explanations for concept subsumption w.r.t. combinations of $\mathcal{EL}$ (resp. $\mathcal{EL}^+$) CBoxes. Our goal is to find explanations for concept subsumptions in such combinations of CBoxes which contain only symbols (concept names and role names) that are common to the CBoxes. For this, we use the encoding of TBox subsumption as a uniform word problem in classes of semilattices with monotone operators for $\mathcal{EL}$ and the $\leq$-interpolation property in these classes of algebras, as well as extensions to these results in the presence of role inclusions. For computing the $\leq$-interpolating terms we use a translation to propositional logic and methods for computing Craig interpolants in propositional logic.

1 Introduction

Description logics are logics for knowledge representation which provide a logical basis for modeling and reasoning about objects, classes of objects (concepts), and relationships between them (roles). They are of particular importance in providing a logical formalism for ontologies. One of the problems arising when creating ontologies is to ensure that they do not contain mistakes that could allow to prove subsumptions between concepts that are not supposed to hold. One situation in which this can happen is when already existing databases (or ontologies) which can be considered trustworthy are extended or when two databases (or ontologies) are put together. Even if the new ontology is still consistent, one needs to make sure that no concept inclusions which are not supposed to be true can be derived. It is therefore important to provide simple explanations for concept subsumptions in such combined ontologies (containing, for instance, only symbols that occur both in the original ontology and in the extension). In this paper we analyze this particular problem for the case in which the two ontologies consist only of TBoxes resp. CBoxes. We restrict the description logics to $\mathcal{EL}$ and $\mathcal{EL}^+$. We use the encoding of TBox subsumption for $\mathcal{EL}$ as a uniform word problem in classes of semilattices with monotone operators and the $\leq$-interpolation property in these classes of algebras, as well as extensions to these results in

CEUR-WS.org/Vol-2663/paper-18.pdf

the presence of role inclusions. (A subset of the axioms needed for deriving a concept inclusion can be determined using an unsatisfiable core computation.) For computing the $\leq$-interpolating terms we use a translation to propositional logic and methods for computing Craig interpolants in propositional logic. We regard these $\leq$-interpolating terms as high-level explanations for the subsumption. If more explanations are needed, they can then be obtained by analyzing the unsatisfiable cores and the resolution derivation of the $\leq$-interpolating terms.

Related work One method for finding justifications in description logics which has been addressed in previous work is the so-called axiom pinpointing. The idea is to find a minimal axiom set, which already has the consequence in question. Similar algorithms for computing minimal axiom sets for $\mathcal{ALC}$-terminologies were given by Baader and Hollunder [4], and by Schlobach and Cornet [18]. They are extensions of the tableau-like satisfiability algorithm for $\mathcal{ALC}$ and the tableau-like consistency algorithm for $\mathcal{ALC}$, respectively, in which they make use of labels to keep track of the axioms that were used during the execution of the algorithms. In contrast to the algorithm in [18], the one in [4] does not compute minimal axiom sets directly, but Boolean formulae from which they can be derived. Possibilities of explaining $\mathcal{ALC}$-subsumption (again based on tableau implementations) are described in [9]. In [17], a general approach to produce axiom pinpointing extensions of consequence-based algorithms is proposed. The methods we propose in this paper are based on resolution and hierarchical reasoning and are restricted to $\mathcal{EL}$ and $\mathcal{EL}^+$.

In [7] and [8] Baader et al. give a similar algorithm for axiom pinpointing in the description logics $\mathcal{EL}$ and $\mathcal{EL}^+$, respectively, in which they modify the subsumption algorithm for $\mathcal{EL}$ (respectively $\mathcal{EL}^+$). Here again labels are used to keep track of the axioms needed and the output is a Boolean formula, from which the axioms can be derived. They show that computing all possible minimal axiom sets may need exponential time, whereas computing one such set can be done in polynomial time. In [8] they consider extensions of TBoxes, i.e. unions of a static TBox (with irrefutable axioms) and a refutable TBox. There are also approaches to finding and enumerating justifications in $\mathcal{EL}$-ontologies and extensions thereof using saturation with respect to a consequence-based calculus [14], using resolution [11], or using other suitable SAT-tools [2]. The approach we present could use these methods for axiom pinpointing or justification computation (here we instead use unsatisfiable core computation).

Possibilities of finding small proofs for description logics have been investigated in [1]. This is not the direct goal of this paper, but might be considered as a further step in an incremental way of generating explanations, after having constructed intermediate terms. In [10], provenance for variants of the description logic $\mathcal{EL}$ is studied; our approach could be extended with such considerations for making the explanations more informative.

Possibilities of detecting differences between ontologies have been studied in [12]; this could be seen as a first step of obtaining the concept inclusions on which our method could be applied.

A_1 :	$\mathsf{Endocardium}$	$\sqsubseteq$	Tissue
A_2 :	$\mathsf{Endocardium}$	$\sqsubseteq$	$\exists\mathsf{part\text{-}of.HeartWall}$
A_3 :	$\mathsf{HeartWall}$	$\sqsubseteq$	$\mathsf{BodyWall}$
A_4 :	$\mathsf{HeartWall}$	$\sqsubseteq$	$\exists\mathsf{part\text{-}of.LeftVentricle}$
A_5 :	$\mathsf{HeartWall}$	$\sqsubseteq$	$\exists\mathsf{part\text{-}of.RightVentricle}$
A_6 :	$\mathsf{LeftVentricle}$	$\sqsubseteq$	$\mathsf{Ventricle}$
A_7 :	$\mathsf{RightVentricle}$	$\sqsubseteq$	$\mathsf{Ventricle}$
A_8 :	$\mathsf{Endocarditis}$	$\sqsubseteq$	$\mathsf{Inflammation}$
A_9 :	$\mathsf{Endocarditis}$	$\sqsubseteq$	$\exists\mathsf{has\text{-}location.Endocardium}$
A_{10} :	$\mathsf{Inflammation} \sqcap \exists\mathsf{has\text{-}location.Endocardium}$	$\sqsubseteq$	$\mathsf{Endocarditis}$
A_{11} :	$\mathsf{Inflammation}$	$\sqsubseteq$	$\mathsf{Disease}$
A_{12} :	$\mathsf{Inflammation}$	$\sqsubseteq$	$\exists\mathsf{acts\text{-}on.Tissue}$
B_1 :	$\mathsf{Ventricle}$	$\sqsubseteq$	$\exists\mathsf{part\text{-}of.Heart}$
B_2 :	$\mathsf{HeartDisease}$	$\sqsubseteq$	$\mathsf{Disease}$
B_3 :	$\mathsf{HeartDisease}$	$\sqsubseteq$	$\exists\mathsf{has\text{-}location.Heart}$
B_4 :	$\mathsf{Disease} \sqcap \exists\mathsf{has\text{-}location.Heart}$	$\sqsubseteq$	$\mathsf{HeartDisease}$
R_1 :	$\mathsf{part\text{-}of} \circ \mathsf{part\text{-}of}$	$\sqsubseteq$	$\mathsf{part\text{-}of}$
R_2 :	$\mathsf{has\text{-}location} \circ \mathsf{part\text{-}of}$	$\sqsubseteq$	$\mathsf{has\text{-}location}$

Fig. 1: Example of an $\mathcal{EL}^+$ ontology

1.1 Illustration

We illustrate our ideas on the example CBox in Figure 1. It is based on an example from [28], which we modified in some points. We changed the CBox in a way that it only contains general concept inclusions and conjunction only appears on the left hand side of an axiom. Furthermore we left out some axioms and concepts, but also added new concepts (LeftVentricle, RightVentricle, Ventricle) and changed some axioms accordingly. (Note that concept names always start with capital letters, whereas role names start with small letters.)

We divided the CBox into three parts: The A-part is our main CBox, which is supposed to be consistent. The B-part is an extension of the main CBox and may introduce some new (and in the worst case even unwanted) consequences. The R-part contains only role axioms RI (we assume that role symbols are always among the shared symbols of the two CBoxes). Thus, the A-part is the CBox $\mathcal{T}_A \cup RI$ and the B-part is the CBox $\mathcal{T}_B \cup RI$, where $\mathcal{T}_A$ and $\mathcal{T}_B$ are TBoxes. We are interested in finding simple explanations for consequences w.r.t. this extended CBox. One such consequence is $\mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$.

Constructor name	Syntax	Semantics
conjunction	$C_1 \sqcap C_2$	$C_1^{\mathcal{I}} \cap C_2^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \mid \exists y((x, y) \in r^{\mathcal{I}} \text{ and } y \in C^{\mathcal{I}})\}$

Fig. 2: $\mathcal{EL}$ constructors and their semantics

Note that $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models \mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$, but $\mathcal{T}_A \cup RI \not\models \mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$ and $\mathcal{T}_B \cup RI \not\models \mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$, i.e. the consequence comes from the extension of the ontology and is not a consequence of one of the parts alone. Our goal now is to find an explanation of the reason why $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models \mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$. Formally this means that we try to find a concept description C which contains only shared symbols of $\mathcal{T}_A$ and $\mathcal{T}_B$ such that $\mathcal{T}_A \cup RI \models \mathsf{Endocarditis} \sqsubseteq C$ and $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models C \sqsubseteq \mathsf{Heartdisease}$. The common concepts in this example are $\mathsf{Disease}$ and $\mathsf{Ventricle}$. We obtain the concept description $C := \mathsf{Disease} \sqcap \exists \mathsf{has\text{-}location.Ventricle}$, which can be regarded as a high-level explanation for $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models \mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$. The details are presented in Section 4.

2 Preliminaries

The central notions in description logics are concepts and roles. In any description logic a set N_C of *concept names* and a set N_R of *roles* is assumed to be given. Complex concepts are defined starting with the concept names in N_C, with the help of a set of *concept constructors*. The available constructors determine the expressive power of a description logic. The semantics of description logics is defined in terms of interpretations $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$, where $\Delta^{\mathcal{I}}$ is a non-empty set, and the function $\cdot^{\mathcal{I}}$ maps each concept name $C \in N_C$ to a set $C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$ and each role name $r \in N_R$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$.

The description logics $\mathcal{EL}$, $\mathcal{EL}^+$ and some extensions. If we only allow intersection and existential restriction as concept constructors, we obtain the description logic $\mathcal{EL}$ [3], a logic used in terminological reasoning in medicine [27, 26]. Fig. 2 shows the constructor names used in the description logic $\mathcal{EL}$ and their semantics. The extension of $\cdot^{\mathcal{I}}$ to concept descriptions is inductively defined using the semantics of the constructors. In [6, 5], the extension $\mathcal{EL}^+$ of $\mathcal{EL}$ with role inclusion axioms is studied. Relationships between concepts and roles are described using TBoxes or, more generally, CBoxes.

Definition 1 (TBox, Model, TBox subsumption). *A TBox (or terminology) is a finite set consisting of* primitive concept definitions *of the form $C \equiv D$, where C is a concept name and D a concept description; and* general concept inclusions *(GCI) of the form $C \sqsubseteq D$, where C and D are concept descriptions.*

- *An interpretation $\mathcal{I}$ is a* model of a TBox $\mathcal{T}$ *if it satisfies:*
  - *all concept definitions in $\mathcal{T}$, i.e., $C^{\mathcal{I}}{=}D^{\mathcal{I}}$ for all definitions $C{\equiv}D \in \mathcal{T}$;*

- *all general concept inclusions in* $\mathcal{T}$, *i.e.*, $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ *for every* $C \sqsubseteq D \in \mathcal{T}$.

– *Let* $\mathcal{T}$ *be a TBox, and* C_1, C_2 *two concept descriptions.* C_1 is subsumed by C_2 w.r.t. $\mathcal{T}$ ($C_1 \sqsubseteq_{\mathcal{T}} C_2$) *if and only if* $C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$ *for every model* $\mathcal{I}$ *of* $\mathcal{T}$.

Since definitions can be expressed as double inclusions, in what follows we will only refer to TBoxes consisting of general concept inclusions (GCI) only.

Definition 2 (CBox, Model, CBox subsumption). *A CBox consists of a TBox* $\mathcal{T}$ *and a set* RI *of role inclusions of the form* $r_1 \circ \cdots \circ r_n \sqsubseteq s$. *Since terminologies can be expressed as sets of general concept inclusions, we will view CBoxes as unions* $GCI \cup RI$ *of a set* GCI *of general concept inclusions and a set* RI *of role inclusions of the form* $r_1 \circ \cdots \circ r_n \sqsubseteq s$, *with* $n \geq 1$.

– *An interpretation* $\mathcal{I}$ *is a model of the CBox* $\mathcal{C} = GCI \cup RI$ *if it is a model of* GCI *and satisfies all role inclusions in* $\mathcal{C}$, *i.e.*, $r_1^{\mathcal{I}} \circ \cdots \circ r_n^{\mathcal{I}} \subseteq s^{\mathcal{I}}$ *for all* $r_1 \circ \cdots \circ r_n \subseteq s \in RI$.
– *If* $\mathcal{C}$ *is a CBox, and* C_1, C_2 *are concept descriptions then* $C_1 \sqsubseteq_{\mathcal{C}} C_2$ *if and only if* $C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$ *for every model* $\mathcal{I}$ *of* $\mathcal{C}$.

In [6] it was shown that subsumption w.r.t. CBoxes in $\mathcal{EL}^+$ can be reduced in linear time to subsumption w.r.t. *normalized* CBoxes, in which all GCIs have one of the forms: $C \sqsubseteq D, C_1 \sqcap C_2 \sqsubseteq D, C \sqsubseteq \exists r.D, \exists r.C \sqsubseteq D$, where C, C_1, C_2, D are concept names, and all role inclusions are of the form $r \sqsubseteq s$ or $r_1 \circ r_2 \sqsubseteq s$, where r, s, r_1, r_2 are role names. Therefore, in what follows, we consider w.l.o.g. that CBoxes only contain role inclusions of the form $r \sqsubseteq s$ and $r_1 \circ r_2 \sqsubseteq s$.

Algebraic semantics for $\mathcal{EL}$, $\mathcal{EL}^+$ and extensions thereof. In [20] we studied the link between TBox subsumption in $\mathcal{EL}$ and uniform word problems in the corresponding classes of semilattices with monotone functions. In [22, 23], we showed that these results naturally extend to the description logic $\mathcal{EL}^+$.
Let $\mathsf{SLO}(\Sigma)$ be the class of all $\wedge$-semilattices with unary operators $(S, \wedge, \{f_S\}_{f \in \Sigma})$, such that, for every $f \in \Sigma$, $f_S : S \rightarrow S$ is a monotone function, i.e. f satisfies:

$$\mathsf{Mon}(\Sigma) = \bigwedge_{f \in \Sigma} \forall x, y \left(x \leq y \rightarrow f(x) \leq f(y)\right).$$

When defining the semantics of $\mathcal{EL}$ or $\mathcal{EL}^+$ with role names N_R we use a class of $\wedge$-semilattices with monotone operators of the form $\mathsf{SLO}^{\exists}_{N_R} = (S, \wedge, \{f_{\exists r}\}_{r \in N_R})$. Every concept description C can be represented as a term $\overline{C}$; the encoding is inductively defined:

– Every concept name $C \in N_C$ is regarded as a variable $\overline{C} = C$.
– $\overline{C_1 \sqcap C_2} = \overline{C}_1 \wedge \overline{C}_2$ and $\overline{\exists r C} = f_{\exists r} \overline{C}$.

If RI is a set of role inclusions of the form $r \sqsubseteq s$ and $r_1 \circ r_2 \sqsubseteq s$, let RI_a be the set of all axioms of the form

$$\begin{array}{ll} \forall x \; (f_{\exists r}(x) \leq f_{\exists s}(x)) & \text{for all } r \sqsubseteq s \in RI \\ \forall x \; (f_{\exists r_1}(f_{\exists r_2}(x)) \leq f_{\exists s}(x)) & \text{for all } r_1 \circ r_2 \sqsubseteq s \in RI \end{array}$$

We will denote by $\mathsf{SLO}^{\exists}_{N_R}(RI_a)$ the class of all semilattices with monotone operators in which all axioms in RI_a hold.

Theorem 3 ([23]). *If the only concept constructors are intersection and existential restriction, then for all concept descriptions D_1, D_2 and every $\mathcal{EL}^+$ CBox $\mathcal{C}$=$GCI \cup RI$ – where RI consists of role inclusions of the form $r \sqsubseteq s$ and $r_1 \circ r_2 \sqsubseteq s$ – with concept names $N_C = \{C_1, \dots, C_n\}$ and set of roles N_R the following are equivalent:*

(1) $D_1 \sqsubseteq_{\mathcal{C}} D_2$.
(2) $\mathsf{SLO}^{\exists}_{N_R}(RI_a) \models \forall C_1 \dots C_n \left(\left(\bigwedge_{C \sqsubseteq D \in GCI} \overline{C} \leq \overline{D} \right) \rightarrow \overline{D_1} \leq \overline{D_2} \right)$.

In [23, 24] we showed that the uniform word problem for the class of algebras $\mathsf{SLO}^{\exists}_{N_R}(RI_a)$ is decidable in PTIME. For this, we proved that $\mathsf{SLO}^{\exists}_{N_R}(RI_a)$ can be seen as a "local" extension (cf. [19]) of the theory SLat of semilattices.

Theorem 4 ([25, 23]). *Let G be a set of ground clauses. The following are equivalent:*

(1) $\mathsf{SLat} \cup \mathsf{Mon}(\Sigma) \cup G \models \perp$.
(2) $\mathsf{SLat} \cup \mathsf{Mon}(\Sigma)[G] \cup G$ *has no partial model A such that its $\{\wedge\}$-reduct is a semilattice, and all Σ-subterms of G are defined.*
Here we denoted by $\mathsf{Mon}(\Sigma)[G]$ *the set of all instances of axioms of* $\mathsf{Mon}(\Sigma)$ *containing only (ground) subterms occurring in G.*

Let $\mathsf{Mon}(\Sigma)[G]_0 \cup G_0 \cup \mathsf{Def}$ *be obtained from* $\mathsf{Mon}(\Sigma)[G] \cup G$ *by replacing (in a bottom-up manner) every term $t = f(c)$ starting with functions in Σ with a fresh constant c_t, and adding $t \approx c_t$ to the set* Def.

The following are equivalent (and equivalent to (1) and (2) above):

(3) $\mathsf{Mon}(\Sigma)[G]_0 \cup G_0 \cup \mathsf{Def}$ *has no partial model A such that its $\{\wedge\}$-reduct is a semilattice, and all Σ-subterms of G are defined.*
(4) $\mathsf{Mon}(\Sigma)[G]_0 \cup G_0$ *is unsatisfiable in* SLat.
(Note that in the presence of $\mathsf{Mon}(\Sigma)$ *the instances* $\mathsf{Con}[G]_0$ *of the congruence axioms for the functions in Σ,* $\mathsf{Con}[G]_0 = \{g{=}g' \rightarrow c_{f(g)}{=}c_{f(g')} \mid f(g){=}c_{f(g)}, f(g'){=}c_{f(g')} \in \mathsf{Def}\}$*, are not necessary.)*

This equivalence allows us to hierarchically reduce, in polynomial time, proof tasks in $\mathsf{SLat} \cup \mathsf{Mon}(\Sigma)$ to proof tasks in SLat (cf. e.g., [25]) which can then be solved in polynomial time. In [23, 24] we proved that similar results hold for the class $SLO_{\Sigma}(RI)$ of semilattices with monotone operators in a set Σ satisfying a family RI_a axioms of the form:

$$\forall x \quad g(x) \leq h(x) \tag{1}$$
$$\forall x \quad f(g(x)) \leq h(x) \tag{2}$$

resp. their flattened version RI_a^{flat}, in which (2) is replaced by (3):

$$\forall x \quad y \leq g(x) \rightarrow f(y) \leq h(x) \tag{3}$$

Theorem 5 ([24]). *Let SL be a local axiomatization of the theory of semilattices. The following are equivalent:*

(1) $SL \cup \mathsf{Mon}(\Sigma) \cup RI_a^{\mathsf{flat}} \models \forall \overline{x} \bigwedge_{i=1}^n s_i(\overline{x}) \leq s'_i(\overline{x}) \rightarrow s(\overline{x}) \leq s'(\overline{x})$;
(2) $SL \cup \mathsf{Mon}(\Sigma) \cup RI_a^{\mathsf{flat}} \wedge G \models \perp$, *where* $G = \bigwedge_{i=1}^n s_i(\overline{c}) \leq s'_i(\overline{c}) \wedge s(\overline{c}) \not\leq s'(\overline{c})$;
(3) $(SL \cup \mathsf{Mon}(\Sigma) \cup RI_a^{\mathsf{flat}})[\Psi_{RI}(G)] \wedge G \models \perp$ *where* $\Psi_{RI}(G) = \bigcup_{i \geq 0} \Psi_{RI}^i$, *with* $\Psi_{RI}^0 = \mathsf{st}(G)$ *and* $\Psi_{RI}^{i+1} = \{f_2(d) \mid f(d) \in \Psi_{RI}^i, (y \leq f_2(x) \rightarrow f_1(y) \leq f(x)) \in RI_a^{\mathsf{flat}}\}$.

An example that illustrates the way the method for hierarchical reasoning can be used in this case is given in Appendix A.

3 *P*-Interpolation Property

Let Pred be a set of predicates. We look at a certain kind of interpolation property which we call P-interpolating. In the following we give a definition for P-interpolation and show that the theory of semilattices has this property.

Definition 6. $\mathcal{T}_0$ *is* P-interpolating *with respect to* $P \in \mathsf{Pred}$*, if for all conjunctions A and B of ground literals, all binary predicates $R \in P$ and all terms a and b such that a contains only constants occurring in A and b contains only constants occurring in B (or vice versa), if $A \wedge B \models_{\mathcal{T}_0} aRb$ then there exists a term t containing only constants common to A and B with $A \wedge B \models_{\mathcal{T}_0} aRt \wedge tRb$.* $\mathcal{T}_0$ *is* strongly P-interpolating*, if there exists such a term t with $A \models_{\mathcal{T}_0} aRt$ and $B \models_{\mathcal{T}_0} tRb$.*[1]

Proving P-interpolability is sometimes easier for theories which are P-convex.

Definition 7. *A theory $\mathcal{T}_0$ is* convex *with respect to the set* Pred *of all predicates (which may include also equality $\approx$) if for all conjunctions Γ of ground atoms, relations $R_1, \dots, R_m \in \mathsf{Pred}$ and ground tuples of corresponding arity $\overline{t}_1, \dots, \overline{t}_n$, if $\Gamma \models_{\mathcal{T}_0} \bigvee_{i=1}^m R_i(\overline{t}_i)$ then there exists $j \in \{1, \dots, m\}$ such that $\Gamma \models_{\mathcal{T}_0} R_j(\overline{t}_j)$.*

We will prove that the theory of semilattices with monotone operators is $\leq$-interpolating. For this we will use the fact that the theory of semilattices is $\leq$-convex.

Lemma 8. *The theory of semilattices is $\leq$-convex.*

Proof: The convexity of the theory of semilattices w.r.t. $\approx$ follows from the fact that this is an equational class; convexity w.r.t. $\leq$ follows from the fact that $x \leq y$ if and only if $(x \wedge y) \approx x$. $\square$

Lemma 9. *The theory* SLat *of semilattices is $\leq$-interpolating.*

[1] This definition is equivalent to the definition, sometimes used in the literature, in which a and b are required to be constants.

Proof: This is a constructive proof based on the fact that $\mathsf{SLat} = ISP(S_2)$ (i.e. every semilattice is isomorphic to a sublattice of a power of S_2), where S_2 is the 2-element semilattice, or, alternatively, that every semilattice is isomorphic to a semilattice of sets. We prove that the theory of semilattices is $\leq$-interpolating, i.e. that if A and B are two conjunctions of literals and $A \wedge B \models_{\mathsf{SLat}} a \leq b$, where a is a term containing only constants which occur in A and b a term containing only constants occurring in B, then there exists a term containing only common constants in A and B such that $A \models_{\mathsf{SLat}} a \leq t$ and $A \wedge B \models_{\mathsf{SLat}} t \leq b$. We can assume without loss of generality that A and B consist only of atoms (otherwise one moves the negative literals to the right and uses convexity - details are given in Appendix B). $A \wedge B \models_{\mathsf{SLat}} a \leq b$ if and only if the following conjunction of literals in propositional logic is unsatisfiable:

$$\begin{array}{llll}
(\mathsf{Ren}(\wedge)) & P_{e_1 \wedge e_2} \leftrightarrow P_{e_1} \wedge P_{e_2} & & P_{g_1 \wedge g_2} \leftrightarrow P_{g_1} \wedge P_{g_2} & \\
(\mathsf{P}) & P_{e_1} \leftrightarrow P_{e_2} & e_1 \approx e_2 \in A & P_{g_1} \leftrightarrow P_{g_2} & g_1 \approx g_2 \in B \\
 & P_{e_1} \rightarrow P_{e_2} & e_1 \leq e_2 \in A & P_{g_1} \rightarrow P_{g_2} & g_1 \leq g_2 \in B \\
(\mathsf{N}) & P_a & & \neg P_b & \\
 & \text{for all } e, e_1, e_2 \text{ subterms in } A & & \text{for all } g, g_1, g_2 \text{ subterms in } B &
\end{array}$$

We obtain an unsatisfiable set of clauses $(N_A \wedge P_a) \wedge (N_B \wedge \neg P_b) \models \perp$, where N_A and N_B are sets of Horn clauses in which each clause contains a positive literal. We can saturate $N_A \cup P_a$ under ordered resolution, in which all symbols occurring in A but not in B are larger than the common symbols. We show that if $A \wedge B \models_{\mathsf{SLat}} a \leq b$ holds, then for the term

$$t := \bigwedge \{e \mid A \models_{\mathsf{SLat}} a \leq e, e \text{ common subterm of } A \text{ and } B\}$$

the following hold:

(i) $A \models_{\mathsf{SLat}} a \leq t$, and
(ii) $A \wedge B \models_{\mathsf{SLat}} t \leq b$.

This means that for the theory of semilattices we have a property stronger than $\leq$-interpolability, but not quite as strong as strongly $\leq$-interpolability.[2]
Every $e \in T = \{e \mid A \models a \leq e, e \text{ common subterm of } A \text{ and } B\}$ corresponds to the positive unit clause P_e (where P_e is a propositional variable common to N_A and N_B) which can be derived from N_A using ordered resolution (with the ordering described above).

It is clearly the case that $A \models_{\mathsf{SLat}} a \leq t$, because $N_A \wedge P_a \wedge \neg P_t \wedge (P_t \leftrightarrow \bigwedge_{e \in T} P_e)$ is unsatisfiable. Thus, (i) holds. For proving (ii), observe that by saturating $N_A \wedge P_a$ under ordered resolution we obtain the following kinds of clauses containing only shared symbols which can possibly lead to $\perp$ after inferences with $N_B \wedge \neg P_b$ (and thus to the consequence $a \leq b$ together with B).

(a) P_{e_k} positive unit clauses s.t. e_k contains symbols common to A and B.
(b) $\bigwedge_{j=1}^{n} P_{c_{ij}} \rightarrow P_{d_i}$, where c_{ij} and d_i are common symbols, such that for all i, j and k we have $c_{ij} \neq e_k$ and $d_i \neq e_k$.

[2] This proof fixes a problem with a claim made in the Appendix of [21] where it is mentioned that the theory of semilattices is strongly $\leq$-interpolating.

Other types of clauses may appear too, but they can not be used to obtain $a \leq b$ (the details are presented in Appendix B).

For the proof of (ii) one needs to consider separately the case in which none of the clauses of type (b) is needed to derive $\perp$ together with $N_B \wedge \neg P_b$ (and thus the consequence $a \leq b$) and the case when some clauses of type (b) are needed. In the first case, $\perp$ is derived already because of a subset $\{P_{e_i} \mid i \in S\}$. In the second case a careful analysis is needed. The details are presented in Appendix B. $\square$

We show how to compute an intermediate term in the theory of semilattices on an example.

Example 10. Let $A = \{a_1 \leq c_1,\ c_2 \leq a_2,\ a_2 \leq c_3\}$ and $B = \{c_1 \leq b_1,\ b_1 \leq c_2,\ c_3 \leq b_2\}$. It is easy to see that $A \wedge B \models a_1 \leq b_2$. We can find an intermediate term by using the methods described in the proof: We saturate $N_A \wedge P_{a_1} = (P_{a_1} \rightarrow P_{c_1}) \wedge (P_{c_2} \rightarrow P_{a_2}) \wedge (P_{a_2} \rightarrow P_{c_3}) \wedge P_{a_1}$ under ordered resolution, in which the symbols P_{a_1}, P_{a_2} are larger than $P_{c_1}, P_{c_2}, P_{c_3}$. This yields the clauses P_{c_1} and $P_{c_2} \rightarrow P_{c_3}$ containing shared propositional variables. $(N_A \wedge P_{a_1}) \wedge (N_B \wedge \neg P_{b_2})$ is unsatisfiable iff $N_B \wedge \neg P_{b_2} \wedge P_{c_1} \wedge (P_{c_2} \rightarrow P_{c_3})$ is unsatisfiable. Indeed $t = c_1$ is an intermediate term, as $A \models a_1 \leq c_1$ and $A \wedge B \models c_1 \leq b_2$. Note that $N_B \wedge \neg P_{b_2} \wedge P_{c_1}$ is satisfiable, so $B \not\models c_1 \leq b_2$. Moreover, we only need $P_{c_2} \rightarrow P_{c_3}$ in addition to $N_B \cup \neg P_{b_2}$ to derive $\perp$, thus $A \wedge B \models c_1 \leq b_2$ and the clause $P_{c_2} \rightarrow P_{c_3}$ obtained from N_A is really needed for this.

Theorem 11. *The theory $\mathsf{SLO}^{\exists}_{N_R}(RI_a)$ of semilattices with monotone operators satisfying axioms RI_a is $\leq$-interpolating.*

Proof (Sketch): The operators of $\mathsf{SLO}^{\exists}_{N_R}(RI_a)$ satisfy the monotonicity condition Mon; the axioms in RI_a are in a class we studied in [21]. Let A and B be two conjunctions of literals (corresponding to two TBoxes), let RI be a set of role axioms and let Mon be the family of all monotonicity axioms for the functions $\{f_{\exists r} \mid r \in N_R\}$. Assume that $A \wedge B \models_{\mathsf{SLO}^{\exists}_{N_R}(RI_a)} a \leq b$, where a is a term containing only constants and Σ-functions occurring in A and b is a term containing only constants and Σ-functions occurring in B. By Theorem 4, $A \wedge B \models_{\mathsf{SLO}^{\exists}_{N_R}(RI_a)} a \leq b$ if and only if (with the notation used in Theorem 4), $A_0 \wedge B_0 \wedge \mathsf{Mon}[A \wedge B]_0 \wedge RI_a[A \wedge B]_0 \wedge \mathsf{Con}_0 \wedge \neg(a \leq b)_0 \models_{\mathsf{SLat}} \perp$. In the presence of monotonicity, Con is not needed. The set $\mathcal{H} = \mathsf{Mon}[A \wedge B]_0 \wedge RI_a[A \wedge B]_0 \wedge \neg(a \leq b)_0$ contains mixed clauses. Using a result similar to one used in [21] (the proof is given in Appendix C, cf. Proposition 13) we can "separate" all clauses in $\mathcal{H}$ as well as $a \leq b$, i.e. we have $A_0 \wedge \mathcal{H}_A \wedge B_0 \wedge \mathcal{H}_B \wedge \neg(a_0 \leq t_0 \wedge t_0 \leq b_0) \models_{\mathsf{SLat}} \perp$, where t_0 contains only constants common to A_0 and B_0. After replacing back the new constants with the terms they represent, we obtain: $A \wedge B \models_{\mathsf{SLO}^{\exists}_{N_R}(RI_a)} (a \leq t \wedge t \leq b)$, where t contains symbols common to A and B.[3] $\square$

[3] As in [21], for function symbols f, g, if f occurs in A and g occurs in B, but they occur together in one of the axioms in RI, they are considered to be shared.

4 $\leq$-Interpolation for High-Level Explanations

In this section we explain our method in detail and illustrate each step of the method on the example from Section 1.1. First we give a formal statement of the problem we are addressing: Let $\mathcal{T}_A$ and $\mathcal{T}_B$ be two $\mathcal{EL}^+$ TBoxes and RI a set of role inclusions. Let N_C be the set of all concept names occurring in $\mathcal{T}_A \cup \mathcal{T}_B$, N_C^A and N_C^B be the set of concept names occurring in $\mathcal{T}_A$ and $\mathcal{T}_B$, respectively, and $N_C^{AB} = (N_C^A \cap N_C^B)$ be the common concept names. Let X be a concept description over N_C^A and Y a concept description over N_C^B such that they do not contain only shared symbols and such that $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models X \sqsubseteq Y$, but $\mathcal{T}_A \cup RI \not\models X \sqsubseteq Y$ and $\mathcal{T}_B \cup RI \not\models X \sqsubseteq Y$. The goal is to find a concept description C, containing only concepts in N_C^{AB} (and possibly also only roles common to $\mathcal{T}_A$ and $\mathcal{T}_B$), such that $\mathcal{T}_A \cup RI \models X \sqsubseteq C$ and $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models C \sqsubseteq Y$. The concept description C can then be seen as a "high-level explanation" for $X \sqsubseteq Y$. Using Theorem 3 and Theorem 11, we can always compute such a concept description. For this we apply the following steps:

1. Translate to the theory of semilattices with monotone operators
2. Flatten, purify and use instantiation
3. Separate all mixed instances of role and monotonicity axioms
4. Compute an intermediate term using P-interpolation

Remark 12. We can make the method more efficient especially for large ontologies, by modifying Steps 2 and 3. Usually, if we have very large TBoxes, only some of their axioms are necessary for obtaining a certain consequence. Therefore it is sufficient to apply Step 2 only on the relevant axioms. Similarly, we can speed up our method by applying Step 3 only on the instances relevant for our problem. For determining which axioms/instances are relevant we can compute a minimal axiom set, for example by using unsatisfiable core computation. We therefore modify the method by including a Step 2a (before Step 2) and a Step 3a (before Step 3) in which we compute a minimal axiom or instance set.

For the ontology from Section 1 we have the following sets of symbols (we indicate also the abbreviations used in what follows):

$$
\begin{aligned}
N_C^A = \quad & \{\mathsf{Endocardium(Em)}, \mathsf{Tissue(T)}, \mathsf{HeartWall(HW)}, \\
& \mathsf{LeftVentricle(LV)}, \mathsf{RightVentricle(RV)}, \mathsf{Ventricle(V)} \\
& \mathsf{Disease(D)}, \mathsf{Inflammation(I)}, \mathsf{Endocarditis(Es)}\} \\
N_C^B = \quad & \{\mathsf{Heart(H)}, \mathsf{HeartDisease(HD)}, \mathsf{Disease(D)}, \mathsf{Ventricle(V)}\} \\
N_C^{AB} = \quad & \{\mathsf{Disease(D)}, \mathsf{Ventricle(V)}\}
\end{aligned}
$$

Therefore the consequence $\mathsf{Endocarditis} \sqsubseteq \mathsf{Heartdisease}$ indeed belongs to the problem described above. We show how to apply steps 1 to 4 (including steps 2a and 3a) in detail:

Step 1: Figure 3 shows the ontology after the translation to the theory of semilattices (SLat). For this we replace $\sqsubseteq$ by $\leq$ and $\sqcap$ by $\wedge$, and write the role

$A_1:$	$\mathsf{Em} \leq \mathsf{T}$		
$A_2:$	$\mathsf{Em} \leq \mathsf{po(HW)}$	$B_1:$	$\mathsf{V} \leq \mathsf{po(H)}$
$A_3:$	$\mathsf{HW} \leq \mathsf{BW}$	$B_2:$	$\mathsf{HD} \leq \mathsf{D}$
$A_4:$	$\mathsf{HW} \leq \mathsf{po(LV)}$	$B_3:$	$\mathsf{HD} \leq \mathsf{hl(H)}$
$A_5:$	$\mathsf{HW} \leq \mathsf{po(RV)}$	$B_4:$	$\mathsf{D} \wedge \mathsf{hl(H)} \leq \mathsf{HD}$
$A_6:$	$\mathsf{LV} \leq \mathsf{V}$		
$A_7:$	$\mathsf{RV} \leq \mathsf{V}$	$R_1:$	$\forall \mathsf{X}\colon \mathsf{po(po(X))} \leq \mathsf{po(X)}$
$A_8:$	$\mathsf{Es} \leq \mathsf{I}$	$R_2:$	$\forall \mathsf{X}\colon \mathsf{hl(po(X))} \leq \mathsf{hl(X)}$
$A_9:$	$\mathsf{Es} \leq \mathsf{hl(Em)}$		
$A_{10}:$	$\mathsf{I} \wedge \mathsf{hl(Em)} \leq \mathsf{Es}$	$M_1:$	$\forall \mathsf{X,Y}\colon\ \mathsf{X} \leq \mathsf{Y} \rightarrow \mathsf{po(X)} \leq \mathsf{po(Y)}$
$A_{11}:$	$\mathsf{I} \leq \mathsf{D}$	$M_2:$	$\forall \mathsf{X,Y}\colon\ \mathsf{X} \leq \mathsf{Y} \rightarrow \mathsf{hl(X)} \leq \mathsf{hl(Y)}$
$A_{12}:$	$\mathsf{I} \leq \mathsf{ao(T)}$	$M_3:$	$\forall \mathsf{X,Y}\colon\ \mathsf{X} \leq \mathsf{Y} \rightarrow \mathsf{ao(X)} \leq \mathsf{ao(Y)}$

Fig. 3: Ontology after translation to SLat with monotone operators

axioms as universal formulae. Note that we use abbreviations for role names (e.g. hl for has-location, po for part-of, ao for acts-on). Also note that we now state the monotonicity axioms for each role explicitly.

Step 2a: Using unsat core computation we get the minimal axiom set $min_A = \{A_2, A_4, A_6, A_8, A_9, A_{11}, B_1, B_4, R_2\}$. This means that for the following instantiation step we only have to consider the role axiom R_2 and none of the monotonicity axioms is needed.

Step 2: Let $\mathcal{T}_0 = \mathsf{SLat}$ and $T_1 = \mathsf{SLat} \cup R_2$ be the extension of T_0 with axiom R_2. We know that it is a local theory extension, so we can use hierarchical reasoning. We first flatten the role axiom R_2 in the following way:

$$R_2^{\mathsf{flat}}:\quad \forall \mathsf{X},\ \mathsf{Y}\colon\quad \mathsf{X} \leq \mathsf{po(Y)} \quad\rightarrow\quad \mathsf{hl(X)} \leq \mathsf{hl(Y)}$$

We have the following instances of this axiom:

$I_1:$	$\mathsf{Em} \leq \mathsf{po(HW)} \rightarrow \mathsf{hl(Em)} \leq \mathsf{hl(HW)}$	$I_9:$	$\mathsf{LV} \leq \mathsf{po(RV)} \rightarrow \mathsf{hl(LV)} \leq \mathsf{hl(RV)}$
$I_2:$	$\mathsf{Em} \leq \mathsf{po(LV)} \rightarrow \mathsf{hl(Em)} \leq \mathsf{hl(LV)}$	$I_{10}:$	$\mathsf{LV} \leq \mathsf{po(H)} \rightarrow \mathsf{hl(LV)} \leq \mathsf{hl(H)}$
$I_3:$	$\mathsf{Em} \leq \mathsf{po(RV)} \rightarrow \mathsf{hl(Em)} \leq \mathsf{hl(RV)}$	$I_{11}:$	$\mathsf{RV} \leq \mathsf{po(HW)} \rightarrow \mathsf{hl(RV)} \leq \mathsf{hl(HW)}$
$I_4:$	$\mathsf{Em} \leq \mathsf{po(H)} \rightarrow \mathsf{hl(Em)} \leq \mathsf{hl(H)}$	$I_{12}:$	$\mathsf{RV} \leq \mathsf{po(LV)} \rightarrow \mathsf{hl(RV)} \leq \mathsf{hl(LV)}$
$I_5:$	$\mathsf{HW} \leq \mathsf{po(LV)} \rightarrow \mathsf{hl(HW)} \leq \mathsf{hl(LV)}$	$I_{13}:$	$\mathsf{RV} \leq \mathsf{po(H)} \rightarrow \mathsf{hl(RV)} \leq \mathsf{hl(H)}$
$I_6:$	$\mathsf{HW} \leq \mathsf{po(RV)} \rightarrow \mathsf{hl(HW)} \leq \mathsf{hl(RV)}$	$I_{14}:$	$\mathsf{H} \leq \mathsf{po(HW)} \rightarrow \mathsf{hl(H)} \leq \mathsf{hl(HW)}$
$I_7:$	$\mathsf{HW} \leq \mathsf{po(H)} \rightarrow \mathsf{hl(HW)} \leq \mathsf{hl(H)}$	$I_{15}:$	$\mathsf{H} \leq \mathsf{po(LV)} \rightarrow \mathsf{hl(H)} \leq \mathsf{hl(LV)}$
$I_8:$	$\mathsf{LV} \leq \mathsf{po(HW)} \rightarrow \mathsf{hl(LV)} \leq \mathsf{hl(HW)}$	$I_{16}:$	$\mathsf{H} \leq \mathsf{po(RV)} \rightarrow \mathsf{hl(H)} \leq \mathsf{hl(RV)}$

We purify all formulae by introducing new constants for the terms starting with a function symbol, i.e. role names. We save the definitions in the following set:

$$\begin{aligned}\mathsf{Def} = \{&\mathsf{po_{HW}} = \mathsf{po(HW)}, \mathsf{po_{LV}} = \mathsf{po(LV)}, \mathsf{po_H} = \mathsf{po(H)}, \mathsf{hl_{EM}} = \mathsf{hl(EM)},\\ &\mathsf{hl_{HW}} = \mathsf{hl(HW)}, \mathsf{hl_{LV}} = \mathsf{hl(LV)}, \mathsf{hl_{HC}} = \mathsf{hl(HC)}, \mathsf{hl_H} = \mathsf{hl(H)}\}\end{aligned}$$

We then have the set $A_0 \wedge B_0 \wedge I_0$, where A_0, B_0 and I_0 are the purified versions of $A = \{A_2, A_4, A_6, A_8, A_9, A_{11}\}$, $B = \{B_1, B_4\}$ and $I = \{I_1, ..., I_{10}\}$, respectively.

Step 3a: Computing an unsatisfiable core yields the following set of axioms: $\{A_2, A_4, A_6, A_8, A_9, A_{11}, B_1, B_4, I_1, I_5, I_{10}\}$. So we have $\mathcal{H} = \{I_1, I_5, I_{10}\}$. Out of these instances the first two are pure A (meaning the premise contains only symbols in N_C^A), but I_{10} is a mixed instance, since $\mathsf{LV} \in N_C^A \backslash N_C^B$ and $\mathsf{H} \in N_C^B \backslash N_C^A$, so $\mathcal{H}_{\mathsf{mix}} = \{I_{10}\}$.

Step 3: To separate the mixed instance $\mathsf{LV} \leq \mathsf{po_H} \to \mathsf{hl_{LV}} \leq \mathsf{hl_H}$ one has to find an intermediate term t in the common signature such that $\mathsf{LV} \leq t$ and $t \leq \mathsf{po_H}$. $t = \mathsf{V}$ is such a term. We get $\mathcal{H}_{\mathsf{sep}} = \{I_{10}^A, I_{10}^B\}$ where

$$I_{10}^A: \quad \mathsf{LV} \leq \mathsf{V} \quad \to \quad \mathsf{hl_{LV}} \leq \mathsf{hl_V}$$
$$I_{10}^B: \quad \mathsf{V} \leq \mathsf{po_H} \quad \to \quad \mathsf{hl_V} \leq \mathsf{hl_H}$$

Note that I_{10}^A is an instance of the monotonicity axiom for the $\mathsf{has\text{-}location}$ role and I_{10}^B is an instance of axiom R_2^{flat}.

Step 4: Note that w.r.t. SLat the formula $A_0 \wedge I_1 \wedge I_5 \wedge I_{10}^A$ is equivalent to:

$$\begin{aligned}\overline{A}_0 = \ & \mathsf{Em} \leq \mathsf{po_{HW}} \wedge \mathsf{HW} \leq \mathsf{po_{LV}} \wedge \mathsf{LV} \leq \mathsf{V} \wedge \mathsf{Es} \leq \mathsf{I} \wedge \mathsf{Es} \leq \mathsf{hl_{Em}} \wedge \mathsf{I} \leq \mathsf{D} \\ & \wedge \mathsf{hl_{EM}} \leq \mathsf{hl_{HW}} \wedge \mathsf{hl_{HW}} \leq \mathsf{hl_{LV}} \wedge \mathsf{hl_{LV}} \leq \mathsf{hl_V}\end{aligned}$$

To obtain an explanation for $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models \mathsf{Endocarditis} \sqsubseteq \mathsf{HeartDisease}$ we saturate the set $\overline{A}_0 \wedge \mathsf{Es}$ under ordered resolution as described in the proof of Theorem 9, where symbols occurring in A and not in B are larger than common symbols. Doing this yields two inferences containing only common symbols: D and $\mathsf{hl_V}$. By taking the conjunction of these terms and translating the formula back to description logic, we obtain $J = \mathsf{Disease} \sqcap \exists \mathsf{has\text{-}location.Ventricle}$. Then $\mathcal{T}_A \cup RI \models \mathsf{Endocarditis} \sqsubseteq J$ and $\mathcal{T}_A \cup \mathcal{T}_B \cup RI \models J \sqsubseteq \mathsf{HeartDisease}$.

5 Conclusion

We analyzed a possibility of giving high-level justifications for subsumption in the description logics $\mathcal{EL}$ and $\mathcal{EL}^+$. For this, we used the encoding of TBox subsumption as a uniform word problem in classes of semilattices with monotone operators for $\mathcal{EL}$ and the $\leq$-interpolation property in these classes of algebras, as well as extensions to these results in the presence of role inclusions. This can be seen as a first step towards providing short, high-level explanations for subsumption. If more explanations are needed, they can then be obtained by pinpointing and analyzing the resolution derivation of the $\leq$-interpolating terms.

There has been work on other forms of interpolation in the family of $\mathcal{EL}$ description logics: a variant of interpolation is proved in [16], possibilities for uniform interpolation are analyzed in [13] and [15] (it is well known that neither $\mathcal{ALC}$ nor $\mathcal{EL}$ allow uniform interpolation). As a plan for future work we would like to analyze possibilities of symbol elimination and abduction in such logics – which are strongly related to uniform interpolation.

Acknowledgements: We thank the reviewers for their helpful comments.

References

[1] C. Alrabbaa, F. Baader, S. Borgwardt, P. Koopmann, and A. Kovtunova. Finding small proofs for description logic entailments: Theory and practice. In E. Albert and L. Kovács, editors, *LPAR 2020: 23rd International Conference on Logic for Programming, Artificial Intelligence and Reasoning*, EPiC Series in Computing 73, pages 32–67. EasyChair, 2020.

[2] M. F. Arif, C. Mencía, A. Ignatiev, N. Manthey, R. Peñaloza, and J. Marques-Silva. BEACON: an efficient sat-based tool for debugging $\mathcal{EL}^+$ ontologies. In N. Creignou and D. L. Berre, editors, *Theory and Applications of Satisfiability Testing - SAT 2016 - 19th International Conference, Proceedings*, LNCS 9710, pages 521–530. Springer, 2016.

[3] F. Baader. Terminological cycles in a description logic with existential restrictions. In G. Gottlob and T. Walsh, editors, *IJCAI-03, Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 325–330. Morgan Kaufmann, 2003.

[4] F. Baader and B. Hollunder. Embedding defaults into terminological knowledge representation formalisms. *J. Autom. Reasoning*, 14(1):149–180, 1995.

[5] F. Baader, C. Lutz, and B. Suntisrivaraporn. Efficient reasoning in $\mathcal{EL}^+$. In B. Parsia, U. Sattler, and D. Toman, editors, *Proceedings of the 2006 International Workshop on Description Logics (DL2006)*, CEUR Workshop Proceedings 189. CEUR-WS.org, 2006.

[6] F. Baader, C. Lutz, and B. Suntisrivaraporn. Is tractable reasoning in extensions of the description logic $\mathcal{EL}$ useful in practice? *Journal of Logic, Language and Information*, 2007. Special issue on Method for Modality (M4M), 2007.

[7] F. Baader, R. Peñaloza, and B. Suntisrivaraporn. Pinpointing in the description logic $\mathcal{EL}$. In D. Calvanese, E. Franconi, V. Haarslev, D. Lembo, B. Motik, A. Turhan, and S. Tessaris, editors, *Proceedings of the 2007 International Workshop on Description Logics (DL2007)*, CEUR Workshop Proceedings 250. CEUR-WS.org, 2007.

[8] F. Baader, R. Peñaloza, and B. Suntisrivaraporn. Pinpointing in the description logic $\mathcal{EL}^+$. In J. Hertzberg, M. Beetz, and R. Englert, editors, *KI 2007: Advances in Artificial Intelligence, 30th Annual German Conference on AI, KI 2007, Proceedings*, LNCS 4667, pages 52–67. Springer, 2007.

[9] A. Borgida, E. Franconi, and I. Horrocks. Explaining $\mathcal{ALC}$ subsumption. In W. Horn, editor, *ECAI 2000, Proceedings of the 14th European Conference on Artificial Intelligence*, pages 209–213. IOS Press, 2000.

[10] C. Bourgaux, A. Ozaki, R. Peñaloza, and L. Predoiu. Provenance for the description logic $\mathcal{ELH}^r$. In C. Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 1862–1869. ijcai.org, 2020.

[11] Y. Kazakov and P. Skocovský. Enumerating justifications using resolution. In D. Galmiche, S. Schulz, and R. Sebastiani, editors, *Automated Reasoning - 9th International Joint Conference, IJCAR 2018, Held as Part of the Federated Logic Conference, FloC 2018, Proceedings*, LNCS 10900, pages 609–626. Springer, 2018.

[12] B. Konev, M. Ludwig, D. Walther, and F. Wolter. The logical difference for the lightweight description logic $\mathcal{EL}$. *J. Artif. Intell. Res.*, 44:633–708, 2012.

[13] B. Konev, D. Walther, and F. Wolter. Forgetting and uniform interpolation in large-scale description logic terminologies. In C. Boutilier, editor, *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 830–835, 2009.

[14] M. Ludwig. Just: a tool for computing justifications w.r.t. $\mathcal{ELH}$ ontologies. In S. Bail, B. Glimm, E. Jiménez-Ruiz, N. Matentzoglu, B. Parsia, and A. Steigmiller, editors, *Informal Proceedings of the 3rd International Workshop on OWL Reasoner Evaluation (ORE 2014) co-located with the Vienna Summer of Logic (VSL 2014)*, CEUR Workshop Proceedings 1207, pages 1–7. CEUR-WS.org, 2014.

[15] C. Lutz, I. Seylan, and F. Wolter. An automata-theoretic approach to uniform interpolation and approximation in the description logic $\mathcal{EL}$. In G. Brewka, T. Eiter, and S. A. McIlraith, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012*. AAAI Press, 2012.

[16] C. Lutz, I. Seylan, and F. Wolter. The data complexity of ontology-mediated queries with closed predicates. *Log. Methods Comput. Sci.*, 15(3), 2019.

[17] A. Ozaki and R. Peñaloza. Consequence-based axiom pinpointing. In D. Ciucci, G. Pasi, and B. Vantaggi, editors, *Scalable Uncertainty Management - 12th International Conference, SUM 2018, Proceedings*, LNCS 11142, pages 181–195. Springer, 2018.

[18] S. Schlobach and R. Cornet. Non-standard reasoning services for the debugging of description logic terminologies. In G. Gottlob and T. Walsh, editors, *IJCAI-03, Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 355–362. Morgan Kaufmann, 2003.

[19] V. Sofronie-Stokkermans. Hierarchic reasoning in local theory extensions. In R. Nieuwenhuis, editor, *Automated Deduction - CADE-20, 20th International Conference on Automated Deduction, Proceedings*, LNCS 3632, pages 219–234. Springer, 2005.

[20] V. Sofronie-Stokkermans. Automated theorem proving by resolution in non-classical logics. *Ann. Math. Artif. Intell.*, 49(1-4):221–252, 2007.

[21] V. Sofronie-Stokkermans. Interpolation in local theory extensions. *Logical Methods in Computer Science*, 4(4), 2008.

[22] V. Sofronie-Stokkermans. Locality and subsumption testing in $\mathcal{EL}$ and some of its extensions. In F. Baader, C. Lutz, and B. Motik, editors, *Proceedings of the 21st International Workshop on Description Logics (DL2008)*, CEUR Workshop Proceedings 353. CEUR-WS.org, 2008.

[23] V. Sofronie-Stokkermans. Locality and subsumption testing in $\mathcal{EL}$ and some of its extensions. In C. Areces and R. Goldblatt, editors, *Advances in Modal Logic 7*, pages 315–339. College Publications, 2008.

[24] V. Sofronie-Stokkermans. Representation theorems and locality for subsumption testing and interpolation in the description logics $\mathcal{EL}$, $\mathcal{EL}^+$ and their extensions with n-ary roles and numerical domains. *Fundam. Inform.*, 156(3-4):361–411, 2017.

[25] V. Sofronie-Stokkermans and C. Ihlemann. Automated reasoning in some local extensions of ordered structures. *Multiple-Valued Logic and Soft Computing*, 13(4-6):397–414, 2007.

[26] K. A. Spackman. Normal forms for description logic expressions of clinical concepts in SNOMED RT. In *AMIA 2001, American Medical Informatics Association Annual Symposium*. AMIA, 2001.

[27] K. A. Spackman, K. E. Campbell, and R. A. Côté. SNOMED RT: a reference terminology for health care. In *AMIA 1997, American Medical Informatics Association Annual Symposium*. AMIA, 1997.

[28] B. Suntisrivaraporn. *Polynomial time reasoning support for design and maintenance of large-scale biomedical ontologies*. PhD thesis, Dresden University of Technology, Germany, 2009.

A Example of local reasoning in $\mathcal{EL}^+$

We illustrate the ideas on the example from Section 1.1. Consider the CBox $\mathcal{C}$ consisting of the following GCI:

$A_1 - A_2$	$\mathsf{Endocardium}$	$\sqsubseteq$	$\mathsf{Tissue} \sqcap \exists \mathsf{part\text{-}of.HeartWall}$
A_4	$\mathsf{HeartWall}$	$\sqsubseteq$	$\exists \mathsf{part\text{-}of.LeftVentricle}$
A_6	$\mathsf{LeftVentricle}$	$\sqsubseteq$	$\mathsf{Ventricle}$
$A_8 - A_9$	$\mathsf{Endocarditis}$	$\sqsubseteq$	$\mathsf{Inflammation} \sqcap \exists \mathsf{has\text{-}location.Endocardium}$
A_{11}	$\mathsf{Inflammation}$	$\sqsubseteq$	$\mathsf{Disease}$
B_1	$\mathsf{Ventricle}$	$\sqsubseteq$	$\exists \mathsf{part\text{-}of.Heart}$
$B_2 - B_4$	$\mathsf{Heartdisease}$	$=$	$\mathsf{Disease} \sqcap \exists \mathsf{has\text{-}location.Heart}$

and the following role inclusions RI:

$$\mathsf{part\text{-}of} \circ \mathsf{part\text{-}of} \sqsubseteq \mathsf{part\text{-}of}$$
$$\mathsf{has\text{-}location} \circ \mathsf{part\text{-}of} \sqsubseteq \mathsf{has\text{-}location}$$

We want to check whether $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{C}} \mathsf{Heartdisease}$. This is the case iff (with some abbreviations – e.g. hl stands for $\exists\mathsf{has\text{-}location}$ and po for $\exists\mathsf{part\text{-}of}$, HW for $\mathsf{HeartWall}$, Em for $\mathsf{Endocardium}$, H for Heart, etc.):

$$\begin{array}{l} SL \cup \mathsf{Mon}(\mathsf{hl}, \mathsf{po}) \cup \{\forall x\ y \leq \mathsf{po}(x) \rightarrow \mathsf{po}(y) \leq \mathsf{po}(x), \\ \qquad\qquad\qquad\quad \forall x\ y \leq \mathsf{po}(x) \rightarrow \mathsf{hl}(y) \leq \mathsf{hl}(x)\} \\ \quad \cup \{\mathsf{Em} \leq \mathsf{T} \wedge \mathsf{po}(\mathsf{HW}),\ \ \mathsf{HW} \leq \mathsf{po}(\mathsf{LV}),\ \ \mathsf{LV} \leq \mathsf{V},\ \ \mathsf{V} \leq \mathsf{po}(\mathsf{H}),\ \ \mathsf{I} \leq \mathsf{D}, \\ \qquad \mathsf{Endocarditis} \leq \mathsf{I} \wedge \mathsf{hl}(\mathsf{Em}),\ \ \mathsf{Heartdisease} = \mathsf{D} \wedge \mathsf{hl}(\mathsf{H}), \\ \qquad \mathsf{Endocarditis} \not\leq \mathsf{Heartdisease}\} \quad \models \quad \perp . \end{array}$$

Then $\mathsf{st}(\mathcal{K}, G) = \{\mathsf{po}(\mathsf{HW}), \mathsf{po}(\mathsf{LV}), \mathsf{po}(\mathsf{H}), \mathsf{hl}(\mathsf{Em}), \mathsf{hl}(\mathsf{H})\}$. To compute $\Psi_{\mathcal{K}}(G)$, note that $\Psi^0_{RI} = \mathsf{st}(\mathcal{K}, G)$, $\Psi^1_{RI} = \{\mathsf{po}(\mathsf{Em}), \mathsf{po}(\mathsf{H})\}$, and $\Psi^2_{RI} = \Psi^1_{RI}$.

Thus, $\Psi_{\mathcal{K}}(G) = \{\mathsf{po}(\mathsf{HW}), \mathsf{po}(\mathsf{LV}), \mathsf{po}(\mathsf{Em}), \mathsf{po}(\mathsf{H}), \mathsf{hl}(\mathsf{Em}), \mathsf{hl}(\mathsf{H})\}$. After computing $(RI_a \cup \mathsf{Mon}(\mathsf{hl}, \mathsf{po}) \cup \mathsf{Con})[\Psi(G)]$ we obtain the following conjunction of (Horn) ground clauses:

G	$(RI_a \wedge \mathsf{Mon} \wedge \mathsf{Con})[\Psi(G)] \wedge SL[\Psi(G)]$
$\mathsf{Em} \leq \mathsf{T} \wedge \mathsf{po}(\mathsf{HW})$	$y \leq \mathsf{po}(x) \rightarrow \mathsf{po}(y) \leq \mathsf{po}(x)$ for $x, y \in \{\mathsf{HW}, \mathsf{LV}, \mathsf{Em}, \mathsf{H}\}$
$\mathsf{HW} \leq \mathsf{po}(\mathsf{LV})$,	$y \leq \mathsf{po}(x) \rightarrow \mathsf{hl}(y) \leq \mathsf{hl}(x)$ for $x, y \in \{\mathsf{Em}, \mathsf{H}\}$
$\mathsf{LV} \leq \mathsf{V}$	
$\mathsf{V} \leq \mathsf{po}(\mathsf{H})$	$xRy \rightarrow \mathsf{po}(x) R \mathsf{po}(y)$ for $x, y \in \{\mathsf{HW}, \mathsf{LV}, \mathsf{Em}, \mathsf{H}\}$
$\mathsf{Endocarditis} \leq \mathsf{I} \wedge \mathsf{hl}(\mathsf{Em})$	$xRy \rightarrow \mathsf{hl}(x) R \mathsf{hl}(y)$ for $x, y \in \{\mathsf{Em}, \mathsf{H}\}$
$\mathsf{I} \leq \mathsf{D}$	$R \in \{\leq, \geq, =\}$
$\mathsf{Heartdisease} = \mathsf{D} \wedge \mathsf{hl}(\mathsf{H})$	
$\mathsf{Endocarditis} \not\leq \mathsf{Heartdisease}$	$SL[\Psi(G)]$

By Theorem 5, $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{C}} \mathsf{Heartdisease}$ iff $\phi = G \wedge (RI_a \wedge \mathsf{Mon} \wedge \mathsf{Con})[\Psi(G)] \wedge SL[\Psi(G)]$ is unsatisfiable. Note that ϕ is a set of ground clauses in first-order logic with equality, containing all instances of the congruence axioms corresponding to the (ground) terms which occur in ϕ. A translation to

Datalog can easily be obtained by replacing the function symbols with binary predicate symbols. Alternatively, we can process the instances in ϕ by replacing, in a bottom-up fashion, all the terms starting with function symbols (which are all ground) with new constants (and adding, separately, the corresponding definitions). We obtain the following set of clauses:

Def	G_0	$(RI_a \wedge \mathsf{Mon} \wedge \mathsf{Con})[\Psi(G)]_0 \wedge SL[\Psi(G)]_0$
$\mathsf{po}(\mathsf{HW}) = c_1$	$\mathsf{Em} \leq \mathsf{T} \wedge c_1$	$a \leq \mathsf{po}_b \rightarrow \mathsf{po}_a \leq \mathsf{po}_b$
$\mathsf{po}(\mathsf{LV}) = c_2$	$\mathsf{HW} \leq c_2,$	$aRb \rightarrow \mathsf{po}_a R\, \mathsf{po}_b \quad a, b \in \{\mathsf{HW}, \mathsf{LV}, \mathsf{H}, \mathsf{Em}\}$
$\mathsf{po}(\mathsf{H}) = c_3$	$\mathsf{LV} \leq \mathsf{V}$	
$\mathsf{po}(\mathsf{Em}) = c_4$	$\mathsf{V} \leq c_3$	$a \leq \mathsf{po}_b \rightarrow \mathsf{hl}_a \leq \mathsf{hl}_b$
$\mathsf{hl}(\mathsf{H}) = c_5$	$\mathsf{Endocarditis} \leq \mathsf{I} \wedge c_6$	$aRb \rightarrow \mathsf{hl}_a R\, \mathsf{hl}_b \quad a, b \in \{\mathsf{H}, \mathsf{Em}\}$
$\mathsf{hl}(\mathsf{Em}) = c_6$	$\mathsf{I} \leq \mathsf{D}$	
	$\mathsf{Heartdisease} = \mathsf{D} \wedge c_5$	$\mathsf{hl}_a, \mathsf{hl}_b$ constants renaming $\mathsf{hl}(a), \mathsf{hl}(b)$
	$\mathsf{Endocarditis} \not\leq \mathsf{Heartdisease}$	$\mathsf{po}_a, \mathsf{po}_b$ constants renaming $\mathsf{po}(a), \mathsf{po}(b)$
		$R \in \{\leq, \geq, =\}$
		$SL[\Psi(G)]_0$

The satisfiability of ϕ can therefore be checked automatically in polynomial time in the size of ϕ which in its turn is polynomial in the size of $\Psi_{\mathcal{K}}(G)$. Hence, in this case, the size of ϕ is polynomial in the size of G.

Unsatisfiability can also be proved directly: $G \wedge (RI_a \wedge \mathsf{Mon} \wedge \mathsf{Con})[\Psi(G)]$ entails the inequalities:

(1) $\mathsf{Endocarditis} \;\leq\; (\mathsf{I} \wedge \mathsf{hl}(\mathsf{Em})) \;\leq\; (\mathsf{D} \wedge \mathsf{hl}(\mathsf{Em}))$;
(2) $\mathsf{LV} \leq \mathsf{V} \leq \mathsf{po}(\mathsf{H})$;
(3) $\mathsf{Em} \leq \mathsf{po}(\mathsf{HW})$, hence also $\mathsf{hl}(\mathsf{Em}) \leq \mathsf{hl}(\mathsf{HW})$;
(4) $\mathsf{HW} \leq \mathsf{po}(\mathsf{LV})$, hence also $\mathsf{hl}(\mathsf{HW}) \leq \mathsf{hl}(\mathsf{LV})$;
(5) $\mathsf{LV} \leq \mathsf{po}(\mathsf{H})$, hence also $\mathsf{hl}(\mathsf{LV}) \leq \mathsf{hl}(\mathsf{H})$;
(6) $(\mathsf{D} \wedge \mathsf{hl}(\mathsf{Em})) \leq (\mathsf{D} \wedge \mathsf{hl}(\mathsf{H})) = \mathsf{Heartdisease}$

Thus, $G \wedge (RI_a \wedge \mathsf{Mon} \wedge \mathsf{Con})[\Psi(G)] \models \mathsf{Endocarditis} \leq \mathsf{Heartdisease}$, which leads to a contradiction, since $\mathsf{Endocarditis} \not\leq \mathsf{Heartdisease}$ is in G.

B Proof of Lemma 9

Lemma 9. The theory SLat of semilattices is P-interpolating for $P = \{\approx, \leq\}$.

Proof: This is a constructive proof based on the fact that $\mathsf{SLat} = ISP(S_2)$, where S_2 is the 2-element semilattice. We prove that the theory of semilattices is $\leq$-interpolating, i.e. that if A and B are two conjunctions of literals and $A \wedge B \models_{\mathsf{SLat}} a \leq b$, where a is a term containing only constants which occur in A and b a term containing only constants occurring in B, then there exists a term containing only common constants in A and B such that $A \models_{\mathsf{SLat}} a \leq t$ and $A \wedge B \models_{\mathsf{SLat}} t \leq b$. We can assume without loss of generality that A and B consist only of atoms: Indeed, assume that $A \wedge B = A_1 \wedge \dots \wedge A_n \wedge \neg A'_1 \wedge \dots \wedge \neg A'_m$, where $A_1, \dots, A_n, A'_1, \dots, A'_m$ are atoms. Then the following are equivalent:

- $A \wedge B \models_{\mathsf{SLat}} a \leq b$

– $\models_{\mathsf{SLat}} A \wedge B \rightarrow a \leq b$
– $\models_{\mathsf{SLat}} \neg A_1 \vee \cdots \vee \neg A_n \vee A'_1 \vee \cdots \vee A'_m \vee a \leq b$
– $\models_{\mathsf{SLat}} (A_1 \wedge \cdots \wedge A_n) \rightarrow A'_1 \vee \cdots \vee A'_m \vee a \leq b$
– $A_1 \wedge \cdots \wedge A_n \models_{\mathsf{SLat}} A'_1 \vee \cdots \vee A'_m \vee a \leq b$

Since the theory of semilattices is convex w.r.t. $\leq$ and $\approx$, it follows that if $A \wedge B \models_{\mathsf{SLat}} a \leq b$ then either (a) $A_1 \wedge \cdots \wedge A_n \models_{\mathsf{SLat}} A'_j$ for some $j \in \{1, \dots, m\}$ or (b) $A_1 \wedge \cdots \wedge A_n \models_{\mathsf{SLat}} a \leq b$. It is easy to see that in case (a), $A \wedge B \models \perp$. Then the conclusion of the theorem follows immediately. We therefore continue the proof for the case when A and B consist only of atoms.

As $\mathsf{SLat} = ISP(S_2)$, in SLat the same Horn sentences are true as in the 2-element semilattice S_2. Thus, $A \wedge B \models_{\mathsf{SLat}} a \leq b$ iff $A \wedge B \models_{S_2} a \leq b$, so we can reduce such a test to entailment in propositional logic.

It follows that $A \wedge B \models_{\mathsf{SLat}} a \leq b$ if and only if the following conjunction of literals in propositional logic is unsatisfiable:

(Ren($\wedge$))	$P_{e_1 \wedge e_2} \leftrightarrow P_{e_1} \wedge P_{e_2}$		$P_{g_1 \wedge g_2} \leftrightarrow P_{g_1} \wedge P_{g_2}$	
(P)	$P_{e_1} \leftrightarrow P_{e_2}$	$e_1 \approx e_2 \in A$	$P_{g_1} \leftrightarrow P_{g_2}$	$g_1 \approx g_2 \in B$
	$P_{e_1} \rightarrow P_{e_2}$	$e_1 \leq e_2 \in A$	$P_{g_1} \rightarrow P_{g_2}$	$g_1 \leq g_2 \in B$
(N)	P_a		$\neg P_b$	
	for all e, e_1, e_2 subterms in A		for all g, g_1, g_2 subterms in B	

We obtain an unsatisfiable set of clauses $(N_A \wedge P_a) \wedge (N_B \wedge \neg P_b) \models \perp$, where N_A and N_B are sets of Horn clauses in which each clause contains a positive literal. We can saturate $N_A \cup P_a$ under ordered resolution, in which all symbols occurring in A but not in B are larger than the common symbols. We show that if $A \wedge B \models_{\mathsf{SLat}} a \leq b$ holds, then for the term

$$t := \bigwedge \{e \mid A \models_{\mathsf{Slat}} a \leq e, e \text{ common subterm of } A \text{ and } B\}$$

the following hold:

(i) $A \models_{\mathsf{SLat}} a \leq t$, and
(ii) $A \wedge B \models_{\mathsf{SLat}} t \leq b$.

This means that for the theory of semilattices we have a property stronger than $\leq$-interpolability, but not quite as strong as strongly $\leq$-interpolability.
Every $e \in T = \{e \mid A \models_{\mathsf{SLat}} a \leq e, e \text{ common subterm of } A \text{ and } B\}$ corresponds to the positive unit clause P_e (where P_e is a propositional variable common to N_A and N_B) which can be derived from N_A using ordered resolution (with the ordering described above).

It is clearly the case that $A \models_{\mathsf{SLat}} a \leq t$, because $N_A \wedge P_a \wedge \neg P_t \wedge (P_t \leftrightarrow \bigwedge_{e \in T} P_e)$ is unsatisfiable. Thus, (i) holds.

For proving (ii), observe that by saturating $N_A \wedge P_a$ under ordered resolution we obtain the following kinds of clauses which can possibly lead to $\perp$ after inferences with $N_B \wedge \neg P_b$ (and thus to the consequence $a \leq b$ together with B):

(a) P_{e_k} positive unit clauses s.t. e_k contains symbols common to A and B, for $k \in \{1, \dots, l\}$.
(b) $\bigwedge_{j=1}^{n_i} P_{c_{ij}} \rightarrow P_{d_i}$, where c_{ij} and d_i are common symbols, such that for all i, j and k we have $c_{ij} \neq e_k$ and $d_i \neq e_k$, for $i \in \{1, \dots, p\}$.

Other types of clauses may appear too, but they can not be used to obtain $a \leq b$:

To see that clauses where some $c_{ij} = e_k$ are not necessary to derive the consequence $a \leq b$, note that if P_{e_k} is a positive unit literal and we have the clause $(P_{e_k} \wedge \bigwedge P_{c_{ij}}) \rightarrow P_{d_i}$, then by resolution we get as an inference $\bigwedge P_{c_{ij}} \rightarrow P_{d_i}$. It is easy to see that $(P_{e_k} \wedge \bigwedge P_{c_{ij}}) \rightarrow P_{d_i}$ is redundant in the presence of $\bigwedge P_{c_{ij}} \rightarrow P_{d_i}$. In the same way, clauses of the form $\bigwedge P_{c_{ij}} \rightarrow P_{e_k}$ (i.e. clauses of type (b) where $d_i = e_k$) are redundant in the presence of $P_{e_1}, \dots, P_{e_l}$. For the proof of (ii) one needs to consider separately the case in which none of the P_{d_i} is needed to derive $\perp$ together with N_B (and thus the consequence $a \leq b$) and the case when some P_{d_i} are needed.

Case 1: None of the P_{d_i} is needed to derive $\perp$ together with N_B (and thus the consequence $a \leq b$). We know that $N_A \models P_a \rightarrow \bigwedge_{k=1}^{l} P_{e_k}$. From this it follows that $A \models a \leq \bigwedge_{k=1}^{l} e_k$.
For $A \wedge B \models a \leq b$ to be true, $\bigwedge_{k=1}^{l} P_{e_k} \wedge N_B \wedge \neg P_b$ must be unsatisfiable, so there has to be a subset $S \subseteq \{1, ..., l\}$ such that $\bigwedge_{k \in S} P_{e_k} \wedge N_B \wedge \neg P_b$. This means that $B \models \bigwedge_{s \in S} e_s \leq b$. But then, since $\bigwedge_{k=1}^{l} e_k \leq \bigwedge_{s \in S} e_s$, it follows that $B \models \bigwedge_{k=1}^{l} e_k \leq b$, and therefore also $A \wedge B \models \bigwedge_{k=1}^{l} e_k \leq b$.

Case 2: Some P_{d_i} are needed to derive $\perp$ from $N_B \wedge \neg P_b$. Again, we know that $N_A \models P_a \rightarrow \bigwedge_{k=1}^{l} P_{e_k}$ (hence $A \models a \leq \bigwedge_{k=1}^{l} e_k$).
For $A \wedge B \models a \leq b$ to be true, i.e. $(N_A \wedge P_a) \wedge (N_B \wedge \neg P_b)$ to be unsatisfiable, there have to be subsets $S_1 \subseteq \{1, ..., l\}$ and $S_2 \subseteq \{1, ..., p\}$ such that $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \bigwedge_{i \in S_2}((\bigwedge_j P_{c_{ij}}) \rightarrow P_{d_i}) \wedge \neg P_b$ is unsatisfiable. Let $N_{AB} := N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \bigwedge_{i \in S_2}((\bigwedge_j P_{c_{ij}}) \rightarrow P_{d_i})$. We know that $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b$ is satisfiable. Assume that there is no c_{ij} such that $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b \models P_{c_{ij}}$. Then for every c_{ij}, $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b \wedge \neg P_{c_{ij}}$ is satisfiable. Since all clauses in $N_b \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b$ are Horn clauses, it follows that $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b \wedge \bigwedge_{i,j} \neg P_{c_{ij}}$ is satisfiable. Every model of $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b \wedge \bigwedge_{i,j} \neg P_{c_{ij}}$ is a model of $N_{AB} \wedge \neg P_b$. It would therefore follow that $N_{AB} \wedge \neg P_b$ is satisfiable, which is a contradiction. Thus, there exists at least one c_{ij} such that $N_B \wedge \bigwedge_{k \in S_1} P_{e_k} \wedge \neg P_b \models P_{c_{ij}}$. We can add $P_{c_{ij}}$ to this set of clauses and repeat the reasoning for the set of clauses obtained this way as long as we still have one clause of the form $((\bigwedge_j P_{c_{ij}}) \rightarrow P_{d_i})$ in N_{AB} such that there exists at least one c_{ij} such that $P_{c_{ij}}$ was not added to N_{AB}.

Then there has to be a sequence $(d_{i_1 j})_{j \in J_1}, (d_{i_2 j})_{j \in J_2}, ..., (d_{i_n j})_{j \in J_n}$ such that:

- $P_{d_{i_1 j}}$ can be derived from $N_{AB} \wedge \bigwedge P_{e_k}$, for all $j \in J_1$,
- $P_{d_{i_2 j}}$ can be derived from $N_{AB} \wedge \bigwedge P_{e_k} \wedge \bigwedge_{k \in J_1} d_{i_1 k}$, for all $j \in J_2$,
- $P_{d_{i_3 j}}$ can be derived from $N_{AB} \wedge \bigwedge P_{e_k} \wedge \bigwedge_{k \in J_1} d_{i_1 k} \wedge \bigwedge_{k \in J_2} d_{i_2 k}$, for all $j \in J_3$, ...

– $P_{d_{i_n j}}$ can be derived from $N_{AB} \wedge \bigwedge_k P_{e_k} \wedge \bigwedge_{k \in J_1} d_{i_1 k} \wedge \cdots \bigwedge_{k \in J_{n-1}} d_{i_{n-1} k}$, for all $j \in J_n$,
– P_b can be derived from $N_{AB} \wedge \bigwedge P_{e_k} \wedge \bigwedge_{k \in J_1} d_{i_1 k} \wedge \ldots \wedge \bigwedge_{k \in J_n} d_{i_n k}$.

But then $A \wedge B \models \bigwedge e_k \leq d_{i_1 l}$, for all $l \in J_1$, hence $A \wedge B \models \bigwedge e_k \leq \bigwedge_{l \in J_1} d_{i_1 l}$, hence $A \wedge B \models \bigwedge e_k \wedge \bigwedge_{l \in J_1} d_{i_1 l} \approx \bigwedge e_k$. Therefore, as $A \wedge B \models (\bigwedge e_k \wedge \bigwedge_{l \in J_1} d_{i_1 l}) \leq d_{i_2 j}$, for all $j \in J_2$, we have $A \wedge B \models \bigwedge e_k \leq \bigwedge_j d_{i_2 j}$. Similarly it can be proved that $A \wedge B \models \bigwedge e_k \leq \bigwedge_j d_{i_n j}$, and finally that $A \wedge B \models \bigwedge e_k \leq b$. $\square$

C A separation result

Proposition 13. *Let $\mathcal{T}_0$ be a theory with signature $\Pi_0 = (\Sigma_0, \mathsf{Pred})$. Assume that $\leq \in \mathsf{Pred}$ is such that $\leq$ is a transitive relation in all models of $\mathcal{T}_0$ and that $\mathcal{T}_0$ is convex with respect to $\leq$ and $\leq$-interpolating.*

Let A_0 and B_0 be conjunctions of ground literals in the signature Π_0^c (the extension of Π_0 with a set of constants) such that $A_0 \wedge B_0 \wedge \mathcal{H} \models_{\mathcal{T}_0} a \leq b$, where a contains only symbols occurring in A_0 and b only symbols occurring in B_0 and $\mathcal{H}$ is a set of Horn clauses of the form $c_1 \leq d_1 \rightarrow c \leq d$ in the signature Π_0^c which are instances of flattened and purified clauses of the form $\mathsf{Mon}[A \wedge B]_0 \wedge RI_a[A \wedge B]_0$ as explained in Theorem 11, i.e. of axioms of the following type:

$$\begin{cases} x \leq g(y) \rightarrow f(x) \leq h(y) \\ \quad x \leq y \rightarrow f(x) \leq f(y) \end{cases} \tag{4}$$

Then the following hold:

(1) There exists a set T of Π_0^c-terms containing only constants common to A_0 and B_0 and a term $t \in T$ such that

$$A_0 \wedge B_0 \wedge (\mathcal{H} \backslash \mathcal{H}_{\mathsf{mix}}) \wedge \mathcal{H}_{\mathsf{sep}} \models_{\mathcal{T}_0} a \leq t \wedge t \leq b,$$

where

$$\begin{aligned} \mathcal{H}_{\mathsf{mix}} = &\{a_1 \leq b_1 \rightarrow a_2 \leq b_2 \in \mathcal{H} \mid a_1, a_2 \text{ constants in } A, b_1, b_2 \text{ constants in } B\} \cup \\ &\{b_1 \leq a_1 \rightarrow b_2 \leq a_2 \in \mathcal{H} \mid b_1, b_2 \text{ constants in } B, a_1, a_2 \text{ constants in } A\} \\ \mathcal{H}_{\mathsf{sep}} = &\{(a_1 \leq t_1 \rightarrow a_2 \leq c_{f(t_1)}) \wedge (t_1 \leq b_1 \rightarrow c_{f(t_1)} \leq b_2) \mid a_1 \leq b_1 \rightarrow a_2 \leq b_2 \in \mathcal{H}_{\mathsf{mix}}, \\ &\quad b_1 \approx g(e_1),\ \ b_2 \approx h(e_1) \in \mathsf{Def}_B,\ \ a_2 \approx f(a_1) \in \mathsf{Def}_A \text{ or vice versa}, \\ &\quad \text{and } t_1, f(t_1) \in T\} = \mathcal{H}^A_{\mathsf{sep}} \wedge \mathcal{H}^B_{\mathsf{sep}} \end{aligned}$$

where $c_{f(t_1)}$ are new constants in Σ_c (considered to be common) introduced for the corresponding terms $f(t_1)$.

(2) $A_0 \wedge B_0 \wedge (\mathcal{H} \backslash \mathcal{H}_{\mathsf{mix}}) \wedge \mathcal{H}_{\mathsf{sep}} \wedge \neg(a \leq t \wedge t \leq b)$ is logically equivalent w.r.t. $\mathcal{T}_0$ with the separated conjunction of literals $\overline{A}_0 \wedge \overline{B}_0 \wedge \neg(a \leq t \wedge t \leq b) = A_0 \wedge B_0 \wedge \bigwedge\{c \leq d \mid \Gamma \rightarrow c \leq d \in \mathcal{H} \backslash \mathcal{H}_{\mathsf{mix}}\} \wedge \bigwedge\{c \leq c_{f(t)} \wedge c_{f(t)} \leq d \mid (\Gamma \rightarrow c \leq c_{f(t)}) \wedge (\Gamma \rightarrow c_{f(t)} \leq d) \in \mathcal{H}_{\mathsf{sep}}\} \wedge \neg(a \leq t \wedge t \leq b)$.

Proof: We prove (1) and (2) by induction on the number of clauses in $\mathcal{H}$.

If $\mathcal{H} = \emptyset$ then the initial problem is already separated into an A and a B part so we are done: We have $A_0 \wedge B_0 \models_{\mathcal{T}_0} a \leq b$ and since we assumed that $\mathcal{T}_0$ is $\leq$-interpolating, there exists a term t containing only constants common to A_0 and B_0 such that $A_0 \wedge B_0 \models_{\mathcal{T}_0} a \leq t \wedge t \leq b$ (we can choose $T = \{t\}$).

Assume that $\mathcal{H}$ contains at least one clause, and that for every $\mathcal{H}_1$ with fewer clauses and every conjunction of literals A_0', B_0' with $A_0' \wedge B_0' \wedge \mathcal{H}_1 \models_{\mathcal{T}_0} a \leq b$, (1) and (2) hold.

Let $\mathcal{D}$ be the set of all atoms $c \leq d$ occurring in premises of clauses in $\mathcal{H}$. As every model of $A_0 \wedge B_0 \wedge \bigwedge_{(c \leq d) \in \mathcal{D}} \neg(c \leq d) \wedge \neg(a \leq b)$ is also a model for $\mathcal{H} \wedge A_0 \wedge B_0 \wedge \neg(a \leq b)$ and $\mathcal{H} \wedge A_0 \wedge B_0 \wedge \neg(a \leq b) \models_{\mathcal{T}_0} \perp$, $A_0 \wedge B_0 \wedge \bigwedge_{(c \leq d) \in \mathcal{D}} \neg(c \leq d) \wedge \neg(a \leq b) \models_{\mathcal{T}_0} \perp$. Let $(A_0 \wedge B_0)^+$ be the conjunction of all positive literals in $A_0 \wedge B_0$, and $(A_0 \wedge B_0)^-$ be the set of all negative literals in $A_0 \wedge B_0$. Then

$$(A_0 \wedge B_0)^+ \models_{\mathcal{T}_0} \bigvee_{(c \leq d) \in \mathcal{D}} (c \leq d) \vee \bigvee_{\neg L \in (A_0 \wedge B_0)^-} L \vee (a \leq b).$$

$\mathcal{T}_0$ is convex with respect to $\leq$ and $(A_0 \wedge B_0)^+$ is a conjunction of positive literals. Therefore, either

(i) $(A_0 \wedge B_0)^+ \models L$ for some $L \in (A_0 \wedge B_0)^-$ (then $A_0 \wedge B_0$ is unsatisfiable and hence entails any atom $c_i \leq d_i$), or
(ii) $(A_0 \wedge B_0)^+ \models a \leq b$, or
(iii) there exists $(c_1 \leq d_1) \in \mathcal{D}$ such that $A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq d_1$.

Case 1: $A_0 \wedge B_0$ is unsatisfiable. In this case (1) and (2) hold for $T = \{t\}$, where t is an arbitrary term over the common symbols of A_0 and B_0.

Case 2: $A_0 \wedge B_0$ is satisfiable and $(A_0 \wedge B_0)^+ \models a \leq b$. Then we can use the fact that $\mathcal{T}_0$ is $\leq$-interpolating and we are done.

Case 3: $A_0 \wedge B_0$ is satisfiable and there exists $(c_1 \leq d_1) \in \mathcal{D}$ such that $A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq d_1$. Then $A_0 \wedge B_0$ is logically equivalent in $\mathcal{T}_0$ with $A_0 \wedge B_0 \wedge c_1 \leq d_1$.

Let $C = c_1 \leq d_1 \rightarrow c \leq d \in \mathcal{H}$ such that $A_0 \wedge B_0 \models c_1 \leq d_1$.

Case 3a. Assume that C contains only constants occurring in A or only constants occurring in B. Then $A_0 \wedge B_0 \wedge \mathcal{H}$ is equivalent w.r.t. $\mathcal{T}_0$ with $A_0 \wedge B_0 \wedge (\mathcal{H} \backslash C) \wedge c \leq d$. By the induction hypothesis for $A_0' \wedge B_0' = A_0 \wedge B_0 \wedge c \leq d$ and $\mathcal{H}_1 = \mathcal{H} \backslash \{C\}$, we know that there exists T' and $t \in T'$ such that $A_0' \wedge B_0' \wedge (\mathcal{H}_1 \backslash \mathcal{H}_{1\mathsf{mix}}) \wedge \mathcal{H}_{1\mathsf{sep}} \models a \leq t \wedge t \leq b$, and (2) holds too.

Then, for $T = T'$, $A_0' \wedge B_0' \wedge (\mathcal{H}_1 \backslash \mathcal{H}_{1\mathsf{mix}}) \wedge \mathcal{H}_{1\mathsf{sep}} \wedge \neg(a \leq t \wedge t \leq b)$ is logically equivalent to $A_0 \wedge B_0 \wedge (\mathcal{H} \backslash \mathcal{H}_{\mathsf{mix}}) \wedge \mathcal{H}_{\mathsf{sep}} \wedge \neg(a \leq t \wedge t \leq b)$, so $A_0 \wedge B_0 \wedge (\mathcal{H} \backslash \mathcal{H}_{\mathsf{mix}}) \wedge \mathcal{H}_{\mathsf{sep}} \models (a \leq t \wedge t \leq b)$, i.e. (1) holds.

In order to prove (2), note that, by definition, $\mathcal{H}_{1\mathsf{mix}} = \mathcal{H}_{\mathsf{mix}}$ and and $\mathcal{H}_{1\mathsf{sep}} = \mathcal{H}_{\mathsf{sep}}$. By the induction hypothesis, $A_0' \wedge B_0' \wedge (\mathcal{H}_1 \backslash \mathcal{H}_{1\mathsf{mix}}) \cup \mathcal{H}_{1\mathsf{sep}} \wedge \neg(a \leq t \wedge t \leq b)$ is logically equivalent to a corresponding conjunction $\overline{A}_0' \wedge \overline{B}_0' \wedge \neg(a \leq t \wedge t \leq b)$ containing as conjuncts all literals in A_0' and B_0' and all conclusions of rules in

$\mathcal{H}_1 \backslash \mathcal{H}_{1\mathsf{mix}}$ and $\mathcal{H}_{1\mathsf{sep}}$. On the other hand, $A'_0 \wedge B'_0 \wedge \neg(a \leq t \wedge t \leq b)$ is logically equivalent to $A_0 \wedge B_0 \wedge (c \leq d) \wedge \neg(a \leq t \wedge t \leq b)$, where $(c \leq d)$ is the conclusion of the rule $C \in \mathcal{H} \backslash \mathcal{H}_{\mathsf{mix}}$. This proves (2).

Case 3b. Assume now that $C := c_1 \leq d_1 \rightarrow c \leq d$ is mixed, for instance that c_1, c are constants in A and d_1, d are constants in B.

(a) Assume that C is obtained from an instance of a clause of the form $x \leq g(y) \rightarrow f(x) \leq h(y)$. This means that there exist $c \approx f(c_1) \in \mathsf{Def}_A$ and $d_1 \approx g(e), d \approx h(e) \in \mathsf{Def}_B$. We know that $A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq d_1$ and that $\mathcal{T}_0$ is $\leq$-interpolating. Thus, there exists a term t_1 containing only constants common to A_0 and B_0 such that

$$A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq t_1 \wedge t_1 \leq d_1. \tag{5}$$

Let $c_{f(t_1)}$ be a new constant, denoting the term $f(t_1)$, and let

$$C_A = c_1 \leq t_1 \rightarrow c \leq c_{f(t_1)} \quad \text{and} \quad C_B = t_1 \leq d_1 \rightarrow c_{f(t_1)} \leq d.$$

Thus, C_A corresponds to the instance $c_1 \leq t_1 \rightarrow f(c_1) \leq f(t_1)$ of the monotonicity axiom for f, whereas C_B corresponds to the rule $t_1 \leq g(e) \rightarrow f(t_1) \leq h(e)$.

As $A_0 \wedge B_0 \models c_1 \leq t_1 \wedge t_1 \leq c_1$ and as $\leq$ is transitive, by (5):

$$\begin{aligned} A_0 \wedge B_0 \wedge C_A \wedge C_B &\models\!\!\!\!\!\dashv_{\mathcal{T}_0} A_0 \wedge B_0 \wedge (c_1 \leq t_1 \wedge C_A) \wedge (t_1 \leq d_1 \wedge C_B) \\ &\models\!\!\!\!\!\dashv_{\mathcal{T}_0} A_0 \wedge B_0 \wedge c \leq c_{f(t_1)} \wedge c_{f(t_1)} \leq d \\ &\models_{\mathcal{T}_0} A_0 \wedge B_0 \wedge c \leq d, \end{aligned}$$

(where $\models\!\!\!\!\!\dashv_{\mathcal{T}_0}$ stands for logical equivalence w.r.t. $\mathcal{T}_0$.)

Hence, $A_0 \wedge B_0 \wedge C_A \wedge C_B \wedge (\mathcal{H} \backslash C) \models_{\mathcal{T}_0} A_0 \wedge B_0 \wedge c \leq d \wedge (\mathcal{H} \backslash C)$. On the other hand, since $A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq d_1$ it follows that $A_0 \wedge B_0 \wedge \mathcal{H}$ is logically equivalent with $A_0 \wedge B_0 \wedge c \leq d \wedge (\mathcal{H} \backslash C)$, so $A_0 \wedge B_0 \wedge C_A \wedge C_B \wedge (\mathcal{H} \backslash C) \wedge \neg(a \leq b) \models_{\mathcal{T}_0} \perp$.

By the induction hypothesis for $A_0 \wedge B_0 \wedge c \leq c_{f(t_1)} \wedge c_{f(t_1)} \leq d$ and $\mathcal{H}_1 = \mathcal{H} \backslash C$ we know that there exists a set T' of terms such that $A_0 \wedge B_0 \wedge c \leq c_{f(t_1)} \wedge c_{f(t_1)} \leq d \wedge (\mathcal{H}_1 \backslash \mathcal{H}_{1\mathsf{mix}}) \wedge \mathcal{H}_{1\mathsf{sep}} \wedge \neg(a \leq t \wedge t \leq b) \models \perp$, and also (2) holds. Then (1) holds for $T = T' \cup \{f(t_1), t_1\}$.

(b) Assume that C corresponds to an instance of a monotonicity axiom $x \leq y \rightarrow f(x) \leq f(y)$. This means that there exist $c \approx f(c_1) \in \mathsf{Def}_A$ and $d \approx f(d_1) \in \mathsf{Def}_B$. We know that $A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq d_1$ and that $\mathcal{T}_0$ is $\leq$-interpolating. Thus, there exists a term t_1 containing only constants common to A_0 and B_0 such that

$$A_0 \wedge B_0 \models_{\mathcal{T}_0} c_1 \leq t_1 \wedge t_1 \leq d_1. \tag{6}$$

Let $c_{f(t_1)}$ be a new constant, denoting the term $f(t_1)$, and let

$$C_A = c_1 \leq t_1 \rightarrow c \leq c_{f(t_1)} \quad \text{and} \quad C_B = t_1 \leq d_1 \rightarrow c_{f(t_1)} \leq d.$$

Thus, C_A corresponds to the instance $c_1 \leq t_1 \rightarrow f(c_1) \leq f(t_1)$ of the monotonicity axiom for f, whereas C_B corresponds to the instance $t_1 \leq d_1 \rightarrow f(t_1) \leq f(d_1)$ of the monotonicity axiom for f. The proof can then continue as the proof of case (a); also in this case we can choose $T = T' \cup \{f(t_1), t_1\}$.

(2) can be proved similarly using the induction hypothesis. $\square$